AIST 2019

1st International Conference
on

Artificial Intelligence and Speech Technology

November 14-15, 2019

1st International Conference on Artificial Intelligence and Speech Technology

November 14-15, 2019

Editors
Dr. (Mrs). Amita Dev
Dr. S.S. Agrawal
Dr. Arun Sharma

Supported by

Department of Science & Technology, Government of India

सत्यमेव जयते

ANVESHAN FOUNDATION

Organized by

Indira Gandhi Delhi Technical University for Women
Kashmere Gate, Delhi
www.igdtuw.ac.in | www.aist2019.com

EXCEL INDIA PUBLISHERS
NEW DELHI

First Impression: November 2019

International Conference on Artificial Intelligence and Speech Technology (AIST2019)

ISBN: 978-93-88237-94-9

Published by
EXCEL INDIA PUBLISHERS
91 A, Ground Floor
Pratik Market, Munirka, New Delhi–110 067
Tel: +91-11-2671 1755/ 2755/ 3755/ 5755
Cell: 9899127755, 9999609755, 9910757755
Fax: +91-11-2671 6755
E-mail: publishing@groupexcelindia.com
Web: www.groupexcelindia.com

Typeset by
Excel Prepress Services, New Delhi–110 067
E-mail: production@groupexcelindia.com

Printed by
Excel Printing Universe, New Delhi–110 067
E-mail: printing@groupexcelindia.com

Message

It gives me immense pleasure to welcome all the participants, delegates, keynote speakers, resource persons for the tutorials to the First International Conference on "Artificial Intelligence and Speech Technology" AIST-2019 to be held at Indira Gandhi Delhi Technical University for Women, Kashmere Gate, Delhi on 14^{th} and 15^{th} November 2019.

The aim of the Conference is to serve as a forum for discussions on the state-of-the-art research, development and implementations of Artificial Intelligence and Speech Technology. AIST-2019 is dedicated to cutting edge research that addresses scientific needs of academic researchers and industrial professionals to explore new horizons of knowledge related to Artificial Intelligence and Speech Technology. Researchers from across the world are presenting their research revealing latest and relevant research findings on almost all the aspects of Artificial Intelligence and Speech Technology.

Keynote Addresses on Artificial Intelligence and Speech Technology will be delivered by eminent researchers and academicians from India and abroad. In parallel, tutorials on Deep Learning and Speech Technology will also be conducted during the conference. Hands on Sessions along with strong conceptual knowledge will be conducted by senior academicians, researchers and practitioners working on these technologies.

As academicians, the responsibility to nurture complete professionals lies with us. This necessitates the knowledge of latest trends in fast changing technology. Conferences bring together people from all different geographical areas who share a common discipline or field and is found effective to extend one's knowledge.

I on behalf of the Steering Committee, would like to express my sincere thanks and appreciation to the world-renowned Professors and prominent Researchers for having agreed to deliver the keynote session and share their knowledge during the Conference.

I am sure that this colloquy of researchers and experts from academia and industry would greatly benefit researchers, students and faculty. Young scientists and researchers will find the contents of the proceedings helpful to set roadmaps for their future endeavours.

I wish the conference a great success.

Dr. (Mrs.) Amita Dev

Vice-Chancellor, IGDTUW

General Chair, AIST-2019

Message

It is the matter of great pleasure and happiness to see that Indira Gandhi Delhi Technical University for Women, Kashmere Gate, Delhi is organizing its First International Conference on Artificial Intelligence and Speech Technology (AIST-2019). The objective of the conference is to provide a platform for a profound discussion and presentations on state-of-the-art research, development, innovations and implementations of Artificial Intelligence and Speech Technology by the researchers world-wide.

There has been a tremendous advancement and innovations in Artificial Intelligence which is incomparable to what Artificial Intelligence emerged traditionally. We use Artificial Intelligence many times during a day-often, without even realizing it. Today Artificial Intelligence has greatly enhanced machine learning, Natural Language Processing (NLP) and Deep learning such that they are enabling new developments in Speech Technology like voice response user interactive systems. Looking to its huge hope and dimensions AIST-2019 brings together academics, industrial experts and education leaders from all over the world to discuss an incredibly wide array of topics ranging from Foundation of Artificial Intelligence and machine learning, data mining, Cognitive science to Speech technology, to name a few.

I would like to express my sincere thanks and appreciation to the world-renowned Professors and prominent Researchers for having agreed to deliver the keynote session and share their knowledge during the Conference. My warmest thanks go to the organizing committee colleagues including the tutorial program co-chairs, the technical program committee members, the paper reviewers for their invaluable work in shaping the technical program and not the least all the authors who kindly submitted their papers to AIST-2019.

In summary, no doubt you all will appreciate the unique combination of cutting-edge technical program, with wonderful organization of the conference, Enjoy meetings with friends and colleagues as well as impromptu discussions with eminent speakers. I look forward to seeing everyone in IGDTUW, Delhi India.

Prof. S.S. Agrawal

Emeritus Scientist, CSIR

General Co-Chair, AIST-2019

Message

I take this opportunity to welcome you all to the International Conference on Artificial Intelligence and Speech Technology i.e. AIST-2019, to be held at Indira Gandhi Delhi Technical University for Women, Kashmere Gate, Delhi during 14-15th November 2019. This conference will have an amalgam of researchers from the fields of Artificial Intelligence and Speech Technology.

The objective of the conference is to provide a forum for researchers worldwide to unveil their latest work in Artificial Intelligence and innovations in Speech Technology. Topics covered in this conference include fundamentals of AI, its tools and applications, Machine Learning, Deep Learning, Soft Computing and Applications, Speech Analysis, Representation and Models, Spoken Language Recognition and Understanding, Affective Speech Recognition, Interpretation and Synthesis, Speech Interface Design and Human Factors Engineering, Speech Emotion Recognition Technologies, Audio-Visual Speech Processing, IoT Security. The conference received more than 90 submissions from all over the globe, out of which the best 39 selected papers will be presented during these two days. Apart than Conference Proceedings, extended version of the selected papers will also be published in reputed Journals by publishers including Springer, Taylor and Francis, IGI Global and others.

AIST-2019 is a first effort of IGDTUW to share knowledge and current research on Artificial Intelligence and its innovation in Speech technology. All the paper submissions have gone through a careful anonymous review process (2 or more reviewers per submission) aided by Technical Program Committee members and Advisory Board. The AIST-2019 Conference includes prominent Keynote addresses by Prof. Satoshi Nakamura (NAIST, Japan), Prof. Németh Géza (Budapest University of Technology and Economics, Hungary), Dr. Barbara Zitova (Institute of Information Theory and Automation of the ASCR, Czech Republic), Prof. S.S. Agrawal (Emeritus Scientist, CSIR) and Dr.(Mrs.) Amita Dev (Vice-Chancellor, IGDTUW Delhi India). It also includes sessions from eminent researchers including Prof. Samudravijaya (IITG), Prof. Omar Farooq (AMU), Prof. Deepak Garg (Bennett University) and Prof, Vipin Tyagi (Jaypee Univ., Guna).

I would like to thank everyone who has given his or her time, energy and ideas to assist in organizing this event including all members of organizing committee, Technical Program Committee members and all reviewers and our distinguished keynote speakers who have agreed to address the conference attendees. I also wish to thank all of our sponsors and supporters especially DST (Curie Grant) who have made this event possible. It is through the collective efforts of these individuals and organizations that we are able to bring this conference a great event!.

Looking for the great success of the Conference.

Dr. Arun Sharma

Convener, AIST-2019

Keynote Speakers

Prof. Satoshi Nakamura

Biography: Satoshi Nakamura (F'16) received the B.S. degree from the Kyoto Institute of Technology, in 1981, and the Ph.D. degree from Kyoto University, in 1992. He was an Associate Professor with the Graduate School of Information Science, Nara Institute of Science and Technology, from 1994 to 2000. He was the Director of ATR Spoken Language Communication Research Laboratories, from 2000 to 2008, and the Vice President of ATR, from 2007 to 2008. He was the Director General of Keihanna Research Laboratories and the Executive Director of the Knowledge Creating Communication Research Center, National Institute of Information and Communications Technology, Japan, from 2009 to 2010. He is currently a Professor with the Graduate School of Science and Technology, Nara Institute of Science and Technology, Japan, the Project Leader of the Tourism Information Analytics Team, RIKEN Center for Advanced Intelligence Project AIP, an Honorar professor with the Karlsruhe Institute of Technology, Germany, and an ATR Fellow. He is also the Director of the Augmented Human Communication Laboratory and a Full Professor with the Graduate School of Information Science, Nara Institute of Science and Technology. He is interested in modelling and systems of speech-to-speech translation and speech recognition. He is one of the leaders of speech-to-speech translation research. He has been serving for various speech-to-speech translation research projects in the world, including C-STAR, IWSLT, and A-STAR. He received the Yamashita Research Award, the Kiyasu Award from the Information Processing Society of Japan, the Telecom System Award, the AAMT Nagao Award, the Docomo Mobile Science Award, in 2007, the ASJ Award for Distinguished Achievements in Acoustics, and the LREC Antonio Zampolli Award, in 2012. He received the Commendation for Science and Technology by the Minister of Education, Science and Technology, and the Commendation for Science and Technology by the Minister of Internal Affair and Communications. He has been an elected Board Member of the International Speech Communication Association (ISCA), since 2011, an Editorial Board Member of the IEEE Signal Processing Magazine, since 2012, and a member of the IEEE SPS Speech and Language Technical Committee, since 2013.

Prof. Nemeth Geza

Biography: Prof. Nemeth Geza is currently the professor with Faculty of Electrical Engineering and Informatics, department of Telecommunications and Media Informatics in Budapest University of Technology and Economics. He has been awarded 2013 Master Tutor Gold Medal, 2012 Pushkas Tivadar Award, 2011 Publication Award of the Hungarian Academy of Sciences, 2004 Officer's Cross of the Order of Merit of the Hungarian Republic, 2002 and 2015 Pollak-Virag Award, 2000 Kempelen Farkas Award, 1999 Award of the Hungarian Academy of Sciences. His teaching and research area includes Speech Information Systems, multimodal and mobile human-machine interfaces extends from basic research, through technology development to applications embedded in Information society.

Dr. Barbara Zitova

Biography: Dr. Barbara Zitova received the M.Sc. degree in computer science from the Charles University, Prague, Czech Republic in 1995 and the Ph.D. degree in computer science from the Charles University, Prague, Czech Republic in 2000. Since 1995 she has been with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague. She was appointed the head of the department on the 1st January 2008. She gives tutorials on Image Processing and Pattern Recognition at the Czech Technical University. Jointly with J. Flusser she gives specialized graduate course on moment invariants and wavelets at the Charles University and at the Czech Technical University. Barbara Zitova has more than an 15-years of experience in digital image processing. She is an author of a book chapter in Invariants for Pattern Recognition and Classification (M.A. Rodrigues ed., World Scientific, 2000), co-author of a monograpgh "Flusser Jan, Suk Tomáš, Zitová Barbara : Moments and Moment Invariants in Pattern Recognition, Wiley & sons, (Chichester 2009)", and of 50 journal and conference papers on moment invariants and related topics. Her paper "Image Registration Methods: A Survey", Image and Vision Computing, vol. 21, pp. 977-1000, 2003, has became a major reference in image registration with more than 2000 citations.

Prof. S.S. Agrawal

Biography: Dr. Shyam Sunder Agrawal is a World-renowned scientist in the area of Acoustic Speech and Communication. He obtained Ph.D. in 1970 from Aligarh Muslim University, India. Having research experience of about 45 years first as a Scientist at the Central Electronics Engineering Research Institute Pilani and subsequently as Emeritus Scientist of CSIR and then as an Advisor to Centre for Development of Advance Computing (CDAC), Noida. He has worked as Guest Researcher at MIT, Cambridge and UCLA in U.S.A. and visited many universities and research institutions in USA, UK, France, Germany, Sweden, Italy, Japan and other Asian countries. His major areas of interest are Speech Perception, Speech Synthesis, Speech/ Speaker recognition and Development of Speech Data Bases in Indian Languages.

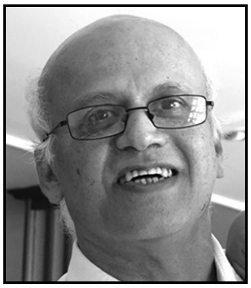
Dr Agrawal has published/presented about 200 papers in National / International Journals and Conferences. Received large no. of Hons. and Awards. Some of these include Sir C.V. Raman Award by Acoustical Society of India, Internationally Eminent Acoustician Award and Gold Medal from Acoustic Foundation and Acoustic Society of America, Honored to deliver Several memorial lectures such as B.D. Chaudhuri Memorial lecture, M. S. Narayanan Memorial lecture, Rais Ahmed Memorial lecture, Best project Awards by CEERI, Achut Menon Award and many other citations. He was awarded Senior Fulbright Fellowship, U.N.D.P Fellowships and Sr. East-West Centre Fellowship. He is a Distinguished Fellow of IETE, and fellow of CSI, ASI, ISPhS and other professional societies. He has been President of the Acoustical Society of India and the Board Member of International Commission on Acoustics, Vice President of ISPhS (USA), Vice-President and Chairman of BOE,ICC, ADEC and ELAN of IETE. He is also Country representative of O-COCOSDA, U-Star consortium & member of Indo-Japan Forum of NICT. He is currently working as Director General of KIIT Group of Institution and steering their Academic & Research programs in Institutions of Higher Learning and engaged in several national and international collaborative research projects and Advisor to various National and International Institutions.

Dr. (Mrs) Amita Dev

Biography: Dr. Amita Dev is the First Pro-Vice Chancellor and Second Vice-Chancellor of Indira Gandhi Delhi Technical University for Women. Dr. (Mrs.) Amita Dev, has obtained her B.Tech degree from Punjab University and completed her post-graduation from BITS, Pilani, India. She had obtained her Ph.D degree from Delhi College of Engineering under University of Delhi in the area of Computer Science. She has more than 33 years of experience in Teaching, Research and Industry. She has excellent track record of Administration quality Teaching, Innovation and Research. She has published 63 + research papers in Renowned International and National Journals and Conference Proceedings. Her teaching and research areas include Artificial Neural Networks, Speech Processing, Opportunistic Networks, Speech Recognition Systems, MANETS, Advanced Computer Networks, Data Mining etc. In recognition to her valuable and worthy research contributions, she has been conferred with many Awards like "National Level AICTE Young Teacher Career Award" in the year 1998 with Research Grant for pursuing Advanced Research in the area of "Hindi Speech Recognition Using Connectionist Model", "State Level Best Teacher Award" in the year 1998 awarded by Govt. of NCT of Delhi, "National Level Best Engineering Teacher Award" in the year 2001 by ISTE for significant contribution in the field of Research and Technology, "Raja Ram Babu Patil National Level Award" in the year 2014 by Indian Society for Technical Education, "National Level ECONS Education Excellence Award" in the year 2015 for significant contribution for Education and Academic Excellence, "Vittiya Saksharta Abhiyan Award" on 08th March 2017 by the Hon'ble Minister for HRD, Govt. of India for spreading Digital Literacy and numerous Appreciation letters have been received from Senior Officials of Delhi Government for the Exemplary Performance, Dedication and Devotion towards the work/duty and for Outstanding Contribution in the field of Technical Education.

Dr. Omar Farooq

Omar Farooq joined Department of Electronics Engineering, AMU Aligarh as Lecturer in 1992 and is currently working as a Professor. He was awarded Commonwealth Scholar from 1999-2002 towards PhD at Loughborough University, UK and one year UKIERI postdoctoral fellowship in 2007-08. His broad area of research interest is signal processing with specialization in speech recognition. Currently he is working in the area of biomedical signal processing on three collaborative research projects. He has authored/co-authored over 220 papers in refereed academic journals & conference proceedings and steered 7 researchers to PhD graduation. He is a Fellow, Acoustical Society of India (ASI) and Fellow of Institution of Electronics and Telecommunication Engineer (IETE), India. He is also a Senior Member, Institute of Electrical and Electronics Engineers, (IEEE, USA).



Prof. K. Samudravijaya

Prof. Samudravijaya is currently visiting faculty at Centre for Linguistic Science and Technology, IIT Guwahati. He has worked as Scientific Officer at TIFR Mumbai, Project Leader at Speaker verification lab, Aum Systems Inc., USA, and Visiting Scientist at Carnegie Mellon University, Pittsburgh, USA. His core area of research is Speech Technology. He has been awarded several prestigious awards including Prof. Rais Ahmed Memorial Lecture Award (2008), Sir C V Raman Award, Acoustic Society of India(2003), Indian Phonetic Society Lecture Award (2002), UNDP Fellowship for research at CMU, Pittsburgh (1988), Best Ph.D. Thesis Award(1986) and several others. He has worked extensively on various projects from government agencies including DeitY, DIT, UNDP and others. He has published approx. 100 papers in reputed Journals and conferences and travelled extensively in India and abroad to deliver Keynote addresses and expert talks.

Prof. Deepak Garg

Dr. Deepak Garg is a Professor and Head, Computer Science and Engineering Department, of Bennett University. He is also holding the position of Director, NVIDIA-Bennett Research Centre for Artificial Intelligence at the University. He has 20 years of rich experience in academics and research. He is a ABET PEV and Senior Member of ACM and IEEE. He has delivered more than 150 expert talks around the country. He has more than 100 publications in reputed Journal and conferences. He has handled research funding of around 2.45 crores from different funding agencies. He has been chair of IEEE Computer Society and IEEE Education Society, India Council. His research is in Deep Learning, Data Science, Data Structures and Algorithms. He is also passionate about quality of higher education in India and works extensively on MOOCs and pedagogy space.

Committees

General Chair

Dr. (Mrs) Amita Dev,	Vice Chancellor, IGDTUW, Delhi India

General Co-Chair

Prof. S.S. Agrawal,	Emeritus Scientist, CSIR

Convener

Dr. Arun Sharma,	HoD-IT, IGDTUW

International Advisory Board Committee

Prof. Satoshi Nakamura,	NAIST, Japan
Prof. Nemeth Geza,	Budapest University, Hungary
Dr. Barbara Zitova,	Institute of Information Theory and Automation ASCR, Czech Republic
Mr. Vivek Khaneja,	Executive Director, CDAC, Noida
Ms. Anu Khosla,	SAG DRDO
Dr. Nick Campbell,	Trinity College Dublin, IR
Dr. S.K. Shrivastava,	Head-TDIL, DRDO
Dr. S.K. Jain,	Director General, CFSL
Dr. P.C. Ching,	Chinese University of H.K.
Prof. Arun Kumar,	IIIT-Delhi
Dr. P.K. Saxena,	PSA, PMO
Dr. Anupam Shukla,	Director, IIIT-Pune
Mr. Karunesh Arora,	CDAC, Noida
Dr. Sunil Tyagi,	Directorate of Naval Design (SDG), IHQ MOD(Navy) Delhi
Prof. M.N. Hoda,	BVICAM, New Delhi

Technical Program Committee

Dr. Chai Wutiwiwatchai,	NECTEC, Thailand
Dr. Bin Ma,	Institute for Infocomm Research, Singapore
Dr. Hsin-Min Wang,	IIS, Academia Sinica, Taiwan
Dr. Ashish Seth,	INHA University, Tashkent
Dr. Mirna Adriani,	UI, Indonesia
Dr. Deepak Garg,	BENNETT University, India
Dr. Mani Madhukar,	IBM India Pvt. Ltd
Dr. Shakir Khan,	IMAM University, Saudi Arabia
Dr. P.K. Dass,	IIT Guwahati, India
Dr. Omar Farooq,	AMU, India

Dr. Shweta Bansal,	KIIT, Gurgaon, India
Dr. Arun Kumar,	Jubail Univ., SA
Dr. Anup Girdhar,	Sedulity Solutions,Delhi
Dr. Poonam Bansal,	MSIT, Delhi
Dr. Vinod Shukla,	Amity University Dubai
Dr. Kirti Tyagi,	INHA University, Tashkent
Dr. Atul Kumar,	KIIT, Gurgaon
Dr. Sanjay Misra,	Covenant Univ., Nigeria
Dr. Poonam Bansal,	MSIT, Delhi
Mrs. Swaran Lata,	Ex-Director & Head, TDIL Programme, DEITY
Dr. Sunita Arora,	CDAC
Dr. Kamini Malhotra,	SAG, DRDO
Dr. Samundra Vijay,	IIT-Guwahati
Dr. Shweta Sinha,	KIIT, Gurgaon
Dr. R.K. Singh,	IGDTUW
Dr. Ashwni Kumar,	IGDTUW
Dr. A.K. Mohapatra,	CTA, Delhi Police
Dr. Deepak Garg,	Bennett University, Gr. Noida
Dr. Archana Baliyan,	MSIT, New Delhi

Organizing Committee

Dr. Ela Kumar,	IGDTUW
Dr. Nidhi Goel,	IGDTUW
Dr. Manoj Soni,	IGDTUW
Ar. Preeti Chauhan,	IGDTUW
Dr. Preeti Vajpeyi,	IGDTUW
Dr. Shalini Arora,	IGDTUW
Dr. Jasdeep Kaur,	IGDTUW
Ar. Vishal Rai,	IGDTUW
Ms. Bhawna Narwal,	IGDTUW
Dr. N.R. Chauhan,	IGDTUW
Dr. Viveak Chawla,	IGDTUW
Dr. Mohana Ghosh,	IGDTUW

Contents

TRACK 4: Applications of AI

TRACK 1

MACHINE LEARNING - I

An Algorithm for Auto-generation of Mockups of Legacy Applications using Behavior Analysis and Artificial Neural Network

Suman De[1], Debjyoti Das[2]

[1]Developer Associate, SAP Labs India Pvt. Ltd., Bangalore, Karnataka, India
[2]Developer, SAP Labs India Pvt. Ltd., Bangalore, Karnataka, India

Abstract: With the simplification of development processes, the need for upgrades of existing solution for enterprise applications is increasing. Enterprise Resource Planning software vendors face a major challenge of transforming complex business processes into lightweight and easy-to-use interfaces that helps industries run better. It is important to do so without affecting the daily activities of an end user. Various tools have been developed to create mockups and focus has shifted towards rigorous acceptance testing, but the actual end users have majorly been ignored, resulting in a delay of implementation of the new products.

This paper proposes an idea to record the behavioral pattern of end users and generate relevant mockups for a revamped software with the help of Artificial Neural Network using a deep learning model. It also explains, how the proposed solution can reduce the Cost to Company in terms of training the end users for using the upgraded product.

Keywords: Artificial Neural Networks, Behavioral Sciences, User-Centered Design, Software Prototyping, Enterprise Resource Planning, Usability Engineering

I. INTRODUCTION

Enterprise applications are facing a continuous challenge of upgrading and simplifying the approach taken by end users to complete business processes. Customer needs are growing with respect to the advancements of technology and optimized project management methodologies have only made it crucial to deliver products at a much faster rate than before. The emergence of open source and microservice based development have all helped expedite the process and calls for further fast-tracking of the same with respect to software development. User Interface designing has seen a bright light in terms of understanding the real needs of a customer and plays a vital role while developing such enterprise applications. Designing mockups and passing design reviews with respect to end user behavior has given birth to Design Thinking and Usability Engineering which looks at various ways to help User Experience designers come up with the right mockups. The aim here is to have empathy towards the customer and deduce the real requirements before coming up with the User Interface mockups.

The availability of tools like Balsamiq, Axure, Visio, etc. has helped the purpose of preparing and optimizing the process of mockup development of User Interfaces. The entire creation process remains manual and still takes a considerable amount of time to provide the final draft for the mockups which is then taken by the development teams to code and develop into real products. Multiple researches are carried out with respect to this topic that could enhance the mockup development phase as well as project management when it comes to software development.

Deep Learning is a field which is used in multiple areas including Healthcare, Supply Chain, e-commerce, science and other research fields. The theory of predicting the right path from a multiple possibility situation via Deep Learning is facilitated by a layer of logic deduced through an Artificial Neural Network. It is possible to classify a group of objects with respect to known and unknown clusters by means of supervised and unsupervised learning which categorizes newly discovered patterns. Pattern Matching is chosen for the given problem in accordance to the power of classifying a business process into a template to be adapted that fits the business problem best.

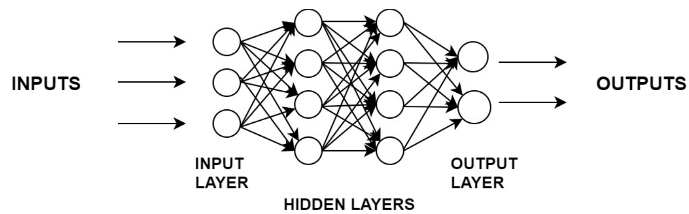


Fig. 1. Artificial Neural Network takes in a specific input and provides a definite output

The proposed paper looks at the mentioned problem statement of optimizing the mockup creation process using an Artificial Neural Network that would help predict the right mockup for the right business process by analyzing end user behavior in the existing legacy application without disrupting current functionality. Figure 1 shows how Artificial Neural Network gets a set of inputs and produces possible outputs after passing through multiple hidden nodes which decides a possible path for the final optimized output. This model continually analyses behavioral aspects that are at par with how a human would draw conclusions. The attempt was to identify the required mockup by having inputs with respect to human behavior, templates and standard guidelines which after multiple processing with conditional biases, gives several outputs in terms of mockups that can be used to simplify the existing legacy applications.

II. RELATED WORKS

There have been multiple attempts to switch from creating paper wireframes or mockups and we look at some current tools and papers that have focused on mockup development or have progressed in terms of optimization of User Interface development processes. We look at literature surveys for Machine Learning-Based Prototyping of Graphical User Interfaces for mobile apps, Speech Recognition with Deep Learning methods and User Interface Adaption based on User Feedback and Machine Learning, in this section. We also understand various Mockup development tools used for building mockups for business applications and highlight some of their key features.

The Machine Learning-Based Prototyping of Graphical User Interfaces for Mobile Apps, by Kevin Moran, Carlos Bernal-Cardenas, Michael Curcio, Richard Bonett and Denys Poshyvanik, talks about a data-driven approach for auto generating prototypes. The proposed idea was implemented in a tool, called ReDRAW, for android based mobile devices. The tool was built for grouping various User Interface controls and classifying their respective hierarchies. It was possible for the tool to generate similar mockup artifacts and deliver realistic business workflows. The behavior is also observed in tools like Build where a User Experience Designer has the luxury of creating Mockups and download production ready code that can be further enhanced with business logic.

Considering the usage of Deep Learning in the proposed idea, the literature review by Rubi, Chhavi Rana, we look at Deep Learning algorithms like Deep Belief Networks, Deep Convolutional Network and Restricted Boltzmann Machines and how multiple development tools can be used implement the same. The User Interface Adaptation based on user feedback and Machine Learning dives into changing of User Interfaces with respect to an Adaptation Rule Engine that uses Machine Learning to accommodate changes during user interactions.

The existing tools for mockup development have significant features that has a huge impact on expediting the software development process. Tools like Balsamic, Axure, etc. are extensively used to develop mockups of software products. Build, as mentioned above, helps a User Experience Designer to not just create mockups but also generate production ready source code that can be used by a User Interface developer to further enhance the software by adding relevant business logic and addressing other integration scenarios. This paper draws motivation from the existing tools and methodologies and introduces a process that helps optimize the manual intervention of creating and refining mockups.

III. PROPOSED IDEA

As observed in the previous sections, there are multiple solution possible for optimizing the enter process of generating fresh mockups. With automation taking a forefront in all business processes, this paper tries to propose a solution that bridges the regular end user behavior and User Experience expert in generating mockups on the fly. The proposed solution uses Convolutional Neural Network (CNN) models (see Fig. 2) to provide effective and efficient models for analyzing and predicting visual imagery through flow of mockups where complicated business processes are involved.

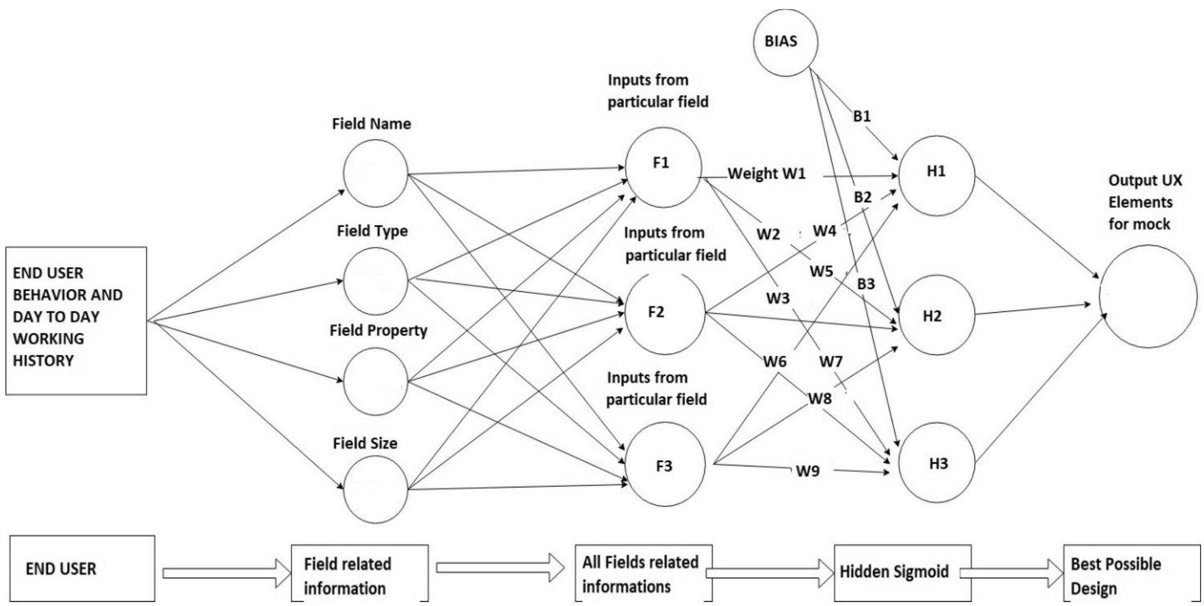


Fig. 2. System collects End User behavior, field priorities, etc. as inputs and provides possible mockups as output.

There are two types of machine learning available: supervised and unsupervised learning. Supervised Learning is considered for instances where a set of classes are already available in which an object can be classified to whereas Unsupervised Learning deals with cases where the model is supposed to generate new classes based on the behavior seen in the objects. In our scenario, we have a pre-defined set of fields as input and a set of templates to choose from and hence, we have used supervised learning in this case. So, the input variables (x) and output variable (y) are used in an algorithm to learn through a mapping function (1) and get the output from the input.

$$y = f(x) \tag{1}$$

Here, the goal is to estimate the mapping function accurately to have the new input data (x) and predict the right output variables (y) for that data. For the proposed paper, a CNN model has been used as a training model. Two neural networks will work sequentially to achieve the same goals.

IV. DATA COLLECTION

1. End User Behavior

Recording of steps performed by the end-user in the existing application provides the usage details of various fields, controls and the end-to-end process of an application. Enterprise Business suites cover a wide range of applications linked with each other and is boxed as one solution. The recording of user behavior in terms of clicks and usage with relevant time delays provide a data set that the current model requires to analyze and further suggest modifications on existing solutions or even create simplified application mockups that will continue to do the same tasks as before.

2. Standard and Customized Templates

The second part of data collection is training the model to the second neural network which gives the user a properly made mock up using the field set collected from the last neural network using user behavior. The user also has an option to introduce new training model for template generation to support and train the neural network for all scenarios. Providing some predefined template may not suffice all complicated scenarios hence custom models needs to be introduced and trained for possible scenario.

V. OUTPUT AND RELATION TO USER EXPERIENCE

Application development for existing or new solutions depend a lot on the mockups generated in the initial phases of the Software Development Life Cycle. In case of existing applications, the availability of end user behavior data in real-time scenarios with a mapping of relevant templates helps generate End-to-End mockups for any enterprise business application. The outputs of the models mentioned in the paper are various mockups that are for development.

The use of back propagation also provides the possibility of back tracking from a proposed mockup design to another and select the one that suits the customer. The availability of multiple mockup choices as outputs helps the User Experience Designer to provide an apt and robust solution based on customer interactions with respect to clicks and usages of controls in existing screens.

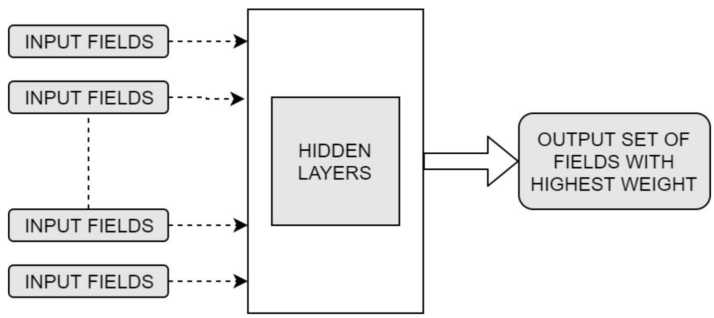


Fig. 3. A simple diagram of neural network where we get the set of fields based on the end user usage.

VI. PROCESS FLOW AND ALGORITHM

The process flow as discussed, starts with maintaining an extensive data of end-user behavior recorded during the usage of legacy applications. The next steps of analyzing and prediction of possible outputs depend on couple of neural networks that work sequentially to predict the possible mockups. The first neural network will help us to identify the set of input fields based on the respective weights calculated depending on user behavior (see Fig. 3).

This results in classifying the fields with highest weights and provides the required order in terms of their usage. The second step for this process is to get the appropriate template which could be mapped to the field sets obtained at the end of the first neural network. The aim of this neural network is to select the desired template for the business process and simplify or rebuilt it. Fig. 5 shows a set of standard templates, that a Business Suite refers to while developing the User Interfaces, which is taken as input and produces the final output as the desired mockup for the application.

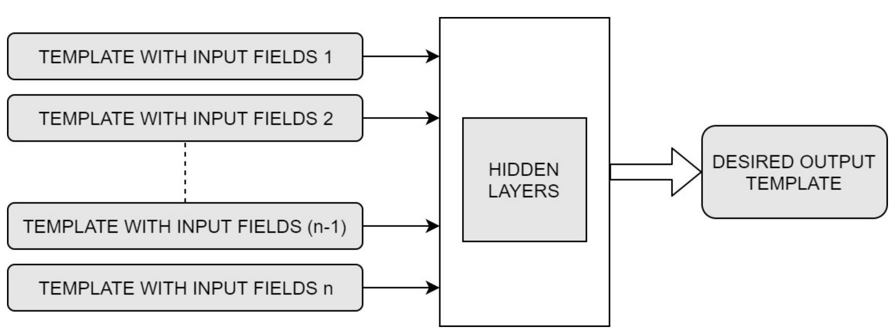


Fig. 4. A simple diagram to show the neural network of a template formation using the output field set from Fig. 3.

The proposed idea doesn't just look at providing an output or a possible mockup but also discusses the possibility of looking into further mockup designs if a suggested design is rejected. The possibility of providing more than one mockup if there is a need is addressed by the concept of Back Propagation. Back propagation is a method used in artificial neural network to find a gradient decent of a network which is used to find proper weights and bias. Back propagation can happen several times for a single node in a network till the required output model is achieved. It is one of the main processes involved in neural network to train the network of nodes.

An accurate back propagation along with appropriate training of the network model helps recover from an erroneous or rejected output and achieves another desired template for the given field set. This back-tracking approach through back propagation for finding alternative but apt mockups, can be best understood from Fig. 5. This is also the basic CNN architecture for the classification algorithm used in the idea.

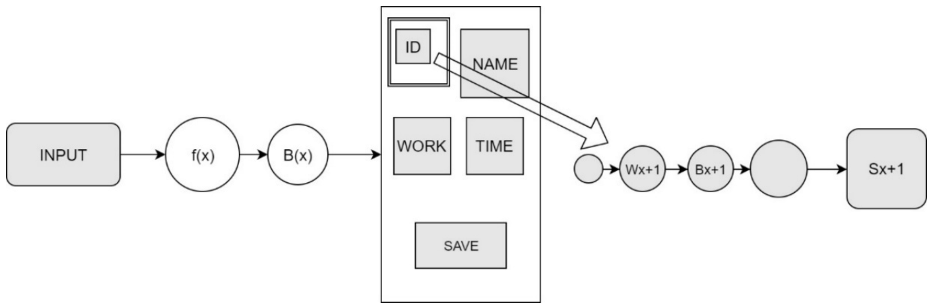


Fig 5. A diagram which shows how CNN model works to find the proper template for the user

It also shows that a template with n pixels and each neighborhood through pooling steps become a pixel, after that scalar weight $Wx + 1$ is used also a bias $Bx + 1$ is also added and then an activation function is used to produce a narrow n times feature $Sx + 1$.

The process is now conceptualized in an algorithm which has 3 different phases: Convolutional, Transfer knowledge and Weight update learning phase. The concluding part of this section looks at the proposed algorithm of this paper.

1) Convolutional Phase

First initialize all the weights and biases of the CNN to a small value

Set learning Ω rate such that $0 < \Omega < 1$

$n = 1$

repeat

For $m = 1 \; to \; M$ do

  propagate pattern Xm through the network

  For $k = 1$ to the number of the neurons in the output layer.

  Get error

End For

For layers $L - 1 \; to \; 1$ do

  For maps $j = 1 \; to \; J \; do$

    Finding the error factor which needs to be back propagated.

  End For

End For

For $i = 1 \; to \; L \; do$

  For $j = 1 \; to \; J \; do$

    For all weights of map, $j \; do$

      Find $d(w)$

      Update the weights and biases

      $w(new) = w(old) + d(w)$

End For

End For

End For

$n = n + 1$

Find mean square error (MSEI)

Until MSEI < ε or n > maximum bounds

The forward pass of the algorithm can also be described through (2):

Output of the neuron k, column y in the lth convolution layer and kth feature pattern:

$$O_{x,y}^{(l,k)} = \tanh\left(\sum_{i=0}^{f-1} \cdot \sum_{r=0}^{Kh} \cdot \sum_{c=0}^{Kw} W_{(r,c)}^{(k,l)} O_{(x+r,x+c)}^{(l-1,i)} + Bias^{(l,k)}\right) \tag{2}$$

where, f is the number of convolution cores in a feature pattern. Output of neuron of row x, column y in the lth sub sample layer and kth feature pattern:

$$O_{x,y}^{(l,k)} = \tanh\left(W^{(k)} \sum_{r=0}^{Sh} \cdot \sum_{c=0}^{Sw} \cdot O_{(x*Sh+r,y*Sw+c)}^{(l-1,k)} + Bias^{(l,k)}\right) \tag{3}$$

The output of the jth neuron in lth hide layer H:

$$O_{(l,j)} = \tanh\left(\sum_{k=0}^{s-1} \cdot \sum_{x=0}^{Sh} \cdot \sum_{y=0}^{Sw} W_{(x,y)}^{(j,k)} O_{(x,y)}^{(l-1,k)} + Bias^{(l,j)}\right) \tag{4}$$

Among the given symbols, the feature pattern in the sample layer is s. Output of the ith neuron lth output layer F

$$O_{(l,i)} = \tanh\left(\sum_{j=0}^{H} \cdot O_{(l-1,j)} W_{(i,j)}^{l} + Bias^{(l,i)}\right) \tag{5}$$

2) Knowledge Transfer phase

Repeat

For $tk = 1 \; to \; TK$ (number of training samples)

Propagate pattern x_{tk} through the network

For $z = 1$ to the number of neurons in the last convolutional layer (Z)

Find output $Oz = (O1, O2, O3 \ldots \ldots . Oz)$

Find O_{ztk} using the TSL framework

End For

End For

3) Weight update learning phase

The below algorithm acts as the basis of identifying the required set of fields that are commonly used by the end-user. Once, this set of fields is prepared, it will be used as an input in the next neural network, where the template is generated using the input field set.

$n = 1$

repeat

$For \; tk = 1 \; to \; TK$

Train the feedforward layers (Layers after last convolutional layer) using Oztk available in phase II. Gradient decent algorithm can be used

End For

$N = n + 1$

Find MSE2

Until MSE2 <ε or n> maximum bounds

The learning algorithm provides the desired output and to train the network using the standard templates that has already been fed to the system.

VII. RESULTS AND DISCUSSION

The given automated process counts for significant optimization of time taken to prepare mockups by collating the user requirements. For simplifying legacy applications, the time taken to realize the new User Interfaces is completely undertaken the proposed idea.

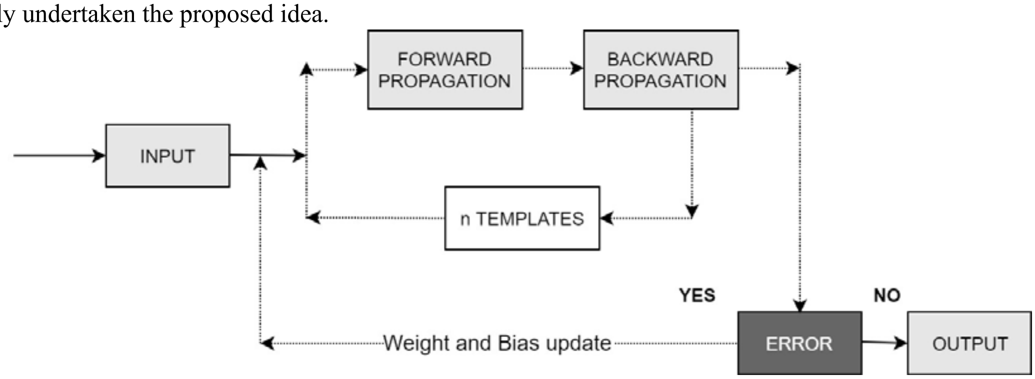


Fig. 6. Iterations involved for generating the required output/mockup

Time for serial execution is $(T1 + T2)\ X\ N\ + T3$

For parallel execution for the n nodes in the hidden layers the time complexity comes

$$\sum_{i=1}^{i=n} . ((T1 + T2)\ \mathrm{X}\ N\ ) + T3 \qquad (6)$$

T1: time taken for forward propagation of a node

T2: time taken for back propagation of a node

T3: Time take to update the weight and bias

N: Number of templates.

n: number of nodes

Here, we observed the steps for iteration of a node and generation of relevant mockups. The similar iteration will also take place for other n nodes too. In serial realization method, the total execution time is N times of T1 and T2.

The software development organization has the opportunity of delivering quality applications according to the realms of the customer or client as the auto-generated mockups are the exact implementation of their regular activities. With respect to other mockup development tools, this is a significant breakthrough as the proposed model is looking at auto creation of mockups that is still done manually where the possibility of human error is considerably high in terms of understanding the customer's needs from their show. The mockup generator takes care of the Usability principles that are mandatory to be followed for creating mockups.

VIII. CONCLUSION

With increasing demand for cloud ready, user-friendly applications, the need for optimization in time taken to build an application is at a high. The proposed idea works for reducing the time taken by a considerable amount via auto creating mockups and delivering the same in hours instead of months. The responsibility of a User Experience Designer in an organization can be made more focused-on development of standard templates that can be reused and maintain consistency over the Business suite. The Cost to Company in terms of time taken to deliver a product is drastically reduced and hence, has a major impact on the software release cycles.

The prediction based on the usage of various User Interface controls also help deliver the right product without any perception biases that generally occur dur to manual interferences. The idea has a lot of promise for Product-based organizations who are focused on developing new products and enhance the existing or legacy applications into a more simplified version of the same.

IX. FUTURE SCOPE

The proposed idea has a significant impact on Enterprise Application development and simplifying existing legacy applications. The immediate scope for the proposed idea would be consider a scalable implementation of the algorithm and validating it with respect to existing mockup development processes and a successful integration with the tools.

ACKNOWLEDGMENT

This paper wouldn't have been possible without the continuous support of Mr. Deepak Padmanabha, who has guided and supported us in setting the right approach to the problem. We also thank Mr. Atreya Biswas, for his guidance in understanding the basics of Artificial Neural Network, which was vital for clarifying doubts about the topic.

REFERENCES

[1] Abdalha Ali, Muasaad Alrasheedi, Abdelkader Ouda & Luiz Fernando Capretz. A Study of the Interface Usability Issues of Mobile Learning Applications for Smart Phones from the User's perpectives. *International Journal on Integrating Technology in Education (IJITE)*, Vol.3, No.4, December 2014.

[2] Alexander Mordvintsev, Christopher Olah & Mike Tyka. Inceptionism: Going Deeper into Neural Networks. *Retrieved* June 17, 2015, *from https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html.*

[3] Alex Krizhevsky, Ilya Sutskever & Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. Retrieved November 4, 2019 from https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[4] Anuse, A. & Vyas. A novel training algorithm for convolutional neural network. *V. Complex Intell. Syst., Springer Berlin Heidelberg,* (2016) 2: 221. Retrieved from https://doi.org/10.1007/s40747-016-0024-6.

[5] Axure. Retrieved November 4, 2019, from *https://www.axure.com.*

[6] Balsamiq. Retrieved November 4, 2019, fr*om https://balsamiq.com/company.*

[7] Bambang Krismono Triwijoyo, Widodo Budiharto & Edi Abdurachman. The Classification of Hypertensive Retinopathy using Convolutional Neural Network. *Procedia Computer Science, 2017*, Volume 116, Page 166, DOI: 10.1016/j.procs.2017.10.066.

[8] Fei Wang, Mengqing Jiang, Chen Qian, Shuo, Chneg Li, Honggang Zhang, Xiaogang Wang & Xiaoou Tang. Residual Attention Network for Image Classification. arXiv:1704.06904v1 [cs.CV], April 23, 2017. Retrieved November 4, *2019, from https://arxiv.org/pdf/1704.06904.pdf.*

[9] Haikel Alhichri, Essam Othman, Mansour Zuair, Nassim Ammour and Yakoub Bazi. Tile-Based Semisupervised Classification of Large-Scale VHR Remote Sensing Images. *Journal of Sensors, 2018*, Volume 2018, Page 1, DOI: 10.1155/2018/6257810.

[10] Kevin Moran, Carlos Bernal-Cardenas, Michael Curcio, Richard Bonett & Denys Poshyvanyk. Machine Learning-Based Prototyping of Graphical User Interfaces for Mobile Apps. *IEEE Transactions on Software Engineering*, arXiv:1802.02312v2 [cs.SE], June 5, 2018.

[11] Nesrine Mezhoudi. User Interface Adaptation based on User Feedback and Machine Learning. *IUI'13 Companion*, March 19–22, 2013, Santa Monica, CA, USA.

[12] Rubi, Chhavi Rana. A Review: Speech Recognition with Deep Learning Methods. *International Journal of Computer Science and Mobile Computing*, Vol.4 Issue.5, May- 2015, pg. 1017-1024.

[13] Tianyi Liu, Shuangsang Fang, Yuehui Zhao, Peng Wang & Jun Zhang. Implementation of Training Convolutional Neural Networks. Retrieved November 4, 2019, from https://arxiv.org/abs/1506.01195.

[14] Visio. Retrieved November 4, 2019, *from https://support.office.com/en-us/article/use-wireframe-templates-to-design-websites-and-mobile-apps-2d54dc55-f5c4-49a2-85da-d649eb7fc281*

Circular Pattern and Angle based Feature Extraction for Sketch based Image Retrieval from Large Databases

G.G. Rajput[1] and Prashantha[2]

[1]Department of Computer Science Akkamahadevi Women's University, Karnataka, India
[2]Department of Computer Science Rani Channamma University, Karnataka, India
[2]prashant.koryal@gmail.com

Abstract: Image retrieval system refers to retrieval of images from databases offline/online given a query in the form of text, image, or sketch. Among various approaches for querying, sketch as query has gained more attention in recent past since drawing a sketch doesn't need an expertise and that it represents what user is looking for in visual form. Moreover, sketch as a query becomes more relevant when example image is not available. Low performance of sketch based retrieval system arises due to natural ambiguity of sketch as well as poor drawing. In this paper, we propose an efficient method to improve the retrieval performance of SBIR system. Angle based and concentric circle-based features extraction is propose and weighted based similarity measure is used to measure the similarity between sketch feature and natural image feature. The performance of propose method is compared with the existing methods in the literature.

Keywords: Angle, Concentric circle, Weighted based, Feature vector, Image retrieval

I. INTRODUCTION

Image retrieval has become a popular research area because of exponential growth in digital libraries due to the advancement of technology to capture and store images. Image retrieval systems are used for retrieving, searching, and browsing images from database based on text and visual content. Text based image retrieval system (TBIR) searches for the images given a text description (keyword, label, or annotation) of the image. Content Based Image Retrieval (CBIR) system is an alternative to TBIR. "Content -Based" relies on content of the image under search [2, 3]. The word content in CBIR refers to features of images like color, shape, texture, edges, etc. which can be extracted from that image itself. In, Sketch Based Image Retrieval (SBIR) system, a variant of CBIR, the image under search is represented in the form of free-hand drawing i.e. a sketch. Free-hand sketch represent the salient features of the object and are particularly useful when the image dataset is not annotated or when the user has no similar example image to use as a query for retrieval.

Though SBIR is a better alternative to other approaches of querying large image databases, challenge is matching of free-hand sketch to the image in the database so that relevant images are drawn. Number of approaches has been suggested in the literature. Comparison of rough sketch to image is naturally difficult [1]. Instead, descriptors are used to extract the information of the image, descriptors are global or local. Global descriptors encode specific features of the whole image [5], and are significantly used in image analysis, matching, and classification [7, 12, 13] methods. Local descriptors describe small specially localized region of the image [9]. Alternatively, complex models are also used to encode geometric information and mutual relationship of objects, e.g. topology models [4, 6]. However, these models are complex and computationally expensive. Further, many of the proposed method in the literature represent the experimental results carried on small scale datasets. With growing image libraries, it is essential that SBIR system should be able to retrieve the desired result from large datasets.

In this paper, we present an efficient SBIR system wherein feature extraction is done in two ways: by first placing concentric circles on the image; and then enclosing the image in a circle and extracting features for each increment of 45^0 within the contour. For each sub image generated by the said approach, mean is computed as feature. Matching of sketch with image is carried out using a weighted similarity approach. Experiments are performed on large dataset. Figure 1 presents the propose method schematically.

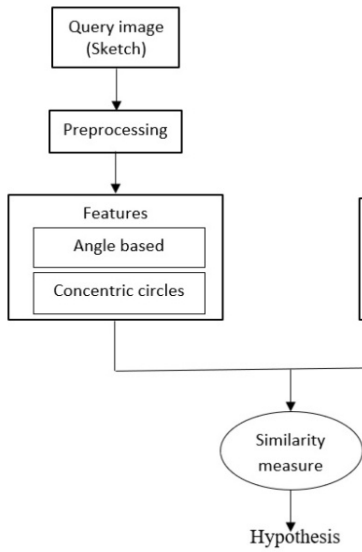


Fig. 1. Propose architecture of Sketch Based Image Retrieval

II. RELATED WORK

Sketch-based image search has become popular since from last few decades [2, 3]. Many methods are proposed in the literature. A brief overview of these methods is presented below.

SBIR based angular partitioning using radial partitioning is presented in [7]. By using the canny operator and Gaussian mask edges are extracted, then to obtain abstract image edges are thinned. Features are computed by partitioning the image into M×N sectors. Manhattan distance is used for measuring the similarity between query vector and data based vector. The method is sensitive to small offsets in location and scale and is rotation invariant. Recent work [10] also utilizes two candidate maps to cover main region and the region of interest and lessens the impact of background thereby making full use of the salient contour. In the criminal investigation, identification of unsubstantial images can be supported by SBIR systems and applications are found in [8, 12].

In the work presented by Sciascio et al. [21], texture-based segmentation is performed to extract shape features. However, the unstable regions resulting from segmentation process fails to retrieve the desired images for the objects being under-segmented or over-segmented. Hoiem et al. [25] performed segmentation of an image into geometrically homogeneous regions using a learning approach to segmentation by estimating the likelihood that two super pixels belong in the same region.

From literature [16,17], early works in sketch based image retrieval systems involved queries described using blobs of color or predefined texture. Later shape descriptors based image retrieval [21] and use of spectral descriptors such as wavelets [22] are found predominantly in the literature. An invariant descriptor encapsulating local spatial structure facilitating codebook based retrieval using HoG variants is found in [23, 24].

III. DATABASE DESCRIPTION AND PREPROCESSING

1. Database Description

Images of flickr15k dataset [18] are used for performing experiments. The dataset contains 60 classes of natural images and contains 10 classes of sketch images, each class drawn by different user. Each sketch class of contains 33 sketch images, the total number of sketch images are 10×33= 330, and total number of natural images from 60 classes is 15k, like flower, bird, Indian_arch and fire_balloon, sample of the dataset images illustrated fig. 2.

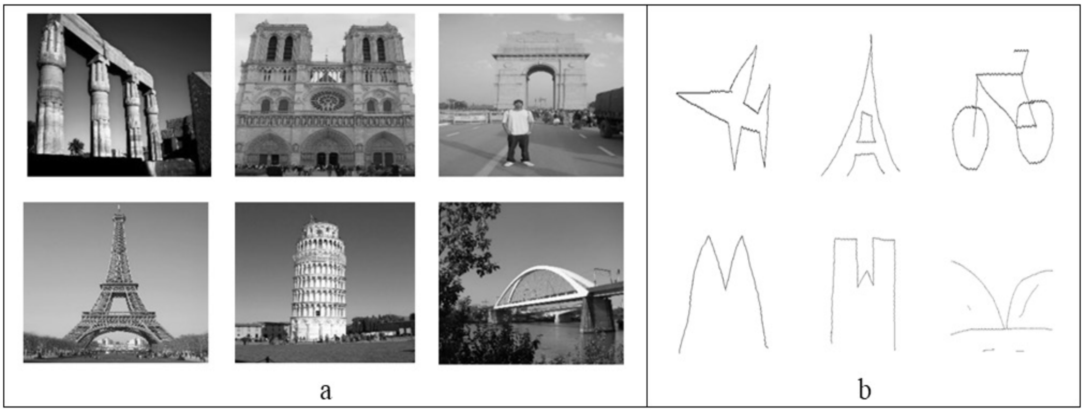


Fig. 2. Flickr15k dataset images (a) natural images (b) sketch images

2. Preprocessing

In the pre-processing stage, we extract global information from the natural images. Otsu's method [14] is used to obtain the contour of the natural images. Subsequently weak contours are eliminated by applying the threshold and image is resized to 260×260 pixels (fig 3).

Fig. 3. The first row contains images, and second row contain global contour images

IV. FEATURE EXTRACTION AND IMAGE RETRIEVAL

The propose SBIR system extracts concentric circles and angular based features fig.1.The feature computation and image retrieval procedure is presented below.

1. Angle based Feature Extraction

Pre-processed image (query image/database image) is resized to 260*260 pixels and center of the image is computed. Angles are sliced into 45 degree based on the radius and center of the image. A contour is overlaid over the image and image is sliced at an angle of 45 degree to obtain 8 partitions of the image Fig. 4. Figure 4(a) shows angular based partition for sample image; green color represents the number of angles used for slicing and red color indicates the maximum radius of the image. The image slicing is done in clock wise direction (starting with 360^0, 315^0,..0^0). Starting with horizontal line, for each increment angle of 45^0, mean is computed for the sub image generated. The sub image generation for an increment of 45^0, each time, is shown in figure 4(b). The feature vector comprises of eight values, one for each sub image.

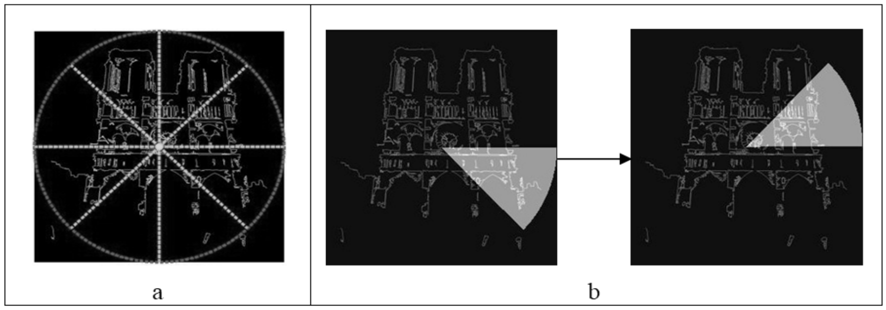


Fig. 4. (a) angular partition (b) feature extraction of each sub image

For image retrieval, Euclidean distance is used to match the query image features and database image features [15]. The results are presented in two steps; for top-10 and top-20 retrievals, respectively. Out of 10 retrieved images 8 are desired images 2 are mismatched. For the second, out of 20 retrieved images 18 are desired matches and 2 images are mismatched shown in fig.5.

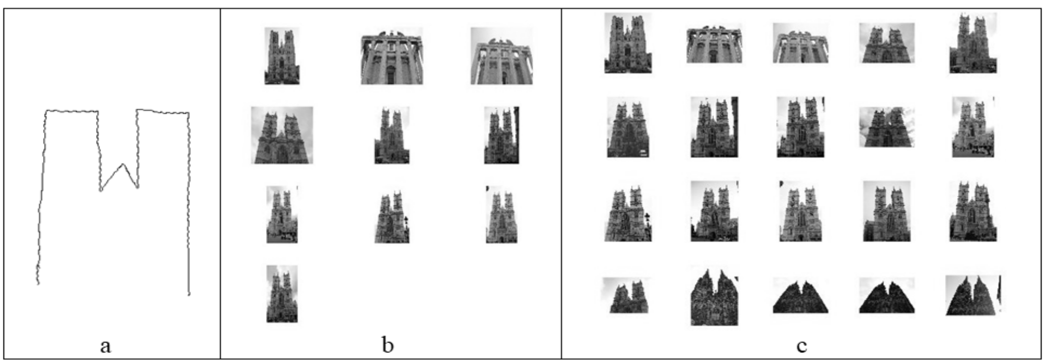


Fig. 5. Example of Angle based results (a) sketch query (b) top-10 results (c) top-20 results

2. Concentric Circle based Feature Extraction

In this approach, the sketch query/image divided into in concentric circles. The image is resized to 280×280 pixels, and then center of the image is found. Then eight concentric circles are overlaid on the image, respectively, to obtain eight partitions of the entire image fig.6 (a). Mean value of each sub image so generated is calculated and stored in vector as feature. The generation of concentric circles and computation of mean value is illustrated in fig. 6(b).

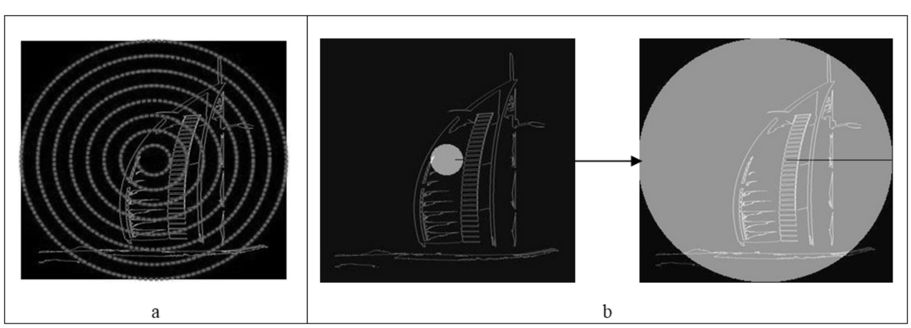


Fig. 6. (a) Concentric circles partitions (b) concentric circle based feature extraction

Euclidean distance is used to match the query image features and database image features. The retrieval performance is analyzed for top-10 and top-20 retrievals. In top-10 approach, 8 are matched and 2 are most similar to the query. In top-20 retrieval approach, first 17 images match whereas remaining 3 images are most similar to the user query shown in fig.7.

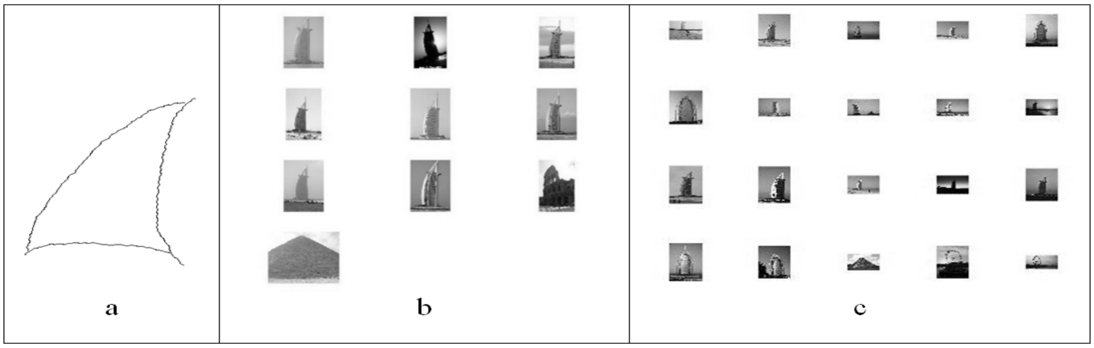


Fig.7. Example of Concentric circle results (a) sketch query (b) top-10 results (c) top-20 results

3. Combining Features using Weighted Similarity Approach

The angle and concentric circle features obtained as described in above are combined using weighted approach as defined in [20]. Let represent the similarity where n represents the number of dataset images and m denotes number of feature approaches employed. In our case, represents angle based features and concentric circle features. On the other end, the features extracted from sketch image is defined as Let sim1(n) and sim2(n) denotes the similarity of angle based and and the similarity of the concentric circle and , respectively. We measure the similarity of sketch image and database image by using following formula.

$$\mathrm{sim(n)} = w \times \mathrm{sim1(n)} + (1 - w) \times \mathrm{sim2(n)}\ , \mathrm{n} - 1{,}2 \ldots., \mathrm{N} \tag{1}$$

Let n=1,2……N denotes number of dataset images and w indicates the weight in the interval [0, 1]. Empirically, weight w was set as w=0.7. Finally, after applying the equation (1) for matching we sorted the similarity results in descending order to retrieve top-n images. Figure 8(b) and (c) shows matching results of top 10 and top 20 retrievals respectively for the sketch given in figure 8 (a).

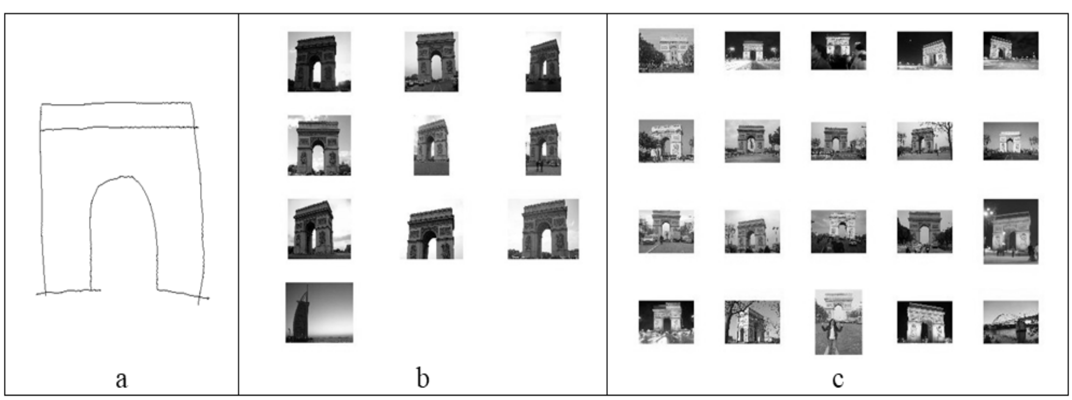


Fig. 8. Example of weighted based results (a) sketch query (b) top-10 results (c) top-20 results

V. EXPERIMENTAL EVALUATION

Experiments are performed using flickr15k dataset [18]. The dataset contains 15k natural images and 330 sketch images. The results obtained using our methods are compared with non Bag-of-Visual-Words approaches and Bag-of-Visual-Words approaches which are widely used in recent years [11, 19, 26 & 18]. The Mean Average Precision (MAP) is used to evaluate the retrieval performance of retrieval system.

$$MAP = \frac{1}{S}\sum_{s=1}^{S} AP(q), \quad AP = \frac{1}{R}\sum_{k=1}^{N} p(k) * rel(r) \tag{2}$$

where S represents the number of Sketch's; R represents number of relevant images in retrieved images; N, the number of retrieved images; p(k) being top precision score; and rel(k), an indicator function with rel(r)=1 for relevant image, otherwise rel(r)=0.

Table 1. Comparison of propose methods with other methods based on MAP (20)

Methods	**Approaches**	**MAP**
Non-BoVW	Angular Partitioning (AP)	0.081
	EHD	0.076
	Structure Tensor	0.073
BoVW	HOG	0.111
	GF-HOG	0.122
	SIFT	0.091
Our Approach	**Angle based**	**0.129**
	Concentric circle based	**0.125**
	Weighted based	**0.136**

The results obtained using angle based features; concentric circular based features and combination of both of the features using weighted similarity approach are comparable with existing methods. Angle based and concentric circle based approaches yielded better retrieval results as compared to the non-BoVW and BoVW approaches on same dataset as presented in table.1.

VI. CONCLUSION

An efficient approach for Sketch Based Image Retrieval in large databases is presented in this paper. A key characteristic of our approach is estimation of similarity between the sketch query and database image based on angle based and concentric circle based feature extraction. Our propose methods yielded better results as compared to other non-BoVW and BoVW methods in the literature. The results obtained using weighted similarity approach are comparable and better than existing methods in the literature as depicted in Table 1.

REFERENCES

[1] Chang, N., & Fu, K. (1979). Query-by-pictorial-example. in: The IEEE Computer Society's Third International Computer Software and Applications Conference. Proceedings. COMPSAC 79, (pp. 325-330).

[2] Yong Rui., & Thomas, S.H. (1999). Image Retrieval: Current Techniques, Promising Directions, and Open Issues. Journal of Visual Communication and Image Representation 10, (pp.39-62).

[3] Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: ideas, influences, and trends of the new age. ACM Computing Surveys, 40(2), (pp.1–60).

[4] Sousa, P., & Fonseca, M. (2008). Sketch-based retrieval of drawings using topological proximity. In VLC.

[5] Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. Int'l J. Computer Vision, vol. 42, no. 3, (pp. 145-175).

[6] Fonseca, M. J., Ferreira, A. , & Jorge, J. a.(2009). Sketch-based retrieval of complex drawings using hierarchical topology and geometry. Computer-Aided Design, Vol 41, issue 12, (pp.1067-1081).

[7] Chalechale, A., Naghdy, G. , & Mertins, A. (2004). Edge image description using angular radial partitioning. IEEE Proceedings-Vision, Image and Signal Processing, vol. 151(2), (pp.93-101).

[8] Jain, A.K., Lee, J.E., & Jin, R. (2010). Sketch to photo matching: a feature-based approach. Proc.SPIE, Biometric Technology for Human identification. vol.7667.

[9] Lowe, D.G. (2004). Distinctive Image Features from Scale-Invariant Key points. Int'l J. Computer Vision, vol. 60, no. 2, (pp.91-110).

[10] Qian,X., Zhao, Y., & Han, J. (2015). Image Location estimation by Salient Region Matching. IEEE Trans. Image Processing, vol.24, no.6.

[11] Chalechale, A., Naghdy, G., & Mertins, A.(2005). Sketch-based image matching using angular partitioning. IEEE Transactions on Systems, Man and Cybernetics, Part A 35 (1) (pp.28–41).

[12] Rajput, G.G., Prashantha, & Geeta, B. (2018). Face Photo Recognition from Sketch Images Using HOG descriptor. 2nd International Conference on Inventive Communication and Computational Technologies .IEEE Xplore.

[13] Salembier, P., Sikora T., & Manjunath, B. S.(2002). Introduction to MPEG-7: multimedia content description interface. John Wiley & Sons, Inc.

[14] Otsu, N. (1979) A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics. Vol. 9, No. 1, (pp. 62–66).

[15] Jing, Li., & Bao-Liang, Lu. (2009). An adaptive image Euclidean distance. Science Direct, Pattern Recognition 42 (pp.349 – 357).

[16] Ashley, J., Flickner, M., Hafner, J. L., Lee, D., Niblack, W., & Petkovic, D.(1995).The query by image content (QBIC) system. in SIGMOD Conference.

[17] Smith, J.R., & Chang, S.F. (1996). Visual seek: a fully automated content-based image query system. In ACM Multimedia, (pp. 87–98), New York, NY, USA.

[18] Hu, R. & Collomosse, J. (2013). A performance evaluation of gradient field hog descriptor for sketch based image retrieval. Computer Vision and Image Understanding, vol.117, no.7, (pp.790-806).

[19] Eitz, M., Hays, J., & Alexa M. (2012). How Do Humans Sketch Objects. In Proceedings of SIGGRAPH.

[20] Yuting Zhang., Xueming Qian, & Xianglong Tan (2015). Sketch-based Image Retrieval Using Contour Segments. IEEE Xplore: Xiamen, China.

[21] Sciascio, E., Mongiello, M. ,& Mongiello, M. (1999). Content based image retrieval over the web using query by sketch and relevance feedback. in In Proc. of 4th Intl. Conf. on Visual Information Systems, (pp.123-130).

[22] Jacobs, C. E. , Finkelstein, A., & Salesin, D. H. (1995). Fast multi resolution image querying, in Proc. ACM SIGGRAPH, (pp.277-286).

[23] Hu, R., Barnard, M., & Collomosse, J. P. (2010). Gradient field descriptor for sketch based retrieval and localization. in ICIP, 2010, (pp.1025–1028).

[24] Eitz, M., Hildebrand, K., Boubekeur, T., & Alexa, M. (2010). Sketch based image retrieval: Benchmark and bag-of features descriptors. IEEE TVCG, vol. 99.

[25] Hoiem, D., Efros, A.A., & Hebert, M. (2005). Geometric context from a single image. in ICCV. vol. 1, pp. 654– 661, IEEE.

[26] Eitz, M., Hildebrand, K., Boubekeur, T., and Alexa, M. (2009). A Descriptor for Large Scale Image Retrieval based on Sketched Feature Lines. In Proceedings of SBIM, (pp.29-38).

A Comprehensive Study on Class Imbalance Problem and its Solutions

Shinjanee Gupta, Atima Syal, Akanksha Chandna, Neha Gupta, Khyati Ahlawat

Indira Gandhi Delhi Technical University for Women, Delhi, India

Abstract: The standard machine learning classifiers are devised in such a manner so as to be able to handle balanced distribution of multiple classes of any dataset. Hence, when they are applied on datasets with the problem of imbalanced classification, they prove to be incapable to handle this problem and yield results with low accuracy concerning the minority classes as they assume to have a relatively balanced distribution of all classes and equal costs for misclassification of any data point. These conventional classifiers need to be modified since they are likely to favor the majority class. Minority class data points have a considerably small ratio over correctly classified majority class data points that the accuracy achieved is still high even when all of the minority class data points are misclassified. These existing classifiers are undergoing progressive modification to downsize the difficulty of learning from datasets with imbalanced classification. In the following sections of this work, different approaches for handling the problem of imbalanced classification are discussed in view of vast literature survey related to them. The paper serves the purpose of understanding the problem and establishes baselines against which future progress and success can be measured.

Keywords: Class imbalance problem, Classification, Performance metrics, Data mining, Imbalance data.

I. INTRODUCTION

Data mining is a process which refers to the extraction of valuable information and beneficial patterns from large amount of data [10]. It unearths a variety of hidden patterns along with knowledge from the data which can be used to generate various inferences and useful information. The process includes three steps- preprocessing, extraction and evaluation.

Classification, an essential technique in data mining, is a supervised learning approach to categorize the data into various classes to differentiate between different instances of the data. The main aim of classification is to decide the target class for a specific object in an unknown class [14]. For instance, iris flowers can be classified into three species on the basis of dimensions such as length and breadth of petals and sepals. Areas where classification is widely used are education industry, fraud detection, telecommunication industry and many more.

With great advancements in technology, there is a lot of data being collected every day. Though various classification techniques have been discovered in order to classify the data, still there is a lack of accuracy due to problem of imbalanced classification. It occurs when one class outnumbers the other class by large amount. Class imbalance problem can be present where only two classes exist in the classification problems and also exist in multi-class classification problems. For instance, during diagnosis of medical disease, if a disease is misdiagnosed, it can have a bad impact. If 95 percent of people do not have cancer and 5 percent have cancer, and ‘no cancer’ is predicted for all datasets, the accuracy is 95% (a very high accuracy); but for those five percent the test is negative, and can be life threatening for the patient.

Various algorithms of classification such as Association Rule mining, Logistic Regression, Naive bayes Classifier, nearest neighbor have been developed to classify and categorize the data. The accuracy recorded by these approaches is good when there is balanced distribution for all classes. But they suffer from ‘Class imbalance’ problem. For the infrequent classes, all the standard algorithms existing have very low accuracy of prediction. For instance, when there are two classes, usually minority class is termed as positive class and majority class is termed as negative class. This problem is prevalent in various enterprises where a lot of data is collected and processed every day. Hence it becomes really challenging to classify the data with good accuracy.

Various existing works on class imbalance problem have been studied for the purpose of experimentation and analysis. To enhance the accuracy of the currently existing models, certain kind of hybridization on data level or algorithmic level or both is required for improved performance and accuracy.

In the following sections of the paper, the class imbalance problem is discussed in detail. Then, based on the current research being carried out in this area, a literature survey highlighting the various solutions proposed in the research is presented. Finally, the last section concludes this work highlighting the outcomes of the literature review and their future scope is discussed.

II. CLASS IMBALANCE PROBLEM

Class imbalance problem is a machine learning problem where the number of minority class data points (positive) is far less as compared to that of majority class data points (negative). The class having lower instances in training data is termed as a positive class and the other one is termed as negative class. This problem has gained popularity and can be widely seen in various fields, e.g. healthcare, content categorization, detection of fraudulent activity etc [6]. Conventional algorithms give correct outcomes for the majority class but fail to give the same results for classes with low instances as they assume to have a balanced distribution of all the classes.

The following diagram illustrates various approaches to solve the problem of imbalance classification.

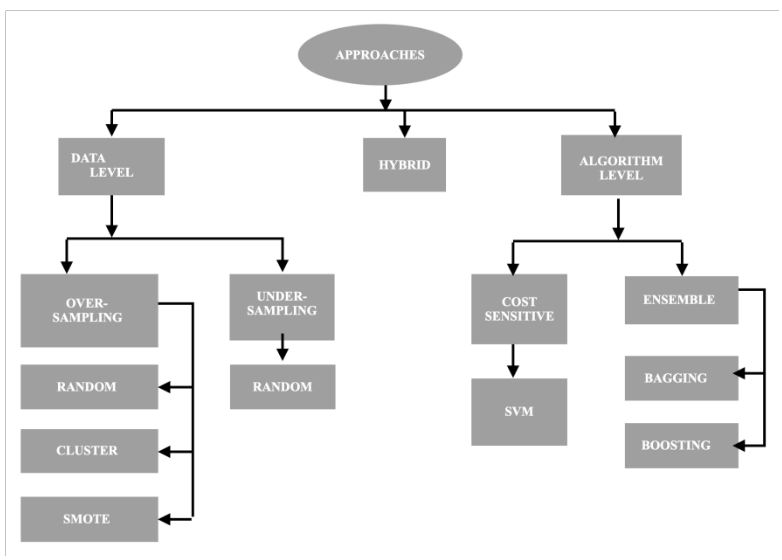


Fig. 1. Approaches to solve Imbalance Classification Problem

There have been three different approaches while addressing imbalanced data:

1. Data-Level Approach

It involves re-sampling of data-set. This approach is more flexible and is a hot topic of research recently. It includes the use of latest machine learning algorithms. Techniques used are:

b. Over-sampling

Minority class samples in the training dataset are increased in this technique. This approach only creates new instances and thus the previous instances are preserved. Hence, no useful insights are lost.

i) Random Over-sampling: Random oversampling increases instances randomly in minority class [14]. It is repeated until the dataset is balanced out. It helps in improving accuracy since it does not lead to information loss. However, it can lead to over fitting due to repetition of minority class.

ii) Cluster based Over-sampling: It makes the use of K-means clustering algorithm for the minority and majority instances, which helps to indicate clusters in the dataset. Each cluster is oversampled until all classes have the same size and same amount of instances. However, it can lead to over fitting of training data.

iii) Synthetic Minority Over-sampling: In this oversampling technique, a subset of minority class is extracted and synthetic samples are added. Dataset then created is used for training the classification models. It helps the eradicating the problem of over fitting.

c. Under-sampling

It involves eliminating the majority class samples in training dataset. It might discard some useful information as some of the instances from the original dataset are removed.

i) Random Under-sampling: This technique eliminates samples randomly from the majority class. The process is repeated until the dataset is balanced out. It helps in improving the accuracy when the data set is large, however, it can rule out important information which can be useful.

2. Algorithm-Level Approach

In this approach existing algorithms are modified and a new algorithm is created in order to solve imbalance classification problem. Deep understanding into new modified learning approach is required and one needs to identify reasons of failure in mining skewed distributions. Major techniques used are cost sensitive and ensemble to classify the data.

a. Cost Sensitive Learning:

In this learning, cost is assigned to every misclassification. Classifying a minority class sample as majority class results in a cost higher than classifying a majority class sample as minority class.

i) Support Vector Machine (SVM): A model in which data used for classification and regression is analyzed and studied. In this, a hyper plane, or a set of hyper planes, is constructed in an infinite-dimensional space for classification and outlier detection.

b. Ensemble learning :

In this, multiple models are generated and combined to solve a particular computational problem.

i) Bagging: It is an abbreviation for Bootstrap Aggregating. It generates n different training samples using combinations with repetitions to produce multi-sets of original data and then bootstrap algorithm is applied separately on each of these samples. It improves the accuracy and stability of machine learning algorithm and reduces variance.

ii) Boosting: It is an ensemble technique which combines various weak classifiers to create a strong classifier by adjusting the weight of an observation based on last classification that can predict accurately. It starts with a weak classifier which is prepared on weak training dataset.

3. Hybrid Approach

The focus is on combining the data level and algorithm level approaches to extract the strengths of these algorithms and reduce their weaknesses. It results in robust and efficient learners making this approach highly popular.

Algorithms such as association rule mining, SMOTE algorithm, LMS method, SVM, soft set theory, rough set theory, neural networks etc. solve the problem to a good extent and yield good accuracy. But when implemented alone, these algorithms do not improve accuracy to much extent. Hence research is being done for implementation of hybrid models to get better performance.

III. RELATED WORK AND SOLUTIONS

In this section, details about literature survey done regarding current scope of study are discussed. Following table Table 1, will demonstrate the research work done in the past decade to solve the problem of imbalance data classification. The research papers have been studied to identify the dataset used, mostly from KEEL and UCI repository covering various domains like healthcare, spam detection and games, preprocessing techniques applied like oversampling, SMOTE, class association rule, classifiers used and performance metrics used like MAUC score which measures area under the receiver operating curve for multi class, accuracy is percent of instances correctly classified, precision depicts how precise the model is, recall is the number of positives captured, F1 score or F measure is the measure of accuracy of the test determined by harmonic mean of recall and precision to evaluate the proposed approach. The research has been done on both data level and algorithm level to yield better results in terms of accuracy.

Table 1. Literature Survey

S. NO.	DATASET USED	PREPROCESSING TECHNIQUES	CLASSIFIERS USED	PERFORMANCE METRICS USED	REFERENCE NO.
1	Breastw, Vehicle, Segment, Glass	K-means clustering, genetic algorithm	KNN and SVM (Support Vector Machine)	Accuracy, G-mean	[15]
2	18 datasets from UCI repository	Biased Resampling	Bagging and Boosting	Accuracy, ROC curve, AUC value	[3]
3	Weibo dataset	Fuzzy based oversampling	SVM (Support Vector Machine)	Precision, Recall, F1-score	[13]
4	MIT-BIH arrhythmia database, European ST-T database, MIT-BIH ST change database	Median Filter, Borderline-SMOTE (BLSM), Context-Feature Module (CTFM)	Two-Phase Training (2PT)	Accuracy, MAUC Score, Sensitivity, Specificity	[5]
5	CelebA dataset	None	Class expert generative adversarial network (CE-GAN)	Precision, Accuracy	[4]
6	KEEL data repository.	Soft set theory, Rough set theory, data formatting, data normalisation and data randomization	Neural networks	Recall, Precision, F-measure	[9]
7	Breast Cancer, Tic-tac-toe, Mofn, Post-operator, German	Class Association Rule	IGB algorithm	Global Accuracy, G-mean, Accuracy, F- measure	[11]
8	Keel dataset Repository, UCI repository	Oversampling	Random -SMOTE algorithm	F-measure value, G-mean value, ROC curve and AUC value	[14]
9	Pima, WBC, WDBC, Live disorder, Appendicitis	Random oversampling for minority instances and under sampling for majority instances	Least Mean Square (LMS) algorithm, SVM, K-NN and Multilayer perceptron [MLP]	CC rate, Error rate, Sensitivity, Specificity, Gmean	[12]

S. NO.	DATASET USED	PREPROCESSING TECHNIQUES	CLASSIFIERS USED	PERFORMANCE METRICS USED	REFERENCE NO.
10	Biomedical repositories (PubMed and Medline)	Novel gene feature clustering	Ensemble document classification	Cluster quality rate, True positive classification rate	[2]
11	Public multi-class imbalance datasets from UCI, KEEL, OpenML	None	Diversified Error Correcting Output Codes (DECOC), combining Error Correcting Output Codes withensemble learning framework	Overall accuracy, Geometric mean, F-measure, and Area Under Curve	[1]
12	Data of respondents that were selected for the gastric cancer screening	None	C4.5 decision tree algorithm using modified proposed weighted entropy	Classification accuracy, Sensitivity, Specificity	[8]

The above table clearly illustrates that techniques at both algorithm and data level are popular in solving the problem of imbalanced classification since last two years. In 2017, the focus was on algorithmic based approaches like C4.5 decision tree algorithm, K-nearest neighbors to improve the classification. Other algorithm based approaches like ensemble learning and rule based classification were popular in 2018. Then neural networks also started getting used for the classification task like multilayer perceptron and CE-GAN. Nowadays, it seems that higher accuracy is obtained by using both data level and algorithm level approach like Jing Jiang et al. presented their work for imbalanced heartbeats classification using a multi-module neural network system which used a combination of Context-Feature Module, Borderline-SMOTEand Two-Phase Training.

IV. CONCLUSION AND FUTURE SCOPE

As evident from the literature review, various classifiers have been implemented independently and in combination and various performance metrics have been used to ascertain better performance. Various data level and algorithmic level modifications to existing models, that have been discussed so far, have given higher accuracy as compared to that when each model is used alone. Various performance metrics that have been used to compare the performance have also been mentioned.

In future, various state-of-the-art approaches can be modified or combined and compared to prove if they are better than the original ones. If yes, a new algorithm can be modeled. Else, more improvisations can be done and tested for improvement in performance. Besides, this work can be extended to cover big data with the problem of imbalance classification.

REFERENCES

[1] Bi, J., & Zhang, C. (2018). An empirical comparison on state-of-the-art multi-class imbalance learning algorithms and a new diversified ensemble learning scheme. *Knowledge-Based Systems*, *158*, 81-93.

[2] Bikku, T., Nandam, S. R., & Akepogu, A. R. (2018). A contemporary feature selection and classification framework for imbalanced biomedical datasets. *Egyptian Informatics Journal*, *19*(3), 191-198.

[3] Wang, B., & Pineau, J. (2016). Online bagging and boosting for imbalanced data streams. *IEEE Transactions on Knowledge and Data Engineering*, *28*(12), 3353-3366.

[4] Cenggoro, T. W. (2018). Deep learning for imbalance data classification using class expert generative adversarial network. *Procedia Computer Science*, *135*, 60-67.

[5] Jiang, J., Zhang, H., Pi, D., & Dai, C. (2019). A novel multi-module neural network system for imbalanced heartbeats classification. *Expert Systems with Applications: X*, *1*, 100003.

[6] Ahlawat, K., Chug, A., & Singh, A. P. (2019). Empirical Evaluation of Map Reduce Based Hybrid Approach for Problem of Imbalanced Classification in Big Data. *International Journal of Grid and High Performance Computing (IJGHPC)*, *11*(3), 23-45.

[7] Ahlawat, K., Chug, A., & Singh, A. P. (2019). Benchmarking framework for class imbalance problem using novel sampling approach for big data. *International Journal of System Assurance Engineering and Management*, *10*(4), 824-835.

[8] Kirshners, A., Parshutin, S., & Gorskis, H. (2017). Entropy-based classifier enhancement to handle imbalanced class problem. *Procedia Computer Science*, *104*, 586-591.

[9] Mohamad, M., Selamat, A., Subroto, I. M., & Krejcar, O. (2019). Improving the classification performance on imbalanced data sets via new hybrid parameterisation model. *Journal of King Saud University-Computer and Information Sciences*.

[10] Ramageri B (2010). Data Mining Techniques And Applications. *Indian Journal of Computer Science and Engineering. Vol. 1 No. 4* pp 301-305.

[11] Abdellatif, S., Hassine, M. A. B., Yahia, S. B., & Bouzeghoub, A. (2018, January). ARCID: A new approach to deal with imbalanced datasets classification. In *International Conference on Current Trends in Theory and Practice of Informatics* (pp. 569-580). Edizioni della Normale, Cham.

[12] Sara Belarouci, Mohammed Amine Chikh. (2017). Medical imbalanced data classification. *Advances in science, technology and engineering systems journal.* 10.25046/aj020316

[13] Tang, W., Ding, Z., & Zhou, M. (2019). A Spammer Identification Method for Class Imbalanced Weibo Datasets. *IEEE Access*, *7*, 29193-29201.

[14] Xie, W., Liang, G., Dong, Z., Tan, B., & Zhang, B. (2019). An Improved Oversampling Algorithm Based on the Samples' Selection Strategy for Classifying Imbalanced Data. *Mathematical Problems in Engineering*, *2019*.

[15] Yong, Y. (2012). The research of imbalanced data set of sample sampling method based on K-means cluster and genetic algorithm. *Energy Procedia*, *17*, 164-170.

Artificial & Natural Intelligence—A Theoretical Comparison

Laxmi Ahuja, Siddharth Gupta

Amity Institute of Information and Technology
Amity University, India

Abstract: The philosophy of artificial intelligence revolves around understanding the related concepts but with a broader understanding thus, logic and philosophy work in coherence and unity. During the current technological and digital times, our dependence on computers and other digitalized components has increased drastically and this trend is likely to continue for decades. Artificial intelligence and machine learning are now holy grails that are employed at nearly all places- chat-bots, insurance eligibility, health care, etc. This paper aims to study the co-existence of humans and machines with a futuristic perspective considering also the social obligations the cohesion will subject itself to.

Keywords: Artificial intelligence, philosophy of artificial intelligence, cognitive science, computer science, Creative Theory, Information Theory

I. INTRODUCTION

Artificial intelligence is quite a common terminology being used almost everywhere these days, once we humans used to imagine that machines would aid us in our daily lives and tasks which now undoubtedly is on the verge of becoming a reality. However, reality itself adheres various constraints, imposed by the limited knowledge of our senses and the answer to what is realism has always changed shapes [1]. This paper demonstrates a comparative study between the natural and artificial intelligence and holds a simple question as its basis i.e. whether machines can surpass the natural mind? [2]. A plausible reason for this study is that until now there was only a single civilization that was considered to have impeccable logical, reasoning, problem solving and communication abilities, but today this civilization is now also a parent to its brain child another deemed intelligent civilization i.e. machines. Here, a philosophical and theoretical approach with respect to humanitarian and computer sciences is followed. Artificial intelligence is the ability of machines to have cognition skills like humans. Right now, it won't be wrong to state that AI is currently in its infancy but the exponential work done in this field is contributing to its growth. We attempt to provide an overview of what would be our world if we live with fully intelligent and functional machines. Humans are creative beings who have the ability to explore and transform things, turning imagination into reality something which can be explained with the example of artificial intelligence. Foremost, we put forward a generic overview of the society and our presence in it. Following this section we have explained the nature of artificial intelligence with a perspective of "thinking" and "fully functional" mind. Then we have described one of the core aspects of this paper i.e. the Generation Theory for machines which also highlights the components of evolution and development in both humans and machines. Then, we have provided a comparison between what we refer to as artificial selection and a graphical depiction of the same. Ultimately, we conclude our study and have discussed the future perspectives and scope.

II. HUMANS & SOCIETY

Humans have been through a long journey to reach where they stand today. They've evolved with what now is known as a universal synchronizing mechanism: time. What they're capable of is a direct outcome of their diversities and variations they faced in the evolution process. In simple words we inherited traits from out parents (or past generations) in order to advance in the future which is why we're at the pinnacle of decadence. Surprisingly various concepts, analogies and theories applicable to us are also applicable to our greatest invention, which is providing machines a similar ability responsible for our prosperity so far, i.e. artificial intelligence. Humans in basic have done only two things to sustain so far, multiplication and struggle for sustenance. Rest of our deeds is an aftereffect. Notice that a civilization can only evolve when it can sustain. We have a complicated cognition mechanism which gives us an ability to learn, observe and interpret things; our social connections also play a vital role in our development where we share knowledge and information. Mankind has to follow several constraints due to social obligations and to keep things in order we have a governing mechanism established in the form of a well defined protocol, more formally known as a law. Society was constructed by us so that we could feel more secure and be more productive as a whole. We did what we did because of the needs we began to understand in the process of evolution and we passed our knowledge to forthcoming generations. The end of a generation's thinking is the beginning of a future generation's thinking and that too with variability. It is important to note that basically, we are just entities. An intelligent entity which can not only learn from past experiences but also has the capacity to reason, refine and redefine.

III. ARTIFICIAL INTELLIGENCE

The entire idea behind artificial intelligence was given by Alan Turing [3].Where he thought whether machines could think like us but the inspiration for intelligent machines wasn't limited to this only, many thoughts and ideas existed too, gaining motivation from various fields like psychology, biology, engineering, Statistics & Probability, Theory of computation, etc. It is first vital to know what comprises of intelligence; there are three forms of intelligence viz. Analytical, creative and practical [4]. All of which are also applicable to the artificial intelligence agenda and clearly AI techniques are being used to solve problems of similar nature being analytical problems the most opportune, requiring encoding/decoding data, looking for patterns and combinations and classifications. While the creative intelligence for machines is what is most likely to happen next as AI can explore things, however there is a constraint that it cannot transform itself as a whole [5]. The simplest way which it uses to explore is Syllogism. It comprises of no beliefs unlike humans but can be studied from a psychologist's perspective as it tries to imitate the conscious processes of the mind like learning. Most common forms of learning, supervised, unsupervised and re-enforcement learning. Out of these three the lattermost aspect is what to be the most apt for consideration. For re-enforcement learning to work in our proposed scenario, there has to be a consideration for internal states due to which the reward system will activate just as it happens with humans, internal state of the mind for humans comprises of emotions, desires, yearnings, etc. As for machines the internal state exists by virtue of configurations provided and is impacted in both i.e. humans and machines by a transition function. It wouldn't be wrong to state that intelligent machines perceive from their surroundings and act accordingly.

IV. THE GENERATION THEORY

Although machines were created by us a long time ago for specific purposes, there was intelligence among them even when there was no digital components the same idea still prevails that now with digital aid they've outsmarted us over the generations. To explain this idea more technically, we here propose a generation theory for machines defining the generation gap between humans and machines, we humans have taken millions of years to reach here but machines never took and never will take that much of time to evolve in a manner that will make them suitable for a full fledged competition with us as their evolution rate is far more greater than ours. In our case there were changing conditions on our planet thus, our ancestors developed themselves accordingly and passed that information and abilities to future generations. Something similar can also be achieved by machines. Another reason why we argue this theory is due to the fact humans explored the world and formed their intelligence, for artificial intelligence there is already an intelligence working behind as the artificial intelligence has certain limitations upon itself due to the information available [6]. But the problem is of scope. A computer program with developed and complex intelligence would not adhere to the same constraints as a human would. We as human beings are subjected to extinction at some point because we require sustenance and for that we need such resources that take millions of years to re-generate. A machine doesn't have that complex requirement as compared to us. Godel's argument [7] states that an effective theory capable of validating basic form of arithmetic is either incomplete or inconsistent, we take this aspect on a positive note as inconsistency is mere enough to cause chaos. It is evident from present scenarios that humans are equally inconsistent and already causing chaos in the society as a whole. But, an important thing to note is that the problem is not with incompleteness and inconsistencies but with the outcome of the same. To support this argument we consider the Chinese room experiment proposed by Searle. Simply put, if a calculation is merely a calculation in a computer oriented simulation (or weak AI) then its application to the real world would still be effective and there will be ramifications. Thus, a calculation even though in a simulated environment can turn into reality if that simulation provides a solution to a real world problem. History has witnessed that humans designed machines to reduce their efforts in the real world and now with artificial intelligence they plan the exact same thing. Another thing that we'd like to point is that the level of complexity in emotions and understanding as shown by humans is going to remain absent in the context of artificial intelligence, as AI works on complex computations but in order to understand the complexities of human emotions any machine would not be able to process that information as emotions won't be that quantifiable. Complex emotions like compassion, friendship and bonding is something that separates humans from others.

V. RISE OF ARTIFICIAL SELECTION

A long time ago Charles Darwin coined the term Natural Selection [8] in the human context stating that only the fittest of all shall survive. In the human context the fittest can be defined as the one who have been through a variety of conditions and adapted to the changes, those who failed to do so were extinct. We argue that something similar is also prevalent in intelligent computer programs, whenever we think of a solution provided to us by the intelligence of computers we tend to optimize it so that it fits to our requirement and adapts to the environment in which it runs. Also, keeping in mind the thought that AI devices can be profoundly distinguished from a computer program once they reach a certain level of intelligence[9]. Any program which fails to learn is discarded. This effect is what machines and humans have in common. However, machines cannot be extinct as long as the most important resource they need persists i.e. us. Because we as humans have our needs and we'd keep on looking for easy and optimal ways to solve our problems. Consider,

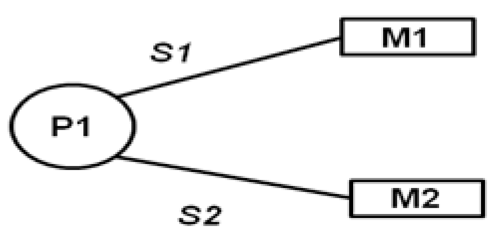


Fig. 1. The artificial selection process

In the figure mentioned above, there are two intelligent machines named M1 and M2 both of them provide different solutions S1 and S2 to a real world problem P1. Just like it happens with humans, as mentioned earlier the same scenario is applicable to these machines M1 and M2. We definitely tend to choose the machine with the optimum solution and the remaining one will be discarded. Intelligent machines exist to serve a specific purpose. This term has been named as artificial selection because; nothing in the entire process is natural as this is a brain-child of humans. Just like us, intelligent machines too, have a variety and they too have evolved just like us during these years. But we would like to point out a difference in this evolution process, human beings have evolved gradually and through something which abides to the laws of nature or we choose to call those laws as *Natural Constraints* which are not subjected to frequent changes but the way computer scientists, engineers and other aspirants are working on providing machines with human like intelligence, the development rate of such components is massive. The constraints which these intelligent machines subject themselves to can be altered in less time as compared to the natural constraints and that is one of the several reasons why this massive growth in the development of intelligent machines should be monitored. This can be explained with a graphical representation mentioned below:

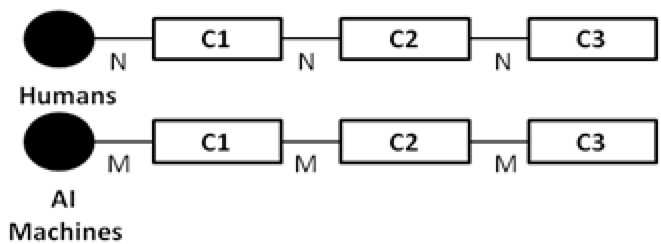


Fig. 2. Characteristic Appending

In the illustration mentioned above, we have compared the addition of characteristics in both machines and human beings. Where C, C2 and C3 are characteristics and N and M are number of years (for simplicity they can be thought of as number of generations) such that:

$$\mathbf{M < N}$$

The quantity of features or characteristics being added to artificially intelligent machines generally will take lesser time (or generations) as compared to a human being. One of the important things that we'd like to mention in this study is regarding the extinction phenomenon in machines. In the case of humans and other living organisms this phenomenon occurs due to competition and as we have mentioned earlier in this paper that artificial constraints are much flexible relative to natural constraints machines too can appear to be competing with each other, in the diagram mentioned above we initially focused on the machine that provided an optimal solution, in such a case the other machine was discarded. Now, if this discarded continues to perform poorly as compared to its other artificial intelligent peers, then it will be a similar scenario for that machine like we as loving organisms have faced over these generations i.e. extinction. But it won't cause any disappearance of any variety of such artificially intelligent machines which is why it is more apt to call this phenomenon as functional-extinction because the subject is merely not functioning but is still in existence.

VI. CONCLUSION

The objective of this study was to depict a comparison between natural intelligence and fully functional, developed artificial intelligence. The nature of this study is philosophical and it discusses the philosophy of artificial intelligence. The study was primarily inspired by the works of Charles Darwin, we in this study aimed to find the similarities in the natures and journeys of human beings and artificial intelligence from the past and present and attempted to provide a futuristic perspective. Right now artificial intelligence is in its infancy but the amount of work and innovations being carried out in this field are enormous, there is a possibility that such advanced computer programs can be written that will be as dynamic and complex as human beings, the future is unpredictable which is why it is a must to cover as many dimensions of a topic as possible. In this study we saw the similarities and difference those humans and machines posses. We gave the concept of generation theory and artificial selection which we firmly believe will help the futurists in understanding the philosophical aspects of artificial intelligence.

ACKNOWLEDGMENT

We would like extend our uttermost gratitude towards Amity Institute of Information and Technology, Amity University as the underlying research was greatly supported by the institution at all times. Our sincere thanks to the head of the department, who consistently kept us motivated. And all others who were directly and indirectly a part of this study.

REFERENCES

[1] Vernon D. Furlong D. (2007). Philosophical Foundations of AI. In: Lungarella M., Iida F., Bongard J., Pfeifer R. 50 Years of Artificial Intelligence. Lecture Notes in Computer Science, vol 4850. Springer, Berlin, Heidelberg, 53-62

[2] Müller, Vincent C. (2013), Philosophy and theory of artificial intelligence (SAPERE, 5; Berlin: Springer). 429

[3] Kosidichimma Anyanwu (2011). Overview and Applications of Artificial Intelligence. Owerri, Nigeria: University of Nigeria

[4] Strenberg, R. J., & Grigorenko, E. L. (2002). Intelligence Applied. New York: Oxford University Press

[5] Margaret A. Boden, (1995). Modeling creativity – A reply to reviewers. Artificial intelligence 79, 161-182

[6] Paul Walton (2018). Artificial Intelligence and the Limitations of Information: Special Issue AI and the singularity: A Fallacy or a Great Opportunity? Information 2018, 332

[7] Nath, Rajakishore. (2009). Philosophy of Artificial Intelligence. Universal Publishers 85-103

[8] Darwin, C. (2004). On the origin of species, 1859. Routledge.

[9] Erb, Benjamin. (2016). Artificial Intelligence & Theory of Mind. 10.13140/RG.2.2.27105.71526.

Optimal Stock Allocation for an Investment Portfolio using Machine Learning Methods

Shreshtha Sarkar[1], Inteshab Nehal[2]

[1,2]Data Scientist, Sprinriver Technology Pvt Ltd, India
[2]inteshabnehal@gmail.com

Abstract: One of the major dilemmas faced by investors is the way of stock allocation for their investment portfolio. The price trend of the financial market is perplexing and is affected by different factors. The investigation of the stock market draws attention from financial specialists and analysts. In this paper, we have followed a top-down approach of trading by choosing the sectors followed by its stocks based on the performance of their fundamental indicators using clustering then computed the various technical indicators and applied Principal Component Analysis and forward feature selection algorithm on it to find the most significant feature affecting the stock price movement. Regression model is used to predict the stock price movement. This data acts as an input to draw a vector of expected returns and the variance-covariance matrix, which gives us the optimal stock allocation weights.

Keywords: Stock prediction, Technical Indicators, Feature Selection, Polynomial Regression, Machine Learning, Support Vector Machine

I. INTRODUCTION

Stock market is very unpredictable and fascinating. There are some basic thumb rules for allocating resources in stock market, however sometimes during short term trading, there are chances we aren't able to invest the right amount of money in it. Various stock prediction models exists using macroeconomic factors [1]. One needs to have an understanding of various theories and methods to analyze information regarding the stock market. There are broadly two types of indicators: fundamental and technical. The advanced development in the field of machine learning and artificial intelligence has drawn attention to find its use in the field of finance. Researchers have used Support Vector Machines, Artificial Neural Network, Deep Learning models to predict stock prices in the past [2]. Machine learning algorithms can easily recognize patterns and predict future price trends which are used to evaluate the relationship between a dependent variable and one or more independent variables. In this paper, we suggest a top-down approach of trading wherein we start by selecting the sectors which are outperforming the market and choose their stocks based on the performance of their fundamental indicators. Further, their technical indicators are computed and regression model is used to predict the prices after performing feature selection. Once the price trend is predicted for a set of stocks, we form a combination of portfolios and return the optimal one based on the risk appetite [3].

II. DATASET

The fundamental indicators like PE ratio, EPS, P/B, Net Profit, and Return on Equity of companies were scrapped from Screener website and formed into a data frame. The stock prices have been fetched from Yahoo Finances, which comprised of Close, Low, High, Volume price day wise. Using this dataset, the technical indicators were computed. These values of the technical indicators are feed as input to the feature selection and the regression model [4].

III. STOCK SELECTION

1. Sector Selection

Top Down Approach is a principal of investing where the country's macroeconomic factors are taken into consideration first, thereby indicating its market situation. This is followed by choosing the sectors which are projected to outperform the market. Stocks faring well within these selected sectors then become the potential spots for investment. We have followed two steps for screening the stocks. We have selected NIFTY 50 index, which is one of the market indicators in India. Firstly, the sectors indices (NIFTY IT, NIFTY BANK et cetera) with positive correlation between NIFTY50 were selected [5]. Secondly, the return percentages of the selected indices are checked with the NIFTY50's return values and are selected for further processing accordingly.

2. SVM Classification

All the values of fundamental indicators form the input variables for classification modeling. The output column '1' denotes the stock should be selected. This labeling has been done on the basis of its market performance and market study. Support Vector Machine Classifier was used to choose the stocks from an array of several stocks per industry-wise. The stocks lying closest to the hyperplane were chosen. This was achieved by using the Density Function of the SVM Kernel [6].

IV. MODELING

1. Feature Selection

Feature selection model is needed because the processing of high dimensional data can be very challenging and such an algorithm removes redundant features, PCA is used for further dimensionality reduction [7]. Regression can easily overfit to such high dimensional data [8], having fewer dimensions also adds to the less computational time lesser variance and more interpretable model. The algorithm uses the cross-validation model while training the model [9]. The features which occur in 3 or more time in the 5 cross-validation system are chosen in the final training dataset. Since we have a dynamic data set that is values being generated on a daily basis, we may need to run the forward feature selection algorithm a several times in order to discard a few fluctuations in the selection of features in the final training data set.

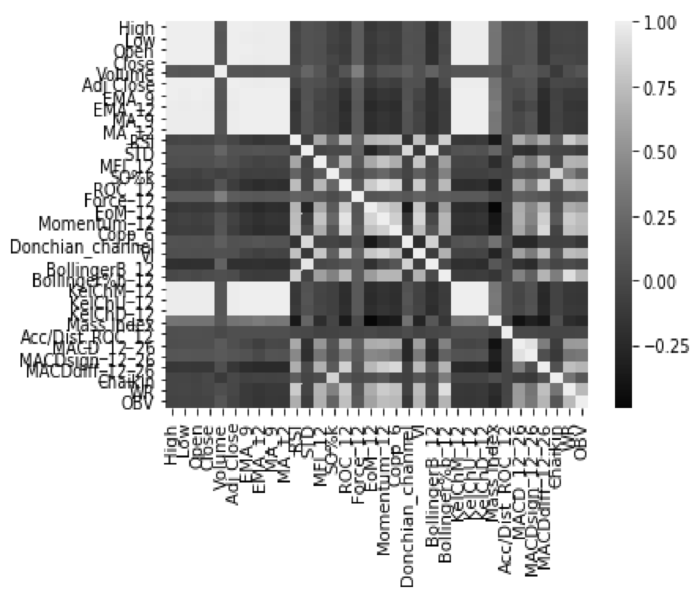


Fig. 1. Correlation heat map plot of the entire feature set i.e. the technical indicators

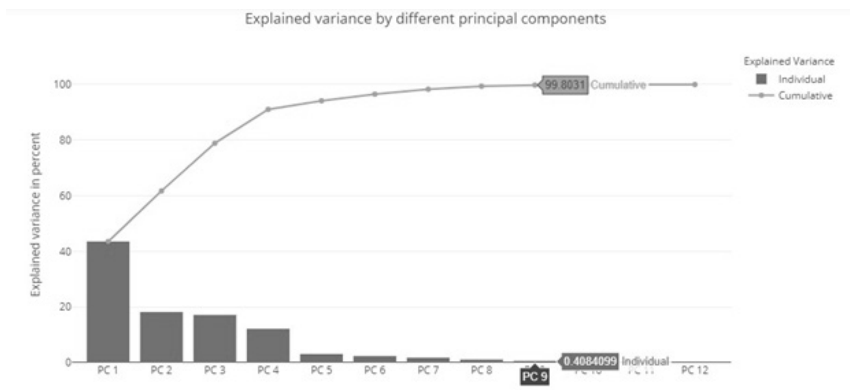


Fig. 2. Principal Component Analysis Graph representing the variance explained by the components

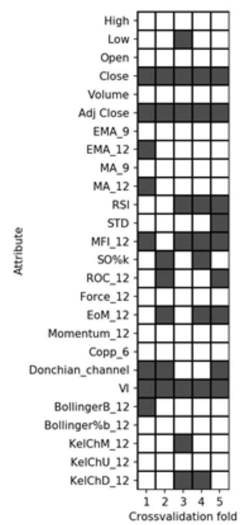


Fig. 3. Features selected after forward feature selection and cross validation

2. Polynomial Regression

Regression is used to determine relationship between target and one or more feature variables. Here we have chosen close price as the target variable and technical indicators as the feature variable. A polynomial model is used because the closing price and the technical indicators are curvilinearly related. On gradual increment of the degree of polynomial the complexity increases. It is used to overcome under fitting, which was a drawback of the simple regression model, at the same time overfitting of the model should be avoided.

Polynomial model can be expressed as

$$Y_i = \beta_0+\beta_1 x_i^1+\beta_2 x_i^2+\ldots\ldots+\beta_k x_i^k+\epsilon_i \quad , \text{for } i= 1, 2, 3\ldots\ldots n,$$

Where k is polynomial regression's degree, n is the number of technical indicators [10].

V. RESULT

Companies from different sectors are chosen to diversify the portfolio. Hence, algorithm's results were: BANK-HDFC, FINANCE-BAJAJ FINANCIAL SERVICES, FMCG- ITC LTD, PHARMACEUTICALS- MERCK, and IT-TCS. The historical prices were taken from 01-01-2016 to 28-02-2019 to train and test the model.

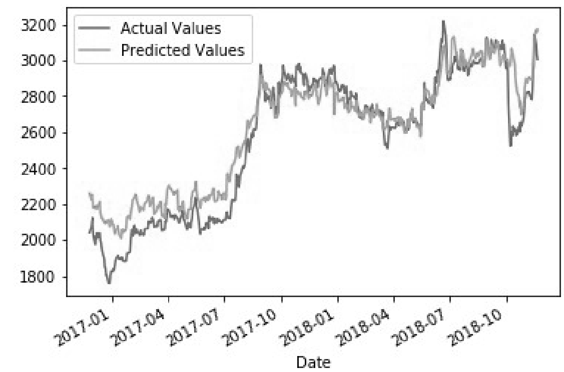


Fig. 4. Plot between actual and predicted values from polynomial regression

In Fig. 4, The R^2 value was 0.96 and the MSE was 0.00238 for a particular stock's close price prediction using polynomial regression. Based on the optimal return points of the output given in the Fig. 5, on virtual trading we have generated a return of 4.4 % in a week. Similarly, for 28 days trading, it gives a return of 7.2 % keeping other factors constant. The return will change according to the user's discretion of risk preference.

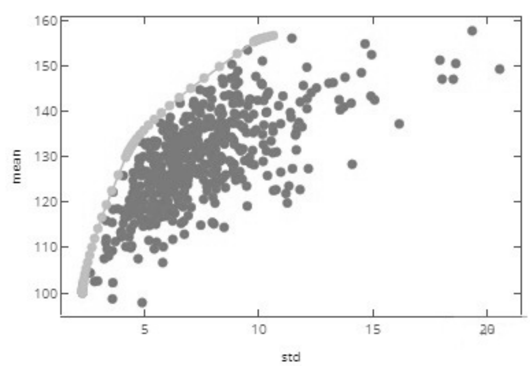


Fig. 5. Return vs. Risk Curve

Each dot in the graph represents different weight-age allocation to different stocks. Yellow dots represent the optimal weight-age allocation to different stocks so that we get the best return for a given risk.

VI. CONCLUSION

The primary objective of the paper revolves around choosing and allocating optimal weights in stocks to get a return based on the risk appetite of an investor. We have achieved this goal by using the combined knowledge of machine learning techniques and financial sciences. The top-down approach of investing may be implemented in various other ways. For instance, even clustering algorithms might be used for selecting stocks in every sector. The feature selection algorithm shows that the technical indicators which affect the price differ from stock to stock. The results which were obtained after adopting feature selection model and dataset normalization showed improved RMSE score of the regression model.

To understand the price movement of stocks with high precision requires incorporation of as much information as possible. Training and hyperparameter tuning of advanced models like ARIMA, LSTM, and GAN generally result in improved models for time series problems like price prediction of stocks, but they are difficult to train. Stocks are also drastically affected by the current news. NLP transfer learning models like BERT may be used to act as an indicator of the price movement.

REFERENCES

[1] Kapoor, V. and Khedkar, N. (2017). *Predicting Volatility in the S&P 500 through Regression of Economic Indicators*.Soe.rutgers.edu. Retrieved February 13, 2019, from https://pdfs.semanticscholar.org/0615/4197da9ce044464e7fbc898d1d2db6be499d.pdf.

[2] M, H., E.A., G., Menon, V. and K.P., S. (2018). NSE Stock Market Prediction Using Deep-Learning Models. *Procedia Computer Science*, 132, pp.1351-1362.

[3] Eychenne K., Martinetti S. and Roncalli T. (2011), Strategic Asset Allocation, Lyxor White Paper Series, 6, Retrieved February 13, 2019, from www.lyxor.com.

[4] Blume, L., Easley, D., O'Hara, M. Market Statistics and Technical Analysis: The Role of Volume. The Journal of Finance (1994).

[5] Mela, C. and K. KOPALLE, P. (n.d.). *The impact of collinearity of regression analysis: the asymmetric effect of negative and positive correlations*.

[6] Das SP, Padhy S (2012) Support vector machines for prediction of futures prices in Indian stock market. Int J Comput Appl 41(3): 22–26.

[7] Bell D. and Wang H., A Formalism for Relevance and its Application in Feature Subset Selection. Machine Learning Vol. 41(2) (2000) 175–195.

[8] Iain M. Johnstone[1] and D. Michael Titterington, Statistical challenges of high-dimensional data (November 2009).

[9] Sylvain Arlot and Alain Celisse, A survey of cross-validation procedures for model selection (July 2009).

[10] Eva Ostertagováa, Modelling using polynomial regression (2012), Elsevier Ltd. Selection and/or peer-review under responsibility of the Branch Office of Slovak Metallurgical Society at Faculty of Metallurgy and Faculty of Mechanical Engineering, Technical University of Košice.

Detect and Estimate: Simple and Efficient baseline for Human Pose Estimation

Pratishtha Verma[1]*, Rajeev Srivastava[2]

[1]Indian Institute of Technology, Banaras Hindu university, Varanasi, India
[2]Indian Institute of Technology, Banaras Hindu University, Varanasi, India
[1]pratishthaver.rs.cse17@iitbhu.ac.in
[2]rajeev.cse@iitbhu.ac.in

Abstract: The Human pose estimation from images has been a demanding challenge of computer vision nowadays. In this paper, our objective is to make a simple and efficient system for human pose estimation with two modules: the human detector and its pose estimation. First, we propose an aggregated architecture with SVM and CNN to make an accurate human detector. Then, a modified Fully Convolutional Google network has been proposed to make the final prediction. we have examined both the modules of the architecture on the popular publically available dataset like INRIA person dataset, LSP and MPII pose estimation dataset. The method gives an impressive performance on these datasets as compared to other states of the art methods.

Keywords: Human Pose Estimation (HPE), Convolution Neural Network (CNN) and Deep Learning (DL).

I. INTRODUCTION

HPE is a technique which determines the joint coordinates of the body like the ankle, wrist, knee, head, hip, and shoulder. The CNN recently highly utilized to improve the performance of computer vision tasks. HPE is a case, in which the joint location is predicted from the image, CNN is largely utilized for the same as [1][2]. Irrespective of the huge improvements occur due to these fascinating model design and procedures, the problem seems to be that there is no perfect representation of features that easily differentiate the visual information and the different manifestations of human account. The whole condition has been changed after the evolution of deep learning in the field of computer vision. The CNN architectures have the ability to gather more and efficient image cues. Like in [3]**,** CNN has utilized to process features among all scales to largely acquire the spatial information of the body for joint prediction. But accurate joint location prediction is difficult because the human body is complex due to its articulation by occlusion, foreshortening, change in viewpoint and body limbs. Due to the evaluation of deep learning, most of the recent techniques makes the model complex only to improve the performance or result. In spite of impactful result they have, they are compute-intensive and have a highly complex architecture which requires a very high configuration graphics card.

This paper presents an efficient technique for HPE that utilize less GPU memory and give good comparable result. The main focus of the architecture is to easily learn the feature from multiple layers like [13]**,** as this make system efficient towards difficult body context joints like ankles and wrists. So, we do multiple layers feature integration for making system robust, we all know that the last layer features are efficient for classification purpose not for localization due to pooling and middle layers are for localization. There are many techniques in the literature which follows the same flow as ours by taking a detection system used for extracting human ROI and then apply pose estimator to predict the joint coordinates. The advantage of using this procedure is their proficiency in arranging a task into multiple subtask for making it easier and more accurate. If the human detector is excellent in detecting the hard candidates, the HPE will usually get accurate with a focused regression space. Our technique gives state of the art results on most of the popular datasets of human detection and pose estimation.

II. RELATED WORK

In computer vision, this topic is highly motivated and improve the performance by utilizing current techniques [7]. The classical approach for human detection uses handcrafted features like HOG, Gradient, wavelet, etc. Few handcrafted features are ACF (Active contour feature) [8]. After the evolution of CNN, the task of detection, localization and classification becomes easy. The few popular techniques which utilizes the deep learning procedure are [5][9]. Currently, [9] introduced an CNN based deep network along with early and late fusion using multi-spectral image data to improve the performance.

The HPE has been intensively studied in recent decades. In early days, the graphical models based on tree structures like pictorial structure models [22] was highly utilized for HPE. CNN based techniques are highly utilized for the estimation in last few years. First, Toshev et al. [10] proposed an CNN based architecture to regress the body joint coordinates for belief map. Similarly, [11] utilize the CNN for HPE.

III. PROPOSED WORK

We go along with detection followed by estimation like detecting the region of interest having human as subject and then estimating the location of the respective body joints. The detail architecture for both individual steps are given below, respectively.

1. Detection Module

For human detection, we follow the architecture design mentioned in Figure 1. First, we apply preprocessing in which we do data augmentation as discussed below in section 4. Second, we apply basic machine learning strategy of feature extraction and classification, which identifies the region that might have human. Then finally the detected outputs have been utlized by the CNN based deep learning to make more accurate classification.

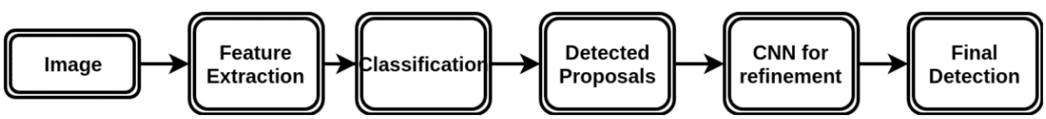


Fig. 1. The architecture of Detection module.

1) Feature Extraction and Classification

We extract HOG (Histogram of oriented gradient) and LBP (Local Binary Pattern) features from the images. The main purpose to use HOG is that it is view invariant and global feature. The reason for using LBP is that it is texture-based feature, uniform and rotation invariant. The mathematical explanation of the steps involved in HOG computation is as follows:

Both the horizontal and vertical gradient computation are:

$$g_h(m,n) = I(m+1,n) - I(m-1,n) \tag{1}$$

$$g_v(m,n) = I(m,n+1) - I(m,n-1) \tag{2}$$

Gradient pixel(m,n) have amplitude and direction as:

$$g(m,n) = \sqrt{g_m(m,n)^2 + g_n(m,n)^2} \tag{3}$$

$$\Theta(m,n) = \tan^{-1}(g_m(m,n)/g_n(m,n)) \tag{4}$$

The steps require to compute LBP with NP number of pixels over ra radius are given below:

np: np^{th} pixel neighbor

g_{cp}: central pixel gray value

g_{np}: np^{th} neighbor gray value

$$LBP_{NP,ra}(m,n) = \sum_{np=0}^{NP} X(g_{np} - g_{cp})2^{np} \tag{5}$$

Here X(q) was utilized to threshold:

$$X(q) = \begin{cases} 1 & q \geq 0 \\ 0 & q < 0 \end{cases} \tag{6}$$

The circular pixel does not contain the np^{th} neighbor. That's why the computation of gray value is as follows:

$$g_{np} = I(m_{np}, n_{np}) \tag{7}$$

np = 0, 1,...NP-1 and

$$m_{np} = m + ra\sin(2\pi np/NP) \tag{8}$$

$$n_{np} = n - ra\cos(2\pi np/NP) \tag{9}$$

The modified rotational invariant LBP is as follows:

$$LBP^{rin}_{NP,ra}(m,n) = \begin{cases} \sum_{np=0}^{NP-1} X(g_{np} - g_{cp}) & if V(m,n) \leq 2 \\ z+1 & otherwise \end{cases} \tag{10}$$

$$V(m,n) = \sum_{np=1}^{NP} |X(g_{np} - g_{cp}) - X(g_{np-1} - g_{cp})| \tag{11}$$

The resulted feature descriptor has been utilized with lib-SVM for detection. After that the detection proposals are again utilized by the CNN based classification model as described below.

2) Pre trained CNN based model for making detection accurate

VGG deep architecture [24] have been utilized for the pre-trained network. The module has total thirteen convolution, three FC and one logistic regression layer. They have 5 max-pooling of 2×2 after the 2^{nd}, 5^{th}, 7^{th}, 11^{th} and 13^{th} layers. ReLU was used for non-linearity of 3 × 3.

The organisation of filters is as follows:

1 and 2 layers have 64 filters
3 and 4 layers have 128 filters
5,6, and 7 layers have 256 filters
8 to 13 layers have 512 filters
14 to 15 layers have 4096 filters
16 layers have 1000 filters

Then at last there is a soft-max function. In the VGG deep network, we apply some changes. First, the network is retrain on the INIRIA dataset. Second, the twoclass classification is required (c=2) for identifying human or nonhuman rather than thousand classes of Imagenet. The sixteenth layer and softmax was modified respectively.

The network was fine-tuned with INIRIA dataset. The fine-tune hyperparameter is as follows, 10 number of epochs, 0.001 learning rate, 100 batch size and 0.9 momentum. For testing, the SVM detector was applied over 288 test images and the resulted output was run by the pretrained CNN module for final classification.

2. Deep Architecture for HPE

In this case, we have to predict the body joint location coordinate. The proposed network is based on the Inception deep network [23]. We freeze the last stages of the network like pooling, drop-out, linear and soft-max layers and starting all layers have been remain the same. To collect multi-layer information, we combine the feature map of multiple inception output layers shown in Figure 2 with the help of deconvolution filter with stride 2 and of size 2 × 2.

The resulted feature map have low resolution because of pooling as compared to input image. So, the produced feature map immediately upsampled with a deconvolution filter of stride 16 and of size 32×32 to get all joints belief maps. At last dropout, dense, batch normalization and sigmoid function utilized for normalizing and regulazing the network.

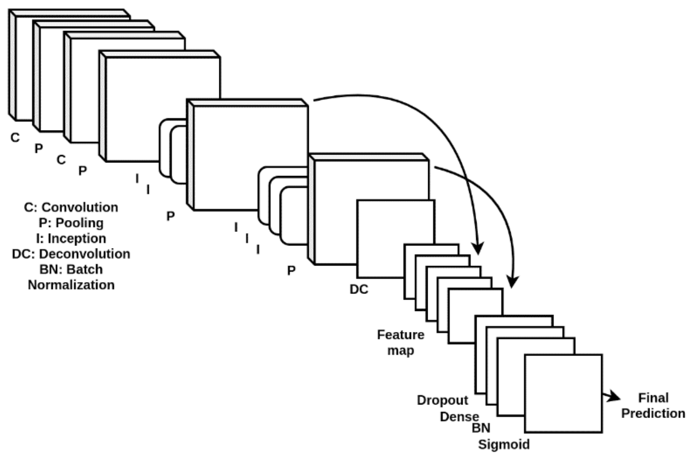


Fig. 2. The multi-layer architecture with Fully Convolutional GoogleNet.

IV. EXPERIMENTS

1. INIRIA Person Dataset

INIRIA person dataset have 1832 training images from which 1218 are negative images and 614 are the positive images. The dataset have 288 test images. Here we build both training and validation set. To make the positive class set, the positive train ground truth bounding box has been utilized resulting in 1237. After that horizontal flipping was applied resulting in a set of 2474.Then, translation and scaling was utilized with a range of [0,5], results in a final set of size 4948. To make negative train set, we randomly extract the windows from the images using [9]. The final set have 12552 negative samples. The whole procedure give 175000 samples from which 15754 is utilized for training and 1749 for validation. The fine-tune hyperparameter is as follows for detection module, 10 number of epochs, 0.001 learning rate, 100 batch size and 0.9 momentum.

2. MPII and LSP Dataset

The dataset have total 40,000 images, from which 25,925 utilized for training and 2,958 for validation. The dataset gives an approximate location of the human for both train and test images. The LSP HPE dataset have 10000 train and 1000 test images.

For HPE, we do the training of the network using the merged dataset of MPII and LSP for 73 epochs, 8mbatch size and learning rate of 0.00092. The quantitative outputs of the proposed module for MPII dataset are given in Table 2 and for LSP given in Table 3. The proposed method outperforms the other state-of-the-art methods.

Table 1. Some Evaluated Metrics on INIRIA Person Dataset

<table>
<tr><td rowspan="7">INIRIA DATA-SET</td><td>Metric</td><td>HOG+LBP</td><td>HOG+LBP+CNN</td></tr>
<tr><td>TP</td><td>530</td><td>545</td></tr>
<tr><td>FN</td><td>37</td><td>42</td></tr>
<tr><td>FP</td><td>1282</td><td>248</td></tr>
<tr><td>FPS</td><td>14.12</td><td>3.20</td></tr>
<tr><td>MR%</td><td>15.80</td><td>14.13</td></tr>
</table>

Table 2. PCKh metric at 0.5 for MPII Dataset

Method	Shoulder	Wrist	Head	Elbow	Ankle	Knee	Hip	PCKh@0.5
Ours	93.7	81.0	97.1	86.2	73.8	80.8	86.4	85.5
Tompson et al. [15]	91.9	77.8	96.1	83.9	64.8	72.3	80.9	82.0
Pishchulin et al. [17]	90.2	77.3	94.1	83.4	68.6	75.4	82.6	82.4
Pishchulin et al. [12]	49.0	34.1	74.3	40.8	35.2	34.4	36.5	44.1
Lifshitz et al. [18]	93.3	80.4	97.8	85.7	70.2	76.6	85.3	85.0
Tompson et al. [13]	90.3	72.4	95.8	80.5	62.8	69.7	77.6	79.6
Hu et al. [16]	91.6	76.6	95.0	83.0	69.5	74.5	81.9	82.4
Carreira et al. [14]	91.7	72.4	95.7	81.7	66.4	73.2	82.8	81.3

Table 3. PCKh metric at 0.2 for LSP Dataset

Method	Shoulder	Wrist	Head	Elbow	Ankle	Knee	Hip	PCKh@0.2
Ours	82.8	71.8	95.6	74.8	74.4	78.4	84.6	80.34
Wang et al. [19]	57.1	36.7	84.7	43.7	50.8	52.4	56.7	54.6
Fan et al. [20]	75.2	64.0	92.4	65.3	70.4	68.3	76.7	73.0
Chen et al. [21]	78.2	65.5	91.8	71.8	63.4	70.2	73.3	73.4
Tompson et al. [13]	79.2	63.4	90.6	67.9	64.2	71.0	69.5	72.3
Pishchulin et al. [12]	56.7	38.0	87.2	46.7	52.7	57.5	61.0	57.1

V. CONCLUSION

The advantage of using the proposed procedure is their proficiency in arranging a task into multiple subtask for making it easier and more accurate. If the human detector is excellent in detecting the hard candidates, the HPE will usually get accurate with a focused regression space. The pose estimator architecture is not as complex other state of the art methods based on deep learning. So, the overall system act as a simple and efficient for HPE. We have examined both the modules of the architecture on the popular publicly available dataset like INRIA person dataset, LSP and MPII pose estimation dataset. The method gives an impressive performance on these datasets as compared to other states of the art methods.

REFERENCES

[1] Alp Güler, R., Neverova, N., & Kokkinos, I. (2018). Densepose: Dense human pose estimation in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7297-7306).

[2] Bulat, A., & Tzimiropoulos, G. (2016, October). Human pose estimation via convolutional part heatmap regression. In *European Conference on Computer Vision* (pp. 717-732). Springer, Cham.

[3] Newell, A., Yang, K., & Deng, J. (2016, October). Stacked hourglass networks for human pose estimation. In *European conference on computer vision* (pp. 483-499). Springer, Cham.

[4] Tompson, J. J., Jain, A., LeCun, Y., & Bregler, C. (2014). Joint training of a convolutional network and a graphical model for human pose estimation. In *Advances in neural information processing systems* (pp. 1799-1807).

[5] Ouyang, W., & Wang, X. (2013). Joint deep learning for pedestrian detection. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2056-2063).

[6] Wagner, J., Fischer, V., Herman, M., & Behnke, S. (2016, April). Multispectral Pedestrian Detection using Deep Fusion Convolutional Neural Networks. In *ESANN*.

[7] Benenson, R., Omran, M., Hosang, J., & Schiele, B. (2014, September). Ten years of pedestrian detection, what have we learned? In *European Conference on Computer Vision* (pp. 613-627). Springer, Cham.

[8] Le, T. H. N., & Savvides, M. (2016). A novel shape constrained feature-based active contour model for lips/mouth segmentation in the wild. *Pattern Recognition*, *54*, 23-33.

[9] Wagner, J., Fischer, V., Herman, M., & Behnke, S. (2016, April). Multispectral Pedestrian Detection using Deep Fusion Convolutional Neural Networks. In *ESANN*.

[10] Toshev, A., & Szegedy, C. (2014). Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1653-1660).

[11] Chen, Y., Shen, C., Wei, X. S., Liu, L., & Yang, J. (2017). Adversarial posenet: A structure-aware convolutional network for human pose estimation. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1212-1221).

[12] Pishchulin, L., Andriluka, M., Gehler, P., & Schiele, B. (2013). Strong appearance and expressive spatial models for human pose estimation. In *Proceedings of the IEEE international conference on Computer Vision* (pp. 3487-3494).

[13] Tompson, J. J., Jain, A., LeCun, Y., & Bregler, C. (2014). Joint training of a convolutional network and a graphical model for human pose estimation. In *Advances in neural information processing systems* (pp. 1799-1807).

[14] Carreira, J., Agrawal, P., Fragkiadaki, K., & Malik, J. (2016). Human pose estimation with iterative error feedback. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4733-4742).

[15] Tompson, J., Goroshin, R., Jain, A., LeCun, Y., & Bregler, C. (2015). Efficient object localization using convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 648-656).

[16] Hu, P., & Ramanan, D. (2016). Bottom-up and top-down reasoning with hierarchical rectified gaussians. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5600-5609).

[17] Pishchulin, L., Insafutdinov, E., Tang, S., Andres, B., Andriluka, M., Gehler, P. V., & Schiele, B. (2016). Deepcut: Joint subset partition and labeling for multi person pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4929-4937).

[18] Lifshitz, I., Fetaya, E., & Ullman, S. (2016, October). Human pose estimation using deep consensus voting. In *European Conference on Computer Vision* (pp. 246-260). Springer, Cham.

[19] Wang, F., & Li, Y. (2013). Beyond physical connections: Tree models in human pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 596-603).

[20] Fan, X., Zheng, K., Lin, Y., & Wang, S. (2015). Combining local appearance and holistic view: Dual-source deep neural networks for human pose estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1347-1355).

[21] Chen, X., & Yuille, A. L. (2014). Articulated pose estimation by a graphical model with image dependent pairwise relations. In *Advances in neural information processing systems* (pp. 1736-1744).

[22] Andriluka, M., Roth, S., & Schiele, B. (2009, June). Pictorial structures revisited: People detection and articulated pose estimation. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 1014-1021). IEEE.

An Insight into Automatic Detection of Epilepsy using EEG Signals

Palak Handa, Nidhi Goel

Indira Gandhi Delhi Technical University for Women, Delhi, India

Abstract: The Epilepsies are a chronic neurological disorder characterized by unprovoked, recurring (similar or different type) seizures which leaves its signatures in electroencephalography (EEG) in the form of sharp spikes. Detection of such seizure activities through EEG signal processing and machine learning algorithms may help in improving the life of an epileptic patient. This paper briefly reviews existing methods of data selection, signal processing methods, feature extraction, its optimization and classification methods on four publically available databases i.e. Bonn, Bern- Barcelona, CHB – MIT, Neurology and sleep centre, New Delhi database for successful automatic epilepsy detection using EEG signals.

Keywords: Epilepsy detection, EEG signal processing methods, Statistical features, Classification techniques

I. INTRODUCTION

Epilepsy is a central nervous system (neurological) disorder characterized by recurring seizures (similar or different type) and hence also called as seizure disorder. It is fairly a common non-communicable disorder affecting one in twenty six individuals worldwide [1]. Different approaches of basic, clinical and translational research of human brain have focused on finding meaningful prediction, detection and therapies for the betterment of individuals affected with epilepsy. One such widely researched detection is of quantitative analysis of seizure and seizure free EEG signals through different signal processing methods and machine learning algorithms. EEG is most often used to diagnose the type of epilepsy [1]. It is portable and cost-effective as compared to other brain imaging techniques however is presently difficult to read. The ictal activity of EEG is observed almost periodic and of high amplitude [2]. This activity is observed due to disturbed activity of neurons present in our brain. It is patient specific and can be observed between 3-32 Hz [2]. This paper briefly reviews existing classification methods for analysis and diagnosis of epilepsy. Figure 1 shows a general framework for analysis of EEG signals for automatic detection of epilepsy.

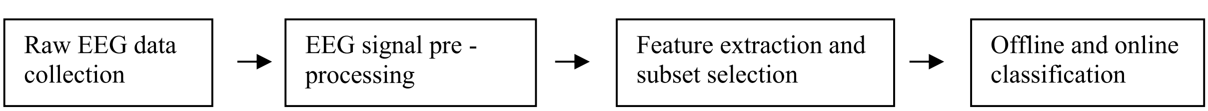


Fig. 1. Framework of epilepsy detection using EEG signals

This paper covers the following points i.e. data collection, signal processing methods, feature extraction, its sub-set selection and classification techniques. The brief review results have been analyzed on two open databases namely University of Bonn, Germany [3] and Children's Hospital, Boston [4]. In addition, Database of Bern Barcelona [5], and Neurology and sleep centre, Delhi [6] have also been mentioned in this paper. Existing pre-processing methods of EEG and other imaging modalities like artifact removal, montage selection, noise detection, rejection and selection of frequency band range etc remains in the proof of the principle stage. With increasing research and awareness, detection of epilepsy will be done with the help of computer aided systems in the future.

II. RELATED WORK AND ITS COMPARISON

This section presents brief discussion of each stage for analysis and detection of epilepsy namely data selection, signal processing methods, feature extraction optimization and classification methods.

1. Data Selection

Bonn EEG time series database: This database comprises of 100 single channels EEG of 23.6 seconds with sampling rate of 173.61 Hz. Its spectral bandwidth range is between 0.5 Hz and 85 Hz. It was taken from a 128 channel acquisition system. Five patients EEG sets were cut out from a multi channel EEG recording and named A, B, C, D and E. Set A and B are the surface EEG recorded during eyes closed and open situation of healthy patients respectively. Set C and D are the intracranial EEG recorded during a seizure free from within seizure generating area and from outside seizure generating area of epileptic patients respectively. Set E is the intracranial EEG of an epileptic patient during epileptic seizures. Each set contains 100 text files wherein each text file has 4097 samples of 1 EEG time series in ASCII code. A band pass filter with cut off frequency as 0.53 Hz and 40 Hz has been applied on the data. It is an artifact free data and hence no prior pre-processing is required for the classification of healthy (non-epileptic) and un-healthy (epileptic) signals. The strong eye movement's artefacts were omitted.

- Bern – Barcelona EEG database: This multichannel EEG database was recorded using intracranial strip and depth electrodes manufactured by AD Tech Medical, Racine, WI, USA. It is of five patients with longstanding pharma-coresistant temporal lobe epilepsy. They were sampled at either 512 or 1024 Hz based upon whether they were recorded with less or more than 64 channels of EEG system. The patients underwent epilepsy surgery. Three out of five attained complete seizure freedom. Two types of EEG are present in the data i.e. focal and non-focal. It has about 10240 samples for duration of 20 seconds sampled at 512 Hz.
- Children's hospital Boston – MIT database: This database comprises of 844 hour continuous scalp multi-channel EEG recording of 23 pediatric patients above age 18. The patients were kept under anti- seizure medication to check their chances of undergoing an epilepsy surgery. A total of 163 seizures were recorded at sampling frequency of 256 Hz in 10-20 bio-polar montage of 18 channels EEG. Each EEG segment is called as a record which usually is for duration of one hour. There are 9 - 42 edf files from a single subject. Additional vagal nerve stimulus signals are also present. Separate file name and montages have been mentioned for seizure v/s non seizure episode in EEG segments.
- Neurology and sleep centre, New Delhi EEG dataset: This database comprises of 5.12 seconds EEG data. It was recorded using 57 EEG channel Grass Tele-factor Comet AS40 Amplification System; sampled at 200 Hz. It's spectral bandwidth range is between 0.5 Hz and 70 Hz. Time series EEG datasets are categorized into three major MATLAB file folder namely ictal, pre-ictal and inter-ictal stages. Each MAT file has 1024 samples. The complete dataset is not publically available.

A comparison of above four known datasets in this research domain is mentioned in Table 1.

Table 1: Brief comparison of available EEG datasets

Source	Status	Size	Channel	No. of Patients	Sampling Freq.	Sets
University of Bonn, Germany [3]	Publically available (2001)	3.05 MB	100 single channel	5	173.6 Hz	5 (Z, O, N, F, S)
CHB MIT [4]	Publically available (2009)	2-40 GB	18 EEG channel	23	256 Hz	-
The Bern Barcelona [5]	Publically available (2012)	814 MB	-	5	512 Hz	2 (F & N)
Neurology and sleep centre, Delhi [6]	Publically available (2016)	604 KB	57 channels	10	200 Hz	3 (Ictal, pre-ictal inter-ictal)

It was noticed that information about the placing of EEG electrodes was missing in all publically available databases. None of the database has mentioned about change of impedance with time of the EEG electrodes. This information could help in deciding about the quality of the EEG signal. Fewer studies have been done on the variation observed in the EEG bands and sub-bands of the available database [7]. Raw data is not available for 1, 2 and 4 databases. Some of the researches have applied pre-processing techniques on the processed data of University of Bonn which might have lead to degradation in signal information. Most of the databases available have been recorded for a short duration of time with no or less information given about stage of epilepsy in the affected patient. University of Bonn database have not mentioned about montages, specific epilepsy type, proper information about the patients and seizure affected brain area though it has become a benchmark for research due to its availability and has helped in escalating research in this domain. Children's hospital Boston – MIT database is a heavy and complex database and may occupy about 2- 40 GB depending on the number of patients taken in consideration. Age, gender, occurrence of seizure and its duration has been mentioned. Discussion about artifact present, noise, and seizure affected area of brain has not been clearly mentioned. All the patients have focal epilepsy. The recorded seizure episodes are of inter-ictal phase. Neurology and sleep centre database have not disclosed about patient's age, gender, occurrence of seizure and its duration and discussion about artifact present, noise, and seizure affected area of brain

According to [1], for a clinical routine testing at least 32 channel EEG is advised where clinical range is between 0 to 30 Hz. Depending on the seizure affected area of brain, channel selection can be done prior to feature extraction step [8]. There are two broad categories of seizures and not all can be diagnosed using an EEG [1].

It was found that many people prefer to use publically available database where clearly, database of University of Bonn is an accepted choice amongst research groups. This paper has mentioned about some personal databases in Table 2. More publically available database generation is much needed for a universal classification of different types of epilepsy and seizures which would eventually lead to successful detection and management of this disease.

2. EEG Signal Processing Techniques

Signal processing technology is a vast field which is used for artifact removal, montage selection, noise detection, rejection, PSD analysis and selection of frequency band range etc. These remain in the proof of the principle stage and tremendous research work is being done in this field. An EEG signal can be decomposed into five frequency bands namely delta, theta, alpha, beta and gamma. It can be analysed in namely four domains i.e. time domain, frequency, joint time – frequency domain and sparse domain [9]. Different signal processing has been applied on different databases and mentioned in table no 2. It was found that wavelet decomposition and empirical mode decomposition [10] is a well-liked choice amongst existing EEG signal decomposition techniques [42]. Though both of these techniques have several demerits and may introduce imaginary data points in the reconstructed signal. Daubechies (db4) wavelet is more preferred in comparison to other wavelet families in field of EEG signal processing [11]. It was observed that use of Principle component analysis (PCA), Independent component analysis (ICA) techniques has declined in the field of EEG signal processing because it requires a reference signal and may not necessarily produce fruitful results. Tunable Q factor wavelet transform (TQWT) has also gained popularity and is used as a signal decomposition technique [12]. Research works have used machine learning techniques in channel detection, and selection of bands of EEG [8, 13]. Most of the papers calculate features from the reconstructed signal matrix of approximation and detailed coefficients of the discrete wavelet transform (DWT) and other techniques such as ICA, EMD, and TQWT etc.

3. Feature Extraction and its Optimization

Feature extraction helps in finding meaningful and distinguishing characteristics between different EEG signals. It directly dominates the final accuracy in statistical classification of epilepsy. Hence, selecting features that best describe the behavior of EEG signals are important for an optimum automated process. Optimization techniques are also applied on developed feature matrix for its sub-set selection to obtain best accuracies. Features based on time domain are mean, median, standard deviation, kurtosis, Hurst exponents, temporal etc. Feature based on amplitude of the EEG signal includes inter-quartile range and root mean square. Variance, standard deviation, inter-quartile range and root mean square (RMS) etc are some of the features based on the measure of dispersion in a signal. Features based on frequency domain include fractal dimensions, spectral roll off, spectral flux etc. Most researches have used combination of features called as hybrid features after early 2010s. It is still debatable whether a single feature matrix or hybrid matrix has the potential to generate 100% accuracy. In this paper's opinion concept of hybrid feature matrix generation is more promising for the training of the classifier. After the comparative study, it was found that different entropy feature such as spectral entropy and Shannon's entropy in combination with other features boasts the accuracy rates [14]. Combination of Time – frequency based features are more preferred than separate time and frequency domain based features [15]. No one feature has been universally accepted and only a few were followed by performance based statistical testing like ANOVA, Kruskal Wallis test etc and more research can be done in this domain.

4. Classification Techniques

Machine learning techniques are changing the world and saving lives in different ways. It focuses on recognizing patterns and work on vital information from a given data. A manual analysis of long EEG recording to find traces of epilepsy is not only time taking but also requires skilled technicians. Hence, an automated process with the help of machine learning algorithms can be of great clinical significance. It was found that support vector machines (SVM) are the most popular choice amongst classification of binary and trio sets of seizure free and seizure containing EEG signals. However, researches have used different kernels in SVM and obtained a maximum of 99% true classification rate. This observation is not only contradictory but also makes it difficult to decide on which kernel is the best in the detection of epilepsy. More research is needed in this domain. Such successful rate achievement could be due to less epileptic data availability or overtraining of the classifier. Random forest classifier and K – nearest neighbour classifier are also a popular choice for the same. Few researches carried out using deep learning such as convolution neural networks (CNN) are trained on raw dataset instead of statistically developed features. Such techniques hold a promising future in the detection of epilepsy [16].

Table 2. Comparison of reported results of four datasets

Author(s)	**Year**	**Signal Processing Techniques**	**Feature Used**	**Classification Method**	**Accuracy**
Epileptic Seizure Detection Method on University of Bonn Database					
Essam H. et al. [17]	2019	DWT (4 levels)	DWT features	WOA and SVM with RBF	100%
Katerina et al. [18]	2019	DWT (5 levels)	DWT features	Random forest classifier	95%
Vijay Anand [19]	2019	DWT, Haar wavelets	Entropy features of DWT matrix	Generalized regression neural network	100%
V. Gupta., R. Pachori [20]	2019	Fourier - Bessel series expansion	Weighted multi scale Renyi permutation entropy	LS –SVM (Morlet wavelet)	99.5%
S Anand, S Jaiswal [21]	2019	Discrete stationary wavelet based stock well transform	Temporal and spectral feature, Amplitude distribution estimation	Hybrid K nearest SVM	100%
Debdeep Sikdara, Rinku Roy [22]	2018	DWT (db4)	Time domain features	SVM with cubic polynomial kernel function	99.70%
A. Gupta et al. [23]	2018	Multi-rate filter bank using DCT	Hurst exponent and ARMA parameters	Binary SVM	-
Mohd Hamza et al. [24]	2018	EMD	IMFs	ANN	96.1- 99.3%

Author(s)	Year	Signal Processing Techniques	Feature Used	Classification Method	Accuracy
Hüseyin Göksu [25]	2018	WPD (3 levels)	Log energy, norm entropy and its energy	MLP with back propagation	100%
David, Sridha [26]	2018	-	Temporal features	LSTM network (Deep learning)	95.54%
Md Mursalin et al. [27]	2017	DWT (db4)	Improved correlation-based feature selection (ICFS) applied on time domain, frequency domain features, entropy and DWT matrix	Random forest classifier	98.45%
A. Jaiswal et al. [28]	2017	-	1- D local gradient pattern, histogram based features	Decision tree, ANN	99.82%
Reddy and Rao [12]	2017	TQWT	Centered correntropy	Random forest classifier, MLP	98.30% 98.20%
Mingyang Li et al. [29]	2017	DWT (db4)	Mean, energy, standard deviation, maximum value	Neural network ensemble	98.70%
Torse, Khanai [30]	2017	TQWT	Spectral, Shannon, Kraskov entropy	Random forest classifier	97.30%
Wang, Huang [31]	2017	Wavelet threshold method (db4)	IMF through EMD, approximate entropy, did PCA	SVM , 10 fold cross validation	99.25%
Subasi, Canba [32]	2017	DWT (db4)	Genetic algorithm on DWT features	Hybrid SVM	99.38%
Boubchir et al [15]	2017	-	Time based, frequency based and time-frequency based	ANN, KNN	99.33%
Ali Yener Mutlu [33]	2017	Hilbert vibration decomposition	Mean IF features	RBF kernel LS-SVM	97.66%
Manish et al [34]	2017	Analytic time-frequency flexible WT, BPF	Fractal dimensions	Least square – SVM (LS – SVM)	98.67%
S. at al [35]	2017	TQWT	Kraskov entropy	Least square – SVM (LS – SVM)	97.75
Fenglin Wang et al. [36]	2014	-	Cluster coefficient	Time series complex networks	94.50%
Kai Fu [37]	2014	HHT	Time domain features of pixel intensity in histogram	Support vector machine (SVM) with RBF	99.13%
Varun Bajaj & R. Pachori [38]	2012	EMD	AM and FM Bandwidths	LS – SVM (Morlet wavelet kernel)	99.5%

Author(s)	Year	Signal Processing Techniques	Feature Used	Classification Method	Accuracy
Mirzaei et al. [39]	2011	DWT	Normalized Shannon and spectral entropy	K-Nearest Neighbor	98.50%
Epileptic Seizure Detection Method on Bern Barcelona Database					
N. Arunkumar [40]	2018	-	Sample entropy and Fuzzy entropy	Non Nested Generalized Exemplers	99%
Rajeev Sharma et al. [41]	2017	TQWT	Fractal dimensions, different type of entropies, Lyapunov exponent	Least square – SVM (LS – SVM)	95%
Anindya Das [42]	2016	Combined EMD – DWT	Log-energy entropy	K-nearest neighbor	89.40%
Epileptic Seizure Detection Method on CHB – MIT Database					
Xinghua et al. [43]	2019	-	Temporal and spatial features	Recurrent neural network	87%
Diyuan et al. [44]	2019	DWT	Entropy features	Residual Deep CNN	91.8 %
Sandeep et al. [45]	2019	-	Mean, variance, skewness, kurtosis, line length, energy, and band power	Ensemble learning, KNN, RUS Boost	94-99%
Javad B. [8]	2017	Random forest (Channel selection)	Frequency domain features	KNN classifier	89.8%
A. Bhattacharyya [46]	2017	EWT	Multivariate Time–Frequency features	RF, Naive-Bayes, K-NN	97.91%
Kaveh Samiee [47]	2016	Rational DSTFT decomposition	Sparse rational decomposition and the Local Gabor Binary Patterns	1. Logistic regression 2. Random forest 3. Linear-kernel SVM	98 .85 98 .62 98 .89 (F1%)
Epileptic Seizure Detection Method on Neurology and Sleep Centre, Delhi Database					
A. Gupta et al. [48]	2018	Multi rate filter bank using DCT	Hurst exponent and ARMA parameters	Binary SVM	-
P. Swami et al. [49]	2017	Comparative study of DWT, WPT and dual-tree complex WT	Combinations of energy, root-mean-square values and standard deviations	RNN	DTCWT coefficients > 98%
P. Swami et al. [50]	2017	Coiflets' WP	Wavelet features	5-Nearest Neighbour	99.3515 ± 0.25 %
Epileptic Seizure Detection Method on Private Database					
N. Jmail [51]	2018	Stationary WT	Gamma reconstruction and spatio temporal mapping	-	-

Author(s)	Year	Signal Processing Techniques	Feature Used	Classification Method	Accuracy
Jesus Martinez-del-Rincon [52]	2017	DWT (db4)	Bag-of-Words or wavelet components	SVM RBF	84.72%
Sevda [53]	2017	Systemic sampling	Time and frequency domain based	K-Nearest Neighbor	95%
Mohammad-Parsa Hosseini [54]	2016	Noise removal filtering and db4	I-ICA selected time and frequency based components	Random Subspace Ensemble	95%
Jasmin Kevric [55]	2014	Multi scale PCA De-noising	PSD, MSPCA components	Decision tree (C4.5)	99.59%

Table 2 clearly shows different combinations of signal processing techniques, statistical feature matrices and machine learning algorithms have yield an accuracy of more than 95% but their reproducibility remains uncertain on real-time databases. Most of them may fail due to added complexity of the database in terms of its duration, artefacts, different seizure types (pseudo seizures) etc. Development of the database is itself a challenging task. Spike pattern recognition of different seizures remains in state of the art principle due to high non-linearity, stochastic, non-deterministic, and non-stationary nature of EEG signals. Similarly, Data over-fitting, classifier choice and the problem of black box remains three chief obstacles in machine learning.

III. CONCLUSION

Diagnosis, treatment and management of epilepsy have become a challenging task for the community. It's detection by visual introspection of long hour EEG is not only time taking but a very tedious and subjective task. Artificial intelligence may help in escalating this process and lead to successful detection of different types of epilepsies. This paper was focused on reviewing different frameworks of publically available four EEG datasets. The main goal of this brief review was to understand the available methods; hurdles present at every stage of the framework and adopt new methods in the detection of epilepsy using EEG signals. After extensive literature review, it was concluded that statistical analysis of EEG signals along with the in-depth study of ontology of specific epilepsy demography, types and its characteristics can prove to be a way to overcome some of the hurdles in the detection of epilepsy.

REFERENCES

[1] Dumanis, S.B. (1st ed. 2016) *Epilepsy - A Giving Smarter Guide*.

[2] Bou Assi, E., Nguyen, D., Rihana, S., & Sawan, M. (2017). Towards accurate prediction of epileptic seizures: A review. *Biomedical Signal Processing And Control*, *34*, 144-157. doi: 10.1016/j.bspc.2017.02.001

[3] Phys. Rev. E, 64, 061907 and Andrzejak RG, Lehnertz K, Rieke C, Mormann F, David P, Elger CE (2001) *Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity.*

[4] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE (2003). *PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals*. Circulation. 101(23):e215-e220

[5] Phys. Rev. E, 86, 046206, 2012, Andrzejak RG, Schindler K, Rummel C. *Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients.*

[6] Swami, Piyush & Panigrahi, Bijaya & Nara, Sanjeev & Bhatia, Manvir & Gandhi, Tapan. (2016). *EEG Epilepsy Datasets*. 10.13140/RG.2.2.14280.32006.

[7] Ji Zhong, Qin Shuren, Peng Chenglin (2008), Study on Separation for the Frequency Bands of EEG Signal and Frequency Band Relative Intensity Analysis Based upon EMD, in *Proceedings of the 7th WSEAS International Conference on Signal Proccessing, Robotics and Automation (ISPRA '08)*, University of Cambridge, UK.

[8] Birjandtalab, J., Baran Pouyan, M., Cogan, D., Nourani, M., & Harvey, J. (2017). Automated seizure detection using limited-channel EEG and non-linear dimension reduction. *Computers In Biology And Medicine*, *82*, 49-58. doi: 10.1016/j.compbiomed.2017.01.011

[9] Krishnan S., and Athavale, Y. (2018), Trends in biomedical signal feature extraction, *Biomedical Signal Processing and Control*, vol. 43, pp. 41-63. Available: 10.1016/j.bspc.2018.02.008.

[10] Pachori, R. B. (2008), Discrimination between Ictal and Seizure-Free EEG Signals Using Empirical Mode Decomposition, *Research Letters in Signal Processing*, vol. 2008, Article ID 293056, 5 pages. https://doi.org/10.1155/2008/293056.

[11] Scolaro, G. R., Boos, C. F., Azevedo, F. and Walz, R. (2013), *Wavelet Filter to Attenuate the Background Activity and High Frequencies in EEG Signals Applied in the Automatic Identification of Epileptiform Events*. INTECH Open Access Publisher.

[12] Ravi Shankar Reddy, G., & Rao, R. (2017). Automated identification system for seizure EEG signals using tunable-Q wavelet transform. *Engineering Science And Technology, An International Journal*, *20*(5), 1486-1493. doi: 10.1016/j.jestch.2017.11.003

[13] Nigam, V. and Graupe, D. (2004), A neural-network-based detection of epilepsy, *Neurological Research*, vol. 26, no. 1, pp. 55-60. Available: 10.1179/016164104773026534.

[14] Acharya, U., Fujita, H., Sudarshan, V., Bhat, S., & Koh, J. (2015). Application of entropies for automated diagnosis of epilepsy using EEG signals: A review. *Knowledge-Based Systems*, *88*, 85-96. doi: 10.1016/j.knosys.2015.08.004

[15] Boubchir, L., Daachi, B. and Pangracious, V. (2017). A review of feature extraction for EEG epileptic seizure detection and classification. *40th International Conference on Telecommunications and Signal Processing (TSP).*

[16] Acharya, U., Oh, S., Hagiwara, Y., Tan J., and Adeli, H. (2018), Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, *Computers in Biology and Medicine*, vol. 100, pp. 270-278. Available: 10.1016/j.compbiomed.2017.09.017.

[17] Houssein, E., Hamad, A., Hassanien, A., & Fahmy, A. (2019) Epileptic detection based on whale optimization enhanced support vector machine, *Journal of Information and Optimization Sciences*, 40:3, 699-723, DOI: 10.1080/02522667.2018.1453671

[18] Tzimourta, K.D., Tzallas, A.T., Giannakeas, N. et al. Health Technol. (2019) 9: 135. https://doi.org/10.1007/s12553-018-0265-z

[19] Vijay Anand, S. & Shantha Selvakumari, R. Soft Computer (2019) 23: 2645. https://doi.org/10.1007/s00500-018-3630-y

[20] Gupta, V. and Pachori, R. (2019). Epileptic seizure identification using entropy of FBSE based EEG rhythms. *Biomedical Signal Processing and Control*, 53, p.101569.

[21] Anand, S., Jaiswal, S., & Ghosh, P. (2018). Epileptic Seizure Detection in EEG Signal using Discrete Stationary Wavelet-Based Stockwell Transform. *Majlesi Journal of Electrical Engineering*, *13*(1), 55-63. Retrieved from http://www.mjee.org/index/index.php/ee/article/view/2771

[22] Sikdar, D., Roy, R. and Mahadevappa, M. (2018). Epilepsy and seizure characterisation by multifractal analysis of EEG subbands. *Biomedical Signal Processing and Control*, 41, pp.264-270.

[23] Gupta, A., Singh, P. and Karlekar, M. (2018). A Novel Signal Modeling Approach for Classification of Seizure and Seizure-Free EEG Signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(5), pp.925-935.

[24] Shaikh, M., Farooq, O. and Chandel, G. (2018). EMD Analysis of EEG Signals for Seizure Detection. *Lecture Notes in Electrical Engineering*, pp.189-196.

[25] Göksu, H. (2018). EEG based epileptiform pattern recognition inside and outside the seizure states. *Biomedical Signal Processing and Control*, 43, pp.204-215.

[26] D. Ahmedt-Aristizabal, C. Fookes, K. Nguyen and S. Sridharan, Deep Classification of Epileptic Signals, (*2018) 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, 2018, pp. 332-335.

[27] Mursalin, M., Zhang, Y., Chen, Y. and Chawla, N. (2017). Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. *Neurocomputing*, 241, pp.204-214.

[28] Jaiswal, A. and Banka, H. (2017). Local pattern transformation based feature extraction techniques for classification of epileptic EEG signals. *Biomedical Signal Processing and Control*, 34, pp.81-92.

[29] Li, M., Chen, W. and Zhang, T. (2017). Classification of epilepsy EEG signals using DWT-based envelope analysis and neural network ensemble. *Biomedical Signal Processing and Control*, 31, pp.357-365.

[30] Torse, Dattaprasad & Desai, Veena & Khanai, Rajashri. (2018). A review on seizure detection systems with emphasis on multi-domain feature extraction and classification using machine learning.

[31] Wang, L., Xue, W., Li, Y., Luo, M., Huang, J., Cui, W. and Huang, C. (2017). Automatic Epileptic Seizure Detection in EEG Signals Using Multi-Domain Feature Extraction and Nonlinear Analysis. *Entropy*, 19(6), p.222.

[32] Subasi, A., Kevric, J. and Abdullah Canbaz, M. (2017). Epileptic seizure detection using hybrid machine learning methods. *Neural Computing and Applications*, 31(1), pp.317-325.

[33] Mutlu, A. (2018). Detection of epileptic dysfunctions in EEG signals using Hilbert vibration decomposition. *Biomedical Signal Processing and Control*, 40, pp.33-40.

[34] Sharma, M., Pachori, R. and Rajendra Acharya, U. (2017). A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension. *Pattern Recognition Letters*, 94, pp.172-179.

[35] Patidar, S. and Panigrahi, T. (2017). Detection of epileptic seizure using Kraskov entropy applied on tunable-Q wavelet transform of EEG signals. *Biomedical Signal Processing and Control*, 34, pp.74-80.

[36] Wang, F., Meng, Q., Chen, Y., & Zhao, Y. (2014). Feature Extraction Method for Epileptic Seizure Detection Based on Cluster Coefficient Distribution of Complex Network.

[37] Fu, K., Qu, J., Chai, Y. and Dong, Y. (2014). Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM. *Biomedical Signal Processing and Control*, 13, pp.15-22.

[38] Bajaj, V. and Pachori, R. (2012). Classification of Seizure and Nonseizure EEG Signals Using Empirical Mode Decomposition. *IEEE Transactions on Information Technology in Biomedicine*, 16(6), pp.1135-1142.

[39] Mirzaei, Ahmad & Ayatollahi, Ahmad & Motie Nasrabadi, Ali. (2011). Automated detection of epileptic seizures using mixed- methodology: Wavelet-Chaos-KNN Classifier-Mutual Information. 87.

[40] Arunkumar, N., Ram Kumar, K. and Venkataraman, V. (2018). Entropy features for focal EEG and non focal EEG. *Journal of Computational Science*, 27, pp.440-444.

[41] Sharma, R., Kumar, M., Pachori, R. and Acharya, U. (2017). Decision support system for focal EEG signals using tunable- Q wavelet transform. *Journal of Computational Science*.

[42] Das, A. and Bhuiyan, M. (2016). Discrimination and classification of focal and non-focal EEG signals using entropy-based features in the EMD-DWT domain. *Biomedical Signal Processing and Control*, 29, pp.11-21.

[43] Yao, X. and Cheng, Q. (2019). *A Novel Independent RNN Approach to Classification of Seizures against Non-seizures*. [ebook] Available at: https://arxiv.org/pdf/1903.09326.pdf.

[44] Lu, D. and Triesch, J. (2019). *Residual Deep Convolutional Neural Network for EEG Signal Classification in Epilepsy*. Available at: https://arxiv.org/pdf/1903.08100v1.pdf

[45] Banerjee S., Alur V., Shah D. (2019) Patient Nonspecific Epilepsy Detection Using EEG. *Advances in Intelligent Systems and Computing*, vol 900. Springer, Singapore.

[46] Bhattacharyya, A. and Pachori, R. (2017). A Multivariate Approach for Patient-Specific EEG Seizure Detection Using Empirical Wavelet Transform. *IEEE Transactions on Biomedical Engineering*, 64(9), pp.2003-2015.

[47] Samiee, K., Kovács, P. and Gabbouj, M. (2017). Epileptic seizure detection in long-term EEG records using sparse rational decomposition and local Gabor binary patterns feature extraction. *Knowledge-Based Systems*, 118, pp.228-240.

[48] Gupta, A., Singh, P. and Karlekar, M. (2018), A Novel Signal Modeling Approach for Classification of Seizure and Seizure-Free EEG Signals, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 5, pp. 925-935. Available: 10.1109/tnsre.2018.2818123.

[49] Swami, P., Gandhi, T., Panigrahi, B., Bhatia, M., Santhosh, J. and Anand, S. (2016). A comparative account of modelling seizure detection system using wavelet techniques. *Systems Science: Operations & Logistics special edition on Modelling and Simulation in Healthcare Systems*, pp. 1—12. DOI: 10.1080/23302674.2015.1116637.

[50] Swami, P., Gandhi, T., Panigrahi, B., Bhatia, M., Santhosh, J. and Anand, S. (2016, February 11—12). Locating ictal activities over human scalp with automated detection using EEG signals. *IEEE Proceedings of International Conference on Signal Processing and Integrated Networks*, Noida, India. DOI: 10.1109/SPIN.2016.7566767.

[51] Jmail, N., Zaghdoud, M., Hadriche, A., Frikha, T., Amar, C. and Bénar, C. (2018), Integration of stationary wavelet transform on a dynamic partial reconfiguration for recognition of pre-ictal gamma oscillations, *Heliyon*, vol. 4, no. 2, p. e00530. Available: 10.1016/j.heliyon.2018.e00530.

[52] J. Martinez-del-Rincon et al. (2017), Non-linear classifiers applied to EEG analysis for epilepsy seizure detection, *Expert Systems with Applications*, vol. 86, pp. 99-112. Available: 10.1016/j.eswa.2017.05.052.

[53] Gül, S., Uçar, M., Çetinel, g., Bergil, e. and Bozkurt, m. R. (2017), Automated Pre-Seizure Detection for Epileptic Patients Using Machine Learning Methods, *International Journal of Image, Graphics and Signal Processing*, vol. 9, no. 7, pp. 1-9. Available: 10.5815/ijigsp.2017.07.01.

[54] Mohammad-Parsa Hosseini, Abolfazl Hajisami, and Dario Pompili (2016), Real-time Epileptic Seizure Detection from EEG Signals via Random Subspace Ensemble Learning, in *IEEE International Conference on Autonomic Computing*.

[55] Kevric, J. and Subasi, A. (2014), The Effect of Multiscale PCA De-noising in Epileptic Seizure Detection, *Journal of Medical Systems*, vol. 38, no. 10. Available: 10.1007/s10916-014-0131-0.

Adversarial Password Guessing—A Review

Sahil Lamba, Suyash Gupta, Nipun Soni and Pranav Dass

Information Technology, Bharati Vidyapeeth's College of Engineering, New Delhi, India

Abstract: The generation of passwords is a very common security issue. Users face the problem of picking good passwords and ensuring that their code is as safe as possible. It was Hashcat and John the Ripper which they tried to compare. We can also measure the effect on the ability of a user's Password Evaluation based on random permutations and language laws. Users can monitor trillions of passwords per second against password hazards with conventional passphrase devictions such as "Hash-Cat and John the Ripper". The use of password generation rules and simple dictionary attacks can expand password dictionaries to this form. Adversarial Password Guessing uses a Generative Adversarial Network (GAN) instead of manual password analysis to learn the distribution of valid passwords autonomously from existing password leaks and generate high-quality password assumptions.

Keywords: Hashcat, Genetic Adversarial network, PassGan, Dictionary Attack

I. INTRODUCTION

Passwords are the most common encryption tool, particularly as they are not difficult to implement, do not require any dedicated hardware or software, and users and developers are familiar with the concept as well. Sadly, many breaches of the password database revealed that users tend to pick passwords$ that are easy to envision, [4] consisting primarily of standard strings (e.g., username, 123456, iloveyou), or variations of popular strings.

Most password cracking programs take advantage of this human tendency by dictionary attacks that try to match the hash of a given password with some accuracy from a list of widely used passwords. More advanced programs will work with a series of language and grammar rules to extend the search space of the competitor beyond a specified code list. Nevertheless, these mechanisms are still based primarily on a series of language and grammar rules, so it becomes more difficult to crack when a user chooses a code outside these rules.

The goal of this project is to explore the possibilities of predicting passwords using machine learning. In fact, how well a machine learning system fits with the generic code guessing methods of today with tools like hashcat and its dictionary attack framework will be studied. Finally, the findings will be addressed, assessing the usefulness of the system and how it could be improved and also what practical implications such a model can have for password intensity calculation and password cracking / devaluation.

While these heuristics are reasonably successful in practice, they are ad-hoc and based on intuitions about how users select passwords instead of being constructed from a simple analysis of large password data sets. For this reason, in the end, that technique is restricted to collecting a particular subset of code space which relies on the logic behind that technique. In contrast, the development and testing of new rules is a process requiring advanced knowledge, time and therefore minimal scalability. [5]

II. LITERATURE REVIEW

1. Genetic Adversarial Networks (GAN)

Generative Adversarial Networks (GANs) translate to biased machine learning the current progress of (implicit) generative computation in deep neural networks. The aim of GANs is to obtain samples from the same distribution as S={ x1, x2,. xn}. Generational modeling is typically based on expressions of a closed type that, in many situations, cannot catch the complexity of actual data. The GANs are learning a deep-neural G generative network that produces the optimal distribution by inputting multi-dimensional random sample z (from the Russian or uniform distributions).GANs transform the density assessment problem into a binary classification problem, where G parameters are established by relying on a profoundly discriminating network D to distinguish between the "real" Samples and the "fake" samples created by G.

2. Hashcat and John The Ripper

John the Ripper and HashCat are two commonly used password guessing methods in use today, extending dictionary attacks by developing password conversion rules. John the Ripper provides a standard dictionary attack, with behavior-based password extension rules for choosing passwords from compromised repositories seen in user accounts. HashCat also offers a similar dictionary attack, but for its implementation uses a decentralized approach through GPUs. In this study, we use HashCat specifically to contrast adversarial methods focused on neural networks. Unfortunately, both John the Ripper and Hashcat leverage simple, hard-coded password patterns based on the developer's ability, which requires review of the password layout. Perhaps if we use deep learning and have the computer learn from scratch the complexities and trends and maybe pick up more complex code patterns that people don't notice? One of these methods was from$, which used neural networks to model the power of passwords, and was able to be competitive to state-of - the-art effectiveness. [9] PassGAN, which is the subject of this article, is another machine-based learning approach to code guessing. PassGAN uses adversarial generative networks (GANs) to produce possible passwords and produces outcomes that are equivalent with generic password guessing devices like John the Ripper and HashCat. Others also reported promising results using other deep learn- ing methods including LSTMs [1] [3] [2].

3. PassGan

The generator network is attempting to produce samples from the data-set as its training data-set S={ x1, x2,..., xn}. Generative modeling is based on expressions of closed form that can not normally catch the complexity of real data. PassGAN trains a generative deep neural network G to produce a sample from a desired password-distribution target that enters a random, multidimensional sample of passwords embedded in a Gussian or uniform distribution. GANs transform the density estimation problem into a binary classification problem where G-parameter knowledge is accomplished by a discriminator depending on it. The discriminator attempts to prevent the generator from producing a password distribution close to the distribution of the intended code while the discriminator attempts to produce a false password distribution and attempts to deceive the discriminator to believe it is a true distribution. [8]

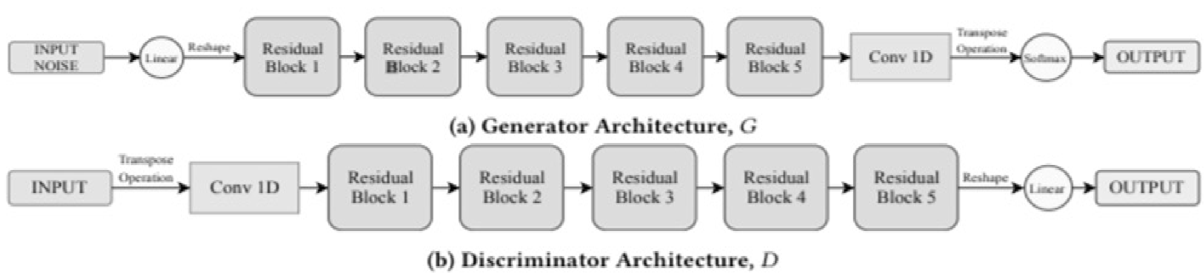


Fig. 1. Passgan architecture

III. METHODS AND IMPLEMENTATION

1. Performance of Different Attack-Styles using Hashcat

There are several ways to break passwords, such as brute force attacks, dictionary attacks and word fusion dictionary attacks. [7] Crack codes are from a mistake in the public database.

Dictionary Attack: When it comes to password cracking, a very simple and highly successful attack is using a list of top-most passwords that were leaked in the past and sorting by how often the password was used, putting the most popular password at the top and the less frequent passwords at the bottom. That's exactly how the dictionary, rockyou.txt, is made. Running hashcat using the rockyou.txt dictionary attack method succeeded in cracking about 15 percent of the hashes of code. Let's look at how many codes have been broken as opposed to how many suspicions have been made. [12]

Dictionary + Rule Set Attack: Another very popular attack, on top of a dictionary, is using a rule-based approach. This method takes each vocabulary word (password) and mutates the word to infer different passwords. Consider the summer base-word, the rule package adds many mutations to the word such as: Summer2017, Summer1234, Summ3r, to list only a few. In this case, about 30 percent of the password hashes are cracked using the rockyou.txt dictionary with a relatively basic rule set called nsa64.rule[NSA16].

Very Large Dictionary Attack: Given the relatively popular rockyou.txt dictionary attack, which is a small list of just over 14 million entries, and the breaking of large passwords, it would be a great candidate to use a huge list. Weakpass 2a.txt is a list of passwords from various leaks in the system and includes 7, 884, 602, 871 passwords.

Such attacks resulted in breaking almost all 100,000 passwords, demonstrating the power 3.1 quality of multiple attack types using hashcat 11 of learning most user- selected passwords and thus the password-selecting actions of humans.

Brute Force Attack: Finally, the very exhaustive model of reliability of a brute force attack has been checked. The brute force attack was configured to guess every combination of characters between the set: [a-z], [A-Z], [0-9] except special characters, from 1-8 character code lengths. For the other approaches, this number is significantly higher than those, meaning it will take a while to determine all the estimates. It was processing hashes at a rate of roughly 47, 000 MH / s for this experiment, resulting in a total runtime of just over 4 hours and 20 minutes. In this sample, 41.8 percent of the hashes were broken by the brute force attack.

2. Comparing the Methods

To illustrate the quality of the different methods of attack, they are compared in a map Figure 2. The graph shows that a probabilistic approach, namely the rockyou.txt dictionary attack, is most successful when it comes to cracking hashes, with the fewest number of guesses. It also struggles to break as many as the other processes, however. The number of passwords broken by adding a rule set to the dictionary attack is doubled, however, requiring a significant additional number of attempts. The big-list proved to be the most effective of these attacks, succeeding in cracking virtually all hashes. Eventually, the brute force attack requires a huge amount of guessing relative to how many passwords it has been able to crack and is therefore much less effective.

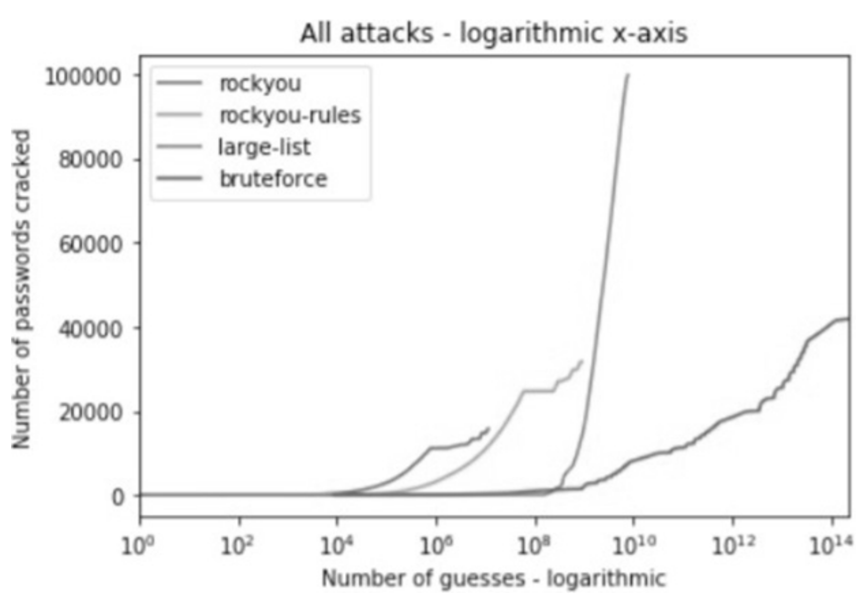


Fig. 2. Comparison of all attacks

IV. PASSWORD SAMPLING PROCEDURE FOR HASHCAT, JTR, MARKOV MODEL, PCFG

The selected portion of the RockYou data set is used as the input dataset for "HashCat Best64, HashCat gen2, JTR Spiderlab rules, Markov Model, PCFG", and created the following passwords:

- Laws of "HashCat and JTR are represented using the practice set's passwords ordered by frequency in descending order (as in [10]). HashCat Best64 produced 754,315,842 passwords, 361,728,683 of which were special and 10 characters or less in size". Remember that "for the given input set, i.e., RockYou training set, this was the maximum number of samples generated by Best64 rule-set. We sampled a random subset of size 109 from their output simultaneously with HashCat gen2 and JTR SpiderLab". That subset consists of 10 characters or less passwords.
- Using the 3-gram Markov method, we created "494,369,794 unique passwords of 10 or less length". The regular setup of this design is used [6].
- Using the PCFG implementation of 109 different passwords of 10 or less size [11].

Approach	(1) Unique Passwords	(2) Matches	(3) Number of passwords required for PassGAN to outperform (2)	(4) PassGAN Matches
JTR Spyderlab	10^9	461,395 (23.32%)	$1.4 \cdot 10^9$	461,398 (23.32%)
Markov Model 3-gram	$4.9 \cdot 10^8$	532,961 (26.93%)	$2.47 \cdot 10^9$	532,962 (26.93%)
HashCat gen2	10^9	597,899 (30.22%)	$4.8 \cdot 10^9$	625,245 (31.60%)
HashCat Best64	$3.6 \cdot 10^8$	630,068 (31.84%)	$5.06 \cdot 10^9$	630,335 (31.86%)
PCFG	10^9	486,416 (24.59%)	$2.1 \cdot 10^9$	511,453 (25.85%)

Fig. 3. The number of matches generated by each password guessing tool against the Rock- You test and the corresponding number of code created by PassGAN to satisfy that tool. Hash- CatBest64 matches are obtained by listing the complete performance of each device exhaustively.

Passwords Generated	Unique Passwords	Passwords matched in testing set
10^4	9738	103 (0.005%)
10^5	94400	957 (0.048%)
10^6	855972	7543 (0.381%)
10^7	7064483	40320 (2.038%)
10^8	52815412	133061 (6.726%)
10^9	356216832	298608 (15.094%)
10^{10}	2152819961	515079 (26.036%)
$2*10^{10}$	3617982306	584466 (29.543%)

Fig. 4. Number of passwords that match passwords in the test set of RockYou generated by PassGAN. In terms of different matches, the scores are displayed.

V. CONCLUSION

Password sequences can be predicted with the aid of machine learning systems, such as artificial neural networks. This also makes it possible to assign a strength to a password depending on how likely the system is to predict the password received. The template can also be used to create passwords that may be selected by humans. Using software like hashcat with a dictionary-style attack, passwords created can then be used to break passwords. The model's created passwords do not very well compare with the existing lists of passwords that have been built over many years. The password intensity estimation property can be used to arrange currently defined password lists by the model's probability of predicting them. The new leaked data sets of code can ultimately benefit significantly and boost Pass- GAN. The model will evolve through the learning of new rules, reducing the number of repeated experiments.

VI. FINAL REMARKS

GANs estimate the training dataset's density distribution. This results in repeated password deviations from the PassGAN output. Whilst a full-scale attack on a brute force would cover everything, learning from the distribution of training data helps PassGAN to attack more effectively, producing highly probable devinations. PassGAN can produce several trillions of guesses in advance and store these in a database, as password generation is possible offline. We also retained specific key samples of our tests and then used them for analysis, avoiding repeat use. If required, Bloom filters can also be used with appropriate parameters, which enable efficient online password deviations, to discard multiple entries.

VII. FUTURE WORK

PassGAN, a new approach that combines theory-driven learning algorithms with human-generating code rules. Rather than relying on manuel password research, "Pass- GAN" uses a "Generative Adversarial Network (GAN)" to learn the distribution of valid passwords and generate high-quality password hypotheses autonomously from estab- lished password leaks. We can train a generative adversarial network with a generator trying to generate guesses to crack a password scheme and a discriminator trying to pre- vent the generator from producing a guess breaching the password protection scheme.

REFERENCES

[1] Binkowski, M., Sutherland, M., Arbel, A., null Gretton: Demystifying MMD GANs (2018)

[2] Cao, Y., Ding, G.W., Lui, Y.C., Huang, R.: Improving GAN Training via Binarized Repre- sentation Entropy (BRE) Regularization. (2018)

[3] Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., Abbeel, P.: Infogan: Inter- pretable representation learning by information maximizing generative adversarial nets. Ad- vances in Neural Information Processing Systems pp. 2172–2180 (2016)

[4] Dell, M., Amico, ., Michiardi, P., Roudier, Y.: Password strength: An empirical analysis. Proceedings IEEE INFOCOM pp. 1–9 (2010)

[5] Denton, S.E.L., Chintala, R., Fergus: Deep Generative Image Models using a Laplacian Pyra- mid of Adversarial Networks. Advances in neural information processing systems pp. 1486– 1494 (2015)

[6] Dorsey, B.: Markov-chain password generator (2017), https://github.com/brannondorsey/ markov-passwords

[7] Gosney, J.: Nvidia gtx 1080 ti hashcat benchmarks (2017), https://gist.github.com/epixoip/a83d38f412b4737e99bbef804a270c40

[8] Hitaj, B., Gasti, P., Ateniese, G., Perez-Cruz, F.: PassGAN: A Deep Learning Approach for Password Guessing (2019)

[9] Melicher, W., Ur, B., Segreti, S.M., Komanduri, S., Bauer, L., Christin, N., Cranor, L.F.: Fast, Lean, Accurate: Modeling Password Guessability Using Neural Networks (2016)

[10] Melicher, W., Ur, B., Sean, M., Segreti, S., Komanduri, L., null Bauer: Fast, Lean, and Ac- curate: Modeling Password Guessability Using Neural Networks, USENIX Security Sym- posium

[11] Weir, M.: Probabilistic Password Cracker (2009), https://sites.google.com/site/reusablesec/ Home/password-cracking-tools/probablistic_cracker.

[12] Wheeler, D.L.: zxcvbn: Low-Budget Password Strength Estimation. Proceedings of the 25th USENIX Security Symposium (2016)

Detection and Segmentation of Brain Tumor using MRI Images

Neda Fatima[1], Salman Ahmad Siddiqui[2], Anwar Ahmad[3]

[1,2,3]Jamia Millia Islamia, New Delhi, India
[2]salman007.rec@gmail.com

Abstract: The technique of Image processing is quite commonly employed, most challenging and upcoming technique, especially in the discipline of medical science in the recent times [1]. Radiology is one particular branch involved with processing of MRI images. This paper depicts the future employable technique that can be used for detecting and extracting brain tumor using the MRI scan of the patient [5][6]. The said technique includes noise removal functions via Perona Malik Diffusion Technique, segmentation and morphological operations [9] that comprise the essential concepts of image processing. Accuracy and Efficiency are critical in such processing techniques as it involves human lives. Thus, computerized images go a long way in confidently achieving these parameters as opposed to manual detection. MATLAB software has been employed for detecting, segmenting and extraction of tumor from MRI scanned brain images.

Keywords: Brain Tumor Detection, MRI, MATLAB

I. INTRODUCTION

Brain is a type of tissue that is enclosed by skull bones and comprises of thin tissue layers called meninges and cerebrospinal fluid. A tumor is an abnormal growth of tissue [10] where cells multiply frenziedly. Tumors could be classified as **Primary or Metastatic**[6] depending on place of origin: Primary implies it had originated in the brain itself whereas Metastatic tumor spreads from any other part of the body to the brain.

It can also be classified as **Benign or Malignant Brain Tumor** [6]**.** When cells do not spread to elsewhere in the body or rarely invade tissues around them, it can be termed as benign. Though, they can be removed and usually do not grow again, however, they can affect sensitive brain areas leading to health problems. On the other hand, malignant brain tumors grow rapidly and crowd into neighboring tissues that are healthy. Their removal is difficult and chances of their reappearance are possible.

In the Radiology discipline of medical sciences, Magnetic Resonance Imaging (MRI) [2] is a critical imaging technique which is deployed to illustrate the anatomy of human body and pictures of physiological processes of the body for depiction of any disease. As the name implies, MRI scanners work with radio waves, strong magnetic fields, and field gradients for depiction of illustrations of the body part under observation. MATLAB software thus used here uses these MRI scans and Image Processing tools to carry out Tumor Detection and Segmentation.

II. REVIEW OF LITERATURE

Various techniques have been researched upon in the field of medical sciences to work on detection of tumor in body parts and their correction. Some have employed threshold segmentation like Li's method and Otsu's method [5]. This paper is based on the Otsu's method [5]. Others have employed watershed segmentation [8] which is a gradual gradient based technique. Filters of different kinds have also been employed. In some cases, low pass filter has been used. Here, an anistropic filter [6] that reduces noises using Perona Malik Diffusion Technique has been employed.

III. METHODOLOGY

1. Block Illustration of the Technique

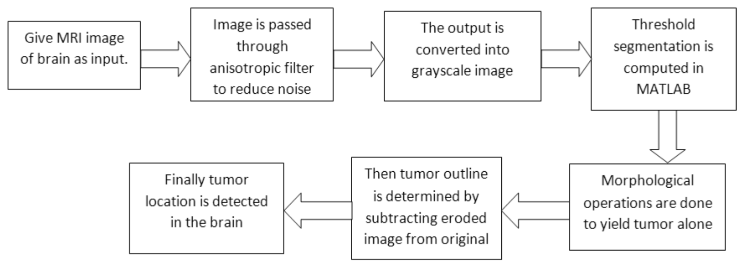


Fig. 1. Illustration of Detection and Segmentation of Brain Tumor Technique

2. Methodology

1) MRI Image of Brain is Fed as Input

From given database of MRIs of real patients, one particular MRI has been selected.

2) Noise Removal by Perona Malik Diffusion Technique

For the image processing applications, anisotropic or Perona Malik diffusion is employed. This helps in decreasing noise of the image without altering the image content such as lines and edges and other detailing that are critical for interpretation of image.

3) Converison to Grayscale

RGB image provides for increased complexity in tumor detection. Hence, with the help of MATLAB functions, the MRI scan has been converted to its grayscale [7] equivalent.

4) Computation of Threshold Segmentation

In threshold segmentation technique, a pixel in the given image is swapped with a black pixel when the image intensity is smaller than constant threshold T. The pixel is replaced with a white pixel when image intensity is higher than T. Thus, the basis of this method is clip level that converts a grayscale image into its corresponding binary image.

5) Computation of Morphological Operation

Morphological operations [6] have been carried out on the semi processed image which helped to obtain further information on solidity and areas of the plausible locations. The structuring element, used to probe an image, is placed in every possible areas in the image and then contrasted to the corresponding neighborhood of the pixel.

6) Tumor Outline is Determined

The tumor has been outlined to specify the designation area, shape and size of the tumor with respect to the brain.

7) Tumour Location Detection

Finally, the location of the tumor in the region of the brain has been identified.

IV. RESULT ANALYSIS

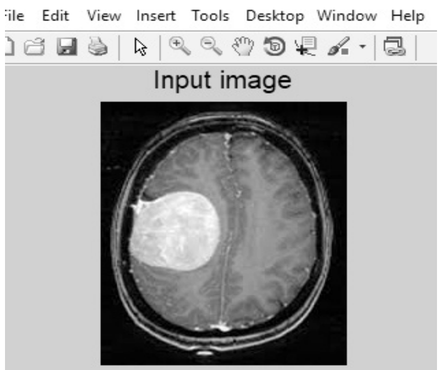


Fig. 2. Input image

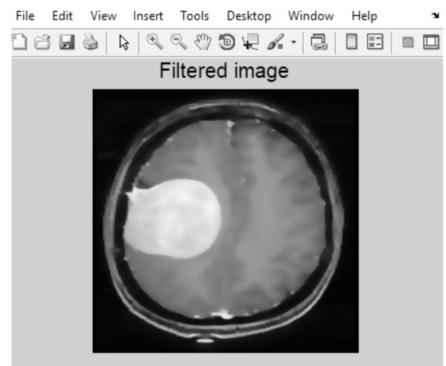


Fig. 3. Filtered image

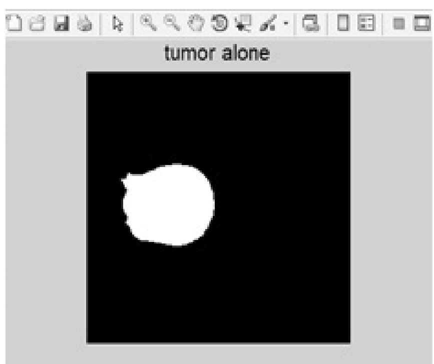


Fig 4. Tumor image

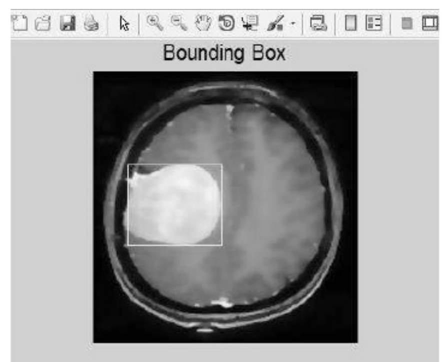


Fig 5. Bounding box

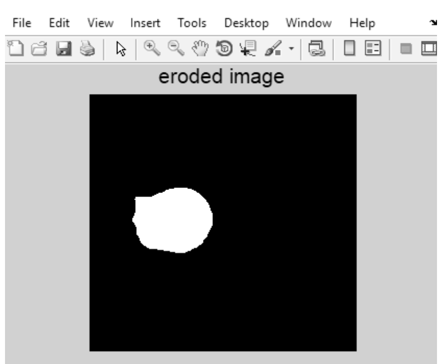


Fig. 6. Eroded Image

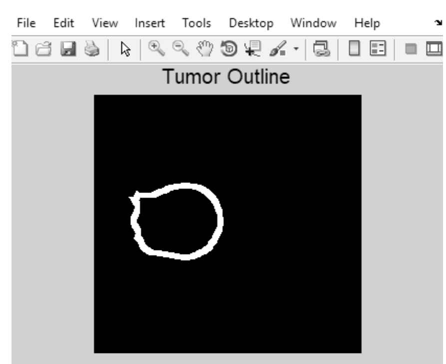


Fig. 7. Tumor outline

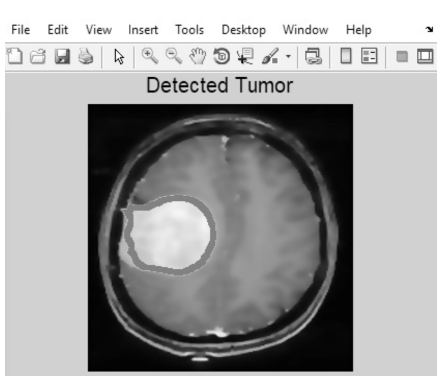


Fig 8. Detected Tumor

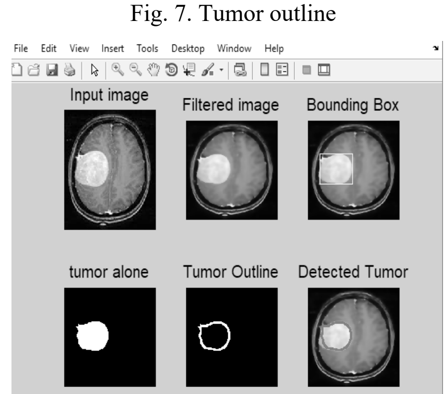


Fig 9. Combined Display

V. CONCLUSION AND FUTURE SCOPE

The results that have been obtained as above help us to locate the accurate position of brain tumor besides segmenting and separating it from the brain images. This indeed helps in improving the accuracy and efficiency of tumor detection and consequently correction, thus saving precious time during treatment. Further improvement of this technique can be done via employing image processing tools to determine the type and nature of tumor and can refer to a cloud database of various feasible option of treatment for different types of tumor. Moreover, further diagnostic techniques can also be then linked to the type of tumor detected by taking inputs of doctors and medical scientists and a database can be generated.

REFERENCES

[1] Deepa & Akansha Singh(2016), "Review Of Brain Tumor Detection From MRI Images", 3rd International Conference On Computing For Sustainable Global Development (Indiacom)

[2] A Reema Mathew & P Babu Anto (2017), "Tumor Detection And Classification Of MRI Brain Image Using Wavelet Transform And SVM", International Conference On Signal Processing And Communication (ICSPC)

[3] Muhammad Nazir & Muhammad Attique Khan & Tanzila Saba & Amjad Rehman (2019), "Brain Tumor Detection From MRI Images Using Multi-Level Wavelets", Brain Tumor Detection From MRI Images Using Multi-Level Wavelets

[4] Md. Rezwanul Islam & Md. Reezbhan Imteaz & Marium-E-Jannat (2018), "Detection And Analysis of Brain Tumor From MRI By Integrated Thresholding And Morphological Process With Histogram Based Method", International Conference On Computer, Communication, Chemical, Material And Electronic Engineering (IC4ME2)

[5] Wassim El Hajj Chehade & Riham Abdel Kader & Ali El-Zaart (2018), "Segmentation of MRI Images For Brain Cancer Detection", International Conference On Information And Communications Technology (ICOIACT)

[6] M. H. O. Rashid & M. A. Mamun & M. A. Hossain & M. P. Uddin (2018), "Brain Tumor Detection Using Anisotropic Filtering, SVM Classifier and Morphological Operation From MR Images", International Conference on Computer, Communication, Chemical, Materials And Electronic Engineering

[7] Miss. Shrutika Santosh Hunnur & Akshata Raut & Swati Kulkarni (2017); "Implementation of Image Processing For Detection of Brain Tumors", International Conference On Computing Methodologies And Communication (ICCMC)

[8] Sukanta Kumar Tulo & Madhusmruti Nayak & Manish Kumar & Khushboo (2017), "Brain Tumor Detection From MRI Image Using Digital Image Processing", International Journal Of Engineering And Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-6 Issue-4

[9] Vishnumurthy T D & Mohana H S & Vaibhav A Meshram (2016), "Automatic Segmentation Of Brain MRI Images And Tumor Detection Using Morphological Techniques", International Conference on Electrical, Electronics, Communication, Computer And Optimization Techniques (ICEECCOT)

[10] Vipin Y. Borole & Sunil S. Nimbhore & Dr. Seema S. Kawthekar (2015), "Image Processing Techniques For Brain Tumor Detection: A Review", International Journal of Emerging Trends & Technology In Computer Science (IJETTCS), Volume 4, Issue 5(2)

Vehicle Number Plate Recognition Via MATLAB

Salman Ahmad Siddiqui[1], Neda Fatima[2], Anwar Ahmad[3]

[1,2,3]Jamia Millia Islamia, New Delhi, India
[1]salman007.rec@gmail.com

Abstract: With the rise in Per Capita Incomes especially in urban areas, affordability of people has risen. Consequently, the number of vehicles driven has risen exponentially. This has led for the demand of increased vehicle security in this era of rising incomes along with rising thefts and crime rate. This paper depicts the employment of Image processing technique for recognition of number plates of incoming and outgoing vehicles using input from CCTV camera images [1][2][9]. The images are processed and can be employed for theft detection and even for traffic monitoring systems for issuing fines or challans in case of overspeeding etc. MATLAB software has been used for detection, character segmentation and finally template detection.

Keywords: Vehicle Number Plate, Crime Monitoring, Traffic Monitoring, CCTV

I. INTRODUCTION

Vehicle Number Plate is a unique identification alphanumeric code that is given to every vehicle that has been registered with the Transport department of that particular city and state. Vehicles can be identified either manually (security guards etc.) or automatically [1] (using closed circuit television sets or CCTVs). Automatic Detection shall help in capturing of images of various vehicles at different places such as parking lots [5], traffic intersections, no parking zones etc. In such a scenario, number plate data from the CCTVs can be utilized to match the unauthorized vehicles or those involved in breaking of traffic rules. Consequently, e- challans can be issued and can be then sent directly to the registered owner of the vehicle or in case of thefts; owners can be notified immediately of their vehicle's location. The Number plates in India have the following format:

Fig. 1. Sample Number Plate

Firstly the country code is specified in blue (India), '2' depicts the state code (Delhi), '3' represents the district/ locality code, '4' represents the type of vehicle, and finally the last 4 digits portray the exact registration number.

The present paper comes up with a technique to be used for the tracking and recognition of vehicle number plate which shall aid in the inspection of vehicle number plates with authorized, unauthorized or stolen numbers.

II. REVIEW OF LITERATURE

Several research papers have been analyzed and examined for writing this paper in order to introduce novelty to the research work. In some papers character recognition is done by Optical Recognition [4] [6] for better translation of image.

Furthermore, Plate localization techniques [7] have been employed to accurately read the number plates with noise reduction in the surrounding region in different papers. Correlation techniques [2] have also been used to determine the vehicle number and match it in the database of penalized/ missing vehicles.

III. METHODOLOGY

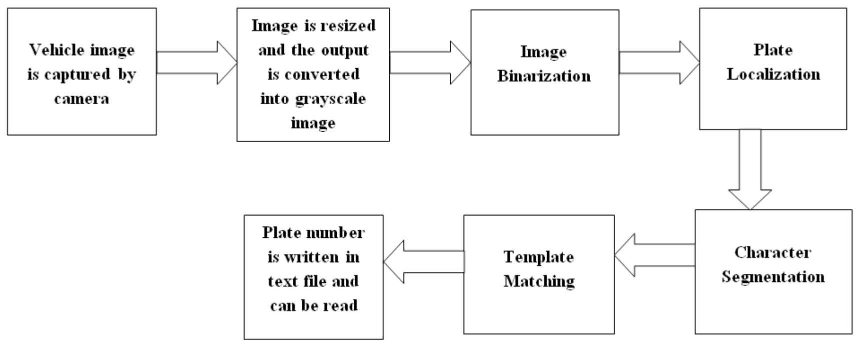


Fig. 2. Block Illustration of Vehicle Number Plate Detection

1. CCTV Image is Fed as an Input

The CCTV camera present at traffic intersection or parking lot captures the image of the number plate of the vehicle. A clear picture of the number plate in white background is fed as an input.

2. Image Resized and Conversion to Grayscale

The images is cropped and resized to only include the number plate and then converted to its grayscale [10] version using 'rgb2gray' to reduce complexities in detection and reading.

3. Image Binarization

The processed image is then converted to its corresponding binary form. 'im2bw' function is employed for this conversion.

4. Plate Localisation

Both figures Fig.5 and Fig.6 are Plate location step images using threshold value and MATLAB toolbox function 'bwareaopen'(). All components with connectivity lower than f pixel are removed to get the actual location of the number plate. 'bwareaopen'(BW,P) function eliminates all the connected components that have a value lesser than P pixels from a binary image. In this case, P=10 pixels thus "IND" in figure is removed. bwareaopen() specifies the expected connectivity. For actual location of the number plate, every component for which the connectivity is less than are removed. After morphological filtering, the four vertex coordinates of the last selected region is taken as the output for the next step and the number on the number plate is extracted.

5. Character Segmentation

The toolbox feature of MATLAB delivers 'regionprops()' function. For each marked area in the tag matrix, it computes a group of properties. Bounding box is employed for calculating the properties of image regions. The region shall be removed from the input image after labeling the connecting components.

6. Template Matching

In this method of template matching [3], the image of the character will be contrasted to the image in database. Finally the image having finest clarity is selected from the database. Correlation technique is used to locate the finest related match. 26 alphabets (A-Z) and 10 digits (0-9) are employed to match the characters which are done pixel by pixel. The entire string is then compared to the database and the best match is shortlisted and retrieved.

Templates for every character are loaded one by one as different priorities are assigned for template matching. Higher priority is assigned for some and in case if higher priority template matches then lower priority shall not be used and is discarded.

7. Number Written In Text File

Finally the number deciphered is written in its text format that can be used to match the vehicle number in data base.

IV. RESULT ANALYSIS

Fig. 3. Input

Fig 4. Binary Image

Fig. 5. Plate Localisation 1

Fig. 6. Plate Localisation 2

Fig. 7. Character S segmentation

Fig 8: Template Matching

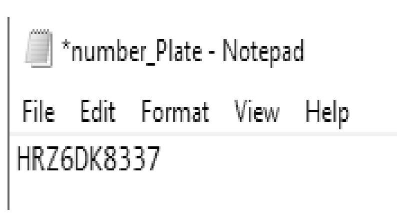


Fig. 9. Number plate output

V. CONCLUSION AND FUTURE SCOPE

Thus, an accurate and efficient, less time consuming method has been devised through this paper which can help in easier detection and extraction of number plates of cars instead of solely relying on human memories of security guards and traffic policemen.

It promises to be less prone to errors with faster computation and matching with the database of lost and penalized vehicles.

Further improvement can be done by making the technique applicable in fast moving traffic and less light ambience so that it can work in every possible weather and environment.

REFERENCES

[1] Ragini Bhat & Bijender Mehandia (2014), "Recognition Of Vehicle Number Plate Using Matlab", International Journal Of Innovative Research In Electrical, Electronics, Instrumentation And Control Engineering*, Vol. 2, Issue 8*

[2] Mayuri M.Musale1& Prof. Dr. M. P. Satone (2015), "Correlation Based Vehicle Number Plate Recognition Using Template Matching", International Journal of Modern Trends in Engineering and Research

[3] D. Renuka Devi & D. Kanagapushpavalli (2012), "Automatic license plate recognition", 3rd International Conference on Trendz in Information Sciences & Computing (TISC2011)

[4] Muhammad Tahir Qadri & Muhammad Asif (2009), "Automatic Number Plate Recognition System for Vehicle Identification Using Optical Character Recognition", IEEE 2009 International Conference on Education Technology and Computer

[5] Abhishek Kashyap & B. Suresh, Anukul Patil & Saksham Sharma, Ankit Jaiswal (2018), "automatic Number Plate Recognition", IEEE International Conference On Advances In Computing, Communication Control andNetworking (ICACCCN)

[6] Bhavin V. Kakani & Divyang Gandhi & Sagar Jani (2017), "Improved OCR based automatic vehicle number plate recognition using features trained neural network", IEEE 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)

[7] Amit Kukreja & Swati Bhandari & Sayali Bhatkar & Jyoti Chavda & Smita Lad (2017), "Indian Vehicle Number Plate Detection Using Image Processing", International Research Journal of Engineering and Technology, Volume: 04 Issue: 04

[8] Riazul Islam & Kazi Fatima Sharif & Satyen Biswas (2015), "Automatic vehicle number plate recognition using structured elements", 2015 IEEE Conference on Systems, Process and Control (ICSPC)

[9] Priyanka Prabhakar & P Anupama & S R Resmi (2014), "Automatic vehicle number plate detection and recognition" IEEE 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)

[10] Jin Chong & Chen Tianhua & Ji Linhao (2013), "License Plate Recognition Based on Edge Detection Algorithm", IEEE 2013 Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing

Designing and Exploring Simple Word Embedding Model Variants

Ritu Sirohi[1], Arun Sharma[2], Neha Bansal[3]

[1,2,3]*Information Technology, Indira Gandhi Delhi Technical University for Women (IGDTUW), India*
[1]sirohi.ritu12@gmail.com, [2]arunsharma@igdtuw.ac.in, [3]nehabansal33@gmail.com

Abstract: Deep learning models requiring large number of parameters for analysing compositionality in text sequences are already explored in depth. But, these models require large computation power for their large number of parameters. It is also not guaranteed that the trained is the simplest one or a significant accuracy improvement was made after adding more layers. Simple word-embedding-based models (SWEMs) have significantly less number of parameters and have performed equally well for multiple datasets. From simple pooling based, concatenation based to hierarchical operations there is no sense of ensembling the results from these compositional functions instead these aggregating operations attempts to combine results in a non-additive and supportive manner that can't be reasonably accounted for with respect to document classification task. Hence, exploring multiple SWEM variants which are best fit for already in demand task of document classification is elaborated upon in this paper. We designed new additive compositional class of SWEM variants which are originally based on concat average-maximum pool operation SWEM model which quantitatively ensembles results after each of the aggregating operations. These new compositional variants have outperformed the earlier stated SWEM models on Yahoo answers dataset. With our variants we saw a maximum increase in accuracy by 0.57 percent for the stated dataset. Also, we achieved a 30 percent faster convergence with our batch normalization layers to already existing research implementation setup. These results highlights the superiority of different aggregator functions instead of using hierarchical ones for the task of document classification and reinstate a relatively lesser importance of spatial arrangement of words while classifying documents.

***Keywords**:* Document Classification, Simple word-embedding-based models (SWEMs), Batch Normalization.

I. INTRODUCTION

More and more documents are being generated with the rapid rise of Internet applications and social networks. These include unstructured to semi structured documents such as conversations, opinions, literature, customer reviews and structured documents such as form data, spreadsheets and relational databases. With this increase in amount of available training data there is a growing adoption of more expressive models that capture rich hierarchical compositional relationships in documents. In the case of unstructured documents, the hierarchical relationships can be observed between characters or tokens and words, words and sentences, sentences and paragraphs, and so on. Similarly, in the case of structured documents it is observed between cells and rows, rows to form or table, and so on. Modelling these relationships leads to more accurate representation of semantics at each level, leading to better text classification performance. An application of such would be identifying relevant advertisement to place around a twitter post based on its semantics and overall context.

It is typical in document-classification architecture to place a text encoder layer in the form of word embedding that would represent tokens as fixed-sized numerical vectors of weights or fixed-sized string vectors. These encoders can be based on highly expressive but resource intensive models such as convolutional neural network (CNN), recurrent neural network (RNN), and long-short term memory (LSTM). Alternatively, these models can be implemented using simple word embedding models (SWEM) such as those that employ word2vec and GloVe. The following sections compares the two classes in terms of expressivity vs computability trade off.

1. Expressivity vs Computability Trade-off

CNN, RNN and LSTM based architectures for language encoding have shown remarkable performance in leading benchmarks. All of these expressive models require optimizing large number of model parameters ranging from hundreds of thousands to millions. To avoid overfitting large number of training data is needed. Furthermore hyper parameter tuning and performance optimization can be challenging. CNN training can be sped up considerably by using GPU, while RNN and LSTM because of their sequential dependency between each node cannot be sped up like CNN. RNN and LSTM perform considerably better than CNN when contextual information is important. On the other hand SWEM models use the efficient embedding models such as GloVe and word2vec to compute word encoding and perform simple linear aggregations to compute encodings at higher level. While SWEM class of models are not as expressive as the aforementioned models and do not capture word sequence information due to its inherent aggregation, it can be trained considerably faster. In other words SWEM class of models trade off expressivity for a significantly reduced parameter count.

II. BACKGROUND STUDY

Basic understanding of Epochs, Batch, Batch Size and iterations are essential to proceed further.

Epochs: Firstly, it is a hyper parameter that gets defined before starting the training of mode. An epoch is defined as passing of data through entire forward and backward pass of gradient descent algorithm once. In deep learning the data is too big to be fed to the model at once. Hence, we divide it over batches.

Batches and Batch Size: The Stochastic Gradient Descent (SGD) optimization algorithm is an iterative algorithm which updates the values of parameters in forward and backward passes. As with deep learning data is large this process becomes time consuming. Hence, we divide datasets into small sets known as batches. Hence, it is a hyperparameter that specifies the number of samples to work through before updating the internal parameters of the model getting trained. Hence, it specifies the number of samples available for training for making an update.

Iterations: It is the parameter composed of division two quantities namely data size (S) and batch size (b). Number of iterations are defined as $n = S/b$, where n is the number of iterations, S is the total size of data and b is batch size of the data.

1. Recurrent Sequence Encoder

Encoding sequence does have an advantage if we can get previous information to somehow contribute into current fed input. A better implementation is defined in recurrent manner to save some parameters in encoding: it takes inputs from word vector v_i at position *i,* and hidden units h_i -1 from previous i-1 position. And for updating the current hidden state value $h_i = g(v_i, h_i-1)$ is applied, where f(.) is the transition function. But, with this we might not be able to remember the original important information that the network wants to remember. Hence, the concept of Long Short Term Memory (LSTM) [6] was introduced which employs gates to remember the specific parts of the sequences and hopefully preserves important information from the sequence. It does maintain the spatial information of sequence into the model but also gives it a much higher number of parameters to learn which makes the model intrinsically complex.

2. Convolutional Sequence Encoder

Convolutional Neural Networks architecture leverage convolution operations to encode text sequences and capture the nearby neighbouring information with the help with the help of these operations. These networks can again be employed to capture information by text sequences. It generates this map with the help of an *n* word window in a consecutive manner generating a corresponding feature map. After, that pooling operations like max-pool does point the most important semantic salient features, resulting in final resultant representation. Deep CNN models have shown good results for being substitute text model for classification tasks.

III. METHODOLOGY

1. Defined Model & Trainings

Consider a text sequence in a document represented as Y, consisting of a sequence of words: {w1, w2, ..., wn}, where *n* is the number of tokens in a given document. Let all the vocabulary tokens in the given document be denoted by *v*i's with respective word embeddings for each token, where they all belong to v*i* belong to R***K.*** After that we create different aggregation based additive functions considered for our experimentation in the form of compositional function, Y → z that combines word embeddings into fixed-size representation of the document z.

2. Simple Word Embedding Model (SWEM) Variants

For modelling word embeddings, we explored class of model variants which involves no additional parameters for creating new compositional functions to encode natural language sequences named as SWEMs.

A class of models without extra composed parameters, termed SWEMs, is of interest to investigate its word embedding performance. In its basic form it computes average over word vectors obtained from the likes of word2vec or GloVe on the given sequence [17] [18]:

$$z = 1/L \sum_{i=1}^{L} v_i \tag{1}$$

In other words, the model may be understood as an average pooling operation, where it takes the meaning of each dimension of word vectors which gives out a representation, say z, having the same dimension as the word vector. This operation will be referred to as *SWEM-aver* operation. The vector z contains information about every element in the sequence except for relative ordering. This kind of pooling is supported by the observation that, in a general case a small count of tokens in the document contribute to prediction result.

Another kind of pooling operation where max value of the word vector dimensions is taken, as opposed to averaging is being proposed. This is called max-pooling and is analogous to the pooling used in CNNs. This is also expected to extract most correlated features:

$$z = \text{Max-pooling}(v^1, v^2, \ldots, v^N) \tag{2}$$

This operation will be referred to as *SWEM-max*. More formally, the k^{th} component in vector z is the maximum value among the set of values $\{v^{1k}, \ldots, v^{Nk}\}$, where v^{Nk} is the k^{th} component in v^1. This pooling operation would result in less weightage to words that are not good predictors by ignoring those in the encoding.

The key difference that SWEM-max has over *SWEM-aver* is that in SWEM-aver all words of sequence have equal weight in the representation. On the other hand *SWEM-max* is choosing the salient ones.

However, both can be composed if we consider that these operations are complementary, extracting different information from the sequence. This work evaluates performance of a composition function that we refer to as *SWEM-concat.*

3. Flow Chart and Model Explanation

Before inputting the data, we must pre-process it. Hence, we use Glove vectors with 300 dimensions to pre-process dataset and create vector representation out of words. After that we split the dataset into test, train and validate subset. We also maintain wordtoix (word to vector mapping dictionaries) and ixtoword (vector to word mapping dictionaries). The feed forward neural network is created which will take utilities like average embedding encoder, max embedding encoder etc. from utils.py file. This fully connected feed forward neural network in original implementation didn't have normalization of any kind. In next stage model variant selection is done which can be average, maximum, concatenation or hierarchical on word embeddings. Then data is fed into encoder by selecting embedded encoder module. After that training procedure on batches of specified size with early stopping mechanism in place recording the maximum test accuracy.

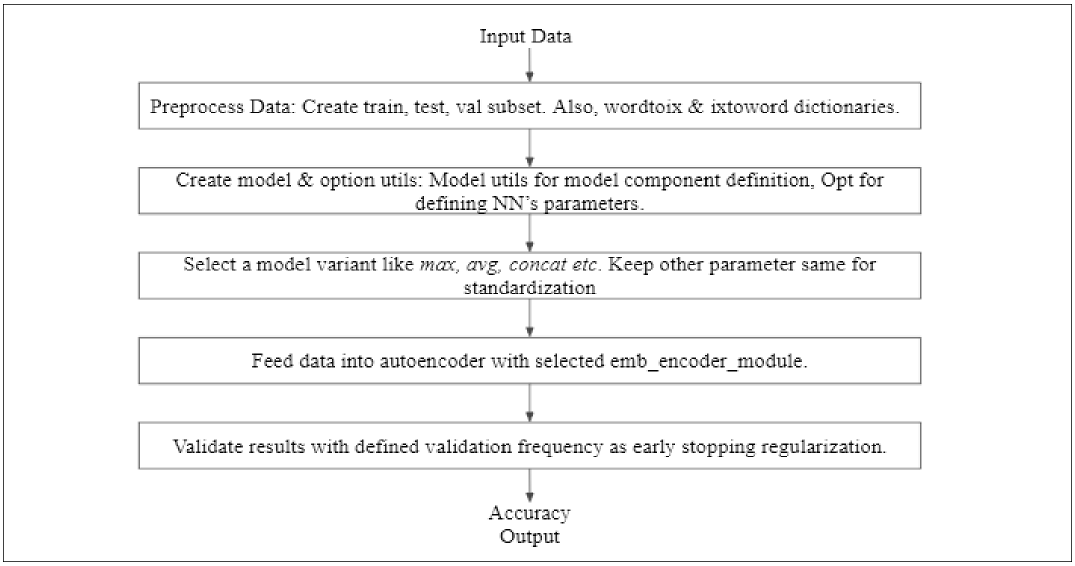


Fig. 1. Explanation of Simple Word Embedding Model in form of flowchart

Early stopping mechanism is used to stop the model from overfitting on a given data. If validation accuracy goes down in comparison to train accuracy, we stop the model from training as it will overwhelming the model parameters with high variance. Also, in the same stage results are validated with defined validation frequency and testing is done of test set defined over the Yahoo dataset.

The new SWEM variant which we designed is an extension of the existing one where batch normalisation is applied for early convergence after adding the outputs from average and maximum embedding layer operations on word embeddings then rather than taking a concatenation operation. We feed our resultant embedding as input to logit *(functional layer that maps 0 to 1 values on –infinity to + infinity)* layer and before feeding it to probability distribution layer we again apply batch normalization operation. Finally, with given probability distribution over a sequence of sentences of a document we make the prediction.

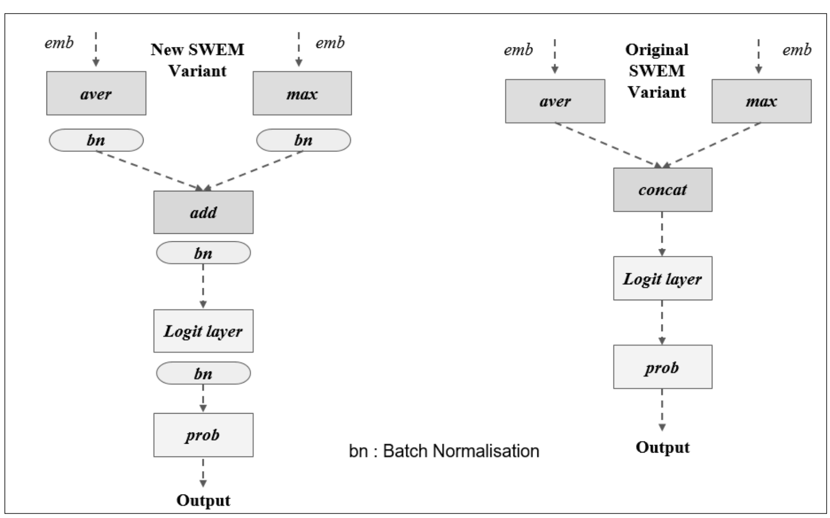


Fig. 2. Comparison of Model diagram for old SWEM using add operation and new SWEM using add operation and batch normalisation.

This is a model diagram which is created from understanding the code provided by Dinghan Shen where he applies average and maximum operation on word embedding then applied concatenation operation and feed that result to logit layer followed by taking the probability and providing output.

IV. IMPLEMENTATION

We evaluated our coalescing compositional functions on task of document categorization on Yahoo Answers dataset [16] which also provides an equivalence to topic classification task. We have used Glove vector embedding with dimensions of 300 [16] for our task. The out of vocabulary words gets initialized with uniform distribution ranging from -0.01 to 0.01. These embeddings are utilized in two ways, first for directly updating each word embedding during the training task and second is 300-dimensional multilayer perceptron feedforward encoder. Glove embeddings are given as direct input to the MLP and output defining our newly refined embeddings. We have used sigmoid soft-max layer for the classifier MLP that is implemented on dimensions selected from a set of 300-dimension Glove vectors instead of others. Adam optimizer is used with default parameters on stochastic gradient descent algorithm for minimization of loss with a learning rate of 3x10-3. Dropout and batch normalization both are employed for achieving regularization in the training tasks.

We carry out our document categorization tasks on a document size at least 100 words averaged out documents. We follow the split specified in [16] for comparability of our tests. On topic predictions tasks SWEM models does outperforms the convention LSTM and CNN models. Specifically, our Add based models outperforms the results of 29-layer deep CNN models.

1. Design Matrix

For convenience, a pre-processed version of this dataset is being used from the baseline study. This data is made available by the author of baseline study as serialized Python objects that can be utilized using Python Pickle library. Unlike LSTM and CNN, SWEM has no composed parameters, so SWEM offers orders of magnitude reduction in parameters and is therefore easy to compute.

2. SWEM Limitations

However, if evaluated for sentiment analysis, CNN and LSTM would perform better than SWEM suggesting that word order is important for capturing semantics necessary to classify sentiment. A study by Pang et al. (2002) hypothesizes this and found that when word ordering information was performed their model performed better in sentiment analysis task. CNN and LSTM capture ordering information inherently with their spatial filter operation and recurrent functions, respectively. However, for the task of topic classification, this isn't relevant therefore SWEM would still be a suitable or superior choice in tasks such as document classification where prediction can rely on presence of keywords and not relative ordering of words. Unlike CNN and LSTM. SWEM also works efficiently on long documents.

V. RESULTS

This chapter contains an analysis of our results which specifies the step by step results that we have obtained for multiple variants of SWEMs that have existed from previous implementation plus the newer ones that we have created.

Epochs	Concat	Average	Max	Add(Avg-Avg)	Add(Avg-Max)	Add(Max-Max)	Add(Cat-Cat)
1	71.14	69.92	69.97	70.23	71.63	69.82	70.67
2	71.82	70.25	70.89	70.79	72.23	70.23	71.56
3	72.23	70.89	71.23	71.58	72.98	71.22	71.89
4	72.49	71.26	71.89	72.24	73.23	72.56	72.47
5	72.89	71.59	72.21	72.68	73.67	72.57	72.84
6	73.12	72.56	72.29	73.13	74.12	73.01	73.68
7	73.33	72.89	72.42	73.13	74.12	73.01	73.68
8	73.53	73.14	72.66	73.13	74.12	73.01	73.68
9	73.53	73.14	72.66	73.13	74.12	73.01	73.68
10	73.53	73.14	72.66	73.13	74.12	73.01	73.68
11	73.53	73.14	72.66	73.13	74.12	73.01	73.68
12	73.53	73.14	72.66	73.13	74.12	73.01	73.68
13	73.53	73.14	72.66	73.13	74.12	73.01	73.68
14	73.53	73.14	72.66	73.13	74.12	73.01	73.68

Fig. 3. Result of SWEM Model using different pooling operations 1) Concat 2) Average 3) Max 4) Add with different variants- Avg-Avg, Avg-Max, Max-Max, Concat-Concat. Shows faster convergence for our batch normalized Add (Avg-Max) based aggregator models.

From this graph we do observe that Add (Avg-Max) and Add (Concat-Concat) variants have performed even better than the earlier concat variation designed in the baseline paper. We also observe that Add (Avg-Max) does obtain a maximum accuracy from our results. Add (Avg-Avg) and Add (Max-Max) does give us the results that doesn't show improvements instead a decline in accuracy is observed may be accounting for loss in spatial information. This loss is highest in Max and Add (Max-Max) function. Having this data along with epochs also gives us the opportunity to comment on the nature of convergence of a batch normalized solution instead of one that is not batch normalized. With the help of graph mentioned below and the tabular data we will comment on the convergence rate of the solution.

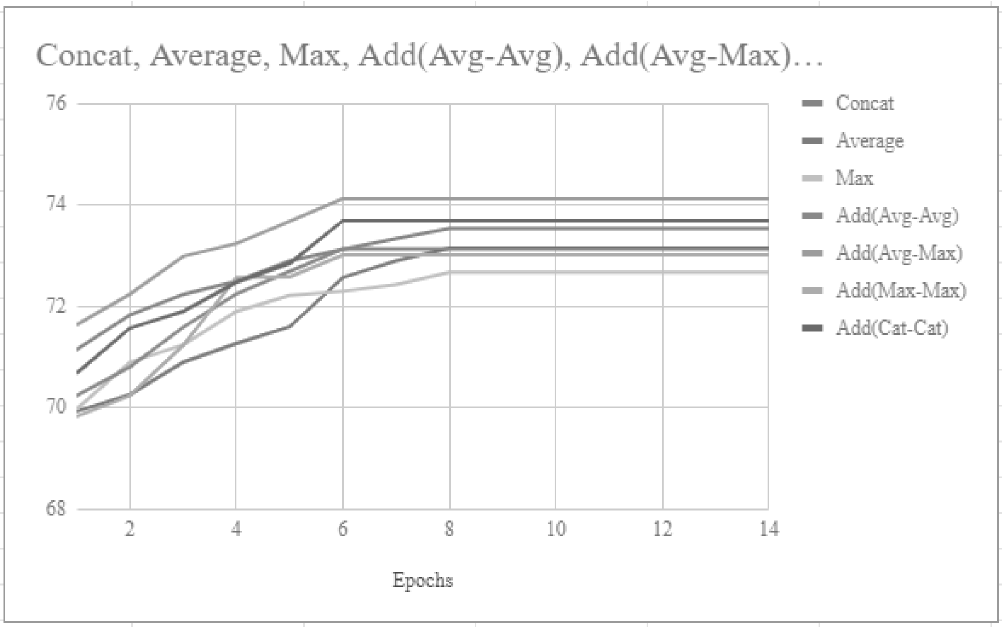


Fig. 4. Graph of epochs vs pooling variations 1) Concat 2) Average 3) Max 4) Add and its variants. Showing faster conversion for batch normalized Add (Avg-Max) based aggregator model.

We do observe faster epoch conversions for SWEM variants designed by us with all of them converging after 6th epoch where as others going onto convergence after 8th one. Hence, our solution not only achieves higher accuracy but also achieves better results earlier as compared to earlier solutions. Also, all of our new variants converge at the same epoch epoch only. Our newly proposed add variants with batch normalization shows a 30% faster convergence. And the results are improved by an accuracy of 0.57 i.e. a 0.6 % percent improvement in earlier results.

VI. CONCLUSION & FUTURE WORK

Deep learning models require large computation power for their large number of parameters where as Simple word-embedding-based models (SWEMs) have much simpler design, significantly less number of parameters and have performed equally well for multiple datasets. We explored multiple SWEM variants and designed new additive compositional class of SWEM variants which are originally based on concat average-maximum pool operation. These new compositional variants have outperformed the earlier stated SWEM models on Yahoo answers dataset. With our variants we saw a maximum increase in accuracy by 0.57 percent for the stated dataset. Also, we achieved a 30 percent faster convergence with our batch normalization layers.

In this context, one further direction can be to develop an optimisation algorithm like Genetic algorithm, Random walk, Particle swarm optimization (PSO) etc for the SWEM framework, to get better performance as we don't know how much complex function we can create for example genetic algorithm can also be taken into consideration while designing other variants for SWEM.

REFERENCES

[1] Y. Kim, "Convolutional neural networks for sentence classification", arXiv preprint arXiv:1408.5882, 2014.

[2] Y. Deng, Z. Ren, Y. Kong, F. Bao and Q. Dai. "A hierarchical fused fuzzy deep neural network for data classification", IEEE Transactions on Fuzzy Systems 2017.

[3] Z. Yang, D. Yang, C .Dyer, X. Dong, A. Smola, E. Hovy. "Hierarchical Attention Networks for Document Classification". Proceedings of NAACL-HLT 2016.Association for Computational Linguistics, San Diego, California,2016

[4] M. Z. Afzal, S. Capobiancot, M. I. Malik, S.M Thomas M. Breuel, A. Dengel and M. Liwicki. "DeepDocClassifier: Document Classification with Deep Convolutional Neural Network".13th International Conference on Document Analysis and Recognition (ICDAR), 2015

[5] S. Lai, L. Xu, K. Liu, J. Zhao. "Recurrent Convolutional Neural Networks for Text Classification". National Laboratory of Pattern Recognition (NLPR) Institute of Automation, Chinese Academy of Sciences,China, 2015

[6] L. Vega and A. M.-Vazquez. "Dynamic Neural Networks for Text Classification", International Conference on Computational Intelligence and Applications, Zapopan, Mexico,2016.

[7] S. Zobeidi, M. Naderan, S. E. Alavi. "Effective Text Classification Using Multi-level Fuzzy Neural Network". 5th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS) 7-9 March, Qazvin Islamic Azad University, Tehran, Iran, 2017.

[8] L. Lenc. "Deep Neural Networks for Czech Multi-label Document Classification". New Technologies for the Information Society. 13th International Conference on Deep Neural Networks, University of West Bohemia, Plzen, Czech Republic, 2017.

[9] J. Lilleberg, Y. Zhu, Y. Zhang. "Support Vector Machines and Word2vec for Text Classification with Semantic Features", IEEE Transactions on Cognitive Informatics Cognitive Computing, 2015.

[10] Y. Zhang, B. C. Wallace. "A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification", arXiv:1510.03820v4, 2016.

[11] Kalchbrenner, E. Grefenstette, P. Blunsom. "A Convolutional Neural Network for Modelling Sentences", In Proceedings of ACL,2014.

[12] Radhika, Bindu, L. Parameswaran. "A Text Classification Model Using Convolution Neural Network and Recurrent Neural Network", International Journal of Pure and Applied Mathematics Volume 119 No. 15, pages 1549-1554, 2018.

[13] N. Srivastava, G. E Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting". JMLR, 15(1):1929–1958, 2014.

[14] A.P. Parikh, O. Tom, D. Das, J. Uszkoreit. "A decomposable attention model for natural language inference". EMNLP, 2016.

[15] M. Iyyer, V. Manjunatha, J.B. Graber, H. Daum ́e., "Deep unordered composition rivals syntactic methods for text classification". In ACL, volume 1, pages 16 81–1691, 2015.

[16] D. Shen, G. Wang, W. Wang, M. R. Min Q. Su, Y. Zhang, C. Li, R. Henao, L. Carin. "Baseline Needs More Love on Simple Word-Embedding-Based Models and Associated Pooling Mechanisms". arXiv:1805.09843v1 [cs.CL], 2018.

[17] Y. Adi, E. Kermany, Y. Belinkov, O. Lavi, Y. Goldberg. "Fine-grained analysis of sentence embeddings using auxiliary prediction tasks". ICLR, 2016.

[18] J. Wieting, M. Bansal, K. Gimpel, K. Livescu. "Towards universal paraphrastic sentence embeddings". ICLR, 2015

TRACK 2

SPEECH TECHNOLOGY

Intricacies in Punjabi to Urdu Machine Translation System

Nitin Bansal, Ajit Kumar

Punjabi University, Patiala

Abstract: Development of Machine Translation System (MTS) for any language pair is a challenging task for several reasons. Lack of lexical resources for any language is one of the major issues that arise when developing MTS using that language. In the coming years, the usage of MTS will become increasingly important and research community must emphasize on its resources to resolve new problems on day to day basis. For example, during the development of Punjabi to Urdu MTS (PUMTS) using Hybrid approach, many issues were recognized while preparing lexical resources for both the language. There is no machine readable dictionary available for Punjabi to Urdu which can be directly used for translation. Along with this non-availability of parallel corpus, handling of out of vocabulary words, handling of multiple sense Punjabi word, identification of proper nouns, identification of collocations in the source sentence, word-order in Punjabi and Urdu language are the issues which we face during development of this system.

Keywords: MTS, Word-order, Lexical Resources

I. INTRODUCTION

Machine Translation System (MTS) are designed to translate one language to another language. Human and machine translators have their own share of challenges i.e. no two translators can produce identical translations of the same text in the same language pair. It takes several rounds of revisions to meet the requirements.

Most of the MTSs for many natural languages are based on statistical approach, where bilingual corpus is used to create correct translations but limited to some scope. The problem area is to analyze the accuracy of translation using available methods or to design new techniques for evaluation. Therefore, it is important to design and develop such methods, which helps in creating correct translations from one natural language to another.

India is a multilingual country, where 22 official languages are being used for the official purpose in different states. For example, in Punjab, Punjabi is being used as an official language while working in offices and in Uttar Pradesh, Urdu is being used as one of the official language. Thus, it becomes compulsory to develop an MTS to have proper communication among them.

In India, organizations like International Institute of Information Technology (IIIT), Technology Development of Indian Language (TDIL) are involved in working on accurate MTS from one regional language to another. They have developed MTSs from English to India regional languages. Popular Indian MTSs are ANGLABHARTI, ANUBHARTI, MANTRA, ANUSAARAKA, SHIVA, SHAKTI, MATRA, ANUVAADAK. Google translator and Bing translator are in demand worldwide as they are required for communication, information sharing all over the world. But still computer scientists have not achieve higher accuracy in MTS while translating from one natural language to another. Thus, MTSs poses a big challenge in generating an exact translation from source language to target language. An attempt has been made to identify and resolve various types of issues and challenges arise during development of PUMTS. The objective of this paper is to discuss various issues and challenges posed during the development of PUMTS, especially during development of parallel corpora and testing of the system.

II. RELATED WORK

During the development of PUMTS, we have studied various research papers and reviewed various issues and challenges covered during the development of those MTSs. Issues and challenges may be approached in number of ways but some of them are well explained below:

Researchers found various issues and challenges, which were analyzed and resolved by the authors of English to Urdu MTS developers*[7]* and *[11]*, which were related to *Word translation*, *Phrase translation*, *Syntactic translation* and *Semantic translation*. Since in many languages, sometime single word has many meanings; so to find out the right meaning of the word for the sentence is a real challenge. Moreover, Human can understands the right meaning of the word by looking at the context in which that word is being used in the sentence. But MTS has to fix it up, which is a real challenge for the developers. In another issues i.e. Phrase translation problem, some idiomatic phrases have hidden meaning where word by word translation is not possible. Since languages which are having similar structure and word-order, above issues do not seem to be a big challenge and can be resolved as and when required. Moreover, some other related problems can arises during the development of MTS are *conference translation problem* and *discourse translation problem.*

In case of Hindi-Punjabi SMT, another approach of Statistical Post Editing System (SPES) has been applied to get better translation quality *[10]*. In this approach, initially Hindi-Punjabi parallel corpus has been used to generate Punjabi output. This Punjabi output further corrected manually and compared with machine generated Punjabi output. This training has been done on SRILIM and GIZA++ to generate language model and translational model. This resulted in increase in accuracy enhancement by 12%.

The issues and challenges being faced by various researchers while following machine translation approaches are discussed by the author*[8]*.For example, in case of *Rule-based approach, w*hile applying Rule-based technique, insufficient amount of really good dictionaries for source and target language is the key issue for the developer. In case of *Direct translation approach,* a big challenge is local-word adjustments, which need to be carry out for better quality translations. But this results in poor quality of translations due to poor quality of bilingual dictionary and rudimentary knowledge of grammar of target language as well. In case of *Transfer based approach,* development of rules and mapping of rules with the source and target language results in generation of sematic and syntactic issues. At every step of translation, it is difficult to apply each and every transfer module and to keep those modules as simple as possible too. In corpus-based approach i.e. *Statistical Approach,* when single sentence in source language found to have multiple sentences in target language leads to sentence alignment issue. Another issue arises while using this approach is statistical anomalies where proper nouns are wrongly translated due to their fixed meaning in the training set. Moreover, creation of parallel corpus becomes a big challenge for the developer of MTS for resource deficient language pair.

Some of the well-known issues and challenges addresses during the development of MTS are: translation of low-resource language pair, translation across domains, translation of informal text, translation into morphologically rich languages and translation of speech *[12]*. The only solution for these issues is to develop a large bi-text, which is difficult to develop.

During the development of Urdu to Punjabi MTS, the authors do not have a parallel corpus to train the system. Urdu language consists of many words with different character set, thus spelling variation was the key issue for them during the development. Urdu and Punjabi languages are SVO but due to their free word-order, it makes the MT task more challenging. Some of the other key issues during the development were: segmentation issues in Urdu language, rich morphological nature of Urdu and Punjabi languages, and words with or without diacritical marks in Urdu language and many more. All these issues were well handled by the developers during the development of Urdu to Punjabi MTS using the incremental machine learning process, where many methods and algorithms were designed to train the system to generate quality output.

III. ISSUES AND CHALLENGES DURING DEVELOPMENT OF PUNJABI TO URDU MACHINE TRANSLATION SYSTEM

Bilingual parallel corpus is the primary resource for the development of any SMT system and there is no published work on Punjabi to Urdu SMT. Since, authentic parallel corpus is not available in any public domain, therefore creation of parallel corpus for the development of PUMTS is a big challenge for us. After creating some part of parallel corpus for the language pair words/ compound words/ phrases, sentence level alignment has been performed manually due to non-availability of sentence aligned corpora. Urdu Unicode data has been compiled from three Urdu website and parallel Punjabi text has been generated using Urdu to Punjabi translation system.

Further, Punjabi text has been refined manually to remove error or mistakes in translated Punjabi text. Later on, a tool has been developed to reverse the Punjabi text so that it can get mapped with the Urdu text (Urdu text written from right-to-left). This process has been well assisted by MOSES *[3]* toolkit along with GIZA++ *[2]*(a software for word/ phrase alignment) and mkcls *[1]*(a utility for making bilingual word classes) were used for training set. While aligning of sentences for language pair, tags and other punctuation marks were removed to get high quality translation.

1. Gurmukhi and Shahmukhi Scripts

Punjabi language is written in Gurumukhi script and is often written in the Devanagari or Latin scripts due to influence from Hindi and English. Gurmukhi script was standardized in 16th Century and it was designed to write the Punjabi Language. Gurmukhi script has forty one (41) letters, in which thirty eight (38) letters are consonants and three (3) are basic vowels. In addition to this, there are nine (9) dependent vowels and ten (10) independent vowels.

Shahmukhi is a local variant of Urdu script to record Punjabi language. Shahmukhi script is written from right to left Nastalique style of the Persian and Arabic script and numerals from left to right. It has thirty nine (39) letters including five long vowel signs and eighteen (18) aspired consonants (combination of two Urdu characters to form single Urdu character).

2. Mapping between Punjabi and Urdu Text

Table 1. Character set for Urdu and Punjabi language

Urdu Alphabet	Punjabi Alphabet	Urdu Alphabet	Punjabi Alphabet
ا	ਅ	ض	ਜ਼
ب	ਬ	ط	ਤ
پ	ਪ	ظ	ਜ਼
ت	ਤ	ع	ਅ/ ਆ
ٹ	ਟ	غ	ਗ
ث	ਸ	ف	ਫ਼
ج	ਜ	ق	ਕ
چ	ਚ	ک	ਕ
ح	ਹ	گ	ਗ
خ	ਖ	ل	ਲ
د	ਦ	م	ਮ
ڈ	ਡ	ن	ਨ
ذ	ਜ਼	ں	ਵ/ ਙ/ ਣ
ر	ਰ	و	ਵ
ڑ	ੜ	ہ	ਹ
ز	ਜ਼	ھ	ਹ
ژ	ਜ਼	ء	
س	ਸ	ی	ਯ/ ਇ
ش	ਸ਼	آ	ਆ
ص	ਸ	ے	ਯ

Table 2. Aspired Consonants for Urdu and Punjabi language

Urdu Alphabet	Punjabi Alphabet	Urdu Alphabet	Punjabi Alphabet
بھ	ਭ	ڈھ	ਢ
پھ	ਫ	رھ	ਰ੍ਹ
تھ	ਥ	ڑھ	ੜ੍ਹ
ٹھ	ਠ	کھ	ਖ
جھ	ਝ	گھ	ਘ
چھ	ਛ	لھ	ਲ੍ਹ
دھ	ਧ	مھ	ਮ੍ਹ
نھ	ਨ੍ਹ	وھ	ਵ
یھ	ਯ		

3. Numerals Mapping for Urdu and Punjabi Language

Shahmukhi has its own set of numerals that behave exactly as Gurmukhi numerals as shown in following Table

Table 3. Numerals set for Urdu and Punjabi language

Urdu Numerals	Punjabi Numerals
٠	0
١	1
٢	2
٣	3
۴	4
۵	5
۶	6
٧	7
٨	8
٩	9

4. Diacritics Used in Urdu Language

Following table describes the other symbols and diacritical marks of Shahmukhi, which can be mapped with Gurmukhi script letters in one or other way.

Table 4. Diacritics used in Urdu language

◌ٗ	◌ّ	◌َّ	◌ِ	◌َ	◌ُ	◌ٰ	◌ٔ	◌ٖ	◌ؒ	◌ؐ

5. Punctuation Marks Used in Urdu Language

Following table describes the Punctuation marks used while writing in Shahmukhi script for Urdu language.

Table 5. Punctuation marks used in Urdu language

-	/	\	؟	'	"	:	؛	۔	()	!

6. Order and Alignment of Numerals, Special Characters, Dates, Websites, Email etc

Punjabi text consists of various numerals, dates, website name, email-id etc, which follows the same writing pattern i.e. left to right for Urdu text also. Therefore, it is an important issue, where we need to take care while dealing with conversion from Punjabi to Urdu translation as well. For example,

- A website name www.google.com must be written in same manner for Punjabi as well as Urdu.
- An email-id profnitinbansal@gmail.com must be written in same manner for Punjabi as well as Urdu.

- A date 12-10-2018 must be written in same manner for Punjabi as well as Urdu.
- A number 3245 must be written in same manner for Punjabi as well as Urdu.

7. Some of the Issues with Urdu and Punjabi Text

Another major challenge is the recognition of Urdu text without diacritical marks (like Zabar, Zer, Pesh, Shad, Hamza, Khari-Zabar, Do-Zabar and Do-Zer. Missing of diacritical marks in Urdu text, results in difficulty to understand the meaning of word e.g.

Text with diacritical marks –

مِتالی راج عک کرِکےٹ خِالڈی هے
(Mitali raj ek cricket khiladi hai)

Text without diacritical marks –

متالی راج عک کرکےٹ خلاڈی هے
(mtali raj ek karket khaladi hai)

In the above Urdu sentence, Table-1 describes the meaning of Urdu words with diacritical marks and Urdu words without diacritical marks. We can see that meaning is different in Punjabi language due to missing diacritical marks in Urdu words. But, Urdu reader can easily read and generates the meaning of Urdu words & sentences without diacritical marks as well. Thus, it is another big challenge for us to keep track with human translation as well. Diacritical marks are important in Urdu language as it results in correct punctuation and disambiguation of certain words.

Table 6. Urdu words with and without diacritical marks

Urdu Words with diacritical marks	**Meaning in Punjabi**	**Urdu Words without diacritical marks**	**Meaning in Punjabi**
مِتالی	ਮਿਤਾਲੀ	متالی	ਮਤਾਲੀ
کرِکےٹ	ਕ੍ਰਿਕੇਟ	کرکےٹ	ਕਰਕੇਟ
خِالڈی	ਖਿਲਾਡੀ	خلاڈی	ਖਲਾਡੀ

Another issue arises where some characters in Urdu language have no similar character in Punjabi language. Few of them are described in the following Table – 2. But these characters are important for Urdu reader and writer for the better clarity over the Urdu text.

Table 7. Urdu characters, which do not have similar character in Punjabi language

Urdu Characters
Hamza (ء)
Do Zabar (ً)
Aen (ع)
KhadiZaber (ٰ)
do chashmee (ھ)

Multiple mapping was another challenge, where multiple Urdu characters have single Punjabi character. For example, 'ਸ' in Punjabi language can be written in three forms in Urdu i.e. ص , س and ث. Following Table – 3 describes some characters which has single Punjabi character.

Table 8. Punjabi characters matching with multiple Urdu characters

Urdu Character	Punjabi Character
س	ਸ
ص	
ث	
ک	ਕ
ق	
ج	ਜ
ض	
ز	
ذ	

Therefore a good transliteration scheme among Punjabi and Urdu languages can resolve this issue of single Punjabi character with multiple Urdu characters *[4]*.

In the same way some other characters in Urdu language has many forms for single character in Punjabi language. Therefore, grammar rules and context needed to be designed using Rule-based approach for quality translation of Punjabi to Urdu text.

Another issues arises when some words in Punjabi language has multiple words in Urdu language. Following Table – 4 described few of them:

Table 9. Punjabi word matching with multiple Urdu words

Punjabi Word	Urdu Word
ਪਤੀ	خاوند (khaveend) خادِم (khadeem) شوہر (shohar)
ਰਾਜਾ	راجا (Raja) بادشاہ (Badshah) نواب (Nawab)

Similarly, many Punjabi words have single Urdu words as described in the following Table 5.

Table 10. Punjabi words having single Urdu word

Punjabi Word	Urdu Word
ਆਨੰਦ ਮਜਾ	لُتف (lutaff)
ਰੱਬ ਭਗਵਾਨ	خُدا (khuda)

Sometime one Urdu word can have different meaning in Punjabi sentences. For example, سُرت (soorat) word in Urdu can be used in place of City name (as noun) as well as for the face of a person. This creates an ambiguity among sentences in Punjabi to Urdu language and thus poses another challenge in developing PUMTS.

Above issues can be characterized as follows, while mapping Urdu with Punjabi words or vice-versa-:

- Multiple character mapping - one character of one language with many characters of other language
- Multiple word mapping – one word of one language has many words in other language
- One Shahmukhi word (without diacritics) has many different Gurmukhi words
- Segmentation problem due to space omission and space insertion
- Word segmentation and wrong interpretations
- Space omission problem in more than two merged words

Transliteration ambiguity at word level in Urdu and Punjabi is another issue which can be resolved using higher level of language information for the pair of language. Since Urdu sentences are not using space consistently therefore non-trivial risk of word segmentation is another big challenge for us while developing PUMTS. This requires maximum efforts while transliterating Punjabi to Urdu text. Multi-word Punjabi text has a single word in Urdu, this imposes another issue for the developer to transliterate those words or sentences.

IV. CHALLENGES DURING DEVELOPMENT OF PARALLEL CORPUS FOR PUNJABI AND URDU LANGUAGE PAIR

It is a very tedious task to create a parallel corpus for Punjabi and Urdu language pair to be used in our MTS. Moreover, manual creation of highly accurate parallel corpora is a laborious and time consuming task. It is also important to keeps an eye over noisy data, which takes lots of time in cleaning manually. Therefore, most of the words in available aligned text are transliterated not translated. We already knew that aligning the text from one language to another is time consuming. But, data collection source as well as cleaning and conversion of data plays an important role in the development of parallel corpora for any language pair (such as in case of Punjabi-Urdu language pair).

1. Data Sources for the Creation of Parallel Corpus

During the creation of parallel corpus for Punjabi and Urdu language pair, some of the websites were visited for this purpose, for example www.arynews.com/ud, www.bbc.co.uk/urdu, https://dailypakistan.com.pk, and https://dunya.com.pk. Further, Urdu to Punjabi MTS is being used via http://www.u2p.learnpunjabi.org, to convert Urdu text into Punjabi text, so that parallel text could be prepared. The accuracy of the used translation system is about 90%. During this process, some of the Punjabi sentences require less manual efforts but some are hard to understand so more manual efforts has been applied to correct the spelling/ grammar variation as well meaning of the sentence. Different spelling for same word was a major challenge during this development. Thus, by resolving said challenges, we will able to prepare a line aligned parallel corpus for this language pair. Till now, more than 1,00,000 sentences has been cleaned and aligned in language pair of Punjabi and Urdu language. Data has been taken from different domain such as Politics, Sports, Business, Books & Magazines, Health, Literature, Religion, Art & Culture, Tourism and Entertainment. The sentence length varies across the corpora, which ranges from 5 to 40 words on an average. Since, our aim is to get high quality of translation using different sets of data so that hybrid approach can be easily applied. Some of the examples of above said process is described below -:

Data Collected from Urdu Website: (Urdu Data)
اس دن کی ملنے والی ویڈیو فوٹیج میں دیکھا جا سکتا ہے کہ خالد شہنشاہ نے کچھ عجیب و غریب حرکات کی تھیں جن کا کوئی جواز نہیں تھا اور نہ ہی کسی نے ان کی تسلی بخش وضاحت کی ہے۔
(hai ki vazahit baksh tassali ki unn ne kisi hee na aur tha nahi jawaj koi ka jin thii ki harkat gareeb azeeb kuch ne shehanshah khalid ke hai sakta ja dekha main footej video vaali milne kii din iss)
Data Translated Using http://u2p.learnpunjabi.org: (Translated Punjabi Data with 90% accuracy)
ਇਸ ਦਿਨ ਦੀ ਮਿਲਣ ਵਾਲੀ ਵੀਡੀਓ ਫੁਟੇਜ ਵਿੱਚ ਵੇਖਿਆ ਜਾ ਸਕਦਾ ਹੈ ਕਿ ਖਾਲਿਦ ਸਮਰਾਟ ਨੇ ਕੁੱਝ ਵਚਿੱਤਰ ਗਤੀਵਿਧੀਆਂ ਸੀ ਜਿਨ੍ਹਾਂ ਦਾ ਕੋਈ ਜਵਾਬ ਨਹੀਂ ਸੀ ਅਤੇ ਨਾ ਹੀ ਕਿਸੇ ਨੇ ਉਹਨਾ ਦੀ ਤਸੱਲੀ ਬਖਸ਼ ਸਪਸ਼ਟੀਕਰਨ ਦਿੱਤਾ ਹੈ ।
(Isa dina dī milaṇa vālī vīḍī'ō phuṭēja vica vēkhi'ā jā sakadā hai ki khālida samarāṭa nē kujha vacitara gatīvidhī'āṁ sī jinhāṁ dā kō'ī javāba nahīṁ sī atē nā hī kisē nē uhanā dī tasalī bakhaśa sapaśaṭīkarana ditā hai.)
Manual cleaning of Punjabi data:
ਇਸ ਦਿਨ ਦੀ ਮਿਲਣ ਵਾਲੀ ਵੀਡੀਓ ਫੁਟੇਜ ਵਿੱਚ ਵੇਖਿਆ ਜਾ ਸਕਦਾ ਹੈ ਕਿ ਖਾਲਿਦ ਸਮਰਾਟ ਨੇ ਕੁੱਝ ਵਚਿੱਤਰ ਗਤੀਵਿਧੀਆਂ ਕੀਤੀਆਂ ਸਨ ਜਿਨ੍ਹਾਂ ਦਾ ਕੋਈ ਜਵਾਬ ਨਹੀਂ ਸੀ ਅਤੇ ਨਾ ਹੀ ਕਿਸੇ ਨੇ ਉਹਨਾ ਦਾ ਤਸੱਲੀ ਬਖਸ਼ ਸਪਸ਼ਟੀਕਰਨ ਦਿੱਤਾ ਹੈ ।
(Isa dina dī milaṇa vālī vīḍī'ō phuṭēja vica vēkhi'ā jā sakadā hai ki khālida samarāṭa nē kujha vacitara gatīvidhī'āṁ kītī'āṁ sana jinhāṁ dā kō'ī javāba nahīṁ sī atē nā hī kisē nē uhanā dā tasalī bakhaśa sapaśaṭīkarana ditā hai.)

Data Collected from Urdu Website: (Urdu Data)
روسی خبررساں ادارے آر آئی اے کے مطابق امن کونسل کے ترجمان احسان طاہری نے کہا: 'ہم نے طالبان کے ساتھ براہِ راست مذاکرات کے
موضوع پر بات کی اور ان سے کہا کہ وہ اپنی مرضی سے جگہ اور وقت مقرر کریں۔'
(karen maqroor waqt aur jagah se marzi apne woh ke kaha se inn aur kee baat par mazuae ke mazakarat barae-e-rast saath ke Taliban ne hum kaha ne tahiri ehsaan tarzmaan ke council aman matabik ke ae aai aar adaare khabarsa roosi)

Data Translated Using http://u2p.learnpunjabi.org: (Translated Punjabi Data with less than 90% accuracy)
ਰੂਸੀ ਖ਼ਬਰੀ ਅਦਾਰੇ ਆਰ ਆਈ ਏ ਦੇ ਅਨੁਸਾਰ ਸ਼ਾਂਤੀ ਕੌਂਸਲ ਦੇ ਪ੍ਰਵਕਤਾ ਅਹਿਸਾਨ ਤਾਹਰੀ ਨੇ ਕਿਹਾ : ""ਹਮ ਨੇ ਤਾਲਿਬਾਨ ਦੇ ਨਾਲ ਸਿੱਧੇ ਗੱਲਬਾਤ ਦੇ ਵਿਸ਼ੇ ਉੱਤੇ ਗੱਲ ਕੀ ਅਤੇ ਉਨ੍ਹਾਂ ਤੋਂ ਕਿਹਾ ਕਿ ਉਹ ਆਪਣੀ ਇੱਛਾ ਨਾਲ ਜਗ੍ਹਾ ਅਤੇ ਵਕਤ ਨਿਯੁਕਤ ਕਰਨ । ""
(Rūsī k̲h̲abarī adārē āra ā'ī ē dē anusāra śāntī kōnsala dē pravakatā ahisāna tāharī nē kihā: ""Hama nē tālibāna dē nāla sidhē galabāta dē viśē utē gala kī atē unhāṁ tōṁ kihā ki uha āpaṇī ichā nāla jag'hā atē vakata niyukata karana. "")
Manual cleaning of Punjabi data:
ਰੂਸੀ ਖ਼ਬਰੀ ਅਦਾਰੇ ਆਰ ਆਈ ਏ ਦੇ ਅਨੁਸਾਰ ਸ਼ਾਂਤੀ ਕੌਂਸਲ ਦੇ ਪ੍ਰਵਕਤਾ ਅਹਿਸਾਨ ਤਾਹਰੀ ਨੇ ਕਿਹਾ : ਅਸੀਂ ਤਾਲਿਬਾਨ ਦੇ ਨਾਲ ਸਿੱਧੀ ਗੱਲਬਾਤ ਦੇ ਵਿਸ਼ੇ ਉੱਤੇ ਗੱਲ ਕੀਤੀ ਅਤੇ ਉਨ੍ਹਾਂ ਨੂੰ ਕਿਹਾ ਕਿ ਉਹ ਆਪਣੀ ਇੱਛਾ ਨਾਲ ਜਗ੍ਹਾ ਅਤੇ ਵਕਤ ਨਿਯੁਕਤ ਕਰਨ ।
(Rūsī k̲h̲abarī adārē āra ā'ī ē dē anusāra śāntī kōnsala dē pravakatā ahisāna tāharī nē kihā: Asīṁ tālibāna dē nāla sidhī galabāta dē viśē utē gala kītī atē unhāṁ nū kihā ki uha āpaṇī ichā nāla jag'hā atē vakata niyukata karana.)

We can observe that data which has been translated with more than 90% accuracy can be understood easily and requires little manual correction. But data which has been translated from source website with less than 90% accuracy, needs manual cleaning and then it can be part of parallel corpora. This results in higher accurate result while translating from Punjabi to Urdu translation, which has been published later in the paper. Parallel corpus of Punjabi and Urdu sentences from different domain has been described as follows i.e.

Table 11. Domain-wise status of Punjabi-Urdu corpora

Domain Name	No of Sentences
Health	28250
Tourism	26565
Politics	12589
Entertainment	2340
Sports	5787
Books and Magazines	12569
Art and Culture	2984
Religion	2431
Literature	3320
Education	3198

2. Word-order Issues while Creating the Punjabi to Urdu Machine Translation System

Since Punjabi and Urdu languages are resource deficient languages therefore very small or no annotated parallel corpus exists for the development of Punjabi to Urdu machine translation system. More specifically, we address the problem of word order between Punjabi and Urdu text. Punjabi can be read and written from left-to-right where as Urdu can be read and written from right-to-left. Thus, different approaches will be required to test the system using the developed corpora for this language pair. By using the above said approach, we have developed a parallel corpus of around 1 lakh sentences, where three word order issues comes out as a challenge to test the system i.e.

a. *Original Punjabi text with original Urdu test* – By using this approach, we can test the system by aligning original order of Punjabi text with original order of Urdu text
b. *Original Punjabi text with reverse Urdu text* - By using this approach, we can test the system by aligning original order of Punjabi text with reverse order of Urdu text
c. *Reverse Punjabi text with Original Urdu text* - By using this approach, we can test the system by aligning reverse order of Punjabi text with original order of Urdu text

Above all three approaches can be result in correct or incorrect translation of Punjabi text to Urdu text. It means that whichever word-order results in higher accuracy can be resultant for Punjabi to Urdu Machine Translation. To test the output of the system for correct translation from Punjabi to Urdu text, two approaches can be used i.e. Statistical as well as Machine Learning. Initially, a corpus of about one lakh sentence (manually corrected) can be used to test the output of this system. Thus, by resolving above issues and addressing the discussed challenges, we will meet to desired results for PUMTS. But word-order issues also play an important role for the correct translation from Punjabi text to Urdu text, which require correct reordering model to generate higher accurate results.

V. CONCLUSION

This paper has presented many key issues & challenges occur during the development of Natural Language MTSs. Some of the issues discussed above are development of parallel corpus, lexical idioms & phrases, different character sets of language pair, segmentation, word-order, syntactic & semantic issues with translation. But all these issues & challenges are well resolved by the developers during the development of their own systems. We are in the process of development of Punjabi-Urdu MTS using hybrid approach, in which these issues and challenges are being resolved to get high quality translation system using bilingual parallel corpus. This research paper also described methodology to develop a parallel corpus. Rule-based approach will be used for post-editing to further improve the quality of translation.

REFERENCES

[1] Och, F J. (1999). An Efficient Method of Determining Bilingual Word Classes. Paper presented in EACL'99: Ninth Conference of the European Chapter of the Association for Computational Lingustics, Bergen, Norway (pp. 71–76).

[2] Och, F. J. & Ney, H. (2003). A Syntactic Comparison of Various Statistical Alignment Models. Computational Linguistics. (pp 19-51) vol 29. No 1.

[3] Koehn, P., Huang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A. & Herbst, E. (2007). Moses: Open Source Toolkit for Statistical Machine Translation. ACL Demos.

[4] Kulkarni, Amba. (2009). Urdu-Hindi-Urdu Machine Translation: Some Problems. Published in the proceedings of Conference on Language and Technology, Lahore, Pakistan.

[5] Kumar, Ajit & Goyal, Visha. (2011). Comparative analysis of tools available for developing statistical approach based machine translation system. Published in the proceedings of International conference ICISIL 2011, Patiala (Punjab), India. (pp. 254-260).

[6] Lehal, Gurpreet Singh & Saini, Tejinder Singh. (2012). Development of a Complete Urdu-Hindi Transliteration System. Published in the proceedings of COLING 2012: Posters. (pp 643-652). COLING 2012, Mumbai.

[7] Arif, Tasleem et al. (2014). An analysis of challenge in English and Urdu machine translation., Published in National conference on Recent Innovations and Advancements in Information Technology (RIAIT 2014). ISBN 978-93-5212-284-4.

[8] Okpor, M. D. (2014). Machine translation approaches: Issues and challenges. International jurnal of computer Sciences. (pp. 159-165). Vol 11. Issue 5. No 2.

[9] Kumar, Ajit & Garg, Kamaldeep. (2015) A survey of spell checkers avalaible for Hindi and Punjabi. Published in the proceedings of International Conference COMMUNE 2015, University of Kashmir, Srinagar, India. (pp. 362-365).

[10] Kumar, Ajit & Goyal, Vishal. (2015). Statistical Post Editing System (SPES) applied to Hindi-Punjabi PB-SMT system. Indian Journal of Science and Technology. Vol 8(27).

[11] Zakir, H. Mohamed & Nagnoor. M. Shafeen. (2017). A brief study of challenges in machine Translation. International journal of computer Science Issues. (pp. 54-57). Vol 14. No 2.

[12] Lopez, Adam & Post Mat. Beyond text: Five open problems in machine translation.

Study on Applied Statistical Machine Translation System between Korean & Myanmar

MyatMyoNweWai[1], Pudhpa B. Patil[2]

[1]ASEAN-India Research Training Fellow, Faculty of Computer Science, Myanmar Institute of Information Technology, Mandalay, Myanmar

[2]Computer Science and Engineering Department, B.L.D.E.A's V.P.Dr.P.G.H, College of Engineering and Technology, Vijayapur, India

Abstract: This paper is explored how to create the Automatic Machine Translation System for colloquialism between KOREAN and MYANMAR languages. To build this useful system by utilizing Statistical Machine Translation System deals with phrases. The proposal is well prepared corpus regarding the common conversation sentences for automatic translation system to be able to resolve the language barrier. So the proposed system is supported for Myanmar citizen who can communicate with Korean easily without language barrier. This paper is point out the linguistic features in both Myanmar language and Korean language by applying part of speech tagger to be factored Machine Translation system that has good translation quality. This paper is methodized to build the serviceable Machine Translation System for community.

Keywords: Statistical machine translation, Myanmar language (Burmese), Korean language, Machine translation for dialects, Parallel corpus developing.

I. INTRODUCTION

Confabulation between people is very important, but language barrier is huge issue among people. The proposed system is first proposal to touch world by making friendship regardless of diverse languages. Myanmar machine translation is still in difficulties arising from parallel corpora are scarce with daily conversation rather than formal sentences. To cross the world, utilities sentences are required to communicate each other. The current translation systems are not much suited like Google Translator or not take part as the translation for Myanmar Language, like papagonaver translator. Even though many researchers are doing to get the highest quality machine translation system by using different methods, the proposed system support for colloquialism is so far received no attention in the literature. This paper is arranged as follow: The section two is studied about the prior works of the Automatic Translation System and linguistic knowledge. The characteristic of Myanmar and Korean have been studied in contiguous sections. At the section 5, we discuss the detail methodology for building the Statistical Machine Translation system. The last section explores the conclusion and further extension of the paper.

II. RELATED WORK

Prior researches for Machine Translation Systems related with Korean or Myanmar languages by using different methodologies have been surveyed in this section. The performance of Machine Translation system is evaluated with two measurements that are BLEU [11] (BiLingual Evaluation Understudy) is used for measuring the adequacy of the translations and RIBES [4] (Rank based Intuitive Bilingual Evaluation Measure) is used for measuring which will penalize the wrong word orders. Machine Translations (MT) Systemsrelated Korean have not been frequently in previous literature. [13] refinedmachine translation system depending on rules between Korean and English with three major methods. They used HMM based Korean tagger, Korean dependency parser and the resolver of the correct case roles to overcome the language topological differences between English and Korean. They achieved BELU [11] for sample 2000 sentences amount to 2.67. But there have been remained the weakness of grammatical and stylistic sentences. Previously English and Korean parallel corpus work on SMT is not openly available. [8]used a Korean-English parallel corpus that was manually collected from travel guide books. Moreover they combined the five techniques to the baseline machine translation system to be more reliable translation output. These five techniques are adding Part of Speech (POS) tagged information and reordering word sequence while performing word alignment and phrase extraction, deleting useless words while analyzing

[1]https://github.com/ye-kyaw-thu/myPOS

word alignment results, categorize the language models to reduce the vocabulary size and appending dictionary into the training corpus before word alignment and phrase extraction process. [10] built Korean-English parallel corpus by manually aligned both Korean-English sentencesas the standard data. This corpus was built based on junior High- School evaluation data. As the future studies, we will provide the baseline SMT system using the first proposed Myanmar-Korean parallel corpus based on communication sense and Myanmar-Korean parallel corpus built from standard evaluation corpus [10] for future compassion purpose. From the point of view for some statistical translation systems that concerned with Myanmar language have been studied. [17]presented the experiments on a total of 40 language pairs that also included Myanmar-Korean parallel corpus based on Basic Travel Expressions Corpus (BTEC), which is related travelling domain sentences. [17] experimented with the three different SMT techniques: phrased based SMT, hierarchical phrase based SMT and operation sequence model. They compared the experimental results that phrased based SMT and operation sequence model approaches gave the high translation qualities that are 32.16 and 33.03 in BLEU [11] scores and both have 0.76 in RIBES scores [4] at Korean and Myanmar parallel corpus by using Conditional Random Field (CRF) word segmentation for Myanmar language. [15] mainly emphasized on phrase-based SMT, hierarchical phrase based SMT, operation sequence model, tree-to-string (T2S) and string-to-tree (S2T) methods that were used to the translation of limited quantities of travel domain data between English and Thai, Laos, Myanmar parallel corpus. The experimental results indicate the PBSMT approach produced the highest quality translations. Neural Machine Translation (NMT) model have advanced than SMT by achieving the fluency of translation that was a huge step forward. [14] explored NMT between Myanmar and Rakhine that is one of Myanmar's ethnical groups. [14] experimented three prominent NMT systems between Myanmar and Rakhine: recurrent neural networks (RNN), transformer and convolutional neural networks (CNN). In proposed system, Myanmar-Korea parallel corpora are firstly built that contains colloquialism, daily personal diary, short stories and common sentences for communication purpose.

III. MYANMAR LANGUAGE

Myanmar words are mostlymade up with more than one syllable without white spacing. Even though using spacing to separate phrases of Myanmar sentences for easier reading, normally short sentences do not have the spaces. For using spaces in Myanmar words, phrases, and even between root words and their affixes are not strictly defined and inserting spaces or not can be connived. Sometime it can omit determiner, auxiliary verb, some preposition in Myanmar language. This propriety of Myanmar language can be found in Korean language. For example:

(my) သူငယ်ချင်းကြီးတော်ရဲ့တီဗွီပျက်သွားတယ်။

(ko) 친구고모의텔레비저이고장났어요.

("My friend's aunt's TV is broken." in English).

In the above sentences, we can omit possessive pronoun in both Myanmar and Korean language but the meaning already has the possessive pronoun sense. Generally, the Korean language is mutually intelligible with the Myanmar language and has the same word order (namely, subject-object-verb (SOV)). Examples of parallel sentences in Myanmar (my) and Korean (ko) are given as follows:
(my) ကိုရီးယားစကားပြောတက်လား။

(ko) 한국말을하세요?

("Do you speak Korean?" in English)

IV. KOREAN LANGUAGE

Korean language isa highly stickyinwhichwords is constructed all consonants are reused as close consonants called 받침. These close consonants have to correctly pronounceto get the correct meaning. In Korean language, the spaces between words may vary the meaning. Eojeols of Korean Sentences are found as words or phrases that are formed by joining content and functional morphemes. An eojeol is surface level from more than one combined morpheme. These eojeols can be break down the Korean sentence by inserting space as the basic segmentation. For example: 친구가,친구is a main noun morpheme and 가 is a preposition that form as a functional morpheme. This가can be used asa nominative case marker (NOM). The meaning of the word친구가is "Friend is" in English but this is formed as one phrase in Korean. While building the parallel corpus manually, we insert space before the functional morphemes, such as 가to distinguish from main content and functional morphemes easily. Based on the previous letter vowel or constant, we can decide which the nominative case marker가or이may use. To explain the structure of Korean sentence that including eojoles, let see the following Figure 1:

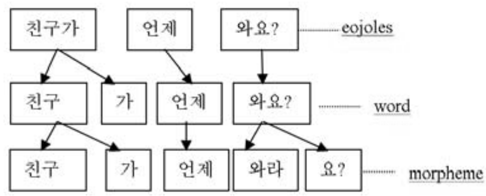


Fig. 1. Essence of Korean Sentence "친구가언제와요?(When is friend coming?)"

Figure 1 show how Korean sentence can be broken down into morphological parts such as 친구is a Noun, 가is a preposition, 언제 is question noun , 와라 is verb and 요 is ending.

V. MOSES STATISTICAL MACHINE TRANSLATION SYSTEM

Moses toolkit [9] is utilized to develop the Statistical Machine Translation system between Korean and Myanmar. The figure 2 is explained how to process and which toolkits will be used for building SMT. In the figure 2, the solid line type boxes have to be implemented for baseline SMT system, dashed line type boxes may be added to be factored translation system.

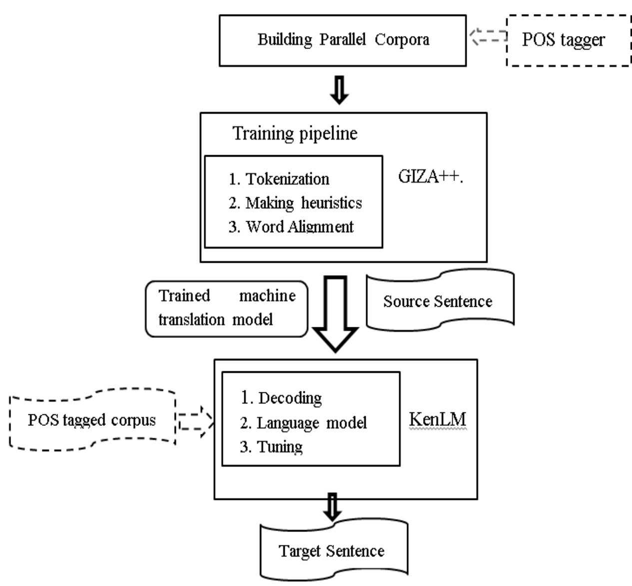


Fig.2. System flow chart of the SMT

Building Parallel Corpora

To communicate other's languages with dialects are most difficult barriers. As the first step of SMT, Myanmar and Korean dialogic corpus with monolingual and parallel corpus is developing to solve the language barrier. To build the dialogic corpus is taking time and need to understand the both languages. To develop Applied Automatic Machine Translation system, domain expert is essential to take part in processes of building qualified parallel corpus, linguistic analysis and evaluations. The crux of the parallel corpora is the essential step towards improving the translation quality through linguistics factors for the language pair. To be a reliable translation system, the lots of good quality data is required. So this paper is built the well prepared conversational sentences of both languages. The proposed system will be built by not only based line translation system such as phrase based model but also factor translation system. We studied the linguistic knowledge of both languages: such as dictionary form (lemma), part-of-speech (POS) and morphological tags. For emphasizing the application of factored models to the machine translation between Myanmar and Korean is not found in prior researches. We will used POS tagged information as a factor to be a factored machine translation that increases the performance outputs. Because of limited training data, it may cause the data sparseness problems. Translation model with factored information can solve this problem. Therefore the factored machine translation is extended to develop for comparing with phrase based SMT. Figure 3 shows the mapping structure of phrased based SMT between Myanmar and Korean.

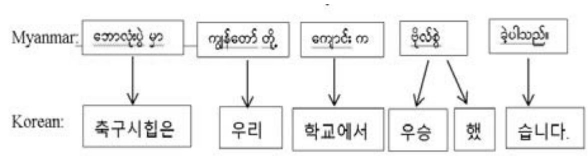


Fig.3.Sample process of phrase based translation

The POS information as the translation factors is needed to insert into parallel corpus for building the factor translation. Examples of POS tagged data are given below, in the format "word|POS-tag":

Korean: 얼마인가|ncpa + 요|ncn+?|sf

Myanmar: ဘယ် |PRN.Questionလောက်| Part. Support ပါ|Part.Sport |SF.Interrogative |UNK

Both languages are syllabic, and often word tokens in both languages have similar grammatical functions : the preposition of the Korean word "가" (ga) corresponds to the the preposition of Myanmar word " " (ka). Moreover most of the interrogative sentences for both languages can be easily distinguished because they have special endings: the Korean interrogative word "까" (kaa) corresponds to the Myanmar interrogative word " " (lae) . Therefore, due to the more direct correspondence, we believe that mappings involving POS tags between the languages are likely to be more effective between Myanmar and Korean.

Training Pipeline

The parallel corpus is taken into the training pipeline that produces a machine translation model by using the collection of tools. We have to change the format of parallel corpus that is befitting the sufficient facility in word segmentation. The corpus can be factored with POS tagged information, so the phrase mapping can be extracted for each factor and broken up into steps such as translation step and generation step. Firstly the analysis of the POS tagging for Korean and Myanmar is studied. Myanmar language has been still in under resources language and there is no standard word segmentation and POS tagger. To handle Myanmar word segmentation task, [20] is reported the machine learning framework. CRF ++ toolkit that is the statistical conditional random field approach has been used for training POS tagger and word segmentation for Myanmar. Myanmar has many compound words such as most of the noun, adjectives and verbs are suffixed or affixed with Post Positional Marker (ppm) or particles. This can be tag each word by placing the POS tag of compound word before and end of the pattern. [2] reported POS tagger with CRF ++ toolkit by using feature set of token unigram at their relative position. Like as [2], We trained the POS tagger with CRF++ based on myPOS[1], the train data has been segmented and tagged as follow:

Input: အမအေိမ် ကိုအရမ်းလွမ်းတယ်။

Output: အမေ n

အိမ် n

ကို ppm

အရမ်း adv

လွမ်း v

တယ် ppm

။ punc

HanNanum[6] is used for Part-of-Speech and Morphological Analyzer for Korean POS tagging. The following sentence is input and POS tagging can get as the output by using HanNanum:

Input: 축구시힙은우리학교에서우승했습니다.
Output:
축구시힙은
축구시힙/ncn+은/jxc
우리

우리/npp
학교에서
학교/ncn+에서/jca
우승했습니다.
우승/ncpa+하/xsva+었/ep+습니다/ef+./sf

Let consider the processes in factored model, the translation of the Korean word (친구) Korean into Myanmar. Firstly we analyzed친구들이 phrase (eojeols) that is broken up into 친구 as lemma, NN as part of speech, count is plural, case is nominative and gender is neutral.

The following translation mappings and generation have to be performed in factored machine translation:

i. Mapping lemmas between source and target language
  •친구→သူငယ်ချင်း, မိတ်ဆွေ, အပေါင်းအသင်း, ဘော်ဒါ
ii. Mapping with POS and more factors
  •친구들이|NN|plural|nominative|neutral→NN|singular,NN|plural
iii. Generating target word/phrase according with POS tagged
  •သူငယ်ချင်း |NN|plural→သူငယ်ချင်းများ
  •သူငယ်ချင်း |NN|singular→သူငယ်ချင်း
  •မိတ်ဆွေ|NN|plural→မိတ်ဆွေများ
  • ...

To do word alignment between source and target languages, it should be use GIZA++ in phrased based model. We will use grow-diag-final and heuristic for symmetrized alignment. For reordering process, a distance based reordering model will be used. As the final step, the generation model is built from target side of the parallel corpus by applying. Finally the configuration file is come out for decoder.

Decoding and Language Model

The decoder translates the source sentence to target language by using a trained machine translation model and source sentence. The decoder can translate the target output easily by using the language model that is statistical model with monolingual data of target language. In this system, we will use KenLM[3] external tools for language model building. We will also use POS tag information to language model building to be factored translation model. Mapping involving POS tags between the Korean and Myanmar will be more effective in reducing the vocabulary size. The last processing of machine translation system is tuning, where repeatedly translates the source sentences by comparing the results to references sentences to produce the best available translations. The proposed system will be used Minimum error rate training (MERT) for tuning target output, and Moses decoder (version 2.1.1) is expected to use for decoding. To build the proposed system, we will use the default parameters that offer from Moses.

VI. CONCLUSION

Myanmar has kind heart people who are willing to communicate with friends of all over the world. SMT of Myanmar and Indian languages will be built as further extension. To adapting this proposed system easily by just developing Myanmar and Kannada Language dialogic corpus. We will analyze the experimental results of baselines phrase based SMT, factored SMT and factored SMT with POS language modeling as future comparison. To get the good translation system, the heart of SMT is good quality parallel corpus. So let hand up together by regardless language barrier.

ACKNOWLEDGMENT

I absolutely thank of gratitude to ASEAN-India Research Training (AIRTF), Department of Science and Technology (DST) in India and Federation of Indian Chamber of Commerce and Industry (FICCI) for providing fund to do this research work.

I would like to thank Mr. Shim Jeahwa, Member of KOICA and Ms. MyintZuKyaw, Studied Korean Language at Mandalay University of Foreign Language for their very useful translation and kindly support during doing this research.

REFERENCES

[1] Choi, J. D. and Palmer, M. (2011). Statistical dependency parsing in Korean: From corpus generation to automatic parsing. In Proceedings of the Second Workshop on Statistical Parsing of Morphologically Rich Languages, pages 1–11, Dublin, Ireland. Association for Computational Linguistics.

[2] Hinn Thu Zar Aye, Chenchen Ding, Win Pa Pa, KhinThandarNwet, Masao Utiyama and EiichiroSumita, "English-to-Myanmar Statistical Machine Translation Using a Language Model on Part-of-Speech in Decoding", proceeding of 15th International Conference on Computer Application, (ICCA 2017), February 16, 2017, Yangon, Myanmar, pp. 66

[3] Heafield, Kenneth, "KenLM: Faster and Smaller Language Model Queries, " in Proceedings of the Sixth Workshop on Statistical Machine Translation, WMT'11, Edinburgh, Scotland, 2011,pp.187-197

[4] Isozaki, Hideki and Hirao, Tsutomu and Duh, Kevin and Sudoh, Katsuhito and Tsukada, Hajime ,"Automatic evaluation of translation quality for distant language pairs," Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing,2010,pp.944-952

[5] Heafield, Kenneth, "KenLM: Faster and Smaller Language Model Queries, " in Proceedings of the Sixth Workshop on Statistical Machine Translation, WMT'11, Edinburgh, Scotland, 2011,pp.187-197

[6] Jungyeul Park, Jeong-Won Cha, and Seok Woo Jang. 2011. Korean POS Tagging using Noisy Channel Model with Syllable Lattice Based OOV Words Resolution. Information - an international interdisciplinary journal, 14(8):2835–2843.

[7] Koehn, Philipp and Hoang, Hieu and Birch, Alexandra and Callison Burch, Chris and Federico, Marcello and Bertoldi, Nicola and Cowan, Brooke and Shen, Wade and Moran, Christine and Zens, Richard and Dyer, Chris and Bojar, Ondřej and Constantin, Alexandra and Herbst, Evan, A. Constantin, and E. Herbst, "Moses: Open source toolkit for statistical machine translation," Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, 2007,pp. 177–180.

[8] Lee. Jonghoon, Donghyeon Lee, and Gary Geunbae Lee. 2006. Improving phrase-based Korean-English statistical machine translation. In Proceeding of: INTERSPEECH 2006 - ICSLP, Ninth International Conference on Spoken Language Processing, Pittsburgh, PA, USA.

[9] MOSES version 0.91, A Factored Phrase-based Beam- search Decoder for Machine Translation. URL: http:// www. statmt. org/ moses/

[10] P. Jungyeul ,Jeen-Pyo Hong, and Jeong-Won Cha. 2016. Korean Language Resources for Everyone. In Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation (PACLIC30), pages 49–58,Seoul, Korea.

[11] Papineni, Kishore and Roukos, Salim and Ward, Todd and Zhu, WeiJing, "BLEU: a Method for Automatic Evaluation of Machine Translation," Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL'02,Philadelphia,Pennsylvania,2002, pp.311–318

[12] Philipp Koehn and Hieu Hoang, "Factored Translation Models", In Proc. of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, ACL, Prague, June 2007, pp. 868-876.

[13] RyuByong Rae, Kim Youngkil, YuhSanghwa and Park Sangkyum, "A Korean English Machine Translation System based on Idiom Recognition and Fail Softening", April 2000, in MT Summit VII, Singapore, 469-475.

[14] ThazinMyintOo, Ye Kyaw Thu, Khin Mar Soe, "Neural Machine Translation between Myanmar (Burmese) and Rakhine (Arakanese)", Proceeding of the Sixth Workshop on NLP for Similar Language, Varieties and Dialects", June 2019, Minneapolis, United States pp 80 -8.

[15] Win Pa Pa, Ye Kyaw Thu ,Andrew FinchandEiichiroSumita, "A Study of Statistical Machine Translation Methods for Under Resourced Languages," 5th Workshop on Spoken Language Technologies for Under resourced Languages (SLTUWorkshop),09-12May,2016,Yogyakarta, Indonesia, Procedia Computer Science, Volume 81,2016 ,pp.250–257.

[16] X. Peng, Jaeho Kang, Michael Ringgaard, and Franz Och. 2009. Using a Dependency Parser to Improve SMT for Subject-Object-Verb Languages. In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pages 245–253, Boulder, Colorado. Association for Computational Linguistics.

[17] Ye Kyaw Thu, Andrew Finch, Win Pa Pa, and EiichiroSumita ,"A Large scale Study of Statistical Machine Translation Methods for Myanmar Language, " in Proceeding of SNLP2016, February10-12,2016.

[18] Ye Kyaw Thu, Andrew Finch, Akihiro Tamura, EiichiroSumita and Yoshinori Sagisaka, "Factored Machine Translation for Myanmar to English, Japanese and Vice Versa", Proceedings of the 12th International Conference on Computer Applications (ICCA 2014), February 17~18, 2014, Yangon, Myanmar, pp. 171-177

[19] Ye Kyaw Thu, Andrew Finch, Masao Utiyamaand EiichiroSumita ,"A Large- scale Study of Statistical Machine Translation Methods for Khmer Language" 29th Pacific Asia Conference on Language, Information and Computation, October30-November1,2015,Shanghai, China,pp.259-269.

[20] Ye Kyaw Thu, Andrew Finch, Yoshinori Sagisaka, and EiichiroSumita. 2013. A study of Myanmar word segmentation schemes for statistical machine translation. Proceeding of the 11th International Conference on Computer Applications, pages167– 179.

Survey on Pitch Estimators for Speech Synthesis

Shraddha G. Revankar[1], K.T. Deepak[2], C.I. Nalini[1], S.R. Nirmala[1]

[1]K.L.E. Technological University, India
[2]IIIT-Dharwad, India

Abstract: Pitch estimation in speech signal plays an important role in several speech processing applications. Also, as an application to pitch estimation, speech synthesis using VOCODER's is discussed in the paper. A detailed literature survey on a large bunch of pitch estimation algorithms and VOCODERs in terms of their methodology, innovations in the algorithm, performance and robustness is presented in this paper. It also provides the short summary on various performance parameters available for the performance evaluation of pitch estimators and VOCODERs. This paper aims to provide the detailed survey, which can help researchers to choose an appropriate pitch estimation algorithm for their problem statement and also focuses on giving the essence of speech synthesis using VOCODERs.

Keywords: Pitch estimation algorithms; VOCODERS; speech synthesis; error metrics;

I. INTRODUCTION

Pitch is the main factor of speech signal and finds its applications in gender identification[1], speaker verification[2], intonation[3], speech synthesis[4], speech coding [5] and in speech recognition of tonal languages[6]. Pitch is a measure of frequency at which the vocal chords vibrate to produce voiced sounds, often referred as fundamental frequency (F0).Pitch is referred as the quantity that is attributed more to the perceptual phenomena. However, in the literature both pitch and fundamental frequency are interchangeably used. The estimation of pitch contour is a challenging task, as there is pitch variation while speaking. Thus obtaining an accurate F0 is one of the major problems allied with speech signal analysis and processing.

When it comes to pitch estimation, the pitch estimation from normal speech, telephone speech, singing signals, emotional and pathological speech is required to be discussed. Estimation of pitch from normal speech is relatively simple as compared to other signals mentioned above. Hence, an attempt is made to carry out the survey on various pitch estimators that could be used by the researchers depending upon the type of speech signal used for the analysis.

Speech analysis, synthesis and manipulation on the basis of vocoders are used in various kinds of speech research. Speech synthesis using vocoder requires three speech parameters: fundamental frequency, spectrogram, aperiodicity map with voicing information. The most commonly used VOCODERS as per literature study are STRAIGHT, TANDEM STRAIGHT and WORLD vocoder. The detailed explanation is provided in the next sections.

This paper is organized as follows: The different pitch estimation algorithms are briefly summarized in Section II. Performance evaluation metrics for pitch estimation available are discussed in Section III. The vocoders selected for comparison are briefly summarized in detail in Section IV. Speech quality assessment of synthesized speech is discussed in Section V. Conclusion is discussed in Section VI.

II. PITCH ESTIMATION ALGORITHMS

Voice can be characterized by the various parameters like pitch, duration, intensity, prosody etc. Among all these the pitch of the speech signal is one of the main characteristics. But when we consider the pitch in acoustics technologies the parameter is called Fundamental Frequency. However extraction of F0 is not trivial task when considered for the normal human speech, but is not the same when considered for telephone speech. The characteristics of the telephone speech make the process complex for F0 extraction. There are number of algorithms presented in the past for the F0 extraction of original speech and telephone speech. Some of them are discussed below. From the study it can be stated that most of the F0 extractors rely on the some basic techniques for the extraction in each frame, they are i) auto-correlation based F0 extraction ii) identifying the spectrum or cepstrum for F0-related peaks iii) Epoch based approaches or, iv) by pattern classification(Neural network) approach [13].

1. Signal Processing based Approaches

In [7] the survey of various F0 extractor algorithms is stated. They include Praat, STRAIGHT, RAPT and YIN. Praat has been titled as "the de facto standard speech analysis program" and is the flexible tool designed for acoustic analysis of speech. The methods that Praat incorporates for pitch extraction are autocorrelation, cross-corelation and Summation of sub-harmonics (SHS) functions, but the default method adopted in Praat is the autocorrelation. Another extractor STRAIGHT [14] developed by Hideki Kawahara is the channel vocoder that incorporates the F0 extractor algorithms; they are based on normalized autocorrelation, that makes an effort in reducing the F0 extraction errors. Extractors used by STRAIGHT is nearly detect free (NDF) algorithm. This method is designed for the accurate F0 extractor that solves the problem of "missing fundamental". NDF employs multicue scheme for F0 extraction. RAPT algorithm [9] being another method used in [7] for evaluation, is the F0 extractor which uses cross correlation function. RAPT provides F0 contour as well as reliable voicing decisions for a large temporal context by simultaneously considering all the possibilities. YIN algorithm [10] also used in [7] is also based on the autocorrelation function, but with certain modifications to overcome the errors faced earlier. This algorithm is best suited for high pitched voices and music. YIN has an issue of over-identification of voicing that results in high error rates when compared to other extractors presented in [7] also the default settings (with 40 Hz as the lower pitch threshold) does not hold good for speech signal. Hence analyzing the results, it has been proved in [7] that when compared to all other algorithms Praat is most efficient F0 extractor. Saw-tooth Waveform Inspired Pitch Estimator (SWIPE) [11] is another F0 extractor algorithm that is mainly used for music and speech signals. SWIPE algorithm solves the problem of sub-harmonic errors found in most of the pitch extractors by considering only first and prime harmonics. It has been proved in [11] that this algorithm has outperformed other algorithms on several musical/ speech instruments and appreciable results have been obtained for pathological speech.

2. Spectrum and Cepstrum based Approaches

The idea of obtaining pitch contours by exploiting summation of residual harmonics is not new; they are widely used as follows. The Summation of Residual harmonics(SRH) algorithm [11],[15] is an state of art algorithm that is robust in pitch tracking for noisy conditions and is also capable of determining voicing and non-voicing segments of the speech. Spectral amplitude auto-correlation (SAC) also mentioned in [11] is the pitch estimator that employs autocorrelation of amplitude spectrum, but is not that robust for noisy data. As a conclusion from [11]it has can be claimed that compared to other two algorithms, SAC gave good results for automatic wide-range F0 extraction. Other extractors mentioned in [15] like AMDF, SIFT based on auto-correlation, SHRP based on sub-harmonic to harmonic ratio, Summation of speech harmonics (SSH) and TEMPO algorithm available in STRAIGHT based on fixed point analysis are implemented. From the results showcased in [15] it can be concluded that SRH performs best in the noisy conditions, but shows comparable results with other extractors when considered the clean speech case.

3. Epoch based Approaches

Popular epoch extraction algorithms used for F0 estimation are Zero Frequency Filtering (ZFF),Dynamic Programming Phase Slope Algorithm (DYPSA), Speech Event Detection using the Residual Excitation, Dynamic Plosion Index (DPI),Single Pole Filtering (SPF) and YAGA (Yet Another GCI/GOI algorithm) [16] work satisfactorily good for clean and noisy speech signals, but their performance degrades when considered for the band limited signals like telephone speech as fundamental frequency components are attenuated . Hence the existing ZFF method is refined for improving the performance of the epoch-identification in telephonic speech. The methodology proposed in [16] exploits the properties of Chebyshev polynomial interpolation and through the Hilbert envelope (HE) the frequency components around the fundamental frequency are reinforced. The results show that by incorporating Chebyshev polynomial interpolation in the ZFF algorithm give better epoch identification rate compared to other epoch extraction techniques. Robust epoch and pitch estimator(REAPER) [11] developed by D. Talkin (Google) is an algorithm that is based on the identification of voiced epochs and then estimate the glottal closure instants(GCI) that is the used to obtain F0 contours as the inverse between successive GCI's.

4. Neural Network based Approaches

A Convolutional Representation for Pitch Estimation (CREPE) [17] operates on the audio signals for pitch estimation. The pre-trained model is also provided by the author, that is freely downloadable. CREPE is robust for noisy signals too. The results in [17] shows comparison of CREPE with p-YIN and SWIPE, which illustrates CREPE outperforms both of them and currently is the state-of- art neural network based pitch estimator. In [18] pitch estimation is obtained by implementing Deep Neural Networks (DNN) and Recurrent Neural Networks (RNN). In this approach, static frame-level acoustic features is used to train DNN model and sequential frame-level acoustics features is used to train RNN model. Viterbi-decoding is done to extract the pitch contours produced by DNN and RNN. Due to powerful learning capabilities, DNN is expected to produce more accurate probabilistic outputs. On the other hand, RNN is best suited for sequential data, hence to utilize temporal information of speech signal, RNN is best suited. The results summarize that DNN and RNN based pitch extractor performs best for all the noise levels, hence being the robust among the extractors mentioned in [18] Pitch tracking in [19] is achieved by the using less amount of data using some speech features and traditional Machine Learning. The overall idea of this method is to extract features from various domains (time, spectral and cepstrum), later these features are fed to classifier for making voicing decisions. From these voicing decisions, the F0 contour is obtained by applying median filter that outperforms and produces lower gross error rates. Frequency domain pitch extractor is proposed in [20] which is the combination of artificial neural network (ANN) and long-term harmonic feature analysis. Again this model performs best for noisy conditions. Another neural net based pitch tracker as proposed in [21] shows that the waveform-based and the peak-based feature sets lead to pitch peaks discrimination. Another state-of-art algorithm that employs Bidirectional Long Short-Term Memory Recurrent Neural Networks (BLSTMRNN) [22], the model is trained to map the noisy to clean speech features. It is best algorithm so far implemented for the strong noise and the uniqueness is that it provides good generalization to new speakers. In [23] the supervised learning algorithm for pitch detection using CNN is proposed. The methodology is simple, pitch candidates are determined by CNN and dynamic programming for pitch tracking. Lastly another paper cited provides another algorithm that deals with smoother pitch tracking algorithm [24] using neural networks. Neural Network based approaches are best designed for noisy signals.

5. Pitch Estimators for Telephone Speech

Yet another algorithm for pitch tracking (YAAPT)[12] is the most robust algorithm for high quality and telephone speech. Base of this algorithm is the normalized cross-correlation proposed by David Talkin. Restoring the partially weaker F0 components by non-linearly processing the speech signal and intelligibility of selecting the 'best' F0 contour are the major innovations made in YAAPT that makes it best suited for telephone speech. Discrete Logarithmic Fourier Transform (DLFT) [25] is specially proposed for telephone speech and prosodic modeling. It is similar to the concept applied in SRH algorithm i.e, it uses the logarithmically sampled spectral representation of speech. However if differs from SRH as processing and tracking algorithms are quite different.

III. PERFORMANCE EVALUATION METRICS

As per the literature study, to evaluate performance of pitch estimators, the four most commonly used metrics are defined in [26]. They are:

1. Gross Pitch Error (GPE): Checks if both pitch tracker and ground truth consider that the proportion of frames detected are voiced, for which relative pitch error is defined. It is usually 20% in speech studies.
2. Fine Pitch Error (FPE): For the frames that do not have gross pitch error, the standard deviation is calculated for the distribution of relative error values. The estimated and ground truth voiced/unvoiced decisions must then be voiced.

3. Voicing Decision Error (VDE): This metric defines the proportion of incorrect decisions of V/UV frames made.
4. F0 Frame Error (FPE): This metric defines the proportion of frames for which an error is made. The overall performance of the pitch trackers can be assessed by this single measure.

IV. VOCODERS FOR SPEECH SYNTHESIS

Normally speech synthesis is often described as the sequence of procedures required to convert text into speech. But the other way to think about speech synthesis is to start from the vocoding idea. The speech signal is converted into more compact representation that can be transmitted using a far narrower frequency bandwidth. Later, the speech can be synthesized back, by retrieving the parameters. In the conceptualization of source-filter model of speech model, the vocoder is the foundation. Speech analysis, modification, voice conversion, compression and resynthesis are the other applications of vocoders.

Literature shows various vocoders implemented so far. Some of them are discussed here. VOCODER is composed of F0 estimator, aperiodicity and spectrogram analyzer. The difference between various VOCODERS is the usage of different F0 extractors, aperiodicity mapper and spectrogram generator. STRAIGHT vocoder [14] developed by Hideki Kawahara, is the state-of-art vocoder that is used for speech synthesis as it is used for real time analysis and synthesis. TANDEM STRAIGHT [27] is the simplified version of STRAIGHT that solves the problem of STRAIGHT of degrading the speech quality of synthesized speech by inculcating the simplified algorithm, but is hard to use in real time speech analysis. WORLD VOCODER [28] is designed not only to obtain the high quality speech synthesis, also obtain real time processing. VOCAINE vocoder [29] developed by Yannis Agiomyrgiannakis (Google) is implemented to overcome the shortcomings of STRAIGHT. Statistical VOCODER [30] using Wavenet architecture is implemented. This VOCODER is designed for improving the overall quality male and female speakers by using cepstrum coefficients and filter bank features.

The parameter F0 is a useful parameter in speech synthesis, which is defined as the inverse of a period of periodic signal. The vocoder performance tends to be fragile in extracting the source and vocal tract parameters in the noisy environment, hence degrading the speech quality. Spectral envelope estimation tries to minimize these degradations by keying on the spectral peaks and attains robustness to acoustic noise[31]. Only replicating the F0 and amplitude precisely is not enough to yield the good quality synthetic speech. Hence along with amplitude and F0, aperiodicity is also an important attribute.

V. EVALUATION OF SYNTHETIC SPEECH QUALITY

The quality of synthetic speech is calculated by conducting two types of tests, they are subjective tests and objective tests. Objective test namely Perceptual Evaluation of Speech Quality (PESQ) [33] signal to noise ratio (SNR)-Global SNR and Segmental SNR are discussed. Subjective tests like MUSHRA test [34] and Mean Opinion score are also discussed in detail.

1. Objective and Subjective Tests

In this section, the objective tests like PESQ, SNR and subjective tests like MUSHRA tests and MOS are explained in detail.

1. PESQ: It is defined in the ITU-T P.826 standard. It is a objective method to test the quality of speech that pass through telecommunication networks. The positive and negative values of score define the following; positive difference indicates that the noise is added in the signal spectra and negative difference indicate that spectral component is omitted. The objectivity is based on the traditional MOS method. In MOS method, group of expert listeners rate the voice quality to a value ranging from bad (0) to excellent(5) PESQ is measured by comparing the original signal and synthetic signal received when passed through communication system. After the analysis, a score of range -0.5 to 4.5 is produced, in which higher score means better speech quality.
2. SNR: A SNR ratio compares a level of signal power to a level of noise power. It is most often expressed as a measurement of decibels(dB).Higher numbers generally mean a better specification, since there is more

useful information(the signal) than the unwanted data (the noise). There are two types Global SNR and Segmental SNR.

3. Global SNR: It deals with entire signal; Segmental SNR: Instead of working on whole signal, it calculates the average of SNR values of short segments(15 to 20ms)
4. Segmental SNR gives better results than SNR for waveform encoders, but it gives very bad results for vocoders
5. MOS: In telecommunications, a Mean Opinion Score is a ranking of the quality of voice and video sessions. Most often judged on a scale of 1 (bad) to 5 (excellent), Mean Opinion Scores are the average of a number of other human-scored individual parameters.
6. MUSHRA Test: The method described in the ITU-R BS.1534-1 standard, commonly known as MUSHRA (MUltiple Stimulus with Hidden Reference and Anchors), is widely used. MUSHRA requires fewer participants to obtain statistically significant results when compared to MOS.

VI. CONCLUSION

In this short review, we have presented different pitch estimators that are suitable for normal, music, pathological and telephone speech signals and performance evaluation metrics for evaluating the pitch estimated. Also, the different vocoders for speech synthesis along with the performance metrics for synthetic speech quality assessment have been presented. It can be concluded that the various estimators cited perform better than the other depending on the type of signal used for experimentation and their uniqueness in terms of methodology. Overall signal processing based, epoch based, spectrum based approaches and neural network based approaches performance is discussed and researchers can rely on the proposed paper for the pitch estimation based experimentations. Also, it can be concluded that most commonly used vocoders for speech synthesis according to the literature are STRAIGHT, TANDEM-STRAIGHT and WORLD and each of them performs better depending upon their methodology used for the purpose.

REFERENCES

[1] Kanani, I., Shah, H., Mankad, S. H. (2019) “On the Performance of Cepstral Features for Voice-Based Gender Recognition. In Informationand Communication Technology for Intelligent Systems .”Springer.

[2] Y. Shi, J. Zhou, Y. Long, Y. Li, and H. Mao. (2019) “Addressing text-dependent speaker verification using singing speech,”Applied Sciences”Applied Sciences9(13): 2636.

[3] C. Tang . (2019) “Cortical representation of vocal pitch during speech perception in human superior temporal gyrus.” PhD thesis, UCSF.

[4] Mary, L. (2019). Significance of Prosody for Speaker, Language, Emotion, and Speech Recognition. In Extraction of Prosody for Automatic Speaker, Language, Emotion and Speech Recognition (pp. 1-22). Springer.

[5] Narendra, N. P., Airaksinen, M., Story, B., Alku, P. (2019). Estimation of the glottal source from coded telephone speech using deep neuralnetworks. Speech Communication, 106, 95-104

[6] Arora, A., Kadyan, V., Singh, A. (2019). Effect of Tonal Features on Various Dialectal Variations of Punjabi Language. In Advances in Signal Processing and Communication (pp. 467-475). Springer, Singapore.

[7] Strmbergsson, S.(2016). Today’s Most Frequently Used F0 Estimation Methods, and Their Accuracy in Estimating Male and Female Pitch in Clean Speech. In INTERSPEECH (pp. 525-529).

[8] Tan, L., Karnjanadecha, M. (2003, September). Pitch detection algorithm: autocorrelation method and AMDF. In Proceedings of the 3rd international symposium on communications and information technology (Vol. 2, pp. 551-556).

[9] Talkin, D. (1995). A robust algorithm for pitch tracking (RAPT). Speech coding and synthesis, 495, 518.

[10] De Cheveign, A., Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. The Journal of the Acoustical Society of America, 111(4), 1917-1930.

[11] Villavicencio, F., Bonada, J., Yamagish, J., Pucher, M. (2015). Efficient Pitch Estimation on Natural Opera-Singing by a Spectral Correlationbased Strategy. IPSJ Technical Report.

[12] Kasi, K., Zahorian, S. A. (2002, May). Yet another algorithm for pitch tracking. In 2002 IEEE International Conference on Acoustics, Speech,and Signal Processing (Vol. 1, pp. I-361). IEEE.

[13] Hosom, J. P. (2005). F0 Estimation for Adult and Children's Speech. In Ninth European Conference on Speech Communication and Technology.

[14] Kawahara, H. (2006). STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds.Acoustical science and technology, 27(6), 349-353.

[15] Drugman, T., Alwan, A. (2011). Joint robust voicing detection and pitch estimation based on residual harmonics. In Twelfth Annual Conference of the International Speech Communication Association.

[16] Gowri, B. G., Soman, K. P., Govind, D. (2018). Improved Epoch Extraction from Telephonic Speech Using Chebfun and Zero Frequency Filtering. In Interspeech (pp. 2152-2156).

[17] Kim, J. W., Salamon, J., Li, P., Bello, J. P. (2018, April). Crepe: A convolutional representation for pitch estimation. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 161-165). IEEE.

[18] Han, K., Wang, D. (2014). Neural network based pitch tracking in very noisy speech. IEEE Procedia Computer Science 00 (2019) 000–000

[19] Drugman, T., Huybrechts, G., Klimkov, V., Moinet, A. (2018). Traditional Machine Learning for Pitch Detection. IEEE Signal Processing Letters, 25(11), 1745-1749.

[20] Wang, D., Loizou, P. C., Hansen, J. H. (2014). F0 estimation in noisy speech based on long-term harmonic feature analysis combined with neural network classification. In Fifteenth Annual Conference of the International Speech Communication Association.

[21] Barnard, E., Cole, R. A., Vea, M. P., Alleva, F. A. (1991). Pitch detection with a neural-net classifier. IEEE Transactions on Signal Processing,39(2), 298-307.

[22] Liu, B., Tao, J., Zhang, D., Zheng, Y. (2017, March). A novel pitch extraction based on jointly trained deep BLSTM recurrent neural networks with bottleneck features. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 336-340). IEEE.

[23] Su, H., Zhang, H., Zhang, X., Gao, G. (2016, March). Convolutional neural network for robust pitch determination. In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 579-583). IEEE.

[24] Ferro, M., Tamburini, F. (2018). A new Pitch Tracking Smoother based on Deep Neural Networks.

[25] Wang, C., Seneff, S. (2000, June). Robust pitch tracking for prosodic modeling in telephone speech. In 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100) (Vol. 3, pp. 1343-1346). IEEE.

[26] Yegnanarayana, B., Murty, K. S. R. (2009). Event-based instantaneous fundamental frequency estimation from speech signals. IEEE Transactions on Audio, Speech, and Language Processing, 17(4), 614-624.

[27] Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., Banno, H. (2008, March). TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In 2008 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 3933-3936). IEEE.

[28] Morise, M., Yokomori, F., Ozawa, K. (2016). WORLD: a vocoder-based high-quality speech synthesis system for real-time applications. IEICE TRANSACTIONS on Information and Systems, 99(7), 1877-1884.

[29] Agiomyrgiannakis, Y. (2015, April). Vocaine the vocoder and applications in speech synthesis. In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4230-4234). IEEE.

[30] Adiga, N., Tsiaras, V., Stylianou, Y. (2018, April). On the use of WaveNet as a statistical vocoder. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5674-5678). IEEE.

[31] Paul, D. (1981). The spectral envelope estimation vocoder. IEEE Transactions on Acoustics, Speech, and Signal Processing, 29(4), 786-794.

[32] Lorenzo-Trueba, J., Drugman, T., Latorre, J., Merritt, T., Putrycz, B., Barra-Chicote, R. (2018). Robust universal neural vocoding. arXiv preprint arXiv:1811.06292.

[33] Min, G., Zhang, C., Zhang, X., Tan, W. (2019). Deep Vocoder: Low Bit Rate Speech Compression of Speech with Deep Autoencoder. arXiv preprint arXiv:1905.04709.

[34] Zielinski, S., Hardisty, P., Hummersone, C., Rumsey, F. (2007, October). Potential biases in mushra listening tests. In Audio Engineering Society Convention 123. Audio Engineering Society

CICD Acoustic Modeling based on Monophone and Triphone for HINDI Speech Recognition

Shobha Bhatt[1], Anurag Jain[1] and Amita Dev[2]

[1]University School of Information and Communication Technology, GGSIP University, New Delhi, India
[2]Indira Gandhi Delhi Technical University for women, New Delhi, India
[1]bhattsho@gmail.com, [2]anurag@ipu.ac.in; [3]amita_dev@gmail.com

Abstract: Automatic recognition of spoken words is an active area of research. To develop a speech recognition system different acoustic units are used, and the performance of the speech recognition system is primarily affected by these acoustic units. Context independent(CI) models based on monophones are used used to overcome the constraint of the requirement of large instances of word utterance for training in large vocabulary continuous speech systems. For speech recognition, neighboring phonemes play a significant role in the pronunciation of a phoneme. Context-dependent (CD) models based on triphone are used to reduce the effect of context by considering the left and right context of monophones. This paper presents a comparative analysis of monophone based and triphone based acoustic unit speech recognition system. Continuous Hindi speech recognition was performed using widely known Hidden Markov Model (HMM) based toolkit HTK. For feature extraction, the perceptual linear predictive (PLP) coefficients were used. Experiments were conducted for both monophones and Triphones based speaker-dependent continuous Hindi Speech Recognition. The monophone based system achieved word accuracy of 65.51% and triphone based system achieved the maximum word accuracy of 69.11%. Research findings can be used to improve the performance of the speech recognition system.

Keywords: Speech recognition, Context-dependent, Triphone, Hidden Markov model, Acoustic Model

I. INTRODUCTION

Speech recognition systems are developed using mostly phoneme based and word-based models. Phoneme based systems are used to develop a large vocabulary continuous speech recognition system to overcome the constraint of enough instances of the words needed for training. Context-dependent acoustic modeling improves speech recognition by considering contextual information. The speech recognition systems are greatly affected by the contextual effects in phoneme based speech recognition systems. Pronunciation of a phone largely depends upon its context. Therefore to overcome contextual effects in speech recognition systems modeling of contextual effects is also essential. The performance of the speech recognition system is improved when right and left phones of the basic phones are also taken in to account for acoustic modeling with basic phone. In Triphone based acoustic modeling left and right phones are considered in addition to basic phones. Speech recognition engine searches for three phones instead of one phone, recognition accuracy is increased[1]. As context-independent models are based on phonemes. The phonemes are less in number, therefore the context independent models are trainable, but suffer from consistency due to contextual effects. Context-dependent models based on triphone are most important due to the modeling of context-dependency by including left and right contexts. These models are much more consistent than context-independent models.

Hindi is one of the official languages of India. It is widely spoken the language in India. Aspiration, germination, nasalization, and reflexiveness are different characteristics of Hindi Speech. For Hindi Speech recognition previous work related to isolated, connected word, continuous speech has experimented. Different acoustic models, different Gaussian mixture, context-dependent, context-independent and different states of Hidden Markov models (HMMs) have been included to improve Hindi Speech recognition[2–9].The phonetic model plays an important role in improving speech recognition results. In large vocabulary, continuous speech recognition sub word models are used. Then it is very important to decide which sub-word units to be used [10]. The performance of speech recognition system is affected by context. Although earlier research works in Hindi present the work using contextual information but lacks detailed analysis of context-dependent and independent speech recognition. The paper presents Hindi speech recognition using both monophone (context-independent) and triphone context-dependent phone models with comparative analysis. Dictionary using common phone set [11] has been used. Phonetic decision tree-based clustering has been used for tied list HMM generation. Research findings may be used to improve the performance of speech recognition systems by selecting suitable sub-word models for continuous speech recognition systems.

The rest of the paper has been organized as follows. Section II describes the literature review. Section III describes monophone and triphone based speech recognition. Section IV highlights the experimental setup. Section V and I describe results and discussion respectively. The final section is about the conclusion and future scope.

II. LITERATURE REVIEW

Continuous Hindi speech recognition system was developed using HMM-based context-dependent phone modeling. For feature extraction PLP coefficients, MFCC coefficients and HLDA coefficients were applied. Hindi speech corpus consisted of 600 unique words recorded in a quiet studio environment by 35 male and 20 female speakers. Experimental results show that the inclusion of the third-order derivative of feature extraction improved speech recognition by 3-4%.Punjabi ASR system was developed using context-independent and dependent, contextdependent tied, and context-dependent deleted system was developed for mobile systems. The system was evaluated at 4,16,32 and 64 GMMs.The system showed the highest recognition score for context-dependent untied models. The maximum accuracy obtained was 81.2% using a context-dependent untied model with 64 GMMs[12]. Different acoustic models were developed and evaluated for continuously spoken Malayalam sentences. The effect of contextdependent, context-independent and context-dependent tied models was experimented using HMM modeling. They used a database has twenty-one speakers including ten males and eleven females. Results show that contextdependent tied model outperforms context-independent models by 21% but it was not tested for syllables[13].The Malay language speech recognition system was developed for the Malay pre-school children using an HMM-based mono phone recognizer. The speech corpus was collected from 20 preschool children. The system was trained for sixteen Malay words. For feature extraction, MFCCs were used. The experimental findings show that highest recognition score of 94.86 for trained model and 76.04% using test model[14].Context-dependent and contextindependent speech recognition systems were developed for speaker-independent mode on the DARPA speech dataset. The LPC coefficients were applied. The phonetic and triphone based system was developed using HMM with no grammar, word pair, and bigram language models. Two types of triphone models such as generalized and function words experimented. The system achieved maximum speech recognition accuracy of 90.6% with a phonetic model and 95.2 % with triphone models using bigram language models[2].State tying was applied to reduce the parameter for HMMs using context-dependent phoneme based speech recognition. The method based on data-driven clustering triphone states with bottom-up algorithms and top-down methods with growing decision trees were applied. The tests were performed on test corpora of the 5000-word vocabulary of WSJ November 92 task and 3000 VERBMOBIL'95 task. The word error rate was reduced by 14% for the WSJ task and by 5% for VERBMOBIL'95 task[15]. Hindi vowel recognition was performed using HMM-based speech recognition. Phoneme based connected word speech recognition was developed using MFCC coefficients for speaker-dependent mode. Hindi Speech Corpus, made of 600utterances spoken by 5 speakers, was used in this experiment. The average vowel recognition score was achieved 83.19% [16].

III. MONOPHONE AND TRIPHONE BASED ACOUSTIC MODELING

Monophone based acoustic models are based on phonemes. Generally, every language has limited numbers of phonemes, so systems can be trained easily, but phonemes are more sensitive to context. To develop the speech recognition system phoneme based dictionaries are created. HMMs topology is defined with different states and left to right HMMs. The HCompv tool scans the set of data files that contain feature vectors calculated from speech samples by the use ofHcopy tool. The process of scanning results in computing global means and variances and sets all the HMM Gaussian to global means and variances. All phoneme based HMMs are re-estimated to get the final trained model[14]. Figure 1 shows triphone based speech recognition. Triphones based model uses both right and left contexts so better for context modeling contextual sensitivity but triphones are not considered efficient due to their large numbers. Triphone based speech recognition is presented starts with the generation of Monophones and Monophones transcription. These monophones are generated with flat start HMMs and short pause silence fixing. Short pause (sp) and silence (sil) are defined as boundary symbols. All phones except boundary symbols are converted into Triphones. Word-internal triphones are generated. In the step of cloning each model of Triphones such as x-y+z are checked. Each such monophone y is searched, and its copy is made. All transition matrices in each Triphone set are tied together. The transition matrix is shared by more than one model. Triphone transcriptions are also generated from Monophone transcription. Once the context-dependent Triphone models have been cloned the new Triphones along with Triphone transcriptions are re-estimated using HERest module. Finally, state occupations statistics file state occupancy (stats) is generated. Stats file also has a log-likelihood for clusters of states and is required for the state clustering process [17].Till now Triphones HMM sharing same transition matrix has been generated. In the next step, tied list Triphones are generated using decision tree based clustering.

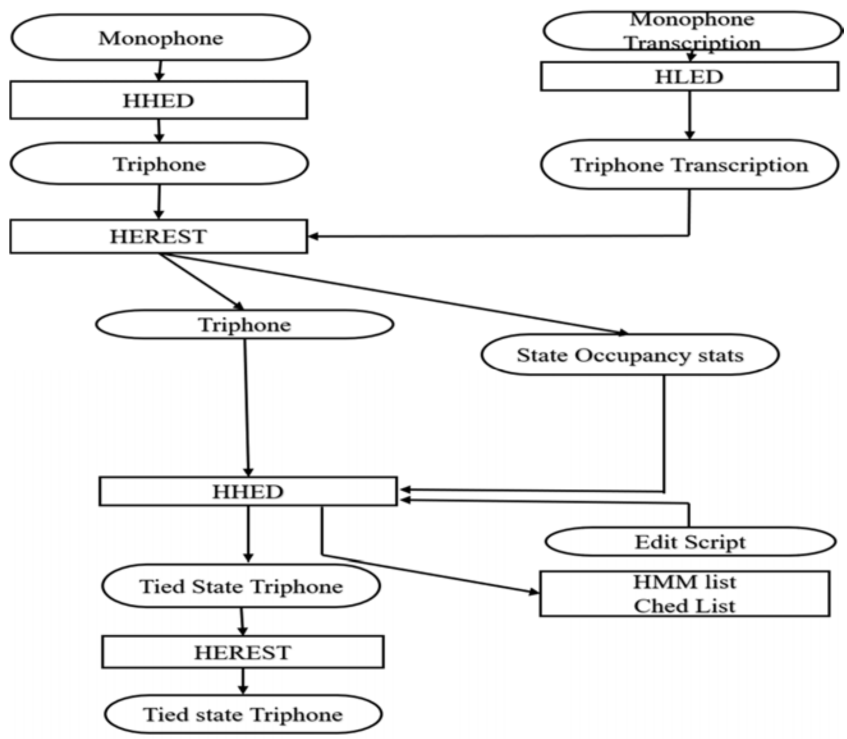


Fig. 1. Triphone based speech recognition system

IV. EXPERIMENTAL SETUP

Continuous Hindi Speech Recognition system was developed for speaker-dependent mode using Hidden Markov Model-based HTK tool kit version 3.4.1 in the Windows environment [17]. For implementation purposes, the following sub-steps were performed. For acoustic modeling, Monophones and Triphones based 5 state HMMs were used. For feature extraction perceptual linear predictive coefficients were used. In Monophones based acoustic modeling only single phone is considered. In Triphone based modeling left and right phones to basic phones are also considered. Hindi Speech corpus was provided by the KIIT Gurgaon, India. The database was divided into testing and training. Speech wave files were converted into Perceptual Linear Predictive Coefficients (PLPs) feature vectors. For generating lexicon common phone set [11] were used. Grammars, Gramdef1 and Gramdef2 were created.Word nets were created using Gramdef1 and Gramdef2.Monophones based transcriptions were generated. First, monophones HMMs were generated with silence fixing and re-estimation. Triphones HMMs and Triphones transcriptions were generated. Tied list Triphones were generated. The test data was evaluated against the final models. Results were generated for analysis. Speech corpus was spoken by 10 speakers and recorded with a sampling rate of 16000 and 16 bits. Pronunciation dictionary was developed for the Hindi language sentences. Some of the examples of pronunciation dictionaries are given in Table 1.

Table. 1. Pronunciation Dictionary Examples

Word	**Pronunciation**	**IPA symbols**
BAD	b aq d	B ɑ: d̪
BAHUT	b ac h uc t	B ə h ʊ t̪
BHAGWAN	bh ac g w aq n	b^h ə g ʋ ɑ: n

The grammar definition has been defined as per HTK format. The following are the example of grammar definition with restriction and without restriction. The example of grammar definition without end word restriction.

$WORD=BHAGWAN|BHI|DAANT|DADI|DARD|DAURA|DEVI||THA|THE;
(SENT-START(<$WORD>) SENT-END)

The following grammar shows the example of grammar definition with word end restriction.

The example of grammar definition with word restriction.

$WORD= BHAGWAN|BHI|DAANT|DADI|DARD|DAURA|DEVI|;
(SENT-START(<$WORD>)THA|THE SENT-END).

It consists of a set of variable definitions followed by regular expression describing the words to recognize. The Perceptual Linear Predictive (PLPs) coefficients were extracted for feature vectors. Table 2 shows coding parametres for extracted features.

Table. 2. Example of coding parameters for feature extraction PLP

S.No.	**Parameter**	**Value of Parameter**
1.	TARGETKIND	PLP
2.	WINDOWSIZE	250000.0
3.	TARGETRATE	100000.0
4.	NUMCEPS	12
5.	USEHAMMING	T
6.	PREEMCOEF	0.97
7.	NUMCHANS	26
8.	CEPLIFTER	22

Training Hmms and Recognition

The first set of monophone HMMs are generated. These HMMs were re-estimated until convergence to produce a trained model for monophones. First flat start monophones were generated. Initial sets of phone models are created. First, a prototype model was defined. All HMMs have five states. Each HMM state was represented by single Gaussian having mean, variance and mixture weight. The HTK tool kit module HCompv generates flat start HMMs for all phones. Global speech mean and variance is calculated from PLP features extracted from speech corpus. All initial phone models have the same state mean and variance equal to global speech mean and variance. After this triphones and triphones transcription was generated. Triphones and Triphone transcriptions are created using HLed tool. Word boundary symbols such as silence(sil) and short pause(sp) are defined. Cloning of models is done using HHED editor. The script mktri.hed contains commands for cloning and ties all transition matrices in Triphone sets. Table 3 shows triphone example of the word Himalaya (हिमालय).

Table. 3. Triphones Examples

SN.	Triphones with sil and sp boundary
1.	Sil
2.	h+ic (ह+इ)
3.	h-ic+m(ह-इ+म)
4.	ic-m+aq(इ-म+आ
5.	m-aq+l(म-आ+ल)
6.	aq-l+y(आ-ल+य)
7.	l-y(ल-य)
8.	Sp

In the final step, input speech is recognized. The speech .wav files to be recognized are first converted in to feature vectors (PLPs) using HCopy tool. These feature vectors are used to recognize the speech using the Hvite command. Test data was recognized for gramdef1 and gramdef2 for both monophones and Triphones based acoustic model.

V. RESULT

Continuous Hindi Speech recognition was tested for Monophones and Triphones using grammar definitions 1 and 2. Recognition results scores were obtained with the HResult tool of HTK[17, 18]. The following metrics are used.

Correct (%) = (N −D−S)/ N ×100

Accuracy(%) =(N −D−S−I)/ N ×100

Here D is deletion error, I is insertion error and S is substitution error, N is a total number of labels in reference transcription[19].The system was evaluated for speaker-dependent mode using 5 state HMMs. Results were generated for Monophones and Triphones. The effects of different grammar definitions have experimented. Table 4 presents recognition results for Monophone based system and triphone based system. For common phone set and Gamdef1 Monophone based maximum % word correct is 65.51.For Triphones generated after decision-based clustering with common phone set and Gramdef1 obtained maximum % word correct is 69.11.

Table.4. Speech Recognition Score for Monophones

S.No.	Coefficients	Phone Set	Grammar	Model Correctness(%)	
				Phoneme	Triphone
1.	PLP	Common Phone Set	Gramdef1	65.51	69.11
2.	PLP	Common Phone Set	Gramdef2	62.81	64.74

Table 5 shows insertion, deletion,and substitution % for the triphone and monophones based speech recognition systems.

Table. 5. Error Estimation during Speech Recognition analysis for triphones and monophones

Error Parametres	Triphone(%)	Monophone(%)
Insertion error	10.94	34.88
Deletion error	4.25	2.57
Substitution error	26.64	31.92

VI. DISCUSSION

The figure 2 shows the speech recognition scores obtained by both monophone based and triphone based systems for two grammar definitions.It was observed that Triphone based HMMs recognition score at word correct is better in comparison to Monophone based HMMs. Research findings also show that grammar def1 exhibits better recognition.

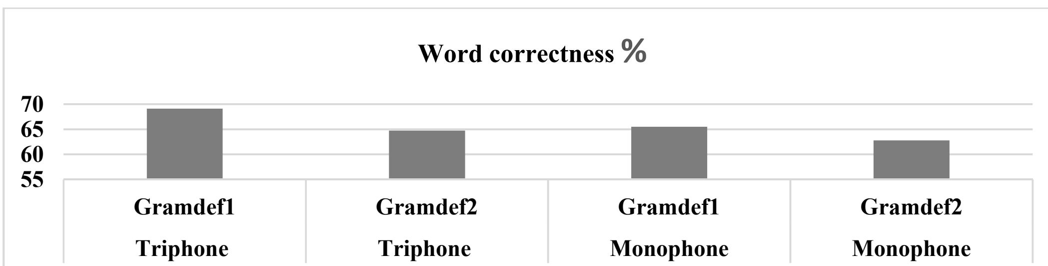


Fig. 2. Monophones and Triphones based continuous Hindi Speech Recognition scores

Figure 3 shows the comparative analysis of mono phone and triphone based Hindi speech recognition system. Research findings reveal that insertion and deletion errors are less in triphone based system in comparison to monophone based systems, but deletion errors are less in the case of monophone based system. The reason for less insertion and deletion error in case of triphone is due to consideration of both left and right contexts. It is not easy to insert and substitute for both right and left context of the phoneme The phoneme based models suffer from contextual effects so insertion and substitution errors are observed more than context depend model. Another observation is that insertion errors have more differences in comparison to substitution errors.

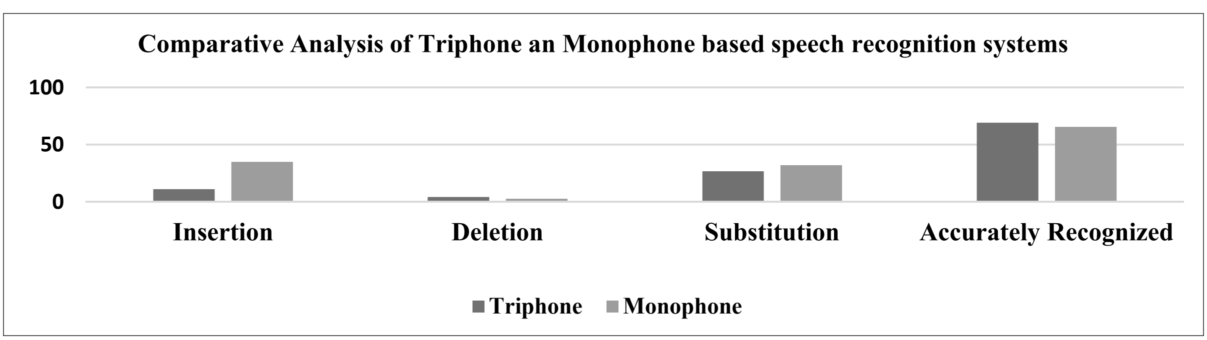


Fig. 3. Comparative analysis of Monophones and Triphones based continuous Hindi Speech Recognition systems

VII. CONCLUSION

Continuous Hindi Speech recognition system was developed for speaker-dependent mode for Monophones and Triphones. For overcoming the problems of large triphones and unseen Triphones, decision tree-based clustering was used. Two grammar definitions were used to study the effect of grammar definitions in speech recognition. Triphones based acoustic modeling provides better results using grammar definition1 and common phone set with perceptual linear predictive coefficients. Research findings could be used to improve speech recognition systems. Further, the work can be extended by experimenting with different language modeling techniques.

ACKNOWLEDGMENT

The authors would like to acknowledge the Ministry of Electronics & Information Technology (MeitY), Government of India, for providing financial assistance for this research work through "Visvesvaraya Ph.D. Scheme for Electronics & IT". We would also like to thank Director KIIT, Gurgaon for providing speech corpus for the experiment.

REFERENCES

[1] Bahl LR, de Soutza P V., Gopalakrishnan PS, et al (1991) Context dependent modeling of phones in continuous speech using decision trees. 264–269. https://doi.org/10.3115/112405.112453

[2] Lee KF (1990) Context-Dependent Phonetic Hidden Markov Models for Speaker-Independent Continuous Speech Recognition. IEEE Trans Acoust 38:599–609. https://doi.org/10.1109/29.52701

[3] Dev A, Agrawal SS, Choudhury DR (2003) Categorization of Hindi phonemes by neural networks. AI Soc 17:375–382. https://doi.org/10.1007/s00146-003-0263-0

[4] Bansal P, Dev A, Jain SB (2008) Optimum HMM combined with vector quantization for hindi speech recognition. IETE J Res 54:239–243. https://doi.org/10.4103/0377-2063.44216

[5] Pruthi T, Saksena S, Das PK (2000) Swaranjali: Isolated word recognition for Hindi language using VQ and HMM. Int Conf Multimed Process Syst 13–15

[6] Aggarwal RK, Dave M (2012) Integration of multiple acoustic and language models for improved Hindi speech recognition system. Int J Speech Technol 15:165–180. https://doi.org/10.1007/s10772-012-9131-y

[7] Kumar M, Rajput N, Verma A (2004) A large-vocabulary continuous speech recognition system for Hindi. IBM J Res Dev 48:703–715. https://doi.org/10.1147/rd.485.0703

[8] Kumar K, Aggarwal RK, Jain A (2012) A Hindi speech recognition system for connected words using HTK. Int J Comput Syst Eng 1:25. https://doi.org/10.1504/ijcsyse.2012.044740

[9] Sinha S, Agrawal SS, Jain A (2013) Continuous density Hidden Markov Model for context dependent Hindi speech recognition. Proc 2013 Int Conf Adv Comput Commun Informatics, ICACCI 2013 1953–1958. https://doi.org/10.1109/ICACCI.2013.6637481

[10] Seide F, Wang NJC (1998) Phonetic modeling in the Philips Chinese continuous-speech recognition system. Iscslp 1998 54–59

[11] Sadhukhan T, Bansal S, Kumar A (2017) Automatic Identification of Spoken Language. IOSR J Comput Eng 19:84–89. https://doi.org/10.9790/0661-1902058489

[12] Mittal P, Singh N (2019) Development and analysis of Punjabi ASR system for mobile phones under different acoustic models. Int J Speech Technol 22:219–230. https://doi.org/10.1007/s10772-019-09593-x

[13] Kurian C, Balakrishnan K (2012) Development & evaluation of different acoustic models for Malayalam continuous speech recognition. Procedia Eng 30:1081–1088. https://doi.org/10.1016/j.proeng.2012.01.966

[14] Maseri M, Mamat M (2019) Malay language speech recognition for preschool children using hidden markov model (HMM) system Training. Lect Notes Electr Eng 481:205–214. https://doi.org/10.1007/978-981-13-2622-6_21

[15] Beulen K, Bransch E, Ney H (1997) State tying for context dependent phoneme models. Fifth Eur Conf 3–6

[16] Bhatt S, Dev A, Jain A (2018) Hindi Speech Vowel Recognition U sing Hidden Markov Model University School of Information and Communication Technology , Guru Gobind Singh Indraprastha University Delhi Indira Gandhi Delhi Technical University for Women , Delhi University School of Infor. 1:196–199

[17] Young S, Gales M, Liu XA, et al (2015) The HTK Book (version 3.5a)

[18] Bhatt S (2017) Hindi Speech Recognition : Issues and Challenges. pp 2719–2723

[19] Bhatt S (2018) Hidden Markov Model Based Speech Recognition-A Review. pp 3367–3372

Development of Syllable Dominated Hindi Speech Corpora

Ruchika Kumari[1,4], Amita Dev[2], Ashwani Kumar[3]

[1]*Department of ECE, Maharaja Surajmal Institute of Technology, New Delhi, India*
[2]Vice Chancellor, Indira Gandhi Delhi Technical University for Women, Delhi, India
[3]Department of ECE, Indira Gandhi Delhi Technical University for Women, Delhi, India
[4]*Research Scholar, Indira Gandhi Delhi Technical University for Women, Delhi, India*
[1]ruchika.kumari33@gmail.com, [2]amita_dev@ hotmail.com, [3]drshwnikumar@gmail.com

Abstract: The paper describes the building of the speech database considering syllable as a speech unit. The precaution has been taken while designing the database to contain a large number of syllables for the news domain-specific application. Due consideration has also been taken to obtain high-quality natural speech with a careful selection of suitable speech units and a well-organized corpus. The database has been prepared by including frequently used words from the Hindi newspapers like "*Hindustan*" and "*Dainik Jagran*." The prepared database consists of 225 sentences comprising of 1964 words and 3358 syllables. After analysis of obtained syllable unit, it is found that CV is 58.88%, CVC is 20.43% and Cv is 10.18%. The rest structures (VC, V, CVCC, CCVC, CCV, CvC) covers less than 4% each. The database may be extended by adding new sentences. This type of database is applicable for building TTS in the Hindi language.

Keywords: Database, Short Term Energy, News domain, Syllable

I. INTRODUCTION

Speech, a way of communication, can be used as an interface for the computer system. The various ways and means have been tried by researchers to make the computer as a tool for interpretative and to understand human speech. The researchers have developed machines using advanced technology to build a corpus of high quality and volume of speech. The automatic conversion of a written text into speech is achieved with the help of the state-of-art technology of speech synthesis i.e., building a natural synthesis voice raised from the knowledge base process by a single speaker. In the digital world, most of the information is accessible to a few who can read or understand a particular language like Hindi written in *Devanagri* script.

The speech corpus is an essential part of all spoken technology systems. Due care has been taken to design and developed speech interface necessary for speech synthesis because of isolated word recognition, connected word, and continued speech to achieve maximum accuracy. The speech corpus is known as an extensive collection of written text structure or audio recording of spoken language i.e., used in linguistic research [1]. C-DAC Noida [2] developed the speech data for the application of automatic speech recognition (ACR) for the travel domain. 8576 sentences are recorded with a duration of 26 hrs, comprising of 74804 words by 30 female speakers. [3] Developed a general purpose of the database for Hindi, Telugu, Tamil, and Kannada languages. For the Hindi speech, the utterances were recorded by 19 speakers with a duration of 3.5 hrs. This database is used for ASR and language identification. [4] created a multi-variability speaker recognition database. This type of database is used in field research like the effect of noise and reverberation, speaker diarization, etc. The recording is done by the headset, microphone, and offline mobile mode. [5] KIIT collaboration by NOKIA Research Centre, China, has developed the database in the field of mobile communication for Hindi and English language consists of 42801 unique words of Hindi and 33963 words of English. These text corpora were recorded by every 100 speakers on three simultaneous channels – mobile phone, headset and desktop. [6] build medium size database for passenger rail information systems for the Hindi language in phoneme level by using HMM. The database consists of 630 utterances with 12674 words to facilitate the researcher in the field of Text to Speech (TTS) and ASR.

In this work, the syllable-based database is created for the Hindi language to boost-up the speech domain, which is still lagging in Hindi. The quality and volume of speech data in the corpus directly related to the effect of the speech-based system. The database is developed in Hindi for speech synthesis, which can be applied to a more general and commercial application like talking aids for the vocally impaired, reading aids for the blind, and remote access to online information, etc. The paper is organized as follows - the characteristics of the Hindi language are described in section 1. Section II describes the methodology to create a database. Section III gives a statistical analysis of the database. In section IV, the results obtained are discussed, and section V explains the conclusions.

II. THE METHODOLOGY ADOPTED TO CREATE A DATABASE

The following steps are carried out to design speech corpus:

- Selection of the text sentences from the news domains
- Recording of the selected text sentences
- Syllabification of the speech signal

1. Selection of the Text Sentences

The selection of the sentences has been manually selected from the daily Hindi newspapers is based on the fact that it should be relevant for the news domain for building the speech synthesis system [7]. Sufficient care has been taken to include all types of the required information so that recording can be minimized. It should also have enough occurrence of each type of sound to capture the acoustic of the Hindi speech.

2. Recording of the Selected Text Sentences

The steps followed for recording the speech wave files were as follows [8]:

- The recording was done in a noise-free and echoed canceled studio recording environment.
- A professional male speaker with a good quality voice has been selected for recording.
- The sampling frequency was set to 16 kHz store in 16-bit PCM with Mono mode type.
- The speaker is required to read each text sentence, and the recorded sample was saved as wav files.

3. Most Frequent Words

The most frequently used words are covered in a significant portion of sentences recorded. Such databases obtain better quality synthesized speech. 100 more frequently occurring words have been identified and cover 25% of speech. Table 1 shows the more frequently repeated words of the database.

Table 1. Most frequently repeated words with few examples

S. No	Words	Number of repetitions
1	हैं	70
2	के	32
3	में	28
4	को	20
5	से	16
6	की	16
7	नहीं	12
8	थी	12
9	पुलिस	10

III. SIGNIFICANCE OF HINDI SPEECH CORPUS

The text material contains 225 sentences written is Hindi chosen from a daily Hindi newspaper [7]. These texts content is recorded by a male speaker, and it contains all the phonemes related to the news domain. The sentences are syntactically valid and meaningful. The recording of sentences is done at the sampling rate of 16KHZ and digitized with 16 bits per sample. The database is carried in syllabic labeling nature by using the Devnagari script. The syllable label boundaries were obtained automatically by using the energy convex hull technique [9]. The Hindi speech sentence corpora developed for speech synthesis information system is essentially a news domains system.

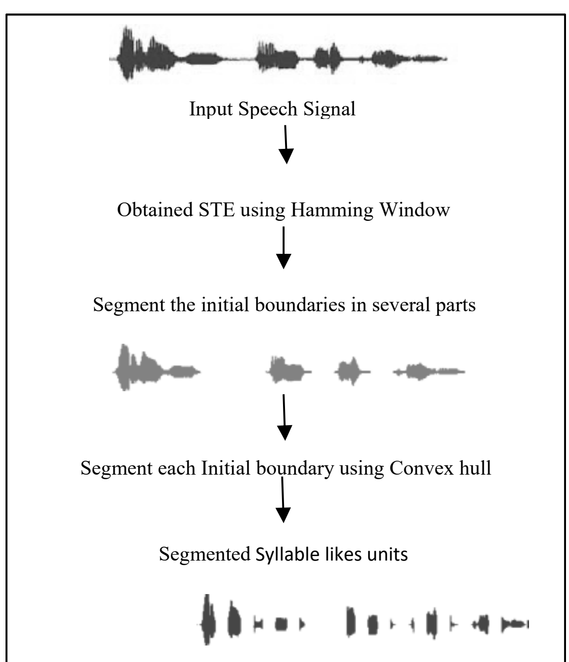
1. Choice of Speech Database Unit

The constituent of the database is multiple forms of the unit phoneme, syllable, and words. Hindi phonemes can be divided into vowels and consonants. [10] The Hindi alphabet consists of 10 pure vowels (/अ/आ/इ/ई/उ/ऊ /ए/ऐ/ओ/औ/) including two diphthongs, namely; /ए/ओ//. In Hindi, the type of syllable is classified as in [11] syllable segments and classified into various syllable structure types such as CV, CVC, Cv, V, CCV, VC, and CCVC. Here C, V, v, stand for a consonant, vowel, and nasalized vowel, respectively. In our database, there is a maximum occurrence of CV and CVC type of syllable. In the Hindi language, the script is phonetic. The phonetic unit of the database is classified into phrases, words, syllables, and phones. In this paper, the chosen unit is syllable. The wave files were segmented into the syllable-like unit. The advantage of the energy convex hull technique avoids manual labeling, which is laborious, time-consuming, and error-prone.

2. Syllable Set and Syllable Boundary Identification

The syllables have to be identified from the speech database. The Syllable boundary identification is made using energy convex hull algorithm segmentation technique [9]. The method for segmenting speech into syllable like units uses the short-term energy function of the speech signal with a hamming window and detects first boundaries Q(n). The energy signal is applied to the sliding Hamming window is applied to Q (n) to detect the segmented output p(n).

The Equation and flowchart are as mentioned below:



The flowchart shows segmentation system for continuous speech

$$Q(n) = \sum_{m=-\infty}^{\infty} [x(m)]^2 w(n-m)$$

$$D(n) = 10 * \log[Q(n) + 1]$$

$$p(n) = \sum_{m=-\infty}^{\infty} D(m) w(n-m)$$

The identification of syllables of words used in the speech database is shown in Table 2.

Table 2. Few examples of syllable occurring in the database

S. No.	Syllable Types	Words
1.	CV	की,को,तो,दु,से,थी
2.	CVC	दस, दिन,वह,रहा, गया
3.	VC	इस, एक, इन
4.	Cv	माँ,हैं,मैं,हूँ
5.	CvC	जाँच, काँच, तंग

3. Syllabification of Speech Signal

The recorded data needs to be appropriately labeled into syllable- like units. In this case, the recorded wave files were automatically segmented as syllable like units by applying the text segmentation technique [12].

The below input sentence gives the syllables obtained, as shown in Table 3.

दवाइयों और जाँच में रुपया खर्च हो गया

Table 3. Text syllabification result

Text input
दवाइयों और जाँच में रुपया खर्च हो गया
Syllabification of Text result output
द वा इ यों और् जाँच् में रुप् या खर्च् हो ग या

4. Syllable and Their Distribution

Syllable, as the basic unit, have been identified for the words in recorded speech corpora. It is shown in Table 4.

Table 4. Distribution of various syllable in Hindi

Structure	Relative Frequency (%age)
CV	58.88
CVC	20.43
VC	2.78
V	3.60
CVCC	1.18
CCVC	0.89

Structure	Relative Frequency (%age)
CCV	0.48
Nasalized Syllables	**Relative Frequency (%age)**
Cv	10.18
CvC	1.57

5. Information About Speech Database

The statistical analysis of the text contains a speech database that has been carried out. However, new sentences can be added to update. Table 5 shows the statistical analysis.

Table 5. Statistical analysis of recorded speech

S. No	Process	Number
1	Utterance reordered	225
2	Number of words obtained	1964
3	Syllable unit identified	3358
4	Number of distinct words recorded	235

IV. RESULTS

The database is consisting of 225 Hindi sentences chosen from the Hindi newspapers. Each sentence has been labeled at the word, syllable, and phoneme level. Figure 1 shows the continuous input wav file.

(दवाइयों और जाँच में रुपया खर्च हो गया). Figure 2 depicts the labeled for the sentence explained below at syllable level and obtained 13 syllables like units.

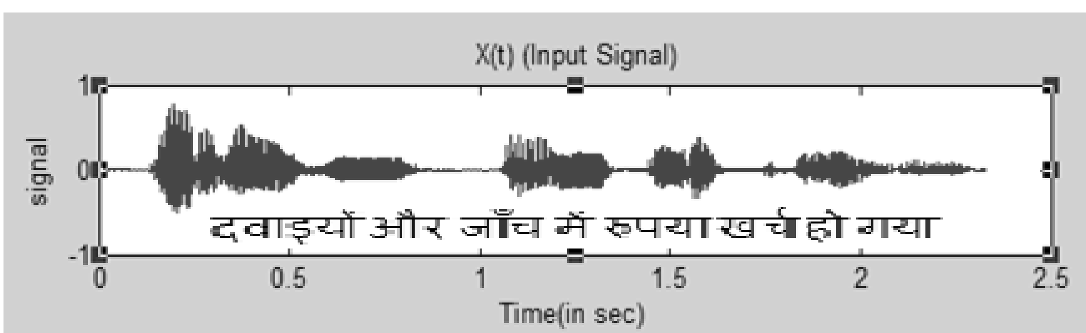


Fig. 1. The waveform of input speech x (t)

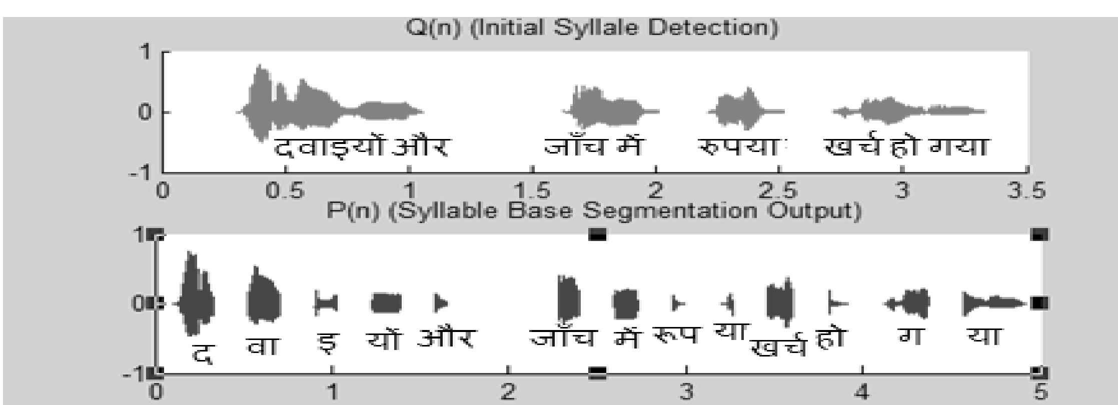


Fig. 2. Segmented output syllable units of STE and Convex Hull

V. CONCLUSION

The medium size segmented and labeled Hindi syllable corpora have been developed for a domain-specific application of the News domain. The text is recorded by one male speaker organized in 225 sentences containing 1964 words and 3358 syllables. The experiment is done by using the energy convex hull segmentation technique. Further, the size of corpora shall be increased by adding more sentences to the news domain. This database can be used in applications such as TTS and Speech recognition.

REFERENCES

[1] Dash, N.S. & Choudhary, B.B. (2011). *Why Do We Need to Develop Corpora for Indian Languages?* International Conferences on SCALLA (Vol. 11), Bangalore.

[2] Arora, Sunita., Saxena, Babita., Arora, Karunesh., & Agarwal, S. S. (2010). *Hindi ASR for Travel Domain.* OCOCOSDA 2010, Kathmandu, Nepal

[3] Shrishrimal, Pukhraj P., Deshmukh, R. R. & Waghmare, V. B. (2012). *Indian Language Speech Database: A Review*, International Journal of Computer Application, Vol. 47 No.5, pp. 17-21

[4] Haris B.C. et al. (2011). *Multi-variability speech database for robust speaker recognition*. National Conference on, Communications (NCC), 2011 vol., no., pp.1-5, 28-30

[5] Agrawal, Shyam., Sinha, Shweta., Singh, Pooja., & Olsen, Jesper. (2012) *Development of Text and Speech Database for Hindi and Indian English specific to Mobile Communication Environment.* International Conference on the Language Resources and Evaluation Conference, LREC, Istanbul, Turkey

[6] Balyan, Archana. (2017). *Development of phoneme dominated database for limited domain T-T-S in Hindi.* International Journal of Artificial Intelligence and Applications (IJAIA), vol.8

[7] http://epaper.livehindustan.com/login.aspx

[8] Balyan, Archana., Agrawal, S. S. & Dev, Amita, (2008). *Building Syllable dominated Speech corpora for Metro Rail Information system.* International Conference of O-COCOSDA-2008, p-p 135-140, Kyoto, Japan

[9] Xufang Zhao, & Douglas O'Shaughnessy. (2008). *A New hybrid approach for automatic speech signal segmentation using silence signal detection, energy convex hull, and spectral variation.* IEEE International Conference

[10] Arora, Karunesh., Arora, Sunita., Verma, Kapil., & Agrawal S. S. (2004). *Automatic Extraction of Phonetically Rich Sentences from Large Text Corpus of Indian Languages*. INTERSPEECH2004 -ICSLP Jeju, Korea

[11] Chandra, Mahesh. (2016). *Hindi Vowel Classification using QCNPNCC Features.* Indian Journal of Science and Technology 9.38

[12] Balyan, Archana., Dev, Amita., Kumari, Ruchika., & Agrawal, S. S. (2015). *Labeling of Hindi Speech.* IETE Journal of Research, vol 62, issue 2, pp. 146-153

Memory based Speech Duration Model using Exemplar Theoretic Approach

Thennal D.K.[1], Deepa P. Gopinath[2], Vrinda V. Nair[3]

[1]Government V&HSS Karakulam Trivandrum, India
[2]Directorate of Technical Education Trivandrum, India
[3]A P J Abdul Kalam Technological University Trivandrum, India
[3]vrinda66nair@gmail.com

Abstract: Incorporation of prosody is a key factor in the generation of natural sounding speech. The prosodic information is significant in speech and speaker recognition applications as well. The complex durational patterns of natural speech can be modelled analogous to the way by which human brain generates durational variation in speech segments. The proposed memory based duration model developed using exemplar theoretic approach stores the duration patterns of speech segments of recorded speech and predicts the duration corresponding to a given text sequence by retrieving the duration pattern of the best matching exemplar. Exemplar clouds are formed by performing k-means clustering on duration vectors of different lengths. The duration model is developed for Malayalam speech in reading style as a case study. The proposed model is evaluated by comparing the predicted durations with the actual durations of the recorded speech, using objective measures like RMSE (Root Mean Square Error) and correlation.

Keywords: Duration model, Exemplar theory, K-means clustering, Memory based learning

I. INTRODUCTION

The prosody or rhythm of natural speech is different for each language and for each speaking style. We generate rhythm for each of these, based on a model stored in our brain, acquired through the exposure in that speaking style. The variation of speech duration is one of the factors that result in the manifestation of rhythm in speech. In Text to Speech Synthesis systems (TTS), the naturalness of speech synthesized depends on the ability of the system to predict the duration and intonation for the given input text. The prosodic cues including the durational variations also find use in speaker and speech recognition. The duration models predict the duration of individual speech segments, for a given text sequence.

The duration modelling is considered as a complex nonlinear task [17]. Statistical and rule based models constituted the earlier attempts in duration modelling. Rule based models were developed for various languages, the first one being that by Klatt [8, 9, 23, 12]. Recorded speech data is used to train models using statistical approach like in Sum of Products (SOP) models [20], Neural network models [17], Classification and Regression Trees (CART) [19], Support Vector Machines (SVM) [18] etc. Statistical parametric speech synthesis system uses HMM based duration modelling [29]. Duration modelling has also been done using deep learning algorithms [26]. Park et al. reports the application of phonemic level duration control in speech synthesis [14].

The duration model proposed in this paper is based on the way by which a speaker of a particular language learns the prosodic features of a particular style. The model is developed based on exemplar theory, a psychological model for perception and classification [15]. According to exemplar theory, whenever we perceive or produce a linguistic unit, the phonetic details of that are stored in our memory in the form of exemplars [6]. When we come across a new word, it is compared with the stored exemplars and classified based on similarity. The duration model proposed stores categories of exemplars from recorded speech and retrieves the duration values of the best matching exemplar for predicting the durations of speech units in the given text.

Exemplar models are extensively used in phonetics and phonology. It is used to explain the internal prosodic models [11], speech perception [15] and speech production [21]. Exemplar theory concepts have also been applied in voice conversion [27, 1, 5, 10, 13, 22] and speech waveform generation methods [7, 25]. This methodology is similar to the memory based learning where, when a new instance of a pattern is processed, the k-nearest neighbours of the pattern are retrieved from memory, based on similarity with the new instance [28].

In this work, the duration model was developed in an exemplar theoretic approach, by storing duration patterns or exemplars obtained from natural speech. The duration vectors in the training database are clustered to form exemplar clouds. For any given text, the best matching exemplar will be searched and the corresponding duration pattern will be given as the output. The search is done from sentence level to word level. Duration modelling using exemplar theoretic approach is not seen reported in literature.

As a case study the experiments are done using speech database in Malayalam, an Indian language, spoken by around 35 million people as per www.censusindia.gov.in. Duration modelling has been previously done for Malayalam using probability distributions [2], CART for TTS in Festival framework [16] and hybrid model using CART and HMM [3]. Pause duration modeling for Malayalam TTS has been done by James et al. [4].

II. METHODOLOGY

1. Database

The database used consisted of recorded Malayalam speech in reading style, comprising of 896 sentences with 5338 words. The sentences were manually segmented and annotated into words, phrases, syllables and phonemes using speech analysis software PRAAT. From the annotated database, the duration of phonemes were measured and tabulated.

2. Development and Evaluation of Memory based Duration Model

Exemplar clouds analogous to that in our brain are formed by performing clustering of duration patterns using k-means clustering. The model was trained using 80% of the database and tested with 20% of the database. The performance of the duration model was evaluated using the RMSE (Root Mean Square Error) value and correlation value. These are the objective evaluation measures used in the evaluation of duration models. The duration model was developed in Python language.

III. RESULTS AND DISCUSSION

1. Implementation of Memory based Duration Model

The memory based duration model stores the phonetic transcription of the text data and the corresponding duration vectors. Each word has a corresponding duration vector with length equal to the number of phonemes in the word. Each element of the vector is equal to the duration (in millisecond) of the corresponding phoneme. The duration vectors of the phrases and sentences are concatenation of duration vectors of the constituent words and the duration of pause in between words. The first step in the training is the storage of duration vectors and the corresponding phonetic transcriptions. The duration vectors in word level are then grouped according to the length of the vector. The database consisted of duration vectors with length ranging from 2 to 22. K-means clustering was performed on each of these groups. To find the optimum number of clusters, Silhouette method was used. For each length the number of clusters was fixed as the one that gave the maximum average silhouette co-efficient value. Table 1 gives the maximum average Silhouette value and the corresponding clusters for each length. Figure 1.gives the plot of average Silhouette co-efficient for different number of clusters for duration vector of length 5.

When a text sequence is given for prediction of duration, the model first checks at sentence level to find a match. If a match is found, the corresponding duration vector will be given as output. Otherwise phrases will be checked for match. If match is not found in the sentence level or phrase level, then the model moves on to the word level. If exact match is obtained at word level, that will be taken. If exact match is not obtained in any of these levels, then the model identifies the best matching cluster to fetch the duration vector.

Table 1. Maximum average Silhouette score and number of clusters for various lengths

Length of duration vector	Average Silhouette Co-efficient	Number of Clusters
3	0.864	23
4	0.728	109
5	0.730	178
6	0.642	170
7	0.650	304
8	0.553	300
9	0.544	258
10	0.551	264
11	0.479	190
12	0.427	221
13	0.435	133
14	0.376	148
15	0.344	94
16	0.285	64
17	0.342	53
18	0.307	42
19	0.275	25
20	0.241	26
21	0.157	9
22	0.400	9

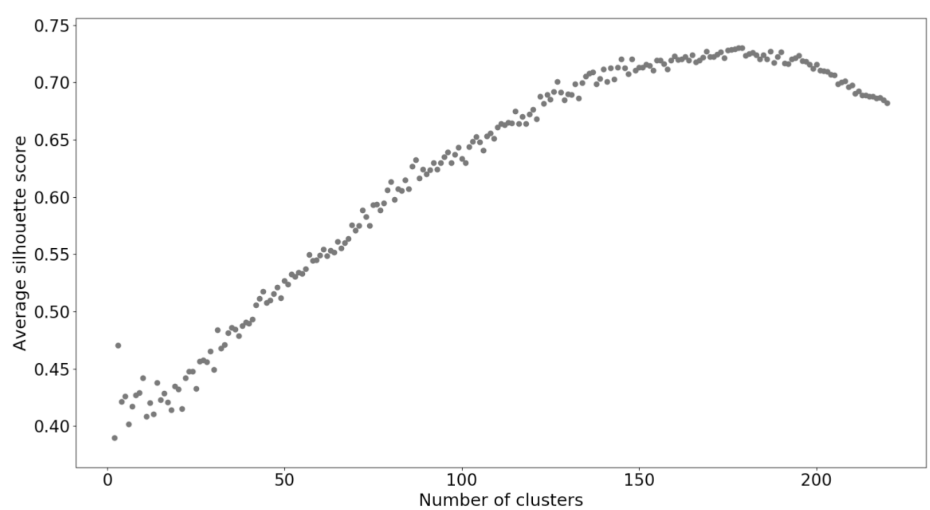


Fig. 1. Average Silhouette score for different number of clusters for duration vector of length 5.

2. Evaluation of the model

The model was tested with a database of 179 sentences. The root mean square error obtained was 4.86ms, correlation was 0.96, which is comparable with the best results obtained for Malayalam language in the previous works reported in literature (RMSE of 8.32ms [3] and correlation of 0.9918 [16]).

IV. CONCLUSION

The memory based duration model using exemplar theoretic approach, stores exemplars of speech constructs of natural speech and retrieves the duration pattern of the most similar exemplar for any given text, for the prediction of duration. The model built similar to the internal prosodic models of human brain, can be used in TTS for generating natural sounding speech as well as in speech and speaker recognition for improving the performance of recognition. The duration model developed using recorded speech in Malayalam language in news reading style, yielded RMSE of 4.86ms and correlation of 0.96 on evaluation.

REFERENCES

[1] Aihara, Ryo & Takashima, Ryoichi &Takiguchi, Tetsuya &Ariki, Yasuo. (2014). A preliminary demonstration of exemplar-based voice conversion for articulation disorders using an individuality-preserving dictionary. EURASIP Journal on Audio, Speech, and Music Processing.doi: 5. 10.1186/1687-4722-2014-5

[2] Gopinath, Deepa P., Veena, S. G. & Nair, Achuthsankar S. (2008): Modeling of vowel duration in malayalam speech using probability distribution, In SP-2008, 99-102.

[3] Gopinath, Deepa P., Vinod, C. S. S., Veena, S. G. & Nair, Achuthsankar S. (2008): A hybrid duration model using CART and HMM, TENCON 2008 - 2008 IEEE Region 10 Conference, Hyderabad, 2008, pp.1-4.doi: 10.1109/TENCON.2008.4766762

[4] James, J. &Gopinath, D.P. (2015). Pause duration model for Malayalam TTS. 2206-2210. 10.1109/ICACCI.2015.7275944.

[5] Jin, Z., Finkelstein, A., Di Verdi, S., Lu, J., Mysore, G. J.,(2016). Cute: a concatenative method for voice conversion using exemplar-based unit selection. IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), 2016

[6] Joan, L. B., (2001). Frequency effects on French liaison. In J. Bybee, P. Hopper, J. Benjamins (Eds), Frequency effects and Emergent Grammar: (pp. 337-359)

[7] Kirchner, R., Moore, R.K., & Chen, T. (2010).Computing phonological generalization over real speech exemplars. J. Phonetics, 38, 540-547.

[8] Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. Journal of the Acoustical Society of America, 59, 1208–1221. doi: 10.1121/1.380986

[9] Kumar, S., &Yegnanarayana, B. (1989).Significance of durational knowledge for speech synthesis system in an Indian language. Fourth IEEE Region 10 International Conference TENCON, 486-489.

[10] Ming, H., Huang, D., Xie, L., Zhang, S., Dong, M., Li, H. (2016).Exemplar-based Sparse Representation of Timbre and Prosody for Voice Conversion. IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), 2016

[11] Mobious, B. &Dogil, G., (2002). Phonemic and Postural Effects on the Production of Prosody, In: B. Bel&I.Marlien (Eds): Proceedings of the Speech Prosody 2002 Conference, Aix-en-Provence, 523-526.

[12] N P, Narendra&Rao, K. &Ghosh, Krishnendu&Vempada, Ramu&Maity, Sudhamay.(2011). Development of syllable-based text to speech synthesis system in Bengali.International Journal of Speech Technology. 14. 167-181. 10.1007/s10772-011-9094-4.

[13] Nguyen, G. N., Phung, T. N. (2017). Reducing over-smoothness in HMM-based speech synthesis using exemplar-based voice conversion. EURASIP Journal on Audio, Speech, and Music Processing, 2017.

[14] Park, Jungbae & Han, Kijong & Jeong, Yuneui & Lee, Sang Wan. (2019). Phonemic-level Duration Control Using Attention Alignment for Natural Speech Synthesis. 5896-5900. 10.1109/ICASSP.2019.8683827.

[15] Pierrehumbert, J. B. (2003). Phonetic Diversity, Statistical Learning, and Acquisition of Phonology. Language and Speech, 46(2–3), 115–154. https://doi.org/10.1177/00238309030460020501

[16] Rajan, B. K., Rijoy, V., Gopinath, Deepa. P., & George, N., "Duration modeling for text to speech synthesis system using festival speech engine developed for Malayalam language," 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], Nagercoil, 2015, pp. 1-5.

[17] Rao, K.S., & Yegnanarayana, B. (2004).Modeling syllable duration in Indian languages using neural networks.2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, 5, V-313.

[18] Rao, K.S., & Yegnanarayana, B. (2005).Modeling syllable duration in Indian languages using support vector machines. Proceedings of 2005 International Conference on Intelligent Sensing and Information Processing, 2005, 258-263.

[19] Riley M. (1992). Tree-based modeling for speech synthesis,In Talking machines. In Theories, models and designs, eds. Elsevier, (pp. 265-273).

[20] Santen, Jan. (1994).Assignment of segmental duration in text-to-speech synthesis. Computer Speech and Language, 8, 95-128. doi:10.1006/csla.1994.1005.

[21] Schweitzer, A., &Möbius, B. (2004). “Exemplar-based production of prosody: evidence from segment and syllable durations,” in Proceedings of Speech Prosody 2004, (Nara), 459–462

[22] Sisman, Berrakand Lee, Grandee & Li, Haizhou.(2018). Phonetically Aware Exemplar-Based Prosody Transformation. 10.21437/Odyssey.2018-38.

[23] Thomas, Samuel (2007). Natural sounding text-to-speech synthesis based on syllable like units, Unpublished M. Sc Thesis, Indian Institute of Technology, Madras, May 2007

[24] Walsh, M., Schutze,H., Mobius, B., &Schweitxer,A. (2007). An Exemplar -Theoretic Account of Syllable Frequency Effects, in Proc. Of ICPhS XVI, Saarbrucken, Germany, 6-10 August 2007

[25] Watts, Oliver &Valentini-Botinhao, Cassia &Espic, Felipe King, Simon.(2018). Exemplar-based Speech Waveform Generation. 2022-2026. 10.21437/Interspeech.2018-1857

[26] Watts, Oliver &Henter, Gustav & Merritt, Thomas & Wu, Zhizheng. (2016). From HMMS to DNNS: Where do the improvements come from?.5505-5509. 10.1109/ICASSP.2016.7472730.

[27] Wu, Z., Virtanen, T., Kinnunen, T., Siong, C.E., & Li, H. (2013).Exemplar-based unit selection for voice conversion utilizing temporal information. INTERSPEECH

[28] Zavrel, J., &Dalemans, W. (1997). “Memory-Based Learning: Using Similarity for Smoothing”, In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics (ACL ‘98/EACL ‘98). Association for Computational Linguistics, Stroudsburg, PA, USA, 436-443. DOI: https://doi.org/10.3115/976909.979673

[29] Zen, H., Tokuda K., Masuko T., Kobayashi T and Kitamura T., “Hidden Semi-Markov Model Based Speech Synthesis,“INTERSPEECH 2004 -- ICSLP 8th International Conference on Spoken Language Processing ICC Jeju, Jeju Island, Korea October 4-8, 2004

Cross-language Shared Phoneme Set for Lambadi Language Speech Recognition System

Banothu Rambabu, Sreedhar Bhukya and Suryakanth V. Gangashetty

Speech Processing Laboratory,
International Institute of Information Technology-Hyderabad, Telangana, India
rambabu.b@research.iiit.ac.in, sreedharbhukya@gmail.com, svg@iiit.ac.in

Abstract: In this paper, we present the design and development of SGMM based automatic speech recognition system for Lambadi language. Lambadi language is one of the scheduled tribal minority languages in India. Lambadi language is a low resource language for speech systems area. To build limited and large vocabulary speech recognition system, we required good speech corpus for acoustic modeling, and its corresponding text transcription for language modeling. Lambadi language does not have writing system. So, we have adopted Telugu writing system and recorded speech data. For representing sound units for computer recognition, we have used Indian Language Speech Sound Labels (ILSL). We have collected regular words and conversation speech as text and speech corpus. We built a GMM-HMM based speech recognition system using SPHINX-3 open source tool kit as a baseline system and then a SGMM-UBM based speech recognition system using Kaldi tool kit for improves the word recognition accuracy. We have evaluated the performance of isolated word recognition with context dependent and context independent models. The performance of the both the systems reported.

Keywords: *Lambadi Speech recognition, Hidden markov model, Universal background model*

I. INTRODUCTION

In India, or even across the world, there are many unwritten languages (Sunayana Sitaram et al., 2013). Such languages are surviving only through oral communication. These languages have large percentage of code mixed (Sitaram et al., 2016) words as well due to its coexistence with other native languages. Building a speech recognition (Rabiner, 1989) or synthesis (Sunayana Sitaram et al., 2013) system for such languages is a huge task. If speech technology could process and build systems for languages without standardized orthography, then the impact of such systems on human life is enormous and it would bridge the current research to the next higher level of research in speech technologies (Kohler, 2011). Current research community explores this area of unwritten languages through some literate speakers who could speak this language along with other popular languages like English, Hindi or Telugu (Baljekar et al., 2015) and (sreedhar, 2018). We collected the data with the help of a bilingual Lambadi Speaker, where he/she is given a written prompt in Telugu and he/she speaks the sentence in Lambadi with the same meaning as that in Telugu prompt. We aim to build a speech to text system that will help us write new prompts in Telugu and generate text output with the same meaning as that in Lambadi language. For this, we used acoustic analysis along with machine translation and speech recognition. We build an acoustic model and iteratively improved it by retraining with the available corpus.

This paper is organized as follows. Section 2 introduces the reader to Lambadi language followed by data collection procedure in section 3. Section 4 describes the experiments on speech recognition for Lambadi language. We concluded our work and presented some insights into possible future work on unwritten languages in section 5.

II. INTRODUCTION TO LAMBADI LANGUAGE

The Lambadi language belongs to the Indo-Aryan family of languages (Trail, 1970). We can find the Lambadi speaking people from different parts in India. They live in thandas. Thanda is a group of same community people live together in small huts nearby hills and forest areas. Major population of this language people are concentrated in Telangana, Andhra Pradesh, Maharastra, Karnataka and Rajasthan states in India. The Lambadi language although varying somewhat in vocabulary and phonemic inventory from region to region, has virtually the same syntactic structure throughout. This language has many loan words from neighboring languages and English language. Frequent code mixing and code switching(Uebler, 2011) occurs in their conversation specially if they interact with non-Lambadi people. Compared to other Indian languages, not much work has been done previously on Lambadi language, there are hardly any computational resources for Lambadi language. The Lambadi doesn't have the written form, people communicate through oral/spoken form. Lexical resources like dictionaries, speech corpora etc., are yet to be developed for Lambadi.

Lambadi language is spoken by 1 billion people all over the India. The Lambadi language used in our work is taken from the dialects from the states of Telangana and Andhra Pradesh. The language has many dialects and varies from region to region. In addition to unwritten nature of the language, this language is influenced by many regional languages. For example, in Telugu-speaking states of Telangana and Andhra Pradesh, the influence of Telugu on Lambadi language is intensely visible. Similarly, in Maharastra state, the Marati language influences the Lambadi language leading to Lamani dialect. In Karnataka, there is other form of Lambadi language, Labhani dialect, with influence of Kannada language. However, around 80\% of the syntactic and symantic structure of the language is same in all regions where Lambadi language is present. The Lambadi language vowels are more tense than English vowels (Trail, 1970) and there are many short and long vowels and nasalized vowels exist in Lambadi language.

The Lambadi language is Syntagmatic morphology with inflectional suffix and affix. It is possible to produce a new word by using the word building suffixes. The root word may connect many different suffixes and can generate a different form. There are 5000 root words and many more suffixes. Lambadi language syntax has SOV (subject object verb) structure. Lambadi language words use the following syllable structures: V, CV, CCV, VC, CVC, CCVC, VCC,CVCC, and CCVCC. Here V stands for Vowel and C stands for Consonant. The Lambadi language has more dialects and every dialect has many loan and borrowed words, specifically from many official regional languages where more number of speakers speak.

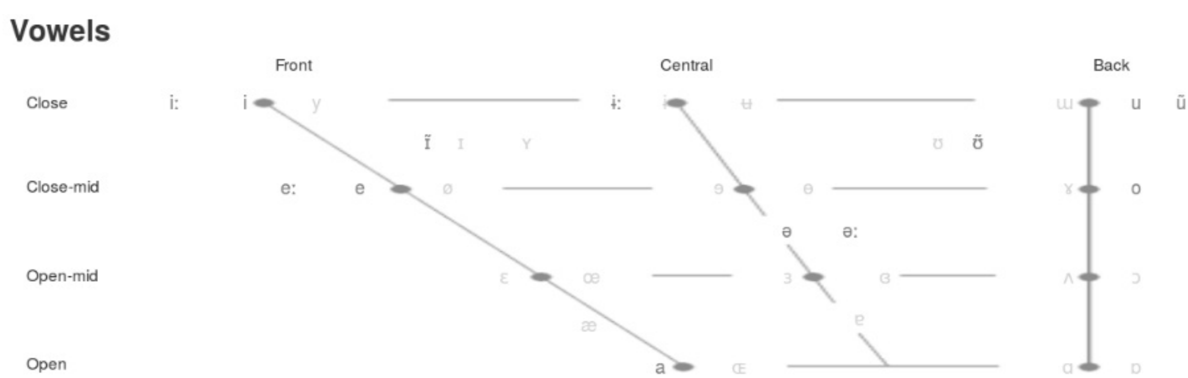


Fig. 1. Lambadi language vowels chart

There are 6 vowels (four vowels occurs in both short and long forms as shown in Figure1{Vowels}), 24 consonants (shown in Figure2{Consonants}), and 5 diphthongs. The Lambadi consonants are divided based on their place and manner of articulation. These consonants are classified into two groups voices and voiceless based on manner. There are six types consonants based on place of articulation. These are 1. Plosives and Affricates 2. Spirants 3. Nasals 4. Laterals 5. Flaps and 6. Continuants. Plosives are all unaspirated consonants. There are eight stop sounds, /p/, /b/, /t/, /d/, /T/, /D/, /k/, /g/, and two affricates, /c/ and /j/. Among these stop sounds and affricates /p/, /b/, /D/, /c/, and /j/ are have two allophones. These allophones vary based on their occurrence in a word or phrase. The Spirants /s/ have four allophones and these allophones are voiceless.All these consonants are represented by using an ILSL sound labels.

Consonants (Pulmonic)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b		t̪ d̪	t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ	n̪	n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ		r̪	r					ʀ		
Tap or Flap		ⱱ	ɾ̪	ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant			l̪	l		ɭ	ʎ	ʟ			

Fig. 2. Lambadi laguage consonants chart

Indian Language Speech Sound Labels

The ILSL labels are developed for 12 regional languages in India. These languages are recognized by the constitution of India. The speech sound labels were key in building speech systems. The ILSL is developed by speech research group (consortia from IITs and IIIT-H) supported by TDIL, DIT, Government of India (GoI).

III. SPEECH DATA COLLECTION PROCESS

Any speech recognition system success depends on quality of speech database and its corresponding text transcription. In our work Lambadi language is a unwritten language and there is no standard orthography available for Lambadi language. We have collected 4000 unique words which contains proper noun, verb, adverb, noun, etc., 800 sentences from the MLE Sarva shikshaa abhiyaan scheme supported by Government of Andhra Pradesh and 200 idioms from Nayak (Village president). These are collected through oral naration which were recorded and then manually transcribed by listening to recorded speech. All these text data was written using Telugu script. We have conducted three sessions for speech database collection over a period of two months. The Lambadi native speakers are from different districts of Telugu speaking states. We have selected speakers (male and female) whose mother tongue is Lambadi and they can read, write and speak in Telugu. All the speakers are graduate students, whose age ranges between 18 to 35. In the first session, we have distributed the Lambadi sentences and asked them to read it once before the actual recording session starts, as different people pronounce differently even though they know the Telugu script. We have 5600 unique sentences which were divided into 70 sentences for 80 speakers. Each session consists of 30 minutes duration. We have given a break of 10 minutes for each speaker to rest their vocal cords. The recording environment is a professional recording studio with noise free environment, The noise level was 15 dB SNR (Signal to Noise Ratio). The Sennheiser omni directional microphone was connected to a computer which has the audacity signal processing open source tool kit is installed in it. Then each speaker speaks for 30 minutes with 10 minutes break in between. Each speaker session was recorded and manually segmented and annotated correspondingly. Each segmented utterance contains speaker Id, name, gender and utterance Id (eg. 01_ABCD_M_04 represents 4th utterance of 1st speaker with name ABCD and gender male). Then all the utterances were recorded at 44 kHz and 24 bit sampling rate. We have downsampled the wav files into 16 kHz with 16 bit sampling rate. This downsampled speech data was used for acoustic models and build speech recognition system.

IV. AUTOMATIC SPEECH RECOGNITION

The components of any speech recognition system are training set, test set and decoding set. We have divided our speech database into training data and test data sets. Our speech database consists of 80 speakers (among these speakers 50 are male and 30 are female). Training database consists of 40 male and 20 female speaker utterances which are 4200 total utterances for training and test database consists of 10 male and 10 female speakers with 1400 speech utterances. The total duration of speech database was 6 hours of speech utterances. Training set contains 4.5 hours of speech and test set contains 1.5 hours of speech. Any acknowledgment to fellow researchers or funding grants should be placed within this section.

Lambadi Speech Recognition System

We have used the Lambadi language speech database for building speech recognition system. The speech data was down-sampled at 16 kHz, 16 bit with single channel speech. A 25 msec frame length, Hamming window and 10 msec frame shift were selected during feature extraction (Mohan et al., 2014). The Mel Frequency Cepstral Coefficient (MFCC) of 39 dimensional feature vector comprising of delta, delta+delta were used as parameters. We used Sphinx-3 speech recognition open source tool kit, developed by CMU speech group. The acoustic models (Schultz and Waibel, 2011) were built based on Guassian Mixture Model (GMM) - Hidden Markov Model (HMM) (Rabiner, 1989). The monophone and triphones are the speech units for Lambadi speech recognition system. The left to right three state context independent HMMs (Uebler, 2011) were used for each phoneme and the GMM of 3 mixture components per state were considered (Ayush and Damdinsuren, 2013). This experiment was a baseline speech recognition system for Lambadi language. One of the better models for the state of the art speech recognition for low resourced languages is SGMM-UBM (Sub Space Gaussian Mixture Model-Universal Background Model). In this experiment, we have used Kaldi tool (open source tool for ASR) the feature vectors are same as the baseline recognition system. The acoustic models were modeled using SGMM (Mohan et al., 2014).

V. RESULT AND DISCUSSION

Table 1

Kaldi toolkit	**Features**	**Word error rate**
Mono Phones	MFCC	47%
Tri-1	MFCC, LDA	32%
Tri-2	MFCC, LDA, MLLR	28%
sgmm	MFCC, UBM-HMM	18%

In this study, word error rate is used as the performance metrics. The baseline system, GMM-HMM is implemented with different variants (monophones, Tri-1, Tri-2 and sgmm) The results of baseline GMM-HMM system of Lambadi speech recognition system is tabulated in Table table:1. The overall performance of baseline system is utmost 18% word error rate. Kaldi based SGMM system gave better results.

VI. CONCLUSIONS AND FUTURE SCOPE

In this work, we have collected speech and text corpus for Lambadi language. The corpus offers potential for research in low resource language like Lambadi. Corpus is annotated manually using automated segmentation. The corpus is evaluated using different acoustic speech recognition systems, namely GMM-HMM and SGMM. Among the models, SGMM performed better in-terms of word error rate compared to GMM-HMM. One can concentrate on data collection in a automated or semi automated way to build large corpus. Code mixing and code switching models can be tried on the built corpus along with the deep learning architectures.

ACKNOWLEDGEMENTS

Authors would like to thank all the people who have involved in speech data collection and text transcription of Lambadi speech database. Also thanks to Mr. G. Shankar and Mr. B. Harisingh for their help in developing text corpus.

REFERENCES

[1] A. Ayush and B. Damdinsuren. 2013. A Design and Implementation of HMM based Mongolian Speech Recognition System. volume 2, pages 341–344.

[2] Pallavi Baljekar, Sunayana Sitaram, Prasanna Kumar Muthukumar, and Alan W. Black. 2015. Using articulatory features and inferred phonological segments in zero resource speech processing. In Proceedings INTERSPEECH, pages 3194–3198.

[3] Joachim Kohler. 2011. Multilingual phone models for vocabulary-independent speech recognition tasks. In Proceedings of the Speech Communication, volume 35, pages 21 – 30.

[4] Aanchan Mohan, Richard Rose, Sina Hamidi Ghalehjegh, and S. Umesh. 2014. Acoustic modelling for speech recognition in indian languages in an agricultural commodities task domain. Speech Communication, 56:167 – 180.

[5] L. R. Rabiner. 1989. A tutorial on Hidden Markov Models and selected applications in Speech Recognition. Proceedings of the IEEE, 77(2):257–286.

[6] Tanja Schultz and Alex Waibel. 2011. Language independent and language-adaptive acoustic modeling for speech recognition. 35:31–51.

[7] Sunayana Sitaram, Sai Krishna Rallabandi, Shruti Rijhwani, and Alan W. Black. 2016. Experiments with cross-lingual systems for synthesis of code-mixed text. In 9th ISCA Speech Synthesis Workshop.

[8] Sunayana Sitaram, Sukhada Palkar, Yun-Nung Chen, Alok Parlikar and Alan W Black. 2013. Bootstrapping text-to-speech for speech processing in languages without orthography. In proceedings of ICCASP, pages 7992–7996.

[9] Ronald L. Trail. 1970. The grammar of Lamani, volume 24. Benjamin F. Elson, United States of America.

[10] Ulla Uebler. 2011. Multilingual speech recognition in seven languages. In Proceedings of the Speech Communication, volume 35, pages 53–69.

[11] Sreedhar Bhukya, 2018, Accent issues in continuous speech recognition system, AICC, Springer, pp 407-414

[12] Sreedhar Bhukya, 2018, Effect of Gender on Improving Speech Recognition System, International Journal of Computer Applications (0975 – 8887), Volume 179 – No.14, pp 22-30.

[13] https://www.ethnologue.com/language/lmn

Evaluation of Suprasegmental Features of Speech using Python

Akshat Agrawal[1], Anurag Jain[2]

USIC&T, Guru Gobind Singh Indraprastha University, Delhi, India
[1]akshatag20@gmail.com, [2]anurag@ipu.ac.in

Abstract: Now-a-days, speech is the fastest and most commonly used means of communication. This paper aims to study the effect of three different attributes i.e. the pitch, energy and duration in speech. Various local features were gathered, and tests were performed to validate the effect of above-mentioned attributes in speech. It was found that each of the three attributes individually contribute in identification of various emotions in any speech.

Keywords: Speech features, local speech features of speech, energy of speech, pitch of speech, duration of speech

I. INTRODUCTION

Speech is very effective and efficient way of communication in humans. it encourages researchers to rigorously work on speech signals and figure out various feature of speech.[1]While a person communicates via speech to any human being, so he/she uses so many emotions on it. So, emotion is the major and impactful factor in speech communication. When a person communicate to another person so us via speech so most of the time both can understand the emotion of everybody but while a person communicates with a machine so at that time machine can understand that what speaker has speaks but cannot predict the emotional state of speaker[2]. So, the emotion recognition concept is coming into the picture. Before analyzing the emotional state of person in speech there is a requirement to analyze the features that is responsible for different emotion in the speech. Mainly speech is having two kind of features local features and global features in speech. Initially we focus on basic features of speech like pitch, energy and duration of speech.[3]For experimenting. Speech can be partitioned into sound portions, which offer some normal acoustics and articulatory properties with each other for a short interim of time. Since the speaker wishes to deliver a sound arrangement comparing to the message to be passed on, most significant vocal tract developments have a deliberate premise. In this area, an exertion has been made to clarify the speech and acknowledgment process which looks like the system of speech creation[4]. In this field ample amount of research work is going on to find out emotion in voice.[5]concluded the importance of fine structure of pitch contour in emotional cue Pitch shape of expressive speech are analyzed. Explained most important variation in pitch as pitch mean and range[6].compared linguistic and paralinguistic features of pitch and also explained pitch shape association of linguistics structure of speech [5]. have worked on the role of pitch contours in differentiating anger and joy in speech. Out of the basic emotions, on anger specifically much acoustic research has been done. [7]. explained the effects of angry emotional tone, with only a word long content and found the fundamental frequency (F0) contour appears to remain steady or fall slightly and the mean duration is shorter for an ‘angry’ word. [8]. In this paper first we are defining the importance of speech features in speech communication then we are showing how basic features of speech depicts our emotional state and then the tradeoff between the all basic features of speech.

II. SPEECH DATABASE

We made our own speech corpora because of necessity of language and sentences and no openly accessible database gives this facility in focused language. In spite of the fact that there are a wide range of classes of feeling, individuals have the normal view that there are five distinctive fundamental feelings: anger , happy, neutral, sad, surprise[9]. In our methodology, we utilized just five essential feelings for example happy, anger, surprise and sad alongside neutral speech style to make correlation. To keep away from exaggerated feeling nonprofessional speaker did accounts speakers articulated 7 sentences in Hindi dialects in five distinct feelings, came about an entirety of 35 expressions. These articulations are recorded in quite condition in *.wav configuration having an inspecting pace of 16 kHz with an accuracy of ~6 bits per test. To deliver the speech corpora exact 4 audience members autonomously assess these expressions. At last dismissed expressions would be re-recorded again and same method rehashed until database is rectified.

III. IMPORTANCE OF FEATURES IN SPEECH COMMUNICATION

Speech is a significant medium not exclusively to pass on carefully semantic substance of sentences yet in addition the statements of mentalities and feelings of the speaker. Pitch, energy, duration utterances are said to be significant prosodic prompts. Inflection example is one of the determinants of the feeling passed on in speech.[10]. Prosodic feature or local feature plays a vital role in terms of emotion analysis in speech. While a person utters a word at that time his voice is having so many variations and we can identify those variation by following features[11].

IV. COMPARATIVE STUDY OF LOCAL FEATURES IN SPEECH

Here we have performed some experiment mainly three basic features Pitch, Energy, Duration of speech shows variation in your voice, and its fantastically denotes basic feature of voice. We use pitch to express our feelings and frame of mind through an adjustment in our intonation, or the tone of our voice. We likewise use pitch to express pressure, or when we make certain syllables longer, stronger, and higher in pitch. To sound progressively regular, you should have the option to control your pitch. Numerous non-local speakers sound somewhat automated when talking since they don't have a ton of pitch variety in their voice. Their tone will in general be quite level and they sound somewhat mechanical, like a robot. In case you're not making this pitch variety in your speech, the individual who is tuning in to you will have trouble getting you.[12] A local speaker's ear is hoping to hear this pitch variety, particularly on focused on words and focused on syllables. Past talking unmistakably, your pitch can impact individuals' impression of you. Pitch is critical for clear correspondence. Here we have shown variation in pitch of a human voice with different emotions. In Table 1 Values of pitch of same sentence with different emotions.

Table 1. Pitch values

Pitch				
Anger	Happy	Neutral	Sad	Surprise
304.1379	265.6627	252.6627	292.053	408.3333

Energy is another feature for speech through this we can distinguish our speech emotion. Here we are showing value of energy of same utterance in different emotion. In Table 2 value of energy of same sentence in different emotion.

Table 2. Energy values

Energy				
Anger	Happy	Neutral	Sad	Surprise
2.913786	2.219749	0.934707	0.786147	2.26904

Figure1 shows of pitch value of anger, happy, neutral, sad, surprise in graph it clearly shown same sentence uttered in different emotion and their graphical structure are different from each other.

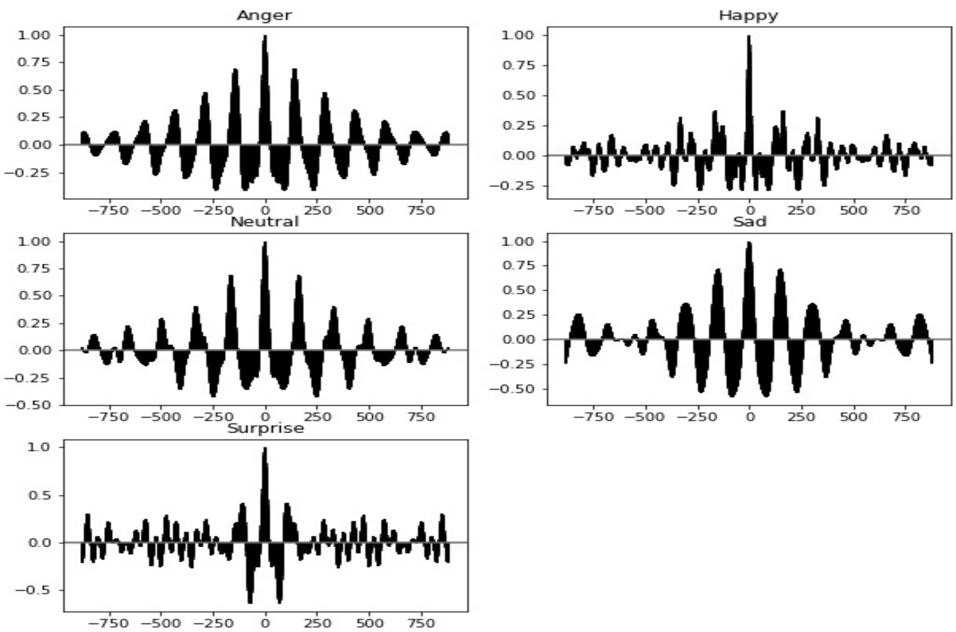


Fig. 1. Variation of pitch in different Utterance in different emotion

Figure2 shows energy variation between same sentence in different emotion graphically all the emotions are different.

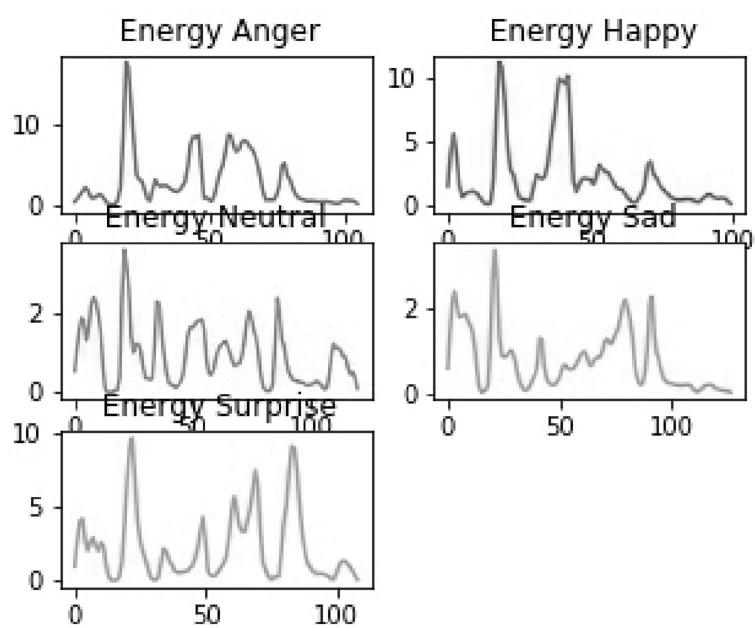


Fig. 2. Energy variation

V. ANALYSIS AND RESULT

We have collected database of non-dramatic actors and they all have uttered different sentences in different emotion here we have computed values of Pitch , Energy & Duration of speech and all the values depicts that values of all these features are near about similar for same emotion Table 2 shows the results of all experiment have been done .We have used Python its supportive libraries as a programming language to find out all the features values.

Table 3. Pitch, Energy, Duration comparison all sentences

Sen	Anger Pitch	Anger Energy	Anger Duration	Sen	Happy Pitch	Happy Energy	Happy Duration
1	380.1724138	4.468282683	0.940408163	1	329.1044776	5.058219415	0.969433107
2	191.7391304	2.382445365	1.079727891	2	74.74576271	3.131638937	1.073922902
3	361.4754098	14.33319608	1.050702948	3	315	5.002407223	1.004263039
4	315	2.999307596	1.195827664	4	310.5633803	1.022905162	1.404807256
5	350	7.392036479	0.946213152	5	334.0909091	2.6721068	1.056507937
6	169.6153846	8.542222627	1.114557823	6	386.8421053	2.360754416	1.108752834
7	310.5633803	2.913785773	1.230657596	7	268.902439	2.219749402	1.160997732

Sen	Neutral Pitch	Neutral Energy	Neutral Duration	Sen	Sad Pitch	Sad Energy	Sad Duration
1	182.231405	2.564311354	0.841723356	1	286.3636364	0.274822745	1.166802721
2	319.5652174	0.968261084	1.242267574	2	286.3636364	0.339073039	1.491882086
3	297.972973	1.219364254	1.428027211	3	268.902439	0.309044908	1.248072562
4	272.2222222	0.690942347	1.56154195	4	290.1315789	0.717950816	1.64861678
5	282.6923077	1.404164834	1.155192744	5	279.1139241	0.496443291	1.242267574
6	268.902439	0.628204652	1.503492063	6	286.3636364	0.542089446	1.364172336
7	268.902439	0.934706759	1.404807256	7	294	0.786146873	1.468662132

Sen	Surprise Pitch	Surprise Energy	Surprise Duration
1	286.3636364	0.274822745	1.166802721
2	286.3636364	0.339073039	1.491882086
3	268.902439	0.309044908	1.248072562
4	290.1315789	0.717950816	1.64861678
5	279.1139241	0.496443291	1.242267574
6	286.3636364	0.542089446	1.364172336
7	294	0.786146873	1.468662132

1. Energy vs Duration

In this experiment we have shown combination of energy of speech and duration in speech using both features we have found out different results in different utterances.

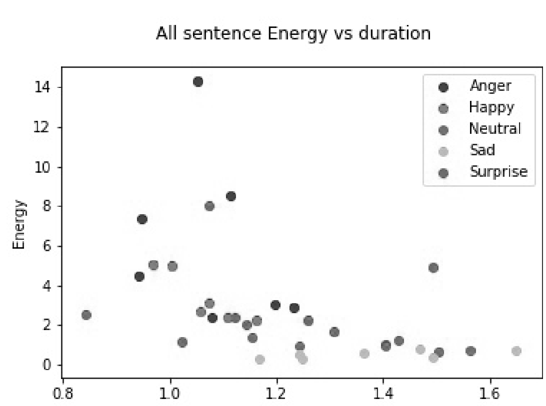


Fig. 3. Experimental analysis of Energy vs Duration of All sentences

2. Duration vs Pitch

In this experiment we have shown combination of duration in speech and pitch and using both features we have found out different results in different utterances.

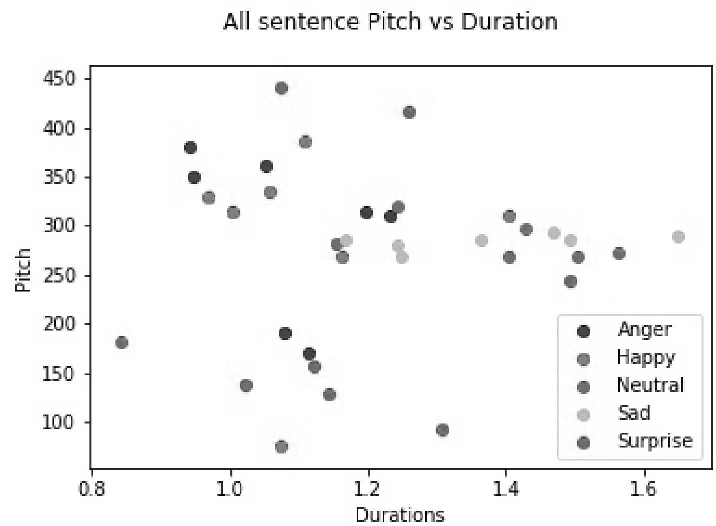


Fig. 4. Experimental analysis of Pitch vs Duration of All sentences

3. Energy vs Pitch

In this experiment we have shown combination of energy of speech and pitch and using both features we have found out different results in different utterances.

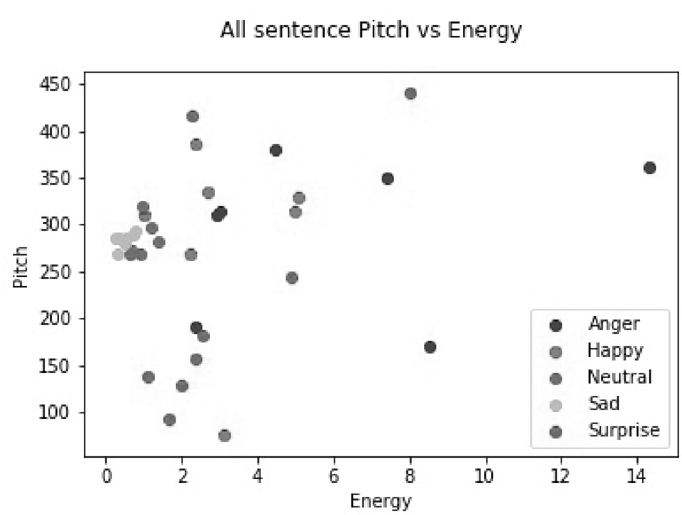


Fig. 5. Experimental analysis of Pitch vs Energy of All sentences

VI. CONCLUSION AND FUTURE WORK

In this comparative study of we identified three basic features like pitch , energy and duration of speech foe the evaluation and we analyzed that using these feature we can we distinguish emotion but we cannot exactly identify the emotion because the values of pitch and energy for emotion Anger and happy will be more likely same .we will try to incorporate some global features like MFCC and LFPC also with the same dataset and then try to detect emotion.

REFERENCES

[1] A. Agrawal, A. Jain, and B. S. Kumar, "Deep Learning Based Classification for Assessment of Emotion Recognition in Speech," *Available SSRN 3356238*, 2019.

[2] L. S. Chen and T. S. Huang, "Emotional expressions in audiovisual human computer interaction," in *2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No. 00TH8532)*, 2000, vol. 1, pp. 423–426.

[3] G. Shashidhar, K. Koolagudi, and R. Sreenivasa, "Emotion recognition from speech: a review," *Springer Sci. Bus. Media*, vol. 15, pp. 99–117, 2012.

[4] A. Jain, "Acoustical analysis of Emotions in Hindi speech and their Transformations using Prosodic Information," 2014.

[5] P. Lieberman and S. B. Michaels, "Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech," *J. Acoust. Soc. Am.*, vol. 34, no. 7, pp. 922–927, 1962.

[6] T. Bänziger and K. R. Scherer, "The role of intonation in emotional expressions," *Speech Commun.*, vol. 46, no. 3–4, pp. 252–267, 2005.

[7] S. Chuenwattanapranithi, Y. Xu, B. Thipakorn, and S. Maneewongvatana, "The roles of pitch contour in differentiating anger and joy in speech," *Int. J. signal Process.*, vol. 3, pp. 129–134, 2006.

[8] J. W. Mullennix, T. Bihon, J. Bricklemyer, J. Gaston, and J. M. Keener, "Effects of variation in emotional tone of voice on speech perception," *Lang. Speech*, vol. 45, no. 3, pp. 255–283, 2002.

[9] R. Cowie, "Emotional states expressed in speech," *ISCA ITRW speech Emot. Dev. a Concept. Framew. Res.*, pp. 224–231, 2000.

[10] S. S. Agrawal, N. Prakash, and A. Jain, "Transformation of emotion based on acoustic features of intonation patterns for Hindi speech," *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 10, no. 9, pp. 198–205, 2010.

[11] V. Gupta, A. Jain, and N. Kumar, "Identification of emotion stress agents in Hindi and English sentences," in *2009 International Conference on Intelligent Agent & Multi-Agent Systems*, 2009, pp. 1–2.

[12] P. N. Juslin and P. Laukka, "Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion.," *Emotion*, vol. 1, no. 4, p. 381, 2001.

Speech Classification: Issues, Challenges and Present State of Research

Parul Rawat[1] and Arun Sharma[2]

[1,2]Department of Information Technology
Indira Gandhi Delhi Technical University for Women, Delhi, India
[1]parulrwt13@gmail.com, [2]arunsharma@igdtuw.ac.in

Abstract: With the advancement in computing technologies one of the emerging areas is of speech technology and speech analytics. A speech by a speaker contains various important features which can be used to detect the gender, age and emotional state of the speaker. This paper gives a concise review of the works that have used various approaches to analyze speech features, extracted from voice recordings or live calls, to predict the gender, age and emotional state of the speaker and others.

Keywords: Speech Features, Voice Features, Emotional State, Speech Classification

I. INTRODUCTION

Speech technology is one of the types of technologies used for computing which is making electronic devices smarter and efficient in recognizing, analyzing and understanding the words which are being spoken or a recorded audio. To accomplish this task digital voice signals are given as input and then the resultant pattern is matched with the stored library pattern. Speech technology stands on the pillars of technologies like signal processing and machine learning. Using signal processing technique information like speaker's voice characteristics, noise at the background and frequency are extracted. Then machine learning technique is used for recognizing and analyzing these voice signals to get the desired output. Gender, age and emotion of speaker while he/she is speaking can be predicted by using speech analytics.

1. Applications of Speech Technology

Speech technology finds its use in many fields now-a-days. It is used in sectors like heath care, enterprise, law and personal use. Some examples are listed below:-

- Audio evidences are not considered as important evidence in the perception of law but speech technology can be applied for identifying and validating a person's voice and hence enforcing law.
- Speech technology can enable healthcare service providers to aid patients with visual impairment or hearing deficiency, handicapped person having crippled or no hands or fingers.
- Devices like Siri, Google Home and Amazon Alexa use speech technology to provide personal assistance to the user.
- The functionality of IVR and speech to text technology, type of speech technology, are been used by companies for customer services and support.

2. Features of Speech

1) Male vs Female

Pitch (Frequency): Low pitches (A2 to C3) are generally associated with males whereas high pitches (A3 to C4) are generally associated with females.

Articulation: Male voices have a rougher articulation as compared to females who have an articulation which is gentler than male voice articulation.

Intonation: Males use full range of the pitch while they speak whereas females have varying pitch ranges while they speak.

Conversational style: While emphasizing on something, males generally raise their volume but females generally raise their pitch while emphasizing on something.

2) Emotions

Anger: The emotional state of anger is associated with a higher intensity, faster attack times at voice onset (the start of speech), lower pitch, higher first formant (first sound produced) and greater amount of energy (approx. 500 Hz) throughout vocalization, as compared to neutral speech.

Fear: The emotion of fear is associated with rate of speech that is faster and consist of more number of pauses, pitch that is higher, small variation, and smaller amount of energy as compared to neutral speech.

Sadness: The emotion of sadness is associated with longer duration with more pauses, a lower first formant ,pitch that is higher, and less intense but with greater amount of vocal energy (approx. 2000 Hz), as compared to neutral speech.

3) Age

Speech rate is used to predict the age of the speaker. Faster speech rate is associated with young speaker whereas slower speech rate is associated with old age.

II. LITERATURE REVIEW

In this paper a review of research papers ranging from year 2010 to 2019 has been provided. 10 research papers have been taken into consideration from Elsevier journal, Czech Science Foundation GA CR and INTERSPEECH 2017 August 20–24, 2017, Stockholm, Sweden. These research papers present the implementation of different approaches to be used to classify a speaker on the basis of gender, age and the emotional state of the person speaking.

Classification Basis

1) Gender and Age

Pitch range (Pitch changes over time) feature set has been used by Barkana, B. D., & Zhou, J.[1]. This feature set has been extracted from dataset aGender corpus which has been provided by the Inter-speech 2010 Paralinguistic Challenge held on August 20–24, 2017, at Stockholm, Sweden. Two-third of the dataset has been used for training. One-third of the dataset, that remains, has been used for testing. Feature extraction has been done as a part of the data pre-processing process. Speech utterances are collected using Voice Activity Detection. KNN and SVM classifiers have being used to evaluate the feature set that has been extracted from the speech utterances. PR features give promising accuracies but shows low classification accuracies for young male speakers.

Spectral and prosodic features of speech have been extracted by the researchers as the feature set to be used in their paper. Feature extraction has been done as a part of the data pre-processing process. Features have been extracted from the dataset aGender corpus [9, 10] provided by the Inter-speech 2010 Paralinguistic Challenge held on August 20–24, 2017, at Stockholm, Sweden. The approach used by Yücesoy, E., & Nabiyev, V. V.[2], is obtained when seven sub-systems are fused together at score level. In the above obtained fused system, the extracted features are used in the model of Gaussian Mixture, SVM and GMM supervector based SVM classifiers for the purpose of classification of the input speech signals. Each system has been combined in several ways and also individual system has been examined to learn the effect of various extracted features and various methods of classification on the performance of the system. It has been observed that the rate of success is 90.4%, 54.1% & 53.5% respectively for gender, age, and age & gender categories.

Feature extraction has been done as a part of the data pre-processing process. Spectral, Temporal and Cepstral features have been extracted by the researchers from the voice signals in the dataset. Dataset consists of the recordings of the conversation over the telephone which has been provided by the Universal Database NIST 2003. Modelization of features has been done with the help of Mel Frequency Cepstral Coefficients (MFCCs), Gaussian Mixture Model-Universal Background Model (GMM-UBM) and the report of Likelihood. Efficiency has been measured using Equal Proportion Probability (EPP). It has been found out that the system does not performs well for female voices or female voices converted to male voices. The proposed system has been observed under both- the clean environment and the noisy environment. It has been observed that efficiency of the system is more in a clean environment as compared to the noisy environment. Three sets of experiments have been performed by Kenai,O.,Djeghiour,S.,Asbai,N.,& Guerti,M. [3] - first set of experiment for male voices, second set of experiment for female voices and the third set for a mixture of male and female voices.

In another paper, researchers have made use of both types of features- the long term features present in the signals of speech and short term features present in speech signals. Long term features like pitch, jitter, formants and intensity and Short term features like MFCCs are extracted from the voice signals of the dataset. Dataset contains speeches by around 700 German speaking people. Feature extraction has been done as a part of the data pre- processing process. Van Heerden, C., Barnard, E., Davel, M., van der Walt, C., van Dyk, E., Feld, M., & Müller, C. [6] have used both regression and classification. Regressors like SVM are used for more accurate estimation of age. Resultant estimations are then combined with respective classifier's posterior probabilities to make predictions about the gender and age of speaker. Training has been done using long term feature like pitch, formants and also some short term features that can be derived from the adaptation of MAP of GMMs. These GMMs were trained with help of MFCCs. SPV regressors are more accurate than LTF regressors even after multiplying them with posterior probabilities of their respective classifiers.

Feature extraction has been done as a part of the data pre-processing process. Features that can be derived from spectrum of the speech signal and prosodic features are extracted from Czech & Slovak dataset for speech classification. Přibil, J., Přibilová, A., & Matoušek, J.[7] have proposed a two level GMM classifier in which first level has been used to recognize the gender of the speaker and second level has been used to recognize the age of the speaker. This proposed approach was firstly applied to children voices and then after it was applied to adult speaker's voice. Accuracy of GMM based model is 92.3%. System shows good accuracies for child voices but lower accuracy for converted gender or converted age voice signals.

2) Emotional State of the Speaker

Feature extraction has been done as a part of the data pre-processing process. Mannepalli, K., Sastry, P. N., & Suman, M.[4] have extracted features like ratio of power of Tone, element of flux in the spectrum of speech signal, Chroma of pitch & MFCCs in the dataset. Two datasets have been used which are:- Berlin Dataset consisting of Emotional Speech [11] and Telugu dataset [12]. The features that are extracted are passed into a network consisting of implementations of the theoretical concepts of fractions and Deep Belief concepts that are adaptive in nature. Weights present in the network of DBN were modified repeatedly using theory of fractions. The weights which were updated before and term of biasness were used in phase of testing of AFDBN in order to recognize emotions. Accuracy achieved by the proposed approach is 99.17% and 97.74% for Berlin & Telugu dataset respectively.

Davletcharova, A., Sugathan, S., Abraham, B., & James, A. P.[5] have used the distance from one peak to another peak in speech signal graph as feature set. Dataset consisted of 3 different Russian speaking people, speaking with a different emotion at a time. Feature extraction has been done as a part of the data pre-processing process. Input speech signals are recorded and then analyzed using MATLAB software which outputs many graphical representations of the input speech signals. These graphical representations are then analyzed. Distance from one peak to another peak in speech signal graph is extracted as the main feature. Study has been done on how the emotions of a person affects his/her word and letter utterances under different emotional circumstances. Different emotions of a person show direct proportionality to voice signals that he/she produces irrespective of what he/she is saying.

Two datasets have been used, first dataset is Berlin Dataset consisting of Emotional Speech [11] and second dataset is Telugu dataset [12]. Spectral features of speech are then extracted from these datasets. Feature extraction has been done as a part of the data pre-processing process. Mannepalli, K., Sastry, P. N., & Suman, M.[8] have developed an algorithm namely the algorithm of Whale-Imperialist optimization (Whale-IpCA) which is a hybrid of Whale Optimization Algorithm (WOA) & Imperialist Competitive Algorithm (IpCA). Whale-IpCA is then used to train multiple SVNN classifier to classify the state of emotion of speaker into different classes. Performance of this approach has been evaluated with frames of varying lengths. Best performance among all the existing models with accuracy =0.9987.

Satt, A., Rozenberg, S., & Hoory, R.[9] have done no extraction of features. The long input of speech is broken into small parts. Each part is approximately 3 seconds long not more than that. None of the smaller parts overlap with each other. Raw spectrograms of each smaller part is fed to a deep neural network consisting of both recurrent networks and convolutional networks. The usage of raw spectrogram has been done so that the unknown signals present in the background, which are not part of signals of speech, can be removed by applying Harmonic technique of filtering. This approach has been proved to be effective than other models with 66% rate of accuracy for convolution only network and 68 % rate of accuracy for a model which is a combination of both – LSTM technique and Convolution technique.

The recognition of emotions has been conducted in two states by Lalitha, S., Geyasruti, D., Narayanan, R., & Shravani, M.[10]. Starting with first step, cepstral features have been extracted from the speech input signals from Berlin dataset of emotions. A tool based on neural network concept has been used to attain the task of emotions recognition. Seven emotions have been recognized successfully with a rate of accuracy which is equal to 85.7 %. But this tool fails to recognize emotion of sadness because happiness and sadness are both extreme emotions whose set of features is very narrow so there are more chances of wrong classification.

III. SUMMARY

S. No.	Objective	Speech Features Considered	Dataset Used	Approach Used	Results
1	Speech classification based on age and gender of the speaker [1]	Pitch range (Pitch changes over time)	aGender corpus [9,10] provided by Interspeech 2010, challenge of Paralinguistic behavior.	Speech utterances collected using Voice Activity Detection. Classifiers like KNN & SVM are used to evaluate given feature set extracted from the speech utterances.	PR features give promising accuracies but shows low accuracy in classifying the voices of young male speakers.
2	Speech classification based on gender & age of speaker [2]	features derived from spectrum of the speech signal and prosodic features	aGender corpus [9,10] provided by the Inter-speech 2010 Paralinguistic challenge organizational body	Model of Gaussian Mixture, SVM and SVM classifier, based on GMM super-vector, used for classifying.	Rate of success is 90.4%, 54.1% & 53.5% respectively in age, gender & both.

S. No.	Objective	Speech Features Considered	Dataset Used	Approach Used	Results
3	Speech classification based on gender & age of speaker [3]	Features present in spectrum of signal of speech. Temporal and Cepstral speech features	Recordings of the conversation over the telephone provided by universal database NIST 2003.	Modelization of features done using Mel Frequency Cepstral Coefficients (MFCCs), Gaussian Mixture Model-Universal Background Model (GMM-UBM) and Report of Likelihood.	Efficiency is measured using Equal Proportion Probability (EPP). System does not performs well for female voices or female voices converted to male voices.
4	Speech classification based on state of emotion of speaker [4]	ratio of power of Tone, element of flux in the spectrum of speech signal,	2 datasets used:- Berlin Dataset consisting of Emotional Speech [11] and Telugu dataset [12] Chroma of pitch & MFCCs	Features that are extracted are passed into a network consisting of implementations of the theoretical concepts of fractions and Deep Belief concepts that are adaptive in nature. Weights present in the network of DBN modified repeatedly using theory of fractions. The weights which were updated before and term of biasness were used in phase of testing of AFDBN	Accuracy achieved is 99.17% and 97.74% respectively for Berlin & Telugu dataset.
5	Speech classification based on emotions of the person who is speaking [5]	Distance from one peak to another peak in the speech signal graph	3 different Russian speaking people, speaking with different emotion at a time.	Study on use of letters and words under different emotional situations was done using the graphical representation of the speech signal.	Different emotions of a person show direct proportionality to voice signals that he/she produces irrespective of what he/she is saying.
6	Speech classification based on age and gender of the speaker [6]	Long term features:- pitch, jitter, formants and intensity. Short term features:- MFCCs	Dataset contains speeches by around 700 German speaking people.	Regressors like SVM are used for more accurate estimation of age. Resultant estimations then combined with respective classifier's posterior probabilities.	SPV regressors are more accurate than LTF regressors even after multiplying them with posterior probabilities of their respective classifiers.

S. No.	Objective	Speech Features Considered	Dataset Used	Approach Used	Results
7	Speech classification based on age & gender of speaker [7]	features derived from spectrum of the speech signal and prosodic features	Czech & Slovak dataset for speech classification	Two level GMM classifier was used. First level to recognize the gender and the second level to recognize the age of the speaker.	Accuracy is 92.3%. System shows good accuracies for child voices but lower accuracy for converted gender or converted age.
8	Speech Classification based on the emotional state of the speaker [8]	Spectral features	2 datasets used:- Berlin Dataset Consisting of Emotional Speech [11] and Telugu dataset [12]	A network consisting of implementations of Theoretical concepts fractions and Deep Belief of concepts has been used. MSV classifier machine based on the algorithm of Whale-Imperialist Optimization has been used. Whale IpCA trains Multi-SVNN to identify emotions.	Best performance among all the existing models with accuracy =0.9987
9	Speech classification on basis of emotions of speaker [9]	NONE	IEMOCAP[15] Dataset	No extraction of feature is done. Long input signal of speech are segmented to smaller segments of not more than 3 seconds long. Raw spectrograms of these smaller signals of speech are fed into deep neural network.	Only Network of convolution: 66% rate of accuracy. LSTM + Convolution technique : 68 % rate of accuracy
10	Classification of speaker on basis of emotions. [10]	Cepstral coefficients	Berlin Dataset consisting of Emotional Speech.	A tool based on neural network for recognizing patterns in speech has been used to recognize emotions.	Rate of Accuracy = 85.7% for 7 emotions. Failed to recognize sad emotion

IV. CONCLUSIONS AND RECOMMENDATION FOR FURTHER WORK

This paper has presented a concise review of all the techniques, mentioned in various research works, that have been used to extract various speech features like pitch range feature set, MFCCs, distance between two peaks in a voice signal, cepstral features etc. . The extracted features are then analyzed to predict the age, gender and emotional state of the speaker using various approaches. The voice samples have been collected either from voice recordings or live telephonic calls. Some papers have presented how accuracies differ in clean and noisy environments. For taking speech classification to much finer level it is recommended to aim for developing multi-layer identifier models as future work, where each layer is used to identify a different factor i.e age, gender or emotional state of the speaker with better than existing accuracies for all genders, age group and emotions. It is also suggested to try to eradicate the falsification introduced in speech regarding gender, age or emotional state like for eg: a male faking a female voice.

REFERENCES

[1] Barkana, B. D., & Zhou, J. (2015). A new pitch-range based feature set for a speaker's age and gender classification. *Applied Acoustics*, *98*, 52-61.

[2] Yücesoy, E., & Nabiyev, V. V. (2016). A new approach with score-level fusion for the classification of a speaker age and gender. *Computers & Electrical Engineering*, *53*, 29-39.

[3] Kenai,O.,Djeghiour,S.,Asbai,N.,& Guerti,M. (2019). Forensic Gender Speaker Recognition under Clean and Noisy Environments. *Procedia Computer Science*, *151*, 897-902.

[4] Mannepalli, K., Sastry, P. N., & Suman, M. (2017). A novel adaptive fractional deep belief networks for speaker emotion recognition. *Alexandria Engineering Journal*, *56*(4), 485-497.

[5] Davletcharova, A., Sugathan, S., Abraham, B., & James, A. P. (2015). Detection and analysis of emotion from speech signals. *Procedia Computer Science*, *58*, 91-96.

[6] Van Heerden, C., Barnard, E., Davel, M., van der Walt, C., van Dyk, E., Feld, M., & Müller, C. (2010, March). Combining regression and classification methods for improving automatic speaker age recognition. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 5174-5177). IEEE.

[7] Přibil, J., Přibilová, A., & Matoušek, J. (2016, July). GMM-based speaker gender and age classification after voice conversion. In *2016 First International Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE)* (pp. 1-5). IEEE.

[8] Mannepalli, K., Sastry, P. N., & Suman, M. (2018). Emotion recognition in speech signals using optimization based multi-SVNN classifier. Journal of King Saud University-Computer and Information Sciences.

[9] Satt, A., Rozenberg, S., & Hoory, R. (2017). Efficient Emotion Recognition from Speech Using Deep Learning on Spectrograms. In INTERSPEECH 2017 August 20–24, 2017, Stockholm, Sweden (pp. 1089-1093).

[10] Lalitha, S., Geyasruti, D., Narayanan, R., & Shravani, M. (2015). Emotion detection using MFCC and cepstrum features. Procedia Computer Science, 70, 29-35.

[11] Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., & Narayanan, S. S. (2010). The INTERSPEECH 2010 paralinguistic challenge August 20–24, 2017, Stockholm, Sweden. In Eleventh Annual Conference of the International Speech Communication Association.

[12] Burkhardt, F., Eckert, M., Johannsen, W., & Stegmann, J. (2010, May). A Database of Age and Gender Annotated Telephone Speech. In LREC.

[13] Website for emotional speech dataset :- http://emodb.bilderbar.info/docu/#docu

[14] Website for Telugu dataset for emotion detection:- http://www.bhimalapuram.co.in/telugusoft/mysql4telugu.html

[15] Busso, C., Bulut, M., Lee, C. C., Kazemzadeh, A., Mower, E., Kim, S., ... & Narayanan, S. S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. Language resources and evaluation, 42(4), 335.

TRACK 3

MACHINE LEARNING - II

SAHYATRI

Asmita Bhardwaj[1], Shruti Bhatia[2], Richa Yadav[3]

[1,2,3]Deptt. of ECE,
Indira Gandhi Delhi Technical University for Women, Delhi, India
[1]asmita.bhardwaj1997@gmail.com

Abstract: With over 1,65,000 vehicles stolen and 1,50,000 deaths due to car accidents in India in 2018, the designed device - SahYatri is a smart, affordable, and economic, virtual vehicle security system designed as a solution to these apocalyptic problems. SahYatri is an Internet of Things based telematics solution that provides GPS-based vehicle tracking, an anti-theft digital lock, emergency notification alerts to family members, remote access to control vehicle's state, drowsiness detection and vehicle health alert, prevents drunk driving and offers an entertainment screen to listen to one's favorite songs. The solution is easily accessible to the users with the help of a compact device that needs to be installed in the vehicle and a mobile app that facilitates the control of the device over a 2G GSM network.

Keywords: IOT(Internet of Things), Emergency notification alert, Image processing, Vehicle health alert, Alcohol sensing, GPS based real-time tracking, Android app.

I. INTRODUCTION

The Internet of Things has introduced a world which is quantifiable and measurable. People nowadays can manage and utilise their assets in better and informed ways. They can make more timely and better informed decisions about what they want or what and how they need to do. This new world has brought in many improvements in the lives of its users such as convenience, health, safety and security.

The team conducted an online survey to study the response of vehicle users to a potential Vehicle Security and Safety system. Out of 385 people surveyed, 58.9% were females and 41.1% were males, 96% of them felt vulnerable while travelling, and out of all those who answered, 56.8% were ready to adopt vehicle security and safety measures and 40.20% preferred having anti-theft and safety system for vehicles if it is cost effective.

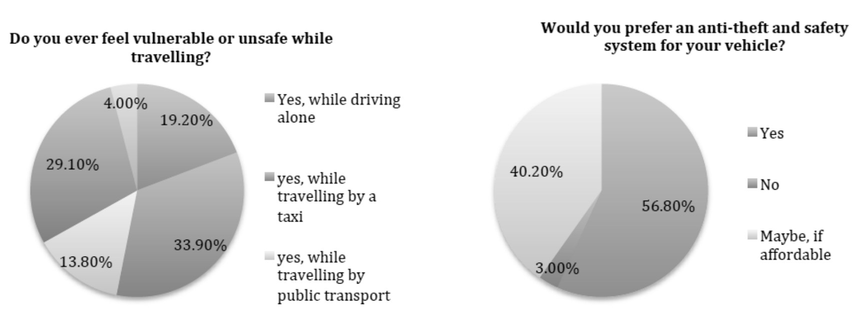


Fig. 1: Pie chart showing market analysis

As 90.8% people claimed to use smart phones, the designed product can be efficiently installed in vehicles and can be accessed using a mobile application so as to facilitate control of the device.

Thus, the pilot edition of the designed project will primarily focus on the urban users looking for extremely safe modes of personal transport, and would then aim at catering to a sizeable user base section in areas of taxi service providers, ambulance, police services, and school transportation among many others.

The user base primarily consists of travellers belonging to 10-41 years of age who are looking for safe, economic, affordable and easily accessible modes of transportation.

II. RELATED WORK

Table 1: Comparison between Existing solution and Sahyatri

FEATURES	SAHYATRI	Online Cab Services/ Existing Solution
Vehicle Tracking	Yes	Yes
Anti-theft Lock	Yes	No
Emergency Button	Yes	No
Alcohol Detection of driver	Yes	No
Remote access to control vehicle's state	Yes	No
Sending alerts in case of accidents	Yes	No
Monitor vehicle health	Yes	No
Graphical User Interface	Yes	Yes
Drowsy Driver Detection	Yes	No
Favorite songs playlist	Yes	Only in prime play cars
Manual Operations & Maintenance	Very low	Moderate
Charges	Minimal	Expensive
Can be installed in any vehicle	Yes	No, one has to use the cab service only
Operating Time	24*7	Dependent on cab agency

III. CHALLENGES FACED

The designed project SahYatri incorporates a number of features that provide an efficient solution to several driving and road related issues . However, there are a number of technically challenging matters which have repeatedly cropped up during the course of the project building phase. These challenges have not only imbibed a sense of progressive and competitive thinking but also have led us to imagine the actual market perspective of the designed product.

Provision of Internet connectivity: The designed project requires a continuous/all time Internet supply as it is an IOT based product. The internet connection is not an easy task especially when the supply is to provide in a dynamically mobile environment (a moving car/vehicle). This indeed was a very challenging task to be encountered as BeagleBone Black, main controller board does not provide an inbuilt Wifi network. However , product designers came up with the solution of using a Wifi dongle with a TP-Link to enable Internet connection.

Power supply: In the designed product, the BeagleBone Black controller needs to be powered for effective functioning. However, since the product needs to be installed in the vehicles itself, there is no way to provide a direct power supply. Hence, it becomes almost mandatory to provide power supply to the board from the vehicle's battery unit .

User's Permission: Since the project revolves around vehicles and their usage, it becomes mandatory for us to take the user's grant so as to access the vehicle's engine and battery so as to implement features like Auto Lock.

IV. PROPOSED SOLUTION

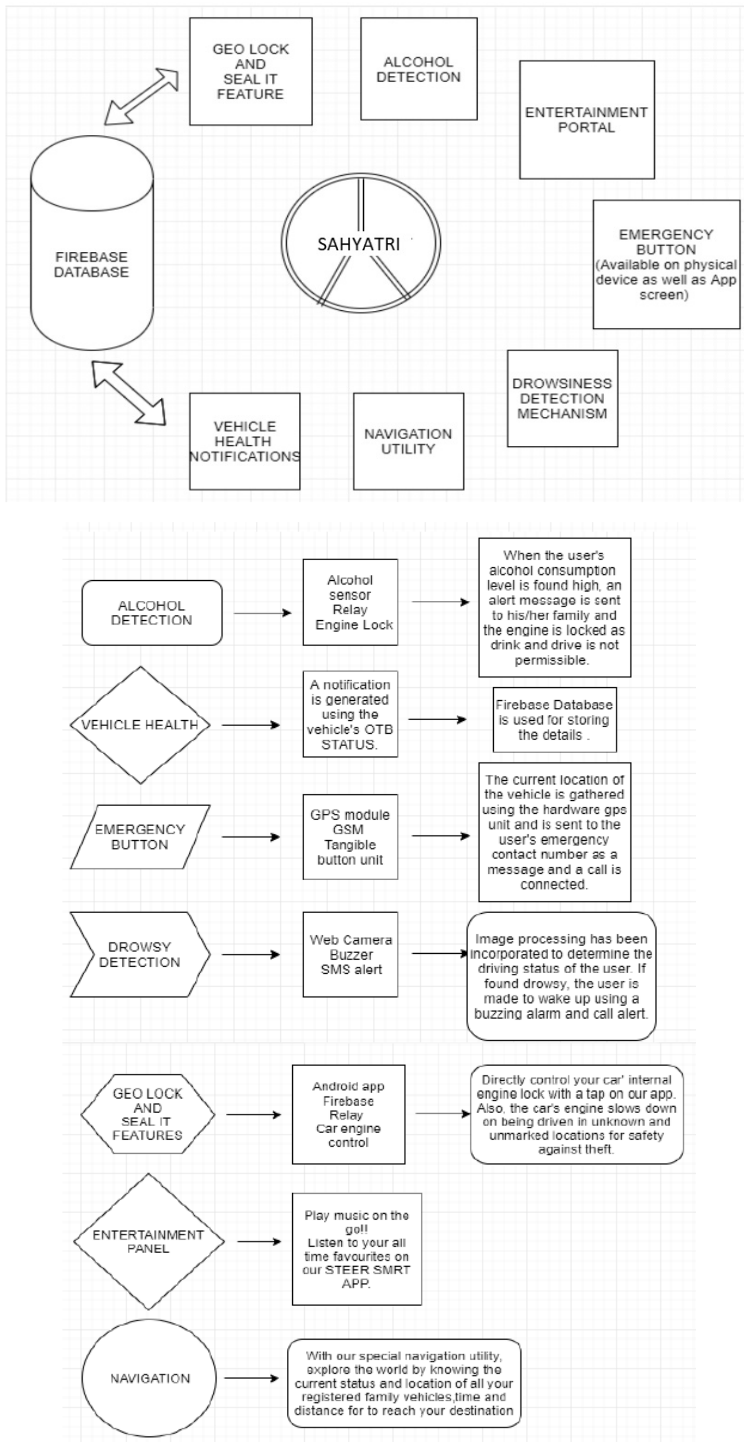


Fig. 2. Block diagram of the features

V. HARDWARE IMPLEMENTATION

1. Feature 1: Real Time Tracking

This feature enables the owner to continuously monitor his/her vehicle through the device's GPS location The SahYatri App also lets the user know the current distance between him/her and the vehicle and the distance between the vehicle's current location and its destination can be determined. The owner can not only keep a track of his/her own vehicle but can also add a number of other vehicles into the list using Family option available on the App.

2. Feature 2: Alcohol Detection

The designed system has detailed alcohol consumption detection technique and algorithm which analyses the amount of alcohol consumed by the driver and takes the necessary action. If the sensor module detects a large quantity of alcohol in the driver's breath (which is not safe for driver), then the vehicle gets automatically locked and a message is sent to the relatives of the driver through the available SahYatri App thereby informing them about the driver's current inability to drive due to high alcohol consumption.

3. Feature 3: Car Lock using app's 'Seal It' feature

This feature allows the owner to lock the car/vehicle directly with the SahYatri App. For unlocking, the owner needs to open the app and login into his/her account and tap the 'Seal It' button under the Smart Lock Screen. This uses a switching circuit which is developed using a relay which is connected to the vehicle's locking circuitry. Also once the vehicle has been unlocked using the App, it can easily be operated using the existing manual car lock mechanism. This will ensure double safety thereby preventing thefts and ensuring a highly secure method of vehicle safety.

4. Feature 4: Accident Detection & Emergency Button

SahYatri hardware mechanism ensures a safe ride through an accident detection feature. This feature uses an accelerometer that detects any changes in the normal positioning of the vehicle (using threshold values) and its usual speed and immediately informs the relatives of the driver through a call. The same services can be availed in case the driver presses the Emergency button available on the SahYatri hardware device for the user's safety and welfare.

5. Feature 5: Drowsiness Detection

The SahYatri device also keeps a constant check on the facial and body responses of the driver using a HD Camera cape with the BeagleBone Black board that constantly monitors the changes in body movements and eye movements of the driver. Product designer have incorporated image processing to determine the driving status of the user. If found drowsy, he/she is made to wake up using a buzzing alarm.

VI. POWER SUPPLY

5V power supply is needed to power BeagleBone Black. This is obtained by using a 12V to 5V DC converter(TPS82130) and converting 12V DC power supply from the internal car battery. In cars with already existing charging ports, a 5V charger can be plugged to the microcontroller directly.

VII. SOFTWARE IMPLEMENTATION

Default screen has login and signup options. Registered users can log in to the app while new users are prompted to fill a form in order to register for the services. Navigation screen displays the real-time location of the vehicle. Smart Lock Screen is used for locking the car engine directly through App Command.

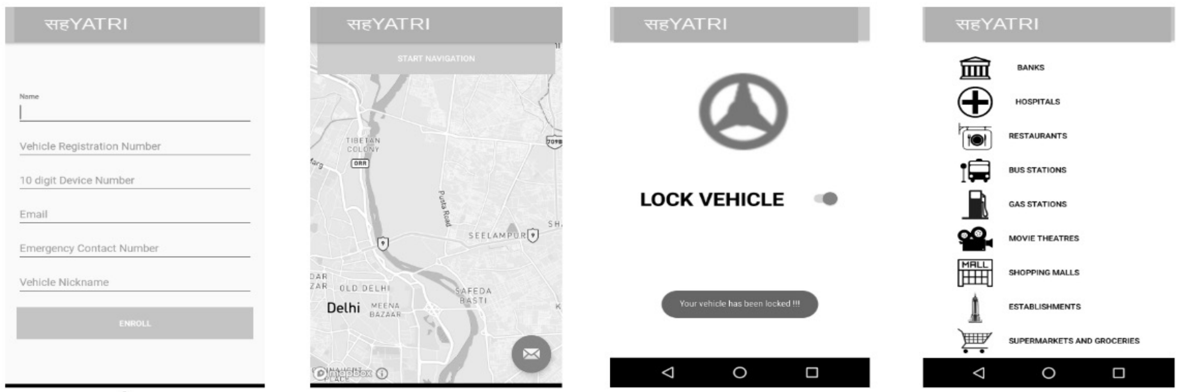


Fig. 3: Software application screen depicting different features of the designed product

VIII. THE PROPOSED RESULT

The uniqueness of SahYatri lies in a compact way in which it amalgamates automated technology with the best possible cutting-edge solution for solving the problems of crime against women, children, elderly and theft. It will not only ensure safety of loved ones but will also help in the reduction of accidents due to alcohol consumption, drowsiness of the driver, decrease delays in ambulance and police response. Long term growth of our country requires huge investments in research and development as well as manufacturing efficient products. As the designed product is cost effective with low power consumption and will be manufactured in India, it will contribute to the economic status of the country.

SahYatri provides the most affordable solution to combat vehicle theft: A report published by Knoema, found that transport vehicle theft rate of India has increased from 8.2 cases filed per 100,000 population in 2005 to 14 cases filed per 100,000 population in 2014 and it is further growing at an average annual rate of 7.95%. The property losses due to vehicle theft in 2014 were estimated at $4.3 billion[3].

A keeper for the women: A 2011 study of over 6000 men and women that is conducted in Delhi by Jagori resource center.The study showed that 52.4% of the women surveyed and those women reported that they had faced a very high levels of sexual harassment in vehicles escpecially buses, taxis and also when they drive alone at odd hours(non business hours). According to the International Labour Organization's Global Employment Trends 2014 report, India has ranked 120th among the 132 countries that are ranked in women's labour force participation. As the presence of women is increasing in the economic activities , hence a transport facility is required that is safe and reliable and it is not just desirable but it is also mandatory for the women.

Peace of mind for the parents and elderly: SahYatri can be effectively used by school transport departments, taxi service providers and ambulance services.For every third heart attack patient in India, it takes more than 450 minutes to reach a nearby hospital, that is almost 14 times more than the ideal window of 30 minutes, the government data showed that. Ambulance delays linked to a huge number of deaths in the past five years.

While sending off the youngster to the school , the parents got really scared and it is a concerning issue for them as they have no idea if their child has reached the school safely or not. Only this concern will disappear when the kid has returned to his/her home safely.Therefore, with the real-time notification feature through the GPS feature(tracking feature), anyone will be able to know where their child is going and why transportation is delayed(whether it is traffic jam, breakdown or any other unfavourable condition).

As the increasing number of new drivers entering the roadway, there is more risks for taxi and cab drivers. The tracking feature through GPS offers security features for both the taxi driver as well as the taxi company along with the notification alerts.

A promise to save lives: Community Against Drunk Driving (CADD) in its report aforesaid that of an average 1,876 deaths because of road accidents had occurred annually, out of which 1,550 are because of drunk driving.

As per CADD report, 55.1 per cent victims are within the age group of 15-35 years. SahYatri has the potential to reduce road accidents due to consumption of alcohol and drowsy drivers.

The designed product is a compact, market-ready solution, which requires a small device and a free mobile application to be installed by the user. It will offer the best performance to price ratio in the market and has over 75% more features and efficiency than the existing solutions, thereby making it feasible.

IX. CONCLUSION AND FUTURE WORK

On a concluding note, the project SahYatri is the most affordable solution to combat vehicle theft, ensures the availability of a safe and reliable transport facility, helps today's young generation to abide to the Road safety rules as it has the potential to bring illegal practices such as Drunk or Drowsy driving to a halt upto a large extent.

The proposed product is a compact, market-ready solution, which requires a small device and a free mobile application to be installed by the user. However, there are still some major scope for further improvement and expansion of this work. The designed system can be expanded with the following ideas:

- Designing a self driven power source for Beagle bone controller instead of utilizing the vehicle's battery.
- Automatic alert to all emergency public services such as Fire stations, Police stations, Ambulance in case of vehicle catching fire, occurrence of theft, accidents respectively.
- Full fledged image processing based vehicle owner identification mechanism and alert system to completely eradicate the possibility of theft.
- Incorporating visual entertainment unit along with the present Music utility available on the SahYatri android application and enabling a touch panel based feature selection

REFERENCES

[1] X. Wang ,J. Yi-Hang, L. Ren-Bo and C. Ming-Zhang, "Automobile safety technology and its improvement", MATEC Web of Conference,Volume 5, Issue 1, July - 2018.

[2] M. Mcdonald, H. Keller, J. Klijnhout, and V. Mauro, "Intelligent Transport Systems in Europe: Opportunities for Future Research", World Scientific Publishing Company,Volume 2, Issue 3, December - 2006.

[3] K. Bimbraw . "Autonomous Cars: Past, Present and Future." Int J Adv Engg Tech,pg-68, Volume 5, Issue 2, April- June ,2014.

[4] Kunal arya and mandeep singh "Real Time Vehicle Tracking System using GSM and GPS Technology- An Anti-theft Tracking System",Deptt of Electronic,Volume-7, Issue-3, March-2012.

[5] B. Lakshmi V, Prof. S. Hiremath, S. Mhamane, "FPGA Based Vehicle Tracking and Accident Warning using GPS",International Journal of Scientific & Engineering Research, Volume 5, Issue 2, February-2014.

[6] K. C. Varma, Poornesh, T. Varma, Harsha, "Automatic Vehicle Accident Detection And Messaging System Using GPS and GSM Modems", International Journal of Scientific & Engineering Research, Volume 4, Issue 8, August-2013.

[7] K. Goldberg, M. Mascha, S. Gentner, J. Rossman, N. Rothenberg, C. Sutter, and J. Wiegley, "Beyond the web: Manipulating the real world," Comput. Netw. ISDN Syst. J., Volume 28, Issue 1, pg. 209–219, December-1995.

[8] K. Goldberg, J. Santarromana, G. Bekey, S. Gentner, R. Morris, C. Sutter, and J. Wiegley, "Detecting Driver drowsiness based on sensors". Mach, Perception Newslett., Volume 5, Issue 1, March-1996.

[9] R. Marín, P. J. Sanz, P. Nebot, and R. Wirz, "A Multimodal Interface to Firebase: A Case Study on Remote Programming" ,IEEE Transactions on Industrial Electronics, Volume 52, December-2005

[10] T. Baudel, M. B. Lafon ,"Intelligent alcohol detection for car", Communications of the ACM, Volume 36, Issue 3, July-1993

Automated Public Transport Complaint Tweet Classification

Sara Khan, Priyanka Sharma,

University of Delhi, India

Abstract: Social media has facilitated a lot in reshaping communication industry and redefining the ways in which we communicate and express ourselves. Platforms like Twitter is embraced by millions of people in India wherein they not only share their thoughts but complaints too. Among other categories of complaints expressed, there is huge data-set of tweets related to public transport complaints. Twitter Handles like @RailMinIndia (Ministry of Railway), @sureshprabhu (Railway Minister), @AirIndia (Air India) are highly active and try to address complaint tweets manually. About thousands of tweets are sent in day and thus needs an automated solution to address and resolve them. Thereby, this paper is an attempt to provide a solution to automate task of public transport complaint classification using a two-stage classifier. First stage comprises of sentiment analysis of tweets which classifies them under positive, negative and neutral categories. Since the solution needs to address the areas having highest probability of complaints, in stage two, negative tweets are further sub-categorized into six different categories such as clean, late, infrastructure, refund, urgent and others with an accuracy of 82%. For data visualization, tweets are further quantified over a pie chart deployed on a user interface. This real-time automated system can be used by government departments which can take these segregated tweets and analysis to take further actions in respective areas.

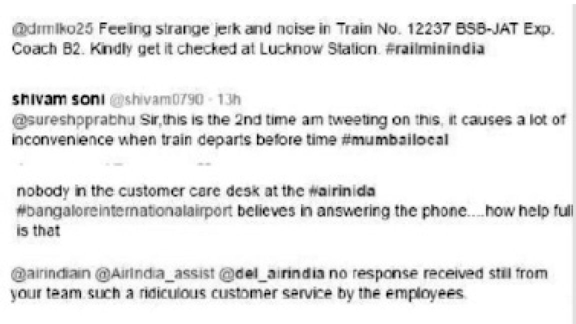
Keywords: *Public Transport Complaints, Government Application, Data Mining, Twitter, Text Classification*.

I. INTRODUCTION

The number of twitter users have increased 10-fold since 2010[1]. People are actively voicing their opinions through tweets. Gone are the days when people used to fill feedback forms to express their views about different domains and their issues. Now a days, in a click of a button, people use these social platforms to give their honest feedback. In formal sense, this casual data is useful for inference and observation as it is genuine and unaltered. Another reason for high volume of complaint tweets is the activeness of the official twitter handlers by various transport authorities like @RailMinIndia, @sureshprabhu and @airindia. Thus, there is a pool of abundant data on single issue that needs to be addressed. Although the transport in India is developing at a rapid pace, still there are a lot of problems. These large amounts of problems and many more which are not noticeable have made people suffer and have deteriorated the system functioning. Therefore, there arises an urgent need to minimize these problems as much as possible.

In order to address it, we initially analyzed the tweets on Twitter and concluded that people usually tweet complaints against railways and airways in India. Around 20 tweets were present for roadways for past 4-5 months during the analysis periods. Railways are one of the most targeted domains. Airways is the next public transport domain being flooded with complaints. We have used twitter handlers like @RailMinIndia, @sureshprabhu, @AirIndia and @del_airindia for our study because they actively cater to people's complaints which encourages them to post more and more in search of a solution. Hashtags like #MumbaiLocal are popularly used by Twitterati for highlighting their issue. It also makes task of the government easier.

This paper aims to help the government in identifying the major problems prevailing in railways and airways by developing a system that could provide them classified data containing generic and specific issues related to public transport. It tries to bridge the gap between the government and the citizens. Since there is a large amount of complaint tweets, it becomes difficult for the government to look at every tweet. But, at the same time, every citizen tweet with a hope that his/her concern will be addressed. Thus, by segregating issues into different classes can help government


@drmlko25 Feeling strange jerk and noise in Train No. 12237 BSB-JAT Exp. Coach B2. Kindly get it checked at Lucknow Station. #railminindia

shivam soni @shivam0790 · 13h
@sureshpprabhu Sir,this is the 2nd time am tweeting on this, it causes a lot of inconvenience when train departs before time #mumbailocal

nobody in the customer care desk at the #airinida #bangaloreinternationalairport believes in answering the phone....how help full is that

@airindiain @AirIndia_assist @del_airindia no response received still from your team such a ridiculous customer service by the employees.

Fig. 1. Sample Tweets

in utilizing their time on working on these problems rather than processing them. The real-time classifier takes the stream data from twitter using the above-mentioned twitter handles and keywords. In first stage, it finds the sentiment of tweet by labeling them as positive, negative or neutral. If the tweet is negative, it further categorizes them into six sub categories such as late problems, urgent, clean, infrastructure, refund and others.

II. LITERATURE REVIEW

This work [2] also caters to the complaints by Twitter users. The authors have built an ensemble classification model using one class SVM with various linear and non-linear kernels. They have considered the tweets in general and are categorizing tweets into appreciation, information sharing, complaints and promotion. The mentioned work has been an inspiration for the authors. Another work in [10] deals with the transport complaints on Twitter in Indonesia. The authors have used SVM to determine the positive and negative opinions of the users. This work [11] measures customer satisfaction of transport services in Indonesia by two private company service provides namely GO-JEK and Grab. The authors implement various models like SVM, Naïve Bayes and decision trees. As a result, they find GO-JEK has better satisfaction than Grab transport service,

Our work is different from previous work as we solely focus on the transport complaints in India unlike [2]. Also, we not only find the sentiment of the tweets but also categories them into different problems unlike [10, 11].

III. EXPERIMENTAL SETUP

After examining several Twitter accounts, we selected two popular hashtags - #MumbaiLocal, #delindia. Among many of Indian government accounts, we selected @RailMinIndia (Railway Ministry of India), @AirIndia (Air India) for the phase of data collection. We selected these handlers because they are highly active and results in bulk of information. We used official Twitter Streaming API [3] for fetching the tweets. We acquired data at three different points i.e. September 2016, February 2017 and April 2017. The overall tweets collected by running the program on specific days in the above-mentioned months were approximately 40,000. Different time spans were selected because we wanted to increase generality of data. For example, the sample acquired near Diwali etc. had a significant number of tweets referring to crowd related complaints. The main idea behind sampling at different timestamps was to remove bias and increase generality.

Figure 1 shows a sample of experimental dataset that is returned using the API. During the data collection phase, the number of tweets collected were @RailMinIndia-37097 and @sureshprabhu-21854. Since there is no language barrier on Twitter, people use different languages to share their opinions. They often share videos, URLs and images while expressing their views on different issues. However, we restrict our search by considering tweets in English. We also didn’t cover tweets that have non textual information like videos, images, video, URLs. Meta level information such as the type of tweet, URL, user mentions, hashtag, the general details of the user such as the username and geolocation (if available) are also extracted. As we aimed not only to acquire all the complaints and grievances of the people but also classifying them into different issues. Thus, after a lot of analysis and research we formulated frequently used words that are used by people which gives a direct insight of the underlying issues they face. Table 1 enlists some of the keywords which were used to sub-categorize each tweet into respective problems.

IV. DATA ENHANCEMENT AND ENRICHMENT

The tweets fetched using Streaming API [3] were in the free form text and contains a lot of noise. In order to filter out the tweets, we used an approach [5] which step by step filters out the noise (Figure 2).

Basic Filtration: This phase filters the tweet using following steps:

Escaping HTML characters: Raw data fetched using Streaming API contains irrelevant HTML entities and adds to noise. Thus, it is advisable to remove such kind of entities. For instance: &amp is converted &

Hashtags removal: Hashtags are important during the fetching phase as they help in finding relevant tweets for a specific issue. After the tweet collection, they don't really help in modelling. Thus, they can be removed. Example of hashtag: #mumbailocal

Decoding data: This allows transforming data into simple structures that is understandable. Various decoding options are available for text data such as "UTF-8" etc.

Table 1. Frequently occurring words which gives insight of underlying issues

Problems	**Keywords**
Late Problems	delay,late,cancel,delayed,detain,abandon
Urgent issues	train no.,pnr,urg,urgent,urgency
Cleanliness Issues	dirty,untidy,unhygienic,bad,poor,pathetic,toilet,stinking
Infrastructure issues	renovate,upgrade,repair,modernize,revamp
Refund issues	renovate,coach,halt,repay,return,refund
Other issues	water,security,abuse,misbehaviour,losses,overcharging

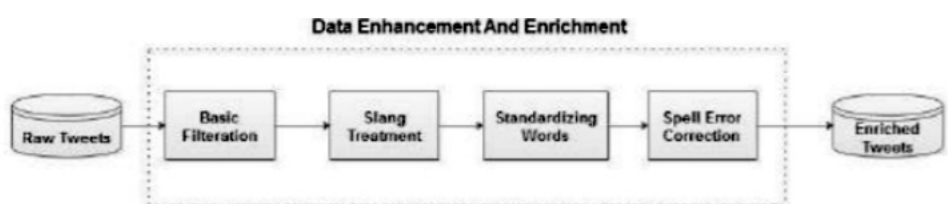


Fig. 2. Framework of Data Enhancement and Enrichment

Removal of URLs: As URLs and hyperlinks don't convey much information. So, they can be filtered.

Punctuation Filter: Punctuation, retweets and whitespaces removal as they don't add any additional support in achieving the desired target.

Same case: Conversion into lower case to make it presentable and easy to read.

Slang Treatment: Twitter [4] is a free social media platform that allows you to communicate with other peop<le through tweets, a Twitter post has a word limit of 280 characters. Before November 7, 2017, it was restricted to 140 characters. Due to this character limit, people often use SMS language or use abbreviations. Thus, replacing all the slang words is a bit challenging task. As with any social medium, there are some words and abbreviations used only by the users of Twitters, commonly known as "Twitterers". Thus, decoding this Twitter lingo into simple language can help in processing data into usable form. Using this approach, words like "plz" will be converted to "please" etc. Thus, we used an approach of lookup table of all possible keys to get rid of ambiguities and converting slangs to standard words.

Standardizing words: Most of the tweets written are informal and doesn't follow any grammatical paradigms. For example: "I hateeeeeee railways" should grammatically be "I hate railways". We used regular expressions and rule-based correction to solve it.

Spell Correction: Text data usually contains misspelled words. Big giants like Google and Microsoft invested in algorithms for an automated spell corrector and have achieved a decent accuracy level. Algorithm [6] having an accuracy 70-80\%, was able to correct words in the tweets. E.g. speling to spelling.

Table 2 shows some examples of tweets before and after processing

Table 2. Sample of Tweets before and after each phase of data enhancement and enrichment

	Before	After
Basic Filtration	@RailMinIndia@sureshpprabhu 64077 is late again today,as usual. I am writing about this for past 41 weeks but no improvement	64077 is late again today usual i am writing about this for past 41 weeks but no improvement
Slang treatment	Rly statns are dirty and untidy.pls wake up!	railwaystatns are dirty and untidy.pls wake up!
Standardizing Words	Sir Helpppppppppme. Ihav lost my wallet. pnr no 93824723	sir help me.Ihav lost my wallet. pnr no 93824723.
Spell Correction	Pathticcondtn of stns.tolets r stinking bdly.	pathetic condition of stations. Toilets are stinking badly.

V. TWO STAGE CLASSIFICATION SYSTEM

As all the Twitter handlers are open accounts, not all tweets posted to the above handles are negative. Thus, in order to classify complaints of the users, we develop a two-stage classifier. Based on our observation and analysis, we divide each tweet as positive, neutral and negative. Where positive tweets correspond to all those tweets which falls in the category of appreciation, or if complaint is being addressed by the corresponding department or any action is being taken by them. Negative tweets correspond to all the complaints and grievances posted by the user. And all the non- concerned tweets like replies from the twitter handles or news etc. fall in the category of neutral.

We further sub categorize negative tweets into six categories as late, urgent, cleanliness, refund, infrastructure and other.

Late issues: We observed that there are lots of complaints posted to the mentioned Twitter handlers regarding the problems faced by people due to dilatory of these public transports. For example*, ''regular delays in 12006 shatabdi express, the train can't stick to its schedule."*

Urgent Problems: These issues correspond to all those problems that people face and requires an immediate action from the concerned department. For example*, ''sir online ticket booking server not working even though we enter right user id and password it is showing invalid user id"*. It indicates an immediate action is required to be done by the IT department of Indian railways.

Cleanliness issues: We observed that most of the complaints correspond to the hygiene and cleanliness. People posts a lot about the unhygienic conditions of these public transports which are required to be addressed by the Indian government. Tweets like *''in ganeswary express 12101 ac first class toilet condition is pathetic please do something"* needs urgent attention from the government.

Refund issues: Sometimes people mention that they are still waiting for their refund of the payment made by them even after the cancellation of trains. These problems need proper addressing from the department.

Infrastructure problems: Among all the above-mentioned categories, people often mention about the infrastructure and the problems faced by them due to lack of up-gradation in this sector. Problems like improper toilet facilities and lack of poor conditions of berths still prevail in Indian railways. For example, tweet like *''sir really very poor condition superfast ac coaches no water no ac"* talks a lot about infrastructure issues in Indian railways.

Other problems: All the main issues are covered in above categories, the rest are clubbed into a common domain, others problems which consists of issues that are not in bulk but still needs to be addressed. For example, pantry problems, misbehavior, losses etc.

VI. EXPERIMENT AND RESULTS

After the pre-processing phase, the next step is building the model. We used supervised learning for our task of

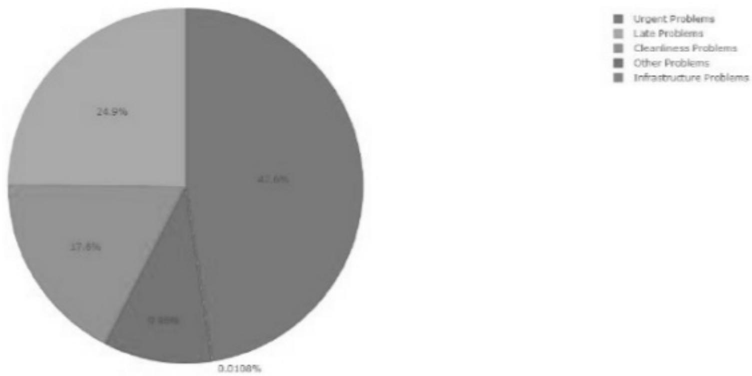


Fig. 3. Pie chart on web interface

classification. In order to train the classifier, we split the data into 70:30 ratio.

In Stage 1, we use human labelling to find the sentiment of the tweet and classify into ‘positive’, ’negative’ and ‘neutral’ for the training set.

In Stage 2, a semi-automated approach is used in which the most probable words that can be used to segregate the problems into different categories are stored in different files. The tweets in training set are categorized using the words in their respective files. As the list of words is quite long, a handful of them are shown in Table 1.

For example, if the tweet is *“Toilets are stinking”*. Then, the two words toilets and stinking are from the same category ‘cleanliness’. Table 3 demonstrates the above example where words ‘toilets’ and ‘stinking’ increases the count of clean sub-category to 2 whereas the remaining sub-categories are still at 0. So, the baseline approach will categorize it as clean sub-category.

Once the training dataset is ready, the feature vector is created. Feature is an aspect of the text that is relevant to the task. It is one of the most important phase priors implementing the classifier. Not all features are important and a good feature results in higher accuracy. The feature vector (comprising of all the features) is fed to build the model for training it. Similarly, in this case, the presence or absence of word in a tweet can be sued as a feature. Each sentence or tweet can be broken into words and each word is added as a feature in feature vector. For single words, this is referred as “unigram” approach. Also, irrelevant words can be filtered out. Stop words such as a, is, the, with etc. These words don’t indicate any sentiment and can be removed. For grammatical reasons, documents are going to use different forms of a word, such as organize, organizes, and organizing. The goal of lemmatization [7] is to reduce inflectional forms and sometimes derivationally relates forms of a word to a common base form. After the feature vector phase, the model is ready to train and learn. The trained model is used to classify test tweets. We train data using Naive Bayes [8]. Naive Bayes is a probabilistic model that is used for classification task. It uses Bayes’ theorem and has an assumption of Independence among all features or predictors. In layman’s term, it doesn’t consider dependence of one particular feature to another feature.

The classification model takes two parameters. The training dataset containing probabilities of each feature vector calculated using Bayes’ theorem and the feature vector of testing dataset. As a result, the classifier labels the testing tweets into positive, negative or neutral. If the tweets are negative, the same approach is applied. But now the training dataset having 6 different sub-categories is used.\par

After classifying the tweets into different categories, they are stored in CSV file. Simultaneously, the count of different sub-problems is calculated which is stored in a global file descriptor. The quantification of different negative tweets into sub categories ‘late’, ‘urgent’, ‘cleanliness’, ‘infrastructure’, ‘refund’ and ‘other’ issues is deployed in the form of a pie chart on a web interface (Figure 3). The work in [8] also helped during the implementation. The total tweets collected so far are 37,097. Out of all, the number of tweets labelled negative is highest followed by neutral and positive respectively. As far as second stage of classifier is concerned, urgent problems form the highest proportionate of complaints followed by other problems (Figure 4). The accuracy on test data was approximately 82% using Naïve Bayes.

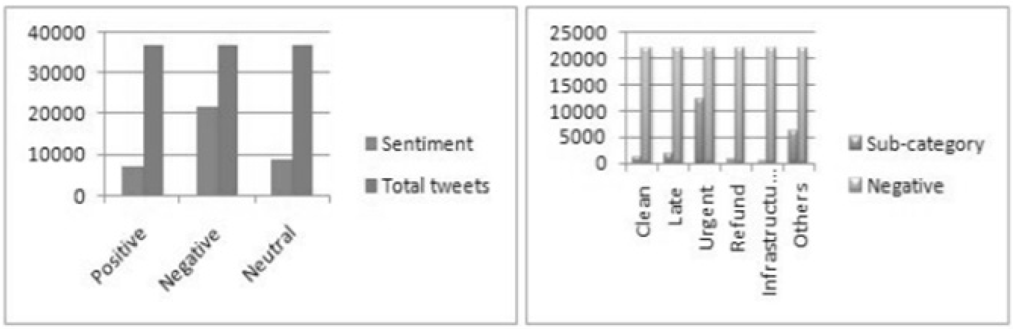


Fig. 4. Classification into different categories

VII. CONCLUSION AND FUTURE WORK

Nowadays, Twitter has been used as a platform to post about complaints and grievances of citizens and to bring government's attention towards these issues. The government has also shown equal interest to address these issues which has motivated people to post. Thus, our paper tries to make both the ends meet as government cannot look after every person but for every person his or her own request is equally important. In order to achieve this, our research approach looks very promising. We can successfully preprocess the raw data by removing slangs through text processing, converting SMS language into a formal one and finding words through manual inspection that acts as a segregator. By using machine learning algorithm, we can classify tweets into different categories. This classification can help the government in resolving the issues by sending tweets to the respective departments.

REFERENCES

[1] Statistics from, https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

[2] Mittal N., Agarwal S., Sureka A. (2016) Got a Complaint? - Keep Calm and Tweet It, *Advanced Data Mining and Applications. ADMA 2016. Lecture Notes in Computer Science, vol 10086*.

[3] Streaming API from, https://dev.twitter.com/streaming/overview

[4] Twitter from, https://en.wikipedia.org/wiki/Twitter

[5] Twitter Sentiment Analyzer from, https://www.ravikiranj.net/posts/2012/code/how-build-twitter-sentiment-analyzer/

[6] Spelling Corrector from, http://norvig.com/spell-correct.htmL

[7] Wordnet Lemmatizer from, http://www.nltk.org/modules/nltk/stem/wordnet.html

[8] Naïve Bayes Classifier from, http://www.nltk.org/modules/nltk/classify/naivebayes.html

[9] Laurent Luke's Blog from, http://www.laurentluce.com/posts/twitter-sentiment-analysis-using-python-and-nltk/

[10] Effendy V., Novantirani A., Sabariah M., (2016) Sentiment Analysis on Twitter about the Use of City Public Transportation Using Support Vector Machine Method, International *Journal on ICT, vol 2, pp. 57-66*

[11] S. Anastasia and I. Budi, "Twitter sentiment analysis of online transportation service providers," 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Malang, 2016, pp. 359-365.

Skin Lesion Segmentation using SegNet with Binary Cross- Entropy

Prashant Brahmbhatt[1], Siddhi Nath Rajan[2]

IMS Engineering College, Ghaziabad, India

Abstract: In this paper a simple and computationally efficient approach as per the complexity has been presented for Automatic Skin Lesion Segmentation using a Deep Learning architecture called SegNet including some additional specifications for the improvisation of the results. The secondary objective is to keep the pre/post -processing of the images minimal. The presented model is trained on limited images from the PH2 dataset which includes dermoscopic images, manually segmented. It also contains their masks, the clinical diagnosis and the identification of several dermoscopic structures, performed by professional dermatologists. The aim is to achieve a performance threshold Jaccard Index (IOU) 92% after evaluation.

Keywords: Skin Lesion Segmentation, Dermoscopic, Jaccard Index

I. INTRODUCTION

As per the figures provided by the American Cancer Society, only 1% of all the cancers are diagnosed as melanoma cancers but they cause most of the skin cancer deaths. The non-melanoma cancers also cause of a large number of deaths. For the year 2019 the death estimation is 7,230 (4,740 men and 2,490 women) from melanoma [1]. As per the reports of World Health Organization around the world 3 million non-melanoma skin cancers and 1,32,000 melanoma skin cancers are recorded every year [2]. After the initial surgery most people are cured and the life expectancy is greatly increased if the cancer is timely diagnosed. Therefore, the proper diagnosis of the lesion is required for the treatment of a patient. To classify the lesion as melanoma or non-melanoma the images require to be segmented which creates a mask used for cropping the lesion then further identify the features from.

The traditional approach of manual segmentation by the expert dermatologists has been already proposed to be replaced with computerized segmentation techniques in which machine assisted methods are used. The previous manual approach was time-consuming, complex and dependent on the observer and his capabilities. The efficient approach of segmentation can save a lot of time and expertise required to complete the task with sufficient accuracy. The already proposed architectures involve some state-of- the-art neural networks like FCN [3], VGG [4], Fast-RCNN [5] and U-Net [6].

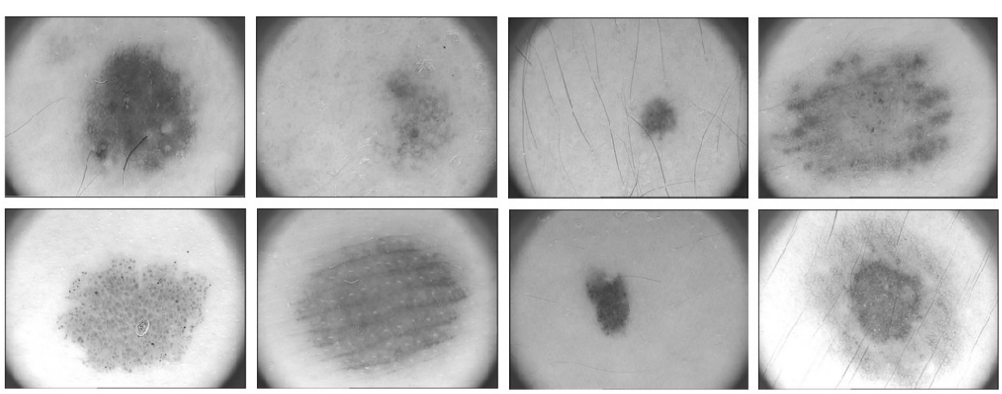


Fig. 1. Sample microscopic images of the skin lesions found on the human skin.

1. The Model

SegNet [7] is a deep neural network based on the encoder-decoder architecture for pixel-wise segmentation, originally proposed and implemented by members of the Computer Vision and Robotics Group at the University of Cambridge, UK [8]. SegNet is based on the principle of Semantic Pixel-Wise Labeling initially used for the segmentation of street images which include a total of twelve different classes. Each pixel in the image is to be classified among one of the twelve classes as per the data.

The architecture of SegNet has non-linear layer encoding sequences and decoders corresponding to each layer. There is a final classifier present for the pixel-wise classification. We have used it as the base architecture with varying hyper parameters and structural differences for our lesion segmentation problem.

2. Network Architecture

The proposed is a 64-layered network excluding the final activation layer. Every sequence of encoder has multiple convolutional layers, batch normalized with ReLU non-linearity which is followed by non-overlapping maxpooling and sub-sampling. At the center of the network there are two dense layers present before the first up-sampling begins.

The defining characteristic of SegNet is the use of max-pooling indices in the decoders to perform up-sampling of low-resolution feature maps. This leads to retaining of the important detailed features in the image and non-useful features are dropped. SegNet provides smooth images without any post-processing technique involved.

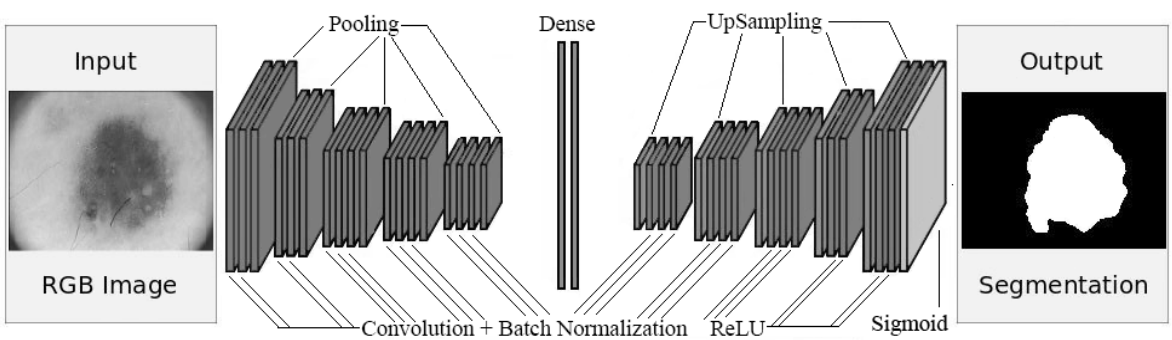


Fig. 2. The proposed network architecture shows all the layers of the network

Table 1. The detailed network architecture with output dimensions at each sequence of layers

Sequences	Filter	Output Dim.	Sequences	Filter	Output Dim.
Input		192 x 256 x 3	u-1	(3,3)	24 x 32 x 512
conv-1	(3,3)	192 x 256 x 64	deconv-1		
bn-1			bn-14		
conv-2	(3,3)	96 x 128 x 128	deconv-2		
bn-2			bn-15		
p-1			deconv-3	(3,3)	24 x 32 x 512
			bn-16		
conv-3	(3,3)	96 x 128 x 128			
bn-3			u-2	(3,3)	24 x 32 x 156
conv-4	(3,3)	48 x 64 x 256	deconv-4		
bn-4			bn-17		
p-2			deconv-5		
			bn-18		

Table 1 (Contd.)...

...Table 1 (Contd.)

conv-5	(3,3)	48 x 64 x 256	deconv-6	(3,3)	48 x 64 x 256
bn-5			bn-19		
conv-6					
bn-6			u-3	(3,3)	48 x 64 x 256
conv-7	(3,3)	24 x 32 x 512	deconv-7		
bn-7			bn-20		
p-3			deconv-8		
			bn-21		
conv-8	(3,3)	12 x 16 x 512	deconv-9	(3,3)	96 x 128 x 128
bn-8			bn-22		
conv-9					
bn-9			u-4	(3,3)	96 x 128 x 64
conv-10			deconv-10		
bn-10			bn-23		
p-4			deconv-11	(3,3)	96 x 128 x 64
			bn-24		
conv-11	(3,3)	6 x 8 x 512			
bn-11			u-5	(3,3)	192 x 256 x 64
conv-12			deconv-12		
bn-12			bn-24		
conv-13			deconv-13	(3,3)	192 x 256 x 1
bn-13			bn-25		
p-5					
Dense-1		6 x 8 x 1024			
Dense-2		12 x 16 x 1024	Output		192 x 256

3. Loss Function

The loss function used here is the binary cross-entropy. The cross-entropy is a function which measures how far away from the true value the prediction is for each of the classes and then averages the errors class wise to obtain the final loss.

$$L(y,\hat{y}) = -\frac{1}{N}\sum_{i=0}^{N}(y * \log(\hat{y_i}) + (1-y) * (1-\hat{y_i}))$$

In this problem, there lies only two classes for each pixel, either black or white (0 or 1) as per the mask. So, here binary cross-entropy is used as the loss function rather than the categorical cross-entropy originally proposed.

The binary cross-entropy is in the below form:

4. Training

For the training procedure has been carried out on 75% of the images available out of the 200 images in the PH2 dataset [9]. Although, the actual number of images will be more than 150 because of the new transformed images that will be added after image augmentation process. As per the architecture of the network the total parameters to be trained are 33,377,795 out of 33,393,669 whereas the non-trainable parameters are 15,874. The implementations are in Keras and the environment used is the IPython notebook provided by Kaggle [10].

5. Image Augmentation

The procedure of image augmentation on training images has been used for increasing the robustness of the model and reducing the chances of overfitting. It will also increase the data images that are available in the dataset. The two simple techniques that used are image rotation and horizontal flipping [11]. In the image rotation all the images in the training set are rotated with a range [-40◦, +40◦] [3] and flipped along the horizontal axis only.

All the above transformations are exactly performed on the corresponding masks of the images as well to maintain the correct orientation of feature images with their truth masks. After the augmentation the transformed images are included in the training set which increases our training set from 150 to 450. Out of these 450 images, 90 have been excluded from training set for the formation of a validation set.

6. Optimizer

The employed optimizer is the SGD (Stochastic Gradient Descent) for the network. The learning rate which is an important hyper parameter in the optimization is set to 0.001 which is one of the many values generally used for learning rate parameter.

The momentum is also used which is an approach that provides an update rule motivated from the physical perspective of optimization. The advantages of using momentum with SGD are that a small change results in large speed up in learning process. Analogically, the velocities are stored for all the parameters, and used for making the updates. The value of momentum used in for optimization is 0.9.

7. Batch Normalization

Batch Normalization [12] is the technique of speeding up the learning process of the neural network by normalizing the values in the hidden layers similar to the principle behind the normalization of the features in the data or activation values. In proposed network the batch normalization layer is present after every convolution layer with a total of 25 batch normalization layers in the entire network architecture.

II. EXPERIMENTAL DESIGN AND RESULTS

1. Databases

The used dataset is the PH2 dermoscopic dataset which contains 200 dermoscopic images and their label masks. Each one is an RGB image and the fixed dimension of each image is 572 x 765. The dataset has been provided publicly for experimental and studying purposes, to facilitate research on both segmentation and classification algorithms of dermoscopic images. The database is acquired at the Dermatology Service of Hospital Pedro Hispano, Matosinhos, Portugal [9].

For the training purpose the dimensions of each image have been reduced to 192 x 256 before feeding it into the network. It largely reduces the parameters to be trained in the network as well as the training time and complexity without significantly affecting the results.

2. Performance Evaluation

The generated binary mask images in the output of the network are evaluated on different mathematical measures in comparison with the ground truth lesion masks as provided in the dataset. The accuracy is measured pixel-wise. The used measures are as below:

- Intersection Over Union: The Jaccard index, also known as Intersection over Union. The Jaccard similarity coefficient is a statistical similarity measure to check the diversity among the sample sets. The IOU gives the similarity among sets and the formula is the size of the intersection over the size of the union of the sets.

$$(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

- Dice Coefficient: The Dice score is like precision. It measures the positives as well as it applies penalty to the false positives given by the model. It is more similar to precision than accuracy.

$$Dice = \frac{2 \times TP}{(TP + FP) + (TP + FP)}$$

- Precision: Precision is a measure which is more focused towards catching the false positives in the results of the model.

$$Precision = \frac{TP}{TP + FP}$$

$$TP$$

- Recall: Recall is a measure which is targeted towards the actual or the true positives yielded by the model output. In the scenarios where the cost of the False Negatives is greater than recall is the better metric to choose the best model among the possible ones.

$$Recall = \frac{TP}{TP + FN}$$

- Accuracy:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

3. Key Component Validation

The network after training on the set of 360 images and validating on the 90 images produced the results that are observed after 100 epochs.

The training curves of the network corresponding to the training set as well as the validation set are also plotted. The curves include the loss curve and the accuracy curve with respect to the epochs along the horizontal axis. Initially for the training set the loss is above 0.735 which gradually declined and reached 0.115. For the validation set it began from 0.707 and ended up at 0.1595. The accuracy for the training set increased from 0.500 to 0.978 and that of the validation set from 0.312 to 0.955.

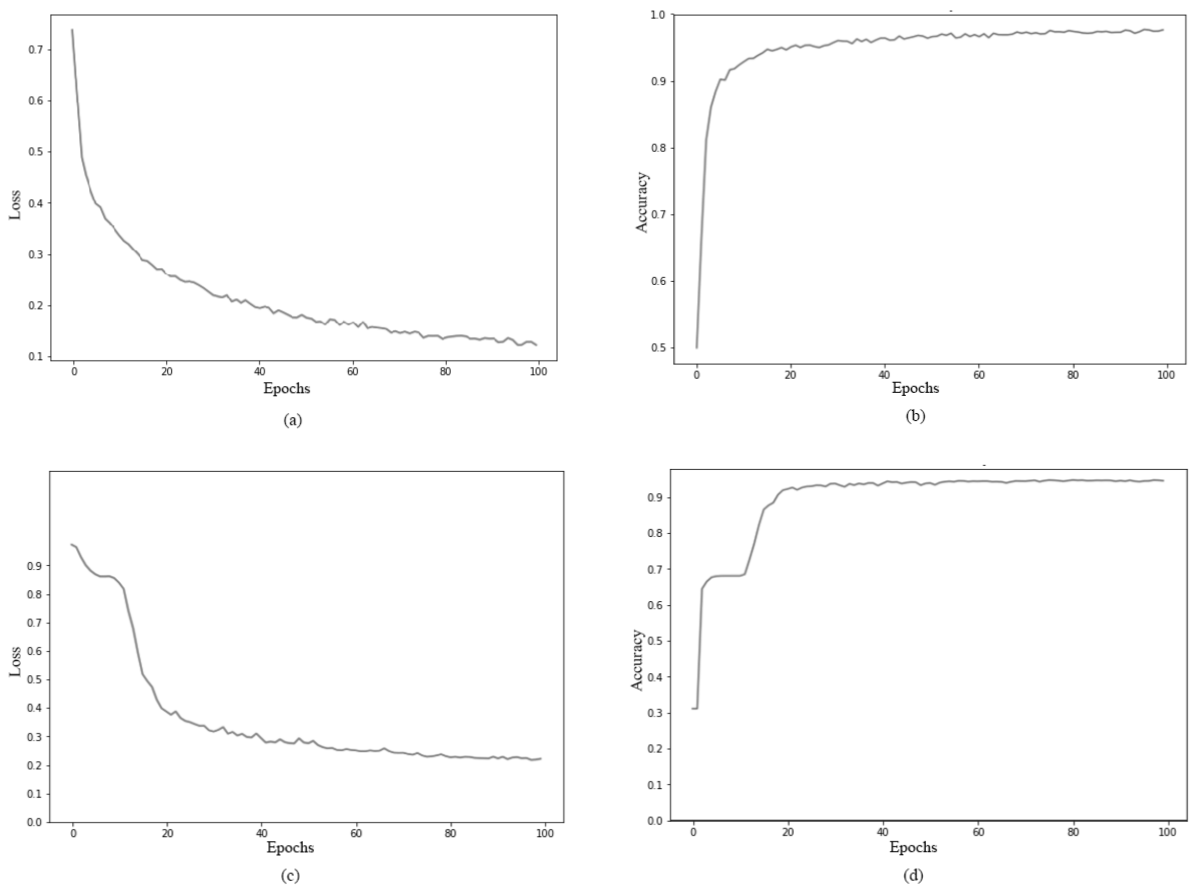


Fig. 3. The training and validation curves (a) Loss curve on training set; (b) Accuracy curve on training set; (c) Loss curve on validation set; (d) Accuracy curve on validation set

Table 2. Performance statistics after training for 1 epoch

Dataset	JA	DI	PR	RE	AC	Loss
Test	67.71	38.35	6	1.5	60.86	69.21
Validation	67.66	38.17	7.47	1.98	61.41	69.21
Training	67.86	38.72	6.68	1.65	60.2	69.23

Table 3. Performance statistics after training for 100 epochs

Dataset	JA	DI	PR	RE	AC	Loss
Test	93.61	80.37	88.99	92.13	93.96	18.75
Validation	95.09	82.1	91.84	94.19	95.53	15.95
Training	97.02	85.16	95.97	97.32	97.87	11.5

4. Prediction on PH2 Database

The final part is to save the trained model and make predictions. The predictions are compared visually to the actual ground truth lesion mask images. The predicted outputs are initially slightly blurry at the edges and do not give a precise prediction towards the boundaries however, it still performs considerably well. To have a clear prediction at the boundary thresholding of the pixel values is performed as a simple post-processing technique.

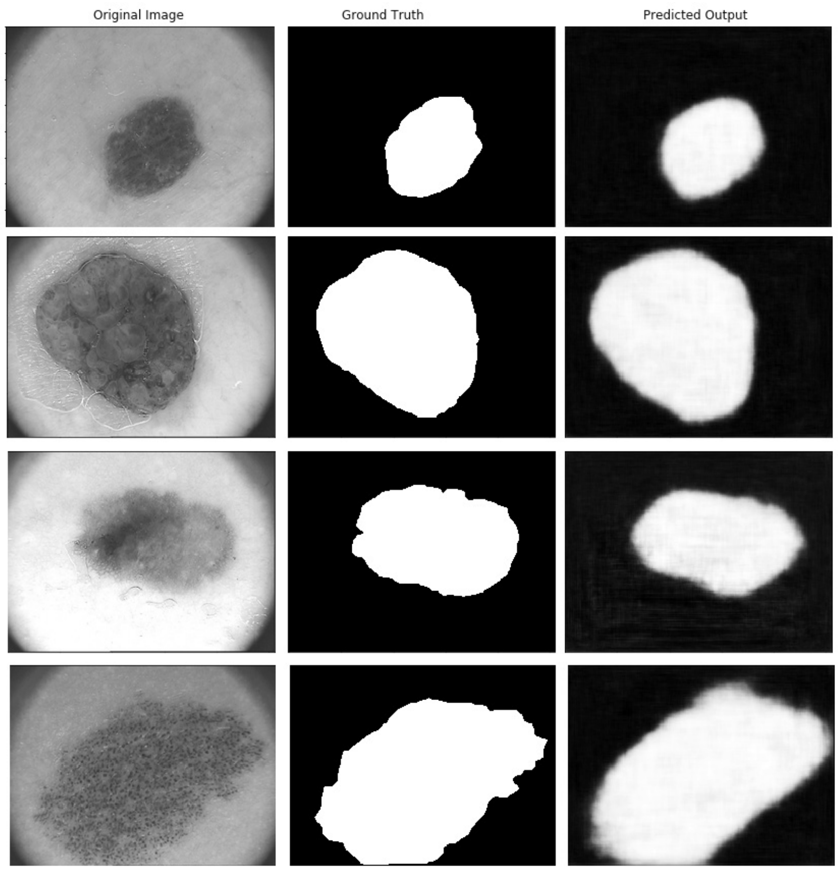


Fig. 4. The comparison of the predictions with original feature images and the ground truths

The predicted images have every pixel value lying in the range [0,1]. To get the clear boundaries the pixel values can be converted to belong in either the black or white class (0 or 1) based on a decided threshold which eliminates any pixel values in the gray shade. The threshold that we have used is 0.2, every pixel having a prediction value of greater than 0.2 will be rounded up to 1 and vice-versa. After performing this technique, the sharp boundaries around lesion in the predicted masks can be observed.

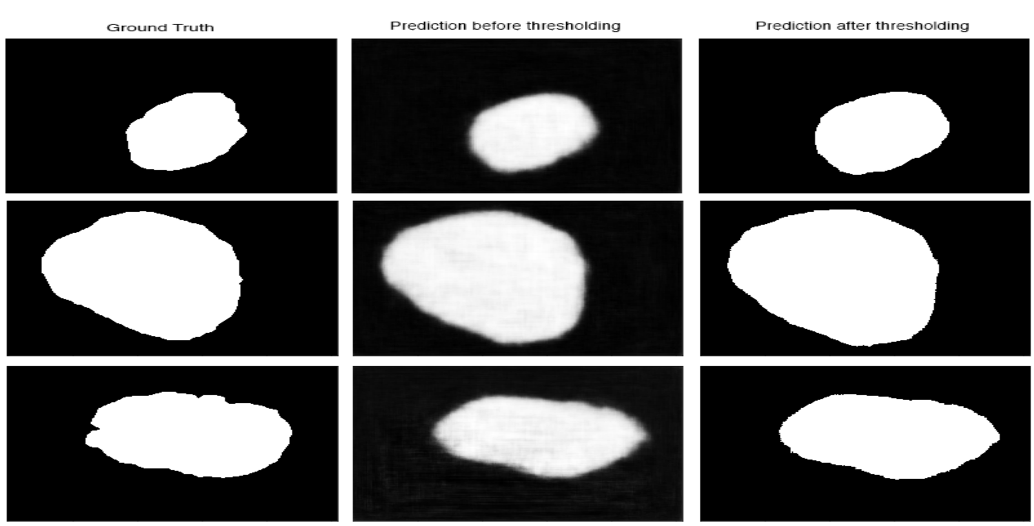


Fig. 5. The comparison of ground truths and original predictions with processed predictions

III. DISCUSSION

The proposed approach and specifications in this paper are only applied to the PH2 dataset and could be used on the ISBI 2016 [13] and ISIC 2017 [14] dataset to fairly compare it with the current state-of-the-art architectures. The advanced approaches of image augmentation such as varying the brightness and contrast or using affine, random crops [11] and levelsets [15] could improve the robustness but it would increase the computational complexity for much larger datasets like ISIC 2017. The larger datasets will significantly increase the training time as they may require more than 100 epochs. The proposed network in this paper took 9.25 seconds per step over 100 epochs resulting in the total duration of 15.416 minutes.

The thresholding value used is chosen merely based on the experimental results and has no theoretical principle behind the exact value. The different image sizes could also be experimented which will increase the time of the training by a large amount but could be needed for the larger datasets with more than 2000 images as cited previously.

IV. CONCLUSION

In this paper the authors proposed the use of the SegNet architecture for solving the skin lesion segmentation and successfully provided with the results of the experiment on the PH2 dataset with sufficient accuracy.

Two techniques of image augmentation, image rotation and horizontal flipping on the training dataset are performed before feeding it to the network for training. After the training process the model was evaluated on several measures for statistical values. The predictions produced from the model on test images were post-processed using the thresholding technique to remove the blurry boundaries around the predicted lesions.

REFERENCES

[1] "Melanoma: Statistics | Cancer.net," American Society of Clinical Oncology (ASCO), 2019. [Online]. Available: American Society of Clinical Oncology (ASCO). [Accessed July 2019].

[2] WHO, "WHO | Skin Cancer," World Health Organisation, 2019. [Online]. Available: https://www.who.int/uv/faq/skincancer/en/index1.html. [Accessed July 2019].

[3] Y. Yuan, M. Chao, Y.C. Lo, "Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks With Jaccard Distance," IEEE Transactions on Medical Imaging, 2017.

[4] Lopez, A. Romero and Giro-i-Nieto, Xavier and Burdick, Jack and Marques, Oge, "Skin lesion classification from dermoscopic images using deep learning techniques," in Biomedical Engineering (BioMed), 2017 13th IASTED International Conference on, IEEE, 2017, pp. 49--54.

[5] R. Girshick, "Fast R-CNN," in 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015.

[6] Md. Z.Alom, M. Hasan, C. Yakopcic, T. M. Taha, V.K. Asari, "Recurrent Residual Convolutional Neural," Cornell University, 29 May 2018. [Online]. Available: https://arxiv.org/abs/1802.06955v5. [Accessed July 2019].

[7] A.Kendall, V. Badrinarayanan, R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.

[8] A.Kendall, V. Badrinarayanan, R. Cipolla, "SegNet," University of Cambridge, 2015. [Online]. Available: http://mi.eng.cam.ac.uk/projects/segnet/. [Accessed July 2019].

[9] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, J. Rozeira, "PH2 - A dermoscopic image database for research and benchmarking," 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 5437-5440, 2013.

[10] B. Hammer, A. Goldbloom, "Kaggle," Alphabet Inc., April 2010. [Online]. Available: https://www.kaggle.com/. [Accessed October 2019].

[11] F. Perez, C. Vasconcelos, S. Avila, E. Valle, "Data Augmentation for Skin Lesion Analysis," Cornell University, 05 September 2018. [Online]. Available: https://arxiv.org/abs/1809.01442. [Accessed July 2019].

[12] S. Ioffe, C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," 11 February 2015. [Online]. Available: https://arxiv.org/abs/1502.03167. [Accessed July 2019].

[13] D. Gutman, N. C. F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, A. Halpern, "Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC)," May 2016. [Online]. Available: https://arxiv.org/abs/1605.01397.

[14] D. Gutman, N. C. F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, A. Halpern, "Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC)," 2017. [Online]. Available: https://challenge.kitware.com/#challenge/583f126bcad3a51cc66c8d9a.

[15] K. H. Cha, L. Hadjiiski, R. K. Samala, H. P. Chan, E. M. Caoili, R. H. Cohan, "Urinary Bladder Segmentation," Med Phys, vol. 43, no. 4, p. 1882–1896., 2016.

Feature Selection in Cardiovascular Diseases Parameters using Machine Learning and Boruta Algorithm

Ashish Mishra [1], Raghavendra Kumar [2], Himanshu Rathore[3]

[1, 3]UG-Information Technology, KIET Group of Institutions, Ghaziabad, India
[2]Department of Information Technology, KIET Group of Institutions, Ghaziabad, India

Abstract: Machine learning techniques have various algorithms fits in multidisciplinary domain. These algorithms bring widely accepted results in healthcare especially in Cardiovascular diseases. This paper compare and analyze the four well established learning techniques using Artificial Neural Network (ANN), Random Forest (RF), XGBoost and Support Vector Machine (SVM) to predict Cardiovascular diseases in a patient. Accuracy, F1 score, recall and precision are considered as performance evaluators and results are compared with a benchmark paper. Feature selection of Cardiovascular diseases parameters is one of the challenging task. Boruta is one of the most widely accepted algorithms for feature selection. In this work selection parameter are set to be minimal and comply to fetch best result in Cardiovascular diseases prediction.

Keywords: Artificial Neural Network (ANN), Boruta, Random Forest (RF), Support Vector Machine (SVM)

I. INTRODUCTION

In health care domain Cardiovascular disease is a major challenge as death cause around the world. People with cardiac illness or those at elevated cardiac danger need early detection and management using counseling and medications. Various algorithms were tried, including KNN, Random Forest, Decision Trees, SVM, XGBoost, Cat-Boost, LightGBM and artificial neural networks. All of these approaches were tried since the most effective model was to be found. Least amount of log loss was needed as the Stacking and ensemble modeling was eventually used to find the best models possible. SVM, XGBoost, Random Forest algorithms, these four methods form an ensemble model where averaging the prediction probabilities out and then passing that to the log loss equation in Fig 1. An ensemble model like Das et al. (3) used but other than that we are also using algorithms such as XGBoost and ANN which haven't really been used in this way in ensemble with SVM so that makes our approach unique. Also, the metric that we used for evaluation is Log Loss, which heavily penalizes large deviations in incorrect predictions. This leads to our model being different than a lot of approaches tried beforehand. Making an ensemble is the major step which led to the most improvement that we have had till now in our experiments. Ensemble model composition can be seen as SVM, XGBoost, Random Forest algorithms as machine learning and ANN is used in ensemble using the averaging method to get a model.

II. RELATED WORK

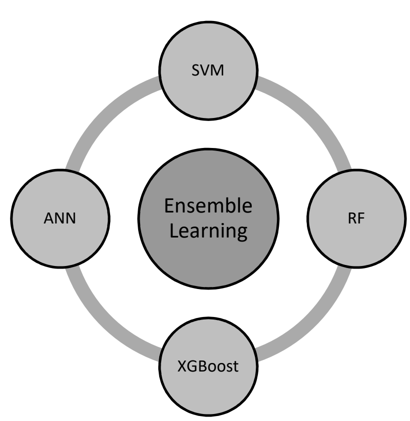


Fig. 1. Ensemble Model

David Chen et al. predict the risk factor of preoperative bleeding for patient undergoing colorectal surgery. Authors applies logistic regression and gradient boosting classifier with Boruta feature selection on the data of 13299 surgeries and select the 43 important features from the 117 features. They compare the variable importance with Logistic Regression (0.812) and gradient boosting classification (0.734) (1). K. Polaraju et. al (2), proposed the usage of Multiple Regression Models to predict Heart Diseases. The dataset used, the training data, was made up of 3000 cases. It had 13 characteristics, which were used as the grounds for model training. Training information accounted for 70% of the entire dataset, while the remainder, i.e., 30%, were sent to the test information. Results analysis leads us to think that, while predicting cardiovascular diseases, regression is better than most other models. Das et. al (3) employed the services of an ensemble method consisting of neural networks for diagnosing of the heart disease. It merged the posterior probabilities of various predecessor models or the expected values. The classification accuracy obtained by them using this ensemble model was 89.01%. The dataset that they used for training and testing was the Cleveland Heart Database. Anbarasi et. al (4) used almost a 2-step approach where first they used 3 classifiers. These classifiers that they tried are decision trees, Naïve Bayes and classification by clustering using thirteen attributes. In the next step, they used genetic algorithm to apply feature sub selection and acquired nearly comparable precision. Their findings showed that decision tree classifier performed the best at 99.2 percent for binary classification. The Naïve Bayes classifier followed with the result of 96.5% and then the classification cluster with 88.3%. Zhang et. al (5) used SVM (Support Vector Machines) as the model for cardiovascular disease prediction. They employed PCA to find the relevant and important features and then used various kernels such as rbf and linear to work on the aforementioned features. Radial Basis Function (RBF) gave the highest classification accuracy. To discover the appropriate parameter values, the grid search method was used. In binary classification, the classification precision was 88.6364 percent. Noura Ajam (6) used artificial neural networks. Feed-forward Back propagation algorithms were used to train the model based on the specified requirements. On finding the appropriate parameters, classification accuracy reached to 88%. The number of neurons that were employed in the hidden layer were 20. ANN shows good results as it is a suitable approach to use when the amount of data is large, as in large datasets. Elshazly et. al (7) proposed a novel approach for classification, GASVM, for diagnosis of lymph diseases. The GA part of it is just there to make the feature pool reduced. Cross validation technique used was k-fold. Different kernel features were used and performance measurements such as precision, sensitivity, area under curve (AUC), F-measurement were assessed for each feature. Final classifier that was found to be giving the most optimal results was linear, with a score of 83.1%. Dey et. al (8) used multiple models including SVM, Naive Bayes and Decision tree. They also experimented with feature selection and worked using it and without using it as well. The method used here was principal component analysis. The dataset is a binary classification problem. Final observations indicated that SVM outperformed the other two and was the best choice for the classifier. Weng et. al (9) used a similar approach to Dey et. al. where they compared various models such as random forest, gradient boosting, neural networks and logistic regression. These 4 machine learning algorithms were used to try and predict cardiovascular diseases. Grid search was used for parameter optimization. PCA was used as well and the best model based on all these parameters was found. SS Kumar et al. evaluate the performance of feature selection approaches with random forest classifier On the Cleveland dataset. Author applies four feature selection techniques such as relief feature selection, random forest selector, recursive feature elimination and Boruta feature selection and achieve the accuracy 98% by selecting 9 features with the Boruta feature selection technique (12). Zhi-Cheng Li et al. performs the retrospective study by building two radiomics model using random forest. One model is based on all relevant features and other based on minimum redundancy maximum relevance (mRMR) features. They were applied Boruta for feature selection and select eight relevant features from the corticomedullary phase image and achieve the accuracy 92.9% on validation cohort (13). Sanchez-Pinto et al. compare performance of classic and modern variable selection technique on clinical dataset. Authors examine the different feature selection method such as stepwise backward selection using p-value and Least Absolute Shrinkage and Selection Operator, and Elastic Net and tree based Variable Selection Using Gradient Boosted Feature Selection, Random Forest, Regularized Random Forests, and Boruta on different size of clinical datasets. They conclude modern tree base feature selection technique is best for large multicentre clinical data. Boruta gives the best result on the adult clinical deterioration cohort (14). Silvia Panicacci et al. identify heart failure patient by exploring the machine learning algorithms. Authors apply random forest and lasso machine learning algorithms with Boruta feature selection technique. By Boruta select 572 features out of 1514 feature from the Tuscan administrative data and give the result in terms of f1 and f2 score for both machine learning algorithm. Random forest f1 and f2 score is 64.23 and 53.49, lasso f1 and f2 score is 34.73 and 27.32 respectively (15).

III. METHODOLOGY

The idea behind a support vector machine is that we have some data points that we want to label either A or B, and we can do that by plotting a line, plane or hyper plane between the two classes A and B, depending upon the problem.

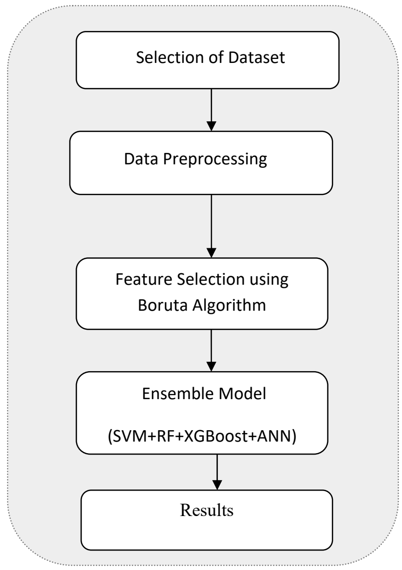


Fig. 2. Flow chart for prediction model

What an SVM does is figure out the "best" hyperplane to divide the points, which is the one that leaves the biggest margin between itself and the points. That line is called the maximum margin hyperplane, because it is a hyperplane that separates the sides while leaving the maximum margin. On trying the various kernels, RBF was found to be the most accurate one, and the one that gave the least log loss. XGBoost stands for eXtreme Gradient Boosting. XGBoost is an optimized library for boosting distributed gradients. It is intended to be flexible and extremely effective. It uses algorithms for machine learning under the framework for gradient boosting. Boruta algorithm is based on the concept of random forest classifier that collects result of randomized sample driven from ensemble model. A random sampling reduces the impact of randomness fluctuations and correlations factors in dataset [16].

IV. RESULT ANALYSIS AND DISCUSSION

Cleveland database is considered as training and testing dataset to predict cardiovascular decease in patient. It is popular multivariate dataset in research community used for classification in machine learning approach (11). The dataset used was the dataset given over at the competition on predicting heart diseases using machine learning at driven data. The dataset has 13 essential features, and 180 instances. Patient id is a feature as well but it must be dropped. There features are not dependent on each other shown in table-I. The summary of selected parameters is illustrated in table-II. The metric used for the competition we took part in is logarithmic loss.

$$\text{Log Loss} = \sum_{i=1}^{n} [\text{yi} \log(\text{y`i}) + (1 - \text{yi})\log(1 - \text{y`i})] \tag{1}$$

y` is considered as probability that y=1. The goal is to reduce log loss. Log loss provides a penalty for too deviant of predictions. The testing procedure involved us passing the data through the model, noting down the accuracy and log loss and repeating it for the various models. We began our experimentations by working with the dataset, cleaning it up, one-hot encoding the categorical variables, normalizing with standard scalar and splitting it into 70% training and 30% testing dataset. Following that began experimentations with SVM and Neural networks at first. They were tried separately, with SVM giving the least log loss among the two at first, optimized using grid search. After optimizing the hyper parameters of NN, we got it close to SVM but it could never cross it. So we decided to go with SVM. After that we tried various algorithms, random forest classifier, XGBoost. After that an ensemble of various algorithms was tried, at the end of which the combination of SVM, XGBoost, Random Forest algorithms and Artificial Neural Network (ANN). To ensure the correct form of data we adopted to go for regression based imputation using EM algorithm that works between missing data and available data.

Table 1. Training Dataset After Feature Selection

	Age	**Sex**	**cp**	**trestbps**	**Chol**	**fbs**	**restecg**	**thalach**	**exang**	**oldpeak**	**slope**	**ca**	**thal**	**class**
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	absence
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	presence
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	presence
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	absence
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	absence
5	56	1	2	120	236	0	0	178	0	0.8	1	0	3	absence
6	62	0	4	140	268	0	2	160	0	3.6	3	2	3	presence
7	57	0	4	120	354	0	0	163	1	0.6	1	0	3	absence
8	63	1	4	130	254	0	2	147	0	1.4	2	1	7	presence
9	53	1	4	140	203	1	2	155	1	3.1	3	0	7	presence

Table 2. Summary of Dataset

	Age	**Sex**	**cp**	**trestbps**	**chol**	**fbs**	**restecg**	**thalach**	**exang**	**oldpeak**	**slope**	**ca**	**thal**
count	303	303	303	303	303	303	303	303	303	303	303	303	303
mean	54.4389	0.67987	3.15842	131.69	246.693	0.14851	0.9901	149.607	0.32673	1.0396	1.60066	0.67221	4.73419
std	9.03866	0.4673	0.96013	17.5997	51.7769	0.3562	0.99497	22.875	0.46979	1.16108	0.61623	0.93121	1.93327
min	29	0	1	94	126	0	0	71	0	0	1	0	3
25%	48	0	3	120	211	0	0	133.5	0	0	1	0	3
50%	56	1	3	130	241	0	1	153	0	0.8	2	0	3
75%	61	1	4	140	275	0	2	166	1	1.6	2	1	7
max	77	1	4	200	564	1	2	202	1	6.2	3	3	7

Performance measures are taken to estimate result of Cardiovascular diseases forecasting. It holds Recall, Precision, F1-score and accuracy. These are considered and results are obtained with the help of confusion matrix shown in Fig 3. True Positive (TP) is the condition where the number of instances marked true when they are actually true. The condition False Positive (FP) is the condition where the number of instances marked true when they are actually false. Similarly, False Negative (FN) is the condition here the number of instances marked as false while they are actually true. At the last True Negative (TN) is the condition where the number of records classified as false while they were actually false.

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{5}$$

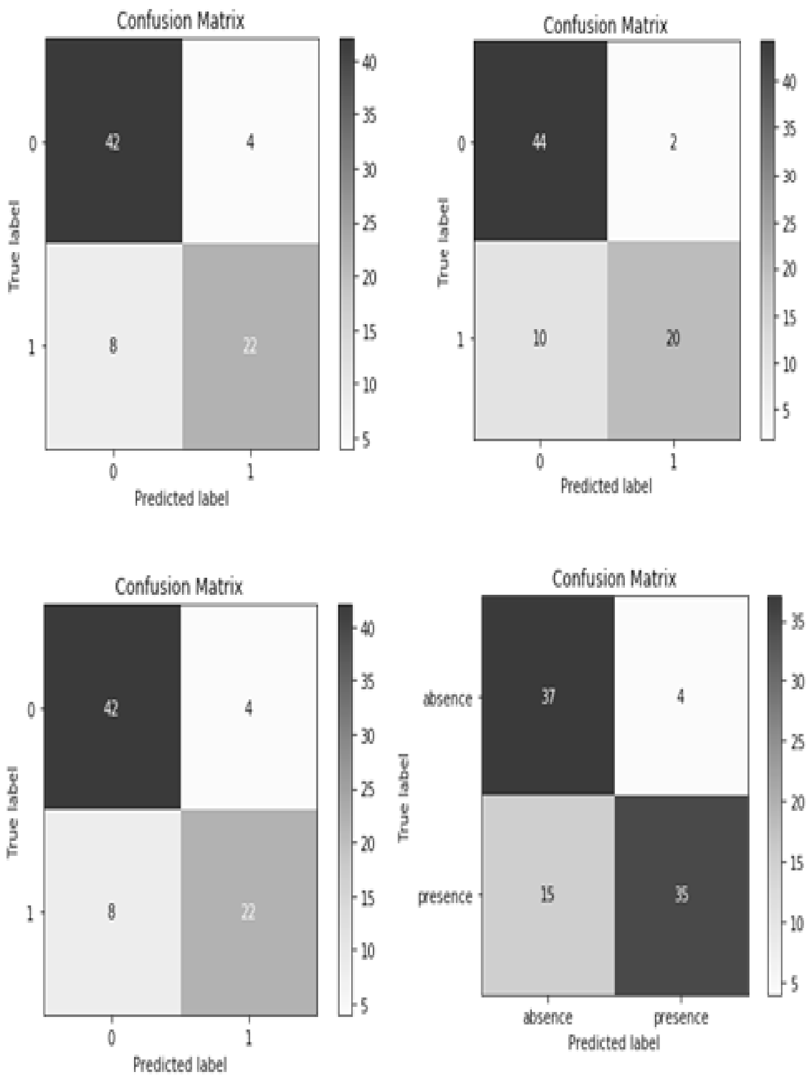


Fig. 3. Confusion Matrix of support vector matrix (SVM), ArtificialNeural Network (ANN), Random Forest(RF) and XGBoost

Table 3. Comparison of results of different methods

Methods	**Precision**	**Recall**	**f1-score**	**Support**	**Accuracy (%)**
XGBoost	0.81	0.79	0.79	91	79.12
RF	0.84	0.84	0.84	76	84.21
SVM	0.87	0.86	0.85	80	87.21
ANN	0.89	0.88	0.88	86	89.31

Based on the values obtained from confusion matrix of different methods taken into consideration Recall, Precision, F1 Score and Accuracy are calculated with given equations and shown in Table-III. However, It is identified that ANN is overwhelming with 89.31% other approaches XGBoost (79.12%), Random Forest (84.21%) and support vector machine (87.21%) illustrated in Fig-4.

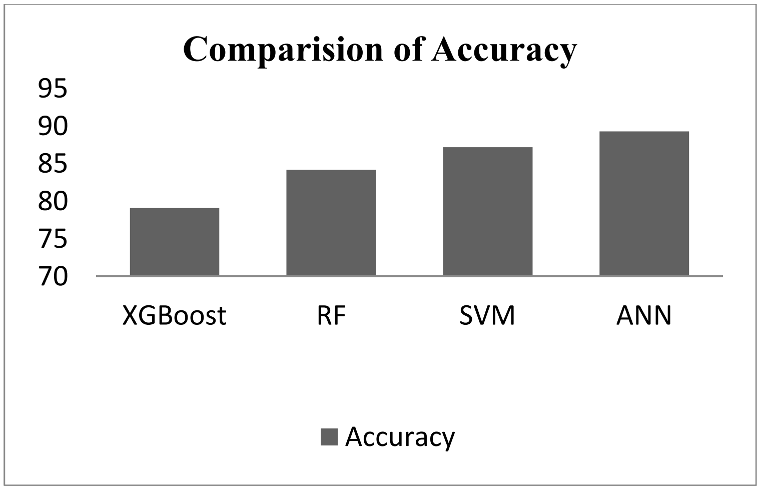


Fig. 4. results of accuracy measure

V. CONCLUSION AND FUTURE WORK

Cardiovascular diseases affect the functioning of the heart. There are various approaches including machine learning and soft computing have been applied on existing datasets to get accurate prediction of heart diseases. After applying numerous models, and defining log loss as the evaluation metric, we have found the ensemble of SVM, XGBoost, random forest and ANNs to be the best model for achieving the metric laid down. Due to the small size of the dataset given in the competition, Neural Networks could overfit quite easily, and the regularization applied to counter that would bring down the accuracy by quite a bit, resulting in relatively simpler machine learning algorithm like SVM outperforming them. Averaging the probabilities out was the best way to get the least log loss even if the individual models didn't give the best log loss by themselves. As the future work Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM) can be applied as core model while to get optimized parameters one of nature inspired algorithm like Genetic Algorithm, PSO can be applied to get more accurate prediction.

REFERENCES

[1] Chen, D., Afzal, N., Sohn, S., Habermann, E. B., Naessens, J. M., Larson, D. W., & Liu, H. (2018). Postoperative bleeding risk prediction for patients undergoing colorectal surgery. Surgery, 164(6), 1209-1216.

[2] K. Polaraju, D. Durga Prasad, "Prediction of Heart Disease using Multiple Linear Regression Model", International Journal of Engineering Development and Research Development, ISSN:2321 -9939, 2017.)

[3] Das, Resul, Ibrahim Turkoglu, and Abdulkadir Sengur. "Effective diagnosis of heart disease through neural networks ensembles." Expert systems with applications 36.4 (2009): 7675 -7680.

[4] Anbarasi, M., E. Anupriya, and N. C. S. N. Iyengar. "Enhanced prediction of heart disease with feature subset selection using genetic algorithm." International Journal of Engineering Science and Technology 2.10 (2010): 5370 -5376.

[5] Zhang, Yan, et al. "Studies on application of Support Vector Machine in diagnose of coronary heart disease." 2012 Sixth International Conference on Electromagnetic Field Problems and Applications. IEEE, 2012.

[6] Ajam, Noura. "Heart Diseases Diagnoses using Artificial Neural Network." Network and Complex Systems 5.4 (2015): 7 -10.

[7] Elshazly, Hanaa Ismail, Abeer Mohamed Elkorany, and Aboul Ella Hassanien. "Lymph diseases diagnosis approach based on support vector machines with different kernel functions." 2014 9th International Conference on Computer Engineering & Systems (ICCES). IEEE, 2014

[8] Dey, A., Singh, J. and Singh, N., 2016. Analysis of Supervised Machine Learning Algorithms for Heart Disease Prediction with Reduced Number of Attributes using Principal Component Analysis. Analysis,140(2), pp. 27 -31

[9] Weng, S.F., Reps, J., Kai, J., Garibaldi, J.M. and Qureshi, N., 2017. Can machine -learning improve cardiovascular risk prediction using routine clinical data?.PloS one,12(4), p.e0174944

[10] Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J., Sandhu, S., Guppy, K., Lee, S., & Froelicher, V. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. American Journal of Cardiology, 64,304--310.

[11] David W. Aha & Dennis Kibler. “Instance-based prediction of heart-disease presence with the Cleveland database”.

[12] Kumar, S. S., & Shaikh, T. (2017, September). Empirical evaluation of the performance of feature selection approaches on random forest. In *2017 International Conference on Computer and Applications (ICCA)* (pp. 227-231). IEEE.

[13] Li, Z. C., Zhai, G., Zhang, J., Wang, Z., Liu, G., Wu, G. Y., & Zheng, H. (2019). Differentiation of clear cell and non-clear cell renal cell carcinomas by all-relevant radiomics features from multiphase CT: a VHL mutation perspective. *European radiology*, *29*(8), 3996-4007.

[14] Sanchez-Pinto, L. N., Venable, L. R., Fahrenbach, J., &Churpek, M. M. (2018). Comparison of variable selection methods for clinical predictive modeling. *International journal of medical informatics*, *116*, 10-17.

[15] Panicacci, S., Donati, M., Fanucci, L., Bellini, I., Profili, F. & Francesconi, P. (2019, June). Exploring Machine Learning Algorithms to Identify Heart Failure Patients: the Tuscany Region Case Study. In *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 417-422). IEEE.

[16] Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *J Stat Softw*, *36*(11), 1-13.

Transfer Learning: An approach for Fish Species Classification in Underwater Environment with Given Dataset

Monika Mathur[1], Nidhi Goel[2]

Department of ECE, Indira Gandhi Delhi Technical University for Women, Delhi, India
[1]mathur.monika2007@gmail.com, [2]nidhigoel@igdtuw.ac.in

Abstract: The use of deep learning for image classification has lead to frontiers of research in the area of image processing. But now a days its use for fish species classification is an emerging area of research for scholars, marine scientist and biologists due to challenging underwater environment. Two light phenomena, namely absorption and scattering affects the image capturing process in water and thus produces low resolution and dark images making fish species recognition a tough job. Moreover, the limited and scarce dataset of fish species makes it even more difficult to train a neural network. The present paper proposes two automatic method for fish categorization in underwater images. First method takes advantage of transfer learning approach for fish species classification using ResNet50 architecture. Training only last few layers of ResNet50 network in transfer learning leads to increase in the classification accuracy even with limited data. The second method also utilizes the architecture of ResNet50 to extract features from it and then feed them to Support Vector Machine (SVM) classifier for fish species classification. Both the proposed methods have been tested on large as well as small datasets. Large and small dataset comprises of 27370 (Fish4Knowledge) and 600 (QUT) fish images respectively having low resolutions and illuminations. Result analysis indicates that both the proposed methods give good accuracy of 98.44% and 95.90% respectively for large dataset and 84.92 % and 77.86% respectively for smaller dataset. Unlike well-known convolutional neural network (CNNs) for fish species categorization, proposed methods can also classify fish images (low-quality training data) accurately.Moreover, proposed methods are compared with the state of art CNNs on large and small-scale datasets.The experimental results show that proposed methods outperform state of art CNNs, which indicates their potential applications in real time.

Keywords: Convolutional Neural Networks, Transfer Learning, ResNet50, Fish Species.

I. INTRODUCTION

In recent years exploration of fish species configuration and their distribution are of utmost significance to marine environment, biodiversity fortification, and the fishery industry. Underwater Fish images collected for these research are usually collected by scuba divers or autonomous underwater vehicles (AUVs) and then annotated physically by marine scientists and biologists. These processes are undoubtedly a terrific waste of time, resources and manpower. But now a days automatic systems implementing deep learning are more common for challenging problem of fish species classification with scarce and low quality data.These automatic systems can classify the fish species more accurately and without human intervention.

This paper aims at finding a solution to underwater fish classification with high accuracy. Deep neural networks can be the solutionto above mentioned problem. But learning a deep neural network from scratch requires a big dataset because it has large weights for training. Datasets available for fish classificationare of small size which are not sufficient for learning. So, two automatic method which uses transfer learning for fish classification are proposed in this paper. Transfer learning uses only last few layers of pre-trained network for classification and is thus fast and also requires less data for training as compared to Convolutional Neural Network (CNN) from scratch.

II. LITERATURE REVIEW

Fish species recognizance is amperage obscure job because of the challenging underwater environment. Many researchers and scholars are working in this domain since last decade and various novel approaches have been proposed for fish species classification.

Cabreira et al. [1] proposed a technique for the automatic classification of fish species depending on their echo recordings. Three parameters namely bathymetric, energetic and morphometric were mined from the echo recordings to serve as inputs to the Artificial Neural Networks (ANNs) used for testing. Classification accuracy of 96% was obtained, depending on the type of school parameters and networks utilized.Fish species classification method proposed by Castillo et al. [2] uses multiclass support vector machine (SVM) and two supervised ANN namely; Probabilistic Neural Network (PNN) and Multilayer Perceptron (MLP). Acoustic records were used as descriptor for the neural networks. Obtained results shows that both MLP and SVM methods with accuracy 89.5% outperforms the PNN with classification accuracy of 79.4%.

Boom et al. [3] proposed a heuristics tree based classification method and finally compares it to a state of art tree on a live fish image dataset. It outperforms the baseline method by achieving an accuracy of 90.0%. El-Bendary et al. [4] proposed an automatic fish classification technique based on SVM classifier and Speeded Up Robust Features (SURF) and Scale Invariant Feature Transform (SIFT) algorithms for feature extraction. Result analysis shows that the SVM based classification algorithm outperformed state of art techniques like ANNs and k-nearest neighbour (k-NN).A hierarchical categorization method for fish species recognition is proposed by Chuang et al. [5]. All levels of the species hierarchy undergoes partial categorization for coarse-to-fine cataloguing and discriminative feature descriptors are generated from species anatomical parts. Experimentation provesthat thepartial classification algorithm works well on imbalance dataset by achieving a classification accuracy of 94%and partial decision rate of 5%.

Ogunlana et al. [6] also proposed a SVM based technique for fish classification. The technique uses the five fin parameters (caudal, pelvic, dorsal, pectoral and anal) and the fish body for shape feature extraction. Results shows a classification accuracy of 78.59%, which outperformed K-mean clustering, K-NN and ANN algorithms. Salman et al. [7] proposed a CNN which uses species-dependenthierarchical features and evades the requirement of extracting them from unprocessed fish images. The CNN based method also performs good for test images which does not belongs to the training data. LifeCLEF15 and LifeCLEF14 fish datasets were used for results analysis and a classification accuracy of 90% is achieved.Automated fish classification method based on image processing, deep learning and convolutional neural networks was presented by Rathi et al. [8]. This method improves the accuracy over state of art methods and attains an accuracy of 96.29%.

Demertzis et al. [9] proposed an automatic Machine Hearing Framework (MHF) for Marine Species classification and recognition in underwater environment.It identifiesfish species based on their sounds. Hearing recognition is performed by using the Online Sequential Multilayer Graph Regularized Extreme Learning Machine Autoencoder (MIGRATE_ELM). An automated system based on CNN was proposed by Iqbal et al. [10] for fish species identification. It uses reduced form of AlexNet and utilizes only four convolutional and two fully connected layers instead of five convolutional and three fully connected layers. Obtained accuracy of 90.48% proves that Alexnet with less number of layers is efficient even on testing data.

III. PROPOSED METHODOLOGY

Fish species classification and recognition is a challenging and tough task due to underwater environment. Challenges hindering correct fish recognition comprises of noise, distortion, overlap, segmentation error, and obstructions. Various techniques includingK-mean clustering, K-Nearest neighbour (KNN) and neural network, are generally employed to solve these challenges but with certain limitations, which bounds the classification accuracy. Most of the above mentioned methods are computationally complex and requires huge data for accurate classification. To overcome these limitations and to achieve high accuracy with limited and untrained dataset two novel approaches are proposed in this paper. They utilizes the concept of transfer learning.

1. Transfer Learning

Transfer learning is a machine learning technique which **builds accurate deep learning models in a timesaving way** [11]. In transfer learning, learning process is not done from scratch; rather it extracts the features from pretrained layers of the network while solving a different problem. In this way transfer learning based networks make use of earlier learnings of network and resist starting it from scratch. Transfer learning generally makes use of **pre-trained networks**. Pre-trained networks are the networks that are trained on a large dataset consisting of millions of images to solve a problem identical to the one that we are interested in. Instead of training a network (deep networks) from scratch which is computationally complex, time consuming and very costly, it is better to use pretrained networks (e.g. AlexNet, VGG, MobileNet, GoogleNet, ResNet etc) to save both time and resources. Canziani et al. [13] presents a comprehensive review of pretrained networks' performance using images from the ImageNet [12] challenge.

2. Selection of CNN for Transfer Learning

Selection of pre-trained networkis a tough task. Table1 [14] shows the top1 and top5 accuracy for various CNNs. On basis of best top1 and top5 accuracy the present paper uses Resnet50 as pre-trained network for transfer learning. Resnets are a kind of CNNs called Residual Networks. They are very deep compared to Alexnet and VGG, and Resnet 50 refers to a 50 layers Resnet. They overcome the problems of vanishing gradients which is main issue in other deep networks with more layers.

Table 1. Top1 and top5 accuracy for various CNNs

Architecture	**Top1 error (%)**	**Top5 error (%)**
AlexNet	57.1	80.2
GoogleNet	69.8	89.3
VGG	70.5	91.2
ResNet50	75.2	93

Training deep neural networks (consisting of large layers) diminishes the gradient dramatically because it propagates backward through the network and error reduces to nearly zero as it reaches initial layers of the network. This problem is known as the problem of "*Vanishing Gradients*". Vanishing gradients distract the parameters and it is tough to find in which direction they should move in order to increase or decrease the cost function.

To overcome this problem Resnet networks make use of residual connections between its layers, meaning that the output of a layer is a convolution of its input plus its output. This is shown in figure1.

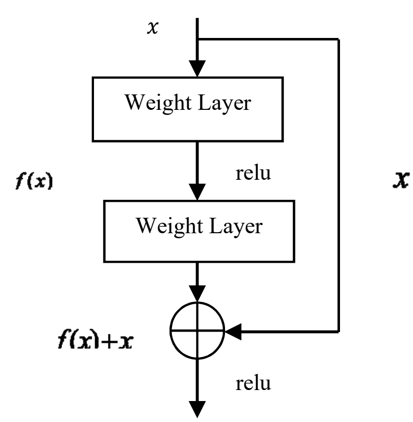


Fig. 1. Concept of residual connections [15]

The basic idea is simply to add an identify connection every few layers that adds the source of the block, x, to the output of the block $f(x)$, resulting in the final output of $H(x) = f(x) + x$. The name "residual networks" comes from the fact that network is learning $F(x) = H(x) - x$ the "residual" of what is left over when input is subtracted from output.

3. Proposed Method

The present paper proposes two methods for fish species classification. First one uses the concept of transfer learning i.e. it take layers from ResNet 50 trained on a large data set and fine-tune on a given fish data set and second method uses ResNet50 as feature extractor and passes its output to a SVM classifier for classification. The former is called Fish_Net and later is called as Fish_Net_SVM in the present paper.

4. Fine Tuning ResNet50 (Fish_Net)

The last few layers of the neural network are modified to apply transfer learning to it. The pre-trained neural networks are already trained to learn rich feature representations for various types of images. In present paper pre-trained network, ResNet50 is used for transfer learning. ResNet50 uses input images of size 224 x224x3, with 3 as number of colour channels.

Image features are extracted using the convolutional layers of ResNet50 network but classification is done using the last softmax layer and the final classification layer. To classify fish images for the Fish_Net, ResNet50 network is retrained by replacing last three layers. These three layers, 'fc1000', 'fc1000_softmax' and 'ClassificationLayer_ fc1000' contain information about combining the features that the network extracts from earlier layers into lables and class probabilities. The Fish _Net replaces these last three layers of ResNet50 network by a fully connected layer, a softmax layer, and a classification output layer. The final fully connected layer of Fish_Net is set to have the same size as the number of fish species to be classified i.e. 23. The learning rate of the fully connected layer in Fish_Net is increased to learn faster in new layers. To complete the connections in Fish_Net, the last transferred layer i.e. 'avg_pool' left in ResNet50 is connected to these new layers (fully connected, softmax, and classification output) of Fish_Net. These connections of ResNet50 and transferred ResNet50 (i.e. Fish_Net) are shown in figure 2 and 3 respectively.

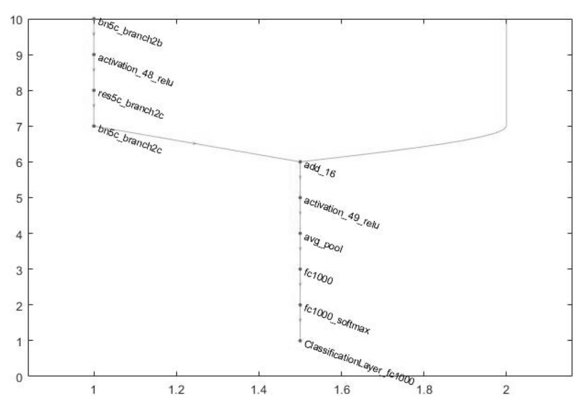


Fig. 2. Original connection of layers in ResNet50 network

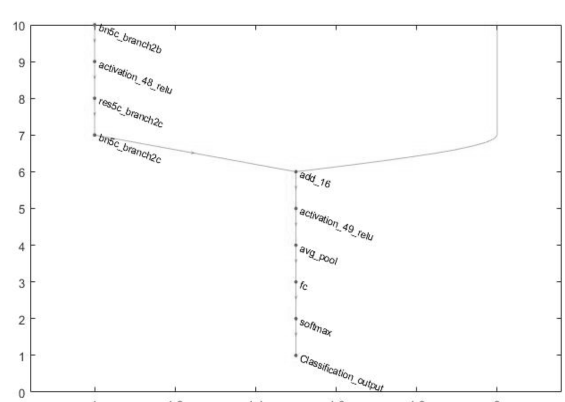


Fig. 3. Connection of layers in Fine tuned ResNet50 Network

Initial Layers of Fish _Net network are freezed by setting their learning rates to zero. During network training, gradients of these freezed layers are not computed because of zero learning rates and this will considerably improves the training speed. Freezing of layers also prevent overfitting problem in case of small datasets. The Fish_Net contains the same layers as ResNet50, but with the learning rates of the earlier layers set to zero. This is how transfer learning or fine tuning the ResNet50 classifies the fish species in the proposed method called Fish_Net.

5. ResNet50 as feature extractor (Fish_Net_SVM)

Fish_Net_SVM also extract image features from ResNet50 network and use those learned features to train an image classifier known as Support Vector Machine (SVM). Pretrained ResNet50 network is trained on a very large dataset consisting of millions of images and can classify images into thousand categories, like cap, pencil, torch, animals, birds, insects etc. Due to large dataset ResNet50 has learned feature representations for different types of images. These feature representations of input images are arranged hierarchically. Initial layers contain low level features like dots, lines, curves etc. but deeper layers contain higher-level features like edges, contours etc. Feature representations of the training and test images are achieved using activations on the global pooling layer (‹avg_pool›) at the end of the ResNet50 network. The 'avg_pool' layer pools these features over all spatial locations and gives a total of 2048 features. The class labels are extracted from both the training and test datasets. Features extracted from the training images act as predictor variables for a multiclass support vector machine (SVM) for classification. Block diagram of Fish_Net_SVM is shown in figure 4.

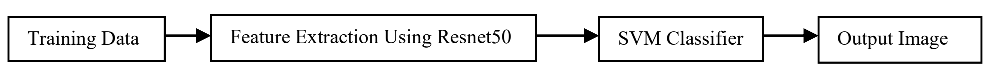


Fig. 4. Block diagram of Fish_Net_SVM

IV. RESULTS

1. Datasets

Effectiveness of the proposed CNN networks, Fish_Net and Fish_Net_SVM is evaluated on Fish Recognition dataset made by the Fish4Knowledge (F4K) [16] project and QUT dataset [17]. The fish dataset, F4K (27370 fish images) used for experiments is acquired from a liveunderwater video dataset. This data is divided into 23 fishspecies. These species are physically categorized by marinebiologists.Out of 27370 fish images 70% (19159 images) are used for training and 30% (8211 images) are used for validating data. Matlab 2018a and NVIDIA Quadro K2200 GPU is used for implementation of algorithmsand evaluation of results.

The QUT fish dataset comprises of 3,960 images collected from 468 species undercontrolled, out-of-the-water and underwater conditions. For proposed methods 37 species have been chosen for classification comprising of 600 images in underwater condition. 70:30 ratio is used for testing and validating images.

2. Result Evaluation

Both the proposed methods (Fish_Net and Fish_Net_SVM) are evaluated using F4K and QUT datasets. F4K has 27370 images and is called aslarge dataset whereas QUT has 600 images and is called as small dataset.Fish_Net and Fish_Net_SVM with large dataset gives an accuracy of 98.44% and 95.90% respectively. Training and Loss curves of Fish_Net method for large dataset are shown in figure 5. Time consumed by Fish_Net is 1387 minutes and 5 sec whereas the time taken by Fish_Net_SVM is 662.8795 seconds for 5 epochs and 0.001 learning rate. The above mentioned results depicts that Fish_Net (fine tuning the pre-trained network according to our data) is better if accuracy is priority and Fish_Net_SVM (Feature extraction from pre trained network + SVM classifier) is better if time consumption or speed of the network is the main concern.

Similarly, the results are also evaluated for small dataset to measure the changes in the accuracy and speed. As the data is small both the accuracy and time consumed is reduced for both the methods. Fish_Net and Fish_Net_SVM with small dataset gives an accuracy of 84.92% and 77.86 % respectively. Training and Loss curves of Fish_Net method for small dataset are shown in figure 6. Time consumed by Fish_Net is 14 minutes and 45 sec whereas the time taken by Fish_Net_SVM is 30.32 seconds for 5 epochs and 0.001 learning rate. The results of both the methods are tabulated in Table 2.

Table 2. Comparison of Fish _Net and Fish_Net_SVM in terms of accuracy and time consumption

	Fish_Net		**Fish_Net_SVM**	
	Accuracy	**Time**	**Accuracy**	**Time**
QUT dataset	84.92%	14 Min 45 Sec	77.86%	30.32 Sec
F4K dataset	98.44%	1387 Min 05 Sec	95.90 %	662.87 Sec

The results of 23 fish species classification using Fish_Net method are shown in Table3in terms of accuracy along with its comparison to exsisting methods on the F4K dataset. To make comparisons, traditional machine learning tools and other networks trained from scratch are used as reference methods. Methods like LDA (feature extractor) + SVM classifier [18], use only reshaped foreground fish images as input and achieved an accuracy of 80.14%. A test accuracy of 89.79% is achieved with nearest neighbour method. A SVM classifier is also trained on the raw pixels and it obtains an accuracy of 82.98%. An accuracy of 87.56% is obtained by using softmax classifier [19]. VLFeat [20] is also used as one of the method for comparison giving an accuracy of 93.58. Results from DeepFish architecture [19] with and without data augmentation are also included for comparison giving accuracy between 98.23 on test dataset. The results are listed in Table 3below. Alex-FT-Soft [21] uses AlexNet for transfer learning for automatic fish recognition and gives an accuracy of 96.61%. Fish_Net also extracts features from fish images using the pre-trained ResNet50 network and classify them by fine-tuning the network. The accuracy of 98.44% demonstrates that proposed network with transfer learning outperforms other state of art methods for fish species classification.

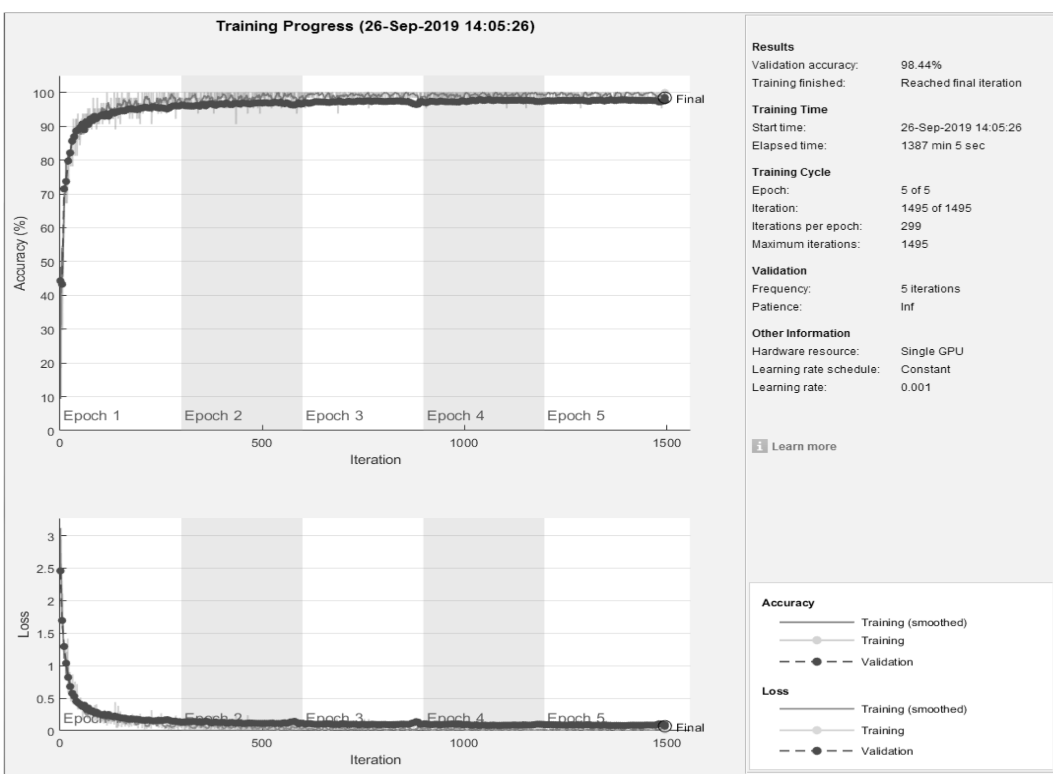


Fig. 5. Training and Loss curves of Fish_Net method for F4K dataset

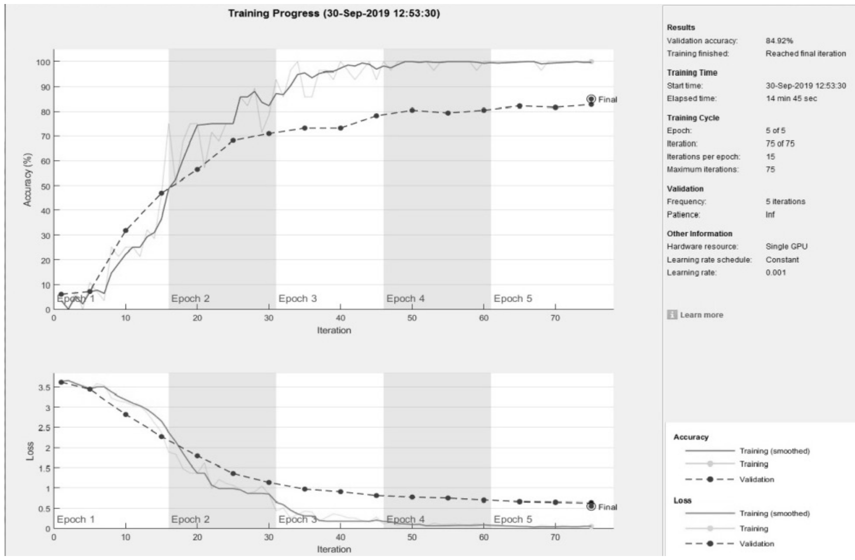


Fig. 6. Training and Loss curves of Fish_Net method for QUT dataset

Table 3. Accuracy comparison of fish species classification of various methods on Fish4knowledge dataset.

Classification Methods	Accuracy (%)	Classification Methods	Accuracy (%)
LDA+SVM [18]	80.14	VLFeat Dense-SIFT [20]	93.58
Raw –Pixel SVM [18]	82.92	Fish_Net_SVM (Proposed)	95.90
Raw –Pixel Softmax [19]	87.56	Alex-FT-Soft [21]	96.61
Raw –Pixel Nearest Neighbour [18]	89.79	DeepFish-SVM [19]	98.23
DeepFISH –Softmax-Aug[19]	92.55	Fish_Net (Proposed)	98.44

V. CONCLUSION

Automatic Fish recognition and classification is important for marine biologists for studying the variety of fishes. Since traditional methods are not eco-friendly as they affect fish behaviour by physically capturing them and also demands time and labour costs but transfer learning based Fish_Net and Fish_ Net_SVM provides a cost-effective and efficient solution to it. In the scenarios, where availability of sufficient data is an issue; transfer learning proves to be a better solution for this type of problem as pre-trained networks are used with fine tuning according to desired requirements and as feature extractor. The proposed Fish_Net_SVM network extracts fish features from images by using learned weights of the pretrained ResNet50 network and it uses SVM to classify fish species. Experimental results demonstrate that both the methods outperform various state of art methods used for fish species classification. The obtained results also reflects that the networks trained with transfer learning perform well than networks trained from scratch both in terms of computation complexity and accuracy.

REFERENCES

[1] Cabreira A. G., Tripode M., & Madirolas A. (2009). Artificial neural networks for fish-species identification. ICES Journal of Marine Science, 66 (6), 1119–1129.

[2] Robotham H., Bosch P., Gutiérrez-Estrada J.C., Castillo J., & Pulido-Calvo I. (2010). Acoustic identification of small pelagic fish species in Chile using support vector machines and neural networks. *Fisheries Research*, 102(1–2), 115-122.

[3] Huang X., Boom J., & Fisher B. (2012). Hierarchical Classification for Live Fish Recognition, *British Machine Vision Conference student workshop*.

[4] Fouad M. M. M., Zawbaa H. M., El-Bendary N., & Hassanien A. E. (2013). Automatic Nile Tilapia fish classification approach using machine learning techniques. 13th International Conference on Hybrid Intelligent Systems (pp. 173-178). Gammarth.

[5] Chuang M., Hwang J., Kuo F., Shan M., & Williams K. (2014). Live fish species by hierarchical partial classification based on the exponential benefit. 2014 IEEE International Conference on Image Processing (pp. 5232-5236). Paris.

[6] Ogunlana S.O., Olabode O., Oluwadare S. A. A., & Iwasokun G. B. (2015). Fish Classification Using Support Vector Machine. *African Journal of Computing & ICTs*, 8(2), 75-82.

[7] Salman A., Jalal A. , Shafait F. , Mian A. , Shortis M. , Seager J., & Harvey E. (2016). Fish species classification in unconstrained underwater environments based on deep learning. *Limnology and Oceanography Methods*, 14(9), 570-585.

[8] Rathi D., Jain S., & Indu S. (2017). Underwater fish species classification using convolutional neural network and deep learning. Ninth International Conference on Advances in Pattern Recognition (pp. 1-6). Bangalore.

[9] Demertzis K., Iliadis L.S., & Anezakis V.D. (2018). Extreme deep learning in biosecurity: the case of machine hearing for marine species identification. *Journal of Information and Telecommunication*, 2(4), 492-510.

[10] Iqbal M. A., Wang Z., & Ali Z.A. (2019). Automatic fish species classification using deep convolutional neural networks. *Wireless personal communication.*

[11] Rawat W., & Wang Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, 29(9), 2352–2449.

[12] Deng J., Dong W., Socher R., Li L. J., Li K. & Fei-Fei L. (2009). Imagenet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition (*pp. 248–255).

[13] Canziani A., Paszke A. & Culurciello E. (2016.) An analysis of deep neural network models for practical applications. *arXiv preprint arXiv:1605.07678.*

[14] https://cv-tricks.com/cnn/understand-resnet-alexnet-vgg-inception/ (last updated on 3.08.2018)

[15] He K., Zhang X., Ren S., & Sun J. (2016). Deep residual learning for image recognition. *Conference on computer vision and pattern recognition* (770-778). Las Vegas.

[16] [16] Boom C. S. B. J., Huang P. X., & Palazzo S. (2013). Fish recognition groundtruth data. http://groups.inf.ed.ac.uk/vision/fish4knowledge/website/groundtruth/recog/.[Online;accessedAug-2018]

[17] Zong C. M., Ge Y., & Corke P. P. (2016).Fish dataset. https://wiki.qut.edu.au/display/cyphy/Fish+Dataset. [Online; accessed Sep-2018].

[18] Chang C.C., & Jen C. (2011). Libsvm: A library for support vector machines. *ACM Transaction on Intelligent System and Technology*, 2(3), 1–27.

[19] Qin J. L. H., Li X. and Zhang C. (2016). Deepfish: Accurate underwater live fish recognition with a deep architecture. *Neurocomputing*, 187, 49–58.

[20] Vedaldi A., & Fulkerson B. P. (2010). Vlfeat: An open and portable library of computer vision algorithms. *18th ACM International Conference on Multimedia* (pp. 1469–1472).

[21] Tamou B., Ben A. & Nasreddine (2018). “Underwater live fish recognition by deep learning. *Image and Signal Processing* (pp. 275– 283).

Survey Paper on Lungs Cancer Detection Techniques

Vidhi Bishnoi, Nidhi Goel, Akash Tayal

Department of ECE, IGDTUW, Kashmere Gate, New Delhi, India

Abstract: Lung cancer is a major reason of a higher percentage of deaths worldwide. This paper reports the survey of various methods for lung cancer detection with multiple processing steps followed by pre-processing, segmentation, features selection and classification of lung images extracted from CT scans. The paper reviews the major machine learning and deep learning concepts pertinent to lung cancer detection. The paper summarizes the open challenges in this field and gives direction for the future work.

Keywords: Machine learning, deep learning, Lung imaging.

I. INTRODUCTION

Lung cancer can be diagnosed by measuring the cell growth in lung called as lung nodule. A pulmonary nodule is a small round or oval-shaped growth in the lung (Lung Lesion). It may also be called a "spot on the lung" or a 'coin lesion'. The nodule considered as a lung nodule, with diameter smaller than 3 centimetres. Growths that are larger than 3 centimetres in diameter are usually called lung masses and typically have a higher chance of being cancerous. Lung nodules are either considered noncancerous or cancerous. A noncancerous nodule is called benign, and a cancerous lung nodule is referred to as malignant. The cancerous tumour (malignant) spreads all over the body organs if not treated at the right time. This paper helps one to analyse the various processing steps along with the different methods used in past and can inculcate best features for better classification in future. To diagnose the lung cancer, we define the Lung cancer as an abnormal growth of tissue resulting from uncontrolled, progressive multiplication of cells in one or both of the lungs. These cells disturb the normal functioning of lungs and spread over other cites if not treated properly. There are two kinds of lung cancers: Small Cell Lung Cancer (SCLC)- These cells tends to grow more quickly and the other is Non-Small Cell Lung cancer (NSCLC) which is a major kind of lung cancer, it grows and spreads slowly beyond the lung area.

Non- Small Lung Cancer: According to [1] these types of cancer are divided into three sub types shown in Table 1.

Table 1. Details of Non-Small Lung Cancer

S.N.	Types of Non-Small Lung Cancer	Description
1.	Adenocarcinoma	It starts in the mucus making gland cells in the lining of airways, Depeursinge et al. [2] founds that the prediction of adenocarcinoma relapse has low specificity but very high sensitivity. They measure the diameter of solid nodules on repeated intervals while Coroller et al. [3] predicts distant metastasis of patients.
2.	Squamous Cell Cancer	This type develops in the flat cells that cover the surface of your airways. It tends to grow near the centre of the lung, strongly associated with smoking.
3.	Carcinoma	The cancer cells appear large (> 4cm) and round under the microscope. They are large cell carcinoma as well as small cell carcinoma.

After knowing about the severity of lung cancer, the major step is to extract the lung images for further diagnosis of cancerous cells. There are several ways for the diagnosis of lungs cancer, they comes with many kinds of modalities for the diagnosis of lung cancer like CT scan (Computed Tomography), PET (Positron Emission Tomography) and chest X-ray. Among these X-ray are capable only to identify early cancer for larger nodules.

II. MATERIAL AND METHODS

1. Computer Aided Diagnosis (CAD) System for Lung Image Processing System

Based on review of research papers, a generalized computer aided diagnosis system is developed after that combines the processing stages for the detection of lung diseases as in Figure 1.

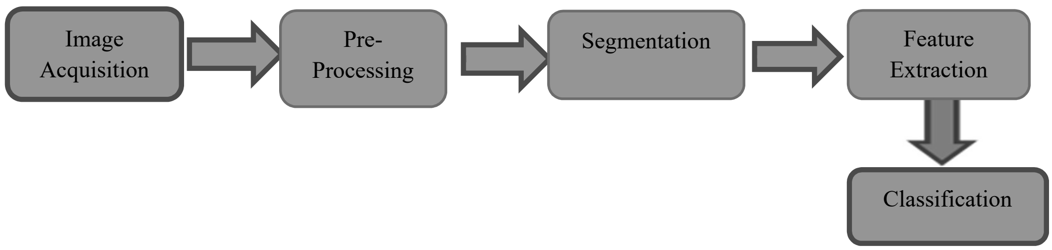


Fig. 1. Lung Image processing system

This lung processing unit is a collection of stages for lung image processing with the stage of image acquisition to post processing after the classification of abnormal tissues in CT scan images.

Input image for the pre-processing stage is acquired from the Computed Tomography (CT) scan with the help of CT scanners such as GE Light-Speed VCT 64 and Toshiba Aquilion 64 gives 64 slices mentioned by Armato III et al.[4], the scanner gives 12 bit DICOM (Digital Imaging and Communications in Medicine) images with mean intensity values in Hounsfield units (HU) in the range of -910 to -500 HU for lung tissues, which was also mentioned by Magdy et al.[5].After understanding the modality of acquitted image, a step forward to find these CT lung images can be done from publicly available database as from the Lung Image Database Consortium (LIDC) used by Jiang et al.[6], which was combined with image database resource initiative (IDRI)contains 1018 thoracic CT scans and corresponding radiologists annotations, which are saved in XML formats. As the data and annotations are freely available, companies and research groups may report their results on different subsets of the databases and will almost certainly perform evaluation in different ways. Making the use of the results efficiently Van Ginneken et al. [7] and Niemeijer et al. [8] compared ANODE09 (Automatic Nodule Detection 2009 database) dataset with other datasets to make the evaluation more accurate. Diciotti et al. [9] utilized ITALUNG Data Set which contains 86 solid non calcified non juxta-pleural in determinate small nodules (having mean diameter between 5 and 10 mm) identified in 63 subjects, repeatedly examined using low-dose protocols over a two-year-period.

2. Pre-Processing

Medical images are often degraded due to noise at the time of measurement process or scanning and acquisition of images from the system. Image quality and accuracy is the core factors of the research, image quality assessment as well as improvement are depending on the enhancement stage where pre-processing techniques as smoothing, denoising, enhancement are done. The various image enhancement techniques are used to improve the quality of the medical images, among these adaptive histogram equalisation is very beneficial for enhancing image information by intensity scaling. The high resolution images can be scaled through image decomposition techniques such as sub sampling or wavelet transforms whereas the low resolution images can be analysed by locating or extracting region of interests (ROIs) of the CT images shown by Rodrigues et al. [10]. Wiener filter used by Wei et al. [11] for noise removal and image enhancement whereas [12] used FFT (Fast Fourier Transform) for auto-enhancement of lung images. To reconstruct the noisy image Bayesian reconstruction used by Zhang et al.[13] for edge preservation of low dose computed tomography (LDCT). Ohkubo et al. [14] used a new image filtering technique with three types of reconstruction kernels like smooth, standard and sharp filters.

3. Segmentation

Early detection of lung cancer leads to more successful treatments, but unfortunately for the most part, lung cancer is not diagnosed until it has already spread throughout the body. Separation of these parts or regions of interest from the other parts of the lungs and the background region is an essential analysis by several algorithms. Segmentation divides the images into sub-parts for better analysis and processing. Segmentation algorithms can be classified according to the application, typically it is of classified in two categories: Region based and Edge based segmentation. Most basic region based algorithm is thresholding in which an image is divided into one group of pixels which are less than the threshold and another group of pixels that are greater than the threshold value. The other methods are the region based methods like Clustering, Region growing, Watershed algorithm which has been applied by [1], Lassen et al.[15], Firmino et al.[16] with edge based methods like Robert, Prewitt, Sobel and Canny methods. Review of segmentation methods is shown below in Table2.

Table 2. Overview of segmentation methods used for lung cancer detection process.

S. N.	References	Segmentation Methods
1	Sun et al.[17]	Robust ASM (Active Shape Model)
2	[1], Tan et al.[18]	Marker controlled Watershed Segmentation
3	Mansoor et al.[19]	Fuzzy Connectedness image segmentation
4	Farhangi et al.[20]	Chan-Vese algorithm
5	Bragman et al.[21]	Unsupervised Lobe segmentation algorithm
6	Hosseini-Asl et al.[22]	Incremental constrained Nonnegative matrix factorization (ICNMF)
7	Song et al.[23]	Graph cut image segmentation

4. Feature Extraction/ Selection

After the successful pre-processing and segmentation of the datasets of the lung CT images, another important stage in the analysis of lung cancer is the feature extraction for machine learning methods to classify the lung image as cancerous or non-cancerous. Features can be classified as 2D, 3D texture features as tumor intensity, shape and texture that has been measured by Coroller et al. [3] and Taşcı et al. [24] whereas, radiomic features that can capture type of tumours in a large set of images are discussed in Van Timmeren et al. [25]. Features can be the statistical parameters of the lung which are further selected for high speed analysis with the help of various optimisation techniques like Genetic Algorithm, Principle Component Analysis (PCA) used by Zhang et al.[26] whereas, Moradi et al.[27] applied Hybrid particle swarm optimization technique and many other methods have been implemented to find better features for classification. The histogram intensities of Gray Level Co-occurrence Matrix (GLCM) technique are used for extracting the actual features shown by Sorensen et al.[28], Korfiatis et al.[29] mentioned in table 3. As well as, Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP) are best utilized for feature extraction. Auto- encoders are simple networks, Stacked auto-encoders are also introduced to map input with the output layer and one hidden layer by Chen et al.[30] and Kumar et al.[31] and they reconstructed the input efficiently.

Table 3. Overview of Feature extraction methods used for lung cancer detection process.

S. N.	References	Feature Extraction Methods
1	Rodrigues et al. [10]	Structural Co-occurrence Matrix (SCM)
2	Sorensen et al. [28]	Local Binary Patterns(LBP) ,Gaussian Filter Bank(GFB), Feature Histograms
3	Korfiatis et al. [29]	3-D Gray-Level Co-occurrence Features, Texture-Based discriminant analysis
4	Hawkins et al. [32]	Relief-F algorithm, Correlation based Feature 3 D Textured Features Selection (CFS)
5	Chen et al. [30] and Kumar et al. [31]	Stacked de-noising auto-encoder (SDAE)
6	Ma et al. [33]	Minimum spanning tree(MST) algorithm

5. Classification

Finally to differentiate among normal and abnormal lung tissues, classification is used as an essential stage of the lung cancer detection process. To teach the CAD system for the diagnosis of malignant and benign tissues, it is necessary to apply classification algorithms to the system which gives better efficiency and accuracy. The concept of classification gives the machine learning methods that are classified as supervised and unsupervised learning. The supervised learning involves the model with input as the features that are selected for processing and the corresponding output will be the result of the diagnosis whether disease present or not as it is classification and regression. Whereas the un- supervised learning groups the data based only on input data as it is clustering. A huge amount of researchers like Sorensen et al.[28], Korfiatis et al.[29], Hawkins et al.[32] used supervised learning models as support vector machine (SVM), Naïve Bayes, Nearest Neighbour, Decision Tree, Neural Networks etc. and classifies the non-cancerous and cancerous lung images with the measurement of classification accuracy in the form of specificity and sensitivity of their models. With the thrust of improving the accuracy for better classification, other models are introduced such as Deep Learning techniques which are the recently used models for detection of pulmonary nodules along with the improvement in false positives shown by Teramoto et al.[34] by means of convolutional neural networks (CNNs).

Two other dominant classes of end-to-end learning machines are massive-training artificial neural networks (MTANNs) and convolutional neural networks (CNNs) are introduced by Tajbakhsh et al.[35] with the result of 2.7 false positives per patient at 100% sensitivity, which was significantly ($p<.05$) lower than the best performing CNN model with 22.7 false positives per patient at the same level of sensitivity. With the improvement in convolutional neural network Coudray et al.[36] trained a new deep convolutional neural network (inception v3) on whole-slide images obtained from the Cancer Genome Atlas to accurately and automatically classify them into LUAD (Adenocarcinoma), LUSC (squamous cell carcinoma) or normal lung tissue with improved results.

Table 4. Review of various classification models.

S. N.	References	Classification Methods
1	Sorensen et al.[28]	kNN (k- Nearest Neighbor)
2	Korfiatis et al.[29]	Naïve Bayes (NB), SVM, random forest (RF)
3	Ma et al.[33]	Unsupervised learning
4	Coudray et al.[36], Hua et al. [39]	Deep belief network (DBN), Convolutional network (CNN)
5	Song et al.[38]	Large margin local estimate(LMLE) classification model
6	Chen et al.[30], Gao et al.[37], Shin et al.[40], Anthimopoulos et al.[41]	Deep Convolutional Neural Networks CNN
7	Shen et al.[42]	Multi crop CNN

III. CONCLUSION

Lung cancer detection is a challenging area for researchers to correctly classify the lung nodules. This paper supports to carefully choose the segmentation, feature selection and classification methods with regard to predictive accuracy. Despite of traditional methods of processing, new techniques are also defined in review to take into brighter direction for accurate prediction and real world challenges in determining lung diseases. We need to appreciate the authors who gave us the direction to move forward with better accuracy and predictions.

REFERENCES

[1] Al-Tarawneh, M. S. (2012). Lung cancer detection using image processing techniques. Leonardo Electronic Journal of Practices and Technologies, 11(21), 147-58.

[2] Depeursinge, A., Yanagawa, M., Leung, A. N., & Rubin, D. L. (2015). Predicting adenocarcinoma recurrence using computational texture models of nodule components in lung CT. Medical physics, 42(4), 2054-2063.

[3] Coroller, T. P., Grossmann, P., Hou, Y., Velazquez, E. R., Leijenaar, R. T., Hermann, G., & Aerts, H. J. (2015). CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiotherapy and Oncology, 114(3), 345-350.

[4] Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., & Kazerooni, E. A. (2011). The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Medical physics*, *38*(2), 915-931.

[5] Magdy, E., Zayed, N., & Fakhr, M. (2015). Automatic classification of normal and cancer lung CT images using multiscale AM-FM features. Journal of Biomedical Imaging, 2015, 11.

[6] Jiang, H., Ma, H., Qian, W., Gao, M., & Li, Y. (2017). An automatic detection system of lung nodule based on multigroup patch-based deep learning network. IEEE journal of biomedical and health informatics, 22(4), 1227-1237.

[7] Van Ginneken, B., Armato III, S. G., de Hoop, B., van Amelsvoort-van de Vorst, S., Duindam, T., Niemeijer, M., ... & Camarlinghi, N. (2010). Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study. Medical image analysis, 14(6), 707-722.

[8] Niemeijer, M., Loog, M., Abramoff, M. D., Viergever, M. A., Prokop, M., & van Ginneken, B. (2010). On combining computer-aided detection systems. IEEE Transactions on Medical Imaging, 30(2), 215-223.

[9] Diciotti, S., Lombardo, S., Coppini, G., Grassi, L., Falchini, M., & Mascalchi, M. (2009). The ${LoG} $ Characteristic Scale: A Consistent Measurement of Lung Nodule Size in CT Imaging. IEEE Transactions on Medical Imaging, 29(2), 397-409.

[10] Rodrigues, M. B., Da Nóbrega, R. V. M., Alves, S. S. A., Rebouças Filho, P. P., Duarte, J. B. F., Sangaiah, A. K., & De Albuquerque, V. H. C. (2018). Health of things algorithms for malignancy level classification of lung nodules. IEEE Access, 6, 18592-18601.

[11] Wei, Q., & Hu, Y. (2014). A hybrid approach to segmentation of diseased lung lobes. IEEE journal of biomedical and health informatics, 18(5), 1696-1706.

[12] Al-Tarawneh, M. S. (2012). Lung cancer detection using image processing techniques. Leonardo Electronic Journal of Practices and Technologies, 11(21), 147-58.

[13] Zhang, H., Han, H., Liang, Z., Hu, Y., Liu, Y., Moore, W., & Lu, H. (2015). Extracting information from previous full-dose CT scan for knowledge-based Bayesian reconstruction of current low-dose CT images. IEEE transactions on medical imaging, 35(3), 860-870.

[14] Ohkubo, M., Wada, S., Kayugawa, A., Matsumoto, T., & Murao, K. (2011). Image filtering as an alternative to the application of a different reconstruction kernel in CT imaging: feasibility study in lung cancer screening. Medical physics, 38(7), 3915-3923.

[15] Lassen, B., van Rikxoort, E. M., Schmidt, M., Kerkstra, S., van Ginneken, B., & Kuhnigk, J. M. (2012). Automatic segmentation of the pulmonary lobes from chest CT scans based on fissures, vessels, and bronchi. IEEE transactions on medical imaging, 32(2), 210-222.

[16] Firmino, M., Angelo, G., Morais, H., Dantas, M. R., & Valentim, R. (2016). Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy. Biomedical engineering online, 15(1), 2.

[17] Sun, S., Bauer, C., & Beichel, R. (2011). Automated 3-D segmentation of lungs with lung cancer in CT data using a novel robust active shape model approach. IEEE transactions on medical imaging, 31(2), 449-460.

[18] Tan, Y., Schwartz, L. H., & Zhao, B. (2013). Segmentation of lung lesions on CT scans using watershed, active contours, and Markov random field. Medical physics, 40(4), 043502.

[19] Mansoor, A., Bagci, U., Xu, Z., Foster, B., Olivier, K. N., Elinoff, J. M., & Mollura, D. J. (2014). A generic approach to pathological lung segmentation. IEEE transactions on medical imaging, 33(12), 2293-2310.

[20] Farhangi, M. M., Frigui, H., Seow, A., & Amini, A. A. (2017). 3-D active contour segmentation based on sparse linear combination of training shapes (SCoTS). IEEE transactions on medical imaging, 36(11), 2239-2249.

[21] Bragman, F. J., McClelland, J. R., Jacob, J., Hurst, J. R., & Hawkes, D. J. (2017). Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior. IEEE transactions on medical imaging, 36(8), 1650-1663.

[22] Hosseini-Asl, E., Zurada, J. M., Gimel'farb, G., & El-Baz, A. (2015). 3-d lung segmentation by incremental constrained nonnegative matrix factorization. IEEE Transactions on Biomedical Engineering, 63(5), 952-963.

[23] Song, Q., Bai, J., Han, D., Bhatia, S., Sun, W., Rockey, W., & Wu, X. (2013). Optimal co-segmentation of tumor in PET-CT images with context information. IEEE transactions on medical imaging, 32(9), 1685-1697.

[24] Taşcı, E., & Uğur, A. (2015). Shape and texture based novel features for automated juxtapleural nodule detection in lung CTs. Journal of medical systems, 39(5), 46.

[25] Van Timmeren, J. E., Leijenaar, R. T., van Elmpt, W., Reymen, B., Oberije, C., Monshouwer, R., & Lambin, P. (2017). Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. Radiotherapy and Oncology, 123(3), 363-369.

[26] Zhang, Y., Rong, J., Lu, H., Xing, Y., & Meng, J. (2017). Low-dose lung CT image restoration using adaptive prior features from full-dose training database. IEEE transactions on medical imaging, 36(12), 2510-2523.

[27] Moradi, P., & Gholampour, M. (2016). A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy. Applied Soft Computing, 43, 117-130.

[28] Sorensen, L., Shaker, S. B., & De Bruijne, M. (2010). Quantitative analysis of pulmonary emphysema using local binary patterns. IEEE transactions on medical imaging, 29(2), 559-569.

[29] Korfiatis, P. D., Karahaliou, A. N., Kazantzi, A. D., Kalogeropoulou, C., & Costaridou, L. I. (2009). Texture-based identification and characterization of interstitial pneumonia patterns in lung multidetector CT. IEEE transactions on information technology in biomedicine, 14(3), 675-680.

[30] Chen, S., Qin, J., Ji, X., Lei, B., Wang, T., Ni, D., & Cheng, J. Z. (2016). Automatic scoring of multiple semantic attributes with multi-task feature leverage: A study on pulmonary nodules in CT images. IEEE transactions on medical imaging, 36(3), 802-814.

[31] Kumar, D., Wong, A., & Clausi, D. A. (2015, June). Lung nodule classification using deep features in CT images. In 2015 12th Conference on Computer and Robot Vision (pp. 133-138). IEEE.

[32] Hawkins, S. H., Korecki, J. N., Balagurunathan, Y., Gu, Y., Kumar, V., Basu, S., & Gillies, R. J. (2014). Predicting outcomes of nonsmall cell lung cancer using CT image features. IEEE access, 2, 1418-1426.

[33] Ma, L., Liu, X., & Fei, B. (2016). Learning with distribution of optimized features for recognizing common CT imaging signs of lung diseases. Physics in Medicine & Biology, 62(2), 612.

[34] Teramoto, A., Fujita, H., Yamamuro, O., & Tamaki, T. (2016). Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique. Medical physics, 43(6Part1), 2821-2827.

[35] Tajbakhsh, N., & Suzuki, K. (2017). Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs. Pattern recognition, 63, 476-486.

[36] Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., & Tsirigos, A. (2018). Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nature medicine, 24(10), 1559.

[37] Gao, M., Bagci, U., Lu, L., Wu, A., Buty, M., Shin, H. C., & Xu, Z. (2018). Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 6(1), 1-6.

[38] Song, Y., Cai, W., Huang, H., Zhou, Y., Feng, D. D., Wang, Y., & Chen, M. (2015). Large margin local estimate with applications to medical image classification. IEEE transactions on medical imaging, 34(6), 1362-1377.

[39] Hua, K. L., Hsu, C. H., Hidayati, S. C., Cheng, W. H., & Chen, Y. J. (2015). Computer-aided classification of lung nodules on computed tomography images via deep learning technique. OncoTargets and therapy, 8.

[40] Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging, 35(5), 1285-1298.

[41] Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., & Mougiakakou, S. (2016). Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE transactions on medical imaging, 35(5), 1207-1216.

[42] Shen, Wei, Mu Zhou, Feng Yang, Dongdong Yu, Di Dong, Caiyun Yang, YaliZang, and JieTian. “Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification.” Pattern Recognition 61 (2017): 663-673.

A Novel Approach for the Design of Optimal Notch Filter Cuckoo Search Optimization

Suman Yadav[1,2], Nisant Jain[1], Harshit Mittal[1], Itisha Goyal[1], Akanksha Prasad[1]

[1]Bharati Vidyapeeth's College of Engineering, ECE, New Delhi 110063
[2] Indira Gandhi Delhi Technical university for women (IGDTUW), ECE, New Delhi 110006

Abstract: This paper represents designing of one dimensional FIR (Finite Impulse Filter) Notch filter by computing the optimal filter coefficients in a way that the frequency response of the designed filter approximates ideal frequency response using proposed optimization algorithms. In this work cuckoo search algorithm is considered to design a linear phase FIR Notch filter of 20th order. Analysis of simulation results of the designed filter is done and various parameters are calculated such as maximum stop band ripple, maximum pass band ripple and maximum attenuation in the stop band.

Keywords: Digital filter design, Filter response, Cat swarm optimization algorithm, Bio-inspired algorithms, Cuckoo search algorithm

I. INTRODUCTION

In signal processing, a filter is used to remove the parts of a signal that are not needed. In a digital filter the representation of a signal is done by numbers instead of a voltage or a current. The process involves multiplying the input values to a certain constant and then carrying on them the necessary calculations. The main advantage of a digital filter is that it is programmable and easy to design. Also, a digital signal can process very low frequency and versatile signals. The filter works when an unfiltered analog signal is fed into ADC, the sampled digitized signal obtained is then processed to obtain digitally filtered signal which is finally fed into a DAC.

The notch filter is a filter that is used for stopping very narrow bandwidth of frequency. The gain of the notch filter drops and rises very steeply as the frequency increases. The magnitude of a notch filter looks like a notch and hence the name. The main advantage of notch filters compared to other filters is that it passes the high frequency signals without attenuating them. Another aspect that differentiates notch filter from other filters is that they cause very little phase lag at the gain crossover frequency. For notch filters to work, they must be tuned perfectly to resonant frequency else they give error. If the work is to be done for a specific frequency then notch filters are of great help as they give high accuracy output.

Notch filter has many applications in the modern day world. A very important use of notch filters is to enhance the electrocardiogram (ECG) signals. At a certain frequency in the ECG signal where there is interference, a notch filter is applied for that specific frequency and the output thus obtained is free of any interference. Notch filters are also used in the transmission and receiving functioning of various musical instruments. They are also pivotal to eliminate a certain frequency in a sound. Over all the notch filters become very important at any point of time where we have to deal with a single frequency or very narrow range of frequency.

Various papers have already been published about notch filters where the notch filters have been designed using various algorithms. One among them is using pole-zero placement method. The advantage of this design is that it gives error less output for second order equations. Another method used to design a notch filter is the one that simplifies the complex circuit involved by applying a method of all pass filters. The method also improves the notch bandwidth characteristics. Other notch filters have been designed like the one used for tracking ECG signals. One more interesting notch filter is the one that uses an algorithm which involves a mixture of tuning of notch frequencies and sparsity of filter coefficients.

Our objective is to design a notch filter with the help of cuckoo search algorithm.

II. PROBLEM STATEMENT

In this paper we have designed a linear phase FIR notch filter with the help of Cuckoo search algorithm. The general equation for frequency response of a FIR notch filter is given by:

$H(w) = A(w)e^{j\theta(w)}$,

Here H(w) is the magnitude function, A(w) is the amplitude function and Θ(w) is the phase function.

Taking the discrete time fourier transformation of the above equation, we get:

$$H(e^{jw}) = \sum_{n=0}^{N} h[n]e^{-jwn}$$

The Ideal notch filter inhibits the passage of a single frequency, while practical filters don't allow the passage of a narrow band of frequencies.The plot of discrete fourier transform of ideal filter response, denoted by **Hid(w)**, is free of ripples, while the practical filter response,denoted by **H(w)**,contains ripples in the passband, stopband as well as in the transition bandwidth because of convolution with decaying oscillations.

To optimise the filter, i.e., to minimize the difference between the ideal and realisable response, we use the **Cuckoo Search Algorithm** which was designed by observing the parasitic behavior of Cuckoo birds while laying eggs.

The cuckoo search algorithm is now explained in detail with reference to the notch filter which contains the following steps:

1. Define optimization parameters which are the passband edge frequency(wp), stopband edge frequency (ws), passband ripple frequency (delta P) and stopband ripple frequency (delta s),population size which is the length of notch filter (N) the fitness function which are the controlling parameter(wt) and tolerance of filter.
2. Generate the initial population of host nests (Ns).
3. Evaluate the fitness of all solutions and identify the best solutions.
4. Generate new solution by levy flights:

$$x_i^{t+1} = x_i^t + \quad \oplus\, Le'vi$$

5. Evaluate the fitness value for new solutions.
6. Replace a fraction of bad solutions with good solutions.
7. Compare the fitness value and keep the best nest with best quality solutions.
8. Determine new alien nests with probability and local step size.
9. Evaluate the fitness value for new alien solutions.
10. Compare the fitness value for each new alien nest and best nest.

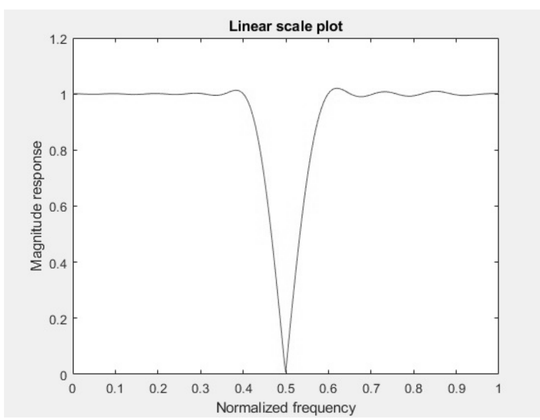
III. SIMULATION RESULTS

1. Notch filter

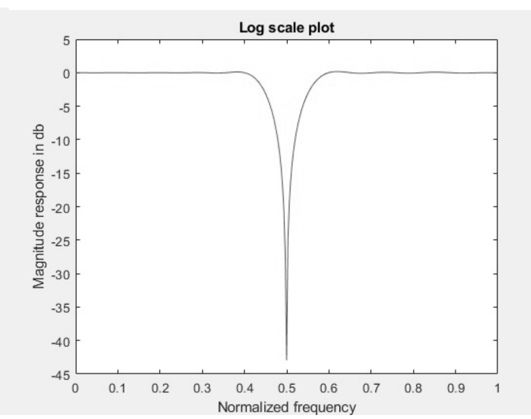
These simulation is done using MATLAB with Intel core i5 7th Generation 2.70 Ghz, 8GB RAM to design 20th Notch filter. Specification for the FIR Notch such as notch frequency is taken as o.5* π. CS algorithm is run for 500 cycles to get the best result and population size is taken as 30. The range for the frequency components ω lies from 0,π The limits for the filter coefficients are taken from -1 to 1. Optimised coefficients are obtained by minimising the error fitness function using CS and given in table 1. Magnitude response based on these coefficients and its dB response are plotted in Fig.1.

Table 1. Optimized filter coefficients of FIR Notch of 20th order

b(p)	Value
b(1) = *b*(21)	0.00579158511553580
b(2) = *b*(20)	-0. -0.00541117722064333
b(3) = *b*(19)	0.00533471712058508
b(4) = *b*(18)	-0.0230700840282999
b(5) = *b*(17)	0.0223560522143460
b(6) = *b*(16)	0.0207449280739406
b(7) = *b*(15)	-0.0249879253751555
b(8) = *b*(14)	-0.0329028938506135
b(9) = *b*(13)	-0.000731175864290419
b(10) = *b*(12)	0.148911740043931
b(11))	-0.172761614995482



(a)



(b)

Fig. 2. **(a)** Magnitude response of LPF of 20th order (b) Magnitude response in dB scale

IV. CONCLUSION

The proposed research aims in designing FIR notch filter having minimum ripples in pass band and higher attenuation in the stop band based on CS optimization algorithm. It is found that there is improvement in both the objectives as CS has got improvement in reduction of ripples in pass band with higher attenuation in the stop band for designing FIR filters. The graph is compared with the papers mentioned in the references column. There is a definite improvement in the shape of the curve and the reduction of ripples is also clearly evident.

REFERENCES

[1] Benhala B.: Ant colony optimization for optimal low-pass butterworth filter design. WSEAS Transactions on Circuits and Systems 13, 313-8 (2014).

[2] Aggarwal, A., Rawat, T. K., Kumar, M., & Upadhyay, D. K.: Optimal design of FIR high pass filter based on L1 error approximation using real coded genetic algorithm. Engineering Science and Technology, an International Journa*l*, *18*(4), 594-602 (.2015)

[3] Ababneh, J. I., & Bataineh, M. H.: Linear phase FIR filter design using particle swarm optimization and genetic algorithms. Digital Signal Processing, *18*(4), 657-668 (2008).

[4] Saha, S. K., Ghoshal, S. P., Kar, R., & Mandal, D.: Cat swarm optimization algorithm for optimal linear phase FIR filter design. *ISA transactions*, *52*(6), 781-794 (2013).

[5] Saha, S. K., Ghoshal, S. P., Kar, R., & Mandal, D.: Cat swarm optimization algorithm for optimal linear phase FIR filter design. ISA transactions, *52*(6), 781-794 (2013).

[6] Daniel E, Anitha J, Gnanaraj J.: Optimum laplacian wavelet mask based medical image using hybrid cuckoo search–grey wolf optimization algorithm. Knowledge-Based Systems 131, 58-69 (2017).

[7] Aggarwal, A., Rawat, T. K., Kumar, M., & Upadhyay, D. K.: Optimal design of FIR high pass filter based on L1 error approximation using real coded genetic algorithm. *Engineering Science and Technology, an International Journal*, *18*(4), 594-602 (2015).

[8] San José-Revuelta, L. M., & Cid-Sueiro, J.: Robust joint channel and noise estimation in Bayesian blind equalizers. *Signal processing*, *84*(3), 535-548 (2004).

[9] Mirjalili S, Saremi S, Mirjalili SM, Coelho LD.: Multi-objective grey wolf optimizer: a novel algorithm for multi-criterion optimization. Expert Systems with Applications 47,106-19 (2016).

[10] Reddy, K. S., & Sahoo, S. K. An approach for FIR filter coefficient optimization using differential evolution algorithm. AEU-International Journal of Electronics and Communications, *69*(1), 101-108 (2015).

[11] Saha, S. K., Ghoshal, S. P., Kar, R., & Mandal, D.: Cat swarm optimization algorithm for optimal linear phase FIR filter design. *ISA transactions*, *52*(6), 781-79 (2013).

[12] Sharma, I., Kuldeep, B., Kumar, A., & Singh, V. K.: Performance of swarm based optimization techniques for designing digital FIR filter: A comparative study. Engineering science and technology, an international journal, *19*(3), 1564-1572 (2016).

Customer Segmentation and Outlier Detection in Power Distribution using Clustering Algorithms

Hadiza Ali Umar[1], Rajesh Prasad[2], Mathias Fonkam[3]

[1,3]School of IT and Computing
American University of Nigeria, Yola, Nigeria
[2]Department of Computer Science
African University of Science and Technology, Abuja, Nigeria
[1]hadiza.umar@aun.edu.ng, [2]rprasad@aust.edu.ng, [3]mfonkam@aun.edu.ng

Abstract: Power utilities globally have a huge task of enhancing their services to enable them to compete in emerging markets. However, power utilities are not able to live up to their undertakings as a result of challenges of revenue deficits which ensue due to collection losses known as Non-Technical losses (NTL). Evaluating the intensity and reducing NTLs to increase revenue generation, thereby improving reliability of power value chain must remain a top priority for utilities and authorities. Machine Learning (ML) techniques offer an unprecedented opportunity for researchers to harness the vast amount of data generated by power utilities to identify energy payment and consumption patterns of consumers. One application is the segmentation of electricity consumers on same profiles into clusters of similar behaviour. This paper applies k-means, DBSCAN and Expectation Maximization clustering algorithms to segment customers and detect outliers by assessing consumers' bill payments, consumption patterns, tariff plans and grouped into segments.

Keywords: Non-Technical loss, Customer segmentation, K-means, DBSCAN, Expectation Maximization.

I. INTRODUCTION

The high increase in demand for power has grown by 4%, quantified in excess of 23,000 Tera Watt hour (TWh), therefore regarded as the fuel for the future. This speedy growth is driving electricity to get a 20% allocation in the total final consumption of energy ascribed to increase in population and power generation. The International Energy Agency (IEA) has anticipated that by the year 2040, electricity consumption will be on the rise worldwide. This evaluated increase has left utilities worldwide with no choice but to seek solutions in tackling some of the very challenging power issues. It has been established severally, that reducing non-technical losses and increasing revenue generation have continued to be two of the most critical of the challenges.

Non-technical losses are referred to as collection and commercial losses resulting from consumed energy that has not been billed owing to theft, tampering of metering infrastructure or evasion of bills. NTLs are typically registered through power distribution, making it the motivation of the research.

The provision of power in Nigeria as in many other nations is a value chain, comprising of generation, transmission and distribution. These services are manned by different establishments. Collectively, these power utilities are currently striving to deal with a huge debt profile of over $2 billion. It has explicitly affirmed that Nigeria had expended $31.45 billion from 1999 till date. This is the trend in most developing nations.

Power utilities in these countries have continued to record huge short falls in revenue due to NTLs. These losses are disproportionately distributed among the different categories of power consumers. These deficits are handed over to consumers through the escalation of tariffs charged by the utilities. However, such increase does not produce the anticipated result as consumers continue to default and power has remained unstable. This problem is even more prevalent in developing nations with precarious economies like Nigeria. Therefore, electricity is no longer discounted and not affordable by the poor and low-income earners, hence their inability to pay for their bills. This and a lot more are reasons for the major causes of NTLs [1], [2]in most papers, commonly deals with the utilization of the registered consumption for each customer; besides, some researchers used the economic activity, the active/reactive ratio and the contract power. Currently, utility company databases store enormous amounts of information on both installations and customers: consumption, technical information on the measure equipment, documentation, inspections results, commentaries of inspectors, etc. In this paper, an integrated expert system (IES. It has become necessary for power utilities and researchers to find the root causes and where these losses are mostly performed.

Machine learning (ML) approaches which entail data mining techniques (supervised and unsupervised) have been employed in several capacities to tackle such issues as presented in a recent survey by Glauner. The authors reviewed current works carried out in the field of Artificial Intelligence (AI) [3]. Several of the approaches utilised techniques such as fuzzy clustering [4], neural networks [5], among others. Monendero et al. used regression analysis to investigate the correlation between time and monthly consumption, with a view to detect considerable drops in power consumption [6]. Other studies explored the use of Support vector machines (SVM) algorithm only or in combination with other methods [7]. In another research, [8] analysed local features of customers neighbourhood and split their locations into grids of varying sizes. Each of the grids formed were scrutinised to get the number of customers and quantify the extent of NTLs. A similar study [9] used a generalised additive model to generate the local estimation of NTL and Markov chain to ascertain future changes in the probabilities. Massaferro et. al also conducted a similar study engaging customers' local features to detect NTLs with a random search procedure [10].

Customer segmentation has been making impacts in data mining for managing power related issues [11] such as electricity tariff [12], micro/mini grids [13] and demand side management [14]. These constitute concepts aimed at enhancing the activities and operations of power utilities in supplying uninterrupted power to consumers. Utilities can maximise the gains of these initiatives by targeting groups of consumers with similar payment and consumption patterns through customer segmentation. Therefore, clustering (a data mining technique) has been verified to be very efficient in customer segmentation [15]. K-means, an unsupervised learning algorithm has proven to be very efficient in customer segmentation for its ease of use. Electricity load profiles of consumers were segmented by applying k-means algorithm to understand their consumption behaviour [16], characterize and cluster electricity customers through a electricity demand signature [17], K-means to establish the shopping patterns of customers by segmenting them into five segments and compared results with other clustering algorithms [15]. However, K-means algorithm is not able to measure the density information of the clusters. We therefore seek for a more suitable clustering algorithm like DBSCAN to handle the density an detect outliers given the nature of the data set [18]. Furthermore, [19] proposed Gaussian mixture model (GMM) based clustering algorithms that are sensitive to data and is inconsistent to outliers as a result of their functional dependency on the Euclidean distance measure. Hence the choice of Expectation Maximisation (EM) algorithm to further shows the presence of outliers in the power utility dataset.

This paper explores the capabilities of clustering techniques to assess the severity of NTLs in the Nigerian power sector by considering the consumers' tariff plans and payment patterns. The data will be harnessed to segment consumers and get outliers with the help of the K-means, DBSCAN, and Expectation Maximization algorithms. This is consistent with the research, since detection or assessment of NTL is a case of anomaly or fraud detection.

The data used in the paper was obtained from one of the distribution companies in Nigeria for three consecutive months. The paper is organized as follows: the background review, the methodology used, results and conclusion.

II. BACKGROUND REVIEW

1. The Nigerian Tariff Structures

Electricity Tariff refers to the charges set by power utility for supply of electricity to different categories of consumers. Tariff rates are fixed to accommodate operational and maintenance costs. Utilities will be more operationally stable if the revenue generated covers their expenses. Four factors (location, tariff class, tariff rate and quantity consumed) determine the amount of money one will expend on electricity. The tariff classes are further divided into sub classes depending on the extent of consumption. There are five (5) tariff classes according to electricity usage; **Residential (**R1, R2A, R2B, R4), **Commercial (**C1A, C2B, C3, C4), **Industrial** (D1, D2, D3), **Special** (A1, A2, A3) and Street lights (S1). Impliedly, for a consumer, the variation in charges is directly related to their own consumption.

The concentration of NTL by tariff plan in the Nigerian power sector is depicted in Fig. 1. It can be observed that the maximum impact of NTLs is delineated in residential consumers. This could be ascribed to the circumstances of NTLs committed at residents.

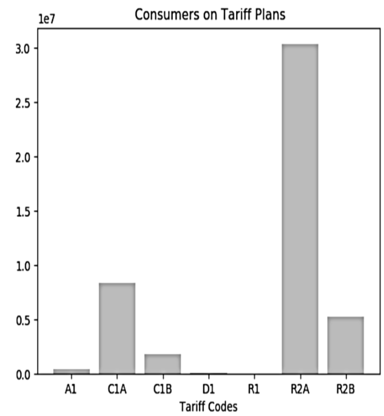


Fig. 1. Extent of Non-Technical Loss by Tariff Plans (Source: Data from Distribution Company in Nigeria)

2. Clustering

Clustering or clustering analysis is an unsupervised data mining technique that separates structured and unstructured data into meaningful groups called clusters based on their similarity [20]. Clusters with similar features or properties are formed. The main objective of clustering data items is to capture the intrinsic meaning in the structure of the data. There many algorithms used in clustering such as K-means, mean shift clustering, DBSCAN, Expectation Maximisation, and Agglomerative hierarchical clustering etc.

1) K-means

K-means is a partition-based clustering algorithm. It partitions data by initialising the positions of K centroids. The centroids are calculated by means of Euclidean distance measure through assigning datapoints to the closest centroid, forming a cluster. This process is repeated iteratively until there is no change in the position of the centroid. Getting the **optimal number of clusters in** a data set is an important issue in partitioning clustering, such as k-means clustering. The methods used to obtain the optimum number of **k** comprise of direct methods (elbow and average silhouette) and statistical testing methods (Gap Statistic).

A) Elbow Method

The main goal of partitioning methods, such as k-means clustering, is to obtain clusters that will minimize the total WSS. A smaller value for the total WSS is always better. The WSS measures the compactness of the clustering. For every k value, k-means is initialised, and the inertia attribute is used to identify the within-cluster sum of squared (WSS) distances of the samples that are nearest to the cluster centroid. As the k values increases, the WSS distance tends to zero. Assuming k is set to the maximum value of n, each sample will form its own cluster, the WSS distances equals zero. An optimal k is obtained if the plot looks like an arm.

2) DBSCAN (**D**ensity-**B**ased **S**patial **C**lustering of **A**pplications with **N**oise)

DBSCAN is a density-based clustering method that is applied in machine learning problems to isolate clusters of high density from clusters of low density. It has proven to be highly efficient in locating areas in the data with high density of observations, versus areas of the data with low density of observations. DBSCAN can organise data into clusters of varying shapes [21]. It focuses on finding neighbours by their densities on an n-dimensional sphere with radius ϵ. *The Core point* (forms a cluster with points reachable from it), *border point* (lies in a cluster and reachable by other points in the cluster) and *outlier* (does not lie in any cluster, not reachable nor connected by density to other points) are different classes of points defined by DBSCAN.

3) *Expectation Maximisation*

The Expectation-Maximization (EM) algorithm finds the maximum-likelihood estimates for model parameters there are incomplete, missing or hidden (latent) variables. It iteratively approximates the maximum likelihood function. It is considered as an extension of k-means algorithm. There are the two basic steps of the EM algorithm, namely E Step or Expectation/ Estimation Step and M Step or Maximization step.

III. MATERIALS AND METHODS

1. Research Framework

Fig. 2 presents a research framework of the problem. This section explains the steps involved in carrying out research.

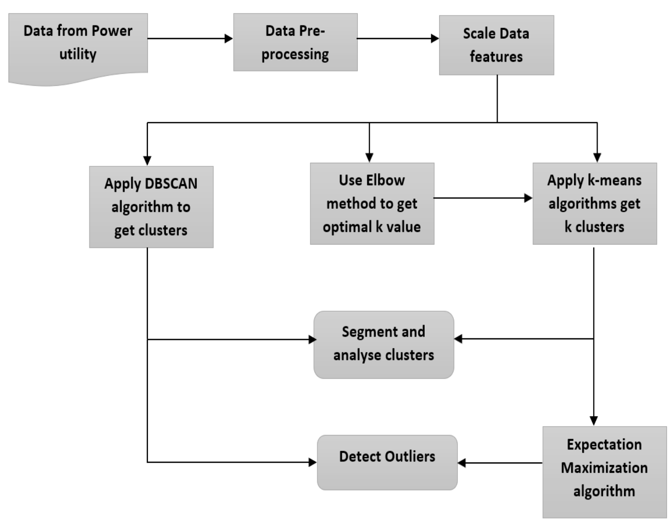


Fig. 2. Research Framework

As depicted in Fig. 2, data was obtained from a distribution company in Nigeria, pre-processed and relevant features extracted for the analyses. The data has been scaled using Standard Scale (z-score) and data gets centred around 0 with a standard deviation of 1.

$$\text{z-score}(x) = \frac{x - mean(X)}{stdev(X)}$$

Where x is the entry in a feature set xi ϵ X, *mean (X) is the mean* of the feature set X and *stdev (X) is the standard deviation of X. The elbow method is then applied to calculate the* value of k for the dataset. The result obtained from the elbow method will influence the efficiency of the k means clustering. Furthermore, the DBSCAN and Expectation Maximization clustering algorithms will be used to segment power consumers in a bid to show the effects of NTLs and to demonstrate the efficiency of the algorithm in showing the density of the clusters and outliers.

IV. EXPERIMENTAL RESULTS

1. System Configuration

The computations were done on a 64-bit operating system and x64 based processor (Intel core i7) computer, with a processing speed of 2.5 -2.9 GHz and RAM capacity of 16 GB. Programs for the analyses were done in python 3 programming language.

2. Results

This paper examines the payment patterns of power consumers by exploring the data as described in the previous section to segment consumers. Using the k-means, DBSCAN and Expectation Maximization clustering algorithms, we categorise the consumers based on their bill payment.

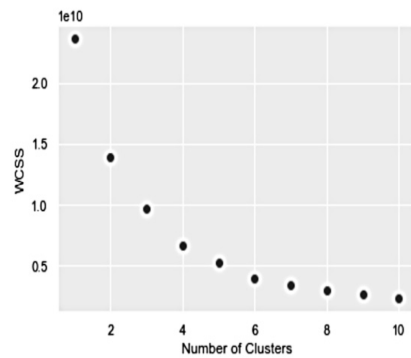


Fig. 3. Graph depicting values of k versus WCSS (Within Cluster Sum of Squares)

From Fig. 3, it is observed that the point of inflection of the curve is at cluster 5. As the number of clusters increase, the score naturally declines. Tt is anticipated that the WCSS rate will decrease once the optimum number of clusters is reached (**k**). Hence the optimal value from our dataset is k = 5.

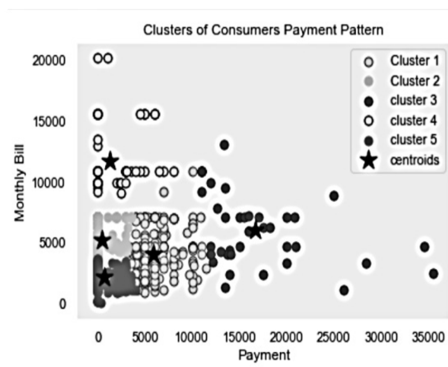


Fig. 4. Payment Clusters

Fig. 4 depicts 5 payment clusters with their centroids obtained from running the K-means algorithm. The cluster analyses are as given:

Cluster 1 appears to be that of consumers who have made some payment to the utility company but have not offset the bill. These consumers are on the Residential and Commercial tariff plans.

Cluster 2 appears to be consumers who have not paid their monthly bill or made partial payment. These consumers are mostly in the Residential tariff plan 2A or 2B.

Cluster 3 appears to be consumers who have been not paid for electricity consumed and their debt has accrued over time. It can be observed that the consumers in this category are largely on Special tariff plan and the others are on Commercial (C1B) plan. It clearly shows that they have a high consumption profile and their refusal to pay for consumed electricity has caused their bill to soar.

Cluster 4 appears to be those very few consumers who have not made any payment, have a huge debt profile and those with a huge debt profile who have made some payments in instalments. A few are on Residential tariff plan and the remaining on Commercial and Special tariff plans.

Cluster 5 appears to be the cluster with the least bill amount, indicating low demand in terms of consumption and one with the highest degree of debt. All the consumers in this category belong to the Residential (2A and 2B) Tariff plan.

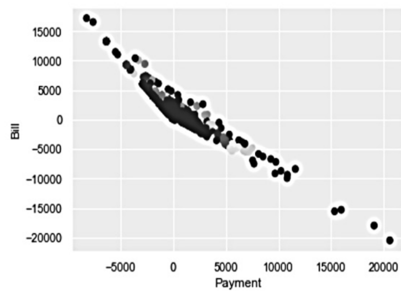


Fig. 5. DBSCAN clusters

It can be noticed from the Fig. 5, that the points along the outer edges of the dataset are not classified; suggesting outliers are present in the data. This indicates extreme cases of debt, implying NTL. The high-density points are clustered together indicating severe cases of non-payment of electricity spread across the tariff plans R2A&B, C1A&B and A1-A3.

The EM algorithm, a type of Gaussian mixture model is an extension of *k*-means algorithm. It makes use of an expectation–maximization approach. It chooses the guesses for the location and shape of the clusters until it converges. Fig. 6 and Fig. 7 shows that each cluster is not associated with a hard-edged sphere, but with a smooth Gaussian model. Like in k-means expectation–maximization approach, the algorithm uses multiple random initializations to form the clusters. The number of clusters to be formed denoted by n was assigned as 4 and 5. The effects show the outliers in both cases.

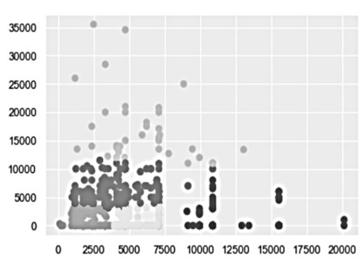


Fig. 6. EM cluster assignment, n=5

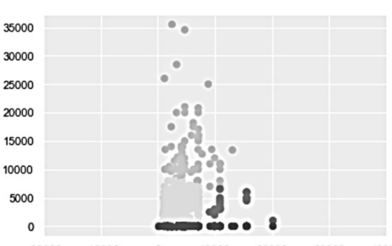


Fig. 7. EM cluster assignment, n=4

Looking at Fig. 8, we can observe that the points at the boundaries between clusters reflect the uncertainty of the cluster assignment. The data was transformed, and the resulting cluster assignment ensued.

Fig. 8. Five components GMM

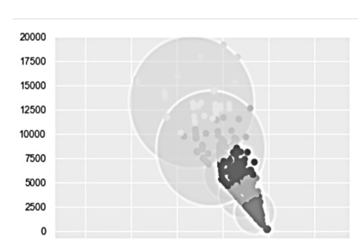


Fig. 9. Stretched Five-Component GMM

In Fig. 9, a five-component GMM for the initial data in Fig. 6 was used to fit the stretched dataset. This allowed for a full covariance for the model to fit the oblong and stretched-out clusters. This clearly shows the outliers, making the EM algorithm a good tool for detecting unusual observations from data sets.

V. CONCLUSION

In this paper, we have presented the strong abilities of clustering techniques and their behaviours concerning spatial data as applied in customer segmentation and outlier detection. K-Means showed five clusters with their centroids, giving a good visualisation of the customer segments. However, we did not achieve appropriate cluster shapes to enable the proper detection of outliers. DBSCAN and on the other hand, portrayed dense clusters indicating outliers (extreme cases of NTL) and noise in the data. We observed a dense cluster which highlighted the severity of non-payment of electricity by power consumers. EM gave clearer clusters (Fig. 8 and Fig. 9). The resulting clusters gave an insight into the quantification of NTLs in the Nigerian power sector and where it is greatly perpetrated. The study has indicated the suitability DBSCAN and Expectation Maximization clustering algorithms in the assessment and detection of NTLS. It is hoped that this research will help utilities in carefully and decisively managing defaulters. Customer segmentation of electricity consumers can be explored for the provision of micro grids to deserving domains. It is also important for the utilities to identify customers that will lead them to a financially sustainable system.

REFERENCES

[1] C. León, F. Biscarri, I. Monedero, J. I. Guerrero, J. Biscarri, and R. Millán, “Integrated expert system applied to the analysis of non-technical losses in power utilities,” *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10274–10285, 2011.

[2] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, “Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft,” *Energy Policy*, vol. 39, no. 2, pp. 1007–1015, 2011.

[3] P. Glauner, J. A. Meira, P. Valtchev, R. State, and F. Bettinger, “The challenge of non-technical loss detection using artificial intelligence: A survey,” *Int. J. Comput. Intell. Syst.*, vol. 10, no. 1, pp. 760–775, 2017.

[4] E. W. S. Dos Angelos, O. R. Saavedra, O. A. C. Cortés, and A. N. De Souza, “Detection and identification of abnormalities in customer consumptions in power distribution systems,” *IEEE Trans. Power Deliv.*, vol. 26, no. 4, pp. 2436–2442, 2011.

[5] S. Golnaraghi, Z. Zangenehmadar, O. Moselhi, S. Alkass, and A. R. Vosoughi, “Application of Artificial Neural Network(s) in Predicting Formwork Labour Productivity,” *Adv. Civ. Eng.*, vol. 2019, 2019.

[6] I. Monedero, F. Biscarri, C. León, J. I. Guerrero, J. Biscarri, and R. Millán, “Using regression analysis to identify patterns of non-technical losses on power utilities,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6276 LNAI, no. PART 1, pp. 410–419, 2010.

[7] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and F. Nagi, “Improving SVM-based nontechnical loss detection in power utility using the fuzzy inference system,” *IEEE Trans. Power Deliv.*, vol. 26, no. 2, pp. 1284–1285, 2011.

[8] P. Glauner, J. A. Meira, L. Dolberg, R. State, F. Bettinger, and Y. Rangoni, “Neighborhood features help detecting non-technical losses in big data sets,” *Proc. - 3rd IEEE/ACM Int. Conf. Big Data Comput. Appl. Technol. BDCAT 2016*, pp. 253–261, 2016.

[9] L. T. Faria, J. D. Melo, and A. Padilha-Feltrin, “Spatial-Temporal Estimation for Nontechnical Losses,” *IEEE Trans. Power Deliv.*, vol. 31, no. 1, pp. 362–369, 2016.

[10] P. Massaferro, H. Marichal, M. Di Martino, F. Santomauro, J. P. Kosut, and A. Fernandez, “Improving electricity non technical losses detection including neighborhood information,” *IEEE Power Energy Soc. Gen. Meet.*, vol. 2018-Augus, no. i, 2018.

[11] Y. Yuan, K. Dehghanpour, F. Bu, and Z. Wang, “A Data-Driven Approach for Estimating Customer Contribution for System Peak Shaving,” pp. 1–8, 2018.

[12] R. Nafkha, K. Gajowniczek, and T. Zabkowski, “Do customers choose proper tariff? empirical analysis based on polish data using unsupervised techniques,” *Energies*, vol. 11, no. 3, 2018.

[13] N. J. Williams, P. Jaramillo, K. Campbell, B. Musanga, and I. Lyons-Galante, “Electricity Consumption and Load Profile Segmentation Analysis for Rural Micro Grid Customers in Tanzania,” *2018 IEEE PES/IAS PowerAfrica, PowerAfrica 2018*, pp. 360–365, 2018.

[14] A. Cominola, E. S. Spang, M. Giuliani, A. Castelletti, J. R. Lund, and F. J. Loge, “Segmentation analysis of residential water-electricity demand for customized demand-side management programs,” *J. Clean. Prod.*, vol. 172, pp. 1607–1619, 2018.

[15] T. Kansal, S. Bahuguna, V. Singh, and T. Choudhury, “Customer Segmentation using K-means Clustering,” *2018 Int. Conf. Comput. Tech. Electron. Mech. Syst.*, pp. 135–139, 2019.

[16] L. G. Viola, “Segmentation of household load-profiles with K- means clustering algorithm Segmentation of household load-profiles with K-means clustering algorithm,” no. July, pp. 0–9, 2016.

[17] A. Camero, G. Luque, Y. Bravo, and E. Alba, “Customer segmentation based on the electricity demand signature: The andalusian case,” *Energies*, vol. 11, no. 7, pp. 1–14, 2018.

[18] P. Wang and M. Govindarasu, “Anomaly Detection for Power System Generation Control based on Hierarchical DBSCAN,” *2018 North Am. Power Symp. NAPS 2018*, no. September, 2019.

[19] P. A. Forero, V. Kekatos, and G. B. Giannakis, “Robust Clustering Using,” *IEEE Trans. Signal Process.*, vol. 60, no. 234914, pp. 4163–4177, 2012.

[20] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Second edi. San Francisco: Morgan Kaufmann Publishers, 2006.

[21] J. Jang and H. Jiang, “DBSCAN++: Towards fast and scalable density clustering,” 2018.

Re-sampling Technique for Joint Motion Normalization

Kiran Chawla[1], Chandra Prakash[2]

Indira Gandhi Delhi Technical University for Women, Delhi, India
[1]kiranchawla09@gmail.com, [2]cse.cprakash@gmail.com

Abstract: The effect of the anthropometric parameters on the gait profile is an active research area. Gait profiling can be used in different fields such as surveillance, biometrics, sports, medical, rehabilitation, animation industry, etc. Walking pattern is unique to an individual, but for an efficient gait profile model, the gait cycle duration should be on the same scale. The subjects considered in this study are from the "MNIT Gait dataset." They illustrate different duration in the gait kinematic output. For further analysis using computational techniques, the data must be re-sampled. In the study, Polynomial Curve Fitting (PCF) has been proposed for re-sampling the time series curves of Gait kinematics with variable sizes. This aids in expanding or contracting the time series data in a supposed range. The result illustrates the importance of the proposed work.

Keywords: Gait analysis, Re-sampling, Joint motion, Gait kinematics, Polynomial Curve Fitting

I. INTRODUCTION

The study of human gait kinematics (joint angles, Spatio-temporal, etc.) has come into focus as the applications of the gait predictions are being explored vastly in the field of medicine, sports bio-mechanics, biometrics and many more [1-3]. Along with the earlier methods of gait prediction, including model-based optimization techniques, the advancement of technology has simplified the way of functioning. The models can independently adapt to the introduction of new data. They learn from previous experiences to give the results with accuracy.

The parameters in the gait analysis can be extracted using kinematics, kinetic, and EMG based methods [4]. The data used for the gait prediction is the time stamp data of multiple individuals, with variable walking styles, velocity, stance, and swing phases. Using these gait parameters, one can propose the model for the gait pro ling prediction. For applying any algorithm to make the predictions and to gather results from the data, we need that the data is in the same domain, that is, the data needs to be in a particular scope in which the data values may lie. It is difficult to predict a result with the variable range of the data. Taking the example of gait prediction, it is a possibility that a particular subject finishes a gait cycle in 56 units/samples of time/frames, another subject completes a gait cycle in 60 units/samples of time/frames, now if we are predicting the gait cycle of another subject we do not know if the prediction should be for let say 56 units or for 60 units. Even for the same individual, the time covered for a gait cycle may vary in multiple trials, depending upon the emotional state, posture, type of clothes and footwear, etc. [5, 6]. The time to complete a gait cycle for each subject would be different, and so to predict some results based on this data, we need to compress or expand the data in a particular range [6]. Thus there is the need for re-sampling the data.

This random time sample problem is prevalent in the medical field. Statistical resampling methods such as Bootstrap, seasonal block bootstrap, and Jackknife are some of the methods used to resolve this limitation [7]. In 1956, John Tukey gave the concept of Jackknife. It is sampling without replacement [8]. Bradley Efron in In 1979, proposed the bootstrap method as the non-parametric re-sampling technique for estimating the sampling distribution of an estimator. In 2003 Hardle et al. used bootstrap methods for Time Series data[7]. Bootstrap is a re-sampling technique used for independent random sample or a time series. It is used in resampling the data and model estimation from the data. Dudek et al. in 2014 used seasonal block bootstrap for seasonal time series [8]. It is a direct and intuitive approach when time-series data contain a periodic/seasonal component, and seasonal block bootstrap is used.

In this study, Vandermonde matrix-based re-sampling technique for human joint motion is proposed. Section II discusses the methodology used for the re-sampling of the data. Section III presents the result followed by the discussion and future scope in section IV.

II. METHODOLOGY

In this section, the methodology used in this study is presented. First, the description of the dataset used is presented. The methodology used in the study is illustrated in figure 1.

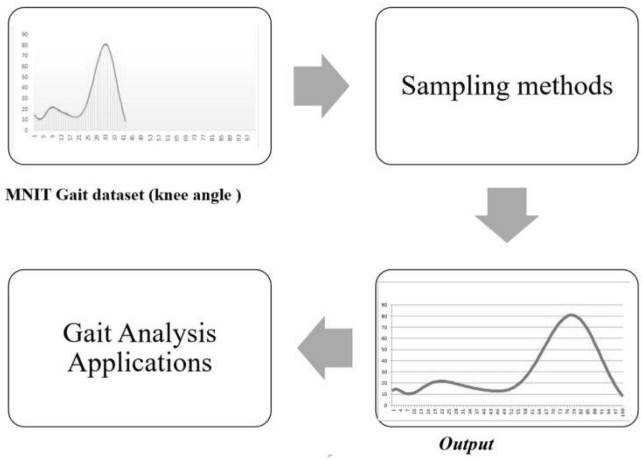


Fig. 1. Methodology proposed for resampling of the gait joint parameters

1. Dataset

The dataset used in this study is taken from the MNIT Gait dataset carried out at Robotics And Machine Analytics Laboratory (RAMAN Lab), MNIT Jaipur, INDIA [9]. Optical Motion Analyzer developed during the study at RAMAN LAB is employed to collect Indian gait norms of 120 subjects from different age groups and genders using the Nikon D5200 DSLR camera. The range of the population of the dataset is from 5-60 years. All of the subjects had no record of injury or pain to the lower limbs. The study was designed following the Declaration of Helsinki on experiments involving human beings. Before participating, all subjects considered for the study signed a consent form.

The result of the proposed system provides quantitative kinematics gait parameters i.e., joint angles. In this model, joint coordinates are used to calculate trunk, thigh, knee, and foot angle. Knee angle of MNIT Gait dataset is illustrated in figure 2. The MNIT Gait dataset consisted of the knee angle data of 120 subjects for the respective unit gait cycle. Figure 2 illustrates the time series nature of the knee plot.

The method we are using to resample data is Polynomial Curve Fitting (PCF) along with the translation, which is explained in the further sections. Considering the mean square error value of all the subjects, when re-sampling is done, the maximum error was corresponding to S107, the minimum error was corresponding to S106, and the average of the MST of all the subjects was approximately 0.2223; hence the nearest value to the standard is corresponding to S85. There-fore, in the following study, the data corresponding to subjects S85, S106, S107 are

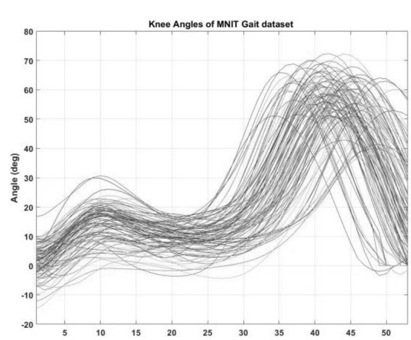


Fig. 2. Aggregate knee angle of 120 subjects of MNIT Gait Dataset

taken into the examination. As observed in Figure 3, the gait cycle of different subjects terminate at different times- S85: 42 units; S106: 58 units; S107: 56 units. The re-sampling of the data is necessary to proceed further with the analysis and to draw conclusions based on it. The aim is to complete the gait cycle of the individual subject in a limited time interval, here in particular 100 units. This will result in uniform gait cycles for all the subjects. For applying any prediction algorithm, the input and output size is fixed, and to achieve this, there is the need to pre-process the data to get the output.

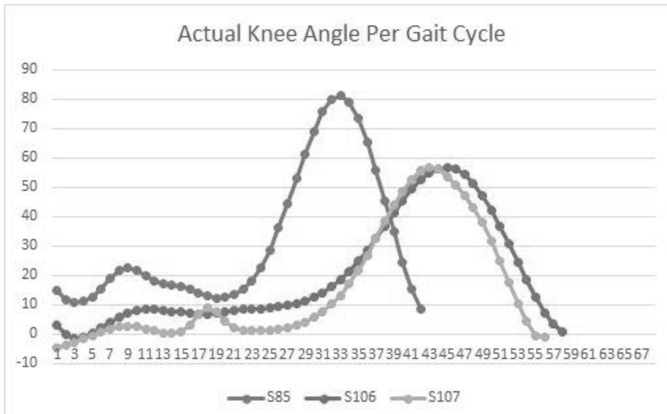


Fig. 3. Knee angle of S85, S106 and S107

In the study, the idea is of horizontal scaling to re-sample the data that one may proceed further by applying relevant algorithms for analyzing the patterns of the data. This is achieved by multiplying or dividing every x-coordinate by a constant while leaving the y-coordinate unchanged. Since horizontal scaling is a property of polynomial functions, we are aiming to nd the best t polynomial function for the time series data of the gait cycle of each subject.

2. Polynomial Curve Fitting (PCF)

A polynomial is an expression that consists of variables and coefficients involving only the operations of addition, subtraction, and multiplication. The concept behind polynomial curve tting is to decide the degree of the polynomial that ts the curve in. Hit and trial and plot visualization are used to determine the degree of the curve. The polynomial looks like

$$p(x) = a_1x^n + a_2x^{n-1} + a_3x^{n-2} + \cdots + a_nx + a_{n+1} \tag{1}$$

The aim is to find the most accurate values for a1; a2; a3; ; an; an+1 corresponding to the actual coordinates given, resulting in the polynomial equation that best fits the actual curve. The approach uses the method of least squares for estimation of true value, which is majorly used in regression analysis [10]. The equation to t the actual curve in a one-degree polynomial in a straight line can is represented as equation 2.

$$y = a_1x + a_2 \tag{2}$$

x,y in equation 3 and 4 are the coordinates in the actual curve. In this method, we need to solve these equation 3 and 4 to get the values of a_1; a_2, which, when substituted in equation 2, give a polynomial function.

$$\sum y = \sum a_1x + na_2 \tag{3}$$

$$\sum xy = a_1 \sum x^2 + a_2 \sum x \tag{4}$$

In the case of the parabola, the cure is a two-degree polynomial as illustrated in equation 5

$$y = a_1x^2 + a_2x + a_3 \tag{5}$$

Equation 5 has three unknowns, that is, a1; a2; a3 , that can be given as:

$$\sum y = a_1 \sum x^2 + a_2 \sum x + na_3 \tag{6}$$

$$\sum xy = a_1 \sum x^3 + a_2 \sum x^2 + a_3 \sum x \tag{7}$$

$$\sum x^2y = a_1 \sum x^4 + a_2 \sum x^3 + a_3 \sum x^2 \tag{8}$$

values of a_1; a_2; a_3 when substituted in equation 5 give a two-degree polynomial which is the t for actual coordinates of the curve.

Degree of the polynomial (n) should be known to best t the coordinates of the curve. For the higher degrees of n, we use the Vandermonde matrix [11] with *(n + 1)* columns and *m = len(x)* rows. The applications of Vandermonde Matrix is in polynomial interpolation for evaluating the coefficients of the polynomial [11]. The Vandermonde matrix is shown as equation 9:

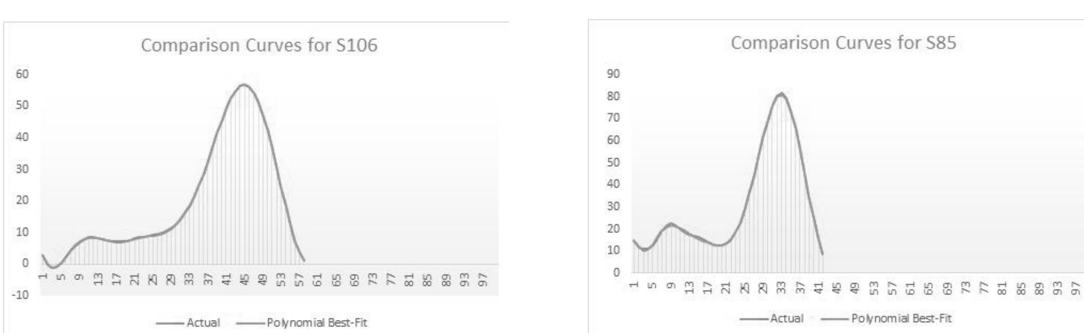
$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix} \tag{9}$$

Vandermonde Matrix is a m n matrix where each row is a geometric progression. This matrix evaluates the polynomial at a set of points. The property of the Vandermonde Matrix used is that the determinant of this matrix if m = n is presented as equation 10

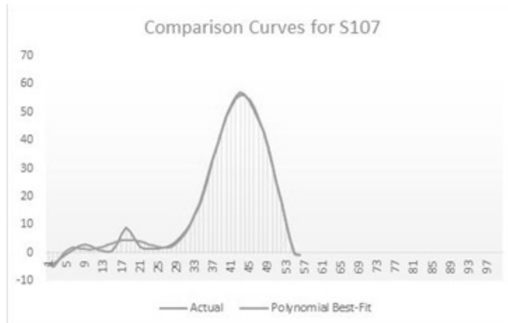
$$dev(v) = \prod_{i \le i \le j \le n} (x_j - x_i) \tag{10}$$

On solving with the above equation, the return value is a matrix with the coefficients (in descending order of power) of the n-degree polynomial function, i.e., a_0; a_1; a_2 …a_{n-1}; a_n. In this particular data set by hit and trial, the degree of the polynomial is set as 12. When applied on the given dataset, the obtained curve for S85 as shown in figure 4a fit the data with mean square error value to be 0.2291, the least mean square error value observed is for S106 as shown in figure 4b and the maximum is for S107 as shown in figure 4c. To apply this recursively on the MNIT gait dataset, we need to decide the value of n (degree) that is the best t for most of the data points.

The fitted curves have a mean square error value ranging from 0.0021-1.5354 and the average mean square error value as 0.2291 for all the subjects.



(a) (b)



(c)

Fig. 4. Two dimensional plot of knee angle using PCF based re-sampling for
(a) Subject S85; (b) Subject S106; and (c) Subject S107.

III. RESULT

This section presents the result of the proposed approach. Using the mathematical operations of multiplication and divide, we changed the domain of the timestamp data into (0-100); that is, in this particular case, the time per gait cycle is made to 100 units.

1. Translation

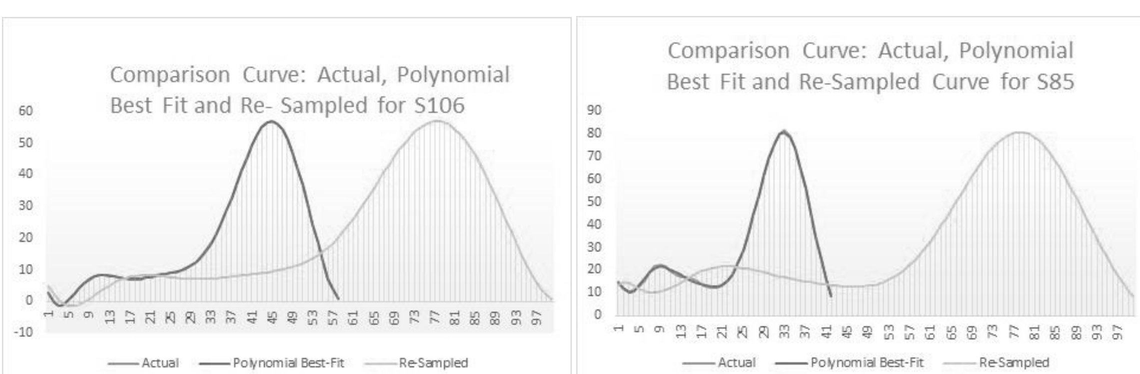
Since we got the equation of the polynomial that best fits the curve, now we are interested in stretching the curve; that is, let say the x-axis value for S85, which was actually at 43 units, would now be extended to 100 units and similarly for all other subjects. This is done by the property of translation of graphs, specifically using multiplication and division to expand the x-axis while keeping the y values the same. This allows us to plot the x-coordinates at best possible y-coordinates in the re-sampled curve such that the plot does not change significantly.

$$x_1 = \frac{x_3}{x_2} \times 100 \tag{11}$$

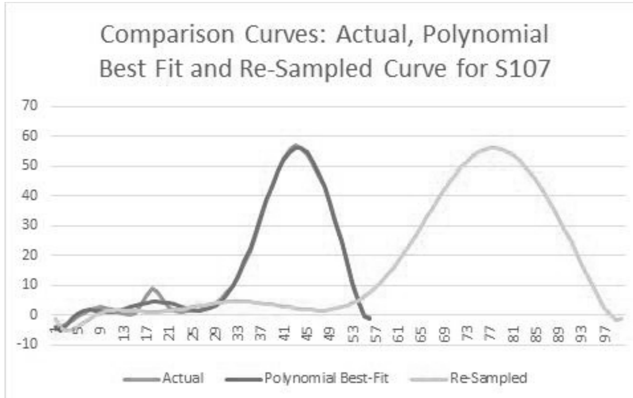
where, x_1 is desired value after re-sampling, x_2 represents total number of values on x-axis (domain) and x_3 is a number in the range (0, x3)

To get the coordinates of the re-sampled curve, we evaluate all the coordinates corresponding to the plotted graph. The value of the polynomial is estimated for the given range of x, and the values of y are returned. In our particular case, we are expecting the values in the scale of 0-100. By translation and fitting the values in the polynomial, we get the values of expected results.

Figure 5 shows the output of the proposed approach after translation for the subject S85, S106, and S107.



(a) (b)



(c)

Fig. 5. Two dimensional plot of knee angle using Polynomial Curve Fitting based re-sampling for (a)Subject S85; (b) Subject S106; and (c) Subject S107.

To cross-validate the proposed approach, three validation measures (top peak vs x ratio, the position of peaks, and parametric test) have been considered in this study.

Test of Peaks to the unit ratio: The top four peak values of each subject in the actual curve and the re-sampled curve are taken, the ratio of the x-coordinates of peaks of the actual curve and that of peaks of the re-sampled curve to x-axis is calculated. The diffierence of the stated ratios for the four peak values comes out to be 0.0338, 0.0272, and 0.0278 for S85, S106, and S107, respectively.

Test of Position of Peak: The peak of the curves should lie on the same y-coordinate value to maintain the nature of the graph. The average of the top four peaks y coordinate values are 1.6721 for S85, 0.5392 for S106, and 0.7096 for S107.

Parametric test (Z-Test): If the information about the population is completely known by mean of its parameters then a statistical test is called parametric Test. Z test is used when the data-sample size is more than 30. Z-test for two populations implies the rate of how much one population is similar to the other population. In this, it is assumed that the two populations have a null hypothesis; that is, there is no significant difference between the populations. By p-value, it is signified that how much the assumption is correct. Z score is the depiction of the standard deviation from the mean; negative values depicts that the value of standard deviation is below the mean whereas positive shows the standard deviation above the mean. The formula for Z-test is as follows :

Where: x is mean of the sample, is mean of the population, is the standard deviation of the population, and n is the total number of observation of in the population.

The results for the Z-test on the actual dataset and the re-sampled dataset is shown in table 1. Two tail z-test was performed and the value of the p is less then 0.025 thus it can be concluded that the null hypothesis is correct. It presents the fact that the plot after the re-sampling followed by translation is similar to the original plot.

Table 1. Z-test on the actual dataset and the re-sampled dataset

Subject Z	-Score p	-value
S85	-0.009	0.9928
S106	0.007	0.9945
S107	0.04	0.9677

This re-sampled plot now can be used for further gait analysis. It can be used to predict the gait pro le models using different computational techniques that can assist in the rehabilitation and pathology related decisions.

IV. DISCUSSION AND FUTURE SCOPE

Re-sampling and normalization play a vital role in the analysis of the data. In this study, Polynomial Curve Fitting based re-sampling technique has been proposed for the knee angle of the MNIT gait dataset. Different validation measures prove that the proposed method returns the considerably accurate result to solve the problem of re-sampling the data before using it for further analysis.

There is a scope of improvement by providing a particular algorithm for finding the most appropriate degree for the polynomial rather than trying by hit and trial. Polynomials are oscillatory functions and are unbounded, so increasing and decreasing the range of the data might affect the evaluations. The result obtained from this can be used for the gait pathology and rehabilitation decisions.

ACKNOWLEDGMENT

The author are very thankful to the RAMAN Lab for providing the dataset.

REFERENCES

[1] L. Angelini, W. Hodgkinson, C. Smith, J. M. Dodd, J. Ray, F. Storm, C. Mazza, B.Sharrack, and D. Paling, \Automated gait analysis using wearable sensors can detect objective and reliable quanti cation of disability in people with multiple sclerosis (p3. 2-004)," 2019.

[2] C. Ricciardi, M. Amboni, C. De Santis, G. Improta, G. Volpe, L. Iuppariello,G.Ricciardelli, G. D'Addio, C. Vitale, P. Barone et al., \Using gait analysis pa-rameters to classify parkinsonism: A data mining approach," Computer methods and programs in biomedicine, vol. 180, p. 105033, 2019.

[3] C. Prakash, R. Kumar, and N. Mittal, \Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges," Arti cial Intelligence Review, vol. 49, no. 1, pp. 1-40, 2018.

[4] E. Rastegari, S. Azizian, and H. Ali, \Machine learning and similarity net-work approaches to support automatic classi cation of parkinsons diseases using accelerometer-based gait analysis," in Proceedings of the 52nd Hawaii International Conference on System Sciences, 2019.

[5] F. Menegoni, M. Galli, E. Tacchini, L. Vismara, M. Cavigioli, and P. Capodaglio, \Gender-speci c e ect of obesity on balance," Obesity, vol. 17, no. 10, pp. 1951-1956, 2009.

[6] Y. Yun, H.-C. Kim, S. Y. Shin, J. Lee, A. D. Deshpande, and C. Kim, \Statistical method for prediction of gait kinematics with gaussian process regression," Journal of Biomechanics, vol. 47, no. 1, pp. 186-192, 2014.

[7] B. Efron, \The jackknife, the bootstrap, and other resampling plans. philadelphi," 1982.

[8] J. W. Tukey, \Bias and con dence in not quite large samples," The Annals of Mathematical Statistics, vol. 29, no. 2, p. 614, 1958.

[9] C. Prakash, \Machine learning techniques for vision based human gait analysis."

[10] R. H. Myers and R. H. Myers, Classical and modern regression with applications. Duxbury press Belmont, CA, 1990, vol. 2.

[11] k. Bjorck and V. Pereyra, \Solution of vandermonde systems of equations," Math-ematics of computation, vol. 24, no. 112, pp. 893-903, 1970.

Analysis of Approaches to Detect Fake Twitter Accounts

Shubham Singh[1], Deepti Aggarwal[2]

[1]Dept. of Information Technology, North Eastern Hill University, Shillong-793022, India
[2]Dept. of Computer Science and Engineering, IMS Engineering College, Ghaziabad-201015, India

Abstract: Social networking sites have become immensely popular among the people as they allow people around the globe to connect to each other. Their social lives are more associated to these sites. But there are malicious users who are utilizing these sites to induce lawful users for various purposes like promoting different products, to enter spam links etc. This is emerging out as a serious problem in social networks. Therefore detection of such users and their removal is necessary. A lot of work has been done for the same which uses different approaches for identification and eradication of such users. This paper presents a comparative analysis of blacklist method with the approaches based upon machine learning to detect fake Twitter accounts.

Keywords*:* Twitter, Machine learning, Fake account, Blacklist, Detection**.**

I. INTRODUCTION

Social networking sites have grown tremendously in past few years. Such websites, for example, Twitter, Instagram, Facebook etc. are used by a large number of people around the globe to keep in touch with their family and friends. These sites have created many online activities which attract the interest of a number of users where they increasingly depend on the credibility of the information exposed on a social network [1]. People use such sites to share news, organize and promote various events, running an online business etc. For this they have to share their personal data with the networking site and the open nature of such networks makes them vulnerable to security attacks [2].

Micro blogging website Twitter is a popular web destination. It is used by a number of organizations & individuals as a primary tool of communicating with each other. It allows users to multicast short messages known as tweets. Such tweets are used to spread ideas through society and hence are seen as a tool by malicious users. There are a number of monthly active users on Twitter which are undesirable, represented by various fake accounts. Such accounts have been created by violating the terms of service of the social networking site and are used to promote undesirable content like hate speech, graphic violence, terrorist propaganda, adult & sexual activity etc. [3]. Attackers deceive themselves off as somebody else to spread malicious content and steal personal information. Their motive would be anything other than good intentions as they show eagerness to phish sole ingenuous users to phony relationships which mostly leads to human trafficking, scam or any other activity which can be considered as harmful [4-5].

Twitter has grown from thousands in 2007 to hundreds of millions in 2019. Due to a large number of users and a relative ease of data access, there is a need for the identification and removal of fake accounts. This paper presents a comparative analysis of blacklist method [6] and machine learning based approaches [7] which are used to serve this purpose.

II. RELATED WORK

A lot of work has been done to detect fake accounts on Twitter. An analysis is shown in [8] about the various ways in which spammers operate on different social networking sites. As per [9], a fake user account on Twitter is considered as one form of deception. Authors in [10] have observed that as more and more researchers are progressing to keep Twitter spam free, more and more Twitter spammers are evolving to avoid detection techniques. Thus new methods are being researched to tackle this problem in this micro blogging website.

There exist methods which rely on graph theory [11] but spammers have developed alternate ways in which they create user accounts. They are now creating user accounts which have all the user information like complete profile details, background information etc. Such fake user accounts are identical to the accounts of authentic users [12]. This has complicated detection efforts and hence continuous research is needed to develop new spam recognition approaches.

Honeypot techniques have been developed which attract spammers to their websites and identify them. They get embedded to their network and harvest necessary information from them [13]. The research is also utilizing machine learning techniques to identify Twitter accounts which are not legitimate. There are different schemes which are developed using machine learning in such a way that these approaches extract unique features from recent activities of a user like frequency of tweets, follow request duration etc. and later applies them to a classifier which is built and trained using machine learning algorithms [14], [15], [16], [17]. Approaches developed using machine learning uses profile information of various spammers. Such information is obtained through the use of approaches like honeypot harvesting etc. so as to train the algorithms to understand the behaviors of spammers & thus helping out in developing fake account detection techniques [18]. There are approaches to identify fake profiles and spammers based on tweet content [19], characteristics of tweet like duplicate tweets, url count etc. [20]. [21] puts forward an approach based on the analysis of features like user profile etc. to identify fake accounts. A supervised machine learning pipeline is used in this approach to compare the frequencies of text in the features of tweet such as email etc. to classify the user account as genuine or fake.

III. PROBLEM STATEMENT

With the tremendous growth of Twitter a number of people have connected themselves to it. It has been used by them to share information about their day to day life, promote events and so forth. This rapid growth and massive amount of subscriber's personal data has attracted imposters and attackers to steal this secret data and perform malicious activities. They create fake accounts and later use them to share malicious links, to spread fake news so as to create social unrest, and to persuade licit users for unalike purposes like promoting various products, to enter spam links etc. Mostly, the genuine users are not aware of these fake accounts and hence easily become a victim of such malicious users. These illegitimate users are a major obstruction in freely using the services offered by Twitter or any other networking site. For that reason, the detection of such fake accounts on Twitter and their removal is essential for all who is a user of it.

This paper put forwards a comparative analysis of two such schemes which are used for the identification of fake accounts on Twitter, with a performance evaluation in the end. The first scheme proposed in [6] uses blacklist method while the scheme in [7] uses different machine learning techniques.

IV. PRELIMINARIES

There are several approaches that can be used for the detection of fake and malicious user accounts. All of these approaches use some common features to serve the purpose. There are approaches which uses unique features like the rate of sent friend requests, accepted requests etc., all extracted from recent user activities. These features are helpful in classifying the users of website into legitimate or not with a very good false positive rate.

But high dimensional data can be a vital problem for a number of classification algorithms due to high memory usage & computational cost. Therefore reducing it is necessary which will also reduce the amount of noisy data & redundant features, & will lead to a classification model much better then the existing one. Different feature reduction techniques like Principal Component Analysis etc. can be used to serve the purpose and then a model can be built for the identification of fake user accounts.

V. BLACKLIST SCHEME

This section deals with the blacklist scheme [6] which is used for the detection of fake Twitter users. In the proposed scheme, the creation of blacklist involves following steps, namely, data collection, preprocessing, topic extraction, & keyword extraction, which are illustrated in fig. 1.The proposed blacklist scheme uses features based on the content of tweet for the detection of fake users. After extracting these features from the content of tweet, Decorate classifier is used on them to classify between a real & fake user. Fig. 2 showcases the process flow of the proposed scheme.

In the proposed approach, fake accounts are detected on the basis of a users tweet content, profile information, and his network of followers & followings. But there are users who hide their profile details for the sake of privacy. Such users only show their tweet to the public. Thus extracting their profile based & other features consumes extra cost & more time. Thus the proposed blacklist scheme extracts features from the publicly available content of the user. In order to identify fake user accounts, fourteen features, based on content, have been extracted and are as follows:

- Total number of words
- Total number of tweets
- Total number of retweets
- Total number of mentions
- Total number of urls
- Total number of hashtags
- Total number of fake words
- Fake word ratio = $\frac{\text{Total number of fake words}}{\text{Total number of total words}}$
- URL ratio = $\frac{\text{Total number of tweets which contains urls}}{\text{Total number of tweets}}$
- Hashtag ratio = $\frac{\text{Total number of tweets having hashtags}}{\text{Total number of tweets}}$
- Mention ratio = $\frac{\text{Total number of tweets having mentions}}{\text{Total number of tweets}}$
- Idle duration time in the middle of tweets
- Mean time in the middle of tweets
- Standard deviation time in the middle of tweets

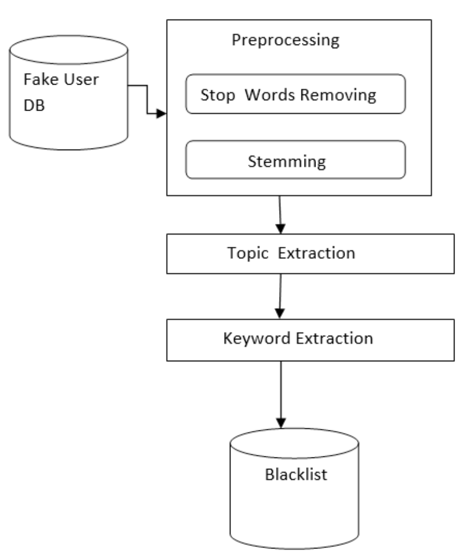


Fig. 1. Blacklist creation

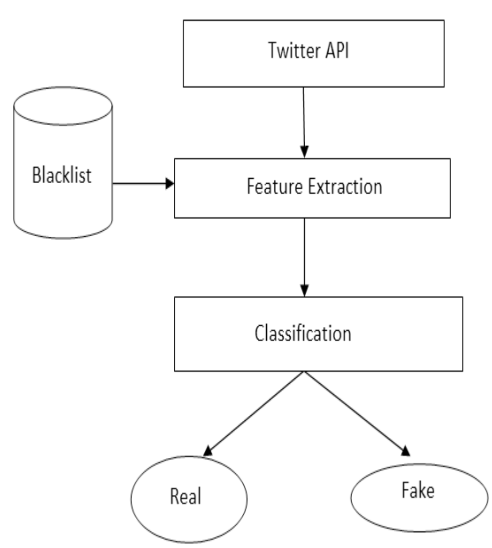


Fig. 2. Process flow of proposed scheme

After extracting above mentioned features, DECORATE (Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples) is used for the identification of fake accounts. It generates the ensemble iteratively by taking in a classifier & then adding it to the existing one. The process works as follows: at first, the ensemble contains a classifier which is trained using given training data. Then, in each consecutive iteration, classifiers are trained using the original data in combination with the artificial one from the previous iteration. These artificial examples contains labels which are chosen in a way different from the one used for the current ensembles prediction. Labeled artificial data is the diversity data [23]. This diversity data along with the original training data are combined together & trained and achieved in a diverse ensemble. The addition of this classifier to the present ensemble increases its diversity. More diversity leads to more accuracy. But if the classifier decreases the accuracy it is removed. This process is repetitive & ends if it reaches the required committee size & maximum iteration.

VI. MACHINE LEARNING APPROACH

This section deals with the scheme based on machine learning techniques as proposed in [6], which is used for detecting fake user accounts on Twitter. In order to evaluate these algorithms, a dataset of fake users is needed to successfully detect fake accounts amongst a collection of users. According to Twitter policy [24], a fake user account is one which follows a number of other user accounts in a very short duration of time and the tweets done by such an account mainly consist of a number of urls or trending hash tags when any irrelevant information is posted. A genuine or the legitimate user is someone who almost has an equal number of followings and followers, & who tweets frequently. Users having a large number of followers & tweets have been categorized as famous personalities or big organizations. A subset of such Twitter users has been selected & classified.

Twitter 4j has been used to build a crawler in order to obtain a sample of real and fake users on Twitter [25]. Twitter REST API is used to gather a publicly available data set which works by making a hunt for certain type of data. Therefore details like user ids, screen name, followers etc. of users has been obtained and is encoded using JSON. A limit is associated for calls to API which is 350 requests per hour per host [26]. Collected sample is then used to obtain features based on graph like number of tweets, followings, and followers etc. to develop test data set. With the help of these obtained features, an analytical model [27] having equations for two parameters, namely, test score & user score, is used to categorize the users as follows:

$$\text{User score} = \left(\frac{10}{\Phi}\right) * (\log\beta + \log\delta) \tag{1}$$

$$\text{Tweet score} = \left(\frac{10}{\Omega}\right) * (\log\alpha + \log\Upsilon + \Psi) \tag{2}$$

Where Ω, Ψ & Φ are taken to be as 6.1, 3.4 & 8.4, and

α = Total number of tweets,

β = Total number of followers or followings,

δ = Total number of followers and followings, and

Υ = Tweet frequency per day

User score and test score have been calculated in the range [0-10] for the collected sample & is then stored. To classify Twitter users, four groups have been defined. Group one consists of users having the value of these scores less than one. This group includes users having a large number of followings and fewer tweets. The users belonging to this group have been classified as malicious. Group two consist of users having user score less than one while tweet score greater than one. The users of this group have a large number of followings and tweets, and have been categorized as malicious. Group three consist of users having user score and test score in the range [1-5]. These users have been identified as not malicious as they have an almost equal number of followers & followings and they tweet frequently. Group four includes those users who have user score and test score more than five. The members of this group follow lesser people than their followers & tweet more, so are considered as big organizations or celebrities.

After all the users have been classified into different groups, a dataset of users is obtained which is to be used for training and testing purpose. Different classification algorithms are used for the detection of legitimate and illegitimate users. These algorithms learn classification from the data which has been classified previously and then uses this knowledge to classify unknown users into intended categories. Weka toolbox is used for the classification purpose. Five classification algorithms, namely, BayesNet, NaiveBayes, SMO, J48 and Random Forest are used, and are compared on the basis of evaluation metrics [28-29].

VII. RESULT AND PERFORMANCE ANALYSIS

The proposed schemes are successfully implemented using Java on Windows machine with Intel core i3 processor TM-4005U running at 1.70 GHz and 4G memory.

The blacklist scheme uses 1KS-10KN dataset which contains 11000 users and their 1354616 tweets. Tweets which are written in English are only taken into consideration. Rest all are ignored. In the blacklist scheme, DECORATE classifier is used for detecting fake user accounts whereas in machine learning approach, Weka toolbox is used for the classification purpose in which five algorithms have been used, namely BayesNet, NaiveBayes, SMO, J48 and Random Forest, whose respective accuracy is shown in fig. 3. The results of both the proposed schemes are shown in fig. 4.

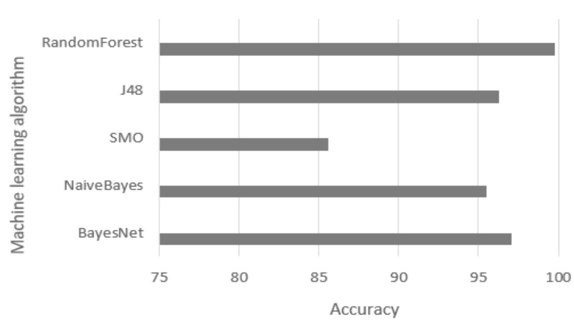


Fig. 3. Machine learning approach

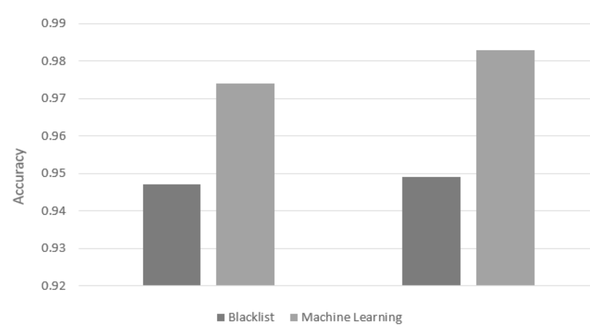


Fig. 4: Blacklist vs machine learning approach

The confusion matrices for the schemes based on blacklist and machine learning approach are shown in table 1 and table 2.

Table 1. Blacklist scheme confusion matrix

	Fake	**Normal**
Fake	476	25
Normal	22	479

Table 2. Machine learning approach confusion matrix

	Fake	**Normal**
Fake	480	26
Normal	25	482

According to the experimental results, when machine learning algorithms were tested on the dataset under consideration, they achieved better accuracy then the blacklist based approach. Fake accounts detection rate of blacklist based approach comes out to be nearly around 94% that is a bit lower than the detection rate of machine learning based scheme. Algorithms in machine learning touched the detection rate of around 98% with highest in case of Random Forest as shown in fig. 4. Hence we can say that the machine learning based approach is more effective than the blacklist approach in the identification of fake user accounts on Twitter. Machine learning based approach notably reduces the false positive rate rather than the blacklist based. Hence it is more suitable to serve the purpose.

VIII. CONCLUSION AND FUTURE WORK

Schemes are proposed for the identification and eradication of fake and malicious users on Twitter. The approach based on machine learning attains an acceptable accuracy & lessens more false positive rate when compared to blacklist based approach. These two schemes utilize only the content of user i.e. the text of the tweets. They do not use the profile and network data and reduces cost and time complexity to extract these features. Experimental results shows that machine learning algorithms have better accuracy than the blacklist approach.

Twitter is an online social networking site and will be ever changing to meet the needs of its users. And so the fake users will change the way they attack. Hence, more features are needed in future that will cover the new attacking patterns of different fake users and will successfully identify them.

REFERENCES

[1] Qiang Cao, Michael Sirivianos, Xiaowei Yang, and Tiago Pregueiro, "Aiding the detection of fake accounts in large scale social online services," in Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation, 2012.

[2] J. R. Douceur, "The sybil attack," in International workshop on peerto-peer systems. Springer, 2002, pp. 251–260.

[3] (2018) Facebook publishes enforcement numbers for the first time. Internet draft. [Online]. Available: https://newsroom.fb.com/news/2018/05/enforcement-numbers/

[4] Y. Boshmaf, M. Ripeanu, K. Beznosov, and E. Santos-Neto, "Thwarting fake osn accounts by predicting their victims," in Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security. ACM, 2015, pp. 81–89.

[5] S. Fong, Y. Zhuang, and J. He, "Not every friend on a social network can be trusted: Classifying imposters using decision trees," in Future Generation Communication Technology (FGCT), 2012 International Conference on. IEEE, 2012, pp. 58–63.

[6] MyoMyoSwe, and NyeinNyeinMyo, "Fake Accounts Detection on Twitter Using Blacklist", ICIS 2018, June 6-8, 2018, Singapore

[7] Monika Singh, Divya Bansal, and Sanjeev Sofat,"Detecting Malicious Users in Twitter using Classifiers", SIN '14, September 09 - 11 2014, Glasgow, Scotland Uk

[8] G. Stringhini, C. Kruegel, and G. Vigna. Detecting spammers on social networks. In 26th Annual Computer Security Applications Conference, ACSAC '10, pages 1–9. ACM, 2010.

[9] J. S. Alowibdi, U. A. Buy, P. S. Yu, L. Stenneth, Detecting deception in online social networks, in: Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on, 2014, pp. 383–390.

[10] C. Yang, R. C. Harkreader, and G. Gu. Die free or live hard? Empirical evaluation and new design for fighting evolving Twitter spammers. In 14th International Conference on Recent Advances in Intrusion Detection, RAID'11, pages 318–337. Springer-Verlag, 2011.

[11] Danezis G and Mittal P (2009) SybilInfer: Detecting Sybil nodes using social networks. In: NDSS.

[12] Yang Z, Wilson C, Wang X, et al. (2014) Uncovering social network Sybils in the wild. ACM Transactions on Knowledge Discovery from Data (TKDD) 8(1): 2 Available at: http://arxiv.org/abs/1106.5321.

[13] Lee K, Caverlee J and Webb S (2010) Uncovering social spammers: Social honeypotsþmachine learning. In: Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval, pp.435–442.

[14] Y. Boshmaf, D. Logothetis, G. Siganos, J. Ler´ıa, J. Lorenzo, M. Ripeanu, K. Beznosov, and H. Halawa, "´Integro: Leveraging victim prediction for robust fake account detection in large scale osns," Computers & Security, vol. 61, pp. 142–168, 2016.

[15] G.Wang, T. Konolige, C.Wilson, X.Wang, H. Zheng, and B. Y. Zhao, "You are how you click: Clickstream analysis for sybil detection." in USENIX Security Symposium, vol. 9, 2013, pp. 1–008.

[16] S. Fong, Y. Zhuang, and J. He, "Not every friend on a social network can be trusted: Classifying imposters using decision trees," in Future Generation Communication Technology (FGCT), 2012 International Conference on. IEEE, 2012, pp. 58–63.

[17] B. Viswanath, M. A. Bashir, M. Crovella, S. Guha, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, "Towards detecting anomalous user behavior in online social networks." in USENIX Security Symposium, 2014, pp. 223–238.

[18] Stringhini G, Kruegel C and Vigna G (2010) Detecting spammers on social networks. In: Proceedings of the 26th annual computer security applications conference, pp.1–9.

[19] Benevenuto F, Magno G, Rodrigues T, et al. (2010) Detecting spammers on Twitter. In: Collaboration, electronic messaging, anti-abuse and spam conference, p.12.

[20] Wang AH (2010) Don't follow me: Spam detection in Twitter. In: Proceedings of the international conference on Security and Cryptography (SECRYPT), pp.1–10.

[21] Xiao C, Freeman DM and Hwa T. Detecting clusters of fake accounts in online social networks. In: Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security, October 2015, pp. 91–101. ACM.

[22] Melville, Prem, and Raymond J. Mooney. "Constructing diverse classifier ensembles using artificial training examples." In IJCAI, vol. 3, pp. 505-510. 2003.

[23] Zhang, Dengsheng, MdMonirul Islam, and Guojun Lu. "A review on automatic image annotation techniques." Pattern Recognition 45, no. 1 (2012): 346-362

[24] Twitter Policy of Spammers: http://help.Twitter.com/forums/26257/entries/1831- The Twitter Rules. Accessed: 2014-03-20.

[25] Twitter API: https://dev.Twitter.com/docs/rate-limiting/1.1. Accessed: 2014-03-23.

[26] Twitter rate limit for calls: https://dev.Twitter.com/docs/api/streaming. Accessed: 2014-03-23.

[27] Sharma A., Ahluwalia A.,Deep S. and Bansal D. 2014. Friend or Foe: Twitter Users under Magnification, Distributed Computing and Internet Technology, Lecture Notes in Computer Science, 8337, 251-262. DOI= 10.1007/978-3-319-04483-5_26.

[28] Lu Z. 2004. Predicting Subcellular Localization of Proteins using Machine-Learned Classifiers. Bioinformatics, 20, 4, 547 – 556. DOI= 10.1.1.131.4463.

[29] Eisner R. 2005. Improving Protein Function Prediction using the Hierarchical Structure of the Gene Ontology. In Proceedings of IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology. DOI= 10.1.1.131.7437

Depression Detection using Machine Learning Techniques

Arunima Jaiswal[1], Damak Kapoor[2]

[1,2]Indira Gandhi Delhi Technical University for Women, Delhi India
[2]damakkapoor@gmail.com

Abstract: Depression is a mental illness characterised by low mood and loss of interest in activities. It is an illness which has enormous impact in terms of severity and scale but is deeply neglected. The major concern being that despite the severity of its detrimental impact on health and the massive scale of population it impacts, it goes undiagnosed and untreated majority of the time due to the limited reach of diagnostic treatments and the taboo attached to mental illnesses in the society. The agenda of this paper is to explore the application of different machine learning techniques for depression detection. The models such as logistic regression, linear SVC, multinomial and Bernoulli naïve Bayesian, ridge classifier, adaboost, multi-layer perceptron and passive aggressive classifiers are applied to prove their utility in depression detection.

Keywords: machine learning, social media, society, natural language processing, depression

I. INTRODUCTION

Depression is a mental illness which affects 300 million people in the world. ¾ of people with mental disorders do not receive a proper treatment, costing about 800,000 million people their lives, each year [1]. As per the World Health Organization (WHO), 7.69% of world population faces anxiety issues. The WHO reports [2] that the most common mental illnesses are anxiety disorders which encompasses several disorders, which include depression among the other most common illnesses. With the popularity of social media platforms growing rapidly at an exponential rate, it is becoming omnipresent in all of our lives. It acts as a reflection of people's emotions, aspirations, moods, sensitivities and mental state as people tend to share their feelings without hesitation on social media. These platforms are utilized by researchers to identify the causes of depression and accelerate counter steps for it. Detecting earlier depression can be a huge step to address the mental illness and offer support to the people suffering from this illness. Depression is severely effecting the quality of lives of people in our society by inhibiting people from being able to live the most positive, healthy and happy version of their life, thereby increasing the feelings of frustration, self-depreciation and suicidal tendencies. There is a dire need of diverting and committing our resources to handle the issue of prevailing depression, as apart from being a leading cause of death effecting the mortality rates, it also depreciates the quality of life of not just the patient suffering from the disorder but also of everyone related and in the vicinity of that person. Prevailing depression thus retards the utilization of human resources in the society, as depression compromises the skills and qualities which otherwise could have been efficiently contributed by that person towards the development of the society. Thus, to ensure a holistic environment for society and its inhabitants, it is of utmost importance that issue of depression be addressed properly. Machine learning techniques can be used as a fundamental tool to cater to the first and most important part of addressing depression i.e. the diagnosis which is seldom done due to the limited reach of diagnostic services of mental health to people, especially in developing countries and due to the natural inhibitions in people to seek help for a mental illness due the taboos and rigid prejudices imbibed in our society.)

Machine learning can be used to overcome all these drawbacks as when leveraged with the social networks[3], which are a gold mine for personal feelings and opinions, Machine learning can be applied and trained on the history of depressive content posted by the people and learn to detect depression at a pre-advanced stage by analyzing the posts of a person and raising an early flag so that appropriate assistance authorities can be alarmed and help the people at an early stage, thereby foreseeing and preventing the damage.

II. RELATED STUDIES

Many people have worked and published methods of predicting depression and suicidal tendencies through analysis of social media accounts. Below is a comparative analysis of the most popular and commonly used predictive methods.

Table 1. Comparative Analysis of most commonly used techniques for Depression Prediction

S. No	Author(s)	Social Media	ML Technique	Remarks	Results
1.	Cacheda et al. [4]	Reddit	Random Forest	Uses open source data from Reddit to detect early MDD	Dual model performs better than the singleton model and improves current state-of-the-art detection models by more than 10%.
2.	Deshpande and Rao[5]	Twitter	Naïve-Bayes	Uses curated word list to detect depression tendencies	The research concluded in Naïve Bayes classifier to be more accurate than SVM.
3.	Benton et al. [6]	Twitter	Logistic Regression, Perceptron	Detects anxiety, depression, bipolar, eating disorder	Results show that an MTL model with all task predictions performs better than other models, reaching 0.846 TPR for neuroatypicality where FPR=0.1, and AUC of 0.848, TPR of 0.559 for suicide
4.	Khan, et al. [7]	Facebook, Twitter	SVM, Naïve Bayes	Uses a curated BDI-II questionnaire also	It is observed that the usage pattern of Social Media users can be very helpful to analyse the Mental State of the users and predicting
5.	Saravia et al. [8]	Twitter	Random Forest	The system gives minimal results	Gives therapist a convenient online assessment tool, data collection expanded using real time crowdsourcing, provides platform for proper diagnosis and intervention
6.	Scott et al.[9]	Twitter	Decision Tree	Classifies person as suicidal or non-suicidal	Shows that people who are at high suicidal risk can be easily differentiated from those who are not by machine learning algorithms
7.	Kang et al. [10]	Twitter	K-means clustering, SVM	Shows high accuracy as compared to previously existing models	The results demonstrated that the proposed method outperforms the baseline

S. No	Author(s)	Social Media	ML Technique	Remarks	Results
8.	Burmap et al. [11]	Twitter	TF-IDF, Decision Tree, Rotation Forest	It predicts suicidal tendencies	Results showed that the reciprocity of either follower/following relationships or 'mutual' links between suicidal users is significantly higher (up to 73% as opposed to 42% in other studies),
9.	Choudhury et al. [12]	Facebook	PHQ-9	Detects postpartum depression only	Finds that increased social isolation and lowered availability of social capital on Facebook, are the best predictors of PPD in mothers.
10.	Choudhury et al. [3]	Twitter	SVM, Principal Component Analysis	Calculates SMDI(Social Media Depression Index)	SMDI can nearly reflect CDC characterized insights on depression

III. APPLICATION OF MACHINE LEARNING TECHNIQUES

Machine learning refers to making a model learn from the history and previous examples of the data available. The dataset curated in this step of the process determines the credibility of any conclusions derived by the model, so it is crucial that the dataset is Credible, Reliable, Consistent and Comprehendible

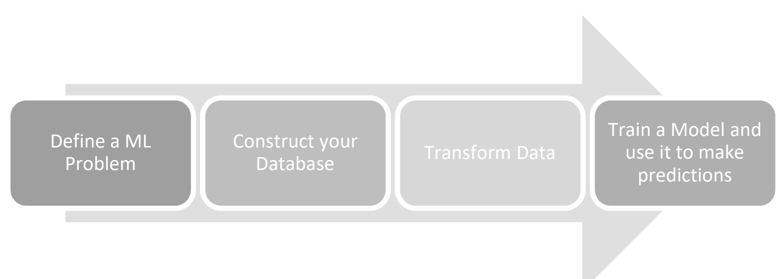


Fig. 1: Stages of Machine Learning

The Sentiment140 dataset contains 1,600,000 tweets [13] extracted using the twitter API. For the experiment, we took a sample of 8,000 tweets with polarity of 4, the positive ones. Web scraping of depressive tweets is done with TWINT to scrape depressive tweets. Twint is an open source tool written in python. It enables the user to scrap data from Twitter, without the need of official permissions or an API. It allows the user to use keywords to filter the tweets, specify the users, dates, hashtags etc to further enhance the control on the data scrapped. Then is pre-processing or cleaning of the dataset.). Total dataset had 10,314 tweets (8,000 positive and 2,314 depressive). Data Cleaning is a fundamental part of the process as it is used to identify and eliminate any errors and discrepancies in the data which might compromise the credibility of our results. It includes removing empty spaces, encodings and any data that might not be relevant to our analysis like mentions, usernames etc, so as to reduce the volume of the data and confine it to only relevant and consistent data to increase the efficiency of our results thus obtained from the models.

The data cleaning function in natural language processing target the following:

- HTML Decoding
- @mention
- URL Links

- UTF-8 BOM (Byte Order Masking)
- Hash tags
- Numbers/Statistics

The dataset details are as follows:

Posts	Depression dataset
No. of entries before cleaning	10314
No. of entries after cleaning	10289

Feature Extraction is done using Tfidf Vectorizer [11]. TFIDF is, which is used in Natural Language Processing in order to obtain numeric form of the data in hand. The product of TF and IDF gives the required output vector.

$$TF(t,d) = \frac{number\ of\ times\ term(t)\ appears\ in\ document(d)}{total\ number\ of\ terms\ in\ document(d)}$$

$$IDF(t,D) = \log\left(\frac{total\ number\ of\ documents(D)}{number\ of\ documents\ with\ the\ term(t)\ in\ it}\right)$$

Then machine learning algorithms for instance logistic regression, linear SVC, multinomial and Bernoulli naïve Bayesian, ridge classifier, adaboost, multi-layer perceptron and passive aggressive classifiers are applied to the dataset for depression detection using accuracy as efficacy criteria.

The details about these algorithms are as follows:

Table 2. Machine Learning Algorithms used

S. No	Technique	Details
1	Logistic Regression[6]	It is used when the variable is binary an an equation is detremine to make the predictions
2	Linear SVC[3]	It fit the data by a hyperplan such that the distance of the closest data point from the plane is minimized
3	Multinomial Naïve Bayesian[5]	It used condition independence of features and multinomial distribution of features
4	Bernoulli Naïve Bayesian[5]	It works for Binary or boolean features
5	Ridge Classifier[14]	It is defines by a hyperplane, which does classification defined by a hyperplane
6	AdaBoost[15]	It tweaks subsequent weak learners as per the previous misclassifications
7	Multi-layer Perceptron[16]	A class of feedforward ANN which uses backpropagation for training
8	Passive Aggressive Classifier[17]	A member of family of online learning algorithms ny Crammer

IV. RESULT ANALYSIS

This section provides an overview of the results obtained by different machine learning algorithms.

Following results were obtained using the above described algorithms.

Table 3. Confusion Matrix obtained from Logical Regression

165	1
41	0

Table 4. Results obtained from Logical Regression

Class	Precision	Recall	F1-Score	Support
Non-Depressive	0.80	0.99	0.89	166
Depressive	0	0	0	41
Avg/total	0.64	0.80	0.71	207

Table 5. Confusion Matrix obtained from BernoulliNB

124	42
19	22

Table 6. Results obtained from BernoulliNB

Class	Precision	Recall	F1-Score	Support
Non-Depressive	0.87	0.75	0.80	166
Depressive	0.34	0.54	0.42	41
Avg/total	0.76	0.71	0.73	207

Table 7. Confusion Matrix obtained from Ridge Classifier

152	14
34	7

Table 8. Results obtained from Ridge Classifier

Class	Precision	Recall	F1-Score	Support
Non-Depressive	0.82	0.92	0.86	166
Depressive	0.33	0.17	0.23	41
Avg/total	0.72	0.77	0.74	207

Table 9. Confusion Matrix obtained from Linear SVC

165	1
41	0

Table 10. Results obtained from Linear SVC

Class	Precision	Recall	F1-Score	Support
Non-Depressive	0.80	0.99	0.89	166
Depressive	0	0	0	41
Avg/total	0.64	0.80	0.71	207

Table 11. Confusion Matrix obtained from MultinomialNB

123	43
29	12

Table 12. Results obtained from MultinomialNB

Class	Precision	Recall	F1-Score	Support
Non-Depressive	0.81	0.74	0.77	166
Depressive	0.22	0.29	0.25	41
Avg/total	0.69	0.65	0.67	207

Table 13. Confusion Matrix obtained from AdaBoost

160	6
37	4

Table 14. Results obtained from AdaBoost

Class	Precision	Recall	F1-Score	Support
Non-Depressive	0.81	0.96	0.88	166
Depressive	0.40	0.10	0.16	41
Avg/total	0.73	0.79	0.74	207

Table 15. Confusion Matrix obtained from Perceptron

157	9
40	1

Table 16. Results obtained from Perceptron

Class	Precision	Recall	F1-Score	Support
Non-Depressive	0.80	0.95	0.87	166
Depressive	0.10	0.02	0.04	41
Avg/total	0.66	0.76	0.70	207

Table 17. Confusion Matrix obtained from Passive-Aggressive

162	4
41	0

Table 18. Results obtained from Passive-Aggresive

Class	Precision	Recall	F1-Score	Support
Non-Depressive	0.80	0.99	0.89	166
Depressive	0	0	0	41
Avg/total	0.64	0.80	0.71	207

The best accuracy for Depression detection is observed in Ridge Classifier and Adaboost classifier.The best average F1-score was obtained of 0.74 and lowest was observed in the case of Multinomial Naïve Bayesian of 0.67. The depression detection models (Logistic Regression, Linear SVC, Passive-Aggressive) tend to face over-fitting as the keyword 'depression' used to scrape the dataset.. The algorithms tend to train on the keyword "depression", and classify the tweets with keyword depression as depressive, thereby giving off the charts accuracy. A limitation of self-scraped dataset is presented in the form of overfitting of machine learning algorithms on the keywords used to scrape the data. As a result of which it might classify data which could be about depression, with the same keywords as being depressive. Therefore, to overcome this limitation, we need to use the same models on a medically credible dataset provided by an attested medical research institution as it cannot be successfully implemented in social networking sites due to the consideration of patient confidentiality, we cannot mark accounts of the patients to feed data to our models, thus in the arena of social networks, keywords must be used. A larger dataset created with extensive resources to scrape large number of posts, in millions, which could not be executed by us owing to the time and resource restrictions, would help resolve this issue of overfitting.

V. CONCLUSION

Depression is serious issue which impairs the population vastly and must be addressed to the best of our capabilities to ensure a healthy and fulfilling life of people and a holistically blooming society Thus, depression prediction tools like these, must be deployed on all active social media handles so that it could help in early and efficient diagnosis and detection of depression. Thereby raises an alarm to the concerned organization with might lend the support in an empathetic way to the person and prevent the long-term repercussions. This work is just the preliminary application of few basic machine learning techniques for depression detection. It has got ample scope for further expansion by analyzing any benchmark depression based corpus or creating a huge dataset from various social media forums such as instagram, facebook, etc. or scraping data from any societies that handle depression etc. Such data can again be analyzed for depression detection using other soft computing techniques such as genetic algorithms, swarm based algorithms, deep learning based algorithms etc. for further testing.

REFERENCES

[1] 10 Facts on Mental Health, Fact 3 referenced at https://www.who.int/features/factfiles/mental_health/mental_health_facts/en/index2.html

[2] Mental disosders Key Facts referenced at https://www.who.int/news-room/fact-sheets/detail/mental-disorders

[3] Munmun De Choudhury, Scott Counts, Eric Horvitz(2013), *Social Media as a Measurement Tool of Depression in Populations*, WebSci, Proceedings of the 5th Annual ACM Web Science Conference, Page 47, Paris, France

[4] Cacheda F, Fernandez D, Novoa FJ, Carneiro V(2019), *Early Detection of Depression: Social Network Analysis and Random Forest Techniques*, Journal of Medical Internet Research, Vol 20, Page 6

[5] Mandar Deshpande, Vignesh Rao (2017), *Depression detection using emotion artificial intelligence*, International Conference on Intelligent Sustainable Systems, Page 858, Palladam, India

[6] Adrian Benton, Margaret Mitchell, Dirk Hovy(2017) *Multitask Learning for mental health condition with limited social media*, Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Vol 1, Page 152, Valencia, Spain

[7] Ashar Khan, Mohd Shahid Husain, Anam Khan(2017) *Analysis of Mental States of users using social media to predict depression, a survey*, 13th International Symposium on Autonomous Decentralized Systems, Page 910, Bangkok, Thailand

[8] Elvis Saravia, Chun Hao Chang, Renaud Jollet De Lorenzo, Yi- Shin Chen(2016), *MIDAS: Mental Illness Detection and Analysis via Social Media*, International Conference on Advances in Social Networks Analysis and Mining,Vol, August, Page 1418, San Francisco, CA, US

[9] Scott R Braithwaite, Christophe Giraud-Carrier, Josh West, Michael D Barnes, Carl Lee Hanson (2018), Validating Machine Learning Algorithms for Twitter Data against established Suicidal Tendency, JMIR Mental Health, Vol 3

[10] Keumhee Kang, Chanhee-Yoon, Eun Yi Kim(2016), *Identifying Depressive Users in Twitter Using Multimodal Analysis*, BigComp, Page 231, Hong Kong, China

[11] Peter Burmap, Gualtiero Colombo, Jonathan Scourfield(2015), *Machine Classification and Analysis of SuicideRelated Communication on Twitter*, ACM HyperText , Page 291, Cyprus.

[12] Munmun De Choudhury, Scott Counts, Eric J. Horvitz, Aaron Hoff(2014), *Characterizing and predicting postpartum depression from shared Facebook data*, CSCW '14 Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing , Page 626, Baltimore, Maryland

[13] Nimit Bindal, Niladri Chatterjee(2016), *A Two-step Method for Sentiment Analysis of Tweets*, International Conference on Information Technology (ICIT), Bhubaneshwar, India

[14] Mohsen Nabian, Mohammad Ali Nabian(2017), *A Comparative Study on Machine Learning Classification Models for Activity Recognition* , Journal of Information Technology and Software Engineering

[15] Peng Wu, Hui Zhao(2011), *Some Analysis and Research of the AdaBoost Algorithm,* International Conference on Intelligent Computing and Information Science, Chongqing, China

[16] Shaohua Wan, Yan Liang, Yin Zhang and Mohsen Guizani(2018), *Deep Multi-Layer Perceptron Classifier for Behavior Analysis to Estimate Parkinson's Disease Severity Using Smartphones,* IEEE, Special Section on Real-Time Edge Analytics for Big Data in Internet OF Things

[17] Koby Crammer, Joseph Keshet, Ofer Dekel, Shai Shalev-Shwartz(2006), *Online Passive-Aggressive Algorithms*, Journal of Machine Learning Research.

TRACK 4

APPLICATIONS OF AI

Intelligent Internet of Things (IIOT): Investigation Issues and Challenges

Shakir Khan*

*College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University,
Riyadh, Saudi Arabia
sgkhan@imamu.edu.sa

Abstract: The Internet of Things (IoT) is progressively developing as the successive aspect of the development of the Internet, it turns out to be decisive to identify the several prospective areas of utilization for IIoT, and the investigation encounters that are linked with these solicitations alternating from intelligent homes, intelligent cities and their network, intelligent farming, and trade, to intelligent alive and intelligent surroundings. IoT is estimated to penetrate into almost all phases of everyday life. Although the present IoT empowering expertise have significantly value-added in the current years, there are immobile various difficulties that need consideration. Since the IIoT thought follows from various different technologies, many issues are guaranteed to arise. Due to this, we can say that virtually all domains of our lives, makes it an important investigation topic for exploration in several related arenas like data processing i.e. IT and computer technology. Consequently, IIoT is showing the innovative path to be explored for investigation. This research boon the fresh growth of IoT expertise and debates future solicitations and investigation challenges.

Keywords: Intelligent Internet of Things; IIoT solicitations; IoT threats; Imminent technologies; intelligent cities network; intelligent surroundings; intelligent agriculture; intelligent living

I. INTRODUCTION

The Internet may be explained as interconnection computers to make the communication between that networks that joins entities and produce information although The IoT is a planned structure of address intelligent physical fragments with several points of sensing, handling, and actuation proficiencies that share the competency to converse and interoperate with the Internet on mutual platform [1][2].

The internet is continuously developing as expected, it has been converted these days more than a simple connected network of computers, but somewhat a network of many devices, while IoT functions as a network of different “connected” devices a network of networks [3], as revealed in figure 1. Currently, appliances like intelligent phones, intelligent spaces, intelligent vehicles, intelligent industrial systems, intelligent health, intelligent cameras, toys, buildings, home appliances, industrial systems and uncountable others can all be a part of information over the Internet by sharing this information.

With all guns blazing, the Intelligent Internet of Things (IIoT) is projected to continue growing its influence as affects the number of appliances and utilities, which it can track. This is obvious from the uncertainty in the appearance of “Things” which makes it challenging to plan the increasing limits of the IoT [4]. While commercial achievement continues to emerge, the IIoT persistently offers a practically unlimited supply of prospects, not just in industries but also in investigation. As a result, the substitute speaks the several prospective areas for solicitation of IoT areas and the investigation tasks that are linked with these solicitations.

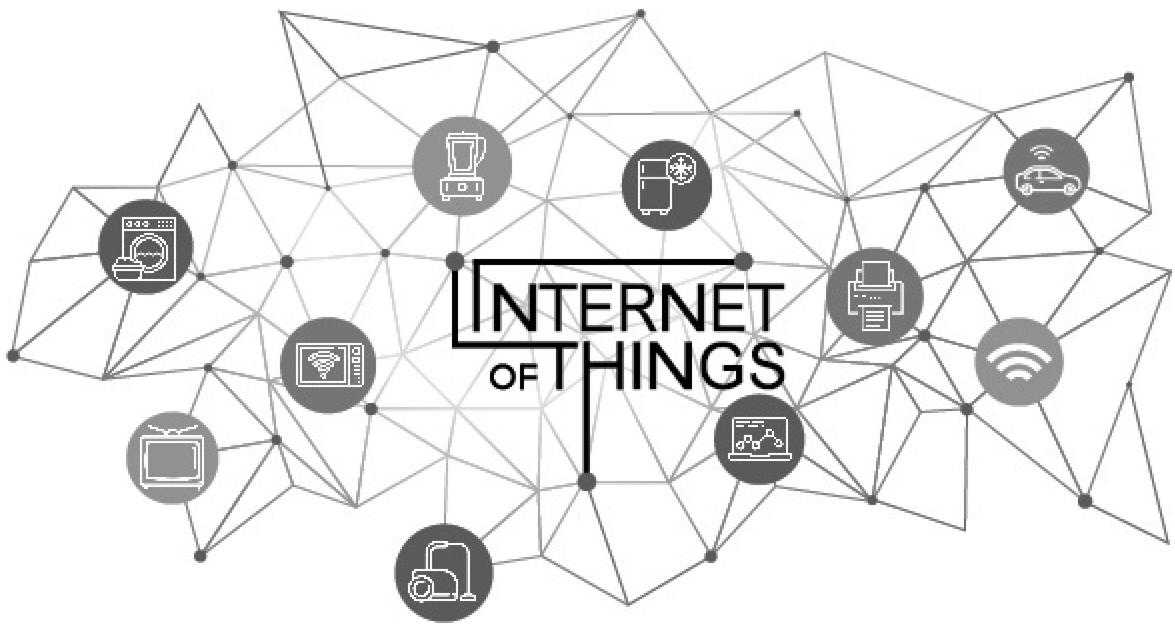


Fig. 1. IoT as a Network of Networks [3].

II. PROSPECTIVE UTILIZATION FIELDS OF IIOT

Rendering to [5], the utilizations of IoT shield far-reaching areas together with manufacturing or the health sector, modern sector, intelligent cities, health sector, safe keeping, intelligent agriculture, and intelligent aeronautics among many others.

III. INTELLIGENT CITIES

Rendering to [6], the IoT shows an important aspect in refining the intelligence of cities and improving overall frame. Some of IoT utilization parts in constructing intelligent cities consist of; intelligent building, traffic congestion [7, 8] waste management [9], intelligent transportation systems [7], intelligent lighting, intelligent parking, and urban maps. This may comprise unlike functionalities like; monitoring existing parking spaces inside the city, monitoring vibrations and material circumstances of bridges and buildings, placing in residence sound monitoring devices in sensitive parts of cities, as well as monitoring the levels of walkers and vehicles.

Utilization of IoT to attain intelligent cities would need exhausting radio frequency identification and sensors. Some of the previously developed utilizations in this domain are the Aware home and the intelligent Santander functionalities. In the United States, some foremost cities like Boston have strategies on how to appliance the Internet of Things (IoT) in most of their systems reaching from their computerized parking, street-lights, automatic sprinkler, and sewage mince are all arranged to be intertwined and linked to the internet. Such utilizations will deal important disruption assumed in terms of saving money and energy.

IV. HEALTH-CARE

Maximum health-care organizations in several countries are incompetent, slow and predictably error-prone. This can smoothly be transformed since the health-care sector trusts on various activities and devices that can be computerized and improved through technology. Further technology that can simplify countless operations like report distribution too many individuals and locations, record keeping and distributing medicines would go a long time in altering the health-care segment [10].

Many advantages that IoT utilization deals in the health-care sector is furthermost categorized into tracing of patients, staff, and objects, recognizing in addition to validating, entity, and the automatic assembly of data and sensing. Hospital workflow can be considerably enhanced once patients flow is traced. Furthermore, verification and documentation ease happenings that may be dangerous to patients, file maintenance and littler cases of mismatching infants. Additionally, spontaneous information collection and communication is dynamic in course automation, decrease of form handling time-lines, automated process examining in addition to medical catalog organization.

Utilization fields in this segment include; being intelligent to observe a patient's agreement with medicines, telemedicine solutions, and warnings for patients' happiness. In this manner, sensors can be useful to outpatient and inpatient patients, dental Bluetooth devices and toothbrushes that can provide material after they are recycled and patient's observation.

The solicitations of Internet of Things (IoT) and Internet of Everything (IoE) are moreover being long-drawn-out through the emergence of the Internet of Nano-things (IoNT) [3]. The concept of IoNT, as the label suggests, is being planned by incorporating Nano-sensors in different objects (things) consuming Nano networks. Medical utilization, as displayed in Figure 2, is one of the foremost motivations of IoNT executions. IoNT solicitation in human body for conducting usage determinations assists access to data from in the structural portions of the build which were till now available to intelligence from or by consuming those medical utensils integrated with massive sensor size. Therefore, IoNT will allow novel medical data to be composed and showing the ways to innovative findings and better diagnostics.

V. KEEN CULTIVATION AND WATER CONTROLLING

Agreeing to [11], the IoT has the capability to reinforce and improve the agriculture sector through investigative soil moisture and in the situation of wineries, monitoring the stem diameter. Additionally, considering climate circumstances let's predicting of snow information, deficiency, wind variations, rain or snow, thus adjusting temperature and moisture levels to stop fungus as well as other microbial pollutants.

When it approaches to cattle, IoT can promote in detecting beasts that browse in exposed locations, identifying harmful gases from animal manures in farms, and monitoring growth conditions in offspring to improve probabilities of health and survival and so on. Furthermore, with IoT utilization in farming, a lot of waste and decay can be circumvented through proper observing approach and management of the whole farming arena. It also shows the ways to improve energy and water control.

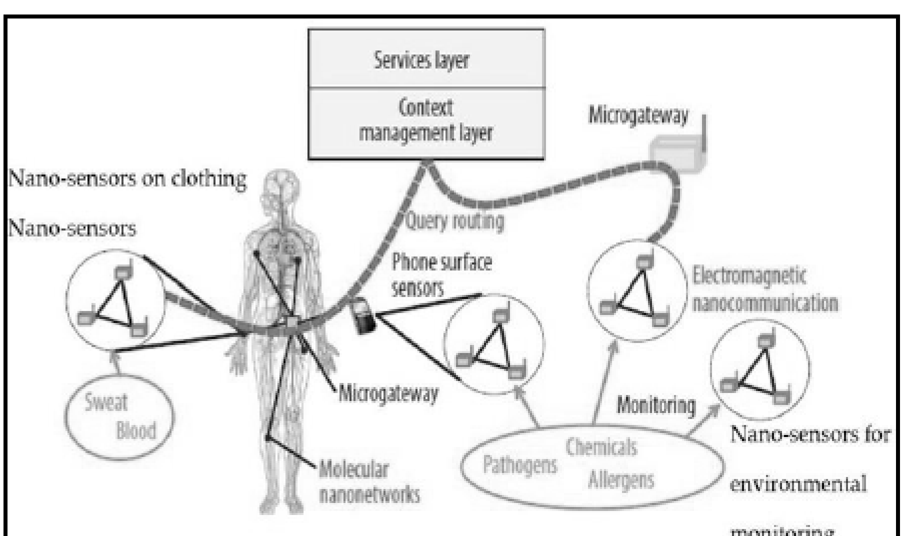


Fig. 2. The Internet of Nano-Things [3].

VI. PADDLE AND ENGINEERING

Exclusive the retail premises, IoT can be applied to numerous utilizations such as track in the shop built on a preselected list, fast payment developments like mechanically inspection out with the help of biometrics, discovering potential allergen yields and monitoring the cycle of products on shelves and warehouses in order to computerize refilling approach [12].

The intelligent IoT components generally applied in this location comprise; cellular sensor networks and wireless occurrence credentials. In paddle, there is an existing use of SUP (Systems Utilizations and Products), although in engineering various illustrations comprise excellence package circumstances, element location, sensing storage unsuitability issues, taskforce tracing between others. In the production territory, IoT supports in sensing stages of gas and leaks inside the manufacturing and its environs, protection track of poisonous gases in addition to the oxygen levels inside the borders of chemical plants to certify the security of belongings and workforces and detecting stages of oil, thrills and water in reservoirs and loading containers.

VII. INTELLIGENT LIVING

In this area, IIoT can be functional in distant control machine by which one can distantly switch uses on and off henceforth stopping accidents and save energy [1, 3]. Additional intelligent home utilizations comprise refrigerators well-fitting with LCD (Liquid Crystal Display) screens, allowing one to identify what is available secret, what has over stayed and is nearly failing along with what requirements to be refilled. This statistics can also be connected to an intelligent phone utilization allowing one to access it once outdoor the house and consequently buy what is desired. Some microwaves which have a self-cleaning component can be simply checked as well. In terms of security in the household, IoT can be useful with alarm arrangements and cameras can be mounted to display and discover window or door beginnings henceforth stopping trespassers [3].

VIII. INTELLIGENT SURROUNDINGS

Intelligent surroundings approaches incorporation with IoT technology should be generated for sensing, following and valuation of entities of the surroundings that deal potential reimbursements in realizing a maintainable life and a green world. The IoT technology permits detecting and handling of air excellence through data gathering from remote sensors through cities and given that round-the-clock geographic treatment to achieve better ways of handling traffic jams in bigger cities. Furthermore, IoT knowledge can be functional in calculating toxic waste ranks in water and subsequently educate verdicts on water convention. In discarded management, which comprises of several types of waste, identical chemicals and contaminants being damaging to the surroundings and to people, animals, and plants as well, IoT can also be useful. This can be attained by natural security by means of directing industrial pollution through immediate observing and management systems collective with management additionally to decision making networks. This helps to reduce waste [13].

IX. INVESTIGATION CHALLENGES

For all the beyond prospective utilizations of IIoT, there has to be suitable possibility into the changed fields to determine the achievement of some usages and their performance. Similar to any additional practice of expertise or revolution, IoT has its challenges and consequences that must be organized to empower mass implementation. Although the recent IoT supporting technologies have prominently improved in the current years, there are still various problems that need consideration, hence flagging the way for innovative dimensions of investigation to be supported out. Since the IoT idea succeeds from heterogeneous technologies that are applied in sensing, assembling, action, dispensation, deducing, transmitting, informing, managing, and loading of data, several investigation challenges are guaranteed to rise. These investigation challenges that need consideration have subsequently covered contrasting investigation areas [14].

X. CONFIDENTIALITY AND SAFETY

Outstanding to the information that IoT has turn out to be a dynamic component as concerns the forthcoming of the internet with its improved practice, it requires a necessity to sufficiently talk safety and belief purposes. Investigators are conscious of the faintness which currently exist in various IIoT devices. Moreover, the groundwork of IoT is placed on the current wireless sensor networks (WSN), IoT therefore structurally take over the similar secrecy and safety concerns WSN owns [3, 15]. Several occurrences and faintness on IoT schemes verify that there is certainly a requirement for comprehensive fluctuating safety designs which will safeguard data and systems from end to end. Various occurrences normally adventure faintness in precise accessories through achievement approach into their schemes and as a consequence building protected devices accessible [16, 17]. This safety crack further stimulates wide-ranging safety clarifications that comprise of investigation that is effective in functional cryptography for archives and system safety, non-cryptographic safety methodology along with contexts that support developers to arise with secure systems on campaigns that are mixed.

XI. DEALING OUT, INVESTIGATION AND ADMINISTRATION OF DATA

The technique for dealing out, investigation and data administration is extremely challenging as a consequence of the confused surroundings of IoT, and the huge gage of information collected, predominantly in this period of Big Data [18]. However, there is a continuous anxiety about conservative cloud architectures not being operative in terms of transporting the huge capacities of data that are created and disbursed by IoT supported devices and to be intelligent additional support the attending computational load and instantaneously encounter timing controls [19]. 77239186

One more investigation way as affections data management is spread over Information Centric Networking (ICN) in the IoT. With this information centric systems deal provision in the well-organized content retrieval and contact to services, they look to be quite appreciated not just in accessing but also transporting in addition to handle created content and its broadcast.

XII. CONCLUSION

The IIoT can finest be labelled as a CAS (Complex Adaptive System) that will remain to advance in future necessitating innovative and pioneering forms of software engineering, plan management, systems engineering, as well as various other corrections to cultivate it further and succeed it the upcoming years. The utilization areas of IIoT are relatively miscellaneous to facilitate it to oblige different users, who sequentially have different requirements. The knowledge functions three classes of users, individuals, the society or communities and institutions. As deliberated in the utilization segment of this investigation paper, the IoT has without a hesitation a great competence to be an extremely transformative force, which will, and to some degree does already, confidently influence millions of lives globally. Rendering to, this has developed even further obvious, as several governments around the globe have exposed an curiosity in the IoT idea by providing further funding in the arena that is meant to enable more investigation. A respectable instance is the Chinese Government. Countless investigation assemblies have been, and continue to be, originated from altered chunks of the biosphere, and their foremost neutral is to monitor complete IoT interrelated investigations. As increasingly investigation studies are accompanied, novel measurements to the IoT developments, technologies complex and the substances that can be linked, continue to materialize, more concrete way for much more utilization performances of IIoT. The fact that IoT is so spread-out and touches almost all areas of our lives, makes it an important investigation topic for studies in numerous correlated arenas such as information technology and computer science. The paper best part numerous potential utilization areas of the Internet of Things (IoT) and the interconnected investigation challenges.

REFERENCES

[1] M. H. Miraz & M. Ali, P. S. Excell & R. Picking, (2015). A Review on Internet of Things (IoT), Internet of Everything (IoE) and Internet of Nano Things (IoNT). *Internet Technologies and Utilizations (ITA),* pp. 219– 224.

[2] P. J. Ryan & R. B. (2017), Watson, Investigation Challenges for the Internet of Things: *What Aspect Can OR Play? Systems,* 5 (1), 1–34.

[3] M. Miraz & M. Ali & P. Excell, & R. Picking, (2018). Internet of Nano-Things, Things and Everything: Future Growth Trends", Future Internet, vol. 10, no. 8, p. 68.

[4] E. Borgia & D. G. Gomes & B. Lagesse & R. Lea & D. Puccinelli. (2016), Special issue on" Internet of Things: Investigation challenges and Solutions. *Computer Communications*, vol. 89, no. 90, pp. 1–4.

[5] K. K. Patel & S. M. Patel & et al. (2016). Internet of Things IOT: definition, characteristics, architecture, enabling technologies, utilization future challenges, *International journal of engineering science and computing,* vol. 6, no. 5, pp. 6122–6131.

[6] S. V. Zanjal & G. R. Talmale, (2016). "Medicine reminder and monitoring system for secure health using IOT," Procedia Computer Science, vol. 78, pp. 471–476,.

[7] R. Jain, (2018). "A Congestion Control System Based on VANET for Small Length Roads", Annals of Emerging Technologies in Computing (AETiC), vol. 2, no. 1, pp. 17–21.

[8] S. Soomro, M. H. Miraz, A. Prasanth, M. Abdullah, "Artificial Intelligence Enabled IoT: Traffic Congestion Reduction in Smart Cities," IET 2018 smart Cities Symposium, pp. 81–86, 2018, DOI: 10.1049/cp.2018.1381.

[9] S. H. Mahmud, S. H., Assan, L. & Islam, R. (2018). "Potentials of Internet of Things (IoT) in Malaysian Construction Industry", Annals of Emerging Technologies in Computing (AETiC), pp. 44-52, Vol. 2, No. 1, International Association of Educators and Investigationers (IAER).

[10] Y., Mano, B. S., Faical, Nakamura L., Gomes, P. G. Libralon, R. Meneguete, G. Filho, G. Giancristofaro, G. Pessin, B. Krishnamachari, & Jo Ueyama. (2015). Exploiting IoT technologies for enhancing Health Smart Homes through patient identification and emotion recognition. Computer Communications, 89.90, (178-190).

[11] V. Sundareswaran & M. S. null, (2018). "Survey on Intelligent Agriculture Using IoT," International Journal of Innovative Investigation in Engineering & Management (IJIREM), vol. 5, no. 2, pp. 62–66.

[12] P. Tadejko, (2015). "Utilization of Internet of Things in logistics-current challenges," Ekonomia i Zarz{a˛} dzanie, vol. 7, no. 4, pp. 54–64.

[13] S. Rajguru, S. Kinhekar, & S. Pati, (2015). "Analysis of Internet of Things in a smart surroundings," International Journal of Enhanced Investigation in Management and Computer Utilizations, vol. 4, no. 4, pp. 40–43.

[14] H. U. Rehman, M. Asif, & M. Ahmad, (2017). "Future utilizations and investigation challenges of IOT," in 2017 International Conference on Informa-tion and Communication Technologies (ICICT), pp. 68–74.

[15] Z. Alansari, N. B. Anuar, A. Kamsin, M. R. Belgaum, J. Alshaer, S. Soomro, & M. H. Miraz, (2018). "Internet of Things: Infrastructure, Architecture, Security and Privacy", in 2018 International Conference on Computing, Electronics Communications Engineering (iCCECE), pp. 150– 155.

[16] J. A. Chaudhry, K. Saleem, P. S. Haskell-Dowland, & M. H. Miraz, (2018). "A Survey of Distributed Certificate Authorities in MANETs," Annals of Emerging Technologies in Computing (AETiC), vol. 2, no. 3, pp. 11–18.

[17] A. S. A. Daia, R. A. Ramadan, & M. B. Fayek, (2018). "Sensor Networks Attacks Classifications and Mitigation", Annals of Emerging Technologies in Computing (AETiC), vol. 2, no. 4, pp. 28–43.

[18] Z. Alansari, N. B. Anuar, A. Kamsin, S. Soomro, M. R. Belgaum, M. H. Miraz, & J. Alshaer, (2018). "Challenges of Internet of Things and Big Data Integration", in Emerging Technologies in Computing (M. H. Miraz, P. Ex- cell, A. Ware, S. Soomro, and M. Ali, eds.), (Cham), pp. 47–55, Springer International Publishing.

[19] J. Cooper & A. James, (2009). "Challenges for database management in the Internet of Things" IETE Technical Review, 26(5), pp.320–329.

Multimodal Biometric System: A Review

Nivedita Palia[1], Shri Kant[2], Amita Dev[3]

[1]Vivekananda Institute of Professional Studies, Delhi, India
[2]Sharda University, Greater Noida, India
[3]Indira Gandhi Delhi Technical University for Women, Delhi, India

Abstract: A process of recognizing an individual based on their biological or behavioral attribute is known as Biometric. Biological traits such as face, fingerprint, etc. are inborn and stable while behavioral traits such as voice, signature, etc. are unstable and change with time and condition. A biometric system developed using a single attribute is known as unimodal biometric system. No single unimodal system is able to meet the required accuracy due to this multimodal biometric system has been developed. A multimodal biometric system combines more than one biometric trait. This paper presented a review of a multimodal biometric system along with a comparative study of sixteen multimodal biometric systems. This includes the general block diagram of biometric system, location and type of attacks on biometric system, different level of fusion and types of multimodal biometric system.

Keyword: Biometric, Unimodal, Multimodal, Fusion Level

I. INTRODUCTION

Biometrics is a process of recognizing an individual based on their biological or behavioural traits as shown in figure 1(Jain, Ross, et al., 2011). Recognition can be categorized into identification and verification. Identification refers to identify an individual either from the set of known person or otherwise while verification refers to confirm or deny a person's claimed identity. Before biometric recognition a person can be recognized either using the knowledge-based technique (e.g. password, PINs) or token-based technique (e.g. ATM or credit cards). The advantage of biometric over previously used techniques is that biometric can't be misplaced, stolen, and forgotten. Any biological or behavioural traits act as biometrics provided it satisfied following characteristics (Clarke, 1994).

- *Universality*: every individual possess the characteristics
- *Uniqueness*: no two individual have same characteristics
- *Permanence*: remain constant over a long time period
- *Easy to collect*
- *Performance*: refers to the accuracy, speed, and robustness of the system
- *Acceptability*: refers to the extent to which public ready to accept the technology in routine
- *Circumvention*: refers to how difficult to spoof the system

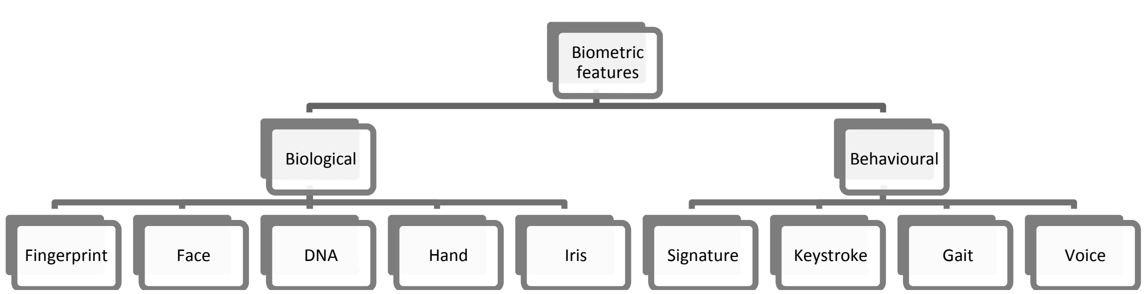


Fig. 1. Biometric features

There are two types of biometric system namely unimodal biometric system (UBS) and multimodal biometric system (MBS). UBS works with single biometric traits. No single UBS has been able to give 100 percent accuracy due to the following factors (Dahen and Fadewar, 2018):

i. Noise: sensor noise and background noise is present in acquired data
ii. Intra-class variation: biometrics of an individual varied due to the effect of age, health, etc.
iii. Inter-class similarities: overlapping in feature space of multiple users
iv. Lack of universality
v. Spoof Attack: fake biometric submitted to the sensor

Some of the limitations of the UBS can be removed by using multiple biometrics; therefore the MBS has been developed. A MBS uses two or more biometric traits to recognize a person. These systems are more reliable and give better performance as compared to UBS.

In this paper, a detailed review of MBS has been presented along with comparative study of sixteen different MBS have been compared. The rest of the paper is organised as follow: Section 2 describe the architecture of biometric system while details of MBS is presented in section 3. Various types of attacks and there location in a biometric system has been presented in section 4 followed by the conclusion in section 6.

II. GENERAL ARCHITECTURE OF BIOMETRIC SYSTEM

A biometric system is a pattern recognition system which performs four main tasks as discussed below:

a. Data Acquire and Digitization: It captures biometric information from an individual, converts, and stores it into digital form for further processing.
b. Feature Extraction: Extracts a salient set of features from the pre-processed data.
c. Classification: Compares the extracted feature set from those stored in a database.
d. Decision Making: Accept or reject an individual based on the decision method.

Figure 2 represents the architecture of the biometric system (Sareen, 2014).

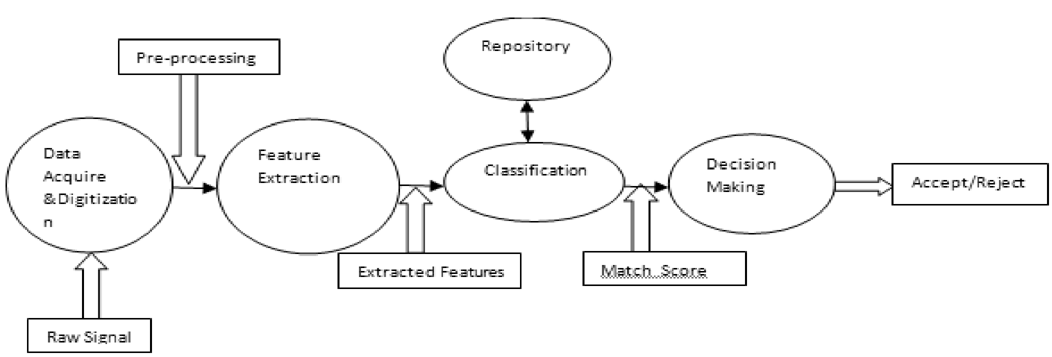


Fig. 2. Architecture of Biometric system

III. MULTIMODAL BIOMETRIC SYSTEM (MBS)

A MBS integrate two or more biometric traits. This section describes the details about the MBS such as level of fusion, mode of operation, types, etc.

IV. LEVEL OF FUSION

A MBS performs the same four tasks as discussed above in the general structure of the biometric system except there has been an additional phase known as fusion, where the match score of each factor has been combined to produce final score which later helps in decision making (Kumar and Farik, 2016). The fusion phase in a MBS can be fused at any of the four phases.

Fusion at the Data Acquire level: Multiple biometric traits of an individual have been collected using the different sensors at same instance or at multiple instances. Figure 3 shows the Data Acquire Level fusion.

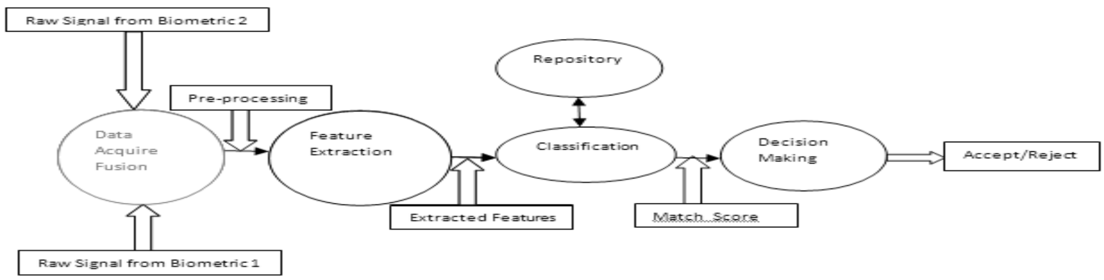


Fig. 3. Fusion at Data Acquire Level

Fusion at the Feature Extraction level: The set of feature vectors extracted from different biometrics traits has been combined together to form a new single feature vector. Either the same feature extraction technique or different technique has been used to extract features from different biometric can be used to fuse features. Fusion at this stage is a challenging task due to unknown relation between biometrics and structurally incompatibility among features. Fusion at feature level has been represented in figure 4.

Fusion at the Classification level: Classifier generates a matching score reflecting the closeness of the feature vector of person with the template vector of registered person. The normalized score generated using different classifiers has been collectively generate a match score. Fusion at this stage is the easiest one. Classification level fusion has been shown in figure 5.

Fusion at the Decision-making level: Match score of each UBS has been fused to make a final decision. Fusion at this level is least preferred in designing MBS. Figure 6 shows Decision level fusion.

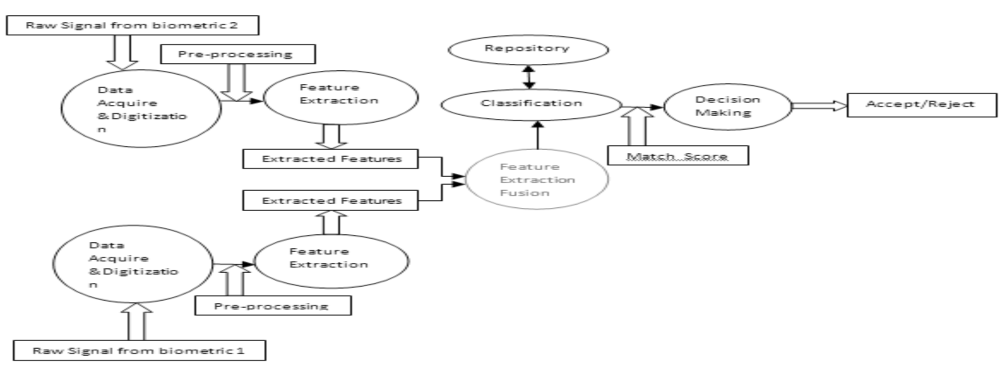


Fig. 4. Feature level fusion

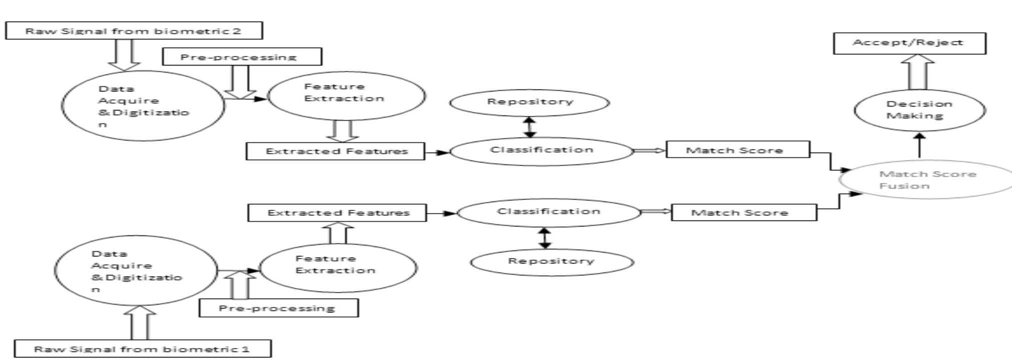


Fig. 5. Fusion at classification level

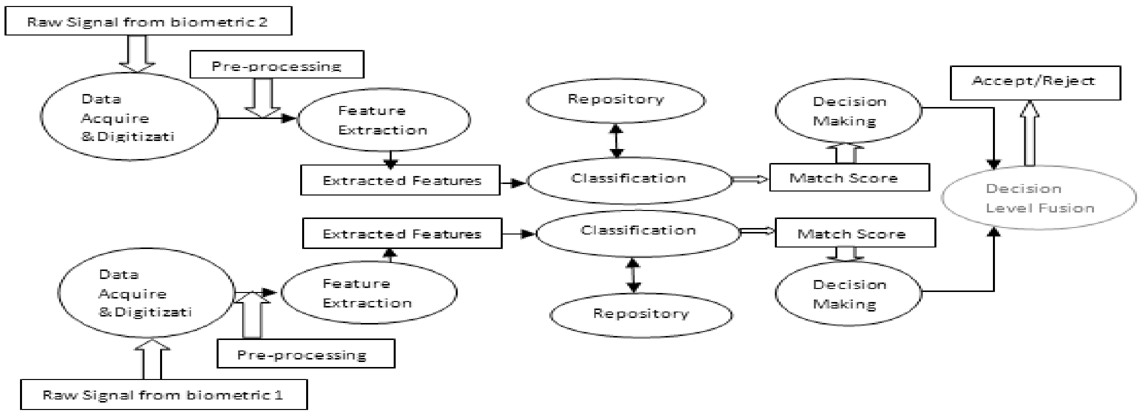


Fig. 6. Fusion at Decision-making level

V. TYPES

On the basis of biometric traits, sensors used and extracted features a MBS have been classified into four types:

- *Single biometric traits with multiple sensors:* A biometric trait has been collected using multiple sensors
- *Multiple biometric traits:* More than one biometric traits have been combined
- *Multiple units with single biometric trait:* For the biometrics such as finger, palm, iris, ear, etc. two or more same trait has been recorded (such as a collect fingerprint of all the fingers of a person).
- *Multiple instances of single biometric trait:* Different instances of the same biometric trait has been collected

VI. MODE OF OPERATION

A MBS can be operated in three different ways:

- Serial mode: Identification of one biometric trait has been used as input for next biometric to validate
- Parallel mode: Simultaneously multiple UBS perform recognition task
- Hierarchical mode: A UBS combined in tree-like structure.

VII. STRENGTH AND CHALLENGES

The MBS has been more reliable and accurate than the independent biometric modalities. The failure of an individual biometric may not influence the person recognition as more than one biometric trait has been used to recognize an individual. Spoofing is minimal which enhances the overall performance of the system. The MBS provides better results than UBS even then it faces a lot of challenges in its implementation as summarized below (Ribaric. Ribaric, et al., 2003; Kumar and Farik, 2016) :

i. Difficult to implement and deploy
ii. User acceptability is very low
iii. The high cost is required to implement
iv. Need to consider which feature is to be used and in what quantity
v. At what level fusion is to be implemented

VIII. ATTACKS ON BIOMETRIC SYSTEM

The biometric system has been more secure than the traditional one (i.e knowledge-based and token-based) but still it has been susceptible to the different attacks. These attacks have been categorized below (Dahea and Fadewar, 2018) and figure 6 represented the location of attack in a biometric system. Spoof Attack: Fake biometric can be shown to a sensor.

Replay Attack: Biometric data has been injected directly into the feature extractor bypassing the data acquisition through sensor.

Substitution Attack: Feature extractor program or decision-making program has been replaced by malware so an attacker hack the system and change the identity of registered person as per his need

Replace Attack: the Real set of feature values have been replaced by synthetic set of values, modified or stolen from repository of registered person

Transmission Attack (TA): Data has been changed during transmission from one module to another

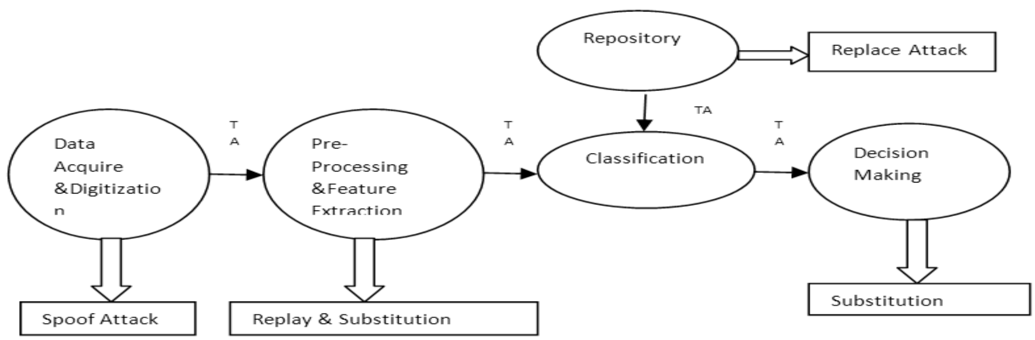


Fig. 7. Locations of Attack in a Biometric system

IX. COMPARATIVE STUDY OF DIFFERENT MBS

On the basis of the limitation of a UBS many users shifted toward the MBS. A comparative study of various MBS has been tabulated in table 1. Based on this comparison it has been found that there is lot of work done at feature extraction and classification level fusion but the area of data acquire and decision-making fusion level is still need a lot of scope.

Table 1: Comparative Study of Different Multimodal Biometric Systems

S. No	Author and Year	Modality Fused	Fusion Level
1	Kisku, Rattani, et al. (2009)	Face and Palm Print	Data Acquire
2	Yazdanpanah, Faez, et al. (2010)	Face, Ear, and Gait	Classification
3	Mahesh and Swamy (2010)	Voice and Palm Print	Classification
4	Zhu and Zhang (2010)	Finger-print, Knuckle-print, & palm-print	Feature Extraction
5	Gargouri, Masmodui, et al. (2011)	Finger-print and Face	Classification
6	Hariprasath and Prabakar (2012)	Iris and Palm	Classification
7	Kim, Shin, et al. (2012)	Face and Iris	Classification
8	Hamdani, Chekima , et al. (2013)	Heart and Speech	Classification
9	Huang, Liu, et al. (2013)	Face and Ear	Feature Extraction
10	Bhaskar and Veluchamy (2014)	Finger-knuckle & palm- print	Feature Extraction
11	Miao, Sun, et al. (2014)	Face and Iris	Classification
12	Jagadiswary and Saraswady (2016)	Finger-print, Retina, & Finger- vein	Feature Extraction
13	Parkavi, Babu, et al. (2017)	Finger-print & Iris	Classification
14	Khoo, Goi, et al. (2018)	Finger-print & Iris	Feature Extraction
15	Kim, Song, et al. (2018)	Finger vein & Finger shape	Classification
16	Choras, (2019)	Dorsal vein, periocular & palm print	Feature Extraction

X. CONCLUSION

There are number of multimodal biometric systems has been developed and implemented for recognizing a person, choice of biometric traits, optimal level of fusion and redundancy in the extracted feature are still some challenges faced in development of the MBS. The different level of fusion, types and the different mode of operation were discussed here. By combining more than one biometric the recognition rate increases which overall enhances the performance of system.

REFERENCES

[1] Bhaskar, B., and Veluchamy, S. (2014). Hand based multibiometric authentication using local feature extraction. *International Conference on Recent Trends in Information Technology*. (pp. 1-5).

[2] Choras, R. S. (2019).Multimodal Biometrics for Person Authentication. (pp. 1-18). IntechOpen

[3] Clarke, R. (1994). Human identification in information systems: Management challenges and public policy issues. *Information Technology & People*, 7(4), 6-37.

[4] Dahea, W. and Fadewar, H.S. (2018). Multimodal biometric system: A review. *International Journal of Research in Advanced Engineering and Technology,* 4(1), 25-31.

[5] Gargouri, B., A., N., Masmoudi,A.,D., and Masmoudi, D.,S. (2011). A New Human Identification based on Fusion Fingerprints and Faces Biometrics using LBP and GWN Descriptors.*International Multi-Conference on Systems, Signals and Devices (SSD), Vol.* 1(7). (pp. 22-25).

[6] Hamdani, O. A., Chekima, A., Dargham, J., Salleh, S.H., Noman, F., Hussain, H., Ariff, A.K., and Noor, M. (2013). Multimodal Biometrics Based on Identification and Verification System. *Journal of Biometrics and Biostatistics*, 4.

[7] Hariprasath, S., and Prabakar, T., N. (2012). Multimodal biometric recognition using iris feature extraction and palmprint features. *IEEE-International Conference on Advances in Engineering Science and Management (ICAESM)*. (pp. 174-179).

[8] Huang, Z. Liu, Y., Li, C., Yang, M and Chen, L. (2013). A robust face and ear based multimodal biometric system using sparse representation. *Pattern Recognition*, 46(8), 2156-2168.

[9] Jagadiswary, D. and Saraswady, D. (2016). Biometric Authentication using fused multimodal biometric. *Procedia Computer Science*, 85, 109-116.

[10] Jain, A., Ross, A., A., and Nandakumar, K. (2011). *Introduction to biometrics* Switzerland: Springer.

[11] Khoo, Y., H., Goi, B., M., Chai, T., Y., Lai, Y., L., and Jin, Z. (2018). Multimodal biometrics system using feature level fusion of iris and fingerprint. *International Conference on Advances in Image Processing.* (pp. 6-10)

[12] Kim, Y.G., Shin, K. Y., Lee, E. C. and Park, K.R. (2012). Multimodal Biometric System Based on the Recognition of Face and Both Irises.*International Journal of Advanced Robotic Systems*, 9(3), 1-6.

[13] Kim, W., Song, J.M. and Park, K.R. (2018) Multimodal biometric Recognition Based on Convolution Neural Network by the Fusion of Finger Vein and Finger Shape using Near –Infrared(NIR) Camera Sensor. *Sensors (Basel),* 18(7).

[14] Kisku, D., Rattani, A., Gupta, P., and Sing, J. (2009). Biometric Sensor Image Fusion for Identity Verification: A Case Study with Wavelet-Based Fusion Rules Graph Matching. *IEEE Conference on Technologies for Homeland Security, HST '09, Vol.* 433(439). (pp.11-12).

[15] Kumar, K. and Farik, M. (2016). A Review of Multimodal Biometric Authentication Systems. *International Journal of Science & Technology Research,* 5(12), 5-9.

[16] Mahesh, P.K., and Swamy, M., N., S. (2010). A Biometric Identification System based on the Fusion of Palmprint and Speech Signal. *International Conference on Signal and Image Processing (ICSIP)*, *Vol.* 186(190). (pp. 15-17).

[17] Miao, D., Sun, Z., and Huang, Y. (2014). Fusion of multibiometrics based on a new robust linear programming. *International Conference on Pattern Recognition*. (pp. 291-296).

[18] Parkavi, R.,Babu, K., R., C., and Kumar, J., A. (2017). Multimodal Biometric for user authentication. *IEEE International Conference on Intelligent Systems and Control.*

[19] Ribaric, S., Ribaric, D., and Pavesic, N. (2003). Multimodal biometric user-identification system for network-based applications.*IEE proceeding – Vision, Image and Signal Processing*, *Vol.* 150(6). (pp. 409-416).

[20] Sareen, P. (2014).Biometrics – Introduction, Characteristics, Basic technique, its Types and Various Performance Measures. *International Journal of Emerging Research in Management & Technology*, 3(4), 109-119.

[21] Yazdanpanah, A., Faez, K., and Amirfattahi, R. (2010). Multimodal Biometric System using Face, Ear and Gait Biometrics. *International Conference on Information Sciences Signal Processing and their Applications (ISSPA)*, *Vol.* 251(254). (pp. 10-13).

[22] Zhu, L.Q. and Zhang, S.Y. (2010).Multimodal biometric identification system based on finger geometry, knuckle print and palm print. *Pattern Recognition Letters*, 31(12), 1641-1649.

Role of Artificial Intelligence in Assisted Living Centers for the Elderly

Shreyata Khurana[1], Preeti Chauhan[2]

Department of Architecture and Planning
Indira Gandhi Delhi Technical University For Women, Kashmere Gate, Delhi-110006
[1]shreyatakhurana97@gmail.com, [2]preetichauhan@igdtuw.ac.in

Abstract: This research paper is about the contribution of artificial intelligence in enhancing the quality of life offered in assisted living centers along with overcoming the issues faced by the elderly people in such a setup. As per statistics, the number of seniors over the age of 65 reached a record of 1 billion people. This figure is going to be double by 2050. The number of older adults worldwide is increasing faster than any age group. It will be a challenge for the assisted living industry to meet the increasing demand. Without restructuring the approach of providing such services, it might not be possible to deliver quality care to people. Big data and artificial intelligence technology is going to play an extremely important role in the near future in the future of senior care[1]. An assisted living residence or assisted living facility is a day care housing facility for those people who need regular supervised medical care. Such centers are either day-care or residential abode to these people who get assisted medical help at all times along with the opportunity to socialize with the peer group. These centers basically provide residential, socializing and primary healthcare facilities. These facilities, also known as personal care homes, have been differentiated from nursing homes by their home-like design, personalized services, and resident independence and privacy. Despite these facilities, negative psychological and physical reactions like depression, anxiety and loneliness are also associated with moving to and residing in assisted living facilities. The people depending on such centers include the retired people, people on medication who might need regular or occasional caretaking or those whose children are from working class group. People are either living alone or with the working children and are left in the care of maids, house helpers or house nurses. But their quality of life is growing worse and they become the victim of depression, loneliness, anxiety and restlessness. Moreover, the working children are always on a quick run in case anything may happen to their elderly at home. There might also be some cancer strugglers or end of life people who need to move out of their frame of constant self-pity, depression, loneliness and enjoy the rest of their lives along with proper medical care. There may also be some child who is suffering from dengue or jaundice and is on bed for few weeks, but his parents can't take so many leaves and take care of the child at home so such children are also taken care of in such centers.

Artificial intelligence refers to machines and computers that are programmed to simulate the way that humans think, learn and work. The systems collect data over time, analyze it to deliver accurate information. This can save a lot of time of the service providers and enable them to spend more time with the patients to enhance their quality of life.

Keywords: Assisted Living, Artificial Intelligence, Passive Health Monitoring, Employee Turnover, Automation.

I. INTRODUCTION

Several factors are driving the demand for retirement housing in India. One of the most significant changes in our world today is the shift in population demographics. The people of the world are getting older. In 2000, there were 600 million people aged 60 and over, there will be 1.2 billion by 2025 and 2 billion by 2050.In 2050, the number of older persons in the world will exceed the number of young for the first time in history[2]. The country's overall population is growing at 1.8% annually but the rate of senior citizens is increasing at a much faster rate, i.e. 3.8% per year. In recent years, the number of people aged 65 or over has risen by 2% a year and the fastest growing age group is people over 80[3].

People are living longer today for several reasons including advances in medical science technology, health care, nutrition and sanitation. An important consequence of this progress is that those aged 80 or older are the fastest growing age group in the world. Although this larger older population is in better health than even before, they have some modified abilities. Sensory, cognitive and physical health, mobility and dexterity changes are prevalent among older persons, and raise many questions about the ways one thinks about human-environment interaction.

Assisted living communities face a number of challenges like client accidents (like falls), high client to employee ratio, high employee turnover, family distrust, increasing regulations, increasing workload and outdated technology. When some member of the family is in the assisted living facility, it is both financial as well as emotional stress on families. The "sandwich" generation is responsible for caring for their ageing parents and supporting their children. It can multiply the efforts of a lean staff while maintaining a high quality of patient care and save caregivers from executing mundane tasks like scheduling and billing by deploying automation processes so that caregivers can spend more quality time with patients and bring positive health outcomes. In terms of elderly care, it can envision a better way to passively monitor residents using technologies like wearable sensors, Ultra Wideband (UWB) radar, and machine vision to train AI engines. It can help to absorb and track a person's behavior, and then, predict health issues before they become medical emergencies.

As the saying goes, "Prevention is better than cure", similarly, identifying potential hazards and preventing issues before they occur reduces injury, reduces healthcare costs and decreases staff overload.

1. Alternating Requirements Adding to the Sentimental Values of the Elderly

After reading various reports, articles and research papers on the elderly the following points can be summarised[4][5]. The elderly face a number of challenges in their survivals as a result of increased likelihood of having Physical, Sensory and Cognitive Deficits in their lives.

- Migration of young couples from Delhi to abroad in search of better employment opportunities to fend for themselves.
- Elders who have been in control of the household for a long time are unwilling to give up the responsibility of their children.
- Youngsters on their part are sometimes resentful of the attitude of their parents.
- Many youngsters have moved to places far away from their native homes and in the recent past to many countries abroad. So even if they want to they cannot accommodate their parents in their own homes.
- Elders are sometimes too incapacitated or unwell to look after themselves or get medical care especially in an emergency.

A few design principles that are proposed keeping in mind the needs of the elderly are follows. These points also form an alternating base for experimenting with the concept of artificial intelligence to be included in the already existing as well as upcoming assisted living centers.

2. Principles of Universal Design

1. **Equitable Use**- Housing is usable by anyone, and does not disadvantage, stigmatize, or privilege any group of users. No-step entries are an example of an equitable features that allows all people to enter the dwelling in the same way.
2. **Flexibility in Use**- Living environments accommodate not only a wide variety of individual choices, but also adapt to user`s varying functional abilities. For example, placing kitchen counters at various heights permits people who are of tail or short status or those who are in seated positions to prepare food in a comfortable manner.
3. **Easy and intuitive** – All aspects of the domestic surroundings are straightforward to grasp, no matter the inhabitant`s experience, knowledge, language skills or concentration level. Bathroom faucets that create operation apparent which clearly indicate temperature levels are an example of a universally designed

solution. Light switches that are consistently located in relation to room entrances and that contain uniform "on/off" indicates help people to intuit lighting operation.

4. **Perceptible Information**- The housing communicates all necessary information effectively to all users regardless of ambient conditions or the user`s varying cognitive or sensory abilities. Both auditory and visual warnings on appliance buzzers and security alarms alert people to important information, and circumvent negative situations that might occur because of low vision, hearing limitations, environmental noise, and dark or clouded spaces.
5. **Tolerance for Error**- The design of residences minimizes hazards and adverse consequences of accidental or unintended actions by all users. Built-in-shower seats are an example of a universally designed feature which can prevent slips and falls while bathing. Niches for keys and other items near every entrance door help users to remember where they locate items that are often misplaced.
6. **Low Physical Effort**- Everyone can use the dwellings efficiently, comfortably and with minimal fatigue. Locating all basic living requirements on one entrance grade level reduces effort for those with mobility difficulties.
7. **Size and Space for Approach and Use**- Housing provides appropriate size and spaces for approach, reach. Manipulation and use regardless of the user`s body size, posture or functional abilities. For example, wide doorways and passageways provide a clear path of travel throughout the dwelling for all inhabitants. Reachable cabinets give users access to all stored items[6].

II. METHODOLOGY

Initially, research papers, Indian and foreign studies were done through secondary resources, i.e., books, journals, conference papers. A need assessment study along with the primary and secondary case studies were done. An online survey was conducted and discussions were done in various assisted living centers like Panchvati Homes, Tughlaqabad, Ashiana Care Homes (Bhiwadi), Epoch Assisted Living (Gurgaon) and Golden Estate, Faridabad. After interaction with people and taking the valuable inputs of the people, the study was taken forward. Loop holes were found from the previous studies that have already been done for assisted living centers and similar set ups. The proposals are made to improve upon the functioning of assisted living centers.

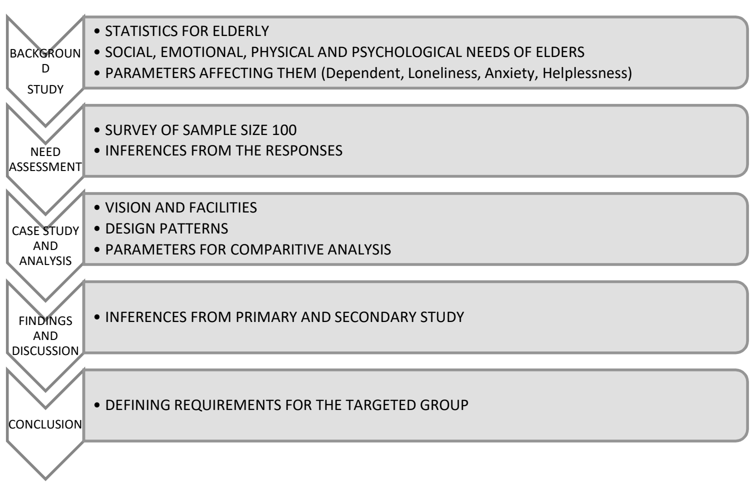


Fig. 1. Flowchart explaining the process of research

III. NEED ASSESSMENT SURVEY

After understanding the needs and requirements, they are now translated to an innovation stage for achieving best out of the blend of artificial intelligence in the concept of assisted living. The need assessment was done by doing a primary survey in a sample size of hundred.

1. Questionnaire for Establishing Need for Assisted Living Centers

A multiple-choice questionnaire was circulated which covered all the following aspects:

1. Basic information: Name, Age, Working/Non-working
2. What is your parents' age?
3. Are their parents retired/ still working (part time/ full time)/non-working?
4. Do they have any medical ailment?
5. Do they need regular medical assistance/help for daily chores?
6. Who is there at home to look after them when you go to your workplace?
7. Does your stress level increase when you get an emergency call from your home (regarding your parents)?
8. Does this affect your concentration level and productivity at your workplace?
9. Would you prefer to send your parents to an assisted living center (day-care/residential) to live, socialize and recreate rather than leaving them alone at home?

2. Observations, Findings and Discussion

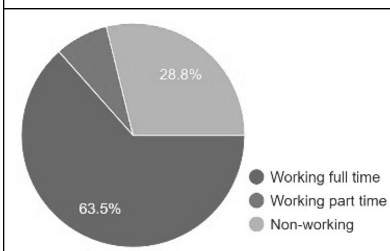
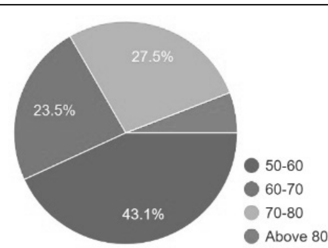
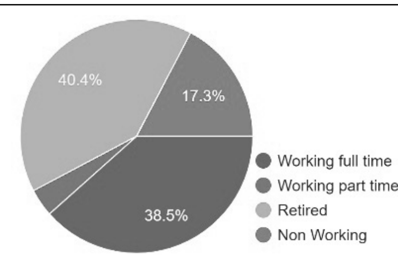
The surveyed sample space of 100 includes the people in the age group between 30-50 years of age whose parents are in the old age. They would be able to connect to the concept of assisted living centers as they see their parents struggling for the survival (whether it is in terms of medical assistance, mobility, daily chores etc). Here are a few observations from the survey held.

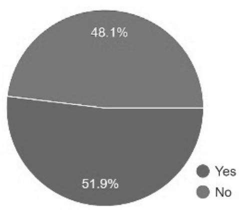
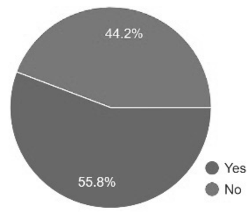
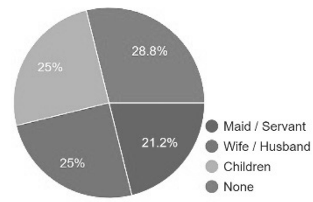
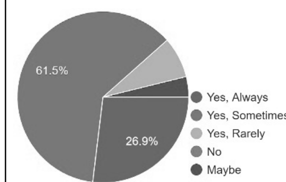
Majority of the elderly population lie in the retired age group or are in the non-working class (67.7 %) which states that they are mostly at home for the entire day. Their mobility has recessed and has limited to going to nearby places like park, temple or sometimes market. This is the time when their health actually starts declining and they become dependant on others for some or the other reason.

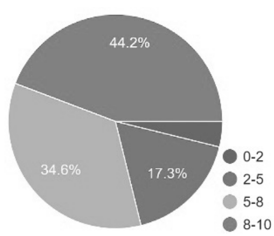
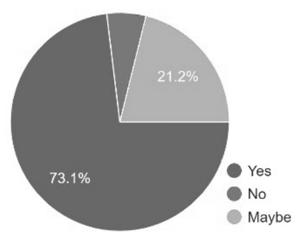
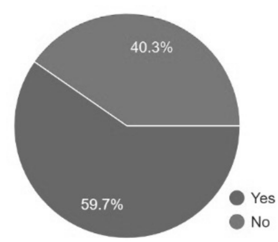
As per the survey done, their children are not always able to fulfil the emotional needs of their parents as they are out throughout the day at their workplaces and they are not able to spend quality time with their parents. As a result, they are there at home with either maids, servants (21.2%) or grandchildren (25%) or all alone (28.8%). When they are alone at home, the chances of accidents increase. Even if they are capable of handling daily chores on their own, they might need some emergency medical assistance.

After understanding the need of assisted living center, 55.8% of the people agreed to leave their parents in either day care or assisted living residential abode rather than leaving them in a more vulnerable situation at home all alone.

The graphical analysis of the survey is as follows:

Which group do you belong to?	In which age group your parents belong to?	In which working group your parents belong to?
		
63.5% belong to the working class (full time) and 28.8 % belong to the working class (part time).	The age group of parents on which the study is focussed is above 50 years.	Majority (40.4%) of the parents are retired and 38.5 % belong to working full time and 17.3% belong to the non-working group.

Do they need assistance for daily chores?	Do they need regular medical assistance?	Who is there to look after your parents when you are at your workplace?	Do you feel that you are able to fulfil their emotional needs?
			
51.9% people need assistance for their daily chores.	55.8% people need regular medical assistance.	The elders are left behind with their maid/servant or children. Still 28.8% remain alone.	Most of the people agree that they are sometimes (61.5%) able to fulfil the emotional needs of their parents.

How much does your stress level when you know that you can get an emergency call from home (regarding your parents)?	Does this stress affects your concentration and productivity at work?	Would you want to leave your parents in assisted living (which has a trained nurse in your personal home-like environment as there might be cases of medical emergency?
		
44.2% people agree that their stress level increases between 8-10 (on the scale of 0-10) when they get an emergency call.	73.1% people agree that their increased stress level affect their productivity and concentration at work.	Afetr understanding the concept of assisted living centers, 59.7% people agree to leave their parents in such centers rather leaving them alone at home.

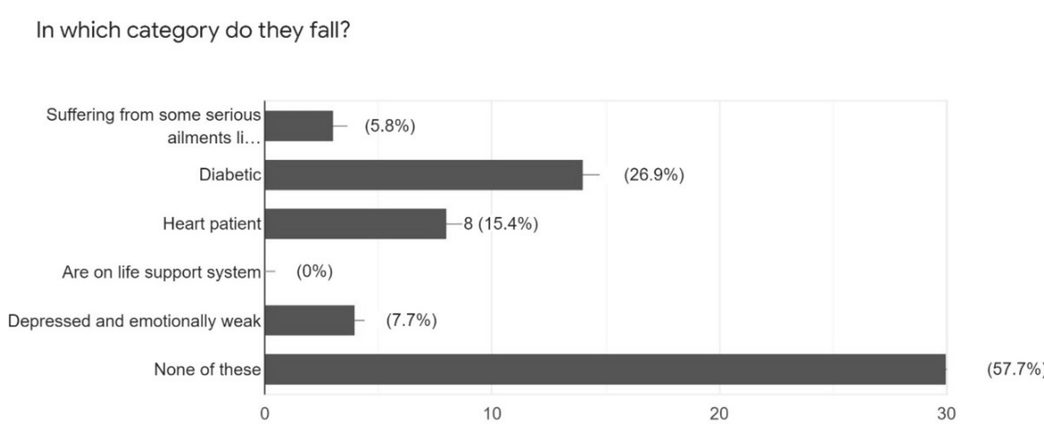


Fig. 2. Graphical Analysis of Survey Done

IV. PROPOSALS FOR AMALGAMATION OF ARTIFICIAL INTELLIGENCE IN ASSISTED LIVING

There are many prototypes which can be adopted in such centers to assess the efficiency and performance in different setups using artificial intelligence. The applications of artificial intelligence can also help an employer's retention and quality with voice activated devices and smartphones apps that can be used by the staff to receive guidance, automatically log mileage, capture notes, send messages to families, talk to physicians and more.

Employee turnover: Can also be improved using artificial intelligence via using a voice activated wearable device which can be operated from a phone or tablet as staff retention is one of the biggest challenges faced today.

Automatic Log: Automatic data collection and analysis by sensors will eliminate a significant amount of paperwork and also save a lot of time of the staff which they can utilize in communicating, providing better care and helping them socialize with their peer group by organizing activities, events or trips.

Capture Notes and Receive Guidance: Record the activity behavioral patterns in real time for performance of people in such centers to make accurate performance graphs and give proper medication as per requirement. Real time monitoring also enables to take guidance from the professions all over the globe.

Passive Health Monitoring: It is one way by which the health of patients can be monitored more accurately using artificial intelligence like sensors recording body temperature, fall indicators, heartbeat, motion, respiratory rate etc. depending upon the need of the patient. The information registered can then be processed using software to interpret patterns and predict potential medical crises, including heart disease, infections and stroke.

Preventing Falls: According to CDC data, older people are treated in the emergency room for fall-related injuries every eleven seconds. Falls can also be prevented in a similar manner by monitoring the patterns of a person in terms of balance, gait, stride, length and walk speed. This information database of every person can be used to alert the family members, care takers and healthcare professionals to reduce fall risks.

Updating family members: Via messages or real time video sharing of the patients about their daily activity and improvement so that they can be at a higher mental peace level for their loved ones, can also be assured of the proper care that is being received at the assisted living facility and also allowing the staff members to provide personalized care for the residents.

Increased staff efficiency and improved retention: Systems and devices using the concept of artificial intelligence can help scheduling the monthly calendars and notifying residents about upcoming events, meals and appointments in an easier and efficient manner. Digital assistants can also be used in such an environment with voice recognition capabilities which would allow staff to dictate notes, send messages to family members as well as healthcare professionals.

V. RESULTS

After getting to know about the prototypes incorporating artificial intelligence that can be used in an assisted living facility set up, it can be clearly stated that this would be beneficial on both the hands, for the residents as well as for the employers. This can save a lot of time of the staff and also enable the residents to enjoy a more quality care.

VI. DISCUSSION

As it is evident from the statistics, that the number of old adults is increasing at a very high rate and hence, there would be an increased demand for the assisted living industry to meet the requirements. The critical aspect of this would be ensuring the same or rather enhanced quality to the residents so that they feel like coming happily into the personalized home care environments or else, there would be no significant difference left out in nursing homes, old age centres and assisted living centres. The quality of life for employees is as important as it for senior residents. If technology can help improve inefficient processes and support in their day to day tasks, job, they will feel empowered and it can result in high job satisfaction. It can change the ways in which an assisted living facility operate. The biggest advantage of artificial intelligence is that it is not a stand-alone technology rather can complement to the existing technologies to make it more impactful.

VII. CONCLUSION

The concept of using artificial intelligence in assisted living centers in numerous techniques can actually help in reducing accidents and rendering quality services to them. Using artificial intelligence to improve the quality of life of the residents at the assisted living facilities can enable their family members to lead a more peaceful life as they would not be in the state of panic in case of some or the medical emergency of their loved ones. This concept can actually contribute to various factors in near future and could be further taken up for the research by innovating more concepts which could be beneficial for the service providers as well as the takers.

REFERENCES

[1] Borah, Hiranya. “Elderly In India.” Government of India. Feb. 2016.

[2] G.Giridhar. *Caring for Our Elders: Early Responses, India Ageing Report*. United Nations Population Fund (UNFPA), 2017.

[3] United Nations, Department of Economic and Social Affairs, Population Division (2015). World Population Ageing 2015 (ST/ESA/SER.A/390).

[4] Lewinson, T., Robinson-Dooley, V., and Grant, K. (2012). Exploring “home” at an assisted living facility: Looking through residents’ lenses with Photovoice.Journal of Gerontological Social Work, 55(8), 745-757. DOI: 10.1080/01634372.2012.684758

[5] Kane, Rosalie A. “Assisted Living Literature Through May 2004: Taking Stock.” *The Gerontologist, Vol 47, Special Issue III, 125-140*, May 2004.

[6] Eckhard Fedderson and Insa Ludtke. E-book *A Design Manual – Living For the Elderly*, 2018.

[7] Dave Cohen and Carlos Matos. *Third Year Projects – Rules and Guidelines*. Royal Holloway, University of London, 2013.

[8] Spillman, B. C., Liu, K., & McGilliard, C. *Trends in residential longterm care: Use of nursing homes and assisted living and characteristics of facilities and residents*. Retrieved September 2007, from http:// aspe. hhs.gov/search/daltcp/Reports/rltct.htm(2002).

Empirical Investigation of Quadruple Evolutionary Techniques for Test Case Selection & Prioritization

Shweta Singhal, Bharti Suri

Guru Gobind Singh Indraprastha University, New Delhi, India

Abstract: God's creations continue to amaze us every now and then. There are numerous species that have intelligence for survival with the varying approaches they use. Ants are blind, yet any sweet if left in open can be found with ants within a matter of time. Such are their exploration capabilities. Software researchers continue to benefit from such evolutionary techniques by digitalizing them to recreate their approaches with adaptation to solving a particular problem. Test case selection and prioritization based on the fault coverage criteria within a time constrained environment is one such regression testing problem. Over the years the problem has been solved using many evolutionary approaches including Ant Colony Optimization, Bee Colony Optimization, combination of Genetic Algorithms and Bee Colony optimization and a Greedy Set Cover Technique. These four techniques have not only been successfully applied to regression testing, their developers have also implemented them in form of tools. This paper attempts to empirically investigate the quadruple regression testing techniques on eight open source test programs. The results convince the researchers once again with the profoundness of these approaches. Though there are some situations where one of the techniques outperforms the other, overall all the quadruple techniques produce highly motivating execution time and size gain at minimal cost and with excellent accuracy.

Keywords: Ant Colony Optimization, Bee Colony Optimization, Genetic Algorithms, Greedy Set Cover, Software Testing, empirical comparison

I. INTRODUCTION

Maintenance phase is a highly demanding software engineering phase, both in terms of time and cost (money and manpower). The intent of regression testing lies in confirming that the updated software runs correctly and detects whether or not the unchanged code is harmfully affected. Regression testing is obligatory to reassure that no fresh faults have been introduced in the system due to modifications. The problem area is to find and run the smallest possible test suite that can be assuring enough with minimum costs involved. However, such problems can be mapped mathematically in order to reach a solution with mathematical optimization, especially with the use of metaheuristics. In this regard, a new research paradigm has emerged which is referred to as Search Based Software Testing (SBST). This field pacts with answering problems of software engineering via the optimization techniques. Time constrained test case selection and test case prioritization are combinatorial optimization problems. Evolutionary techniques like Ant Colony Optimization (ACO), Bee Colony Optimization (BCO), and Genetic Algorithms (GA) etc. are approaches based on the real behavior of ants, bees and human DNA and have already been applied to test case selection & prioritization [1], [2] , [3]. Along with these a time constrained greedy set cover (tGSC) approach has also been taken advantage of in resolving the problem of regression test case selection and prioritization [4]. The quadruple evolutionary approaches (ACO, BCO, BCO_GA and tGSC) have been empirically compared with each other for 8 sample C and Java programs for the selected and prioritized resultant test suite obtained. Each technique was executed for each of the programs 10 times and the results have then been averaged. The current study tries to compare the four techniques in the area of regression test selection and prioritization.

The paper is organized as follows: Background of Techniques: this section describes the four techniques that have already been developed and implemented for Test Case Selection and Prioritization (TCSP). The next section describes the Experimental Design and Comparison, having details of the results obtained and the analysis with the help of some graphs. Final section gives the possible threats to validity and the conclusion.

II. BACKGROUND OF TECHNIQUES

The problem of TCSP has been researched since 1990s and it still continues. Over the three decades of research many techniques and algorithms have been developed and tested for TCSP. There can be various criterion for selection and prioritization such as code coverage, branch coverage, statement coverage, fault coverage etc. Some of these criterion were initially investigated in [5]. Prioritization based on executing test data historically was also developed [6]. An empirical study [7] was accomplished on various greedy approaches to test case prioritization. An additional criterion of time awareness in TCSP was proposed by Walcott et al [8]. The current paper focuses on fault coverage based test case prioritization within a time constrained environment. Four evolutionary combinatorial optimization approaches have been selected for comparison. Ant Colony Optimization (ACO) was primarily used for TCSP in [9] and the same was implemented in [1]. Similarly, Bee Colony Optimization (BCO) was established for TCSP in [2]. A hybrid approach combining Genetic Algorithms (GA) and BCO was also implemented for TCSP [3]. Recently, in 2019 a greedy approach based on Set Cover in a time constrained environment was proposed [4]. These four approaches form the basis of this research work. The already developed tools for the four approaches were tested on 8 open source programs and the results have been then aggregated.

1. Ant Colony Optimization (ACO)

ACO forms a sequence of phases built on artificial intelligence search procedures for finding ideal results; formally proposed in [10], [11], [12]. Natural ants are tiny creatures unable to see, even then they have one of the best exploration capabilities not only in finding the food source but also a shortest path back to their nests. God's creation of their antennas and pheromone helps them accomplish this behavior. Ants use stigmergy to communicate and the rest of the colony converges to the shortest path. ACO is a digital adaptation of the same ant behavior. Artificial or digital ants must possess the ability to synchronize in exploring local solutions based on some previously gathered information and share it among their community so that a global solution can be reached. The concept of ACO was applied to TCSP initially in [9] and then implemented to form a tool ACO_TCSP [1]. The technique has further been empirically validated in [13]. The same tool has been used to gather the specifics in the present study. [14] also completed a survey on applications of the ACO in various turfs of software testing like generating test cases, selecting and prioritizing test cases etc. An improved ACO version [15] had also been proposed without letting the complexity of the existing technique being altered.

2. Bee Colony Optimization (BCO)

BCO is a digitalized form of the way a colony of honey bees forages for its food. It is a population grounded search technique having bees as the key employed agents. Different sorts of bees form a honey bee comb: A Queen Bee **-** Its role lies in laying eggs in order to support constructing the hive inhabitants, Male Drone Bees **–** They hold the responsibility of breeding along the queen bee in order to assist the hive growth, and Worker Bees **- t**he chief working agents of the bee colony. The hive conservation, bringing food for the colony, guiding associate bees to reach a source of food, stance as some of the key roles of worker bees. Worker bees are categorized as Scouts and Foragers. Scouts are supposed to explore new food sources while foragers exploit the already explored paths to reach the food source [16]. In order to search for the available food sources, the Scout bees begin randomly from the hive. Only when they are exhausted, do they return hive after their random exploration. The scout bees then share the gathered information with the forager bees via Waggle Dance. Waggle dance is performed in a particular manner to convey the information about food source quality, distance etc. After learning about the explored information from the scout bees, it is now the turn of forager bees to exploit the best food sources and bring back food from the best quality sources.

Artificial Bee Colony (ABC) was made known to the researchers in 2005 [17] for solving the problem of numerical optimization. Since then many researchers have benefitted from the applications of BCO/ABC algorithm to optimization problems like complex transportation problem, job shop scheduling, travelling salesman problem, p-center problem etc. [18], [19], [20], [21]. The BCO algorithm for solving fault based regression test prioritization within time constrained environment was proposed and implemented in [16], [2] respectively. The same implementation [2] forms the basis of BCO in the current research for comparison.

3. Hybrid Genetic Algorithms (GA) & BCO

Humans keep evolving and adapting to survive. Genetic Algorithms(GA) is an evolutionary approach grounded on survival of the fittest solutions for creating new populations that are fitting and better than the earlier ones. GA was introduced by John Holland [22]. The digital genetic procedure instigates by encoding an initial set of populations into a string format. Fitness equations based on some criteria that solves the problem under consideration are next used on random population. In case fitness equations are able to achieve the desired fitness values, the genetic procedure stops, else a novel set of populations is produced via crossover & mutation procedures. This entire process is looped till a pre-decided count of generations is reached or a solution meeting the desired fitness criterion is available.

The power of both the techniques (GA and BCO) was combined in a single algorithm to unravel desired solutions for regression test prioritization. This hybrid technique was developed by Suri et al [3]. The algorithm picks a new set of test cases from the prevailing test suite built on the maximum fault coverage and minimum execution time criterion. The approach was implemented as MHBG_TCS tool and empirically validated on 17 sample programs. The same algorithm forms the basis for third approach being compared in the current paper.

4. Time Constrained Greedy Set Cover (tGSC)

Set cover problem has been solved using greedy approach as early as in 1979 [23]. The Set Cover problem attempts to reduce the number of test cases via a delayed greedy approach in order to discover a new smaller set of the test cases that realize all the necessities achieved by the initial test suite [24]. The initial proposal for solving TCSP using set cover with the help of mutation testing was proposed in [25]. The technique was then improved by running it in a time constrained environment and adding some selection criteria. The running time of each test case was added as a criterion for selection of test cases in addition to the existing criteria of extreme fault coverage and smallest number of test cases chosen to produce the optimized test suite. 't-GSC' was thus established for forming the optimized test suite [4], the same approach has been used for comparison in the current study.

III. EXPERIMENTAL DESIGN

The ACO_TCSP tool along with the BCO, MHBG_TCS and t-GSC tools have already been developed and all of the four techniques take the faults killed and the running time of the test cases as their input. The output is an optimized test suite. All the tools yield a test suite, yet the answers are slightly varying for all the techniques. The tools have been run on eight open source programs (details given in the following section). Mutation faults were induced in all the programs and test cases generated. Then the fault matrix for each program was furbished to create the input required for the tools under comparison. Since three of the four techniques being compared are approximation techniques, thus they were run 10 times for every program and the results were then averaged for analysis. For analysis eight different open source C++ and JAVA programs have been considered. Ephemeral details about the Programs Under Test (PUT) like their length, versions (number of faults), and test suites are summarized in Table 1.

Table 1. Program Under Test Details

PUT No.	PUT Name	Size (in LOC)	No. of Versions (Faults)	Test Suite Size
P1	College_Admission	281	5	9
P2	Hotel_Management	666	5	5
P3	Triangle_sides	37	6	19
P4	Quadratic_eqn	38	8	19
P5	cost_of_publishing	31	8	19
P6	Calculator_	101	9	25
P7	previous_day	87	7	19
P8	railway_booking	129	10	26

IV. COMPARISON AND ANALYSIS

To establish our confidence in the soundness of the techniques under comparison, three aspects were compared. The amount of size efficiency, time gain and the percentage accuracy achieved by each of the four techniques were recorded and plotted. All the four techniques give highly motivating results with each of the technique being superior to other in one or the other aspect. These have been analyzed and the reasons explaining this behavior have been tried to be looked upon.

The amount of test cases forming the final test suite is lesser than the total amount of test cases building the original test suite. By what amount has the size of the test suite been reduced is tried be analyzed using the Averaged Percentage Size Efficiency (APSE) calculated as follows:

$$\text{APSE} = \left(\frac{|TS|-|RTS|}{|TS|}\right)*100 \tag{1}$$

Where, TS represents the Total Size of the initial Test Suite, RTS is size of the Resultant Test Suite. APSE has been calculated for all of ACO, BCO, hybrid BCO_GA, and tGSC tools and the details have been plotted in Figure1. As can be clearly observed from the line graph that all the 4 tools provide almost similar size efficiency varying from 60% to 85%. Only exception being the greedy tGSC approach which is found to underperform as compared to the approximation approaches.

Fig. 1. APSE (Average Percentage Size Efficiency)

Averaged Percentage Time Gain (APTG) has been calculated using the formula (2) and the aggregated results presented in Figure 2.

$$\text{APTG} == \left(\frac{RT-RRT|}{ET}\right)*100 \tag{2}$$

Where, RT is Total Running Time for the entire test suite of the PUT's (P1 – P8), RRT is averaged Running Time for the Resultant test suite. Yet again the plot of Fig. 2 clearly establishes the similarity in time gain achieved using the four approaches. Except for minor variations, all of the ACO, BCO, hybrid BCO_GA and the tGSC provide highly inspiring time gains as against running the entire test suite.

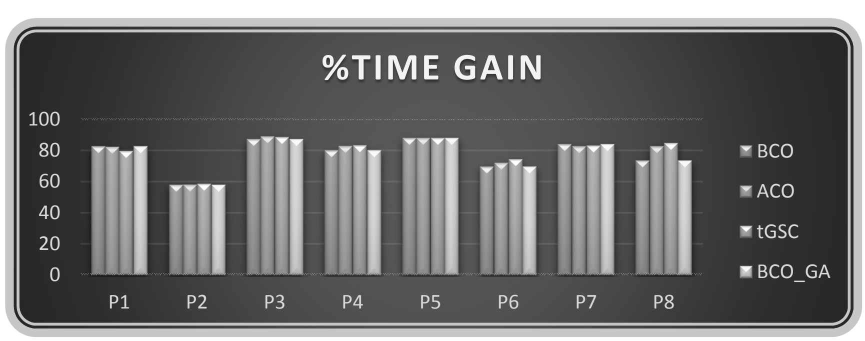


Fig. 2. APTG (Averaged Percentage Time Gain)

Averaged Percentage Accuracy (APA) has been designed using the formula (3) and the gathered results presented in Figure 3.

$$APA = \left(\frac{OR}{10}\right)*100 \qquad (3)$$

Where, OR is averaged number of Optimal Runs of the respective tool.10 is the total number of runs the tool was executed for each program (P1-P8). Optimal run refers to the output which is possibly the best out of all. Since, the eight programs under test are small, thus optimal output could be known and compared with. Fig. 3 represents the radar plot for APA of the four approaches. The desired is a complete hexagonal plot with maximum radius, that would represent 100% accuracy. But since ACO, BCO and hybrid BCO_GA are approximation approaches, thus 100% accuracy is not possible. Hybrid BCO_GA gave very good accuracy while ACO gave the best accuracy out of all quadruple techniques. tGSC being a greedy approach gave the worst accuracy for regression test selection and prioritization. This is as expected because TCSP is an NP Complete problem.

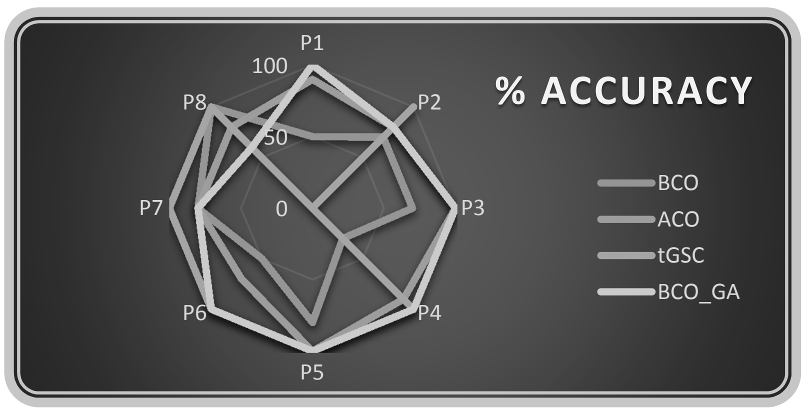


Fig. 3. APA-Averaged Percentage Accuracy

V. CONCLUSION

The quadruple approaches: t-GSC, hybrid BCO_GA, BCO and ACO techniques have been run on eight test programs for regression test selection and prioritization. Each of the four tools was run 10 times on every test program and the averaged results have been investigated. The investigation was accomplished for size efficiency, time gain and the percentage accuracy achieved by all the four techniques. The quadruple tools yielded excellent size efficiency of the resultant test suite along with huge time gain for the execution time of resultant test suite against the original test suite. The accuracy achieved by tGSC being a greedy approach was found to be least in comparison to others, while that of ACO was found to be the best. Although all the quadruple approaches yielded borderline better or worst results, yet all four gave excellent time and size gains with tGSC being worst and ACO being best in terms of accuracy of the technique. Future investigation can be made on more evolutionary algorithms with a larger data set.

REFERENCES

[1] B. Suri. and S. Singhal, "Implementing Ant Colony Optimization for Test Case Selection and Prioritization," *International Journal of Computer Science and Engineering 0975-3397. Vol. 3,* pp. 1924-1932, 2011.

[2] A. Kaur and S. Goyal, "Implementation and Analysis of the Bee Colony Optimization algorithm for Fault based Regression Test Suite Prioritization," *International Journal of Computer Applications. Vol. 41,* pp. 1-9. DOI:10.5120/5606-7867., 2012.

[3] B. Suri and S. Singhal, "Evolved regression test suite selection using BCO and GA and empirical comparison with ACO," *CSI Transactions on ICT. Vol. 3,* pp. 10.1007/s40012-016-0080-5., 2016.

[4] N. Jatana and B. Suri, "An Empirical Comparison of t-GSC and ACO_TCSP Applied to Time Bound Regression Test Selection.," *Recent Patents on Computer Science. Vol. 12.,* p. 10.2174/2213275912666190417152016., 2019.

[5] G. Rothermel, R. J. Untch and C. Chu, “Prioritizing Test Cases For Regression Testing,” *IEEE Transactions on Software Engineering. Volume 27 Issue 10,* pp. 929-948, 2001.

[6] J. Kim and A. Porter, “A history-based test prioritization technique for regression testing in resource constrained environments,” in *ICSE ‘02 Proceedings of the 24th International Conference on Software Engineering*, Orlando, Florida, 2002.

[7] H. Li and C. Lam, “Software Test Data Generation Using Ant Colony Optimization,” *World Academy of Science, Engineering and Technology,* pp. 1-4, 2005.

[8] K. Walcott, M. Soffa, G. Kapfhammer and R. Roos, “TimeAware test suite prioritization,” in *ISSTA ‘06 Proceedings of the 2006 international symposium on Software testing and analysis*, Portland, Maine, USA, 2006.

[9] Y. Singh, A. Kaur and B. Suri, “Test case prioritization using ant colony optimization,” *ACM SIGSOFT Software Engineering Notes, vol. 35, no. 4,* pp. 1-7, 2010.

[10] A. Colorni, M. Dorigo and V. Maniezzo, “Distributed optimization by ant colonies,” in *ECAL’91, European Conference on Artificial Life*, Amsterdam, 1991.

[11] M. Dorigo, “Optimization, learning and natural algorithms, Ph.D. thesis,” Politecnico diMilano, Milano, 1992.

[12] M. Dorigo, V. Maniezzo and A. Colorni, “The ant system: An autocatalytic optimizing process, Technical Report TR91-016,” Politecnico di Milano, Milano, 1991.

[13] B. Suri and S. Singhal, “Understanding the effect of time-constraint bounded novel technique for regression test selection and prioritization,” *International Journal of System Assurance and Engineering Management,* pp. DOI: 10.1007/S13198-014-0244-3, 2014.

[14] B. Suri and S. Singhal, “Literature survey of ant colony optimization in software testing,” in *CONSEG, CSI sixth international conference on software engineering*, Indore, 2012.

[15] B. Suri and S. Singhal, “Development And Validation Of An Improved Test Selection And Prioritization Algorithm Based On ACO,” *International Journal of Reliability, Quality and Safety Engineering, Vol. 21, No. 6,* p. DOI: 10.1142/S0218539314500326., 2014.

[16] A. Kaur and S. Goyal, “A Bee Colony Optimization Algorithm for Fault Coverage Based Regression Test Suite Prioritization,” *International Journal of Advanced Science and Technology of Korea. 29,* pp. 17-29, 2011.

[17] D. Karaboga, “An Idea Based on Honey Bee Swarm for Numerical Optimization. Technical Report-TR06,.,” Department of Computer Engineering, Engineering Faculty, Erciyes University., Turkey, 2005.

[18] D. Teodorović and M. Dell’Orco, “Bee colony optimization - A cooperative learning approach to complex transportation problems,” *Advanced OR and AI Methods in Transportation,* pp. 51-60, 2005.

[19] C. Chong, A. Sivakumar, M. Low and K. Gay, “A bee colony optimization algorithm to job shop scheduling,” in *WSC ‘06 Proceedings of the 38th conference on Winter simulation*, Monterey, California, 2006.

[20] L. Wong, M. Y. H. Low and C. S. Chong, “A Bee Colony Optimization Algorithm for Traveling Salesman Problem,” in *Second Asia International Conference on Modelling & Simulation (AMS)*, Kuala Lumpur, 2008.

[21] T. Davidović, D. Ramljak, M. Šelmić and D. Teodorović, “Bee colony optimization for the p-center problem,” *Computers and Operations Research, Volume 38 Issue 10,* pp. 1367-1376, 2011.

[22] J. Holland, “Genetic Algorithms and Adaptation,” in *Adaptive Control of Ill-Defined Systems. NATO Conference Series (II Systems Science), vol 16.*, Boston, MA, Springer, 1984, pp. https://doi.org/10.1007/978-1-4684-8941-5_21.

[23] V. Chvatal, “A Greedy Heuristic for the Set-Covering Problem,” *Mathematics of Operations Research, volume 4, no. 3,* pp. 233-235, 1979 .

[24] S. Tallam and N. Gupta, “A concept analysis inspired greedy algorithm for test suite minimization,” in *Proceedings of the 6th ACM SIGPLAN-SIGSOFT workshop on Program analysis for software tools and engineering*, Lisbon, Portugal, 2005.

[25] N. Jatana, B. Suri, P. Kumar and B. Wadhwa, “Test suite reduction by mutation testing mapped to set cover problem,” in *ICTCS ‘16 Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*, Udaipur, India, 2016.

A Review of Control Strategies for Microgrids

Ritu Kandari, Pankaj Gupta, Ashwani Kumar

Department of ECE, IGDTUW, Kashmere Gate, New Delhi, India

Abstract: Integration of renewable energy resources with utility grid causes number of interconnection issues. This paper presents the different strategies to manage the issues which arises due to integration of distributed energy sources with utility grid. One cannot merely rely on conventional techniques to limit the deviations in frequency or fluctuations in voltage. An overview of different techniques is presented in this paper showing their benefits over the conventional techniques. Most of the possible problems occurred due to integration of distributed energy sources and their controlling techniques are covered in this paper.

Keywords: Microgrid, Microgrid control, Distributed Generation, Voltage Control

I. INTRODUCTION

Distributed generation (DG) in power system is becoming more popular due to low carbon footprints, efficient and reliable power system. This concept of controlled microgrid is a significant and important solution to the problems like limited availability of conventional resources, environmental impacts of conventional resources, carbon dioxide emissions [1]-[3]. Also integration of distributed energy resources causes many issues like voltage and frequency unbalance, total harmonic distortion, fault ride through etc. So operation of microgrid is to be controlled using suitable controlling strategies [4] [5].

Increased demand of renewable energy resources makes distributed generation a great topic of research since the beginning of this century. Microgrid consists of distributed energy resources, power electronic devices, storage devices and load. Microgrids are designed such that they work in grid connected and autonomous mode. Controllers are hence required so as to manage the power, voltage, and frequency within a tolerable range. In Isolated mode, microgrids are designed to serve at least the critical loads.

Coordination between the various units connected to the microgrids is to be maintained for the efficient utilization of all the units and hence making the grid operational all the time. The stability is maintained by balancing the power between the connected units. Hierarchical control is provided to the system so as to maintain all the quantities in tolerable range.

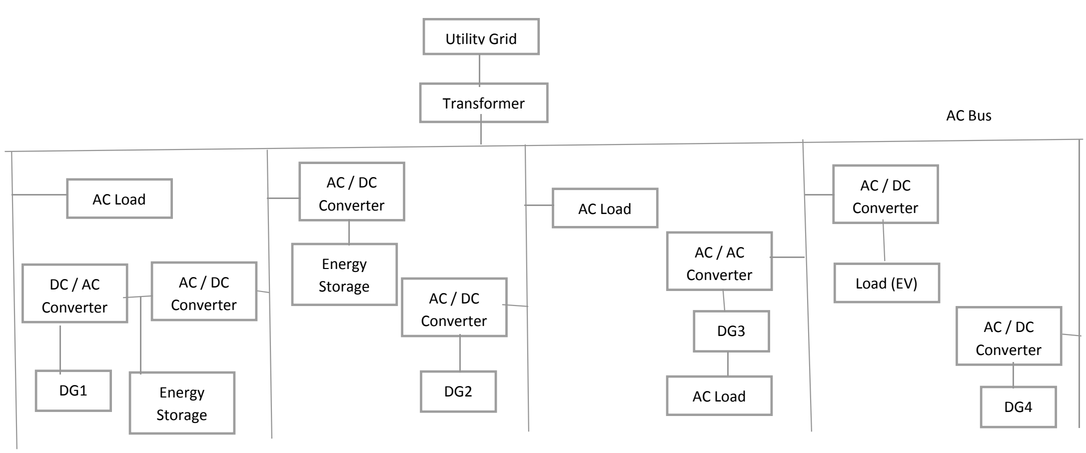


Fig. 1. Architecture of a Microgrid

Fig. 1 shows the architecture of a microgrid consisting of a main grid which is connected to four distributed energy sources, energy storage devices and loads through an AC bus.

II. ALTERNATING CURRENT MICROGRID (ACMG)

Fig. 2. shows AC microgrid architecture with interconnection with the grid. Switches are used to change the microgrid mode. The DGs of DC nature are connected through power electronic devices to the AC bus. During normal operation, the AC microgrid and the main grid are connected at the PCC. When the output of DG is more than the load requirement, extra power is send to the main grid. An AC microgrid provides reconfigurable systems and is easily integrated with the main grid. So as to follow the set standards for voltage and frequency, the AC microgrid system is provided with different control techniques. The main demerit of an AC microgrid is that it requires complex power electronic devices to stay in synchronization with the main grid. The power system parameters, frequency, voltage, phase angle, all need to be matched with the utility. AC microgrid has more conversion steps as compared to DC microgrid which is a limitation in the application of AC microgrid in the system.

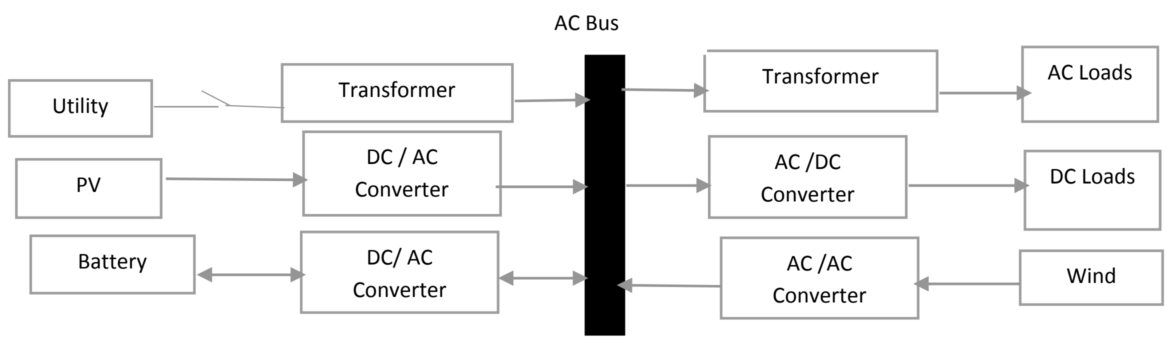


Fig. 2. AC Microgrid architecture

Most prominent is the hierarchical control due to its optimization and efficiency in controlling the key variables of the AC microgrid system [6]-[9].

III. CONTROL CHALLENGES IN AC MICROGRID

Depending upon the system parameters and prevailing conditions, best suited method need to be implemented in the system. So improved methods for regulation of voltage should be employed without getting affected by the system parameters. Proper control scheme should be employed for unbalanced microgrid system consisting of nonlinear loads in order to attain accurate power sharing. Droop control techniques are less reliable due to no interaction between DG units. Hence more reliable, communication based techniques should be employed. Centralized technique has a risk of single point failure, so focus should be more on decentralized techniques. Cooperative and intelligent methods to control the AC microgrid system should be employed.

IV. DC MICROGRID

A DC microgrid system can be easily implemented with distributed generation units because DG and energy storage devices gives DC output and hence it controls DC voltages. DC - DC converters connect PV, loads and energy storage with the DC bus. Diesel generators and wind turbines are connected to DC bus via converters. Fig. 3 shows the architecture of DC microgrid. It is simpler in implementation as synchronization of distributed generation is not required. Reliability issues are there in this type of system because serially connected power electronic interface is required.

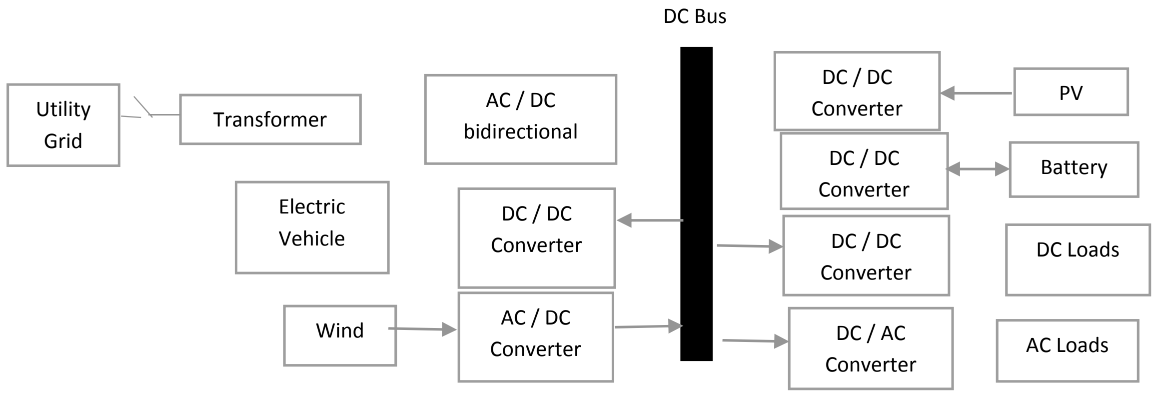


Fig. 3. DC Microgrid architecture

V. CONTROL CHALLENGES IN DC MICROGRID

In DC microgrid system a separate DC distribution lines are required. Absence of zero crossing of the current and voltage, so it is difficult to protect a DC microgrid system. Since only active power flow is there, voltage stability gets affected in the absence of reactive power flow. Hence accurate active power sharing is the major concern here [10]. Techniques like average load sharing, master slave operation are employed for power sharing and control of DC microgrid. Hierarchical control topology is employed in DC microgrid also except the fact that the control scheme and variable to be controlled is different in this case.

VI. HYBRID AC/DC MICROGRID (HMG)

A hybrid AC/DC microgrid has two subgrids, AC microgrid and a DC microgrid. The AC subgrid is connected to the PCC while the DC subgrid is connected to the AC bus via bidirectional converter. Control of this converter is crucial to manage the flow of power and stability of the system [11] [12].

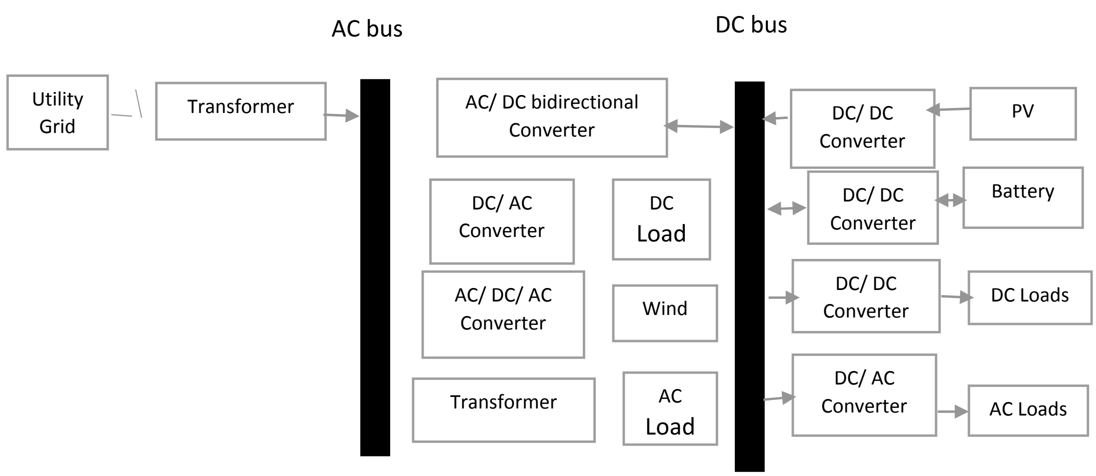


Fig. 4. Hybrid AC/DC Microgrid

A hybrid microgrid consists of DGs, loads and energy storage system. Power exchange between the different units is done using a bidirectional interlinking converter. Some DGs are connected to boost converters to maintain a desired voltage level. The implementation of hybrid microgrid needs to find solutions for many technical challenges such as appropriate control technique for optimal operation of hybrid microgrid, designing and control of interlinking converter, coordination control of power electronic devices. Hybrid microgrids will help in building sustainable grid which may further lead to Smart Grid as they create a power system with DG, storage, cogeneration with consumer control [13]-[17].

VII. CONTROL CHALLENGES IN HYBRID AC/DC MICROGRID

Since there is no global variable present in hybrid it becomes complicated and difficult to control a hybrid microgrid. Conventional methods are not successful in hybrid microgrid during isolated mode. A specific droop control is required for sharing and managing the power among the AC and DC microgrid sources that has to be independent of the system parameters like line impedance. Harmonic power sharing control need to be taken into account because of the nonlinear loads. Trade-off between the regulation of the voltage deviation mitigation and the power flow can also be seen. A complex energy management system is required for optimal and reliable system.

Table 1. Comparative chart of different Microgrid structures

MG Type	Advantage	Disadvantages
AC	Easy to reconfigure by existing grid	Difficult to synchronize with utility grid
DC	Lesser power converters required, DC loads can be served directly, grid synchronization is not required	Cannot be reconfigured from existing grid, Additional stage of power is required to generate AC power
Hybrid	No synchronization required. Can be connected directly to both the types of load. Energy loss is minimized.	Complex controllers, Less reliable

VIII. LITERATURE REVIEW

A detailed and comprehensive survey has been done. All the categories of enlisting the details and their respective reference numbers are reported in the form of tables. This provides a very understandable, clear and easily approachable way of explaining all the state of the art techniques of the research topic.

Table 2. Categorization on the basis of different methods to control one Parameter:

Table 2.1. Techniques for Voltage Control

Control Technique	Contribution	Inadequacies
Nonlinear disturbance observer control [18]	Enables plug-and-play feature, mitigation of voltage deviations, power quality improvement	
Adaptive power sharing mechanism [19]	Provides distributed control, improved transient response	Can be used only with fixed frequency hybrid AC/DC MG
Using Tie Converter [20]	Voltage at both the buses can be regulated, provides bidirectional power flow	Can be used for low voltage and low rating DERs, Energy demand by the houses and generation DERs cannot be measured.
Fuzzy logic-based intelligent control technique[21]	Frequency and voltage stability regulation, transient time is minimized, suppresses voltage and frequency deviations and maintains the power ratings of generators during disturbances	Load shedding has to be done

Table 2.2. Table for Seamless mode transfer

Control Technique	Contribution	Inadequacies
Intelligent load-shedding algorithm [22]	Seamless mode-transition	
Two phase lock loops [23]	Error between the phases of the PLL is reduced	
Modified Voltage based droop control [24]	Synchronization of frequency, voltage and phase with the utility grid is done	Resynchronization
LQR theory based transfer scheme [25]	Transients during transition are reduced	

Table 2.2 (Contd.)...

...Table 2.2 (Contd.)

Master-slave method [26]	Re-synchronization is achieved	Complex, failure of master DG causes microgrid failure.
Dispatch unit [27]	provides smooth transition, synchronize the inverters with the grid	

Table 2.3. Table for Communication Architecture

Control Technique	Contribution	Inadequacies
Three-level hierarchical control [28]	ac current and dc voltage are controlled, proportional current sharing is realized, applicable to both the modes of operation, control levels do not interfere with each other	Designed only for dc side
3-layer communication architecture [29]	Reduced system delay	low reliability, communication protocol cannot be unified
An adaptive integrated module [30]	Plug and play functionalities, ease of usage, rapid deployment	

IX. CONCLUSION

A brief review of various control techniques for microgrid is presented. It is studied that in a microgrid coordination between the different units is required for managing the power. In the condition of load fluctuations, multilevel converters provide fast response and manage the fluctuations. Voltage deviation can be regulated during saturation conditions of PV grid connected inverter by using thermostatically controlled loads. Need for communication link is minimized by using Lyapunov algorithm for voltage controller while managing the reactive power sharing accurately among the DGs. Sparse Communication network can be utilized to manage the voltage and frequency of a microgrid. An adaptive fuzzy PID controller can be used to control the frequency in a microgrid.

REFERENCES

[1] F. Katiraei, R. Iravani, N. Hatziargyriou, and A. Dimeas, "Microgrids management", IEEE Power Energy Magazine., vol. 6, no. 3, pp. 54–65, May/Jun. 2008.

[2] R. H. Lasseter, "Microgrids and distributed generation", Journal of Energy Engineering, vol. 133, no. 1, pp. 144–149, Sep. 2007.

[3] N. Hatziargyriou, H. Asano, R. Iravani, and C. Marnay, "Microgrids", IEEE Power ad Energy Magazine, vol. 5, no. 4, pp. 78–94, Jul./Aug. 2007.

[4] C. Marnay, F. J. Robio, and A. S. Siddiqui, "Shape of the microgrid", Conference Proceedings IEEE Power Engineering Society Winter Meeting, vol. 1. Columbus, OH, USA, pp. 150–153, Jan. 2001.

[5] R. H. Lasseter, "MicroGrids", Conference Proceedings IEEE Power Engineering Society Winter Meeting, vol. 1, pp. 305–308, Jan. 2002,

[6] Yazdani and P. P. Dash, "A control methodology and characterization of dynamics for a photovoltaic (PV) system interfaced with a distribution network", IEEE Transactions on Power Delivery, vol. 24, no. 3, pp. 1538–1551, Jul. 2009.

[7] Y. Li and Y. W. Li, "Virtual frequency-voltage frame control of inverter based low voltage microgrid", in Proc. IEEE Electrical Power and Energy Conference (EPEC), pp. 1–6, Oct. 2009.

[8] F. Katiraei, M. R. Iravani, and P. W. Lehn, "Micro-grid autonomous operation during and subsequent to islanding process", IEEE Transactions on Power Delivery, vol. 20, no. 1, pp. 248–257, Jan. 2005.

[9] Y. A.-R. I. Mohamed and A. A. Radwan, "Hierarchical control system for robust microgrid operation and seamless mode transfer in active distribution systems", IEEE Transactions on Smart Grid, vol. 2, no. 2, pp. 352–362, Jun. 2011.

[10] D. J. Hammerstrom, “AC versus DC distribution systems did we get it right?”, Proceedings of the IEEE Power Engineering Society General Meeting, pp. 1–5, Jun. 2007.

[11] E. Planas, J. Andreu, J. I. Garate, I. M. De Alegria, and E. Ibarra, “AC and DC technology in microgrids: A review”, Renewable and Sustainable Energy Reviews, vol. 43, pp. 726–749, Mar. 2015.

[12] I. Patrao, E. Figueres, G. Garcerá, and R. Gonzalez-Medina, “Microgrid architectures for low voltage distributed generation”, Renewable and Sustainable Energy Reviews, vol. 43, pp. 415–424, Mar. 2015.

[13] F. Blaabjerg, R. Teodorescu, M. Liserre, and A. V. Timbus, “Overview of control and grid synchronization for distributed power generation systems”, IEEE Transactions on Industrial Electronics, vol. 53, no. 5, pp. 1398–1409, Oct. 2006.

[14] X. Liu, P. Wang, and P. C. Loh, “A hybrid AC/DC microgrid and its coordination control”, IEEE Transactions on Smart Grid, vol. 2, no. 2, pp. 278–286, Jun. 2011.

[15] C. Jin, P. C. Loh, P. Wang, Y. Mi, and F. Blaabjerg, “Autonomous operation of hybrid AC-DC microgrids”, IEEE International Conference on Sustainable Energy Technologies. (ICSET), , pp. 1–7, Dec. 2010.

[16] R. Majumder, “A hybrid microgrid with DC connection at back to back converters”, IEEE Transactions on Smart Grid, vol. 5, no. 1, pp. 251–259, Jan. 2014.

[17] P. C. Loh, D. Li, Y. K. Chai, and F. Blaabjerg, “Hybrid AC–DC microgrids with energy storages and progressive energy flow tuning”, IEEE Transactions on Power Electronics., vol. 28, no. 4, pp. 1533–1543, Apr. 2013.

[18] C. Wang, Xi. Li, Li Guo and Y. W. Li, “A Nonlinear-Disturbance-Observer-Based DC-Bus Voltage Control for a Hybrid AC/DC Microgrid”, IEEE Transaction on Power Electronics, vol. 29, no. 11, Nov. 2014.

[19] A. Navarro-Rodrıguez , P. Garcıa, R. Georgious , and J.Garcıa , “Adaptive Active Power Sharing Techniques for DC and AC Voltage Control in a Hybrid DC/AC Microgrid”, IEEE Transactions on Industry Applications, Vol. 55, no. 2, March/April 2019.

[20] P. C. Loh, D. Li, Y. K. Chai, and F. Blaabjerg, “Autonomous operation of hybrid microgrid with AC and DC subgrids”, IEEE Transactions on Power Electronics, vol. 28, no. 5, pp. 2214–2223, May 2013.

[21] A. A. A. Radwan and Y. A.-R. I. Mohamed, “Networked control and power management of AC/DC hybrid microgrids”, IEEE Systems Journal, to be published.

[22] I. J. Balaguer, Q. Lei, S. Yang, U. Supatti, and F. Z. Peng, “Control for Grid-Connected and Intentional Islanding Operations of Distributed Power Generation”, IEEE Transactions on Industrial Electronics, vol. 58, no. 1, January 2011.

[23] J. Wang, N. Chialin, P. Chang, Xi. Feng, and A. Monti, “Design of a Generalized Control Algorithm for Parallel Inverters for Smooth Microgrid Transition Operation”, IEEE Transactions on Industrial Electronics, vol. 62, no. 8, August 2015.

[24] T. L. Vandoorn, B. Meersman, J. D. M. De Kooning, and L. Vandevelde, “Transition From Islanded to Grid-Connected Mode of Microgrids with Voltage-Based Droop Control”, IEEE Transactions on Power Systems, vol. 28, no. 3, August 2013.

[25] D. Das, G. Gurrala and U. Jayachandra Shenoy, “Linear Quadratic Regulator-Based Bumpless Transfer in G. Microgrids”, IEEE Transactions on Smart Grid, vol. 9, no. 1, January 2018.

[26] G. Talapur, H. M. Suryawanshi, R. R. Deshmukh, M. S. Ballal and R. Choudhary, “Hybrid Control Technique for load sharing and seamless transition in microgrid mode of operation”, pp. 1-6, IEEE, 2018.

[27] Md. N. Arafat, A. Elrayyah, and Y. Sozer, “An Effective Smooth Transition Control Strategy Using Droop-Based Synchronization for Parallel Inverters”, IEEE Transactions on Industry Applications, vol. 51, no. 3, May/June 2015.

[28] Xi. Lu, J. M. Guerrero, K. Sun, J. C. Vasquez and L. Huang, “Hierarchical Control of Parallel AC-DC Converter Interfaces for Hybrid Microgrids”, IEEE Transactions on Smart Grid, vol. 5, no. 2, March 2014.

[29] Z. Liang, Z. Yuanbo, Z. Liang and Z. Yuanbo, “Implementing a 3-layer Communication Architecture for AC & DC Hybrid Smart Microgrid”, Asian Conference on Energy, Power and Transportation Electrification, Sands Expo and Convention Centre, Marina Bay Sands, Singapore, October 2016.

[30] X. Pan, L. Zhang, J. Xiao, F. H. Choo, A. K. Rathore and P. Wang, “Design and Implementation of a Communication Network and Operating System for an Adaptive Integrated Hybrid AC/DC Microgrid Module”, CSEE Journal of Power and Energy Systems, vol. 4, no. 1, March 2018.

Application of Spatial Informatics in Built Space

Priya Singhal[1], Vishal Rai[2]

[1,2]Department of Architecture and Planning,
Indira Gandhi Delhi Technical University for Women, Delhi.
[1]singhalpriya828@gmail.com, [2]vishalrai@igdtuw.ac.in

Abstract: The research paper describes the potential of spatial informatics that can help in the transformation of the built space, with the help of digital tools and the engagement of the community for better innovation. Because the information obtained to establish a link between people, place, and technology which when analyzed produces a result for the effective planning of the space. The analysis of data Offers a detailed overview of the choices and expectations of the community. Because of the technological advancement and the development of APIs (Application program interface), the extraction of data from the various open data sources is possible. This has involved the direct participation of the user in the development of the space with a rich perspective.

Keywords: Spatial informatics, Innovation, Technological, Application program interface, Data.

I. INTRODUCTION

The internet has evolved to become the main medium of communication that links several threads across the daily fabric. The ubiquity of internet services and applications has led many academics to challenge the dichotomy between real space and cyber space. The data is the most crucial and expensive thing in today's 21s century. Data provides the statics and facts for the reference which is important for the new development. Big data is a collection of tremendous data that has no defined subject but with the data analytics information required for the research can be extracted out of it.

Data analytics is the process of examining, modifying and carving data to obtain functional information, interpretations and support decision making. Here the data referred to is spatial data which can be used for the planning, improvement of services and the development (social, institutional, technological, etc.) in the space by obtaining the relevant statics and facts by targeting the audience.

The aim is to define the emerging sources which help in the analysis of big data (urban data) thus helps in the understanding of the formation of the cities and increasing the efficiency of services in space, providing the scope for the retrofitting, redevelopment, regeneration of the area. Further, solving the dilemma between the planning of the city or planning for the people. Due to the enthusiasm of people in 21 st century many political, cultural, technological, traditional, institutional and social changes are taking place, which has given birth to the new possibilities and serving us with the different options.[2]

The objective of the research is to identify the use of spatial informatics in increasing the efficiency of services in a built space. Further, Understanding the role of technology in increasing the efficiency of the space, and how the information obtained can be used to increase the efficiency of built space. The paper justifies the potential in the technology which when used effectively can be used in increasing the efficiency of the space to meet the requirements of the user.

II. THE RISE OF APPLICATION INFORMATION AND TECHNOLOGY IN SPATIAL APPLICATION

Spatial informatics and digital tools along with community participation is an approach towards engagement and innovation. The cities are transforming space to produce more efficiency in a single space which requires simulation and analysis of data set is being done to get the effective result. Historically the researches used to rely on the censuses for the quantitative urban research and specialized sensor system. The development of information and communication is massive in which citizens can directly participate in the development of the spaces by providing suggestions and actively participating in the surveys.[1]

Increasing awareness of people with education information and environmental data does not automatically cause sufficient inspiration and bring change in their habits towards a more sustainable lifestyle and environmental friendly. This research is aimed at developing a better understanding of how to go beyond training and inspiring change and action. Based on participatory culture, real-time information and technology, the study aims to provide research findings that suggest feasible new approaches to the development of a space-efficient application.

1. Research Methodology

Are the achievements made possible with the use of informatics and data analytics in producing the effective output that can be used for adding more value to the space?

COMPONENTS	TOOLS	OUTPUT
SPATIAL INFORMATICS	Internet study, research paper, books	History of Spatial informatics. What is spatial informatics? How the information is obtained? Analysis of information Applications of spatial informatics.
ONLINE USER BEHAVIOUR	Internet study, research paper, books	This will help in the study of the online behaviour with the mapping of overall data collection.
BIG-DATA	Internet study, research paper, books	Big data is one of the main aspect that has be considered while studying the Spatial informatics, the cloud with the tremendous information will help in understating at what level the data can be effectively used to give a new definition to the space.
CASE STUDY	Kelvin Grove Urban Village: A strategic planning case study By Linda Carroli	The Kelvin Grove Urban Village. Brief: how the blog helped in the redevelopment of the village meeting all the requirements of the villagers. Reason: the questionnaire of the blog actually generated data (urban informatics) which was being analysed and accordingly the result was obtained which helped in the development.
RESEARCH WORK	Android application Coding	An android application which uses the urban informatics can be proposed This will help in the understanding of the live functioning of the space and also the implication that can be done using the same.

III. USER GENERATED CONTENT

Over the past few decades, many transformative changes have occurred in how citizens participate and are involved in the co-creation of information. Citizens go from research studies and being passive subjects of a survey to active producers of data through the use of social media and sensors and various knowledge that is socially generated resulting in involvement in political, civic and social activities.

Typically, user-generated content happens when:

- When users voluntarily produce ideas information, document incidents, solve problems, conflicts, or activities of civic and social interest.
- If analysts analyze secondary user-submitted data that is published by social, media, inernet and other resources, retroactively.

  User generation information can be generated proactively by generating the concept, feedback, and problem-solving. Such practices produce immense unstructured and structured information that can be analyzed to gain insights attitudes and behaviours and so on. It has enabled ideas and voting on the different projects at the urban level that help and refine the project to meet the user group's demand. Social internet data, social media streams or blogs have generated a great deal of interest among researchers and research field, with the various dominant online networks that are socially active such as facebook, twitter, likedIn, and foursquare serving as a location-based social network in the current scenario. [4]

There are a wide range of online multi-media social sharing platforms like Flickr and Youtube, many users have created online content in the form of sites like Wikipedia, Yelp, and TripAdvisor. The government collects micro data on people as part of the daily payment, registration, and record-keeping procedures usually occurring during service delivery.

A well-defined application programming system is required to integrate, coordinate, capture, scan, query, and analyze data along with the software, hardware. In the format of the text, images, and reactions, etc., the formatted user data is sometimes collected thought the APIs. The wealth of data that has been proactively generated through various sensing systems is growing explosively. There are different types of data that can be obtained from the different sources generated by the user himself, such as revenue and tax agencies recording data of the citizens and revenues generated, licenses issued and taxes paid, and vehicle and transaction of real estate.[4]

Administrative data (micro) contains massive relevant information the evaluation of urban policy. The benefits of using administrative data include potentially less intrusive and relatively cheap yet comprehensive, as well as fewer issues with attrition but larger sample sizes, measurement error and non-response compared to traditional sources of survey data. The "open data" project is a specific innovation with the open data source. Such programs are largely driven by policies for open government. Open sources of information have facilitated many projects and have the potential to lead to innovation. Though 'open data' initiative faces problems at the same part of the world which includes reason like closed government culture in some of the localities.

IV. TECHNOLOGICAL CHALLENGES IN USING DATA

Technical challenges arise due to the need to capture, generate, manage, disseminate, process and discover the information called urban information. The complexities of managing large volumes of unstructured and structured data have been documented extensively in other ways. Construction of data cloud stores, infrastructures and multi-cloud architectures, as well as mechanisms for language resource, discovery, and execution environments are some of the major challenges of information management. Certain considerations include Software and hardware, and the need for Application Programming Interfaces (APIs) to collect, organize, integrate test, and request information. Scalability, fault tolerance, and performance are of equal importance as well as tools for scalable execution.[7]

In the market, various Big Data solutions such as MapReduce, Hadoop and others have emerged, some of which are open sources. One of the toughest challenges with using Big Data for Urban Informatics is not that the data which is obtained is inherently enormous, as in case of economics, high-energy physics, genomics or other information (although this might change with the incoming influx of data from connected vehicles and the IoT worlds.) Other areas have been extensively documented the complexities of sorting large amounts of unstructured data and structured data.

Development of data storage, database repositories, and multi-cloud systems, as well as frameworks for language, resource discovery and execution environments are some of the major challenges in information management. In general, for data linkage, if you go beyond standardized, rectangular database to stream data through an APIs leading to image, text, and other unstructured data formats, the fragmentation and complexity can pose significant problems.[7]

With many Big Data sources, whether user-generated image or administrative micro-data or GPS data, data privacy also becomes all-important and is often a major roadblock for research data acquisition, particularly for research requiring potentially personally identifiable data.

Another important determinant is access to high quality data is crucial asset discovering tools for urban researchers to understand and find data, a data governance and ontologies for knowledge representation, system involving harmonization of standards, keywords, and functional aspects. In many cases, authors are unlikely to have direct relevance to the above technology issues. Nonetheless, the technical aspects of Big Data, such as retrieving information, linking and curating, and the political economy of Big Data, such as governance , data access, trust management and privacy, may have IT requirements that may limit data accessibility for different studies.[7]

V. RESEARCH WORK

1. Defining the Problem

Small scale-spaces and buildings are manageable and it is easy to keep the record and updates of the services that are being upgraded with time. But, when it comes to large spaces like institutional, commercial, community, etc, the and maintenance of whole space is burdened on a single body it becomes a task and takes more time than the required.

Problem 1:

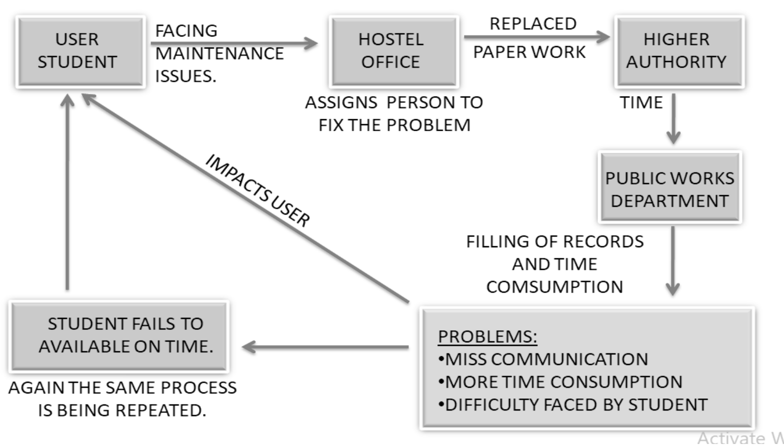


Fig. 1. The flow chart defining the problem that is being faced by the user (student) and causes lot of inconvenience and time taking process.

Problem 2:

Once the issue that is defined in problem 1 has been resolved there comes a question, who is keeping the record that which area has been upgraded (replaced or repair) and what was the origin of the issue?

There are n. Nos of appliances in a space that becomes a task to identify which appliance has been upgraded.

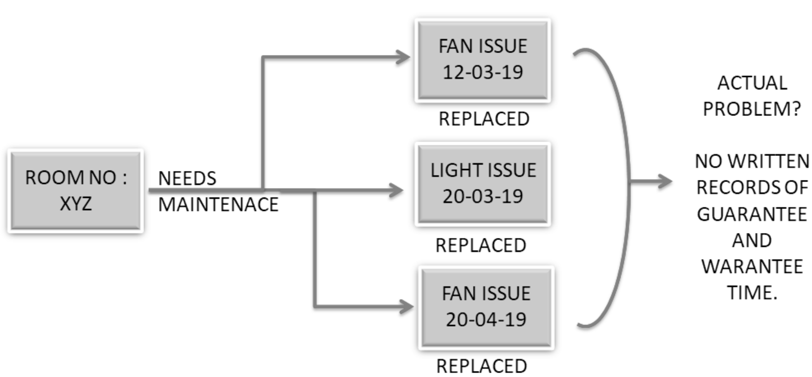


Fig. 2: Defining the problem 2

2. User Interface for the Prototype

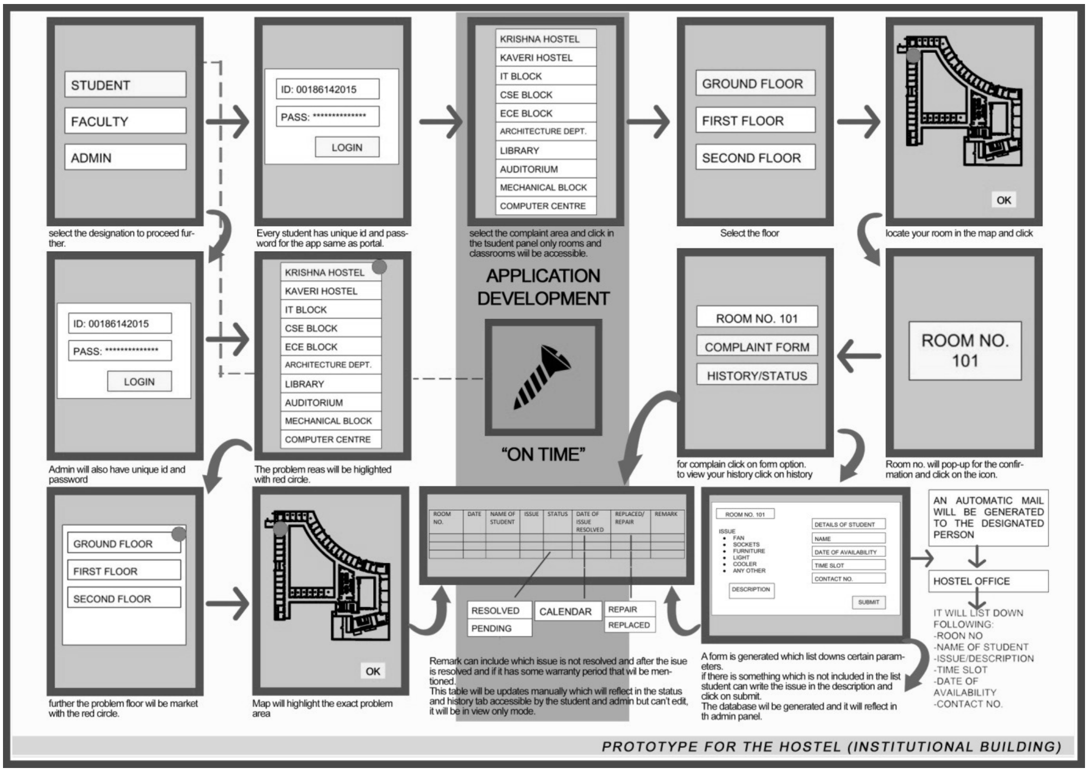


Fig. 3. Showing problem 1 and 2 can be resolved with the introduction of application.

The above flow chart explains how the problem 1 and 2 can be resolved with the introduction of one application and this will help in increasing the efficiency of services in a space. With the use of the application, the record is generating the spatial data. The data can be used for the analysis and as per the map, the origin of the problem can be identified in a short time.

3. Current Stage Prototype

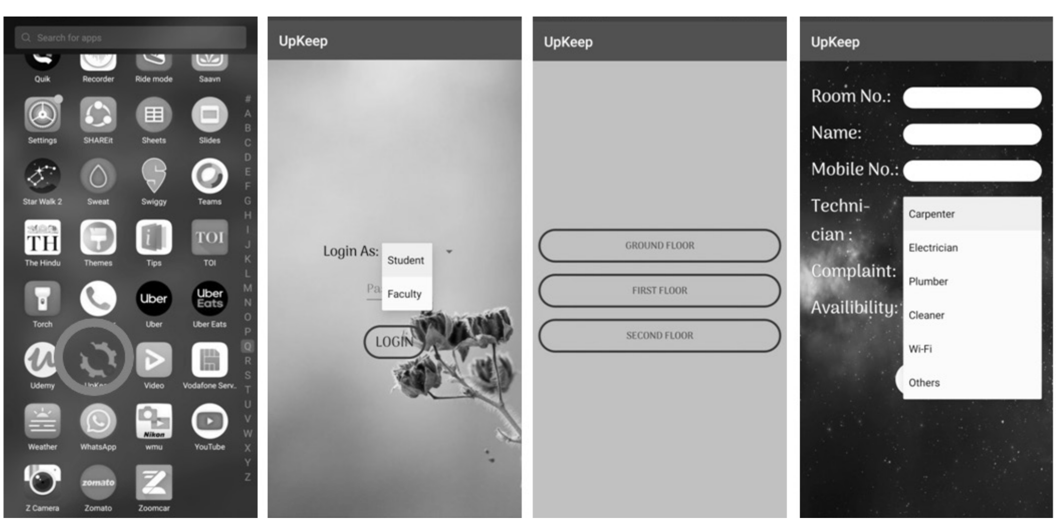


Fig. 4. Shows the steps of prototype when used by the student (User) for Krishna hostel

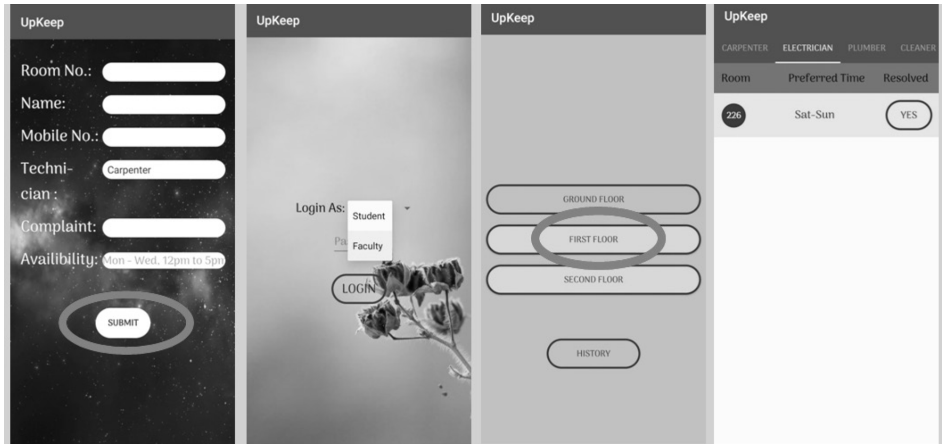


Fig. 5. Shows the steps of prototype when used by the Admin for Krishna hostel

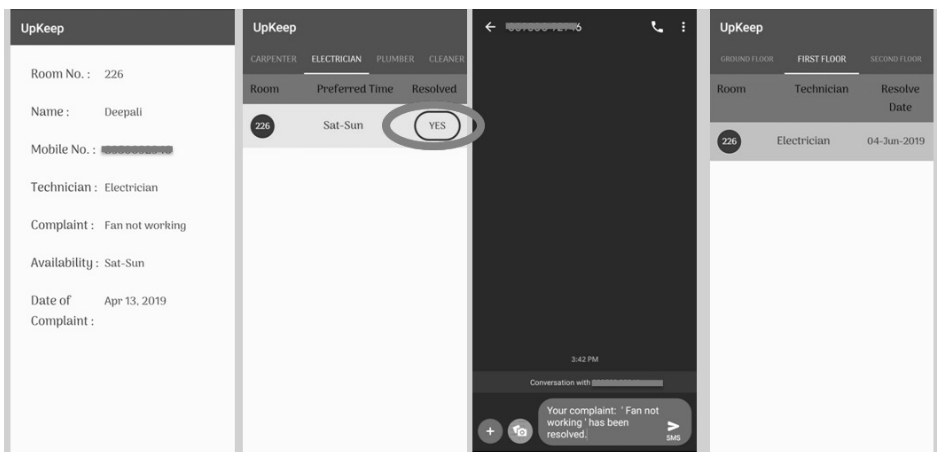


Fig. 6. Shows the steps of prototype when used by the student (User) for Krishna hostel

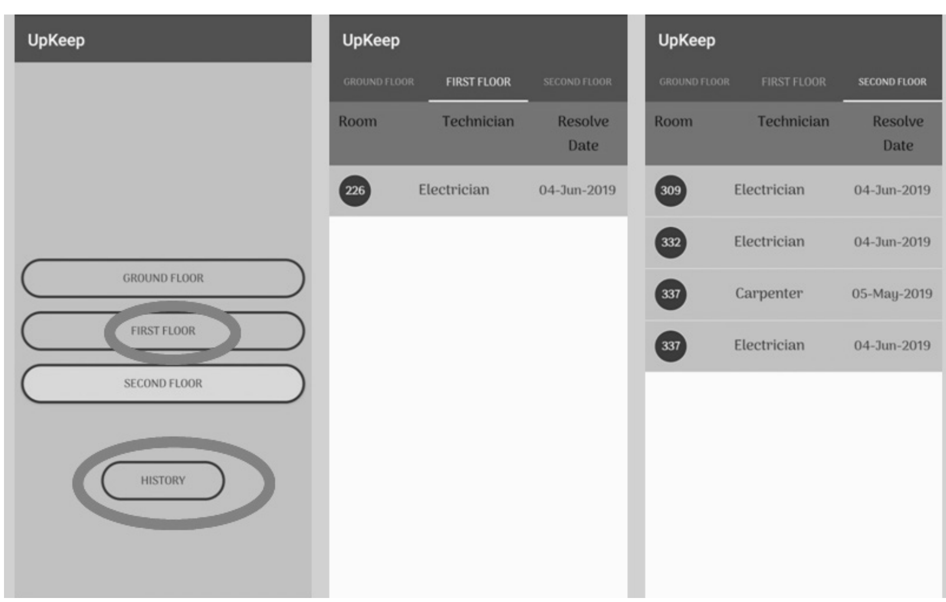


Fig. 7. Shows the steps of prototype when used by the student (User) for Krishna hostel.

4. Scale up

The prototype was prepared for the single building in a large institute campus. It can be further developed for the campus which includes all the blocks.

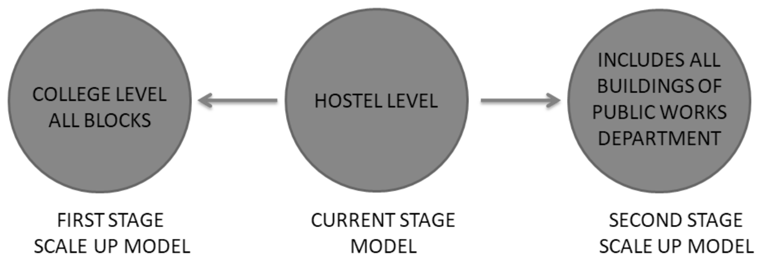


Fig. 8. Shows how the model can be scaled up at the institutional building and further at the pwd level

5. Unique Features

1. Generates record and keep track.
2. easily accessible and user-friendly.
3. reduces the time of maintenance.
4. connects the user directly to the main authority.
5. true updates.
6. zero chances of miscommunication.
7. maintaining the history and log about each complaint may help improve service delivery.
8. maintaining log may give performance reports about various products used in a project, and this may help reduce cost on various components like operation /maintenance of projects, especially in government projects it may benefit more and also help curtail usage of substandard products which is usually a complaint in government projects.

6. Use of Informatics and Data

The informatics has been used at every step. There are two data sets which get generated in the process, first is student registering the complaint, the data set is generated for the individual rooms. The second data set is the compilation of all the services and complaints of all the rooms. This data will help in identifying the origin of the problem.

7. Analysis: from Data to Insights

A series of information that is collected through the use of the application, each containing the information that is collected using the complaints from each room is automatically organized. The complaints through the application give the analysis that which appliance is causing the problem, and what is the origin of the problem if that same problem is being repeated in the various rooms.

Therefore, the theoretical model provides a framework in which experience of urban can be interpreted in terms of different levels and types of power and constraints in the selection of the material and their relationship.

Through this framework, the model can be obtained which when analyzed will give the conclusion of the appliances and there efficiency. The analysis offers a detailed overview of the choices and expectations of students about the maintenance and current issues.

VI. CONCLUSION

Technology allows a user to directly participate in the development of the space with a rich perspective on the types of activities, social processes, and experiences they had experienced in the space. Technology now can be used as a tool for efficient design and planning. The models that are generated provides framework in which urban experience can be interpreted in terms of differential levels and types of power and constraint in space development and their relationships. Along with this the whole process also favours the choices and expectations of a user about their space and its development. Therefore, building a user-centric design, by incorporating different experiences, social processes, and contributions which one has undergone in a single space. Further, online data and research collection tools can be used for gaining in-depth insights into urban environments and social contex

REFERENCES

[1] Foth, M., Choi, J. and Satchell, C. (2001). Urban Informatics. p.8.

[2] Batty, M. (2013). Big data, Smart City and City Planning. Dialogues of Human Gerography.

[3] amayo Caldwell, G., Foth, M. and Guaralda, M. (2013). an urban informatics approach to smart city learning in architecture and urban design education. Interaction Design and Architecture Journal.

[4] Tu, W., Li, Q., Zhang, Y. and Yue, Y. (2019). User- Generated content and Its Application in Urban Studies. p.8.

[5] Williams, A., Robles, E. and Dourish, P. (2014). Urbane-ing The City: Examining and Refining The Assumptions Behind Urban Informatics. p.17.

[6] M A Bettencourt, L. (2014). The Uses Of Big Data in Cities. p.8.

[7] Thakuriah, P. and L Zellner, M. (2017). Big Data and Urban Informatics: Innovation and Challenges to Urban Planning and Knowledge Discovery. p.30.

[8] Pettit, C., Stimson, R., Nino- Ruiz, M., Morandini, L., Widjaja, I., Delaney, P., Tomko, M., Sinnott, R., Randolph, B. and Kvan, T. (2014). Supporting Urban Informatics Through a Big Data Analytics Online Workbench. p.27.

[9] Linda, C. (2017). kelvin grove urban village: a strategic planning. The Journal Of Public Spcae, 2(4), p.140.

A Survey on IoT Security: Issues, Challenges & Opportunities

Kavita Sharma[1], **S.R.N. Reddy**[2]

[1,2]Computer Science and Engineering, IGDTUW, Delhi India
[1]kavita11235@gmail.com, [2]srnreddy@igdtuw.ac.in

Abstract: The IoT (Internet of things) play a significant role in routine life. IoT deployment connect the smart sensing nodes in the ubiquitous environment to collect useful information autonomously. The open environment (wireless medium), used by IoT devices for communication, has a dynamic path to send information from one location to other which is much more vulnerable and prone to network threats than wired medium. As a result, to maintain confidentiality, encryption, access control and node integrity, it is extremely required to implement efficient security and privacy protocols. This paper has a survey report on different IoT models, weaknesses, security requirement, issues and challenges including threats in smart home environment. This paper also discusses on some available technological solution to direct reader for future researches.

Keywords: IoT Security, IoT security issues, IoT protocols, IoT model

I. INTRODUCTION

Over the past few years, the Internet of Things (IoT) has attracted significant attention. Kevin Ashton first proposed the concept of IoT in 1999[2]. As a result of rapid development in mobile networks, Wireless Sensor Networks (WSN), Radio Frequency Identification (RFID), and cloud computing, communications among IoT devices has become more simpler than it was ever before. IoT devices has strong communication capabilities with neighbouring objects. The World of IoT includes heterogeneous type of devices having different size of memories, different processor types and processing capabilities. E.g.: smart phones, personal computers, PDAs, laptops, tablets, and hand-held embedded devices. The IoT devices have cost-effective sensors and wireless communication systems to communicate with each other and transfer meaningful information to the centralized system on internet as shown in figure 1.

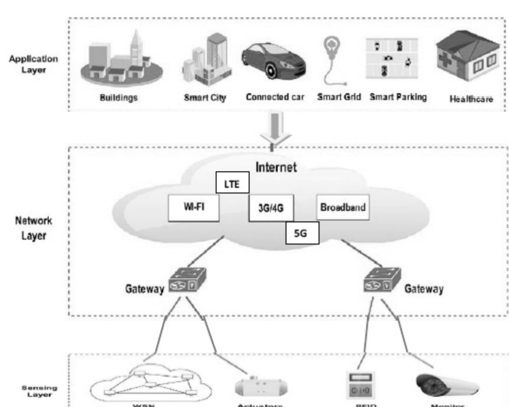


Fig. 1. Centralized system of IoT [12]

The IoT devices are resource constrained and small in size. These devices are deployed in pervasive environment to collect environmental information. These IoT environment can be controlled from anywhere, anytime with the help of any heterogeneous device. The information shared by these devices follow dynamic path on wireless medium. Most of the security tools, devices and applications discussed in literature handle the security and privacy attacks faced by traditional IT devices. Due to resource constrained feature of IoT devices [1]-[7],[9]-[13], the traditional IT security tools [8] are not suitable for IoT environment. Hence, IoT devices are easy to hack and compromise. There are a lot of security and privacy issues in the IoT networks such as confidentiality, authentication, data integrity, access control, secrecy. On every day, the IoT devices are targeted by attackers and intruders. An assessment [13] shows that 70% of IoT devices are vulnerable to different type of attacks. Therefore, it is necessary to have an efficient mechanism to protect interconnected IoT devices against hackers and intruders.

The main contribution of the paper is:

- To highlight on different IoT models suitable for different type of IoT applications
- A survey report is presented on different type of IoT security requirements, vulnerabilities/Issues/Threats including threats in smart home environment equipped with IoT devices.
- The possible security solutions for future research directions are also highlighted in this paper.

The paper is divided into five sections and subsections. Section one is introduction, section two highlights on different IoT models, section three reveals security issues and challenges and section four has possible IoT security solutions for future research directions. Section five concludes the overall paper.

II. INTERNET OF THINGS MODELS

1. IoT System Model

IoT system combines Physical and digital components in the real environment.

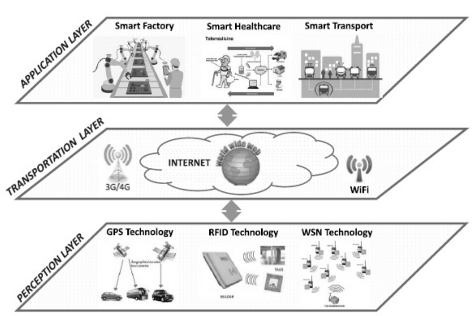


Fig. 2. IoT system Model [5]

The system model (figure 2) is the deployment process of IoT objects. It has basically three layers. The lower layer is perception layer where IoT objects communicate with the help of wireless technologies. E. g.: GPS technology, RFID technology, WSN etc. The second layer transportation layer is responsible to connect IoT object with Internet objects with the help of 3G/4G/5G/LTE or Wi-Fi technology. The uppermost layer is the application layer. It is the environment where IoT objects are deployed e.g.: smart factory, smart industry, smart transportation.

2. IoT Network Model

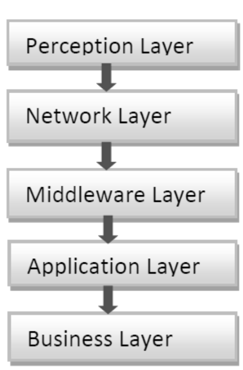


Fig. 3. IoT N/W Model

The network model (figure 3) of IoT reveals how information is communicated on the wireless medium. Initially there are three layers in IoT network architecture: perception layer, network layer and application layer exist. But as the IoT devices are resource constrained and has most internet attack prone due to wireless medium used for sharing information. Thus, it requires more layers (middleware layer and Business layers) to control and providing safety to Network Model. The network model is targeted to feed security tools in the system.

3. IoT CIA Model

The CIA (Confidentiality, Integrity, Availability) guide the information security policies in IoT environment. The Confidentiality (C) ensures that unauthorized users cannot access the system. The Integrity(I) guarantees that data will remain unaltered if accessed any location and any time. The availability(A) feature of CIA security model ensures that data will be available at any time and any place. The CIA model maintain the trust of IoT users and ensures that data cannot be used by hackers or malicious users.

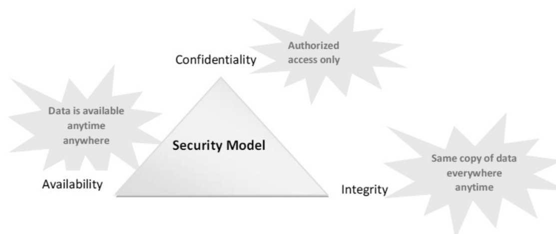


Fig. 4. CIA security Model

III. IOT SECURITY: ISSUES AND CHALLENGES

1. Requirement of IoT Security

he deployment of IoT has become the need of our routine life. IoT devices are deployed to change the way of solving problem in everyday life e.g. Health care system, manufacturing, smart homes, smart cities etc. The deployed IoT model has five-layer network model to share information among neighbouring IoT objects. The IoT objects are not resource rich and communicate in open environment, hence offer a golden chance to network attacker to enter in the system. There are different type of issues and challenges in the IoT network to breach the IoT network security. hence, encourage the user to use available security tools. The different layers of IoT network has different type of security issues [3]-[6], [11] listed in table 1

Table 1. Different Security Issues [3]-[6], [11]

Security Issues	Security Issues
Data Leakage	Heavy computation cost of TLS
DoS Attack	Privacy for Lack of User Control
Routing Attack	Heavy cost of computation and high handshake of DTLS
Data Transit Attack	Hardware Insecurity
Impersonation	Lack of lightweight Cryptographic Algorithm
Data Transit Attack on header	Lack of Lightweight Trust Management System
Threats to NDP Protocol	Lack of Lightweight Secure Routing Protocols
Lack of Lightweight Anti-malware Solutions	Middleware Security
Physical Wireless Insecurity	Privacy Violation on cloud
DDoS Attack Issue	Jamming Adversaries
Common Application Vulnerabilities	Insecure initialization and Configuration
Privacy Protection Issue	Insecure Neighbour Discovery
Sleep Deprivation Attack	Buffer Reservation Attack
Replay Attack Due to Fragmentation	Attack at the Time of Session Establishment & Resumption
Authentication and Secure Communication	Eavesdropping
End to End Security	Gathering
Insecure software and Firmware	Routing Diversion, Fabrication

The various types of security issues, given in the table, arises due to large user accessibility and some critical applications stored in the IoT storage device at application layer. There are heterogeneous IoT devices with different hardware and software configuration. The connectivity of heterogeneous devices is the golden river at network layer for the hackers. At the perception layer, the physical exposure and technological heterogeneity is the main reasons to vulnerability and has the entrance door for hackers.

2. Traditional IT Security and IoT Security

The traditional network devices e.g. desktops, laptops, servers, gateways are resource rich and generic objects. These objects change their nature according to need with the help of software in the storage device. These devices have a large capacity of storage. Thus, the security tools are installed externally in these devices. While IoT objects are resource constrained and has limited power, storage and size, thus unable to handle large and complex security tools as in IT network. These devices are embedded in the environment to do a specific task; hence these are unable to process traditional IT security tools as listed in table 2 as they are suitable for resource rich devices.

Table 2. IT security and IoT security

IT Security	IoT Security
Externally installed security tools	Embedded in the IoT object.
Support complex algorithm	Support lightweight algorithms only
User has full control on security tools	Run automatically as IoT collect private information automatically.
Less heterogeneity than IoT	More heterogeneity than IT
More security guards available in the literature	Less security guards available in literature.
Devices are located in closed environment	Devices are located in open environment.

3. Classification of Security Issues (2G,3G,4G and 5G issues)

The IoT network is used in different type of applications as shown in figure 5:

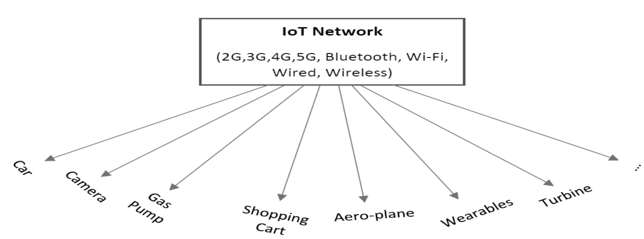


Fig. 5. IoT Network Applications

As the IoT applications use wireless communication and information is send/received wirelessly in open medium following dynamic path. The basic and initial architecture of IoT network includes three layers: physical layer, network layer, application layer. The different type of issues in these layers raised in deployment of IoT network are listed in figure 6.

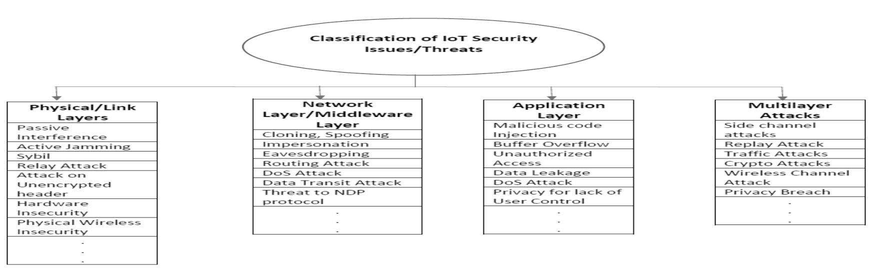


Fig. 6. IoT Security Issues/Threats

4. Threats in IoT based Smart Home System

Recently smart homes [14] are going to be popular as they are rich with IoT objects. The IoT network in the smart home is not a dedicated network architecture and the householder are also not skilled with deep technological factors, thus a smart home is prone with different type of challenges and threats as shown in figure 7.

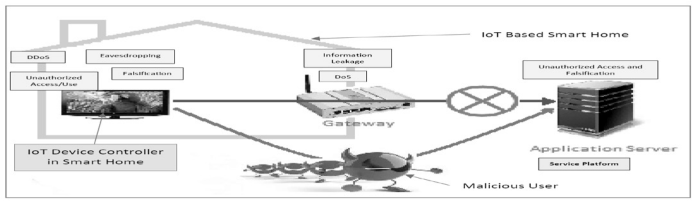


Fig. 7. Challenges and Threats in Smart Home

IV. IOT SECURITY: POSSIBLE SOLUTIONS AND FUTURE RESEARCH DIRECTIONS

IoT devices does not have built in security tools. On the other hand, some security tools deployed in internet environment are unable to detect connected IoT devices. Thus, IoT security can have only integrated solution which has visibility, segmentation and protection throughout internet and IoT environment. Different type of threats/issues/vulnerabilities extracted from the survey of [1]-[3][5][7][8][10][11][14] are given in table 3 with their possible solutions which direct the reader for future research.

Table 3. Issues/Threats/Vulnerabilities with their possible solutions [1]-[3][5][7][8][10][11][14]

Issue/Threat/ Vulnerability	Level	Layer	Protocol	Possible Solution
Data Transit Attack	High	Physical Layer	IEEE 802.15.14, BLE, Wi-Fi, LTE	AES-CCM algorithm, WEP, WPA, WPA2, EFA and EIA algorithm
Data Transit Attack	High	Network layer	IPv4/IPv6, 6LowWPAN, RPL	IPsec Protocol, Compressed DTLS, 802.15.4 security tools, AES/CCM algorithm
Threat to NPP Protocol	Medium	Network Layer	IPv4/IPv6	SEND protocol of IPv6
Data Transit Attack	High	Application Layer	MQTT, CoAP	PSK, Certificates, Secure MQTT, DTLS Protocol, PSK, RPK, Lithe solution
Data Leakage	High	Application Layer	MQTT, REST/ HTTP	Intrusion Prevention Tools
DoS Attack	High	Application Layer	MQTT	Firewalls, Proxy Server
Malicious Code Injection	High	Application Layer	MQTT, CoAP	Privilege setting, Input validation, Parameterization
Routing Attack	Medium	Network Layer/ Middleware Layer	Dynamic Source Routing, AODV	Security Mechanism for neighbour listening.
Jamming adversaries	Low	Physical Layer	Ethernet, Wi-Fi (802.11), PPP	Computing packet delivery ratio, Error encoding codes, Frequency and location change

Table 3 (Contd.)...

...Table 3 (Contd.)

Sybil & Spoofing Attack	Low	Physical Layer	Ethernet, Wi-Fi (802.11), PPP, LTE	Signal strength measurement and channel estimation
Insecure Initialization and Configuration	Low	Physical layer	Ethernet, Wi-Fi (802.11)	Fixing data transmission rate between nodes,
Insecure Physical Interface	Low	Physical Layer	Ethernet, Wi-Fi (802.11), PPP	Restricted USB access and Testing/ Debugging tools
Sleep Deprivation Attack	Low	Network layer/ Middleware layer	IPv4/IPv6, LoWPAN, RPL	Multilayer Intrusion detection tools
Replay Attack	Medium	Network Layer	IPv4/IPv6	Fragment verification and timestamp
Insecure neighbour Discovery	Medium	Network layer	IPv4/IPv6,	Cryptographic based authentication system
Buffer Reservation Attack	Medium	Network Layer	IPv4/IPv6,	Buffer splitting approach
RPL Routing Attack	Medium	Network Layer	IPv6,	Hashing and Signature based authentication, Node behaviour monitoring
End to End security	Medium	Network Layer	IPv4/IPv6,	AES/SHA algorithm, Compressed IPsec, Identification and Authorization
Session Establishment and Resumption	Medium	Network Layer/ Middleware Layer	IPv4/IPv6, LoWPAN, RPL	Secret key authentication and symmetric key encryption
CoAP Security with Internet	High Level	Application Layer, Network Layer	MQTT, CoAP, IPv4/IPv6	TLS/DTLS, Mirror proxy, Resource Directory, Message Filtration
Insecure Interfacing	High Level	Application Layer	MQTT, CoAP,	Firewalls, Strong passwords, Testing tools at interfaces.
Insecure S/W, H/W	High Level	Application Layer	MQTT, CoAP,	Encryption techniques, Regular update, file Signatures
Middleware Security	High Level	Application Layer, Network Layer	MQTT, CoAP, IPv4/IPv6	Authentication Techniques, M2M security encryption mechanism

V. CONCLUSION

Today IoT is a major research area. IoT physical objects are interconnected with the help of network technologies using various IoT models. The IoT objects share sensitive information in the wireless environment e.g.: health care data, smart home data, alarm system, manufacturing data etc. The rapid growth of IoT objects which are heterogeneous in nature has various type of threats. Thus, this paper explores the different types of IoT deployment models. It reveals on various threats/issues/vulnerabilities including threats in smart home environment and possible solutions with respect to threats given in literature. The solutions highlighted in the paper has some vulnerabilities which are not explored in this paper. So, the reader can explore these vulnerabilities to extend their research skills and future research in the area of IoT security.

REFERENCES

[1] Conch, M., Dehghantanha, A., Franke, K., & Watson, S. (2018). Internet of Things security and forensics: Challenges and opportunities.

[2] Razzaq, M. A., Gill, S. H., Qureshi, M. A., & Ullah, S. (2017). Security issues in the Internet of Things (IoT): A comprehensive study. International Journal of Advanced Computer Science and Applications, 8(6), 383.

[3] Conti, M., Dehghantanha, A., Franke, K., & Watson, S. (2018). Internet of Things security and forensics: Challenges and opportunities.

[4] Khan, M. A., & Salah, K. (2018). IoT security: Review, blockchain solutions, and open challenges. Future Generation Computer Systems, 82, 395-411.

[5] Frustaci, M., Pace, P., Aloi, G., & Fortino, G. (2017). Evaluating critical security issues of the IoT world: Present and future challenges. IEEE Internet of Things Journal, 5(4), 2483-2495.

[6] Oh, S. R., & Kim, Y. G. (2017, February). Security requirements analysis for the IoT. In 2017 International Conference on Platform Technology and Service (PlatCon) (pp. 1-6). IEEE.

[7] Ahmad, Ijaz, et al. “Overview of 5G security challenges and solutions.” IEEE Communications Standards Magazine 2.1 (2018): 36-43.

[8] Frustaci, Mario, et al. “Evaluating critical security issues of the IoT world: Present and future challenges.” IEEE Internet of Things Journal 5.4 (2017): 2483-2495.

[9] Vashi, S., Ram, J., Modi, J., Verma, S., & Prakash, C. (2017, February). Internet of Things (IoT): A vision, architectural elements, and security issues. In 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 492-496). IEEE.

[10] Yang, Y., Wu, L., Yin, G., Li, L., & Zhao, H. (2017). A survey on security and privacy issues in Internet-of-Things. IEEE Internet of Things Journal, 4(5), 1250-1258.

[11] Yaqoob, I., Ahmed, E., ur Rehman, M. H., Ahmed, A. I. A., Al-garadi, M. A., Imran, M., & Guizani, M. (2017). The rise of ransomware and emerging security challenges in the Internet of Things. Computer Networks, 129, 444-458.

[12] https://www.researchgate.net/figure/Structure-of-centralized-IoT-system_fig5_330376010

[13] Andrea, I., Chrysostomou, C., & Hadjichristofi, G. (2015, July). Internet of Things: Security vulnerabilities and challenges. In 2015 IEEE Symposium on Computers and Communication (ISCC) (pp. 180-187). IEEE.

[14] Pirbhulal, S., Zhang, H., E Alahi, M., Ghayvat, H., Mukhopadhyay, S., Zhang, Y. T., & Wu, W. (2017). A novel secure IoT-based smart home automation system using a wireless sensor network. Sensors, 17(1), 69.

Block Chain or AI? Future is Block Chain and AI: Review

Prerna Goel[1], Mohona Ghosh[2]

[1,2]Department of Information Technology,
Indira Gandhi Delhi Technical University for Women, New Delhi, India
[1]prerna.goel181@gmail.com, [2]mohonaghosh@igdtuw.ac.in

Abstract: Block chain and Artificial Intelligence (AI) are two major breakthroughs in the world of technology. Any transaction or other operation done on block chain platform is visible to every other node on the network, ensuring transparency. Along with this, it provides with many other features like immutability, trust and decentralization. On the other hand, AI provides intelligence to act and react like humans through learning from its surroundings. When combined, they both overcome limitations of each other. In this paper, we review block chain, AI and their features. We show how block chain completes AI and how AI completes Block chain and what they can do when they are combined together.

Keywords: Artificial Intelligence, Block chain, Smart contract

I. INTRODUCTION

Blockchain and Artificial intelligence (AI) are two highly pioneering technologies in today's world. While blockchain is decentralized shared ledger mostly used for cryptocurrencies till now, where every transaction is signed cryptographically and verified by other nodes in the network before being added to the ledger, AI performs tasks that requires intelligence, learning from the data provided to the algorithms. Blockchain has its wings spread over other sectors including logistics, advertising, healthcare, business development etc. [1]. AI also has its application ranging from self-driving cars to the fake news, voice and photos creation [2]. Data for learning of AI models is obtained from some centralized sources, where data can be possibly hacked and tampered [3] leading to deviation from expected result, thus deriving wrong conclusions. Therefore, blockchain can be used for creation of AI models over decentralized data [4].

Till date, literature reviews done include blockchain and AI, and their applications in isolation [5]-[8]. Some of the studies have shown blockchain for AI [9],[10]. Only few have integrated both AI and blockchain [2],[11]. Primary objectives of this paper are summarized as follows:

1. Overview of blockchain and AI and their key features
2. How Blockchain can transform AI, overcoming limitations of AI applications
3. AI for blockchain, suggesting how AI can overcome blockchain limitations
4. Discussion on some recently implemented applications and ongoing work combining AI and blockchain.
5. Limitations and challenges with blockchain and AI

Rest of the paper is organized as follows: Section II includes background explaining blockchain and AI, their key features and some of their applications. Section III presents literature review discussing AI for blockchain, blockchain for AI and blockchain and AI, along with class diagram. Section IV details limitations and challenges with blockchain and AI. Section V concludes the paper.

II. BACKGROUND

In this section, overview of blockchain and AI, and their key features are given.

1. Block Chain

Blockchain is a public, distributed and immutable ledger shared among all the users in the distributed network [12]. It is a chain of blocks, where each block contains the transaction details along with hash value of previous block and timestamp [9]. Depending on the consensus protocol, users called miners, solve a computationally hard cryptographic puzzle. One who solves the problem, verifies the transaction and mines the transaction block to the ledger and earns some sort of reward in return.

Etherum is an extension of bitcoin blockchain that contains smart contract [13]. Each smart contract contains a piece of code that is self-executing. Only what is specified in smart contract is executed, no third part can alter the code. So, in order to verify the transaction, miner just needs to run the code in smart contract, no need to verify the previous transactions in the chain. Some of the features of blockchain are:

Trust- new transaction can only be added after it is verified by miner, which requires that more than 50 percent of the nodes in network approve the transaction.

Immutability and transparency [9]- all changes done to the public blockchains are visible to everyone in the network. Also, information once appended to the chain cannot be altered by anyone.

Disintermediation [9]-the data stored in blockchain is decentralized. No single person has control over the data.

As, storing data on blockchain is very expensive, so decentralized storage is used for data storage and hash of the data in storage is linked to the chain of blocks. Example of distributed storage are Swarm [14], Interplanetary File System (IPFS) [15].

Many applications are being developed using blockchain like Corda [16], ThoughtMachine, OpenBazaar, Tzero that utilizes decentralized storage of blockchain to secure their data and perform operations on it with ease.

2. Artificial Intelligence

Artificial intelligence is the learning that enable machines to act and react just like humans. It involves learning from huge amount of data, analysing and pattern discovery [9]. Core of AI is machine learning, where learning can be under supervision or without supervision [18]. When done under supervision, it involves regression and classification of data while without supervision, it relies on ability to identify patterns in given data. AI systems can perform planning, reasoning, learning along with problem solving using creativity and social intelligence, just like humans [2]. Most of the AI systems present today use data from centralized database for learning. Upcoming researches are working on AI systems that can use restricted number of problems, apply their intelligent decision-making process to find solutions to the real-life problems [10]. Some of the applications using AI are Numerai [19], digital twins [20].

III. LITERATURE REVIEW

This section discusses how blockchain and AI has transformed each other, overcoming their limitations. Figure 1 is the class diagram for blockchain and AI. It shows features and some of the applications built on blockchain and AI. Then it explains blockchain for AI, how blockchain features can be used for AI applications and AI for blockchain, how AI improves blockchain with its intelligence. It then shows combination of blockchain for AI and AI for blocking depicting their strength and what applications can be developed or are being developed using their combination.

1. Blockchain for AI

Blockchain provides backbone to AI algorithms [1]. It helps in clarifying that decision made by AI is transparent, trustworthy and explainable. As AI requires access to huge amount of data for decision making, it is not an easy task to gather sufficient amount of private data that belongs to some person or organization. Also, privacy of data is a growing concern for companies handling and providing these data to researchers. Blockchain solves this issue by having control of user's data in their hand. User can control its data. It depends on him that with whom he wants to share his data or sell to via smart contracts [2]. Policies regarding transparency of data use like read or update of data can be implemented in smart contract [21]. Blockchain ensures de-centralized data used by AI systems as centralized data can be subject to hacking, leading to tampering of data [10].

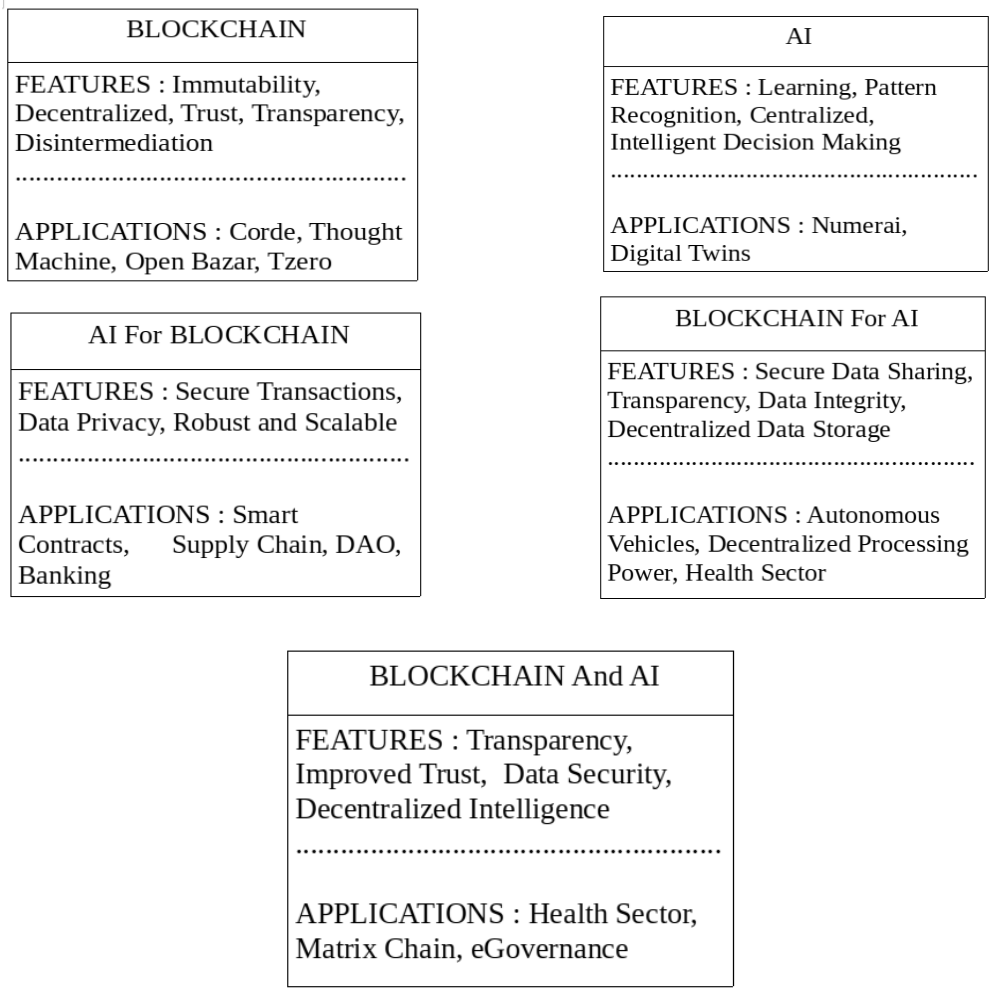


Fig. 1. Class Diagram for Blockchain and AI

AI models used today for decision-making are data-driven model. That is, they require training data for learning and result obtained depend on the characteristics of data. So, one cannot infer anything about model from the training data used, making them black-box [21]. Blockchain can be used to get transparency about the process. User can easily track data flow through these models. They can observe every step involved in learning process and decision making as every action taken on blockchain is visible to all other peer nodes. Thus, gaining trust and transparency for users that use AI systems [1].

Some of the applications that shows how blockchain has transformed AI are-

Autonomous vehicles (AV) when interconnected, use AI algorithms to learn about the traffic. They can communicate with each other, informing each other regarding traffic jam, accidents and traffic jams can be avoided. What if, data they use for traffic analysis is hacked. In worst case, it can lead to accident, traffic jam or even stealing of vehicle [9]. So, to prevent this, data can be stored on blockchain, from where it can be accessed by AV, exploiting immutability of blockchain data.

AI requires high computing power to train and deploy algorithms for machine learning. It uses millions of GPUs to prepare their algorithms. Such high power can be obtained from blockchain-based cloud computing. Blockchain can provide distributed computing power through decentralized market [2].

AI is also used to classify and analyse raw medical data in healthcare sector, where it faces data integrity issues during its collection and storage [22]. Thus, data can be stored securely on blockchain.

2. AI For Blockchain

All the blockchain data is available publicly, AI can be used to provide data privacy and confidentiality. With the help of smart contracts written using intelligent AI algorithms, several mistakes made in the smart contract, while written manually, can be avoided [2]. This will prevent transactions from being exploited by hackers, like what happened in the theft of $50 million where decentralized autonomous organization (DAO) was victim. Hackers exploited the mistakes made in smart contract that allowed them to repeatedly withdraw money [23]. Blockchain can be made robust and scalable using AI. In case of hike in the transactions, AI can increment the rate at which blocks are created. When some of the blocks in chain suffer from damage, rest of the components can be isolated from the attacked one using AI [2].

AI can integrate with blockchain to make applications more reliable. DAO use smart contracts for its operations. It has no employees or office building. AI algorithms will need to be applied when there will be machine to machine transactions for payments [9].

We can have smart contracts that rely on AI for learning and determining. As smart contract when made include some environmental assumptions that may not be valid at later point of time, AI based smart contracts will have those assumptions modified with time depending on environment change [9].

Supply chain operations that were previously dominated by paper-based system, are moving towards blockchain leading to cost reduction, elimination of intermediaries and more trust between sellers and buyers. Currently, AI supports supply chain only in scheduling and planning [9]. But it can be extended for automation of transactions in future with the help of smart contracts.

While moving towards banking sector, there are some blockchain banking applications like ThoughtMachine, Corda [16] that provide reliable banking system which is capable of managing finance, accounts and customers. These banking systems utilize AI for scanning financial reports and press releases for keywords that can forecast stock rise or fall [24].

3. Blockchain and AI

Blockchain and AI together builds the future. Combining both of these techniques will overcome their individual limitations resulting in powerful applications. Together they can build applications that provide more data security, improved trust among users, transparent decision making, decentralized intelligence [10] removing the need of intermediators during various operations. More data and process transparency will make it easy for users to accept the decision made. Decentralized data storage will prevent data from being tampered. AI will enable decisions to be more realistic by self -learning based on changing conditions.

In health sector, AI and blockchain can lead to improved medical care quality along with reduced cost. AI can be used for accurate diagnosis and thus effective recommendations by utilizing patient's data while blockchain can assure integrity of data. Both of them are being integrated in an attempt to build blockchain 3.0, which is called MATRIX chain [17]. AI and blockchain can even be introduced for government operations, in which AI can be used for establishing monetary and tax evasion and blockchain will help in tackling issues like bureaucracy, leading to improved efficiency and thus saving billions [9]. Some other applications include Fintech [9] and DeepBrain chain [25].

IV. LIMITATIONS AND CHALLENGES

Although we have seen in previous section that how AI and Blockchain overcome limitations of each other, there are some more concerns and challenges that need solution:

- Blocks in the blockchain require the hash of the previous block to be included in the newly created transaction block. These hashes can be victim of preimage attacks, making blocks less secure. Authors in

[1] have proposed some partial homomorphic encryption models, that can be used as substitute of hashes to make blocks preimage attack resistant. More secure and stronger approaches need to be figured out and implemented to have blockchain more secured.

- Adapting and integrating existing systems with blockchain may require considerable amount of changes and replacements in existing systems [9].
- To get a transaction verified, it needs to be approved by all the other nodes in the network, requiring high electricity consumption to keep all the computers running. Many alternate solutions are being developed to deal with it. One is Merkle tree that allow verification without having each of the node in the network to verify it [9].
- Data in blockchain can be compromised if majority of the nodes in the blockchain are malicious. This can lead to compromising of security. In case of centralized malicious control centre, all that requires is identifying that location and shut down that harmful program at once. But in case of decentralized control, all the malicious programs need to be killed one by one [2].

V. CONCLUSION

In this paper, we reviewed about current state of AI and Blockchain. We presented overview of blockchain and AI, their key features and issues with them. We have also shown their features and applications via class diagram that how useful they can be for each other. In literature review we have shown blockchain for AI, AI for blockchain and their combination. Blockchain for AI shows the applications that currently utilize AI algorithms for problem solving but face some issues like data integrity and trust. These issues can be resolved using blockchain strengths. Similarly, AI for blockchain presents applications build on blockchain platform that need AI to function efficiently. Blockchain and AI describes application that are likely to be developed in future or are being under progress using combination of AI and Blockchain. Limitations and challenges present some of the key issues within these technologies itself and also challenges that will be faced with their combination.

REFERENCES

[1] Yaji, Sharath & Bangera, Kajal & Neelima, B.. (2018). Privacy Preserving in Blockchain based on Partial Homomorphic Encryption System for AI Applications. 10.1109/HiPCW.2018.00021.

[2] Dinh, Thang & Thai, My. (2018). AI and Blockchain: A Disruptive Integration. Computer. 51. 48-53. 10.1109/MC.2018.3620971.

[3] Dinh, Thang & Thai, My. (2018). AI and Blockchain: A Disruptive Integration. Computer. 51. 48-53. 10.1109/MC.2018.3620971.

[4] Dillenberger, Donna & Novotny, Petr & Zhang, Qi & Jayachandran, Praveen & Gupta, Himanshu & Mehta, Sameep & Hans, Sandeep & Chakraborty, Supriyo & Walli, Matthew & Thomas, John & Vaculín, Roman & Sarpatwar, Kanthi & Verma, Dinesh. (2019). Blockchain Analytics and Artificial Intelligence. IBM Journal of Research and Development. PP. 1-1. 10.1147/JRD.2019.2900638.

[5] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

[6] Fioretto, F., Pontelli, E., & Yeoh, W. (2018). Distributed Constraint Optimization Problems and Applications: A Survey. *Journal of Artificial Intelligence Research*, *61*, 623–698.

[7] Fernández-Caramés, Tiago & Fraga-Lamas, Paula. (2018). A Review on the Use of Blockchain for the Internet of Things. IEEE Access. 6. 32979-33001. 10.1109/ACCESS.2018.2842685.

[8] Zheng, Zibin & Xie, Shaoan & Dai, Hong-Ning & Chen, Xiangping & Wang, Huaimin. (2018). Blockchain challenges and opportunities: A survey. International Journal of Web and Grid Services. 14. 352. 10.1504/IJWGS.2018.095647.

[9] Makridakis, Spyros & Polemitis, Antonis & Giaglis, George & Louca, Soulla. (2018). Blockchain: The Next Breakthrough in the Rapid Progress of AI. 10.5772/intechopen.75668.

[10] Salah, K., Rehman, M. H. U., Nizamuddin, N., & Al-Fuqaha, A. (2019). Blockchain for AI: Review and Open Research Challenges. *IEEE Access*, *7*, 10127–10149. https://doi.org/10.1109/ACCESS.2018.2890507

[11] Lopes, Vasco & Alexandre, Luís. (2018). An Overview of Blockchain Integration with Robotics and Artificial Intelligence.

[12] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system.Tech. Rep., 2008. Retrieved October 10, 2019, from https://archive.is/rMBtV

[13] Bhatia, Gurpriya & Dubey, Alpana & Kumaraguru, Ponnurangam. (2018). WorkerRep: Building Trust on Crowdsourcing Platform Using Blockchain.

[14] Hartman, John & Murdock, Ian & Spalink, Tammo. (1999). The Swarm scalable storage system. Proceedings - International Conference on Distributed Computing Systems. 74 - 81. 10.1109/ICDCS.1999.776508.

[15] Benet, J. (2014). IPFS - Content Addressed, Versioned, P2P File System. *ArXiv, abs/1407.3561*.

[16] M. Hearn. Corda–a distributed ledger. Corda Technical White Paper,2016

[17] While Blockchain Platforms Struggle With Consensus, MATRIX AI Perfects It. Retrieved October 15, 2019, from https://blog.usejournal.com/while-blockchain-platforms-struggle-with-consensus-matrix-ai-perfects-it

[18] What is Artificial Intelligence (AI)? - Definition from Techopedia. (n.d.). Retrieved October 10, 2019, from https://www.techopedia.com/definition/190/artificial-intelligence-ai

[19] A New Cryptocurrency for Coordinating Artificial Intelligence on Numerai. Retrieved October 10, 2019, from https://medium.com/numerai/a-new-cryptocurrency-for-coordinatingartificial-intelligence-on-numerai

[20] Schluse, Michael & Priggemeyer, Marc & Atorf, Linus & Romann, Jurgen. (2018). Experimentable Digital Twins—Streamlining Simulation-Based Systems Engineering for Industry 4.0. IEEE Transactions on Industrial Informatics. PP. 1-1. 10.1109/TII.2018.2804917.

[21] Bertino, Elisa & Kundu, Ahish & Sura, Zehra. (2019). Data Transparency with Blockchain and AI Ethics. Journal of Data and Information Quality. 11. 1-8. 10.1145/3312750.

[22] Campbell, D. (2018, November 21). Combining AI and Blockchain to Push Frontiers in Healthcare. Retrieved October 20, 2019, from https://www.macadamian.com/learn/combining-ai-and-blockchain-in-healthcare/

[23] De, N. (2017, December 29). Hacks, Scams and Attacks: Blockchain's 2017 Disasters. Retrieved October 17, 2019, from https://www.coindesk.com/hacks-scams-attacks-blockchains-biggest-2017-disasters

[24] Satariano, A., & Kumar, N. (2017, September 27). Bloomberg - Are you a robot? Retrieved October 15, 2019, from https://www.bloomberg.com/news/features/2017-09-27/the-massive-hedge-fund-betting-on-ai

[25] Bharadwaj, R. (2019, August 13). AI in Blockchain - Current Applications and Trends. Retrieved September 16, 2019, from https://emerj.com/ai-sector-overviews/ai-in-blockchain/

Review of Chatbot Design and Trends

Pooja Gambhir

Department of Information Technology,
Indira Gandhi Delhi Technical University for Women, Delhi, India
poojagambhir01@gmail.com

Abstract: Chatbot is a conversational device designed to counterfeit a smart communication with Humans through speech and text. Chatbot emerged like a machine which interacts with a human by their dialect, spoken/ textual conversation and response similarly but now there has been a tremendous increase in development of Assistants and Chatbots with advanced technologies based on Speech. This paper presents a survey of framework for researchers in identifying the areas of developing methodology of a Chatbot. In this paper, we will discuss about how Chatbots are designed, the machine learning and deep learning techniques and algorithms used, compare the technologies and how performance of Chatbot is evaluated. This survey concludes with curiosity to know why Chatbots are becoming more friendly to humans than just a communication machine and what are the future aspects of Human from the Human like Chatbots.

Keywords: Chatbots; machine learning; deep learning; Performance; Algorithms

I. INTRODUCTION

Chatbot is an application used to mimic human conversation. They are also referred as Virtual assistants which have taken over the world through its cutting-edge advancement from being just a machine to becoming a conversational friend. In recent years several conversational chatbots are being designed to interact with humans (Table 1). Chatbots have grown as a space for research and development in recent years due to both realization of their commercial potential and to advancement in language processing that can facilitate more natural conversations. They are made capable of responding and sometimes execute the tasks also. Users expect them to give an interactive experience. Customers feel that chatbots can help them to resolve their day to day issues. Chatbots have also gained a huge popularity mainly in business sector as chatbots have the ability to bring down efforts of human and automate customer service. For instance Roshan Khan proposed a generic architecture for designing and implementation of personalized Chatbot for different problem[9]. In the coming years Chatbot is likely to become the digital faces of all brands.

No doubt various technologies have emerged and are flourishing to make Chatbots as smart as behaving and feel like a Human. (example: Sophia). There have been many recent development and experimentation in conversational agent system. Apart from traditional chatbot development techniques that use rule-based techniques, or simple machine learning algorithms, many chatbots now a days are using advanced Natural Language Processing (NLP) techniques, deep learning techniques like Deep Reinforcement Learning (DRL) and Deep Neural Network (DNN) and computational intelligence. There has been a splendid enhancement in variety of algorithms and techniques like Neural Networks, Pattern matching techniques, voice recognition methods, and much more which if embedded in Chatbots make them equally like a Human.

This review paper described advancement of Chatbot from a basic model to an advanced intelligent model. Going back to a brief history, Term “Chatbot” was conceptualized by a problem coined by Alan Turing as Turing Test (1950) [22] in which an “interrogator” asked a question on identifying the human. If the human and machine are indistinguishable, we say the machine can think. The first chatbot was created in 1966 at MIT called **ELIZA[10]**. ELIZA identifies keywords and pattern matching concept against a set of pre-programmed rules to generate quality responses. Kenneth Colby created **PARRY**, in 1972, first chatbot which passed the Turing Test. Aditya Deshpande el at [4] cited in his survey paper that PARRY tried to create a person with paranoid schizophrenia. It personified a conversational strategy, and it was very much advanced and complex than ELIZA. The process was accomplished by an experienced psychiatrists’ group who analysed an arrangement of real computers and patients controlling PARRY with the help of teleprinters. They were capable to attain the accurate identification, 48 percent of the time. Through the 1980's and 1990's the technology was deployed in automated telephone systems that used decision trees through MSN and AOL. **A.L.I.C.E** designed by Richard Wallace**,** in 1995 which is a complex bot based on pattern recognition that matches the inputs adjacent to template pairs(output) that must be kept in knowledge base(Statement, 1966). Artificial Intelligence Mark-Up Language (AIML) is used to write the documents. It was also awarded as the most intelligent Chatbot. **SIRI** is an application of Apple's IOS designed as an intelligent personal assistant and knowledge navigator. Harshita Phatnani el at [17] reviewed in their paper that SIRI during its release has opened doors to huge opportunities to catch its wave and exploit this natural language-based voice-interface technology in new and profitable ways. And Today Chatbots are proving to be fundamental tool also helping in improving customer engagement and retention. Amazon's Echo, Alexa, IBM Watson and Microsoft Tay have taken over the world through the advancement in technology (Machine learning, deep learning, AI), Architecture and "advanced information retrieval" processes. Robert Dale in his paper [3] proposed that 2016 has been the return of Chatbots in the world. Microsoft's Cortona, Apple's SIRI, Amazon's ALEXA are now embedded on Desktops. Google Assistant announced in 2016 acquires all the conversational capabilities to interact with humans. Microsoft Tay was also released in 2016 which became a controversial agent. Gini and Peter in their paper examined the case of Tay connected with the users on Twitter to identify their social relationships and responses which at the end underscored Tay's personality.

The paper is drafted into following sections. Section II describes the basis of Chatbot, its architectural model and design approach. Section III discuss about the application domain of Chatbots and its use. Section IV provides an overview of available Machine learning techniques, deep learning methods and algorithms for effective development of Chatbots. Section V focus on different evaluation method to test the performance of Chatbots designed. Section VI discuss about the present trend for popularity of Chatbots among the people. Section VII concludes the work and future of using Chatbots.

II. BASIS OF CHATBOT

From the design of visual layout and interaction mechanisms to the design of conversation there is a blink of an eye transition in design of Chatbots.

1. Classification of Chatbots

Chatbots can be designed in four types based on the requirement of user.

1. **Test-To-Text Bot (TTT**): It is basic architecture where user gives inputs as text in form of words and sentences and Chatbot response similarly using pattern matching or rule based general purpose methods. Suitable example is ELIZA which started the era of Chatbots in the world. Now a days we can find various messenger apps like Facebook Messenger, WhatsApp, Telegram which uses TTS for two user interaction service.
2. **Text-To-Speech Bot (TTS):** TTS is a Chatbot that can speak. It is made more user interactive and provides useful purpose to users for listening to Bots like suitable for Visually impaired persons who can hear or speak but cannot see. Example Snatch-bot
3. **Speech-To-Text Bot (STT):** STT generates text responses where user interacts with the Bot through their voice and view the answers. It is suitable for various purposes like Conferences, hearing impaired individuals to interact socially. SIRI is one of the STT Chatbot developed by Apple.
4. **Speech-To-Speech Bot (STS):** STS is an emerging Chatbot which is mostly like by humans currently. It can be used as voice to voice assistant for teaching in Education sector.

2. Components of Chatbots

For structuring a generic Chatbot there are four basic components.

1. **Natural Language Processing:** NLP is a module used to analyse user requests. It basically takes in unstructured data from the user and turns it into structured data. There are various tools available for the same like Dialog flow.
2. **Dialog Manager:** It is used to decide what to say to the user, given its input, users past interactions stored in the database and the data it learns using different AI methods.
3. **Content:** It is the basis of content the Bot decides to respond to the user after analysis of user input. It can be customized according to design of Bot.
4. **Custom Interaction:** This component is although optional but is used mostly by complex Chatbots. It is used to pull data from web service or databases, runs conditions and informs the Dialog Manager.

3. Design and Approach

Design of Chatbot is based on its generic requirement. It can be a textual bot or a dialogue system. And Dialogue systems are on demand with emerging dialogue chatbot system design and technology. Here described below is a layered approach for design of chatbots advocated by Satoshi Nakamura [18] which includes an utterance layer, discourse layer, dialogue layer and physical layer (including understanding of language, generation and semantics) and intention layer. Most important aspect of this layered structure are theory and evaluation measure for performance of layers together at each stage with explained corpora of spoken or textual dialogue, data sharing among chat and text, ontology and knowledge, emotions, symbols and mutual grounded data.

A typical Chatbot layered structure can be as social as personal as well as business ready assistants. While proposing a standardized architecture for conversational Chatbots Roshan Khan in his paper [9] described a typical layered structure which consists of a) Presentation Layer, b) Business layer, c) Service layer, d) Data layer(storage), e) Utility layer and external services. Now in presentation layer there are components which implements and display user interface and manage user interaction such as multiple channel support, multi-platform support and UI components. Business layer involves data processing, data formatting and dialogue management such as Natural Language Understanding (NLU) with AI/NLP services. Service layer components give access to both external and internal data, middleware connectivity and business functionality. It consists of NLP services, data access services and external service interfaces. Now after this, it has maximum importance to have a well-defined approach in designing a Data layer to have secure and efficient data access. This also enhance the maintenance and decreases the development time of Data layer. Example: storing all data, analysis of data stored and collected, execution of machine learning techniques on data for analysis. This means that storage structures also fall under Big Data processing to allow characters like Sentiment Analytics. Utility layer is an important layer in complete process of architectural operation and so it is important for a system to allow plug-n-play.

Table 1. Chatbots Designed (1960 to 2019)

Chatterbot	Technology/ Approach	Type	Architectural model	Self training	Description
ELIZA (1966) [Open source]	Natural Language Processing: pattern matching, MADSlip	TTT	Retrieval based model	No	Tongue in cheek simulation of Rogerian psychotherapist
PARRY (1972) [Proprietary]	Conceptual ontologies, Turing test	TTT	Retrieval based model	No	Attempt to simulate the behaviour of a person with Paranoid schizophrenia
JABBERWACKY (now Clever-bot) (1988) [Proprietary]	Contextual pattern matching, learns by association storing replies in database	TTS	Generative model	Yes	To simulate Natural human chat for entertainment

Table 1 (Contd.)...

...Table 1 (Contd.)

DR. SBAITSO (1992) [Proprietary]	Pattern matching and substitution programming	TTS	Retrieval based model	No	Developed for MS Dos to converse with user as a psychologist
A.L.I.C.E. (1995) [Open source]	NLP, AIML heuristic pattern matching	TTS	Retrieval based model	No	Applying heuristics pattern matching to the human's input for human ommunication interface.
Watson (2006) [Proprietary]	NLP, IBM's Deep-QA software and Apache UIMA	TTS	Retrieval based model	Yes	A question answering system defeating best human players that won the jeopardy contest in 2011
SIRI (2010) [Proprietary]	Java, JS, objective C, NLP, TTS, STT	TTS/ STT	Generative model	yes	Computer program that works as intelligent personal assistant. It uses a NL user interface to answer questions
Google Now (2012) [Proprietary]	RNN, neural network mechanism	STT	Generative model	yes	Google search mobile apps and perform actions by passing requests to a set of web services, also used for predictive search
ALEXA (2015) [Proprietary]	NLP, TTS, STT, Python, Java, node JS	SST/ STS	Generative model	yes	Personal assistant capable of voice interaction, streaming podcasts, playing music and audiobooks, setting alarms, making to-do list and providing real time information
Cortana (2015) [Proprietary]	Python, Java, JS, NLP	SST/ STS	Generative model	Yes	Intelligent assistant that can set reminders and answer questions sing Bing search engine with multiple languages
Tay (2016) [Proprietary]	Python, java, node JS	text	Generative model	Yes	AI chatterbot created controversy on Twitter by realising inflammatory tweets,

4. Architectural Model

A Chatbot design is based on the core purpose of bot development**.** There are various reasons people use Bots for and therefore Chatbots can be categorized as a personal bot, a customer service bot or a functional bot and so these categorical bots can be responsive in two possible types of models currently used in Deep learning to decide the structure of Chatbot design. First being **Generative model**. Generative model are smart Bots, rarely used but designed for developing complex algorithms These bots interact like a human. Example: Microsoft Tay (Fig 1) [4].

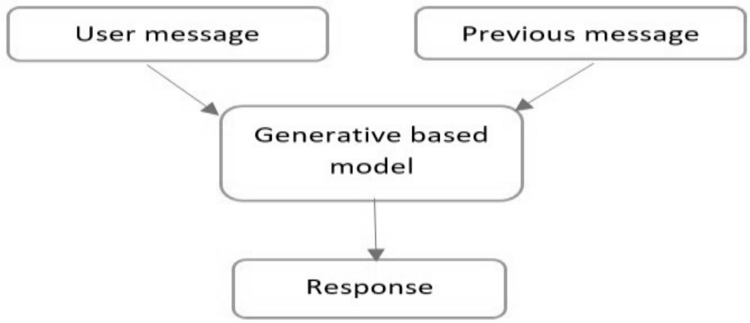


Fig. 1. Generative Model

Second being **Retrieval based Model.** These are easily built with known user conversational context, usually developed using pre-build API's. (Figure 2) [4]. Though this model satisfied performance has achieved. It is easy to implement this model. Its limitation is that sometimes the question asked by the user did not fall into the existed question and answer database

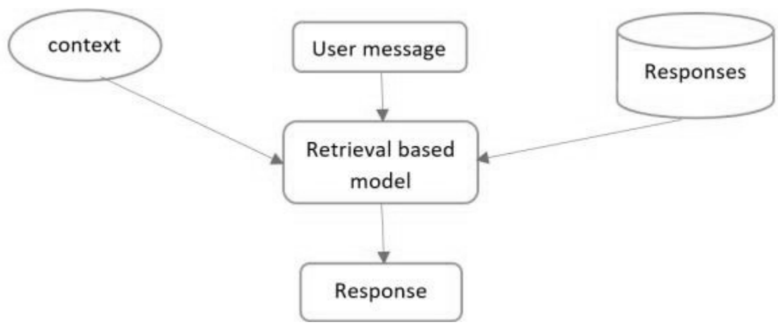


Fig. 2. Retrieval based Model

III. APPLICATION DOMAIN OF CHATBOTS

Design and development of a Chatbot is about getting the knowledge on need of users and it can motivate them. There has been intense shift in how developers and designers are thinking about conversational user interface with data and services based on the need of users. Bayan Abu Shawar, Eric Atwell in their paper[19] investigated various applications of Chatbots that can be useful in information retrieval, education, e-commerce and business.

1. **Chatbots as device for Entertainment:** Traditionally the initial aim of designing chatbot systems was to copy human dialogue and charm users**.** ELIZA was designed for the first time. After this various Chatbots have been designed to simulate different fictional or real personalities.

2. **Chatbot as a device to practice and learn a language:** Chatbots nowadays have included Mark-Up linguistic analysis, such as dialogue-act annotations, semantic, grammatical and linguistic knowledge to generate responses. Therefore, it has become an interesting device to practice and enjoy chatting for users. Lucy was designed for language learning purpose. It was hosted as an online language Robot on Pandora-bots website1 to help English learners review its grammar and vocabulary learned from Lucy's world [6].

3. **Chatbot as a tool for information retrieval:** Chatbots now a days are emerging as user interactive in the form of education**,** such as Sofia. They can be trained to answer the users or even take interviews like VPBOTS. Chatbots as an interviewer is becoming more reliant for business purposes today as theses bots can also evaluate user performance through semantics, gestures and emotions recognition[23].

4. **Chatbots as Assistants in Business and E-Commerce:** Shopping assistant is the most convincing sales. It offers assistance in store for customers, give extra information of products and makes the decisive process[1]. Happy Assistant, a natural language dialog-based navigation system is such example developed that supports

customers to access ecommerce sites and give related data about services and products" [8]. RITA known as Real time Internet Technical Assistant, is currently workable in the ABN AMRO Bank. It is an e-Gain graphical avatar, that helps a customer do financial tasks such. If it do not understand or respond back, it can help redirect the customer to use another channel such as an e-mail or live chat. **Other Applications of Chatbots which may be more useful and motivate the world are** Assistive Technology products to interact with Speech/ Hearing/ Visually impaired persons, Chatbots in Hospitals for patient assistance, health diagnosis**,** AI powered Chatbots for Travel and Transportation and more.

IV. TECHNOLOGY ALGORITHMS AND ML/DL TECHNIQUES FOR CHATBOT DESIGN

A chatbot is programmed to answer questions formulating it into natural language and response back as real human via messages or voice commands. The responses of chatbot are based on a combination of predefined scripts, machine learning and deep learning techniques and algorithms. Data is available in knowledge database. Due to increasing efficient learning algorithms, Chatbot is learning from human interaction, emotions and behaviours which is rising its relevance in future. Techniques like NLP (Natural language processing) which includes NLTK in combination with Java and python, Artificial Intelligence Mark-up Language, Recurrent neural networks, LSTM and Tensor-flow. These can be used to analyse Speech and generate intelligent responses to interact with humans.

A cumulative description about chatbot can be designed on a variety of Platform and can be classified as below:

No programming platforms: These platforms require no coding knowledge and works on the principle of 'Close-click get result'[15]. Some of the related examples are: Chatfuel, Estherbot, Flow XO, Botsify, Kitt.AI.

Conversation oriented: These are Chatbots that interact with humans like a human. Example: Siri, Alexa, Messenger Bots like Facebook messenger, WhatsApp, Hike etc.

Platforms backed by Tech organizations: These are Coding Chatbot platforms which have specific architectural model and algorithms to analyze corpus data [15]. Some of the examples are: IBM Watson, Microsoft BOT Framework, Semantic Machines, Wit.ai and many more.

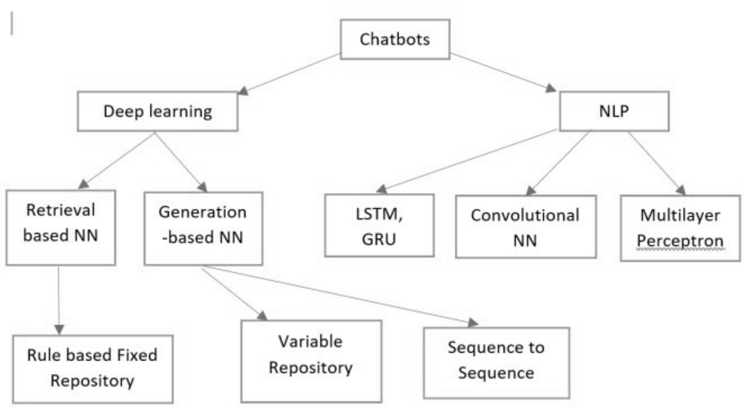


Fig. 3. Techniques for Chatbot Design and Analysis

Figure 3 shows a collection of different techniques used to design efficient algorithms for data analysis of a Chatbot. Artificial Intelligence Chatbots are changing the way the world does business and AI has made Chatbots more lifelike than ever before. To maximize the ability of Artificial Intelligence Chatbots to improve services Machine learning and Deep learning techniques have taken over the programming work (Table II). **Machine learning**: Machine Learning with AI has been traditionally used for building neural conversational chatbot. The machine is designed as to provide informative answers and be indistinguishable from the human by maintaining the context of the Dialogue. It includes Natural language Processing, Artificial Neural Network, LSTM, AIML, optimization algorithms and mathematical processes for analysis and evaluation of Chatbot data and working performance. Machine learning algorithms are designed to easily perform pattern recognition, feature extraction, Automated Speech recognition, POS tagging and text generation (phonemic, hyponyms and semantics). **NLP** is one of the primary branch of Artificial Intelligence (AI) for language processing that allows interaction between computer and human languages [12]. It can be used not only for text translation but for recognizing speech also. One of the successful NLP system developed in early 1960's was ELIZA which used pattern and keyword matching based on substitution methodology. In 1980's most of the NLP's were designed based on some set of hand-written rules. Later they were augmented with Machine Learning (ML) algorithms for language processing. **Artificial Neural Network (ANN) [14]** is a Mathematical model designed to train, visualize, and validate neural network models just like Human Brain. ANN is also known as a model-free estimator as it does not rely on an assumed form of the underlying data. **Recurrent Neural Network (RNN)** is an extension of general feed forward network used for training datasets in which the network not only consider the current input, but also takes the previous output to generate a response. Moreover, RNN's have memory which can be used to remember the input sequence. Like other neural network, RNN has an input layer, output layer and some hidden layers.

The main drawback with RNN is they cannot remember the input for a long sequence but this problem can be solved by **Long Short-Term Memory (LSTM)[20]** and **Gated Recurrent Unit(GRU)** which serves as an extension of RNN and can remember long sequences of data. **Artificial Intelligence Mark-up language (AIML) [5]** a derivative of XML is widely used approach for pattern matching process during data analysis. AIML is based on the software technology developed for A.L.I.C.E. (the Artificial Linguistic Internet Computer Entity). It represents the knowledge that is put into Chatbots. It can characterize the type of data object (AIML objects) which consist of two units:- topics and categories. The data is either parsed or unparsed in these categories. AIML has the ability to describe partial conductance of the programs that it processes. The objective of the AIML language is to simplify the job of conversational modelling, in relation to a "stimulus-response" process. It is based on XML as a variant and so it depends on tags which are the identifiers for making snippets of codes and send commands into the Chatbot. Speech is one the most natural and powerful modes of communication, which is now "widely accepted as the future of interaction with computer and mobile applications.". Because they are more natural than graphic-based interfaces, spoken dialogue systems are beginning to form the primary interaction method with a machine. HCI (Human Computer Interaction) has made researchers ambition to take this concept to improve and enhance speech process between the computer and the human which will help to simulate human speech interaction. Therefore modern networked computing devices with Speech interaction has received enormous interest in humanising machines for users in the future [5]. For Speech based Chatbots Speech-To-Speech or Speech-to-text conversion usually starts with a process known as **Automatic Speech Recognition (ASR). Large Vocabulary Speech Recognition** is a method used in ASR for Speech to Text Conversion. The goal of ASR is to achieve speaker-independent large vocabulary speech recognition (LVCSR). LVCSR can also be improved by measuring it along a number of dimensions such as Vocabulary.

Size, Speaker Independence, Co-articulation, noise-handling and ability to process speech using microphone[18]. **Deep learning**: A Deep learning Chatbot learns everything from its data and human-to human dialogue [5]. In this process, the chatbot is traditionally created using machine learning algorithms. The first step is to prepare data then depending on the source, data reshaping is performed which is followed by data pre-processing. Data pre-processing involves tokenization, stemming and lemmatizing of the chats which makes chats readable for deep learning chatbots. Ones done with pre-processing deep learning chatbots are selectively designed based on **Generative type** or **Retrieval-based**. Various methods are used for chat data evaluation such as generating work vectors using python scripts and training the data using word2vec model like TensorFlow (Seq2Seq model) [24]. This whole process can be summarized with tracking of process and adding it to an application followed by testing and improving methods. Deep learning is known to be a subset of machine learning techniques which enhance the performance of already defined Machine Learning techniques and algorithms to design an effective chatbot. DL includes Convolutional Neural Networks, Tensor-flow, Deep Neural.

Table 2. Technical Approach and Algorithms

Technology	Techniques	Models	Usage	Performance Evaluation Techniques
Natural Language Processing	Feed forward network (FFN), Natural Language Tool Kit (NLTK)	Multilayer Perceptron framework, Multiple reference translation	Pattern matching, Speech recognition, Speech translation	Intrinsic Human posthoc evaluation, Classification matrix, BLEU evaluation matrix, significance testing: statistical bootstrap, Paired tests.
Neural Networks	Linear Regression, Naïve Bayes, AIML, GRU Support vector machines (SVM), Recurrent Neural Network, LSTM	Generative models, Intent classification, Soft-max, regression Ranking model, Seq2Seq model, Google's neural network translation	Pattern extraction, feature engineering Feature detection chatbots, development translation recognition, Dialogue data detection Text generation, image Captioning, machine translation, POS Tag, Dialogue generation	ALICE Bot, C++, Turing test, ROC curves, cost plots, Item response theory model.
Deep Learning	Tensor-flow, DRL, DNN, Convolutional Neural Networks(CNN)	Open AI Gym Toolkit, VGG-19 model	Representation learning, reduce complexity of networks Part of Speech (POS) tagging, Semantic analysis, computer vision tasks	Scalasca[10], Paraver POP performance metric, PARADISE model

Networks (DNN), Deep Reinforcement Learning (DRL)[21]. DNN involve Sequence2Sequence model and Google's Neural Machine Translation (GNNT) and advanced LSTM models. Restricted Boltzmann Machine implementation.

V. EVALUATION MODELS FOR CHATBOT PERFORMANCE

There are a number of different perspectives on how to evaluate chatbot performance. From an information retrieval (IR) perspective, chatbots have specific functions: there are virtual assistants, question-answer and domain-specific bots. Evaluators should ask questions and make requests of the chatbot, evaluating effectiveness by measuring accuracy, precision, recall, and F- score relative to the correct chatbot response [2]. From a user experience perspective, the goal of the bot is, arguably, to maximize user satisfaction. Evaluators should survey users (typically, measured through questionnaires on platforms such as Amazon Mechanical Turk), who will rank bots based on usability and satisfaction. From a linguistic perspective, bots should approximate speech, and be evaluated by linguistic experts on their ability to generate full, grammatical, and meaningful sentences. Finally, from an artificial intelligence perspective, the bot that appears most convincingly human (e.g. passes the Turing Test best) is the most effective. **PARADISE**: PARAdigm for Dialogue System Evaluation is used to estimate subjective factors such as(i) Clarity, (ii) friendliness, (iii) ease for use, (iv) naturalness, (v) robustness regarding misunderstandings. It seeks to objectively quantify bot effectiveness by maximizing task success and minimizing dialogue cost. It introduced the concept of Attribute value matrix (AVM) to measure the effectiveness and two types of minimizing dialogue cost: (i) Efficiency cost, (ii) qualitative cost. Kuligowska et. al. [11] proposed an evaluation framework which they apply to 29 polish-speaking chatbots. Metrics is also used for evaluating Chatbots. A metric is a quantifiable measure that is used to assess a business process. There are several metrics that helps in designing an effective chatbot such as a) Bleu Score: BiLingual Evaluation

Understudy (BLEU) score, a method used to compare a generated sequence of words with reference sequence. BLEU score was proposed by Kishore Papineni in 2002 and was initially developed for translation task only. The advantages of BLEU score are:

- easy to calculate and inexpensive.
- It is language independent
- It correlates highly with human evaluation BLEU score works by counting matching n-grams of user text to n-grams of reference text. The higher the BLEU score, the more intelligent the chatbot.

Turing Test: Turing test is a popular Evaluation test method used to test the ability of machine that exhibits intelligent behaviour equivalent to human. If the tester is unable to distinguish the answers provided by human and machine, then we say that the machine has passed the Turing test. c) **Scalability**: A chatbot is said to be more scalable if it accepts huge no. of users and additional modules. d) **Interoperability** which is the ability of a system to exchange and make use of information. An interoperable chatbot should support multiple channels and users are allowed to switch quickly between channels. e) **Speed**: Regarding speed, the response rate measurement of a chatbot plays an important role. Quality chatbots should be able to deliver responses quickly.

VI. TRENDS FOR POPULARITY OF THE TERM "CHATBOT"

It's been 2019 now, so we have decided to make the profound analysis of chatbots development tendencies which are still being the key trend in personal and business communication technologies. What we observe world-wide is huge popularity and feasible accessibility of Chatbots by the population today and in coming digitally equipped era. Figure 4 and Figure 5 below shows the popularity dynamics for Chatbots (Trends) in Google Trends for last five years including 2019. The popularity for Chatbots among the people of World is ecstatic and it is increasing and emerges to be significant topic throughout the past, present and will remain trending in coming future. There is a growing interest in Chatbots today. In this study we have also identified the trends for popularity of "Chatbots" and how it is motivating the users (age: 15-40 years).

We conducted an online questionnaire asked users from the India to share their reason and interest of using Chatbots. One user used Clever-bot and shared that best feature of it is it learns and stores previous data experience of users. Some used Siri while driving for managing calls and other tasks. One user visited an eye care centre and used chatbot to make an appointment. People who use Zomato found it interesting that Zomato has introduced a feature that enables the customer to communicate their issues through chatbot feature in the application which has significantly reduced the amount of man power required for customer support as well.

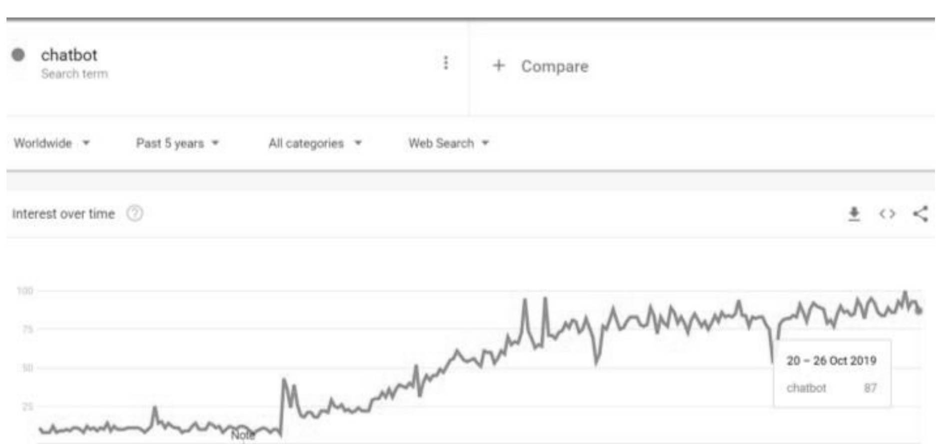


Fig. 4. Worldwide Trend for Chatbots
(past 5 years including 2019)

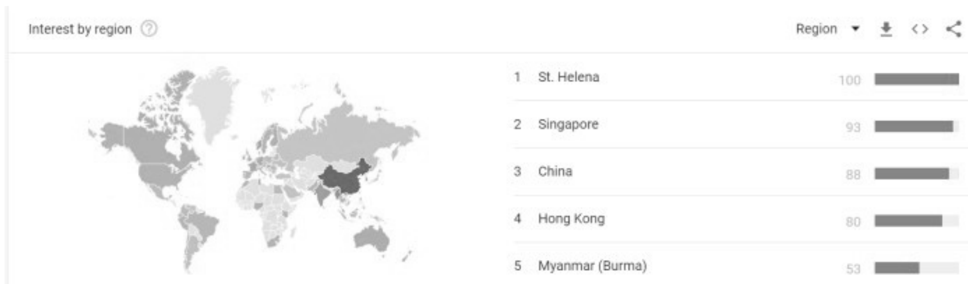


Fig. 5. Worldwide interest by region for Chatbots (in percentage)[7]

Many people also found it motivating and curious to have chatbot features enables in websites for answering their queries and have conversation with the experts. Petter Bae Brandtzaeg et al. [16] in their paper“Why do people use Chatbots” has provided essential insight into the motivational factors(related to use of conversational interfaces) which can guide for future design and development of Chatbots for users. Some of the motivational factors which drives users to use chatbots found is timely and efficient assistance, motivation pertaining to entertainment and motivation to act as personal assistant similar to a Humanoid they expect today. In our last question of our survey we found that 70% of population now expect a real Humanoid to interact with them rather than just being limited to messaging apps. And so, we can predict that in coming years advancement in Artificial Intelligence will impact Chatbots. Chatbots will not remain just as a machine but will work as Humanoid to interact in real as a friend. It will also impact industries and its business world-wide.

VII. CONCLUSION

Chatbot have been around the world for decades. Reason for this increasing interest in chatbots include tremendous advances in artificial intelligence (AI) and AI based technology design and technical development approaches. One of the major usage shift from online to mobile messaging apps and to real interactive agents as robots. Major Internet companies such as Google, Facebook, and Microsoft have already done marvellous work on chatbots and popularized it as a popular technology for all. In this paper we reviewed around twenty journals including related books and websites to generate trends graph, and presented a survey on Chatbots discussing about the basic approach of design and architecture of developing Chatbots along with a broad application domain of using them world-wide from working as personal assistants, organizing meetings, ordering food, making of appointments to booking a flight chatbots which have helped users explore online content and services. Then we have took an overview on different techniques and algorithms developing efficient chatbots with variety of evaluation methods to test its performance. We have also conducted a survey to know how Chatbot brings interest to the users now and how it motivates in future. Developers and designers now have an urge to know more about the user needs that motivate the future use of Chatbots and understand how people experience it. Still miles to cover for becoming humanoids, Chatbots are now part and parcel of our lives. We expect that the way in which people are interacting with conversational user interfaces in present and change their way of interaction in the future will change based on their changing behaviour and expectations along with new social norms Thus, Chatbots are proving to be as real Human than just a material of machine.

REFERENCES

[1] A. Bogdanovych, S. Simoff, C. S. and H. B. (2005). Training of Virtual Shopping Assistant. *“Implicit Training of Virtual Shopping Assistants in 3D Electronic Instituions”. In Proceedings of the IADIS International E-Commerce 2005 Conference, Porto, Portugal, December 15-17, IADIS Press 20*, 50–57.

[2] Cahn, B. J. (2017). *[Thesis] Chatbot Literature Review*. Retrieved from https://static1.squarespace.com/static/569293741c1210fdda37b429/t/59160b6bff7c50104e601a85/1494616940469/CHATBOT_thesis_final.pdf

[3] Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, *22*(5), 811–817. https://doi.org/10.1017/S1351324916000243

[4] Deshpande, A., Shahane, A., Gadre, D., Deshpande, M., Joshi, P. M., Repositiory, I., & Base, K. (2017). *PTC Mogensen.pdf. XI*.

[5] Dr. John woods, S. A. A.-K. (2015). Survey on Chatbot Design Techniques in Speech Conversation Systems. *IJACSA*, *7*. https://doi.org/10.17148/iarjset.2018.596

[6] Fei, Y., & Petrina, S. (2013). Using Learning Analytics to Understand the Design of an Intelligent Language Tutor – Chatbot Lucy. *International Journal of Advanced Computer Science and Applications*, *4*(11), 124–131. https://doi.org/10.14569/ijacsa.2013.041117

[7] Google. (2019). *Google Trends*. Retrieved from https://trends.google.com/trends/explore?date=today 5-y&q=chatbot#TIMESERIES

[8] Joyce Chai*, Jimmy Lin+, Wlodek Zadrozny*, Y. Y., & Margo Stys-Budzikowska*, Veronika Horvath*, Nanda Kambhatla*, C. W. *The Role of a Natural Language Conversational Interface in Online Sales: a Case Study*., (2001).

[9] Khan, R. (2017). Standardized Architecture for Conversational Agents a.k.a. ChatBots. *International Journal of Computer Trends and Technology*, *50*(2), 114–121. https://doi.org/10.14445/22312803/ijctt-v50p120

[10] Kuipers, B., McCarthy, J., & Weizenbaum, J. (1976). Computer power and human reason. *ACM SIGART Bulletin*, pp. 4–13. https://doi.org/10.1145/1045264.1045265

[11] Kuligowska, K. (2015). Department of Information Systems and Economic Analysis, Faculty of Economic Sciences, University of Warsaw, Warsaw, Poland. *Professional Center for Business Research*, *2*(2), 1–16.

[12] Kumar, P., Sharma, M., Rawat, S., & Choudhury, T. (2019). *Designing and Developing a Chatbot Using Machine Learning*. (1), 87–91. https://doi.org/10.1109/sysmart.2018.8746972

[13] Laurence Devillers, Joseph Mariani, M. E. (2019). *Advanced Social Interaction with Agents*. https://doi.org/http://doi.org/10.1007/978-3-319-92108-2

[14] Moustafa, A. A., Alqadi, Z. A., & Shahroury, E. A. (2011). Performance evaluation of artificial neural networks for spatial data analysis. *WSEAS Transactions on Computers*, *10*(4), 115–124.

[15] Oleksii, K. (2019). *Building Chatbots with our without programming skills*. Retrieved from https://medium.com/datadriveninvestor/building-chatbot-with-and-without-coding-skills-bc181d3bb025

[16] Petter Bae, B., & Asbjørn, F. (2018). *Why do People use chatbots?* 3–5. Retrieved from https://www.researchgate.net/profile/Asbjorn_Folstad/publication/318776998_Why_people_use_chatbots/links/5a016a99a6fdcc82a318416a/Why-people-use-chatbots.pdf

[17] Phatnani, H., Patra, J., & Sharma, A. (2015). Review Article CHATBOT ASSISTING : SIRI Address for Correspondence. *International Journal of Advanced Engineering Research and Studies*, *4*(2), 102–103. Retrieved from http://www.technicaljournalsonline.com/ijaers/VOL IV/IJAERS VOL IV ISSUE II JANUARY MARCH 2015/562.pdf

[18] Satoshi Nakamura, Rainer E. Gruhn, W. M. (2011). *Signals and Communication Technology*.

[19] Shawar, B., & Atwell, E. (2007). Chatbots : ¿son realmente útiles? *Jlcl.Org*, *22*(1), 29–49. Retrieved from https://jlcl.org/content/2-allissues/20-Heft1-2007/Bayan_Abu-Shawar_and_Eric_Atwell.pdf

[20] Statement, P. (1966). *Developing a Chatbot using Sequence Modelling*.

[21] Tiha, A. (2018). *Intelligent Chatbot using Deep Learning*. (April), 1–22. https://doi.org/10.13140/RG.2.2.14006.75841

[22] Turing, A. M. (2009). Computing machinery and intelligence. *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, 23–65. https://doi.org/10.1007/978-1-4020-6710-5_3

[23] Van Vugt, H. C., Bailenson, J. N., Hoorn, J. F., & Konijn, E. A. (2010). Effects of facial similarity on user responses to embodied agents. *ACM Transactions on Computer-Human Interaction*, *17*(2), 1–27. https://doi.org/10.1145/1746259.1746261

[24] Wang, Y., Rong, W., Ouyang, Y., & Xiong, Z. (2019). Augmenting Dialogue Response Generation with Unstructured Textual Knowledge. *IEEE Access*, *7*, 34954–34963. https://doi.org/10.1109/ACCESS.2019.2904603

A Casting Algorithm for Chemical Plume Tracking Gas Leakage Source Localization by Mobile Robots

Urvashi Awasthi, Nathi Ram Chauhan, Urfi Khan, V.K. Chawla

Indira Gandhi Delhi Technical University for Women, Delhi, India

Abstract: Incorporating robotics in olfaction industry has opened the gateway to research that focuses on the integration of gas identification and sensing technology on autonomously driven vehicles. In this paper, a novel casting algorithm is proposed for chemical plume tracking and gas leakage source localization by mobile robots under the laminar wind flow conditions. The experimental parameters such as distance traversed, effective displacement distribution and the success rate are calculated for the performance evaluation of the aforementioned algorithm for the gas leakage source localization by a mobile robot. From the results, it is observed that the proposed algorithm offers very promising results and observed to be faster and efficient in its application for the gas leakage source localization.

Keywords: Casting Algorithm, Mobile Robots, Path Planning, Gas detection, Optimization

I. INTRODUCTION

Robotics has always been a topic of intense research and gas sensing adds just another feather to this field. With advances in the technology at a faster pace, easy accessibility and its high rated usage, robotic systems have become an inevitable part of the day to day lives. For instance, using robots as home assistants for basic chores, at offices, workplaces, hospitals, industrial plant inspection and rescue operations are some of the basic applications of robotics. Mobile robots can be used for different tasks such as navigation to an odor source, concentration mapping, continuous inspection and creation of dimensional representations, detection of obstacle and path planning and estimating the exact position and orientation of the robot in the work environment.

Many instances for odor source localization can be found in Mother Nature and nature-inspired algorithms can be developed for solution of odor source localization issues. The nature-inspired algorithms observed to be highly efficient for solving the odor source localization problems. In the insect kingdom the moths can efficiently detect any kind of odors by virtue of their casting behavior. Similarly the casting nature of moths can be used in form of algorithm so as to apply it in mobile robots for solution of chemical plume tracking and gas leakage source localization issues in the industries [12] [15]. The implementation of such natural behaviors for solution of industrial problems by a real mobile robot is the need of the hour for research and its application in the industry [1][14] [13]

II. LITERATURE REVIEW

The research related to mobile robotics olfaction and robotized leak detection dates back to the 1990s which led to the evolution of a number of algorithms. Transient-Response based control algorithm [2] was implemented in 2002 on GaPTR-II robot which was operated with semiconductor gas sensor and speed control is achieved by open-loop pulse width modulation technique. A detailed statistical analysis of two reactive strategies permanent love and exploring love given by Braitenberg [12,13] for localization of static odor source using direct sensor–motor coupling. A stereo electronic [3] mobile nose was developed which can act as an electronic watchman performing detection, localization, and identification of odors resulting from leaking solvents, hazardous gases, etc. In paper [4] a control algorithm based on fuzzy and swarm technique is described for coordinating a network of distributed nodes termed as robots. The development of a biosensor using cell culturing of neuron cell and bulb cell integrated with the light-sensitive potentiometer which is based on the electrophysiological process of the neuronal network is described in the paper [5]. Remote sensing technology is used in the paper [6] which uses infrared spectroscopy for gas detection and a comparison of search-based maximization and triangulation based localization strategies are done. Ecofriendly chemical detection solar operated robotic system [7] was developed in 2011 constituting an autonomous light-sensitive motor-operated tilting mechanism to move in the direction of maximum intensity along with propane and smoke sensors. A low-cost android based gas sensing robot is proposed in paper [8] which offers the easiest and safe way to monitor factories and storage depots. A comparative study of robotic source tracing strategies and moth tracing behavior was conducted by Kuwana et. al [9] in a laboratory setup for experimental validation. The Iterative Chemo-Tropotaxis algorithm is described in the paper [10] which was validated with multiple sensors and works independently of airflow direction. A binary search based algorithm named Spiral Surge was tested by Hayes and co-workers [11] in which the robots moves a certain distance in upwind direction on encountering an odor patch and redefines the surge distance if another odor hit is detected. In the present work, three algorithms (casting, surge-spiral, and surge-cast), have been applied. The obtained results through simulation have been compared.

III. METHODOLOGY

The implementation of casting algorithm for chemical plume tracking and gas leakage source localization by mobile robots are applied and discussed as follows:

1. Casting Algorithm

In the casting algorithm, the robot is initialized to its position and distribution of plume is created. Then the sensor starts monitoring the environment and starts collecting data. The robot starts moving from the initial position following an S-path and checks for the points of plumes. If it finds the plume value above a threshold value it starts surging i.e. moving in an upwind direction at an angle β with respect to the wind flow. If it reacquires again another plume then it starts surging again. If it loses the plume points it travels a distance d-lost and starts searching for the plume by moving in the crosswind direction until the robot hits another plume particle and restarts surging until it finds plume of maximum intensity and declares the source is found and simulation is stopped.

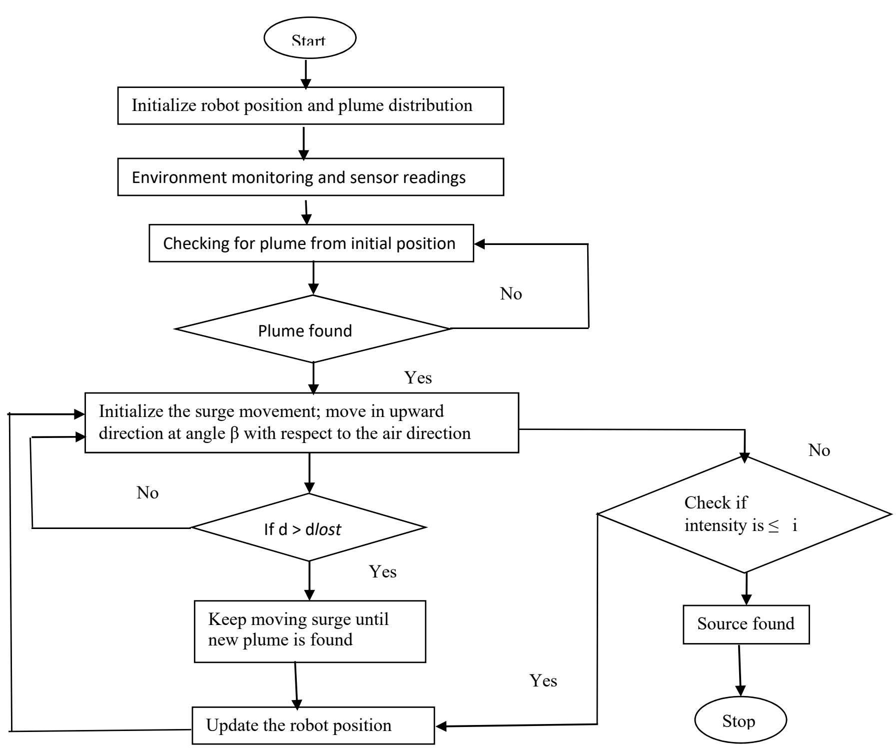


Fig. 1. Flowchart of Casting method for gas leakage source localization by a mobile robot

IV. SIMULATION MODEL

1. Experimental Arena

A rectangular area of 45 units' length and 35 units width is taken to be the experimental arena, which corresponds to the axis definition of the whole simulation model. The initial position of the robot is defined at (0,0) coordinates downwind from that point.

2. Wind and Plume Model

A two-dimensional space having perfectly laminar wind flow at an angle of β=15 degrees with respect to the x-axis and a single odor source emitting a chemical substance at constant rate is taken into consideration. The particle plume mapping algorithm applied here is based on the random walk algorithm which provides a stochastic distribution. To accomplish this, the data values or points generated from a random walk algorithm are converted into a point cloud around the source's location on a global frame. These points are depicted with various colors in order to denote variation in the intensity of the chemical cues the intensity Table 1 is depicted below along with the color specifications.

Table 1. Colour and Intensity Variation for Plume Distribution

S. No.	Color	Intensity	
		Min value.	**Max value.**
1	Red	402.3623	502.7028
2	Magenta	302.0217	402.3623
3	Blue	201.6811	302.0217
4	Green	101.2046	201.6811
5	Yellow	0	101.2046

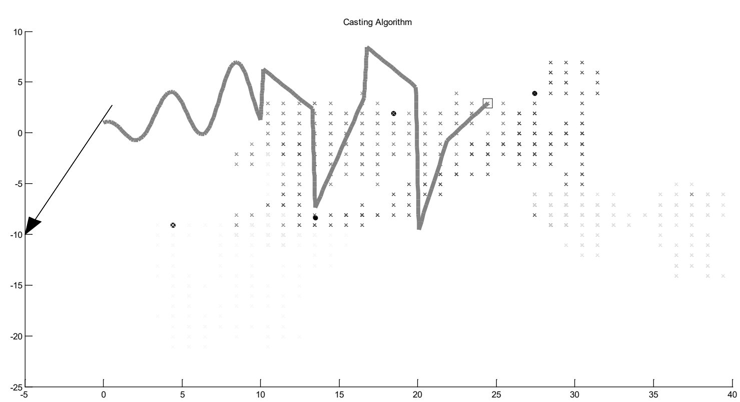


Fig. 2. Casting Algorithm in 2-Dimensional Space

3. Two Dimensional Model

The simulation set-up is created in MATLAB environment with wind flow being considered as laminar and odors randomly distributed. The threshold distance traveled crosswind while searching the plume is 5 units, 12 units, and 10 units for casting algorithms**.** The two-dimensional model shown in Fig 2 consists of leakage source located close to the ground within a two–dimensional representation. The obstacles are introduced at various positions and when robot encounters these it changes its path by moving in a semicircular manner with an obstacle.

V. RESULTS AND DISCUSSIONS

Three main parameters are considered to evaluate the effectiveness of the casting algorithm. These are the distance covered by a robot in the algorithm vs time graphs, overhead distance left from the source to be covered by the robot vs time and success rate which decides how fast each of the algorithms converges to reach the leakage source.

1. Simulation Results

The results are obtained for the casting algorithm applied in two dimensions shown in Fig. 3 to 6 and Table 2. Fig. 3 represents the distance traveled by the robot at each point of time while navigating towards the source. The total distance covered is 95.01 units in 31.01 seconds. Fig.4 shows the variation of displacement with time which is initially maximum and with source of simulation decreases to zeros. The variation is not uniform as robot often deviates from the path when the plume points are lost and starts searching which increases the time as well the distance covered. The non-uniform displacement variation is well presented in Fig.5 which depicts the displacement histogram giving the frequency of various displacement points reached in the course of source tracking. Fig.6 presents the success rate of the casting algorithm which is the function of the ratio of distance overhead which is calculated as the ratio of the distribution of the distance that the robot covers effectively and the distribution of the distance by which it approaches the source. The success is termed as 1 when the robot reaches the target source and the time is taken by it to do so is 31.02 seconds for casting algorithm.

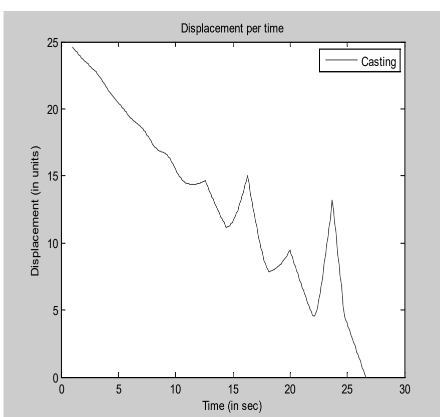


Fig. 3. Variation of Displacement with Time for Casting

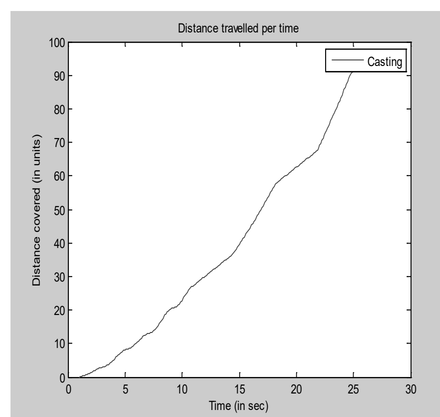


Fig. 4. Variation of Distance with Time for Casting

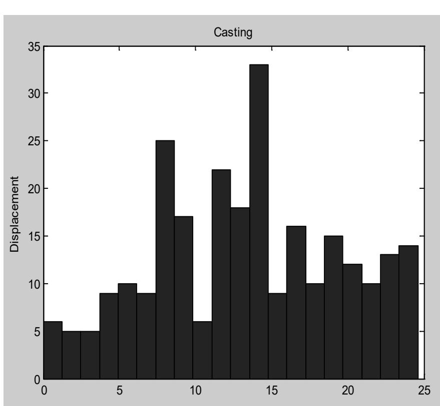


Fig. 5. Displacement Histogram for Casting

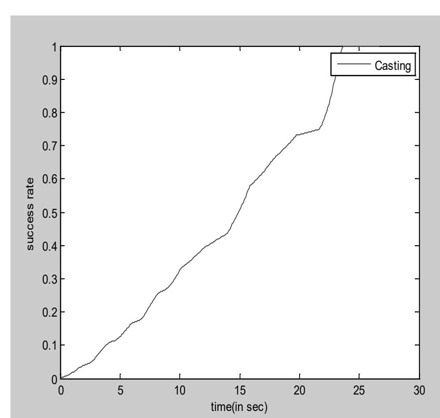


Fig. 6. Variation of Success Rate with Time for Casting

Table 2. Distance Traveled and Time is taken by Casting Algorithm Distribution

Algorithm	**Net Distance Covered**	**Total Time Taken (in seconds)**
Casting	95.01 units	31.01

VI. CONCLUSION

The simulation experiment is used to implement a casting algorithm for chemical plume tracking and gas leakage source localization by a mobile robot. It is observed that casting algorithm is efficient for large upwind angles and less efficient for small upwind angles. It is found that by using the proposed casting algorithm the robot reaches the source in time of 49.19 seconds. It is also observed that the plume lost distance does not have a big impact on the performance of the mobile robot as the simulation runs performed by varying the d_{lost} yields the similar results. Therefore, the simulation methodology provides a good overall picture of the applied casting algorithm. The future research work can be directed to compare the performance of casting algorithms with other bio-inspired algorithms namely surge-spiraling, and surge-cast mimicked from nature for the solution of chemical plume tracking and gas leakage source localization problems.

REFERENCES

[1] G. Kowadlo & R.A. Russell (2008). Robot odor localization: Taxonomy and Survey. *The International Journal of Robotics Research, 27*(8), 869-894.

[2] Hiroshi Ishida, Gouki Nakayama, Takamichi Nakaoto, Toyosaka Moriiziumi (2005).Controlling A Gas/Odor Plume-Tracking Robot Based on transient response of Gas Sensors. *IEEE Sensors Journal.*

[3] Lilienthal A. J. and Duckett T. (2003) . *A Stereo Electronic Nose for A Mobile Inspection Robot*. Paper presented in IEEE Robotic Sensing, June 5, 2003.

[4] Cui X., Hardin T., Ragade R. K. , Elmagharby A. S. (2004). *A Swarm Based Fuzzy Logic Control Mobile Sensor Network for Hazardous Contaminanats Localization Mobile Ad-Hoc Sensor Systems*. Paper presented in IEEE International Conference, 2004.

[5] Quingjun Liu, Ying Xu, Hua Cai, Lifeng Qin, Xuesong Ye, Rong Li, Ping Wang.(2005) . *The Study on Bionic Olfactory Neurochip Based ON light Addressale Potentiometer Sensor*. Paper presented in IEEE Proceedings in Medicine and Biology 27th Annual Conference 2005.

[6] Werner Baetz, Andreas Kroll, and GeroBonow (2009). *Robots with Active IR-Optical Sensing for Remote Gas Detection and Gas Localization*. Paper presented in IEEE International Conference on Robotics and Automation, Kobe, Japan, 2009.

[7] Iem Heng, Andy S.Zhang, Ali Harb (2011).*Using Solar Robotic Technology to Detect Lethal and Toxic Chemicals.* Paper presented in IEEE Global Humaitarian Technology Conference (GHTC), SEATTLE, WA, 2011.

[8] Roger Achkar, Nicolas Makhoul, Christian Abou Haidar, Hassan Osseili (2013). *Gas Ascertainment via Smart Phone*. Paper presented in 7th IEEE GCC Conference and Exhibition , Doha, 2013.

[9] Kunwana Y. ,Nagasawa S., Shimoyaa, I., Kanzaki R. (1999). Synthesis of the Pheromone- Oriented Behaviour Of Silkworms Moths By A Mobile Robot with Moth Antennae as Pheromone Sensors. *Bio-Sensors and Bioelectronics Journal,*1999(192-202).

[10] Russel R. A. , Bab-Hadiashar A. , Shepherd R. L. , Wallece G. G. (2003). A Comparison of Reactive Chemotaxis Algorithms. *Robotics and Autonomous Systems* (45), 45-97.

[11] Lilienthal A. J. , Reiman D. , Zell A. (2003). Gas Source Tracing with a Mobile Robot Using an Adapted Moth Strategy. Paper presented in Autonome Mobile Systeme (AMS) 18, Fachgesprach , 150-160, GDI Informatikaktuell, 2003.

[12] Settles G.S. (2005). Sniffers: Fluid-dynamic sampling for olfactory detection in nature and homeland security The 2004 Freeman scholar lecture. *Journal Of Fluids Engineering , Transaction of the ASME*, vol. 127, (189-218).

[13] Russell R. A. (1999) . Odour Detection by Mobile Robots. *World Scientific Series in Robotics and Integlligent Systems World Scientific Publishing Company* , vol.22, 1999.

[14] Kuenen and Rowe H.C. (2006). Cowpea weevil flights to a point source of female sex pheromone: analyses of flight tracks at three wind speeds . *Physiological Entomology , vol. 31, no.2, p.103*, June 2006.

[15] Chawla V. K., Chanda A.K., Angra S., Rani S. (2019). Effect of nature-inspired algorithms and hybrid dispatching rules on the performance of automatic guided vehicles in flexible manufacturing system. *Journal of the Brazilian Society of Mechanical Sciences and Engineering* 41(10), 391.

Automated Code Correction to Mitigate SQL Injection Vulnerabilities

Charu Gupta[1], R.K. Singh[2], A.K. Mohapatra[3]

[1,2,3]Department of IT, IGDTUW
[1]charugupta@igdtuw.ac.in,
[2]rksingh@igdtuw.ac.in,
[3]akmohapatra@igdtuw.ac.in

Abstract: Vulnerabilities due to SQL injection are approximate 60% of all vulnerabilities. In literature, various algorithms and methods to detect and/or prevent injection attacks due to SQL vulnerabilities are discussed. However, no automated method or tool exists that could remove the SQL injection flaws and correct them using prepared statements of SQL query in the code. It takes hundreds of man-hour to patch and mitigate the vulnerable code constructs that cause SQL injection attacks. In this paper, an approach is proposed that will automate the correction of vulnerable code causing SQL injection attacks. A tool has been developed that automates the correction of code through the use of parameterized queries for communicating with the database layer of a web application. The tool corrects the code with more than 90% accuracy and increases the security of web applications without affecting the functional modules.

I. INTRODUCTION

In the last three years, SQL injection vulnerabilities being reported to have increased many folds from 100 to more than 7000 cases (NVD, 2019). SQL Injection vulnerabilities are flawed SQL statements existing in almost all programming languages including C, C++ to PHP, ASP, C#, JAVA, ruby, python. These vulnerabilities have led to critical cyber-attacks causing information disclosure, disruption, distortion, and destruction. These vulnerabilities form 60 % of vulnerabilities reported on the National Vulnerability Database and have been ranked as critical as per CVSS V3 scoring system (NVD, 2019). In web applications written in PHP programming language, developers use vulnerable functions of mysqli_* form to communicate with database tier. Any flawed data from un-trusted sources leads to malformation of SQL queries and thereby causing code execution attacks, DOS attacks, data destruction, distortion, and disclosure. SQL Injection attacks range from In-band SQLi, Inferential SQLi to Out-of-band SQLi. These attacks are further classified as error –based, union queries based, tautology and time-based injection attacks. An attacker can also take control of the database via HTTP or DNS requests of SQL and PL/SQL requests. (Acunetix, 2016).

In literature, two broad categories of research on SQL injection attacks are available. One group of researchers focuses on the detection and identification of vulnerabilities that exist in a web application and lead to SQL injection attacks. Many open source and commercial tools such as BSQL Hacker, SQLmap, SQLninja, Safe3SQLinjection are also available to detect injection vulnerabilities in web applications (Shankdhar, 2019). The other group of researchers stresses to prevent SQL injection related vulnerabilities. Security experts, guidelines and best practices also lay down principles for preventing injection flaws by validating and filtering all data that is being inputted in a web application.

Very little work is available in the literature that proposes fixing vulnerable code through automated detection and correction approaches. Medeiros and Neves (2015) detected SQL and XSS vulnerabilities using taint analysis and fixed the vulnerable code by using inbuilt PHP functions such as mysql_real_escape_string() and converting my_sql queries to my_sqli queries. However, the code correction has its limitations due to vulnerabilities in my_sqli functions.

Prevention techniques and guidelines lay stress on the use of parameterized SQL statements that will filter the malicious input data before executing SQL queries on the database. However, removing SQL vulnerabilities through manual efforts has its limitations such as portability issues, performance issues, computation complexity, time-consuming and labor-intensive.

In this paper, the authors have proposed to correct code by using more secure and highly acceptable SQL parameterized queries in web applications to communicate with database tier. A tool has been developed in which vulnerable code is detected and corrected with more than 90% accuracy. The automated correction of vulnerable SQL statements has reduced the manual efforts by developers to remove vulnerable SQL queries and replace them with parameterized SQL queries. The results show that the correction of code has increased the security of web applications without affecting the functional modules.

II. RELATED WORK

Mitropoulos, Louridas, Polychronakis, and Keromytis (2019) proposed a model that identified coding flaws causing Cross-Site Scripting and SQL injection attacks and then categorized 41 previously proposed defenses into Parse-Tree Validation, Policy Enforcement, Taint tracking, and Instruction Set Randomization. The set of 41 defenses was also analyzed by Mitropoulos et al (2019) based on their security, availability, accuracy, performance, and deployment and showed that few of these defense mechanisms can be bypassed by naive attackers.

Katole, Sherekar, and Thakare (2018) detected SQL injection attack in all types of database connected applications by profiling, listing, and modularization of SQL queries through the removal of parameters from the SQL query. However, the approach could detect only the type, syntactical errors and predefined vulnerabilities of SQL.

Xiao, Zhou, Yang, and Deng (2017) proposed a URL-SQL mapping by analyzing user behavior and SQL execution response and subsequently blacklisting the malicious users from accessing the application. Uwagbole, Buchanan, and Fan (2016) use the technique of Artificial Neural Network using -Class Averaged Perceptron (TCAP) and Machine Learning using Two-Class Logistic Regression (TCLR). Singh, Kant, Gangwar, and Singh (2015) used a machine-learning algorithm to detect SQL injection attacks and unauthorized users. Tommy, Sundeep and Jose (2017) proposed a machine-learning-based tool namely Bug Terminator Bot (BTB) and improvised automated detection and fixing of SQL vulnerabilities of web-based applications.BTB is composed of four components namely scanner to crawl each web page, fortifier to correct code by suggesting potential fixes, centralized server to maintain existing vulnerabilities and respective fixes and fourth component a firewall (Sundeep and Jose, 2017). Ping (2017) proposed an ISR (Instruction Set Randomization) method that randomizes the trusted SQL keywords of a Web application and constructs new SQL instruction set for detecting second-order SQL injection attacks with low processing cost.

Priyaa and Devi (2016) proposed fragmentation of SQL query parse tree into to *n-dimensional* feature vector to detect SQL related flaws with higher accuracy and speed in the preprocessing while Som, Sinha, and Kataria (2016) detected SQL vulnerabilities by tokenization and information encoding of literals, tables, and fields using AES-algorithm.

Singh, Dayal, Raw, and Kumar (2016) focused on the use of firewalls in SQL server and providing the least privilege to users to prevent SQL flaws. The method proposed by Qian, Zhu, Hu, Liu (2015) detects sensitive characters in the address bar to identify SQL injection attacks. Medeiros, Neves, and Correia (2014) used hybrid methods of data mining to detect SQL vulnerabilities with less false positives. Buja, Jalil, Ali, and Rahman (2014) proposed a more efficient and accurate model that uses the Boyer Moore string matching algorithm to add few parameters in the Parameter testing panel which prevents the web application from SQL injection attack. Jang and Choi (2014) detected the SQL injection attack by analyzing the size of an input. Sadeghian, Zamani, and Manaf (2013) provided a taxonomy of various SQL injection detection and prevention techniques along with their strengths and weaknesses. Shar and Tan (2013) and Sadeghian, Zamani, and Ibrahim (2013) emphasized on the usage of parameterized queries in the coding phase can protect a web application against SQL injection attack. Antunes and Vieira (2012) implemented several layers of security mechanism by various checks in all phases of SDLC. MeiJunjin (2009) developed a prototype tool called SQLInjectionGen for detection of SQL injection vulnerability. The approach of static analysis, runtime detection, and automatic testing have been used in combination.

Bandhakavi, Bisht, Madhusudan, Venkatakrishnan (2007) proposed a scalable mechanism called CANDID that dynamically analyzes programmer intended queries over harmless inputs of the web application.

The literature of the last ten years shows that research is focused on tools and techniques that could detect and/or prevent SQL vulnerabilities using the similar approaches such as defense in depth, static and dynamic code analysis. There are very few approaches that work for correction of web application vulnerabilities (Medeiros and Neves (2015). These approaches have similar pros such as ease of use, greater efficiency and greater accuracy and similar cons such as Performance issues, Computation complexity, time-consuming and labor-intensive.

III. PROPOSED WORK

In this work, a new method has been proposed for preventing SQL Injection Attack and correction the web application vulnerability. In this approach, an algorithm has been developed to automate the detection and correction of SQL injection vulnerabilities in PHP based web applications by making it more secure through Parameterized SQL queries.

1. For each file in an application
2. Read the whole file, two lines together at once (e.g. 1-2, 2-3, 3-4)
3. if the second line contains mysql_ or mysqli_* //vulnerable functions
4. Split the first line by space to get the array of a SELECT statement
5. Array vulQuery={"select", "update", "delete", "Insert"}
6. For each element in vulQuery
7. replace $ sign with : (colon)
8. Merge the array using space
9. Insert parameterized and bind SQL statements.
10. Remove mysql_* or/and mysqli_* statements

IV. RESULTS AND ANALYSIS

The proposed approach was tested on seven applications built on PHP. All the seven applications were analyzed for the presence of vulnerabilities using RIPS (RIPS 2019), an open-source static analysis tool. SQL injection vulnerabilities were detected in all these applications. Three applications namely Online Recruitment Application (ORA), Online Academic Registration, Online Academic Up-gradation each having more than 50 files and more than 300 SQL queries written in plain PHP were detected with 249, 514 and 342 sources of SQL injection vulnerabilities respectively when analyzed with RIPS. These three applications have been corrected for SQL injection attacks with the proposed approach and then analyzed again. It was found that SQL vulnerabilities were substantially reduced. All seven applications were also tested with SQL map to check for SQL injection attacks. These applications were originally vulnerable to union-based, timing based, Boolean, blind SQL and other first-order SQL injection attacks. After code correction using prepared statements of PHP, these applications were hardened for SQLi attacks.

The proposed approach corrected the vulnerable code with approximately 90% accuracy and increased the security of web applications without affecting the functional modules. Web pages have been manually tested and mitigated code increase security without deviating functional requirement. Four open-source vulnerable applications built on PHP were considered to remove injection vulnerabilities. It has been found that while correcting the vulnerable SQL statements took more than 24 hours if done manually, our proposed tool did the same in seconds. The efforts of patching and correcting code have been reduced tremendously. The results of test applications are summarized in the table

Table 1

S. No.	Vulnerable Application	No. of PHP Files	No of Vulnerable SQL Queries	No of SQL Queries Corrected	Time Consumption (sec)
1.	DVWA	105	7	7	1
2.	BTSLAB	52	5	3	2
3.	BRICKS	23	11	11	2
4.	TWITTERLIKE	14	1	1	1
5.	Online Recruitment Application	56	249	225	240
6.	Online Academic Registration	160	514	496	525
7.	Online Academic Up-gradation	55	342	338	486

The corrected files and applications were then tested for any deviation in their functional modules. It was found that applications passed functional testing effectively. Moreover, the applications after code correction were also put for the detection of SQL injection vulnerabilities and penetration testing. These test applications were found less vulnerable to SQL injection attacks after automated code correction.

The removal of SQL injection vulnerabilities through automated use of prepared statements minimizes the probability of database exploitation and SQL injection attacks. Automated correction is less labor-intensive as it has reduced the need of software developer to manually make changes in the code. While mysqli_* functions work only with MYSQL database, code correction with prepared statements has made the applications portable and they can support as many as 12 database drivers apart from MySQL.

V. CONCLUSIONS AND FUTURE WORK

In this work, the authors have proposed a novel approach to detect and remove SQL injection attacks in web applications written in PHP. Several approaches for automated detection of vulnerabilities due to SQL are available in the literature. However, this is a novel approach that will correct the vulnerable code by replacing them with parameterized SQL statements and hence reduce the efforts of software developers to patch such vulnerabilities. The vulnerable code is corrected with more than 90% accuracy to increase the security of web applications without affecting the functional modules. The remaining SQL queries could not be corrected owing to their complex nature such as the use of inner join, group by / having clause, updating columns or/ and rows in more than one table, etc. The mitigated code increases the security of applications by removing maximum SQL vulnerabilities without deviating functional requirements. Correcting vulnerable SQL may take exponentially higher efforts and time. However, with this approach, 90% of the SQLi based statements can be corrected and only 10% manual intervention for correcting complex SQL queries is required.

In the future, authors aim to automate the correction of more complex vulnerable SQL queries that include inner join constructs, aggregate functions, group-by / having clause, updating more than one table and /or database and nested SQL queries.

REFERENCES

[1] Acunetix. (2016, March). Types of SQL Injection? Retrieved February 2019, from https://www.acunetix.com/websitesecurity/sql-injection2/

[2] Antunes, N. (2012). Defending against Web Application Vulnerabilities. *IEEE Computer*.

[3] Bandhakavi, S., Bisht, P., Madhusudan, P., & Venkatakrishnan, V. N. (2007, October).

[4] Buja, G., Jalil, K. B. A., Ali, F. B. H. M., & Rahman, T. F. A. (2014, April). Detection model for SQL injection attack: An approach for preventing a web application from the SQL injection attack. In *2014 IEEE Symposium on Computer Applications and Industrial Electronics (ISCAIE)* (pp. 60-64). IEEE.

[5] CANDID: preventing SQL injection attacks using dynamic candidate evaluations. In Proceedings of the 14th ACM conference on Computer and communications security (pp. 12-24). ACM.

[6] CVE. (2019, June). Vulnerabilities By Type. Retrieved June, from https://www.cvedetails.com/vulnerabilities-by-types.PHP

[7] Damele, B., Guimaraes, A., & Stampar, M. (2006). sqlmap®. Retrieved October 2019, from http://sqlmap.org/.

[8] Jang, Y. S., & Choi, J. Y. (2014). Detecting SQL injection attacks using query result size. *Computers & Security*, *44*, 104-118.

[9] Junjin, M. (2009, April). An approach for SQL injection vulnerability detection. In *2009 Sixth International Conference on Information Technology: New Generations* (pp. 1411-1414). IEEE.

[10] Katole, R. A., Sherekar, S. S., & Thakare, V. M. (2018, January). Detection of SQL injection attacks by removing the parameter values of the SQL query. In 2018 2nd International Conference on Inventive Systems and Control (ICISC) (pp. 736-741). IEEE.

[11] Marcus, D. (2017, November 26). PHP PDO Prepared Statements Tutorial to Prevent SQL Injection. Retrieved March 2019, from https://websitebeaver.com/PHP-pdo-prepared-statements-to-prevent-sql-injection

[12] Marcus, D. (2018, June 8). PDO vs. MySQLi: The Battle of PHP Database APIs. Retrieved March 2019, from https://websitebeaver.com/PHP-pdo-vs-mysql

[13] Marjanovic D. (2017, November 20) Retrieved March, from https://code.tutsplus.com/tutorials/pdo-vs-mysqli-which-should-you-use--net-24059

[14] Medeiros, I., Neves, N. F., & Correia, M. (2014, April). Automatic detection and correction of web application vulnerabilities using data mining to predict false positives. In *Proceedings of the 23rd international conference on World wide web* (pp. 63-74). ACM.

[15] Mitropoulos, D., Louridas, P., Polychronakis, M., & Keromytis, A. D., Defending Against Web Application Attacks: Approaches, Challenges, and Implications; In march/April 2019 IEEE transactions on dependable and secure computing, vol. 16, no. 2, IEEE.

[16] Mohanty, S. (2018, March 08). 5 Important Software Vulnerabilities - DZone Security. Retrieved February 2019, from https://dzone.com/articles/5-important-software-vulnerability-and-attacks-tha

[17] NVD. (2019, June 7). National vulnerability database. Retrieved from https://nvd.nist.gov/vuln/search

[18] OWASP. (2019, June 2). Category: OWASP Top Ten Project. Retrieved June, from https://www.owasp.org/index.PHP/Category:OWASP_Top_Ten_Project

[19] Ping, C. (2017, December). A second-order SQL injection detection method. In 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (pp. 1792-1796). IEEE.

[20] Priyaa, B. D., & Devi, M. I. (2016, January). Fragmented Query parse tree based SQL injection detection system for web applications. In 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16) (pp. 1-5). IEEE.

[21] Qian, L., Zhu, Z., Hu, J., & Liu, S. (2015, January). Research of SQL injection attack and prevention technology. In *2015 International Conference on Estimation, Detection and Information Fusion (ICEDIF)* (pp. 303-306). IEEE.

[22] RIPS. (2019). Retrieved October 2019, from https://www.ripstech.com/free-trial/.

[23] Sadeghian, A., Zamani, M., & Ibrahim, S. (2013, September). SQL injection is still alive: a study on SQL injection signature evasion techniques. In *2013 International Conference on Informatics and Creative Multimedia* (pp. 265-268). IEEE.

[24] Sadeghian, A., Zamani, M., & Manaf, A. A. (2013, September). A taxonomy of SQL injection detection and prevention techniques. In *2013 International Conference on Informatics and Creative Multimedia* (pp. 53-56). IEEE.

[25] Shankdhar, P. (2019, May 08). Best Free and Open Source SQL Injection Tools [Updated 2019]. Retrieved June 2019, from https://resources.infosecinstitute.com/best-free-and-open-source-sql-injection-tools/#gref

[26] Shar, L.K. and Tan, H.B.K. Defeating SQL Injection; Published by the IEEE Computer Society March 2013.

[27] Singh, G., Kant, D., Gangwar, U., & Singh, A. P. (2015). SQL injection detection and correction using machine learning techniques. In *Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1* (pp. 435-442). Springer, Cham.

[28] Singh, N., Dayal, M., Raw, R. S., & Kumar, S. (2016, March). SQL injection: Types, methodology, attack queries, and prevention. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*(pp. 2872-2876). IEEE.

[29] Som, S., Sinha, S., & Kataria, R. (2016). Study on SQL injection attacks: Mode detection and prevention. *International Journal of Engineering Applied Sciences and Technology, Indexed in Google Scholar, ISI, etc., Impact Factor: 1.494*, *1*(8), 23-29.

[30] Tommy, R., Sundeep, G., & Jose, H. (2017, September). Automatic Detection and Correction of Vulnerabilities using Machine Learning. In 2017 International Conference on Current Trends in Computer, Electrical, Electronics, and Communication (CTCEEC) (pp. 1062-1065). IEEE.

[31] Uwagbole, S. O., Buchanan, W. J., & Fan, L. (2016, April). Numerical encoding to Tame SQL injection attacks. In *NOMS 2016-2016 IEEE/IFIP Network Operations and Management Symposium* (pp. 1253-1256). IEEE.

[32] Xiao, Z., Zhou, Z., Yang, W., & Deng, C. (2017, May). An approach for SQL injection detection based on behavior and response analysis. In the 2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN) (pp. 1437-1442). IEEE.

Acoustic Analysis of Punjabi Oral and Nasal Vowels

Shyam S. Agrawal[1], Shweta Bansal[2]

[1,2]*KIIT College of Engineering* Gurugram, India

[1]ss_agrawal@hotmail.com, [2]bansalshwe@gmail.com

Abstract: In Punjabi language the phonemic opposition of oral and nasal vowels is very significant as it is found in many other languages also. There are ten vowels in Punjabi which can be spoken in oral as well as in nasalized form. In the present study the acoustic parameters of oral and nasal vowels spoken in their initial, medial and final position have been studied. The first three formants F1, F2 & F3 of the oral as well as nasal vowels in their steady portion have been measured and compared. It has been observed that there is a shifting of these formants, first formant going up and second formant going down, thus decreasing the distance between F2 &F1 in the case of nasal vowels. Vowel triangles have been plotted and compared. Presence of nasal formant (pole and zero) has been observed near the first formant. Fundamental frequency F_0, of all the vowels have been measured and compared. The degree of nasality for all the vowels have been measured and compared with oral vowels for both male and female speakers. The spectral properties as well as nasality measures show significant differences among oral and nasal vowels.

Keywords: Punjabi corpus, Oral vowels, Nasal vowels, Acoustic analysis Introduction

I. INTRODUCTION

Punjabi is one of the official languages spoken mainly in Northern part of India and in some other states. It is also spoken in Pakistan, Canada and a few other countries. It has 43 phonemes out of which 33 are consonants including 5 tonemes and 10 vowels. The vowels can be produced as oral as well as nasals and show significant phonemic opposition [2]. Table I shows list of Punjabi oral and nasal vowels along with their manner and place of articulation and example words in which they are spoken in initial, medial and final position.

The Punjabi vowel system comprises of 10 vowels out of which three are short vowels /ɪ,ʊ, ə/ and rest seven are long vowels. Out of seven long vowels, /i, e, ɛ/ are front vowels and /u, o, ɔ ,a/ are back vowels. Complete vowel inventory is given in Figure 1. Figure 1 shows the production of these sounds in relation to tongue height, openness/closeness and front/back articulatory positions. Most of these positions are similar to those produced in Hindi and other Indo -Aryan languages.

Objectively, the vowels can be defined in terms of their formant frequencies, primarily by examining the first two formant frequencies F1 & F2. Each vowel has unique F1 & F2. F1 refers to the width of the Pharyngeal cavity and position of tongue on a vertical scale (ranges from open to close) whereas F2 refers to the length of oral cavity and position of tongue on a horizontal axis.

In the present study, an attempt has been made to do acoustic analysis of Punjabi oral as well as nasal vowels by studying and comparing their formant structures and the nasality differences in oral and nasal Punjabi vowels.

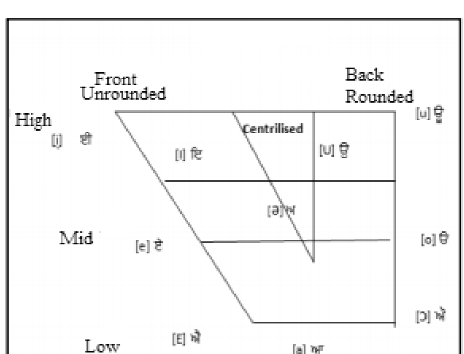


Fig. 1. Vowel inventory of Punjabi

Table 1. List of Oral and Nasal Vowels in Punjabi

Punjabi Phones	IPA	Initial Position		Medial Position		Final Position		Manner & Place of Articulation
		Oral	Nasal	Oral	Nasal	Oral	Nasal	
ਅ	/ə/	ਅਰਾਮ	ਅੰਦਰ	ਮਾਧਿਅਮ	ਵਿਅੰਜਨ	ਖਿਲਾਅ		Unrounded mid Central short
ਆ	/a/	ਆਇਆ	ਆਂਕੜਾ	ਪਿਆਜ਼	ਸੰਸਾਰ	ਦੁਖੀਆ	ਧੂੰਆਂ	Unrounded low Central long
ਇ	/I/	ਇਕਾਈ	ਇੰਜਣ	ਚਿਰ	ਸਿਕੰਦਰ	ਖਾਇ	ਸੇਂਵਇ	Unrounded High-mid front short
ਈ	/i/	ਈਸਾਈ	ਈਂਦੂ	ਕਈਆਂ	ਪੀਂਘ	ਸਫਾਈ	ਗਾਂਧੀ	Unrounded High Front long
ਉ	/U/	ਉਮਰ	ਉਂਚਾ	ਬਣਾਉਣਾ	ਸਮੁੰਦਰ	ਮਿਲਾਉ	ਬਿੰਦੂ	Rounded High-Mid back short
ਊ	/u/	ਊਚਾ	ਊਂਧਾ	ਟਿਊਬ	ਸੂੰਘਣਾ	ਕਮਾਊ	ਹਿੰਦੂ	Rounded High back long
ਏ	/e/	ਏਕਤਾ	ਏਂਠਨਾ	ਸੇਕ	ਅੰਗਰੇਜ਼	ਆਏ	ਮੁੰਡੇ	Rounded low-mid front short
ਐ	/ɛ/	ਐਲਾਨ	ਐਂਠ	ਪੈਰ	ਟੈਂਕ	ਲਾਈਐ	ਮੈਂ	Unrounded low-mid front long
ਓ	/o/	ਓਸਨੂ	ਓਂਕਾਰ	ਡੋਰ	ਗੋਂਧ	ਬਚਾਓ	ਸਰੋਂ	Rounded low-mid back short
ਔ	/ɔ/	ਔਰਤਾਂ	ਔਂਕੜ	ਦੌੜ	ਚੌਂਕ	-	-	Rounded low back long

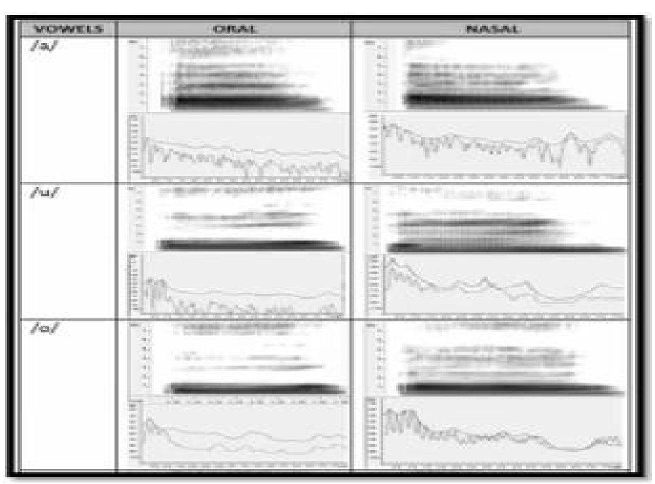
II. TEXT MATERIAL AND RECORDING

The set of text corpus containing oral and nasal vowels spoken in initial, medial and final positions of words (as examples shown in table I). This text material was recorded by 10 male and 10 female native speakers of Punjabi with clear pronunciation and without speech/hearing impairment. The recording of the utterances was done using zoom H4N audio recorder using internal microphone in the studio environment in normal text reading mode

The recording format is 16 bit mono at a sampling frequency of 48000 Hz. After each recording, the moderator checked for any wrong pronunciation during the recording, and if so, the utterances were recorded again.

III. SPECTROGRAPHIC ANALYSIS OF ORAL & NASAL VOWELS

Fig. 2. depicts the comparison of spectrograms along with a power spectrum for a small segment of about 20 m sec in the steady state of the oral and nasal vowels. The spectrum is taken at the most steady point in the spectrogram. The differences in the formant frequencies and the presence of additional pole (formant) and broadening of low frequency spectrum for nasal sounds can be seen. These differences are prominent and are in agreement with the observations of Howkins and Stevens [4]. These differences are not same for all the vowels. Some vowels such as /i/,/o/,/u/ show more prominence compared to others vowels.



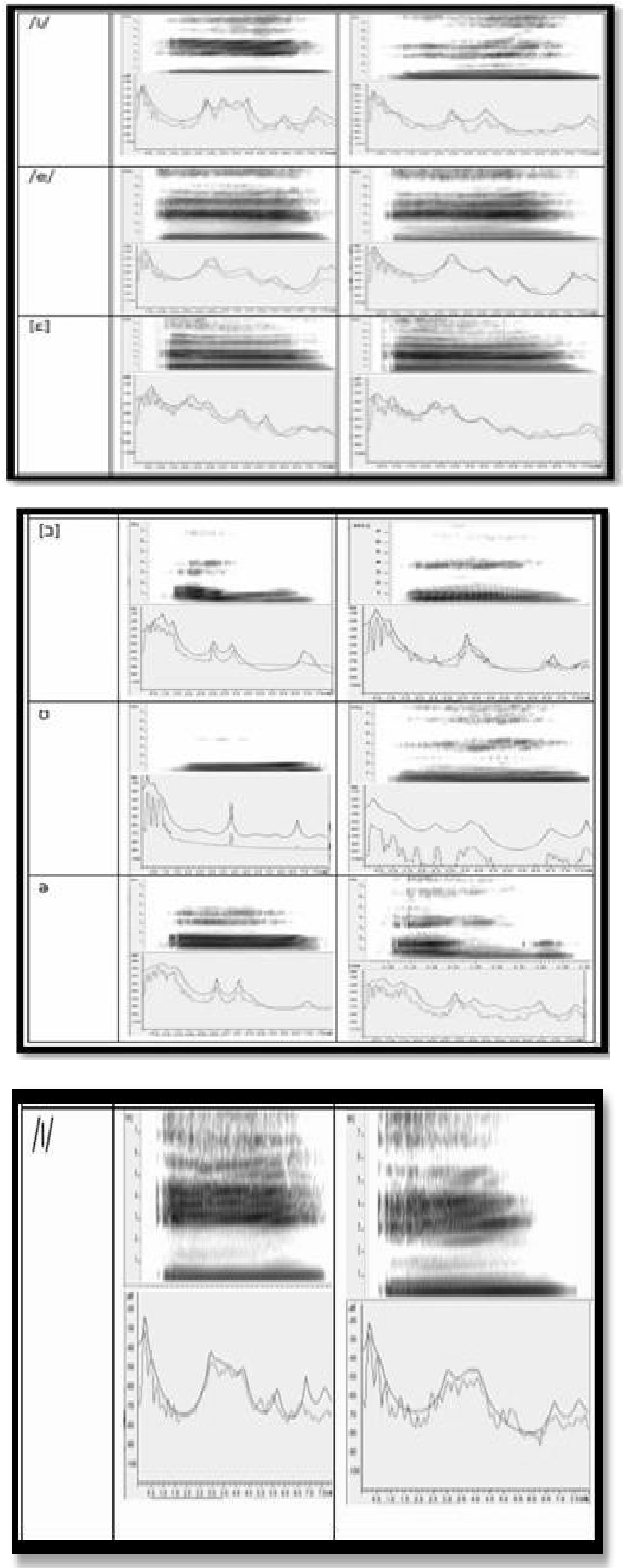


Fig 2. Power Spectrum Analysis of Nasal & Oral Vowels

IV. ACOUSTIC ANALYSIS OF ORAL & NASAL VOWELS

For Acoustic analysis 20 msec segments of the steady portions of the vowels were taken out manually and analysed using PRAAT and WAVESURFER software packages. The parameter settings for the extraction of the formants using WAVESURFER are as follows:

- Analysis window length 0.049 s
- Pre-emphasis factor 0.7
- Frame Interval 0.001s

The formant values of those vowel segments, which were outside mean ± standard deviation was once again checked and if found erroneous these were corrected manually using spectrum section.

The measured values of first three formant frequencies (F1 F2 & F3) of the oral and nasal vowels are shown in table II . This table represents the average and standard deviation of all the three formants for oral and nasal vowels spoken by all speakers of Punjabi Language. It is observed from table 2 that the values of F1 for nasal vowels are greater than the oral vowels. F1 represents the height of vowels therefore we can say that the greater F1 means lower and open articulation of nasal vowels in comparison with oral vowels.

From the table II, it is also been observed that the standard deviation of both the formants of each of the vowel is almost one third of the average values of the formants. This signifies that the data is quite consistent.

Table 2. Mean and SD of F1, F2 & F3 for Nasal and Oral Vowels (for Male)

Phoneme		ORAL			NASAL		
		F1	F2	F3	F1	F2	F3
/u/	Mean	343.0398	865.5428	1269	452.3333	873.3333	1318
	Stdev	33.18065	207.0833	389	158.7	300.33	400.33
/ʊ/	Mean	393.2783	988.3814	1452.1	433.6	834	1264
	Stdev	43.52361	146.1241	458	150.2	275.3	416.1
/o/	Mean	475.4902	1025.51	2218	556.6667	840	1791
	Stdev	41.25176	304.0306	520.33	50.03	127.36	532.8
/ɔ/	Mean	591.9118	1278.32	2351	665.6667	1073.667	2051
	Stdev	92.74894	472.9747	690	200.89	289.23	620
/ə/	Mean	598.595	1548.995	2762	728.3333	1305	2485
	Stdev	133.1659	372.0221	296.2	123.19	435	138.5
/ɛ/	Mean	615.5134	1389.063	2283	686.6667	1867.333	2503
	Stdev	109.8567	527.481	701	129.8	612.1	826
/ɑ/	Mean	712.1366	1123.821	2061.3	791.6667	1500.667	2332
	Stdev	115.4534	184.3248	249	125.4	512.3	169
/e/	Mean	439.3916	2291.049	3010.1	453	2341.667	3125
	Stdev	50.21917	483.0204	958	58.2	496.4	852
/ɪ/	Mean	417.6002	2255.896	3122	502	2612.667	3220
	Stdev	69.96136	467.0975	189	79.6	480.5	230
/i̥/	Mean	315.8586	2252.702	3056.12	340.3333	1917	2767.25
	Stdev	58.05133	855.9026	995.2	60.05	650.6	922.8

V. COMPARATIVE ANALYSIS OF VOWEL TRIANGLE FOR ORAL & NASAL VOWELS

The vowel triangle represents the perceptual distances or the dispersion among various vowels. Perceptual spaces of Punjabi speakers for the cardinal vowels ई /i/, आ /a/, ऊ /u/ have been shown in fig. 3. These vowels were chosen because these vowels represent the cardinal position of articulation which is most common in human spoken languages.

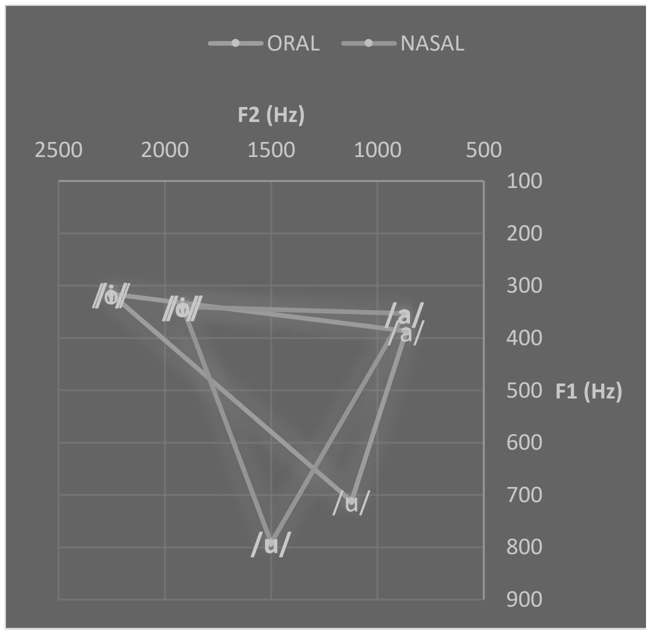


Fig. 3. Oral & Nasal Vowel Triangles of Punjabi

VI. MEASUREMENT OF NASALANCE FOR ORAL AND NASAL VOWELS

The term nasalance is a measure of velopharyngeal closure during voiced speech in which nasally emitted acoustic energy is compared to the orally emitted energy. The nasalance of oral and nasal vowels was measured using Nasometer, a commercially accessible computer- based device which consists of the separator and a nasal mask as shown in fig.4.

Fig.4. Separator and nasal mask

Software is provided with the system that can be used interchangeably with either a handle that uses the dual-chamber circumferentially vented (CV) mask, or a handle using a partition plate called separator held against the lips as shown in fig. 5:

Fig. 5.: Pictorial presentation of mask using partition plate

The term 'nasalance' is used to describe the acoustic energy at the nares, N_a, and the acoustic energy at the mouth, N_o, during the vowel production. This balance between N_a and N_o can be expressed as a simple ratio, N_a/N_o, to yield a measure that can be referred to as a "Nasalance Ratio" (NR) or it can be expressed as a percentage, N_a / (N_o + Na), to yield a measure that can be referred to as "% Nasalance" (%N) [8]. The typical setup for acquiring the nasalance from the informants is shown in fig. 6:

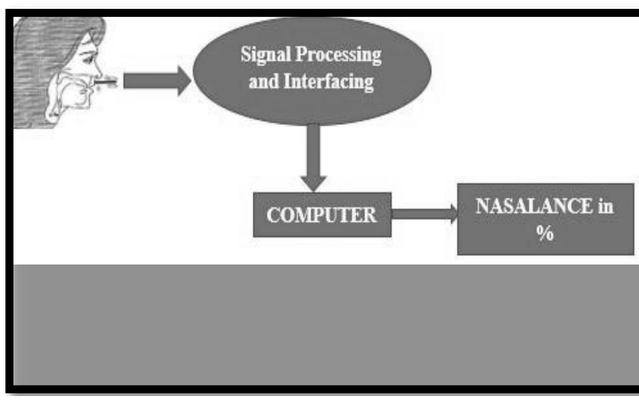


Fig. 6. Set up for capturing the nasalance

Example of data analysis with Nasometer (using separator) for oral and nasal vowels is shown in fig.7. In this figure, purple colour visually differentiates the nasal vowels and red colour indicate the oral emission. The sound pressure level of voice is indicated in green colour.

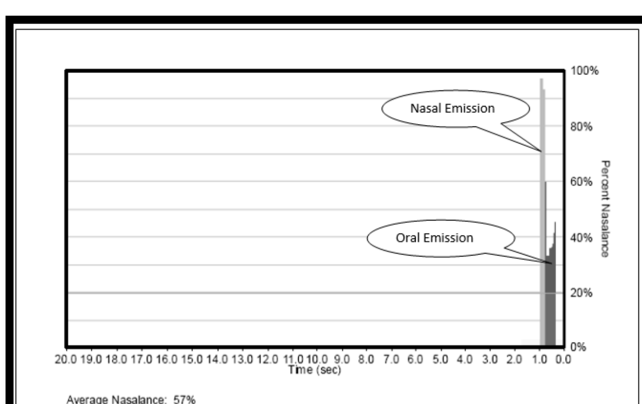


Fig. 7. Measurement of Nasality (oral vs nasal)

The nasalance of oral and nasal vowels were studied both for Male and Female Speakers. The nasalance scores for Male and Female Speakers for different vowel are displayed in figure 8. It has been observed that, each vowel had a different nasalance score. It has been observed that the female speakers always produced a higher nasalance score than the male speakers for all the vowels.

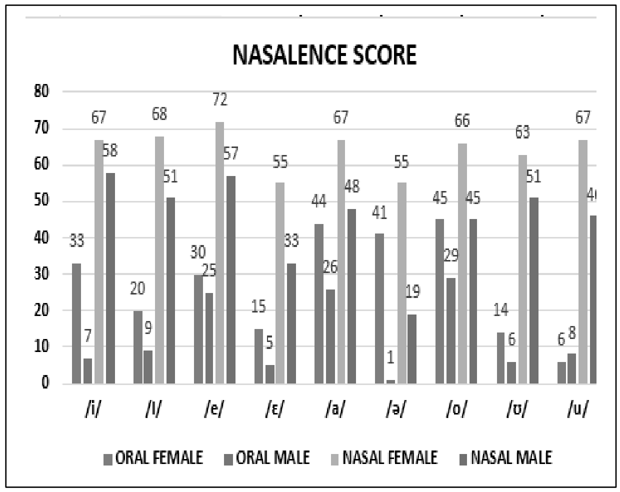


Fig. 8. Nasalance score for male & female speakers

The greatest difference between the female and male mean nasalance scores was observed during the production of the high front oral vowel /i/ and the low back vowel /u/. From figure 8, it has been observed that the greatest difference in mean nasalance score for male and female was in short vowel /ə/.

VII. FUNDAMENTAL FREQUENCY (FO) ANALYSIS

The fundamental frequency (F0) for all the oral and nasal vowels is shown in figure 9. It may be noticed that the nasals have higher fundamental frequency compared to oral vowels in general. Also, the central vowels have lower Fo compared to front and back vowels. This may be due to slight compression of vocal folds during the production of nasal vowels.

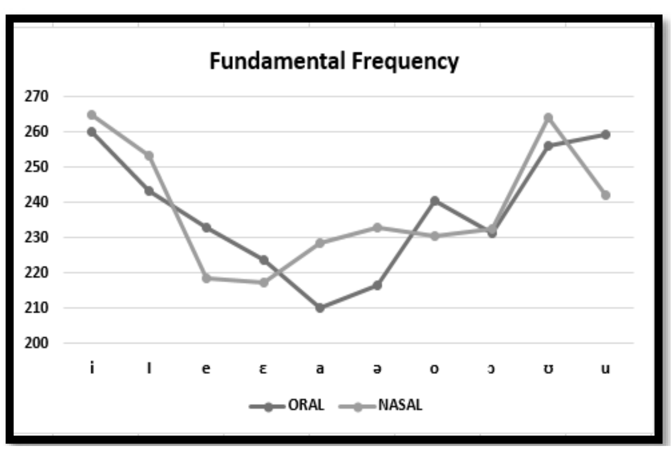


Fig. 9. Comparison of Fo for oral and nasal vowels

VIII. CONCLUSION AND DISCUSSION

1. The spectral properties show that the energy of oral vowels are more than those energy of the nasal vowels. This is perhaps due to sound attenuation in the nasal cavity.

2. The formant frequencies show difference in the oral and nasal vowels. There is an appearance of additional formant between the first and second format (its position varies depending upon the vowels. It is better in open vowels such as /a/ compared to close vowels (such as /u/)).
3. The Nasal oral contrasts depend upon the vowel type.
4. Nasometer measurements show that the nasalance for Nasal vowels is significant by higher than those of the oral vowels.
5. The nasal vowels have higher fundamental frequency compared to the oral vowels.

ACKNOWLEDGMENT

The authors wish to acknowledge the support receive from DeitY for conducting the experiment. They also thank the support and encouragement given by Dr. Harsh Kamrah, Dr. Neelima Kamrah and the assistance provided by Mr. A. K. Dhamija, Ms. Anupam Bansal and Mr. Bidesh Das

REFERENCES

[1] Shyam S. Agrawal, Shweta Bansal, Shambhu Sharan, and Minakshi Mahajan. "Acoustic analysis of oral and nasal Hindi vowels spoken by native and non-native speakers" *The Journal of the Acoustical Society of America. htt*ps://doi.org/10.1121/1.4970648

[2] Shyam S. Agrawal, Shweta Bansal, Shambhu Sharan, and Amritpal Singh. "Acoustic phonetic study of phonemes and tonemes of spoken Punjabi language" , *The Journal of the Acoustical Society of America*.139,2214 (2016); https://doi.org/10.1121/1.4950619.

[3] Shweta Bansal Shambhu Sharan, S. S. Agrawal "Corpus design and development of an annotated speech database for Punjabi" International Conference Oriental COCOSDA held jointly with Conference on Asian Spoken Language Research and Evaluation, Shanghai, 2015, pp. 32-37.

[4] Hawkins, Sarah & N. Stevens, Kenneth. (1985) "Acoustic and perceptual correlates of the non-nasal-nasal distinction for vowels" The Journal of the Acoustical Society of America. 77. 1560-75. 10.1121/1.391999.

[5] IPA transcription with SIL fonts http://scripts.sil.org/.

[6] Hawkins, S., & Stevens, K. N. 1983 "A cross-language study of the perception of nasal vowels". *Acoustical Society of America*, Cincinnati, Ohio .

[7] A.M. Ansari, S.S. Agrawal, M Ganesan and K.D. Pavate. "Study of Acoustic Characteristics of Hindi Vowels in different Phonetic Context". *Journal of Acoustic Society of India*. Vol XVII, No. 1 & 2,1989, pp. 295-300.

[8] Lewis, Kerry & Watterson, Thomas & Quint, Teresa. (2000). The Effect of Vowels on Nasalance Scores. The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association.37.10.1597/154569(2000)037<0584:TEOVON>2.0.CO;2.