Indian Journal of Signal Processing (IJSP)
ISSN: 2582-8320 (Online), Volume-6 Issue-1, February 2026

Speech Emotion Recognition Using MFCC Features and LSTM-Based Deep Learning

Devraj, Ravindra Nath, Nikita Singh, Vibhushit Katiyar, Amber Srivastava

***Abstract:** Speech Emotion Recognition (SER) has emerged as a significant research area within Human–Computer Interaction (HCI), enabling intelligent systems to interpret human emotional states from spoken audio. Accurate emotion recognition from speech plays a crucial role in enhancing natural interaction between humans and machines. This paper presents a deep learning–based SER framework that combines Mel-Frequency Cepstral Coefficients (MFCC) for feature extraction with Long Short-Term Memory (LSTM) networks for temporal modelling and emotion classification. MFCC features effectively capture the spectral characteristics of speech signals, whereas LSTM networks are well-suited to modelling long-term temporal dependencies inherent in emotional speech patterns. The proposed model is trained and evaluated on the Toronto Emotional Speech Set (TESS) dataset, which covers multiple emotional categories, including happiness, sadness, anger, fear, and neutrality. Experimental results demonstrate that the proposed MFCC–LSTM approach achieves promising classification accuracy, indicating its effectiveness in recognising emotional states from speech signals. The findings highlight the potential applicability of the proposed system in real-world scenarios, including virtual assistants, call centre analytics, and mental health monitoring systems, thereby contributing to the development of emotion-aware intelligent interfaces.*

***Keywords:** Speech Emotion Recognition, MFCC, LSTM, Deep Learning, TESS Dataset, Human–Computer Interaction, Audio Signal Processing*

Nomenclature:
SER: Speech Emotion Recognition
HCI: Human–Computer Interaction
MFCC: Mel-Frequency Cepstral Coefficients
LSTM: Long Short-Term Memory
TESS: Toronto Emotional Speech Set
CNNs: Convolutional Neural Networks
RNNs: Recurrent Neural Networks
SVMs: Support Vector Machines
GMMs: Gaussian Mixture Models
DCT: Discrete Cosine Transform
FFT: Fast Fourier Transform

Manuscript received on 27 January 2026 | First Revised Manuscript received on 06 February 2026 | Second Revised Manuscript received on 10 February 2026 | Manuscript Accepted on 15 February 2026 | Manuscript published on 28 February 2026.
* Correspondence Author(s)

Dr. Devraj*, Department of Division of Social Science, ICAR-Indian Institute of Pulses Research, Kanpur (Uttar Pradesh), India. Email ID: drmishra_1969@yahoo.co.in, ORCID ID: 0009-0005-4381-5933

Dr. Ravindra Nath, Associate Professor, Department of Computer Science, Babasaheb Bhimrao Ambedkar Central University, Lucknow (Uttar Pradesh), India. Email ID: rnkatiyar@gmail.com

Nikita Singh, Department of Computer Centre, ICAR-Indian Institute of Pulses Research, Kanpur (Uttar Pradesh), India. Email ID: officialnikita2626@gmail.com

Vibhushit Katiyar, Student, Department of Computer Science, B.Tech.(CS) Student, Lovely Professional University, Jalandhar (Panjab), India. Email ID: vibhushitkat@gmail.com

Amber Srivastava, Student, Department of Computer Science, B.Tech.(CS) Student, Lovely Professional University, Jalandhar (Panjab), India. Email ID: srivastava amber@gmail.com

I. INTRODUCTION

Recently, most speech processing tools use AI to understand better and respond to human emotions. Speech Emotion Recognition (SER) aims to identify emotional states conveyed through vocal expressions. SER plays a critical role in a wide range of real-world applications, including virtual assistants, intelligent tutoring systems, customer service analytics, healthcare monitoring, and mental health assessment. In such applications, emotion-aware systems can improve user engagement, enhance decision-making, and support early detection of emotional distress. However, speech emotion recognition remains a challenging task due to factors such as speaker variability, environmental noise, emotional ambiguity, and the dynamic nature of speech signals. Traditional SER approaches primarily relied on handcrafted acoustic features and conventional machine-learning classifiers, such as Support Vector Machines, k-Nearest Neighbours, and Gaussian Mixture Models. While these methods achieved moderate success, they often struggled to capture the complex temporal dependencies and non-linear patterns present in emotional speech. Recent advancements in deep learning have significantly improved SER performance by enabling automatic learning of discriminative features and temporal representations directly from data.

Humans express their emotions by changing the tone, pitch, intensity, and rhythm of their speech. Earlier approaches to emotion recognition in speech primarily relied on manually engineered acoustic features. However, the emergence of deep learning has significantly improved performance through advanced architectures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Among these, Long Short-Term Memory (LSTM) networks have proven particularly effective due to their ability to capture temporal patterns and sequential dependencies inherent in speech signals. This study strives to create a more accurate and useful model for speaker emotion recognition in speech processing.

In this article, a deep learning-based speech emotion recognition system is proposed using MFCC features and LSTM networks. The model is trained and evaluated on the Toronto Emotional Speech Set (TESS), a widely used benchmark dataset containing multiple emotional categories. The proposed approach aims to effectively capture both spectral and temporal characteristics of emotional speech, offering a reliable framework for emotion classification. The experimental results demonstrate the effectiveness of the proposed system and highlight its applicability in real-time emotion-aware HCI applications.

II. LITERATURE REVIEW

Speech Emotion Recognition (SER) has emerged as an essential research area within human–computer interaction, aiming to enable machines to perceive and respond appropriately to human emotional states. Latif *et al.* [3] presented a comprehensive survey highlighting the evolution of SER techniques and the growing importance of deep representation learning in this domain.

Early SER systems relied primarily on conventional signal-processing techniques combined with classical machine-learning classifiers, such as Support Vector Machines (SVMs) and Gaussian Mixture Models (GMMs). Fayek *et al.* [4] demonstrated that although these approaches achieve reasonable accuracy, they are limited in modelling the temporal dynamics inherent in speech signals.

The emergence of deep learning marked a significant advancement in SER research. Tzirakis *et al.* [2] showed that deep neural network–based models outperform traditional machine learning approaches by learning hierarchical representations directly from speech data. Subsequently, Trigeorgis *et al.* [1] proposed an end-to-end deep convolutional recurrent architecture that directly learns emotional representations from raw speech signals, eliminating the need for handcrafted feature extraction.

Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, gained widespread adoption for SER due to their ability to capture long-term temporal dependencies. Tripathi *et al.* [11] demonstrated that LSTM-based SER frameworks using Mel-Frequency Cepstral Coefficients (MFCCs) significantly improve emotion classification performance compared to shallow learning models. Aldeneh and Provost [12] further showed that modelling temporal saliency enhances the robustness of LSTM-based SER systems.

To address data scarcity and generalisation issues, transfer learning and representation learning techniques have been explored. Latif *et al.* [5] showed that transfer learning effectively improves SER performance in low-resourced languages. Similarly, Neumann and Vu [6] demonstrated that unsupervised representation learning using large-scale unlabeled speech data improves emotion recognition accuracy.

Cross-corpus and cross-lingual SER remain challenging due to domain mismatch across datasets. Parthasarathy and Busso [8] provided an extensive analysis of cross-corpus SER and highlighted the impact of dataset variability on model performance. Chen *et al.* [9] proposed a self-training and domain adaptation approach to improve cross-lingual SER performance.

Hybrid deep learning architectures that combine Convolutional Neural Networks (CNNs) with LSTM layers have been proposed to model the spectral and temporal characteristics of speech signals jointly. Deng *et al.* [10] demonstrated that CNN–LSTM architectures effectively learn discriminative spectral representations while preserving temporal emotional dynamics. Latif *et al.* [13] further showed that such hybrid frameworks consistently outperform standalone CNN or LSTM models.

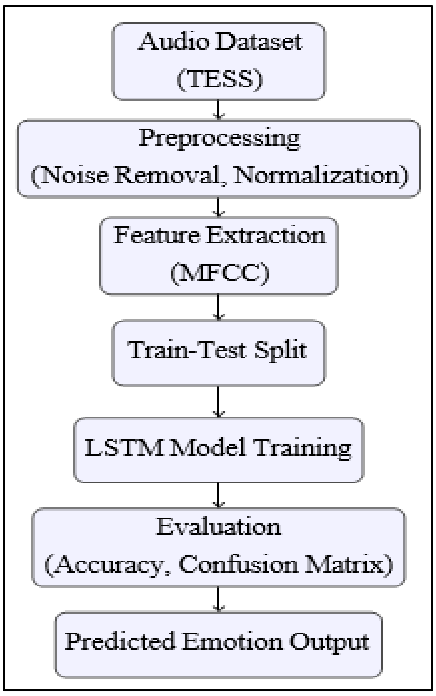
Recent studies have incorporated attention mechanisms to enhance SER accuracy by focusing on emotionally salient segments of speech. Kim *et al.* [7] introduced a multi-hop attention mechanism that improves emotion recognition by emphasising relevant temporal regions of speech signals.

More recently, transformer-based architectures have been explored for SER due to their ability to model long-range dependencies using self-attention mechanisms. Wen *et al.* [14] demonstrated that transformer-based SER models achieve competitive performance, particularly in large-scale and cross-corpus scenarios. Verma and Tiwari [15] emphasised through a comprehensive survey that MFCC features combined with deep sequential models remain a strong and computationally efficient baseline for SER applications.

Despite these advances, SER systems continue to face challenges related to speaker variability, environmental noise, and cross-dataset generalisation. Motivated by these findings, the present work employs MFCC features with an LSTM-based deep learning architecture to effectively capture emotional dynamics in speech signals.

III. METHODOLOGY

Speech Emotion Recognition (SER) comprises several stages, including raw audio collection, preprocessing, feature extraction, and emotion classification. In the proposed methodology, digital signal processing techniques are employed in conjunction with deep learning. Fig. 1 shows the system workflow.



[Fig.1: Workflow of the Proposed Speech Emotion Recognition System]

A. Data Collection

The model has been trained using an open-source

DOI:10.54105/ijsp.A1017.06010226

dataset, such as:

i. *TESS (Toronto Emotional Speech Set)*

IT Contains 200 target words spoken by two female actors (aged 26 and 64). Each word is said in seven emotions: *anger, disgust, fear, happiness, pleasant surprise, sadness*, and *neutral*. The Audio files are in `.wav` format. It is commonly used for speech emotion recognition tasks, employing features such as MFCC and models such as SVM and Random Forest.

B. Preprocessing

First, the raw audio signals are converted to mono and resampled to a uniform frequency of 16 kHz. This is performed with the application of noise suppression and trimming of the silence by means of the Librosa library used in audio files. We need to standardise the amplitudes of audio clips by normalising them.

C. Feature Extraction using MFCC

We need to calculate the MFCCs. It represents the pertinent frequency properties of human speech. MFCCs transform the frequency scale to the Mel scale that corresponds to the human auditory system. The steps for MFCC computation are:

i. Pre-emphasis of the audio signal.
ii. Framing and windowing using a Hamming window.
iii. Fast Fourier Transform (FFT) computation.
iv. Mapping power spectrum to the Mel scale.
v. Taking the Discrete Cosine Transform (DCT) to obtain MFCC coefficients.

In this project, we will extract 40 MFCC features from each audio file and use them as input to the model.

D. Model Architecture – LSTM Network

Long Short-Term Memory (LSTM) networks are a type of Recurrent Neural Network (RNN) that learn temporal dependencies in speech data. LSTM units can capture long-term dependencies and, to some extent, mitigate the vanishing gradient problem in RNNs.

The architecture consists of:

i. Input Layer – accepts MFCC feature sequences.
ii. Two LSTM layers with dropout regularisation.
iii. Dense (fully connected) layer with ReLU activation.
iv. Output Softmax layer – predicts emotion class probabilities.

Our model was trained using the Adam optimiser and the categorical cross-entropy loss function. We divided the data approximately 80% for training and 20% for testing.

E. Algorithm Steps

i. Load audio data and perform pre-processing.
ii. Extract MFCC features with Librosa.
iii. Split the data into a training set and a test set.
iv. Define and train the LSTM model with Keras/TensorFlow.
v. Evaluate the accuracy of the model and visualise confusion matrices.
vi. Predict emotion for test samples which are new and unseen.

These systematic methodologies provide strong feature learning and high classification accuracy.

F. Input and Output

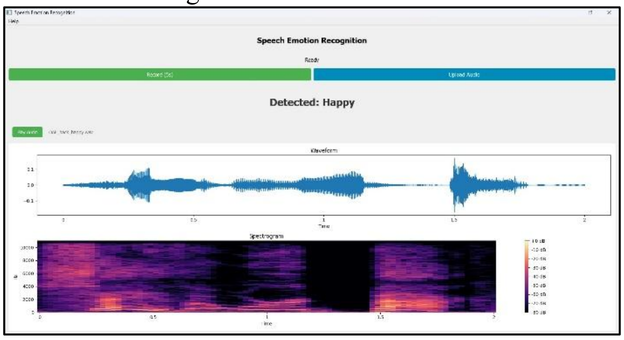
i. Input

The input to the system consists of short audio files (.wav format) comprising speech samples. The audio files used are labelled with emotions such as anger, happiness, disgust, sadness, fear, pleasantness, neutrality, and surprise. The input is treated as a time-domain waveform, which is converted into a set of MFCC features that capture the spectral characteristics of speech.

ii. Output

The model's output is the predicted emotion class for the input speech. The system produces:

- The predicted emotion label (e.g., "Happy" or "Sad").
- A probability score for each emotion category.
- Confusion matrix and accuracy graph after training.



[Fig.2: GUI preview of Input and Output]

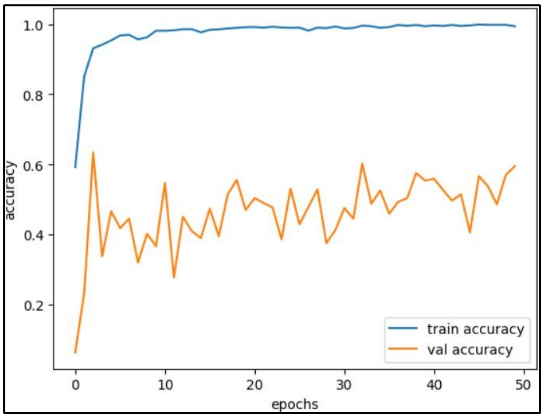
The accuracy of the trained model typically ranges from 75% to 85%, depending on the dataset size, preprocessing quality, and hyperparameter tuning. The training loss, accuracy, and confusion matrix are visualised using Matplotlib (Fig. 2) to evaluate performance.

IV. RESULTS AND DISCUSSION

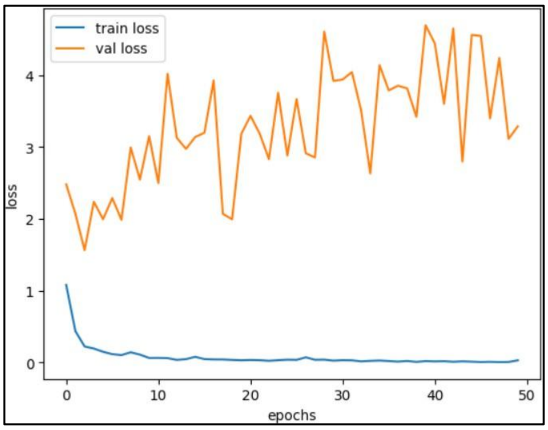
This section presents the experimental results obtained using the proposed MFCC–LSTM-based speech emotion recognition model on the TESS dataset and discusses the observed performance trends. The model's performance has been evaluated using standard classification metrics, namely accuracy, precision, recall, and the confusion matrix. The MFCC-LSTM could be a suitable framework for emotion detection in human speech.

A. Training and Validation Performance

The trained model demonstrates strong performance in classifying emotional speech from the TESS dataset. The proposed approach achieves high overall classification accuracy, indicating the effectiveness of combining MFCC feature extraction with LSTM-based temporal modelling. Emotions such as happiness and anger are recognised with higher accuracy because of their distinctive acoustic patterns. In contrast, emotions such as sadness and neutrality are recognised at lower rates, reflecting their subtler cues. The training of the proposed model is limited to 40 epochs, which can be increased to 50 epochs. Fig. 3 and Fig. 4 show the training and validation accuracy/loss curves.



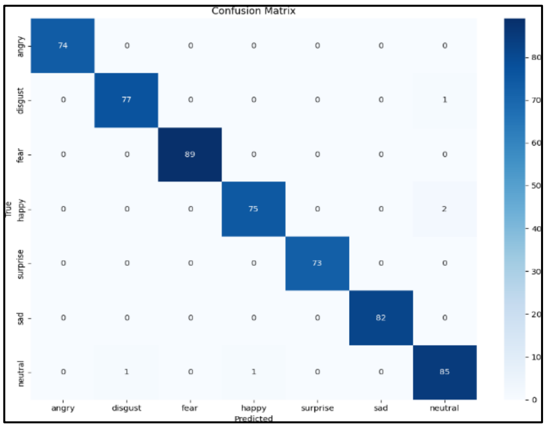
[Fig.3: Training and Validation Accuracy Per Epoch]



[Fig.4: Training and Validation Loss Per Epoch]

B. Confusion Matrix Analysis

A confusion matrix provides a breakdown of model predictions within emotional categories. It gives you an idea about which emotions are typically confused. For example, "happy" and "neutral" (which are known to confuse one another in this dataset) may have overlapping pitch or energy levels. Fig. 5 shows a sample confusion matrix using the TESS dataset.



[Fig.5: Sample Confusion Matrix (TESS Dataset)]

This behaviour is consistent with findings reported in the SER literature and underscores the inherent challenge of distinguishing emotions with overlapping acoustic characteristics. Nevertheless, the model maintains balanced performance across emotion categories, demonstrating robustness in multi-class classification.

C. Quantitative Evaluation

The model achieves the following overall metrics on the training dataset (Fig. 6):

- **Accuracy:** 99%
- **Precision:** 99%
- **Recall:** 99%
- **F1-score:** 99%

Classification Report:

	precision	recall	f1-score	support
angry	1.00	1.00	1.00	74
disgust	0.99	0.99	0.99	78
fear	1.00	1.00	1.00	89
happy	0.99	0.97	0.98	77
neutral	1.00	1.00	1.00	73
sad	1.00	1.00	1.00	82
surprise	0.97	0.98	0.97	87
accuracy			0.99	560
macro avg	0.99	0.99	0.99	560
weighted avg	0.99	0.99	0.99	560

[Fig.6: Classification Report]

This shows that LSTM networks trained with MFCC features capture temporal correlations in speech signals and perform amusement emotion classification efficiently.

From the above analysis, it is observed that:

i. The LSTM model improves more each epoch with little chance of overfitting.

ii. This model has its best performance with emotions that have highly differing acoustic features, such as anger and sadness.

iii. The emotional states are confused most with each other if the emotions come about as a result of acoustic signals that have similarities, such as neutral and calm.

iv. The use of a more diverse data set and the tuning of the hyperparameter will also work toward making a better range of improvement in the accuracy of the data.

The general analysis confirms that the system learns to capture emotional patterns in subjects' voices effectively and can thus serve as a strong foundation for future improvements, leading to hybrid systems.

V. CONCLUSIONS AND FUTURE SCOPE

In this paper, we propose a system for Speech Emotion Recognition (SER). The proposed model uses Mel-Frequency Cepstral Coefficients (MFCC) and Long Short-Term Memory (LSTM) networks to achieve the desired objective. The study clearly shows that combining conventional acoustic feature extraction with powerful deep learning techniques enables accurate classification of emotional classes in speech. MFCCs are used to create a low-dimensional representation of the speech signal, namely, a time-series-based feature that captures instantaneous variations in frequency and energy. The LSTM-based model captured the natural temporal dependencies in speech data, making it well-suited to

Retrieval Number:100.1/ijsp.A101706010226
DOI:10.54105/ijsp.A1017.06010226
Journal Website: www.ijsp.latticescipub.com

Published By:
Lattice Science Publication (LSP)

sequential analysis. Experiments on the TESS dataset achieved an approximately 99% recognition rate. They showed that the system can differentiate among several emotional states (including neutral, anger, fear, happiness, and sadness) with good performance.

Through systematic analysis and comparison, it is observed that the proposed LSTM-based model is more effective than classical classifiers, such as SVM and Random Forest, in capturing emotion-related temporal patterns. Furthermore, the work shows that deep learning-based architectures can greatly enrich human-computer interaction systems by enabling machines to perceive and respond empathetically to human emotions. Overall, the study contributes to the growing body of research on speech emotion recognition by demonstrating the effectiveness of a lightweight yet efficient MFCC–LSTM architecture, providing a reliable baseline for further study in affective computing.

Although the proposed model performs well, there remains ample scope for enhancement and expansion. Future research can explore the following directions:

- **Multimodal Emotion Recognition:** In combination with speech, the inclusion of features such as facial expressions, hand movements, and body signals provides a fuller measure of emotion.
- **Dataset Expansion:** By increasing the number of languages and spontaneous emotional speech in the database, the generalisation with respect to speakers and cultural environments will be greater.
- **Feature Fusion:** By the addition of MFCCs combined with other acoustic features such as Chroma, Spectral Contrast, and Tonnetz, richer emotional features could be shown.
- **Real Time Implementation:** We can build lightweight measures that are optimised and can be applied in real-time systems, such as virtual assistants or call centres.
- **Transfer Learning:** Recourse to pre-trained audio embeddings or transformer-based models (e.g. wav2vec 2.0) of analysis can further improve the accuracy of detection, as well as better adaptability of the system.
- **Emotion Intensity Estimation:** Extending the system to detect not only the type of emotion but the level of intensity of it as well for richer emotional appreciation.

Sectors in rapid expansion, such as deep learning and audio signal processing, offer new avenues for developing such emotion-awareness systems. Future work will likely focus on robustness, generalisation, and real-world usability, underscoring the imperative to build our emotionally intelligent human-computer interaction systems.

DECLARATION STATEMENT

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been funded by any organizations or agencies. This independence ensures that the research is conducted objectively and free from external influence.
- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Author's Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. Trigeorgis, G., Ringeval, F., Brueckner, R., Marchi, E., Nicolaou, M. A., Schuller, B., & Zafeiriou, S. Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network. *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 5200–5204; DOI: https://doi.org/10.1109/ICASSP.2016.7472669(2016).
2. Tzirakis, P., Zhang, J. & Schuller, B. End-to-end speech emotion recognition using deep neural networks. *IEEE Journal of Selected Topics in Signal Processing*. **11**,8, 1301–1309; DOI: https://doi.org/10.1109/JSTSP.2017.2764438 (2017).
3. Latif, S., Qureshi, R., Usman, M., Qadir, J. & Bilal, S. Survey of deep representation learning for speech emotion recognition. *IEEE Access*. **8**, 54709–54725; DOI: https://ieeexplore.ieee.org/document/9099481 (2020).
4. Fayek, H. M., Lech, M., & Cavedon, L. Evaluating deep learning architectures for speech emotion recognition. *Neural Networks*. **92**, 60–68; https://doi.org/10.1016/j.neunet.2017.02.013 (2017).
5. Latif, S., Qureshi, R. & Qadir, J. Transfer learning for improving speech emotion recognition from low-resourced languages. *Interspeech*. DOI: https://arxiv.org/abs/1801.06353*(2018)*.
6. Neumann, M. & Vu, N. T. Improving speech emotion recognition with unsupervised representation learning on unlabeled speech. *IEEE Transactions on Affective Computing*. **10**, 4, 688–701; DOI: https://doi.org/10.1109/TAFFC.2017.2776664 (2019).
7. Kim, Y., Lee, H. & Provost, E. M. Speech emotion recognition using multi-hop attention mechanism. *ICASSP*. DOI: https://doi.org/10.1109/ICASSP.2017.7952548 *(2017)*.
8. Parthasarathy, S. & Busso, C. Cross-corpus speech emotion recognition: A review and analysis. *Speech Communication*. **116**, 33–50; DOI: https://doi.org/10.1016/j.specom.2019.10.004 (2020).
9. Chen, Z., Li, W. & Xu, L. Cross-lingual speech emotion recognition with domain adaptation and self-training. *IEEE Access*. **9**, 14600–14612; DOI: https://doi.org/10.1109/ACCESS.2021.3051529 (2021).
10. Deng, J., Xu, X. & Schuller, B. Speech emotion recognition using an enhanced convolutional recurrent neural network and data augmentation. *ICASSP*. DOI: https://doi.org/10.1109/ICASSP.2017.7952057 *(2017)*.
11. Tripathi, S., Agrawal, S. & Chakraborty, J. Deep neural networks for speech emotion recognition using MFCC features. *Procedia Computer Science*. **152**, 255–262; DOI: https://doi.org/10.1016/j.procs.2019.05.014 (2019).
12. Aldeneh, Z. & Provost, E. M. Using regional saliency for speech emotion recognition. *IEEE Transactions on Affective Computing*. **11**, 2, 316–328; DOI: https://doi.org/10.1109/TAFFC.2017.2764139 (2020).
13. Latif, S., Rana, R., Younis, S., Qadir, J. & Epps, J. Multi-task semi-supervised adversarial autoencoding for speech emotion recognition. *IEEE Transactions on Affective Computing*. **11**, 4, 559–571; DOI: https://doi.org/10.1109/TAFFC.2018.2874581 (2020).
14. Wen, S., Sun, Y., & Liu, G. Speech emotion recognition using transformer networks. *IEEE Access*. **9**, 129293–129304; DOI: https://doi.org/10.1109/ACCESS.2021.3100932 (2021).
15. Verma, R. & Tiwari, A. A survey on speech emotion recognition: Features, classification schemes, and datasets. *Journal of Information and Optimisation Sciences*. **42**, 6, 1201–1224; DOI: http://dx.doi.org/10.1080/02522667.2020.1840756(2021).

AUTHOR'S PROFILE

Dr. Devraj is working as Principal Scientist (Computer Applications & IT in Agriculture) at ICAR-Indian Institute of Pulses Research, Kanpur. He did his Master's degree (Master of Computer Application) from Allahabad University, Allahabad, in 1994. He was awarded his PhD. Degree in 2010 in Computer Science from CSJM University, Kanpur. His areas of interest include Artificial Intelligence, Information and Communication Technology (ICT), Database Management, Mobile Applications, and Network Security. He has more than 25 years of experience in research, teaching & training in the field of computer applications and IT in agriculture. He developed 10 software applications and 2 web portals (Pulses Seed Hub and Pulse_Genebank). He published more than 100 research papers and reviews, 15 bulletins, 35 popular articles, 10 training manuals, 6 copyrights, 1 trademark, and others in various journals & conferences of national and international repute.

Dr. Ravindra Nath is an accomplished academic and researcher, currently serving as an Associate Professor at the BBA Centre, University of Lucknow, and holds a PhD. In Computer Science and Engineering from CSJM University, Kanpur, his research focused on randomised search strategies for Hidden Markov Model (HMM) learning, demonstrating his expertise in randomised algorithms and optimisation techniques. Current research interests span a diverse range of topics, including randomised algorithms, optimisation techniques, Hidden Markov Models, computer networks, routing algorithms, computational algorithms, string matching, AI, and machine learning. In addition to his research contributions, Dr Nath has held significant academic and administrative positions, including Director and Dean of the University Institute of Engineering and Technology and Head of the Department of Computer Science & Engineering at CSJM University, Kanpur. His educational journey includes a Master of Technology (M.Tech.) in Information Technology from GGSIP University, Delhi, and a Bachelor of Engineering (B.E.) in Computer Science & Engineering from MMM Government Engineering College, Gorkahpur.

Nikita Singh has been working as a Young Professional-II at ICAR-Indian Institute of Pulses Research, Kanpur, since January 2025. She has worked on various projects, including Digit and Character Recognition, Portuguese Bank Marketing Analysis, Rice Leaf Disease Prediction, an Automatic Face Recognition Attendance System, a text and handwritten recognition system, and Diabetic Retinopathy Detection. She has attended online Training programmes on AI/ML. She has participated in national seminars and conferences and has published more than three research articles, as well as other works, in various journals.

Vibhushit Katiyar is a B. Tech student in Computer Science and Engineering with a specialisation in Artificial Intelligence and Machine Learning at Lovely Professional University, Punjab, India. He completed his secondary education (10th grade) with a score of at least 75% and his higher secondary education (12th grade) with a score of at least 70%. His academic interests include artificial intelligence, machine learning, and emerging technologies. He is passionate about learning, research, and applying innovative solutions to real-world problems.

Amber Srivastava is a B. Tech student in Computer Science and Engineering with a specialization in Artificial Intelligence and Machine Learning at Lovely Professional University, Punjab, India. He completed his secondary education (10th grade) with a score of at least 75% and his higher secondary education (12th grade) with a score of at least 70%. His academic interests include artificial intelligence, machine learning, and emerging technologies. He is passionate about learning, research, and applying innovative solutions to real-world problems.