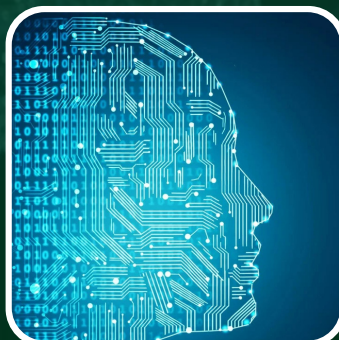




6th INTERNATIONAL CONFERENCE ON Artificial Intelligence & Speech Technology AIST 2024 13th - 14th November, 2024



Centre of Excellence – Artificial Intelligence
(Supported by Department of Science and Technology, GOI)
Indira Gandhi Delhi Technical University for Women
(Established by Govt of NCT of Delhi),
Kashmere Gate, Delhi-110006
www.igdtuw.ac.in



6th International Conference on Artificial Intelligence and Speech Technology
AIST 2024

13th November – 14th November, 2024

Guest of Honour



Dr. L Venkata Subramaniam,
Quantum India Leader - IBM

Chief Guest



Prof. Lalit K. Awasthi
Hon'ble Vice Chancellor
Sardar Patel University, HP

General Chair



Prof. Ranjana Jha
Hon'ble Vice Chancellor
IGDTUW, Delhi

TPC Chair



Prof. S.S. Agrawal
Emeritus Scientist
CSIR

Organizing Chair



Prof. Arun Sharma
Dean Academic Affairs
IGDTUW

Organizing Co-Chair



Prof. R.K. Singh
Registrar
IGDTUW



6th International Conference on Artificial Intelligence and Speech Technology
AIST 2024

13th November – 14th November, 2024

KEY NOTE SPEAKERS



Prof. Bjoern Schuller
Faculty of Engineering
Imperial College London,
England



Dr. Omid Mahdi Ebadati
Kharazmi University
Tehran



Dr. Samiya Khan
University of Greenwich
United Kingdom



Prof. Subramaniam Ganesan
Oakland University
USA



Prof. Fayad, Mohamed E
San Jose State University
USA



6th International Conference on Artificial Intelligence and Speech Technology AIST 2024

13th November – 14th November, 2024

Message from the General Chair

I am delighted to welcome all participants, delegates, Keynote Speakers, and Resource Persons to the 6th International Conference on “Artificial Intelligence and Speech Technology” (AIST-2024), taking place at Indira Gandhi Delhi Technical University for Women, Kashmere Gate, Delhi, on November 13-14, 2024.

Building on the success of the previous five conferences, we are excited to see an even greater number of high-quality submissions from academicians, professionals, and researchers worldwide.

This year, AIST-2024 provides an excellent opportunity to engage with renowned international Keynote speakers and learn from their expertise.

The conference aims to serve as a platform for discussing the latest research, developments, and applications in the fields of Artificial Intelligence and Speech Technology. AIST-2024 will spotlight cutting-edge topics such as Machine Learning, Deep Learning, Speech Synthesis, and Speech Recognition, encouraging innovative exploration and collaboration.

We extend our heartfelt gratitude to our distinguished keynote speakers for their willingness to share their knowledge with us. We believe that this gathering of experts from academia and industry will greatly benefit researchers, students, and faculty alike. Young scientists and researchers will find valuable insights to help shape their future endeavors.

I wish the conference great success!



Prof. Ranjana Jha

Hon'ble Vice Chancellor, IGDTUW
General Chair, AIST 2024



6th International Conference on Artificial Intelligence and Speech Technology AIST 2024

13th November – 14th November, 2024

Message from the Technical Program Committee Chair

It is a great pleasure to acknowledge that Indira Gandhi Delhi Technical University for Women, Delhi, is hosting the 6th International Conference on Artificial Intelligence and Speech Technology (AIST-2024). This conference aims to provide a platform for in-depth discussions and presentations on state-of-the-art research, development, innovations, and implementations in Artificial Intelligence and Speech Technology by researchers worldwide.



Today, AI significantly enhances machine learning, Natural Language Processing (NLP), and Deep Learning, leading to innovations in Speech Technology, such as voice-responsive interactive systems, smartphones, and smart home devices. Given this potential, AIST-2024 brings together academics, industry experts, and educational leaders to discuss a wide array of topics, including the foundations of Artificial Intelligence, machine learning, data mining, cognitive science, and Speech Technology.

I would like to express my sincere gratitude to the esteemed professors and prominent researchers who will deliver keynote addresses and share their insights during the conference. Special thanks are also due to the organizing committee, including the co-chairs, technical program committee members, and paper reviewers, for their invaluable work in shaping the program. Additionally, I appreciate all authors who submitted their papers to AIST-2024; I am pleased to note that accepted papers will be submitted to Springer for publication, which is indexed by SCOPUS.

In summary, I am confident that you will find the combination of a cutting-edge technical program and excellent organization to be rewarding. Enjoy reconnecting with colleagues and engaging in meaningful discussions with our distinguished speakers. I look forward to seeing everyone at IGDTUW, Delhi, India!

Prof. S.S. Agrawal
Emeritus Scientist, CSIR
Technical Program Committee Chair, AIST 2024



6th International Conference on Artificial Intelligence and Speech Technology AIST 2024

13th November – 14th November, 2024

Message from the Organizing Chair

I take this opportunity to welcome you all to the 6th International Conference on Artificial Intelligence and Speech Technology (AIST-2024), to be held at Indira Gandhi Delhi Technical University for Women, Delhi, on November 13-14, 2024.

Topics covered in this conference will include the fundamentals of AI, its tools and applications, Machine Learning, Deep Learning, Speech Analysis, Representation and Models, Spoken Language Recognition and Understanding, Affective Speech Recognition, as well as various applications such as NLP, Computer Vision, Robotics, and Medical technologies. We have received over 400 submissions from around the globe, and the best 55 selected papers will be presented over the course of the two days. The Conference Proceedings will be published by Springer.



AIST-2024 will feature prominent keynote addresses from distinguished speakers, including Professor Bjoern Schuller, Imperial College London, Dr. Samiya Khan, University of Greenwich, United Kingdom, Prof. Subramaniam Ganesan, Oakland University, USA, Dr. Omid Mahdi Ebadati, Kharazmi University, South Mofatteh Ave., Tehran and Prof. Fayad, Mohamed E, San Jose State University, USA.

I would like to extend my gratitude to everyone who has contributed their time, energy, and ideas to help organize this event, including all members of the organizing committee, the Technical Program Committee, the reviewers, and our distinguished keynote speakers. I also wish to thank our sponsors and supporters, especially American Express and DST, whose contributions have made this event possible. It is through the collective efforts of these individuals and organizations that we are able to host this significant conference. I look forward to the great success of AIST- 2024!

Prof. Arun Sharma

Professor - IT & Dean Academic Affairs, IGDTUW
Organizing Chair, AIST 2024



6th International Conference on Artificial Intelligence and Speech Technology AIST 2024

13th November – 14th November, 2024

Message from the Chief Guest

It is a tremendous honor to participate in the 6th International Conference on Artificial Intelligence and Speech Technology organized by IGDTUW. This conference stands as a testament to the remarkable advancements in technology and the endless possibilities that lie ahead. The convergence of Artificial Intelligence and Speech Technology signifies a paradigm shift in operations, interactions, and problem-solving through the power of machines.



Scientists, researchers, and innovators are at the forefront of shaping a future where intelligent systems collaborate seamlessly with human capabilities, leading to groundbreaking discoveries, improved efficiencies, and new innovations.

The applications of Artificial Intelligence offer numerous opportunities for advancing defense technologies, with ongoing developments not only in functionality but also in safety. As the conference unfolds with presentations, discussions, and exchanges of ideas about current work, results, and achievements, it is crucial to remain focused on the possibilities of tomorrow. Additionally, the ethical implications of AI must be carefully considered to ensure that innovations are channeled for the betterment of humanity.

I encourage all participants to share new ideas, debate the best paths forward, forge new collaborations, and create an environment where innovative products can take shape. May the ideas exchanged during this conference plant new seeds of knowledge that will grow and flourish. I eagerly anticipate the insights that will emerge from our discussions and the impact they will have on shaping the future of AI and Speech Technology.

Bestwishes to all participants and organizers for a successful conference!

Prof. Lalit K. Awasthi
Hon'ble Vice Chancellor
Sardar Patel University, Mandi

6th International Conference on Artificial Intelligence and Speech Technology AIST 2024



13th November – 14th November, 2024

Innaugration Schedule 13th November, 2024

Time	Schedule
10:00 AM – 10:05 AM	Kulgeet of the University
10:05 AM – 10:10 AM	Lighting of Lamp
10:10 AM – 10:15 AM	Welcome of Guests
10:15 AM – 10:20 AM	Welcome Address by Organizing Co-Chair, AIST-2024 and Registrar, IGDTUW
10:20 AM – 10:25 AM	Highlights of Technical Program by Technical Program Chair, AIST-2024
10:25 AM – 10:30 AM	Presidential Keynote Address by General Chair, AIST-2024 and Hon'ble VC, IGDTUW
10:30 AM – 10:35 AM	Address by the Guest of Honor
10:35 AM – 10:45 AM	Address by the Chief Guest
10:45 AM – 10:47 AM	Release of Conference Souvenir
10:47 AM – 10:50 AM	Presentation of Mementoes to the Guests
10:50 AM – 10:55 AM	Vote of Thanks by Organizing Chair, AIST-2024
10:55 AM – 11:00 AM	National Anthem
11:00 AM – 11:45 AM	High Tea

Financially Supported by



Publication Partner



Indira Gandhi Delhi Technical University for Women
(Established by Govt. of Delhi vide Act 09 of 2012)
NAAC A+ Accredited Kashmere Gate, Delhi-110006
Women Education | Women Enlightenment | Women Empowerment

6th International Conference on Artificial Intelligence and Speech Technology AIST 2024



13th November – 14th November, 2024

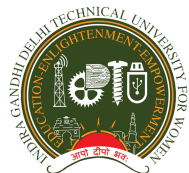
Master Program Schedule

Date: 13 th November, 2024		
10:00 AM - 11:00 AM	Inauguration Session Auditorium	
11:00 AM - 11:45 AM	High Tea	
11:45 AM - 1:15 PM	Technical Session 1 Seminar Hall - Offline	Session Chair: Dr. Devender Kumar, NSUT Session Co Chair: Dr. Deepak Kumar Sharma, IGDTUW
11:45 AM - 1:15 PM	Technical Session 2 Conference Hall - Offline	Session Chair: Dr. S. S. Agrawal, GGSIPU Session Co Chair: Dr. Shweta Jindal, IGDTUW
1:15 PM – 2:00 PM	Lunch	
02:00 PM - 03:15 PM	Technical Session 3 Conference Hall - Offline	Session Chair: Dr. Dinesh Vishwakarma, DTU Session Co Chair: Dr. Ritu Rani, BPIT
02:00 PM - 03:15 PM	Technical Session 4 Seminar Hall - Offline	Session Chair: Dr. Monica Mehrotra, JMI Session Co Chair: Dr. Alongbar Wary, IGDTUW
3:15 PM – 4:00 PM	Keynote Address 1 Conference Hall-Online	Speaker: Dr. Samiya Khan, Univ. of Greenwich, UK Session Chair: Dr. Ritika Kumari, IGDTUW
04:00 PM– 04:45 PM	Keynote Address 2 Seminar Hall - Online	Speaker: Dr. Subramaniam Ganesan, Oakland University, USA. Session Chair: Dr. Kamal Kumar, IGDTUW
Date: 14 th November, 2024		
10:30 AM –11:15 AM	Keynote Address 3: Conference Hall-Online	Speaker: Dr. Bjoern Schuller, Imperial College London Session Chair: Dr. Arun Solanki, GBU/Ms. Karuna Kadian, IGDTUW
11:15 AM - 12:45 PM	Technical Session 5 Seminar Hall - Online	Session Chair: Dr. Manju Khari, JNU Session Co-Chair: Dr. Mohona Ghosh, IGDTUW
11:15 AM - 12:45 PM	Technical Session 6 Conference Hall-Online	Session Chair: Dr. Pradeep Tomar, GBU Session Co Chair: Dr. Himanshu Mittal, IGDTUW
12:45 PM – 01:30 PM	Keynote Address 4: Seminar Hall - Online	Speaker: Dr. Omid Mahdi Ebadati, Kharazmi University, South Mofatteh Ave., Tehran Session Chair: Ms. Bhawna Jain, IGDTUW
1:30 PM – 2:30 PM	Lunch	
02:30 – 04:00 PM	Technical Session 7 Seminar Hall - Online	Session Chair: Dr. Garima Jaiswal, Bennett University Session Co Chair: Mr. Rahul Sachdeva, IGDTUW Dr. Ritu Rani, BPIT
4:00 PM - 4:30 PM	Valedictory Session	

Financially Supported by



Publication Partner



Indira Gandhi Delhi Technical University for Women
(Established by Govt. of Delhi vide Act 09 of 2012)
NAAC A+ Accredited Kashmere Gate, Delhi-110006
Women Education | Women Enlightenment | Women Empowerment

6th International Conference on Artificial Intelligence and Speech Technology
AIST 2024



13th November – 14th November, 2024

Technical Session: 01

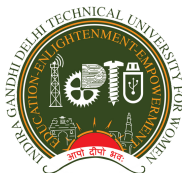
Date: 13th November, 2024 (Offline Mode) Time: 11:45 AM – 01:15 PM

S No	Paper ID	Paper Title	Authors
1	61	Analysis of Online Hate Speech Detection for Multilingual and Multimodal Data using Artificial Intelligence	Navya, Anjum, Neha Rani, and Riya Singh
2	63	DeepFake Classification using Fine-Tuned Wave2Vec2.0	Subham Sarkar, Ankita Ghosh, Anshul Gupta
3	80	Emotional Speech Corpus: A Review	Krishna Rohilla, Aaditya Sharma, Nilakshi Srivastav, Nivedita Palia
4	141	Speech Summarization Diverse Challenges and Advancements	Riya Ahlawat, Samriddhi Tiwari, Muskan Singh, Shweta Jindal
5	142	Identifying Disaster-Related Tweets using Natural Language Processing and Core Machine Learning	Aditi Mehta, Arshiya Garg, Ritu Rani, Garima Jaiswal, Arun Sharma
6	202	Machine Learning and Neural Networks for Enhanced Autism Spectrum Detection ASD Detection with ABIDE preprocessed Dataset	Gaytri Mohapatra, Ritu Rani
7	247	Translating Heritage: A Transformer-Oriented Method for Neural Machine Translation from Sanskrit to Hindi	Nandini Sethi, Amita Dev, Poonam Bansal
8	277	Advanced Speech Recognition: Techniques, Challenges, and Applications	Laxmi Ahuja, Saurav Vinod Nair, and Ayush Thakur

Financially Supported by



Publication Partner



Indira Gandhi Delhi Technical University for Women

(Established by Govt. of Delhi vide Act 09 of 2012)

NAAC A+ Accredited Kashmere Gate, Delhi-110006

Women Education | Women Enlightenment | Women Empowerment

6th International Conference on Artificial Intelligence and Speech Technology
AIST 2024



13th November – 14th November, 2024

Technical Session: 02

Date: 13th November, 2024 (Offline Mode) Time: 11:45 AM – 01:15 PM

S No	Paper ID	Paper Title	Authors
1	69	Rule-based Named Entity Recognition for Hindi	Devanshi Suri, Nimisha Malviya, Nisheeth Joshi
2	81	Investigating AI Explanations in Medical Diagnosis	Nikita Garg, Manya Joshi, Nidhi Goel, Palak Handa
3	116	A Metric-Driven Comparative Study of Text Summarization Model: Insights from State-of-the-art LLMs	Shivang Gaur, Aman Karunik, Swapnali Naik
4	134	Automatic Assessment of Program Code using CodeBERTScore: A Transformer-based Approach	Hunny Gaur, Devendra K Tayal, Amita Jain
5	139	Survey on the Availability of Datasets for Machine Translation Systems in Indian Languages	Ayushi, Amita Dev, Poonam Bansal
6	267	SHiTraD: Sanskrit-Hindi Translation Dataset	Nandini Sethi, Ayushi, Amita Dev, Poonam Bansal
7	292	Emotional Arousal of Female Politicians: Distinguishing Enthusiasm and Anger Using Data Augmentation and Machine Learning	Hilla Goren Barnea, Vered Silber-Varod, Elishai Ezra Tzur

Financially Supported by



Publication Partner



Indira Gandhi Delhi Technical University for Women

(Established by Govt. of Delhi vide Act 09 of 2012)

NAAC A+ Accredited Kashmere Gate, Delhi-110006

Women Education | Women Enlightenment | Women Empowerment

6th International Conference on Artificial Intelligence and Speech Technology
AIST 2024



13th November – 14th November, 2024

Technical Session: 03

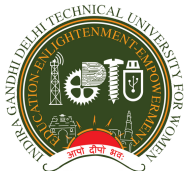
Date: 13th November 2024 (Offline Mode) Time: 02:00 - 03:15 PM

S No	Paper ID	Paper Title	Authors
1	56	Harnessing Machine Learning for Feature Extraction in Plant Imaging and Analysis: A Review	Anshika Jain, Akansha Khandelwal, Vandana Singh, Astha Sharma, Ashwni Kumar
2	70	Deepfake Detection: A Comprehensive Analysis of Modern Techniques	Deepanshi Jalan, Shradha Jain, Anandita Tuli, Vanshika Chaudhary, Astha Sharma, Ashwani Kumar
3	71	Transfer Learning for Leaf Disease Image Classification via CNN Model	Kirti Dabas, Anuradha Chug, Amit Prakash Singh, Ravinder Kumar, Veerubommu Shanumugam
4	82	Assessing ChatGPT in Medical Domain	Aikanshi Rohilla, Palak Handa , Nidhi Goel
5	138	Comparative Analysis of Deep Learning Techniques on Tomato Plant Disease Detection	Riya Sharma, Ayushi Pasrija, Saloni Gupta, Poonam Bansal
6	187	Recent Advances in Denoising Techniques for Hyperspectral Image Enhancement	Priyanka Gupta, Garima Jaiswal, Ashish Kumar
7	305	Predictive Models in Biomedical Applications: A Machine Learning Approach	Kanika Wadhwa, D. Ananya, Himanshu Mittal, Arun Sharma, Swaty Wadhwa
8	386	Echoes of Concern: Security Challenges in Voice Assistant Systems	Rakesh Kumar Singh, Charu Gupta

Financially Supported by



Publication Partner



Indira Gandhi Delhi Technical University for Women

(Established by Govt. of Delhi vide Act 09 of 2012)

NAAC A+ Accredited Kashmere Gate, Delhi-110006

Women Education | Women Enlightenment | Women Empowerment

6th International Conference on Artificial Intelligence and Speech Technology
AIST 2024



13th November – 14th November, 2024

Technical Session: 04

Date: 13th November 2024 (Offline Mode) Time: 02:00 - 03:15 PM

S No	Paper ID	Paper Title	Paper Title
1	78	Real-Time Detection of Household Objects Using Single-Shot Detection with Mobile Net	Pallavi Sharma, Isha Malhotra, Palak Handa, Nidhi Goel
2	90	CTC-Based Automatic Speech Recognition for Hindi language	Arushree Mishra, Suhani, Amita Dev, Poonam Bansal
3	225	Bridging the Gap: Real-Time ASL Fingerspelling to Sentence Translation	Manya Chandna, Abantika Dasgupta, Bhumika Gupta, Ravinder M, S.R.N. Reddy, Rishika Anand
4	244	Dictionary-Based Lexical Resource for Somali Homograph Disambiguation	Abdullahi Mohamed Jibril, Abdisalam Mahamed Badel
5	283	Comparative Evaluation of 3D U-Net and SegResNet Architectures for Brain Tumor Segmentation Using Adam and Ranger21 Optimizers	Nisha Purohit, C P Bhatt, Subodh Kumar
6	289	Advancements in Alzheimer's Disease Detection: A Comprehensive Review of Deep Learning Approaches in MRI Imaging	Khushi Jindal, Seeja K.R, D.K. Tayal
7	306	Image Captioning Using Deep Learning Models: A Comprehensive Overview	Nisha, Shailesh D Kamble, Himanshu Mittal

Financially Supported by



Publication Partner



Indira Gandhi Delhi Technical University for Women

(Established by Govt. of Delhi vide Act 09 of 2012)

NAAC A+ Accredited Kashmere Gate, Delhi-110006

Women Education | Women Enlightenment | Women Empowerment

6th International Conference on Artificial Intelligence and Speech Technology

AIST 2024

13th November – 14th November, 2024



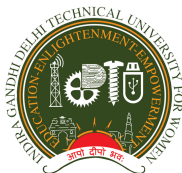
Technical Session: 05

Date: 14th November, 2024 (Online Mode) Time: 11:15 AM – 12:45 PM

S No	Paper ID	Paper Title	Authors
1	45	Speech Emotion Classification with Acoustic Features using 1D CNN and LSTM Networks	Yesha Vyas, Shofiya Bootwala, Neha Soni
2	62	Unveiling Voices: Deep Learning Based Noise Dissipation with Real-Time Multi-Speech Separation and Speaker Recognition	Chaitra V, Meera Reji, Meghana Aithal, Varshini Gopal, Shylaja S S
3	74	Identification of Speaker-Specific Features to Minimize the Mismatch Outcomes for Speaker Recognition using Anger and Happy Emotional Speech	Shalini Tomar, Shashidhar G. Koolagudi
4	89	Bilingual Speech Translation Between Tamil and Telugu with Elimination of Dependency on Any Intermediate Languages	Kaarthik R, Rithuh Subhakkirith S, Yaraswini Dharmavarapu, Vignesh Kumar A, Dhanya M Dhanalakshmy, Karthikeyan N G
5	92	Turbocharging Pull Request Reviews: Exploring Generative AI for Code Review	Sourav Datta, Anjali Sharma, Jaymin Rajgor
6	114	NITK-TIEKLS: A Text-Independent Emotional Kannada Language Speech Dataset for Speaker Recognition	Shalini Tomar, Shashidhar G. Koolagudi
7	117	Audioverse: Personalized Story-telling Using Voice Cloning	Kasey Ann Britto, Harini J, Shivanii M, Pranavi Karthik, Kalpana B N, Karthi M
8	155	Customer Churn Prediction Using Artificial Neural Networks	Manish Saini, Saurav Singha, Brijesh Kalakoti, Hashmat Fida

Financially Supported by

Publication Partner



Indira Gandhi Delhi Technical University for Women

(Established by Govt. of Delhi vide Act 09 of 2012)

NAAC A+ Accredited Kashmere Gate, Delhi-110006

Women Education | Women Enlightenment | Women Empowerment

6th International Conference on Artificial Intelligence and Speech Technology
AIST 2024



13th November – 14th November, 2024

Technical Session: 06

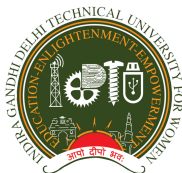
Date: 14th November, 2024 (Online Mode) Time: 11:15 AM – 12:45 PM

S No	Paper ID	Paper Title	Author
1	84	Human Evaluation of Some Machine Translation Toolkits for English-Hindi Language Pair	Nisheeth Joshi, Riya Lonchenpa, Anju Krishnia, Pragya Katyayan, Mhasilenuo Vizo
2	135	Optimizing Speaker Recognition through Feature Extraction Techniques: A Focus on Gujarati Dialects	Meera M. Shah, Hiren R. Kavathiya
3	222	A.R.I.S.E. Artificial Recognition and Intelligent Sign-language Eyewear - A Multifunctional Smart Glasses System for Enhanced Accessibility and User Assistance	Roshi Bhati, Rahul Thambi, Nikhil Nerurkar, Ketan Shah
4	294	Comparing Audio Speech Recognition and Video Speech Recognition using Neural Networks	Atharva Swami, Neetigya Bisen, Darshak Savant
5	335	NLP-Driven Sentiment Analysis for Fake News Identification: The Responsive Truth	Arpita Singh, Arpit Dubey, Rabins Porwal
6	344	Sign Language to Speech Conversion in Regional Language (Tamil)	Anitha Julian, Pavan Pavan Kumar A B, Janarthanan S, Priyanka S
7	385	A Framework Using Feature Maps Improving Speech Delivery In Preschoolers	Arindam Bindlish, Jitendra Singh Goyal, Vaibhav Gujral, Tanu Dua, Gaurav Dubey

Financially Supported by



Publication Partner



Indira Gandhi Delhi Technical University for Women

(Established by Govt. of Delhi vide Act 09 of 2012)

NAAC A+ Accredited Kashmere Gate, Delhi-110006

Women Education | Women Enlightenment | Women Empowerment

6th International Conference on Artificial Intelligence and Speech Technology
AIST 2024



13th November – 14th November, 2024

Technical Session: 07

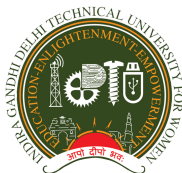
Date: 14th November, 2024 (Online Mode) Time: 02:30 PM - 04:00 PM

S No	Paper ID	Author	Paper Title
1	32	Exploring State-of-the-Art approaches for Heart Disease Detection: A Detailed Analysis	B. Shamna, C. P. Maheswaran
2	33	A Comprehensive Analysis of Recent Methodologies for Irregular Heart Beat Prediction	K. Rema, C.P. Maheswaran
3	49	Next-Gen Imaging: The Power of Hyperspectral Data and Autoencoders	Pallavi Ranjan, Sushma Hans; Salih Ismail
4	50	Redefining Vision Tasks: The Power of Transformers in Classification, Detection, and Segmentation	Sushma Hans, Pallavi Ranjan, Salih Ismail
5	94	Leveraging Depth Data and Parameter Sharing in Vision Transformers for Improved Face Anti-Spoofing	Aashania Antil, Chhavi Dhiman
6	103	Comparative Analysis of Multiple Embedding Models for Text Based Document Similarity	Shreya Khosla, Shresth Jain, Anupama MA, Rama Subba Reddy Thavva
7	107	Electricity Demand Prediction using Machine Learning	Ali Sadiq Khawja
8	156	AI-Driven Essay Evaluation: Enhancing Accuracy and Fairness in Automated Scoring Systems	Kuldeep Kumar Singh, Gaurav Gupta, Sapna Chaudhary, Sajal Singh Masand, Swati Saran, Shivani Sharma
9	268	CodeBERT-BiGRU for Software Defect Prediction	Priya Singh, Ruchika Malhotra
10	280	AccuRep: A Real-Time Android Application for Exercise Tracking Using MoveNet Pose Detection and Joint Angle Analysis	Atharva Swami, Ishi Prashar, Tanmay Sahasrabudhe, Preeti Agarwal

Financially Supported by



Publication Partner



Indira Gandhi Delhi Technical University for Women

(Established by Govt. of Delhi vide Act 09 of 2012)

NAAC A+ Accredited Kashmere Gate, Delhi-110006

Women Education | Women Enlightenment | Women Empowerment

ABSTRACTS**Speech Emotion Classification with Acoustic Features using 1D CNN and LSTM Networks**

Yesha Vyas, Shofiya Bootwala, and Neha Soni

Abstract. Speech emotion recognition, with its wide range of applications in robotics, customer service, and HCI (human- computer interaction), is being continuously improved using new models, features, and feature extraction mechanisms. This study proposes 3 deep learning models: 1D CNN, 1D CNN combined with LSTM, and 1D CNN combined with BiLSTM, these models exhibit remarkable performance on a merged dataset comprising RAVDESS, TESS, CREMA-D, and SAVEE emotional speech corpora. Furthermore, they demonstrate strong performance when evaluated on individual databases. For reference, experiments using conventional machine learning algorithms, such as SVM, logistic regression, and ensemble methods are also performed on the datasets. The proposed deep learning models leverage a fusion of acoustic features including MFCC, RMSE, and ZCR, for training. Augmentation techniques like noise addition and pitch scaling were employed to enhance the models' robustness. Notably, the proposed 1D CNN model achieves an impressive accuracy of 96% on the combined datasets, outperforming prior studies on merged datasets.

Keywords: Speech emotion recognition, Human computer interaction, Convolutional neural networks, Long-Short Term Memory, Emotion recognition

Audioverse: Personalized Story-telling Using Voice CloningKasey Ann Britto¹, Harini J, Shivanii M, Pranavi Karthik, Kalpana B N, and Karthi M

Abstract. The surge in global podcast and story listeners, now totaling 504.9 million and comprising 23.5% of internet users, highlights the growing demand for advanced narrative tools. This study explores the transformative potential of voice in storytelling by developing a deep learning model designed to generate diverse narratives based on user prompts. Utilizing advanced voice cloning techniques, the model aims to embed nuanced emotions within these narratives, merging human emotion with technological capabilities to create a sustainable tool for enhancing individual well-being. The model was constructed using PyTorch, TensorFlow, and Convolutional Neural Networks (CNN), with voice cloning powered by BARK, an open-source transformer-based model. Audioverse promises to significantly enrich human connections and revolutionize the storytelling landscape.

Keywords: Audioverse · Deep Learning · Story Generator· Pytorch· CNN· Tensorflow· Genre· Voice Cloning · google/flan-t5-base.

Unveiling Voices: Deep Learning Based Noise Dissipation with Real-Time Multi-Speech Separation and Speaker Recognition

Chaitra V, Meera Reji, MeghanaAithal, Varshini Gopal, and Shylaja S S

Abstract. Humans have a remarkable ability to focus on speech even in noisy environments. To replicate this capability computationally, recent research has developed advanced speech enhancement and separation algorithms. This paper presents a comprehensive pipeline integrating several top-notch models for efficient background noise removal, speaker separation, and recognition in audio recordings. Furthermore, our approach exhibits remarkable proficiency in speaker diarization and identification, facilitated by sophisticated feature extraction and classification techniques. Our results demonstrate substantial improvements in audio clarity, speaker separation, and accurate speaker identification. This enhances the capability of our approach for real-time applications in telecommunications, hearing aids, and other audio processing domains. The effectiveness of our method provides a strong proof-of-concept for its real-world applicability, showcasing its potential to revolutionize acoustic applications.

Keywords: Audio classification, Background noise extraction, Speech enhancement, Speaker separation, Speaker Recognition

Identification of Speaker-Specific Features to Minimize the Mismatch Outcomes for Speaker Recognition using Anger and Happy Emotional Speech

Shalini Tomar and Shashidhar G. Koolagudi

Abstract. A vital component of digital speech processing is Speaker Recognition (SR). However, variation in speakers' emotional states, such as happiness, anger, sadness, or fear, poses a significant challenge that compromises the robustness of speaker recognition systems. It appears to be challenging to distinguish between emotions like "anger" and "happy," according to research on SR using emotive speech. The study looks at prosody-related speech characteristics to determine how to distinguish between "anger" and "happy" emotional speech for SR tasks. The goal is to explore speaker-specific features. The experiment outcomes demonstrate that, as speaker-specific features for the SR task, Intensity, Pitch, and Brightness (IPB) variables can distinguish between angry and happy emotional speech. Combining IPB and MFCC (IPBCC) feature extraction with the Hybrid CNN-LSTM combined with an attention mechanism approach achieves an SR accuracy of 95.45% for anger and 96.22% for happy emotional speech.

Keywords: Speaker Recognition using Emotional Speech, Intensity, Pitch, Brightness

NITK-TIEKLS: A Text-Independent Emotional Kannada Language Speech Dataset for Speaker Recognition

Shalini Tomar and Shashidhar G. Koolagudi

Abstract. Speaker recognition systems have traditionally relied on the consistency of speech content to identify individuals. However, text independent speaker recognition, irrespective of the spoken content, presents a more flexible and robust alternative, especially in real-world scenarios. This research focuses on enhancing text-independent speaker recognition by incorporating a dataset for the Speaker Recognition (SR) task. The dataset is named the National Institute of Technology Karnataka - Text-Independent Emotional Kannada Language Speech (NITKTIEKLS) dataset. The 200 natives of the Karnataka state of India have recorded emotional speech in the Kannada language for the proposed dataset. The neutral text-independent speech consists of a 4-minute speech duration for each speaker. The two emotional speech utterances, from any two of the emotions anger, happiness, sadness, and fear, are text-independent speech utterances that consist of 2 minutes. The total duration is approximately 30 hours. The proposed study includes developing, processing, analyzing, acquiring, and evaluating the proposed dataset. The suggested dataset consists of performance evaluations of the SR system through deep learning techniques with the proposed Wavelet Mel Spectrogram.

Keywords: Speaker Recognition, Text-independent emotional speech, Speaker Recognition in Emotional Environment, Mel Spectrogram, Zero Crossing Rate, Pitch, Tempo, Wavelet Spectrogram

Identifying Disaster-Related Tweets using Natural Language Processing and Core Machine Learning Algorithms

Aditi Mehta, ArshiyaGarg, Ritu Rani, GarimaJaiswal, Arun Sharma

Abstract. In today's digital era, social media platforms like Twitter are used by people across the globe to share information and updates, such platforms serve as crucial means of communication especially during emergencies. However, the most prominent challenge faced by authorities is the large volume of data and the consequent complexity of extracting crucial information from it. This research aims at finding an optimal solution to tackle the stated problem. The authors used machine learning approaches, such as Naïve Bayes and Support Vector Machines (SVM), to identify tweets that needed to be addressed right away. The models used for final evaluation were trained using a dataset of about 8,000 tweets out of which just 3,229 were analyzed to be disaster related.

The goal is to accelerate information flow and enable authorities to react to urgent needs more rapidly and efficiently by precisely recognizing these tweets. This method helps with better resource distribution and faster action during emergencies by not just speeding up disaster response but also ensuring that important information isn't lost in the shuffle. The results of this research indicate that SVM (Support Vector Machine) performs better, which makes it a preferable option for catastrophe prediction based on tweets.

Keywords: Natural Language Processing, disaster, twitter, SVM, tweet classification, social media analysis

Analysis of Online Hate Speech Detection for Multilingual and Multimodal Data using Artificial Intelligence

Navya, Anjum, Neha Rani, and Riya Singh

Abstract. Online hate speech (OHS) on social networking sites such as Facebook, Instagram, Twitter(now X), WhatsApp, Reddit,etc. has emerged as a growing concern due to its negative psychological effects on individuals as well as communities. This issue is amplified by the enhancement of voices, misinformation, anonymity, and the lack of accountability on social media, where users may not face immediate or tangible consequences for their words. Recognising the detrimental impact of OHS on individual well-being, there is growing interest in developing effective mechanisms for its detection and mitigation. This study presents a comprehensive survey of existing literature on OHS detection for multimodal and multilingual data

available on social networking sites. The analysis involves examining available datasets and reviewing 23 most recent research papers published between year 2020 and 2024. The surveyed papers are categorised based on the datasets and approaches used such as traditional machine learning algorithms, deep learning architectures and other approaches like rule-based and hybrid methods are also explored.

Keywords: social media, hate speech, dataset, techniques, multilingual, multimodal, online hate speech (OHS)

Emotional Speech Corpus: A Review

Krishna Rohilla, Aaditya Sharma, Nilakshi and NiveditaPalia

Abstract: Speech Emotion Recognition is an emerging area within speech processing. It enhances human machine interaction by empowering systems to detect and responds according to human emotion. Despite significant development in speech recognition, the ability of computers to recognize the emotion through speech signal remains a challenging task due to varying and dynamic nature of emotion. Extracting emotional content from the speech sample a benchmark dataset is required. This paper provides a comprehensive review of thirty different Emotional Speech Databases. Out of these 30 datasets 16 dataset developed for Indian languages. Each corpus contains speech samples collected under different emotional conditions. A brief description of each corpus and its area of application is discussed. During this study, we discovered that most common emotion recognized are sad, anger, neutral, fear, happy, surprise and disgust in decreasing frequency of appearance.

Keywords: Speech Emotion Recognition, Speech Emotion Databases

Bilingual Speech Translation Between Tamil and Telugu with Elimination of Dependency on Any Intermediate Languages

Kaarthik R, Rithuh Subhakkirith S, Yasaswini Dharmavarapu, Vignesh Kumar A, Dhanya M Dhanalakshmy, and Karthikeyan N G

Abstract. Usage of English as an intermediate language in translating one language to another is prominent, in most of the existing works. Many languages share a common parent language, and when such languages use English as an intermediate language for translation, there can be contextual differences, which could be avoided if there existed a direct source language to target language translation, instead of an external language dependency. Translation of Tamil speech to Telugu speech is discussed here. There are two main approaches: direct translation approach, and cascaded translation approach. A direct Speech-to-Speech Translation approach using discrete units is compared with a cascaded speech translation system approach that integrates Automatic Speech Recognition, Machine Translation, and Text-to-Speech modules. Several combinations of the cascaded translation approach are compared with the direct Speech-to-Speech Translation approach by training and testing with Few-shot Learning Evaluation of Universal Representations of Speech (FLEURS) dataset. Bilingual Evaluation Understudy (BLEU) and Perceptual Evaluation of

Speech Quality (PESQ) evaluatory scores show that direct Speech-to-Speech Translation is better than some of the cascaded models in the quality of speech translated, but all cascaded Speech-to-Speech Translation approaches are better than the direct Speech-to-Speech Translation approach in preserving the context. It can be concluded that when the right mix of models is cascaded, the translation of Tamil to Telugu becomes better than using a direct Speech-to-Speech Translation.

Keywords: Natural Language Processing, Direct Speech to Unit Translation, HuBERT, VITS, WAV2VEC2, WHISPER, SEAMLESSM4T

Speech Summarization Diverse Challenges and Advancements

Riya Ahlawat, Samriddhi Tiwari, Muskan Singh, ShwetaJindal

Abstract. Speech Technology is increasingly popular due to the high demands of voice-activated systems and security purposes. As the volume of speech data grows, obtaining important information becomes increasingly difficult. There comes the need to generate information-rich summaries. Even with high-effort research, the results have become stagnant over the years. The current study aims to deeply analyze the shortcomings in the available speech summarization techniques. It was found that this could be due to the homogeneous datasets, unconstrained reference summaries, structural bias in the training data, limitations of the evaluation metrics, or the papers' limited discussion of implementation specifics. After analyzing the findings from several papers, it was found that these drawbacks could be addressed by 1) limiting the "inverted pyramid" structure in the training data, 2) introducing new bias-free datasets with constrained summaries across domains and languages, 3) implementing evaluation metrics that are in line with user preferences, or 4) closely examining implementation details. This would immensely help in choosing, building, or modifying speech summarization tools in the future.

Keywords: Speech Summarization, Large Language Models, Graphs-Oriented Approaches, Datasets, Evaluation Metrics, State-of-art

Optimizing Speaker Recognition through Feature Extraction Techniques: A Focus on Gujarati Dialects

Meera M. Shah and Hiren Kavathiya

Abstract. Speaker recognition has proved to be an important field of research in the context of automatic speech recognition systems contributing applications, ranging from security authentication to speech analysis. In this paper, optimizations of speaker recognition systems by advanced feature extraction techniques are discussed in particular with the Gujarati dialects. Gujarati is one of the dialects majorly spoken in western India. These dialects, which are difficult and distinguishable due to variations in phonetics, intonation, and articulation make this attempt distinctively challenging when implementing speaker recognition techniques. In this study, we would assess various feature extraction methods; this includes Mel-Frequency Cepstral Coefficients (MFCC) to Linear Predictive Coding (LPC) and Fast Fourier Transform (FFT) is compared in detail with Discrete Wavelet Transform (DWT), Perceptual Linear Prediction (PLP), and Relative Spectral (RASTA-PLP) methods and their impacts on speakers of those dialects. Accuracy results indicate a significant improvement when dialect-specific features are incorporated into the system, thus opening a new field toward more robust and adaptive speaker recognition systems. The results of this work will therefore be useful for more accurate and reliable construction of speaker recognition systems for Gujarati as well as other underrepresented languages. In bridging this gap in dialect-specific research in speaker recognition, it contributes to increasing the availability and accuracy of automatic speaker recognition systems across varying linguistic conditions.

Keywords: MFCC, Feature Extraction, Speaker Recognition, Gujarati Dataset, Voice Recognition, LPC

CTC-Based Automatic Speech Recognition for Hindi language

Arushree Mishra, Suhani, Amita Dev and Poonam Bansal

Abstract. Speech Recognition(SR) is a process in which spoken words or sentences are converted to text. In this process an acoustic signal, captured by a microphone , is converted to text [1]. The previous couple decades have seen such rapid developments in automatic speech recognition systems. They are being used in various industries such as healthcare, agriculture, speech dialling, directory services, education, and autos, among others, most commonly being virtual assistants like Siri, GoogleAssistant or Cortana. Even after so much advancements these ASR systemsare not commonly built for Indian regional language. Since, majority of the population of India speaks Hindi it becomes important for us to have an ASR framework for Hindi to improve the situation of the service sector [2].

This paper presents an innovative approach for 'Automatic Speech Recognition' that uses a modified

DeepSpeech2 model adapted for the Hindi language. The model integrates "Recurrent Neural Networks" (RNNs) and "Convolutional Neural Networks" (CNNs) for feature extraction with Gated Recurrent Units (GRUs) for sequence modelling. To further enhance performance, an attention mechanism is incorporated, enabling the model to concentrate on pertinent audio sequence segments. The purpose of this paper is to improve speech recognition performance for the Hindi language by utilising a modified DeepSpeech2 model along with some fine-tuning. The proposed method effectively models the Hindi language and reduces the training and validation loss with the help of efficient pre processing and fine tuning of the model [3].

Keywords: Automatic Speech Recognition, Hindi, RNN, CNN, CTC loss

Comparing Audio Speech Recognition and Video Speech Recognition using Neural Networks

Atharva Swami, Neetigya Bisen, and Darshak Savant

Abstract. The comparative analysis of recognition techniques using the GRID dataset enables us to evaluate the performance of models on visual lip movements and auditory signals. Therefore, we construct a common architecture for video and audio tasks. Similar model construction blocks are utilized to perform different tasks such as lip-reading from a sequence of video frames and speech recognition from audio waveforms using Mel-Frequency Cepstral Coefficients (MFCCs). In this study, we aim at determining how well these models can process the distinct characteristics of visual lip movements and auditory signals by training both with the GRID dataset. This approach also provides an opportunity to assess the effectiveness and adaptability of our models across various modalities directly. As such, this paper discusses how well the shared architecture deals with temporal and spatial characteristics, impacting advancements in video-based recognition, audio-based technologies, and analyzing cross-domain applications.

Keywords: Mel-Frequency Cepstral Coefficient, Lipreading, Convolution Neural Network, Long Short-Term Memory, Recognition Techniques, Unified Architecture, Performance Evaluation, Audio and Video based Recognition

Sign Language to Speech Conversion in Regional Language (Tamil)

Anitha Julian, Pavan Kumar A B, Janarthanan S and Priyanka S

Abstract. By employing hand gestures, individuals with hearing or speech disabilities are able to communicate their ideas without the need for spoken words. The work envisaged in this paper is to create an affordable yet real-time sign language to speech conversion system for the Tamil speaking regions. The system makes use of a web camera to capture hand gestures which are processed according to Media Pipe and then classified using Convolutional Neural Networks. The recognized gestures are converted to human voice in Tamil using gTTS API for Tamil language, thus providing instant voice output in Tamil language. The proposed system is highly beneficial for the hearing-impaired population in the Tamil-speaking areas and its accuracy stands remarkable at 92%. The system has been developed enough for 12 alphabets of the Tamil language which can as well be updated in the future to cover all the 247 alphabets.

Keywords: Sign language, speech conversion, Tamil language, gTTS, hand gestures, regional languages, CNN

A Framework Using Feature Maps Improving Speech Delivery In Preschoolers

Arindam Bindlish, Jitendra Singh Goyal, Vaibhav Gujral, Tanu Dua, and Gaurav Dubey

Abstract. Proper speech helps preschoolers' cognitive and social development. A hybrid model including Wav2Vec 2.0 for audio feature extraction, Bi-LSTM, and an attention mechanism improves preschoolers' speech competency categorization. LSTM, Bi-LSTMwith Attention Mechanism, and Proposed Bi-LSTMwere examined. All tests confirmthe Proposed model's98.26% accuracy, precision, recall, and F1 score. Confusion matrix: Ordinary 98.55%, Incompetent 100%, Proficient 96.0%. The recommended model has a 0.98 ROC AUC, compared to 0.95 for Bi-LSTM and 0.90 for LSTM, suggesting excellent class difference. Adding attention processes to Bi-LSTM enhances toddler linguistic proficiency. For the Proposed approach, speechdelivery classification efficiency increases early childhood speech therapy.

Keywords: Wave2Vec 2.0, Bi-LSTM,Attention Mechanism, Preschoolers, Speech

Advanced Speech Recognition: Techniques, Challenges, and Applications

Laxmi Ahuja, Saurav Vinod Nair, and Ayush Thakur

Abstract. This paper provides a comprehensive overview of modern speech recognition technology, focusing on advanced techniques, current challenges, and diverse applications. We explore the fundamental principles of Automatic Speech Recognition (ASR) systems, including acoustic modeling, language modeling, and decoding algorithms. The paper delves into cutting-edge approaches such as deep learning-based models, attention mechanisms, and transfer learning techniques that have significantly improved ASR performance. We discuss persistent challenges in the field, including robustness to noise, handling low-resource languages, and real-time processing. The paper also examines the wide-ranging applications of speech recognition across various domains, from voice assistants and automotive systems to healthcare and education. Finally, we consider the societal impact of this technology and potential future directions for research and development in speech recognition.

Keywords: Automatic Speech Recognition, Deep Learning, Natural Language Processing, Voice User Interfaces, Acoustic Modeling

Machine Learning and Neural Networks for Enhanced Autism Spectrum Detection ASD Detection with ABIDE preprocessed Dataset

Gaytri Mohapatra, Ritu Rani, Garima Jaiswal, Arun Sharma and Nidhi Bisla

Abstract. The goal of the study is to use deep learning and machine learning models to enhance early identification and diagnosis of autism spectrum disorder (ASD). The ABIDE preprocessed dataset is used in this study, along with other classification approaches such as Decision Trees, Random Forest, Support Vector Machines, K-Nearest Neighbors, and Feedforward Neural Networks. A stacked ensemble model was designed to combine the capabilities of separate classifiers achieving an accuracy of 99.77 %. The study discovered that deep learning captures complex patterns in neuroimaging data, implying that hybrid models can make a major contribution to early detection and prompt therapies. Future research should look at larger datasets and more advanced neural networks.

Keywords: Autism Spectrum Disorder, ABIDE Preprocessed, fMRI, Deep Learning, Machine Learning

Echoes of Concern: Security Challenges in Voice Assistant Systems

Rakesh Kumar Singh and Charu Gupta

Abstract. Voice Assistant systems like Alexa, Echo Dot, Siri, Google Assistants, and Cortana are widely used for the ease of being hands-free and being controlled through voice commands. Voice assistants utilize speech recognition, speech identification, and other machine learning algorithms to perform a task, connecting smart home devices, and smart vehicles, searching the web and even playing a game. The components of Voice assistants are exposed to attackers to gain unauthorized control through varying methods of injecting audio, speech, and inaudible voice commands. The attacks through the voice commands pose security and data privacy risks that need urgent attention. In this paper, different security challenges, and vulnerabilities in the Voice assistants are outlined. The study provides a brief overview of possible solutions to minimize the security and privacy risks associated with using Voice Assistants.

Keywords: Speech Recognition, Voice Assistants, Privacy, Security, Human-Computer Interaction

Assessing Chat GPT in Medical Domain

Aikanshi Rohilla, Palak Handa, Nidhi Goel

Chat Generative Pre-trained Transformer (ChatGPT), an advanced linguistic framework, holds immense potential for transforming various fields, including healthcare. In light of its versatility and capabilities, this systematic research delves into the exploration

of ChatGPT for addressing pressing challenges in healthcare and beyond. The research investigates applications of ChatGPT in various disease diagnoses, evaluating its accomplishments and the challenges they present. Significantly, the research accentuates the pivotal role of ChatGPT across various medical disciplines including Neurology, Gynecology, Gastroenterology, and Radiology. By enhancing responses to inquiries from non-experts, refining Artificial Intelligence (AI) generated reports, and evaluating its efficacy within these fields, ChatGPT emerges as a transformative force in medical practice. Moreover, the research not only offers valuable insights into the dynamic realm of medical research but also underscores the critical importance of ongoing advancements in diagnostic methodologies. As a road map for future investigations, this research inspires further exploration, providing a comprehensive perspective on the evolving medical landscape and fostering contributions to the continuous progress in disease diagnosis and treatment.

Keywords: ChatGPT, Medicine, Neurology, gastroenterology, Gynecology, Radiology

DeepFake Classification using Fine-Tuned Wave2Vec 2.0

Subham Sarkar, Ankita Ghosh, Anshul Gupta

Abstract. As deepfake attacks pose an increasing threat to biometric voice security, developing effective countermeasures is essential. This study presents a novel approach to detect audio deepfakes by utilizing the Wave2Vec2 framework. A customized post-processing technique was incorporated during inferencing on the In-the-Wild dataset. Additionally, five layers “dropout, dense, tanh, dropout, and output projection” were integrated into the Wave2Vec2 architecture, leading to enhanced classification accuracy that is on par with peer models. To further ensure the reliability of predictions, a unique algorithm was developed, which compares the model’s logits against a set threshold and iteratively repeats the process until a confident outcome is obtained. The fine-tuned Wave2Vec2 model demonstrated remarkable results in both configurations: the Grounded Wav2Vec2, which fine-tunes only the custom classification layers, achieved an accuracy of 99.42% and an equal error rate of 1.53%, while the AdaptiveWav2Vec2, which fine-tunes the entire model, reached an accuracy of 99.59% and an equal error rate of 1.4%. These findings align closely with state-of-the-art results for deepfake detection on similar datasets.

Keywords: Biometric Security, Deepfake Detection, Wave2Vec2

Turbocharging Pull Request Reviews: Exploring Generative AI for Code Review

Sourav Dutta, Anjali Sharma, and Jaymin Rajgor

Abstract. This paper examines the application of generative AI in automating pull request (PR) reviews, focusing on large-scale pre-trained models. The study highlights the potential of AI-driven tools to generate comprehensive PR descriptions,

identify code issues, and enhance developer productivity. Early results indicate that these tools, such as Code Evaluator, can streamline the review process, reducing the time spent on routine tasks and improving the consistency of code quality checks. Despite the benefits, challenges such as the accuracy of AI-generated content and the necessity for human oversight are discussed, emphasizing the need for balanced integration of AI in software development workflows.

Keywords: Generative AI, Pull Request (PR) Review, Code Review Automation, Large Language Models (LLMs), AI in Software Development, PR Description Generation, Developer Workflows, AI-driven Code Review, Software Quality Assurance, Machine Learning in Code Review.

Customer Churn Prediction Using Artificial Neural Networks

Manish Saini, Saurav Singha, Brijesh Kalakoti and Hashmat Fida

Abstract. The prediction of customer churn is crucial for any firm, but particularly for a banking firm as it is ordinarily cheaper to maintain an existing customer than obtain a new one. This paper aims at the use of Artificial Neural Networks (ANN) in modeling customer churn with analysis based on both static and dynamic characteristics. Criticizing prevalent models that depend on instantaneous data, this study incorporates time series data to enhance predictive precision. To control class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) were used in improving the performance of the models. Compared to Decision Tree, Random Forest, XG Boost and SVM, the proposed ANN gives the highest accuracy of 89.5%, precision of 91.0%, and F1-score of 89.2%. The outcomes of this study imply that the ANN model, with its desired low error rate, is suitable for identifying customers planning to churn in a timely manner and provides recommendations for more efficient anti churn measures.

Keywords: Customer Churn Prediction, Artificial Neural Networks (ANN), Class Imbalance, SMOTE (Synthetic Minority Over-sampling Technique), Credit Score Analysis, Churn Analysis.

Investigating AI Explanations in Medical Diagnosis

Nikita Garg, Manya Joshi, Palak Handa, and Nidhi Goel

Abstract. The lack of comprehensive explanations accompanying the results generated through automated diagnosis systems has encountered widespread mistrust among practitioners, patients, and stakeholders. This study aims to elucidate different Explainable Artificial Intelligence (XAI) strategies that can aid patients and practitioners in comprehending the results better. A set of biomedical images with different modalities served as the dataset for testing, predicting, and interpreting the predicted results. A pre-trained AI model was employed to generate predictions, which were subsequently explained using XAI techniques. Conventional criteria including accuracy, precision, recall, and F1 score were used to assess biological image analysis. Additionally, with the aid of visualization tools like feature maps and heat maps produced by XAI techniques such as gradient-weighted class activation mapping, local interpretable model agnostic explanations, and shapely additive explanations, interpretability was observed.

Keywords: Explainable Artificial Intelligence · Interpretability · Evaluation matrices

Automatic Assessment of Program Code using CodeBERTScore: A Transformer-based Approach

Hunny Gaur and Devendra K Tayal Amita Jain

Abstract. Evaluating student programming assignments is a labor intensive and time consuming activity for evaluators. A programming question can have several approaches to get the desired solution. During evaluation process, it is not feasible to check the functionality of all approaches adopted by the students to verify the results. However, automation of program code evaluation can reduce this burden. The paper proposed a web tool for automatic program code evaluation using transformer- based model CodeBERT and evaluation metric CodeBERTScore. The proposed web tool was tested for six programming languages, namely, C, C++, Java, Java Script, C# and Python, respectively. Out of these programming languages, C++ program code evaluation outperformed with a precision, recall, F1score and F3score of 0.9478, 0.932, 0.9406 and 0.9328, respectively.

Keywords: Code Evaluation, Natural Language Processing, Transformer Model, BERTScore

Rule based Named Entity Recognition for Hindi

Devanshi Suri, Nimisha Malviya, Nisheeth Joshi

Abstract. Named Entity Recognition (NER) is a fundamental task in Natural Language Processing (NLP) that identifies and recognises named entities in large structured or unstructured texts. These named entities can be names of persons, locations, organisations, date etc. We developed a POS-tag-rule based system that performs NER for Hindi. Rules were created using the POS tokens and gazetteers were used for identifying named entities. We also focused on creating gazetteers for person names, locations and organisations in Hindi language. The system is capable of identifying and classifying 5 named entities: person, location, organization, date, and time. Our system was evaluated on the benchmark Naamapadam dataset for three entities-person, location, and organisation. The accuracy achieved for the three categories was 0.9521, 0.9032 and 0.9701 respectively.

Keywords: Named Entity Recognition (NER), Rule-based systems, POS tagging

SHiTraD: Sanskrit-Hindi Translation Dataset

Nandini Sethi, Ayushi, Amita Dev, and Poonam Bansal

Abstract. Sanskrit is a culturally rich but low resource language. Presently, it is under represented and there is no digitized corpus available for Sanskrit-Hindi language pair publicly. We present a dataset, SHiTraD that consists of 45,500 parallel Sanskrit-Hindi sentences, collected from various sources online or offline through web or scanned from books. The source- target paired sentences were created using manual translations. This dataset is then used to train models and evaluate the translations and the results are shown using BLEU score evaluation metric comparing three architecture models 1. Encoder decoder with attention mechanism, 2. Self attention based transformer mechanism and 3. Multi-head based transformer mechanism.

Keywords: Machine translation, Sanskrit, Hindi, Corpus, Parallel corpus, Sanskrit-Hindi parallel corpus

Translating Heritage: A Transformer-Oriented Method for Neural Machine Translation from Sanskrit to Hindi

Nandini Sethi, Amita Dev and Poonam Bansal

Abstract. The globe is more technologically, socially, and culturally united when languages are used. Interlanguage information translation is essential for the interchange of ideas and information because different native speakers speak different languages. Even though Sanskrit is an old Indo-European language, it still requires a lot of information processing work to be thoroughly studied and utilized to open up new possibilities in computer science and computational languages. This paper describes a machine translation system that can convert Sanskrit to Hindi. The presented method uses a multi-head attention mechanism to train a transformer-based neural machine translation system. This is a novel strategy that can be applied to any low-resource language with rich morphology. It is a multi-field, universal system with minimal need for human interaction. Additionally, we constructed a parallel corpus for the language pair Sanskrit and Hindi. The corpus now contains more than fifty thousand parallel sentences. The BLEU performance metric was used to automatically assess the system. With an automatic assessment metrics-derived BLEU Score of 67.8%, the proposed approach outperformed the existing solutions. The results show that the developed and recommended strategy outperforms earlier studies for this language combination.

Keywords: Machine Translation System, Sanskrit, Hindi, Neural machine translation, multi-head attention, Transformer

Survey on the Availability of Datasets for Machine Translation Systems in Indian Languages

Ayushi, Amita Dev, and Poonam Bansal

Abstract. Machine translation is crucial to sharing knowledge using one language to any other language. It facilitates in communicating breaking language barrier. With the advent of neural networks, Machine translation has improved with good quality translations achieving humanlike translation. Availability of high quality dataset remains a challenge. Since most of the Machine Translation work has been done on high resource languages like English and Chinese, very less dataset is available for Machine translation in Indian Languages. This paper tries to present an extensive list of existing dataset available for machine translation in Indian languages. These dataset are mostly pivoted to English language or not open accessible or include limited languages. We have discussed the importance of creating an open-source dataset that contains parallel sentences for multilingual translation models that includes underrepresented languages.

Keywords: Dataset, Parallel Corpus, Machine Translation

Emotional Arousal of Female Politicians: Distinguishing Enthusiasm and Anger Using Data Augmentation and Machine Learning

Hilla Goren Barnea, Vered Silber-Varod, Elishai Ezra Tzur

Abstract. This study aims to distinguish between anger and enthusiasm, two high-arousal emotions with negative and positive valence, respectively, using prosodic features of speech. The dataset consists of plenary speeches from ten female members of the Israeli Parliament (Knesset), with a total of 482 utterances. These utterances were labeled as anger, enthusiasm, or "other" by 70 human labelers in an online perception test. We employed eXtreme Gradient Boosting (XGBoost), Support Vector Machines (SVM), and Random Forest to conduct acoustic-prosodic analysis. Among the models, XGBoost achieved the highest accuracy. The results showed that the label "other" was easier to distinguish compared to anger or enthusiasm. However, when anger and enthusiasm were combined into a single "high-arousal" label, the distinction from "other" became more pronounced. Frequency and intensity parameters were found to be crucial across all categories. Additionally, significant differences in prosodic attributes were observed among different politicians, reflecting individual emotional styles in speech. A correlation was also found in the feature importance rankings across all three models, with fundamental frequency, rhythm, and intensity identified as the most important features.

Keywords: Emotions, Speech, Acoustic-Prosodic Features, Politicians.

A Metric-Driven Comparative Study of Text Summarization Model: Insights from State-of-the art LLMs

Shivang Gaur, Aman Karunik , Swapnali Naik

Abstract. The exponential growth of textual data in the digital era has necessitated the development of efficient Text Summarization systems to distill critical information from extensive documents. This study presents a Metric Driven comprehensive comparative analysis of three state-of-the-art Large Language Models (LLMs) for Text Summarization: ChatGPT-4o, Gemini 1.5 Pro, and Cohere Command. Leveraging advanced evaluation metrics such as ROUGE, BERTScore, METEOR, and Cosine Similarity, we assess the performance of these models in generating accurate and coherent summaries. Our findings indicate that while all three models demonstrate significant capabilities, ChatGPT-4o consistently outperforms its counterparts in terms of semantic similarity and structural coherence, as evidenced by its superior scores across multiple metrics. The results of this study not only underscore the potential of LLMs in enhancing text summarization tasks but also provide critical insights into the strengths and limitations of the current LLM model.

Keywords: Text summarization, Large Language Model (LLM), Natural Language Processing (NLP), ChatGPT - 4o, Gemini 1.5 Pro, Cohere command.

A.R.I.S.E. Artificial Recognition and Intelligent Sign-language Eyewear - A Multifunctional Smart Glasses System for Enhanced Accessibility and User Assistance

Rahul Thambi, Roshi Bhati, Nikhil Nerurkar, and Ketan Shah

Abstract. The rapid advancement of smart wearable technology has created an opportunity to integrate various high-utility functions into a single device. This research introduces A.R.I.S.E. (Artificial Recognition and Intelligent Sign-language Eyewear), a versatile smart glasses system that combines cutting-edge technologies, including speech recognition and generation capabilities, to enhance user interaction and accessibility. A.R.I.S.E. offers a range of essential functions such as real-time object recognition, text extraction, AI-driven conversational support, facial recognition, and gesture control. This research details the design, integration, and implementation of these diverse software functions into a cohesive wearable device, highlighting challenges in creating a seamless user experience. We also explore potential applications in accessibility and communication, emphasizing the role of large language models in conversational AI to facilitate effective human-machine interaction.

Keywords: Smart Glasses, Wearable Technology, Artificial Intelligence, Computer Vision, Accessibility.

NLP-Driven Sentiment Analysis for Fake News Identification: The Responsive Truth

Arpita Singh, Arpit Dubey, Rabins Porwal

Abstract. In the current era of technology news can often be considered more valuable than money. However, this information needs to be genuine, which is often not the case, leading to a pressing need to distinguish real news from fake news. News, as a form of information, is subjective and relies heavily on proof and source credibility. People can typically discern real news from falsehoods through their innate ability to apply logic and recognize dubious sources. However, having a few trusted sources for fact-checking is essential. There is an urgent demand for software solutions that can quickly address the spread of false news, making this a highly researched area. As part of Information Retrieval, it has garnered significant attention from researchers worldwide seeking real-time solutions. In this article, we analyze various research and survey articles to provide readers with a concise overview of fake news, its various forms, characteristics, and identification basics. We can reasonably conclude that the majority of fake news is spread with the purpose of instilling hatred and resentment in Indian society.

Keywords: Machine Learning, Sentiment Analysis, NLP, TextBlob, Hugging Face Transformers

Human Evaluation of Some Machine Translation Toolkits for English-Hindi Language Pair

Riya Lonchenpa, Anju Krishnia, Pragya Katyayan, Mhasilenuo Vizo, Nisheeth Joshi

Abstract. This work examines machine translation systems from a human perspective. the modern era of neural machine translation and its applicability. A scalar quality metric comprising 11 factors was employed in the study, and it was graded on an 0-4 scale by humans. IndicTrans2 and Facebook's NLLB MT Engines were the two MT engines that were taken into consideration for the research. In order to determine the common mistakes that each MT engine encountered, the human evaluators conducted assessments using eleven parameters. The results showed that, of the two, IndicTrans2 MT Engine yielded superior outcomes.

Keywords: Human Evaluation, Subjective Evaluation, MT Development Process.

Next-Gen Imaging: The Power of Hyperspectral Data and Autoencoders

Pallavi Ranjan, Sushma Hans, and Salih Ismail

Abstract. Hyperspectral imaging (HSI) captures detailed spectral and spatial information across hundreds of contiguous wavelengths, providing unprecedented data for analysis. However, HSI data is high-dimensional, imposing challenges for storage, processing and analysis. Autoencoders (AE) are unsupervised deep learning models that learn efficient data encoding in lower dimensional spaces while capturing salient features. Integrating HSI and AE provides a powerful solution by leveraging AE's dimensionality reduction and feature learning capabilities on HSI's rich data. This paper reviews the integration of these two technologies, covering background, motivation, applications in classification, unmixing and anomaly detection, hyperparameter tuning, and future research directions. The combined HSI-AE approach unlocks new possibilities across domains like agriculture, medical imaging, remote sensing and environmental monitoring.

Keywords: autoencoders, hyperspectral, classification, segmentation, detection

Redefining Vision Tasks: The Power of Transformers in Classification, Detection, and Segmentation

Anshika Jain, Akansha Khandelwal, Vandana Singh, Astha Sharma and Ashwni Kumar

Abstract. Plants exert a vital influence on the survival of life on Earth by producing oxygen and supplying the necessities for food, medicine, and a variety of industrial applications. Given an enormous array of different plant species, each with its own unique features such as form and texture, manually identifying them for diverse applications becomes challenging. Consequently, the use of artificial intelligence-driven feature extraction techniques becomes critical in order to speed the process. Feature extraction involves identifying and capturing specific characteristics from leaf images. This paper aims to offer a comprehensive overview of the related work done by the researchers in the previous years. For feature extraction, Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP), and Grey Level Co-Occurrence Matrix (GLCM) are preferably employed, followed by classification using neural networks, and so on. HOG excels at extracting details concerning the distribution of edge directions or intensity gradients within an image. LBP performs well in describing local patterns of pixel intensities, making it useful for work involving texture. Different techniques are used depending on the specifications. The foremost objective of our work is on investigation on various feature extraction techniques that can be utilized.

Keywords: HOG, LBP, GLCM, machine learning, feature extraction, PCA, classification, RGB

Electricity Demand Prediction using Machine Learning

Ali Sadiq Khawja

Abstract. Since decades prediction of electricity demand has been of utmost interest for better grid management and cost control. An accurate demand prediction can help reduce the emission of greenhouse gases by better planning of extracting energy from cleaner sources. It also allows grids to integrate renewable energy sources like wind and solar energy with nonrenewable sources according to the demand. Another advantage of demand prediction is stabilizing the grid by adjusting the generation and distribution according to the forecast as sudden highs and lows in demand can destabilize the grid. This paper explores different machine learning algorithms to predict electricity demand like gradient boosting methods, support vector machine, neural networks, etc. This project focused on comparing different algorithms based on their performance. Traditionally, demand is predicted using statistical methods. An advantage of machine learning models over statistical models is the capability to capture non-linear relationships between data. The data consisted of different hourly weather variables along with the demand. The results showed neural networks and tree-based models performed better than other models of which neural networks were the best. This is because of their capability to handle non-linear data as well as being robust to outliers. The performance of the models was measured using different metrics like R2 score, Mean squared error and root mean squared error.

Keywords: machine learning, gradient boosting, random forest, demand prediction, neural networks

Deepfake Detection: A Comprehensive Analysis of Modern Techniques

Deepanshi Jalan, Shradha Jain, Anandita Tuli, Vanshika Chaudhary,
Astha Sharma, Ashwani Kumar

Abstract. Given the recent progress in artificial intelligence over the last few years, various capabilities to manipulate multimedia to make people look and sound like other people have been developed. Though the technology has a lot of potential usages in the entertainment, education and activism industries, it has been leveraged by malicious users for illegal or harmful purposes. Apart from creating skepticism and spreading false information, it also pose threats to the privacy and security of individuals or groups, owing to its convincing ability to impersonate anyone. High-quality, realistic videos and images that have been digitally manipulated are referred to as 'deepfakes.' Various researches are carried out in the field of deepfake images and certain techniques have been devised to detect the same. This paper aims to summarize the history of deepfakes, the existing generation methods, possible threats it poses in near future and why it is essential to detect the deepfake content.

Further, we provide an overview of 25 research papers from 2016 to 2024 that have presented various methodologies to detect deepfake images. The analysis has been performed by grouping the techniques, to detect deepfake images, into several categories, namely, machine learning and deep-learning techniques. Moreover, various methods to detect deepfake videos have also been analyzed. Additionally, some other research questions relevant to the context have also been answered and the performance of each of the methods have been compared to conclude that deep learning methods outperform other techniques for deepfake detection.

Keywords: Deepfake image detection, digital image manipulation, deep learning, convolution neural networks, deepfake video.

A Comprehensive Analysis of Recent Methodologies for Irregular Heart Beat Prediction

K. Rema and C. P. Maheswaran

Abstract. To extract pertinent knowledge from the vast amounts of medical data that are available, real data analysis methods must be used. Heart disease has been the leading global cause of death over the past ten years. The prevalence of irregular heartbeat disorder is higher in less developed and developing countries as a result of expensive diagnostic procedures. Cardiopulmonary arrest and heart failure are the two frequent names for heart conditions. In developing and impoverished nations, heart disease is more common due to expensive diagnostic techniques. The optimization and reasoning methods for diagnosing heart illness are showing excellent results in the identification and prediction of irregular heartbeats. The effects of integrating numerous strategies that have previously shown better results in the detection of cardiac disease have been the subject of research. However, there hasn't been much attention placed on implementing the suggested approaches to identify effective treatments for heart disease. The relative study are approved out in the methods as AlexNet technique to calculate the best accuracy in diagnosing heart disease as 97 %.

Keywords: PSO- Particle Swarm Optimization, RFSO – Random Forest Swarm Optimization, XG Boost optimization, Heart disease diagnosis.

Recent Advances in Denoising Techniques for Hyperspectral Image Enhancement

Priyanka Gupta, Garima Jaiswal, Ashish Kumar

Abstract. HSI is a robust technology that covers the spectral information in detailed manner over the wide range of wavelengths, makes it successful in the domain of medical imaging, agriculture and majorly in remote sensing. But sometimes hyperspectral images are targeted by noise due to the environmental reasons, imperfections in sensors and data transmission errors.

Effective denoising of images is must so that the quality of hyperspectral images can be maintained which assures the accuracy of subsequent investigations. This paper introduces the concept of noise in images and explores various types of noise. It further reviews existing denoising methods specifically applied to hyperspectral images, categorizing them into four major groups. Additionally, it outlines the challenges in hyperspectral image denoising, such as selecting appropriate models and parameters, managing spectral distortion and band-wise normalization, accurately estimating noise variance, and addressing mixed noise scenarios. This review aims to be a valuable resource for beginners to the field of hyperspectral image denoising, offering insights into denoising strategies for HSI applications.

Keywords: Hyperspectral Images, Denoising, spectral, spatial

Predictive Models in Biomedical Applications: A Machine Learning Approach

Kanika Wadhwa, D Ananya Himanshu Mittal, and Arun Sharma

Abstract. In biomedical applications, predictive modeling makes predictions about future events and behaviors by analyzing massive patient data sets through machine learning. Given its application in medical contexts, this approach must be employed with utmost caution; otherwise, the potential consequences could jeopardize human life. This paper highlights the necessity of incorporating domain-specific information into the training processes of machine learning algorithms, a factor that is crucial for numerous applications, particularly in the realm of biological data analysis for breast cancer detection, which affects women globally and is associated with a significant mortality rate. By presenting a novel predictive model designed to enhance the accuracy of breast cancer classification through various machine learning (ML) approaches, we conducted experiments using the WDBC dataset within the Jupyter framework. According to the findings of a thorough comparison investigation, the suggested model is superior to the most advanced machine learning algorithms in terms of accuracy, precision, recall, and F- measure.

Keywords: Biomedical Application, Machine learning, Disease diagnosis, Breast cancer, WDBC dataset

Comparative Analysis of Multiple Embedding Models for Text Based Document Similarity

Shreya Khosla, Shresth Jain, Anupama MA and Rama Subba Reddy Thavva

Abstract. Document similarity is a foundational task when it comes to applications such as plagiarism detection, information retrieval, and other Natural language processing domains. Traditional similarity measures, primarily based on lexical comparison, such as Jaccard index and cosine similarity, have constrained capability of capturing semantic relationship.

Recent progression in embedding models, including FastText, BERT, RoBERTa, and SBERT, have significantly enhanced the capability to represent semantic similarity within vector spaces. This study focuses on consolidating the model preference based on use case correlating embedding models and similarity metrics while considering their effectiveness for capturing semantic document similarity. We employ advanced metrics such as Soft Cosine Similarity and Jensen-Shannon distance to assess the performance of these embeddings on benchmark datasets. Our findings indicate that the Soft Cosine Similarity metric, when coupled with the SBERT paraphraseMiniLM-L6-v2 model, demonstrates superior performance in accurately capturing semantic document similarity.

Keywords: Document Similarity, Embeddings, Semantic Similarity, Soft Cosine Similarity, Jensen-Shannon Distance.

Transfer Learning for Leaf Disease Image Classification via CNN Model

Kirti Dabas, Anuradha Chug, Amit Prakash Singh, Ravinder Kumar,
Veerubommu Shanumugam

Abstract. The agriculture industry greatly aids a nation's economic maturation. Although, plant diseases are a main factor causing crop devastation, which lowers the amount and quality of agricultural production and results in significant financial loss. Therefore, it is essential to detect plant illnesses well in advance to prevent crop devastation. Automated and timely identification of these illnesses will boost output and facilitate system digitization. Our primary goal is to identify potato diseases using sophisticated machine-learning algorithms on leaf image data. This paper uses image-based Convolutional Neural Network (CNN) models with a transfer learning strategy to study and compare performance using pre-trained models via VGG16, ResNet50, and MobileNetV2. In this work, our primary focus is on potato diseases consists of 2152 images of three classes: Early Blight, Late Blight and Healthy. It is ascertained through experimentation that MobileNetV2 outperformed other models. The proposed method achieved a 90.71%, 65.89% and 93.27% accuracy for potato leaf images. This model helps the farmer detect disease early and help them in taking preventative action well on time to reduce their loss.

Keywords: Diseases, Transfer Learning, VGG16, ResNet50, MobileNetV2

Leveraging Depth Data and Parameter Sharing in Vision Transformers for Improved Face Anti-Spoofing

Aashania Antil, Chhavi Dhiman

Abstract. With the rapid advancements in face recognition (FR) technology, current systems perform well in unconstrained scenarios. However, detecting face spoofing attacks remains a significant challenge, making face anti-spoofing (FAS) a critical research area. Despite the development of various anti-spoofing models, their ability to generalize to unseen attacks often deteriorates in the presence of challenging variations such as background, lighting, diverse spoofing materials, and low image quality. To address these limitations, we propose a novel bi-branch FAS framework that leverages a pre-trained Vision Transformer (ViT) with RGB and depth data as input. The ViT's self-attention mechanism excels at capturing intricate image contexts, making it particularly effective for Presentation Attack Detection (PAD) tasks. To enhance computational efficiency, we introduce a parameter-sharing technique within the dual-branch ViT network, substantially reducing the computational burden while maintaining robust feature learning. Our method achieves state-of-the-art performance on the Replay-Attack, CASIA-FASD, and OULU-NPU benchmarks across intra- and cross dataset testing, while also enhancing computational efficiency.

Keywords: Face anti-spoofing, presentation attack, vision transformer, parameter sharing

Comparative Analysis of Deep Learning Techniques on Tomato Plant Disease Detection

Riya Sharma, Ayushi Pasrija, Saloni Gupta and Poonam Bansal

Abstract. Plant diseases are harmful conditions that affect plants, leading to symptoms like wilting, discoloration, and reduced crop yields. Microorganisms like fungi, bacteria, viruses, nematodes, and phytoplasmas typically cause these diseases. Given that agriculture serves as a key source of income and employment for many, the detection of plant diseases becomes a critical task. To address this issue, the authors have implemented Artificial Intelligence algorithms on a visual dataset of Tomato plant leaves containing 8443 RGB images of 4 most common diseases in Tomato crop to make early disease detection easier. The paper comprehensively elucidates the entire process, encompassing image preprocessing, model training, testing procedures, and ultimate deployment, along with a comparative analysis of four distinct Machine Learning and Deep Learning models, namely Random Forest Classifier, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Visual Geometry Group Neural Network (VGG). After extensive research, it has been concluded that RNN outperformed all the models for the entire dataset.

Keywords: Deep learning, plant disease detection, convolutional neural network, random forest.

Image Captioning Using Deep Learning Models: A Comprehensive Overview

Nisha, Shailesh D Kamble, and Himanshi Mittal

Abstract. Person with impaired vision have faced issues while understanding and interacting with pictures and videos. Image captioning enhance their capacity to deal with visual content by creating descriptive language, that makes the images more accessible and engaging. This technique generates relevant captions from visual data by connecting visual feature extraction and language modeling. It acts as a key connector between computer vision and natural language processing, that allows machines to analyze and describe pictures. This article examines the developments of models, such as encoder-decoder frameworks and attention processes, that combine visual feature extraction with language creation and focusing on how object identification improves caption accuracy and contextual relevance. The study discusses on various deep learning models and also highlights various evaluation metrics like BLEU, ROUGE, and CIDEr, which measure the quality of generated captions. Moreover, the paper addresses potential improvements and prospects in image captioning systems.

Keywords: Image Captioning, Attention Mechanism, Object Detection

Bridging the Gap: Real-Time ASL Fingerspelling to Sentence Translation

Manya Chandna, Abantika Dasgupta, Bhumika Gupta, Ravinder M,
S.R.N. Reddy, Rishika Anand

Abstract. Sign language represents an essential communication tool that benefits the Deaf and hard-of-hearing community, yet the lack of effective communication with non-signers often leads to social isolation. This research emphasizes on the development of an application using Tkinter for recognizing words in American Sign Language (ASL) through fingerspelling and turning them into sentences. It presents a novel approach to converting American sign language into text in real time, utilizing cutting-edge computer vision and deep learning techniques. The proposed method uses a webcam-based self-made dataset and a custom model with the InceptionV3 architecture as the basic model to categorize the letters via transfer learning. The recognized letters were then progressively combined to form words, which were further computed into meaningful sentences. Our approach focuses on improving recognition accuracy while maintaining real time processing for minimizing the communication gap and contributing to promote social inclusivity of the Deaf community.

Keywords: Deep Learning, Sign Language Recognition, InceptionV3

Real-Time Detection of Household Objects Using Single-Shot Detection With Mobile Net

Pallavi Sharma, Isha Malhotra, Palak Handa, and Nidhi Goel

Abstract. Real-time household object detection presents a formidable challenge due to the requirement for rapid computation in identifying objects promptly. Popular object detection models, such as Region-based Convolutional Neural Networks (R-CNN), Fast R-CNN, Faster R-CNN, You Only Look Once (YOLO), Spatial Pyramid Pooling Network (SPPNet), Single Shot Multi-box Detection (SSD), and others, each come with their advantages and disadvantages. The present work proposes to use a Convolutional Neural Network (CNN) model known as SSD, as it processes images in a single step. Additionally, to address the computational constraints prevalent in devices such as mobile phones and laptops, a lightweight network model utilizing depth-wise separable convolution, termed Mobile Net, is incorporated into the proposed methodology. The SSD-Mobile Net model offers a mean Average Precision (mAP) of 21%, with the model weights ranging between 25 to 30 Mega Bytes (MB), and achieves a processing speed of 30 milliseconds (ms) per image. Experimental findings demonstrate that integrating Mobile Net with the SSD model significantly enhances accuracy in real-time household object identification. Furthermore, this research extends beyond theoretical exploration, showcasing the deployment of the SSD-Mobile Net model within a flutter application, providing a user-friendly interface for seamless interaction with real-time detection outcomes.

Keywords: Object Detection, Flutter, Deep Learning, SSD, Mobile Net, Household

AccuRep: A Real-Time Android Application for Exercise Tracking Using Move Net Pose Detection and Joint Angle Analysis

Atharva Swami, Ishi Prashar, Tanmay Sahasrabudhe, and Preeti Agarwal

Abstract. AccuRep is an Android-based fitness tracking app which incorporates human pose evaluation that offers real-time feedback on workout form. AccuRep assists users in ensuring appropriate posture during workouts, which lowers the chance of injury and encourages consistency and progress monitoring. With live camera input, the app uses the TensorFlow Lite's MoveNet model to compute key joint angles for exercises like bicep curls, shoulder presses, tricep extensions, and squats. The developed AccuRep include features such as meal planning assistance, BMI/BMR calculations, and real-time body landmark identification with a repetition counter that increments with every correct proper form. For users of all fitness levels, this app provides a comprehensive exercise tracking solution together with an easy-to-use UI. AccuRep's early studies yield encouraging results with an average accuracy of 90.2% showcasing precise joint angle recognition and a smooth user interface.

This software provides users a complete fitness tool by effectively tracking their exercise form and progress.

Keywords: Edge Computing, Android Fitness App, Pose Estimation, Real-time exercise tracking, Exercise Counter, Tensor Flow Lite, Body landmark detection, Artificial intelligence in fitness, Joint angle calculation, Move Net

Exploring State-of-the-Art approaches for Heart Disease Detection: A Detailed Analysis

B. Shamna and C. P. Maheswaran

Abstract. The availability of enormous volumes of medical data necessitates the employment of actual data analysis technologies in order to derive relevant knowledge. In the last 10 years, heart disease has been the top cause of mortality worldwide. Due to costly diagnostic methods, heart disease has an elevated incidence in underdeveloped and poorer nations. The two common terms for heart disorders are cardiopulmonary arrest and heart failure. Due to costly diagnostic methods, heart disease has an increased incidence in underdeveloped and poorer nations. To diagnosing the heart disease the optimization and intelligent methods are showing promising results is in the detection of diseases. Researchers have been examining the impact of combining many techniques that have shown improved outcomes in the identification of heart disease previously. However, little emphasis has been paid to adopting the proposed methods to find a good treatment for people with heart disease. The relative study are approved out in the optimization method as Improved SVM to calculate the best accuracy in diagnosing heart disease as 94.97%.

Keywords: PSO- Particle Swarm Optimization, RFO – Random Forest Optimization, ANN – Artificial Neural Network, XG Boost optimization, Heart disease diagnosis.

Comparative Evaluation of 3D U-Net and SegResNet Architectures for Brain Tumor Segmentation Using Adam and Ranger21 Optimizers

Nisha Purohit,C P Bhatt and Subodh Kumar

Abstract. Accurate segmentation of brain tumors from MRI scans is a critical task in medical imaging, essential for precise diagnosis and effective treatment planning. This study investigates the application of SegResNet and 3D U Net, an advanced deep learning architecture, for brain tumor segmentation using the MONAI framework.

The performance of these two architectures is evaluated and compared by using two different optimizers (a) Adam, a well-established optimization algorithm in deep learning, and (b) Ranger21, a newer optimizer designed to enhance training stability and generalization. The study presents a comprehensive technical workflow starts with data preprocessing, model architecture design, training procedures, and evaluation metrics such as Dice Score, Intersection over Union (IoU) and precision, Recall and F1 Score. The study employed to comparative analysis of model's performance and generalizability across different model and optimizer combination. The results demonstrate the effectiveness of the SegResNet and 3D U Net architecture, yielding high accuracy and robust segmentation performance, comparing its potential for improving clinical out comes in tumor segmentation.

Keywords: Deep learning, 3D UNet, SegResNet, Segmentation, MONAI, Ranger21, Adam, Optimizers

Creating and Analyzing a Dictionary-Based Lexical Resource for Somali Homograph Disambiguation

Abdullahi Mohamed Jibril and Abdisalam Mahamed Badel

Abstract. In natural language processing (NLP), machine learning models often encounter challenges due to the lexical ambiguity of languages, particularly in tasks such as machine translation, Text-to-Speech (TTS), and information retrieval. Homographs present significant challenges in this context. While many studies focus on homographs in resource-rich languages, such works are lacking for languages with limited resources. In this paper, we present a lexical resource for Somali homograph Disambiguation based on a dictionary. We employed various natural language processing techniques, including sentence embeddings and machine learning clustering algorithms, to analyze 4,809 homograph entries, representing 1,592 unique homographs. We created distributions of homographs based on their semantic relationships. Our research provides insights into the distribution, semantic relationships, and clustering patterns of Somali homographs, contributing to the understanding of lexical ambiguity in this understudied language.

Keywords: homograph, Resource, Somali, Machine learning

CodeBERT-BiGRU for Software Defect Prediction

Ruchika Malhotra and Priya Singh

Abstract. Software defect prediction (SDP) has a key role in the realm of software engineering. It allows for early identification of defects so that timely action can be taken to correct them. Preliminary research in the field includes techniques based on machine learning that focus only on the simple code characteristics followed by deep learning(DL) models that also capture the semantics of the code. Recently, the surge in the popularity of large language models (LLMs) have piqued interest among researchers to apply them for SDP. In the current study, we have proposed a novel model based on CodeBERT and BiGRU, that first extracts semantic information using CodeBERT, combines it with the simple code characteristics with an advanced merger mechanism, and collectively uses them to detect defects in the code. Experimental results of the proposed approach on multiple open-source datasets compared with the commonly used DL models signify its superiority. The findings of the work would be helpful in guiding future researchers and practitioners to best utilize the power of LLM in conjunction with the appropriate DL model for SDP.

Keywords: CodeBERT, Software defects, Deep learning, Word embeddings, Natural language processing

AI-Driven Essay Evaluation: Enhancing Accuracy and Fairness in Automated Scoring Systems

Kuldeep Kumar Singh, Gaurav Gupta, Sapna Chaudhary, Sajal Singh Masand, Swati Saran, Shivani Sharma

Abstract. Over the last few years, Automated Essay Scoring (AES) technology systems have seen increasing focus. Researchers focus on AES to reduce the load of educators and provide consistency along with equity in grading essays. This paper designs and develops an AES system that makes use of NLP techniques and ML algorithms for assessing student essays. It evaluates language features-including grammar, coherence, structure, and content relevance-for scores that approximate human judgments as closely as possible. In this paper, we evaluate several scoring models, both traditional statistical and deep learning approaches, to select the single model that performs best on essay grading tasks. Another aspect is the estimation of the bias present in automated grading and how it might be ensured that fairness is guaranteed at the same time as transparency on the scoring procedure. The proposed AES system shows very promising results with high accuracy compared to the human evaluators and has the potential to bring revolution into the landscape of educational assessment by offering quick, reliable, and scalable solutions in grading.

Keywords: Automated Essay Scoring, Natural Language Processing, Machine Learning, Essay Evaluation, Linguistic Features, Deep Learning, Fairness, Bias Mitigation, Educational Assessment, Scalable Grading Solutions

Advancements in Alzheimer's Disease Detection: A Comprehensive Review of Deep Learning Approaches in MRI Imaging

Khushi Jindal, Seeja. K., R, and D. K. Tayal

Abstract. Timely and adequate identification of Alzheimer's disease (AD) is essential for its proper treatment. Lately, progress in deep learning and neuro-imaging has played considerable roles in facilitating and improving diagnosis. This study represents several attempts and methods based on deep learning techniques for detecting Alzheimer's disease such as CNNs, hybrid models, ensemble methods, transfer learning, and complex architecture discrimination. In this setting, CNNs have played a central role in detecting AD by facilitating the examination of spatial hierarchies in MRI images. Hybrid and ensemble models approach the limitations of an individual model by combining many algorithms to increase accuracy and robustness. Using pre-trained models enables transfer learning to improve the diagnosis capability of the disease in situations with scant data. Furthermore, novel technologies such as transformers and multi-stage frameworks complement and enhance existing diagnostic methods. The research proved that by applying such integrated approaches, detection accuracy and their early appearance significantly increased, which illustrates the effect of innovative deep learning technologies on the treatment of Alzheimer's disease.

Keywords: Alzheimer's Disease, Deep Learning, Medical Imaging, Convolutional Neural Networks, Hybrid Models, Transfer Learning



INDIRA GANDHI DELHI TECHNICAL UNIVERSITY FOR WOMEN
(ESTABLISHED BY GOVT. OF DELHI VIDE ACT 09 OF 2012)
ACCREDITED NAAC A+ GRADE
KASHMERE GATE, DELHI-110006

WOMEN EDUCATION | WOMEN ENLIGHTENMENT | WOMEN EMPOWERMENT

**6TH INTERNATIONAL CONFERENCE ON
ARTIFICIAL INTELLIGENCE AND SPEECH TECHNOLOGY**

13th - 14th November, 2024

AIST 2024

After the grand success of five versions of International Conference on Artificial Intelligence and Speech Technology (AIST2019, AIST2020, AIST2021, AIST2022 and AIST 2023), Indira Gandhi Delhi Technical University for Women, Delhi is all set and ready for hosting **ANRF(Formerly known as SERB) - Sponsored 6th International Conference on Artificial Intelligence and Speech Technology (AIST2024)** on 13th - 14th November, 2024 at IGDTUW Campus, Delhi. Last five Conference witnessed a number of Expert Researchers, Academicians and Industry practitioners in AI and Speech Domain from premier Institutions in India and abroad including Japan, Australia, France, Hungary, Myanmar, Uzbekistan, Austria, Vietnam and others.

ABOUT IGDTUW

Indira Gandhi Delhi Technical University for Women (IGDTUW) was established by the Govt. NCT of Delhi in May, 2013 vide Delhi Act 09 of 2012, as a non-affiliating University to facilitate and promote studies, research, technology, innovation, incubation and extension work in emerging areas of professional education among women, with focus on engineering, technology, applied sciences, architecture and its allied areas with the objective to achieve excellence in these fields.

IMPORTANT DATES

Paper Submission: 10th October, 2024
Acceptance Notification: 15th October, 2024
Registration: 21st October, 2024
Conference Dates: 13th - 14th November, 2024

CALL FOR PAPERS

Papers are invited in these tracks:

- Track 1** - Spoken Language Processing
- Track 2** - Speech Recognition Signal Processing
- Track 3** - Trends and applications in speech technology
- Track 4** - Multi-lingual language translation and assessment
- Track 5** - NLP and AI
- Track 6** - Understanding large language models
- Track 7** - Computer vision and AI

TUTORIAL SESSION

Hands-on Sessions by Renowned Researchers

- Tools and Techniques for Speech Processing and Database Development
- Deep Learning and Machine Learning models for NLP and Speech

SPONSORS



SUBMIT PAPER HERE

All accepted and presented papers of the conference by duly registered authors will be submitted to Springer CCIS for Publication.

PUBLICATION



Indexed by Scopus and WoS

ORGANISING COMMITTEE



PROF. RANJANA JHA
HON'BLE VICE CHANCELLOR, IGDTUW
GENERAL CHAIR



PROF. S.S. AGRAWAL
DG KIIT, EMERITUS SCIENTIST, CSIR
TECHNICAL PROGRAM COMMITTEE CHAIR



PROF. ARUN SHARMA
DEAN ACADEMIC AFFAIRS, IGDTUW
ORGANIZING CHAIR



PROF. R.K. SINGH
REGISTRAR, IGDTUW
ORGANIZING CO CHAIR

ORGANIZING CO-CHAIR:

- DR. SYED ZULKARNAIN IDRIS, UNIVERSITI MALAYSIA PERLIS, MALAYSIA
- DR. RITU RANI, BHAGWAN PARSHURAM INSTITUTE OF TECHNOLOGY, DELHI(GGSIPU)

