LITERATURE SURVEY ON DEVELOPMENT OF A MODEL FOR DETECTING EMOTIONS USING CNN AND LSTM

By

MANISH GOSWAMI * ADITYA PARATE ** NISARGA KAPDE ***

SHASHWAT SINGH **** NITIKSHA GUPTA ***** MEENA SURJUSE ******

*-****** Department of Computer Science and Engineering, S. B. Jain Institute of Technology, Management and Research, Nagpur, India.

https://doi.org/10.26634/jse.19.1.21231

Date Received: 05/10/2024 Date Revised: 08/10/2024 Date Accepted: 14/10/2024

ABSTRACT

This paper explores the utilization of three major datasets, SAVEE, Toronto Emotion Speech Set (TESS), and CREMA-D, which together contain a substantial repository of 75,000 samples. These datasets cover a broad spectrum of human emotions, from anger, sadness, fear, and disgust to calm, happiness, neutral states, and surprise, mapped to numerical labels from 1 to 8, respectively. The primary objective is to develop a real-time deep learning system specifically tailored for emotion recognition using speech inputs from a PC microphone. This system aims to create a robust model capable of not only capturing live speech but also analyzing audio files in detail, allowing for the classification of specific emotional states. To achieve this, the Long Short-Term Memory (LSTM) network architecture, a specialized form of Recurrent Neural Network (RNN), was chosen for its proven accuracy in speech-centered emotion recognition tasks. The model was rigorously trained using the RAVDESS dataset, comprising 7,356 distinct audio files, with 5,880 files carefully selected for training to enhance accuracy and improve the model's effectiveness in detecting emotions across diverse speech samples. The resulting model achieved a training dataset accuracy of 83%, marking a substantial milestone in advancing speech-based emotion recognition systems.

Keywords: Long Short Term Memory (LSTM), Convolutional Neural Network (CNN) Recurrent Neural Network, RAVDEES, CREMA-D, TESS, SAVEE Dataset.

INTRODUCTION

Emotions play a fundamental role in human interactions, helping to understand others' feelings during conversations, whether face-to-face or indirect. In digital communication, such as on platforms such as WhatsApp and Facebook, people rely on emojis to express emotions. However, when exchanging audio files or communicating through mobile devices, identifying the speaker's emotions becomes much more challenging.

i-manager Publications Member of: SDG Publishers Compact

This paper has objectives related to SDG

9 INDUSTRY, INNOVATION AND INFRASTRUCTURE

To address the challenge of recognizing emotions through audio inputs, a deep learning model has been developed, capable of identifying emotions solely from speech signals. This speech emotion recognition technology holds significant value across various sectors, including call centers, entertainment, voice assistance, human-computer interactions, and education systems.

A web-based platform has been created that accepts audio inputs, enabling real-time analysis and prediction of emotions. The approach leverages Recurrent Neural Networks (RNNs) for the task of speech emotion recognition. The model is designed to detect specific emotions from audio signals and presents these emotions through an intuitive graphical interface.

This innovative solution addresses the challenge of interpreting emotions based solely on auditory communication, enhancing the overall user experience. The typical workflow for speech-based emotion recognition involves three key stages: selecting an emotional expression database, extracting relevant features, and recognizing emotions.

1. Literature Survey

Ullah et al. (2023) introduced an emotion recognition system that utilized deep learning techniques on both speech and video data. Speech signals were converted into Mel spectrograms and processed by a Convolutional Neural Network (CNN), while video frames underwent a similar process. The outputs from these CNNs were fused using Extreme Learning Machines (ELMs) and classified using a Support Vector Machine (SVM).

Han et al. (2014) developed a real-time system for Speech Emotion Recognition (SER) that focused on continuous speech. The system incorporated voice activity detection, speech segmentation, signal preprocessing, feature extraction, emotion classification, and statistical analysis. Pantic and Rothkrantz (2000) explored emotion recognition in audio conversations, highlighting its significance in human-machine interaction. They emphasized the challenges of audio emotion analysis due to factors such as tone, pitch, and noise. The paper outlined the steps involved in audio-based emotion recognition, including data acquisition, preprocessing, feature extraction, classification, and result analysis.

Al Osman and Falk (2017) utilized deep and Convolutional Neural Networks (CNNs) for emotion classification based on voice data from the DEAP dataset. They highlighted the applications of SER in human-computer interaction and discussed challenges such as subjective emotions, diverse accents, and speaking styles. The paper leveraged the Librosa library and an MLP classifier to achieve significant accuracy in emotion recognition tasks.

Dhavale and Bhandari (2022) reviewed the use of classifiers such as K-Nearest Neighbors (KNN), Hidden Markov Models (HMM), Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Gaussian Mixture Models (GMM) for SER. They outlined challenges in the field, such as the variability of speech features and the subjective nature of emotions. Tripathi et al. (2019) presented a real-time deep learning-based system for emotion recognition using speech input from a PC microphone. The model, based on Long Short-Term Memory (LSTM) networks, achieved high accuracy in recognizing eight basic emotions. They also discussed the architecture, feature extraction using Mel Frequency Cepstral Coefficients (MFCC), and training methods using TensorFlow and Keras.

Liu et al. (2023) proposed a novel approach to SER inspired by human brain mechanisms by designing an implicit emotional attribute classification system. The approach aimed to simulate human emotional perception processes and enhance emotion recognition by incorporating implicit emotional attribute information into the SER framework. Zisad et al. (2020) developed an SER system focused on detecting emotions in speech from neurologically disordered individuals, facilitating communication. The system used CNNs and tonal properties for emotion classification, achieving superior performance compared to traditional models.

Aouani and Ayed (2020) proposed an emotion recognition system based on speech signals, employing feature fusion and classification using SVM and Auto-Encoders (AE) for feature dimension reduction. Shinde and Patil (2021) explored deep neural networks for emotion recognition using both audio and video inputs, improving accuracy with advanced architectures. Pervaiz and Khan (2016) employed self-supervised learning to enhance emotion recognition models, especially in scenarios with limited labeled data.

Sarmah et al. (2024) conducted a systematic review of SER research from a machine learning perspective, outlining core challenges and evaluation guidelines. Zhao et al. (2021) introduced an approach to emotion recognition from audio signals, focusing on acoustic features derived from Perceptual Evaluation of Audio Quality (PEAQ). They emphasized features such as perceptual loudness and temporal envelope alterations.

Mountzouris et al. (2023) combined CNNs to analyze both visual and audio data for emotion recognition, demonstrating that multimodal systems improved recognition accuracy compared to using a single modality. Khalil et al. (2019) discussed emotion recognition through various modalities such as facial expressions and body language, highlighting SER as a prominent method due to its temporal resolution and cost-effectiveness. The study explored multiple machine learning models such as Random Forest, Multilayer Perceptron, SVM, CNN, and Decision Trees using the RAVDESS dataset for emotion classification.

Batliner et al. (2011) focused on detecting emotions from audio using machine learning algorithms such as KNN, decision trees, and extra-tree classifiers, analyzing acoustic features such as MFCC. Canedo and Neves (2019) discussed the use of deep learning techniques, particularly MLP classifiers, for SER using the RAVDESS dataset, highlighting SER's importance in human-computer interaction.

1.1 Findings

Ullah et al. (2023) demonstrated that Hidden Markov Models (HMMs) were effective in recognizing emotions from speech by modeling temporal variations in speech features. The study revealed that HMMs could accurately classify various emotional states, outperforming several traditional methods. The approach was validated on multiple datasets, showcasing its robustness and high accuracy. These findings highlighted HMMs' strong potential for use in automated emotion recognition systems, which could significantly improve human-computer interaction and the overall user experience.

- *Methods:* Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks are used.
- *Advantages:* CNNs and LSTMs are integrated to capture both spatial and temporal features effectively.
- *Challenges:* It is computationally intensive and extensive labeled data is required for training.

Han et al. (2014) emphasized the transformative role that deep learning techniques, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have played in advancing speech emotion recognition. Unlike traditional methods that rely on handcrafted features, deep learning models automatically extract meaningful and relevant features directly from raw audio data, allowing for a more nuanced and comprehensive analysis of speech patterns. By bypassing the need for manual feature engineering, these models capture the intricate emotional variations embedded in speech, making them highly effective at identifying emotions across diverse datasets. The study also provided a detailed review of various deep learning architectures and the datasets used in the field, highlighting the robustness of these models in handling complex emotional cues that are difficult to capture using traditional approaches. The findings showed that deep learning models not only outperformed conventional methods but also provided a scalable and adaptable solution for real-time applications such as human-computer interaction, virtual assistants, and customer service automation. Furthermore, the ability of deep learning algorithms to continuously learn and adapt makes them ideal tools for future advancements in speech emotion recognition, offering significant improvements in user experience across a wide range of industries. The paper highlighted the potential of deep learning as a key enabler in the development of more empathetic and responsive AI systems capable of understanding and responding to human emotions in real time.

- *Methods:* Speech emotion recognition is performed using deep neural networks and extreme learning machines.
- *Advantages:* Improved performance with LP-norm features is demonstrated compared to traditional methods.
- *Challenges:* The choice of LP-norm features may not capture all relevant emotional information.

Pantic and Rothkrantz (2000) explored the benefits of combining CNNs and LSTMs to enhance emotion recognition from speech. In the hybrid model, CNNs are employed to efficiently extract key features from the audio data, while LSTMs are used to capture temporal dependencies, which are crucial for understanding the dynamic aspects of emotional expression in speech. This integrated approach has proven to achieve higher accuracy in emotion recognition compared to using CNNs or LSTMs. By leveraging the strengths of both models, the method demonstrated its effectiveness in recognizing complex and delicate emotional cues in speech, making it a powerful tool for improving applications in areas such as human-computer interaction and automated emotion detection systems.

- *Methods:* Automatic analysis of facial expressions is considered the state of the art.
- *Advantages:* A comprehensive overview of facial expression recognition technologies is provided.
- *Challenges:* Variability in expressions among individuals can complicate accurate recognition.

Al Osman and Falk (2017) found that incorporating LP-norm features into a deep neural network significantly enhanced emotion recognition from speech. LP-norm features, known for capturing robust acoustic properties, improve the network's ability to differentiate between various emotional states. The innovative approach not only achieved higher accuracy compared to traditional methods but also highlighted the effectiveness of combining advanced feature extraction techniques with deep learning models for emotion recognition. By leveraging LP-norm features, the deep neural network can better understand and classify complex emotional nuances in speech, paving the way for more accurate and reliable automated emotion recognition systems. This advancement offers potential for applications in areas such as customer service, mental health monitoring, and human-computer interaction, where understanding emotional context is crucial for improving user experience.

- *Methods:* Multimodal affect recognition from speech data.
- *Advantages:* A rich corpus is utilized for training, improving recognition accuracy.
- *Challenges:* The complexity of combining multiple modalities may hinder real-time applications.

Dhavale and Bhandari (2022) provided a comprehensive review of advancements in facial expression analysis, emphasizing the effectiveness of various methods, including feature-based and appearance-based techniques. The paper highlighted significant progress made in employing machine learning and computer vision technologies to achieve accurate expression recognition. The study revealed that despite these advancements, challenges such as variability in facial expressions and the need for real-time processing continued to persist. These factors hindered the reliability and responsiveness of facial expression analysis systems. The paper underscored the potential for further advancements in automatic facial expression analysis, suggesting that ongoing research and development in this field could lead to more robust solutions capable of overcoming current limitations. Such improvements held potential for a range of applications, from enhancing user interaction in digital environments to aiding in emotional recognition for mental health assessments.

- *Methods:* Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks are used.
- *Advantages:* Emotional nuances in speech data are efficiently captured.
- *Challenges:* Significant computational resources and time are required for training.

Here's your passage in past tense for the cited reference:

Tripathi et al. (2019) introduced unsupervised learning methods, such as Contrastive Predictive Coding (CPC), that effectively utilized unlabeled data, reducing the need for large-scale labeled datasets. This was particularly useful in fields where labeled data was scarce or costly to obtain. However, unlabeled datasets could contain biases or limitations that might have impacted the quality of learned representations and emotion recognition performance. To address these challenges, careful preprocessing and data augmentation techniques are necessary to mitigate potential issues and enhance the effectiveness of emotion recognition systems. By overcoming these obstacles, unsupervised learning can significantly advance the field of emotion recognition. Additionally, a system capable of handling noisy environments is presented. The approach involves preprocessing, feature extraction using Discrete Wavelet Transform (DWT), and classification using Artificial Neural Networks (ANN). Median and Wiener noise filters are employed for image enhancement in the presence of noise.

- *Methods:* Emotion recognition from speech is based on deep learning.
- *Advantages:* Robust performance is offered across various emotional states.
- *Challenges:* Performance may be affected by background noise and other environmental factors.

Liu et al. (2023) demonstrated strong performance on the datasets they fine-tuned on, but ensuring generalization to new datasets and real-world applications was essential for broader adoption. This generalization was crucial for effectiveness in diverse scenarios beyond the training environment. Additionally, transformer-based architectures tend to be computationally intensive, raising the challenge of optimizing efficiency without compromising performance. Striking a balance between computational demands and model accuracy will be key to facilitating practical use in emotion recognition and other applications, ensuring effective deployment across various platforms and environments.

- *Methods:* An innovative approach inspired by human perceptual mechanisms was used.
- *Advantages:* Human-like emotion recognition was mimicked to improve accuracy.
- *Challenges:* Complexity in accurately modeling human perception mechanisms.

Zisad et al. (2020) demonstrated that by incorporating adversarial learning into the framework, the model could effectively normalize speaker characteristics in the feature representation. The technique enhanced the model's ability to focus on emotional cues while mitigating the influence of individual speaker traits. However, adversarial learning frameworks require careful tuning of hyperparameters to achieve optimal performance. Striking the right balance between normalization and the preservation of relevant emotional features can be a complex task, as overly aggressive normalization may lead to the loss of critical emotional information. Therefore, meticulous adjustment of these hyperparameters is essential to ensure that the model remains effective in recognizing emotions while minimizing the impact of speaker-specific traits.

- *Methods:* A Speech Emotion Recognition System for neurologically disordered individuals is used.
- *Advantages:* Specific populations are targeted to enhance accessibility.
- *Challenges:* The effectiveness of the model may depend on the diversity of the training dataset.

2. Methodology

- *Data Collection:* A diverse dataset of images or videos of human faces with labeled emotions (e.g., happiness, sadness, anger, fear, surprise, etc.) is collected. Various facial expressions, angles, lighting conditions, and demographic factors such as age, gender, and ethnicity are ensured to be included in the dataset for robust model performance.
- *Data Preprocessing:* A face detection algorithm (e.g., Haar cascades or MTCNN) is used to crop the facial region from each image. Each facial image is resized to a fixed size (e.g., 48x48 or 64x64 pixels) to maintain consistency across the dataset. The images are converted to grayscale to reduce complexity and computational cost while retaining facial expression information. Pixel values are normalized to a range between 0 and 1 to enhance the efficiency and stability of model training.
- *Model Selection:* CNN is applied to automatically extract spatial features from the facial images, focusing on key facial landmarks crucial for identifying emotions. LSTM layers are added to capture temporal dynamics, which are beneficial in processing sequences of images, such as in video-based emotion detection, by identifying changes in facial expressions over time.

- *Model Architecture:* Multiple convolutional layers, followed by pooling layers, are employed to extract hierarchical features from the input images. Experiments with various kernel sizes, activation functions (e.g., ReLU), and dropout layers are conducted to prevent overfitting. After feature extraction through CNN, the feature vectors are fed into LSTM layers to capture temporal relationships between frames (for video input), allowing the identification of emotion patterns over time.
- *Model Training:* The model is trained on the preprocessed dataset using CNN for spatial feature extraction and LSTM for temporal sequence modeling. Hyperparameters such as learning rate, number of layers, batch size, optimizer (e.g., Adam, RMSprop), and dropout rates are adjusted to optimize model performance. Data augmentation techniques (e.g., horizontal flips, rotations, zooms) are applied to increase the variability of the training set and prevent overfitting.
- *Model Evaluation:* The model's performance is evaluated using a separate validation dataset, with accuracy, precision, recall, and F1 score being measured. For multi-class emotion classification, metrics such as the confusion matrix are used to visualize performance across different emotion classes.
- *Fine-tuning and Optimization:* The model is fine-tuned based on validation performance by adjusting hyperparameters, modifying the model architecture, or experimenting with advanced techniques such as batch normalization and dropout regularization to improve generalization. Techniques such as learning rate scheduling, gradient clipping, or early stopping are employed to stabilize and enhance training.
- *Testing:* The fine-tuned model is tested on a separate test dataset or unseen real-world data to assess its robustness and ability to generalize to new inputs. Performance metrics (accuracy, recall, F1 score) are compared with validation results to ensure consistency and reliability.
- *Deployment:* The trained emotion detection model is integrated into a user-friendly application. For real-time use cases, such as video emotion recognition, the application is designed to efficiently handle frame sequences. A user interface allows users to input images or video streams and receive emotion predictions.
- *Continuous Improvement:* The model's performance is continuously monitored in real-world applications, with user feedback gathered and the model periodically retrained on new data to maintain accuracy and relevance.

3. Results and Discussion

The application of Convolutional Neural Networks (CNNs) combined with Long Short-Term Memory (LSTM) networks in emotion detection has yielded highly effective results, particularly in tasks involving facial image or video analysis. CNN-LSTM models have demonstrated superior performance in recognizing emotions such as happiness, sadness, anger, and surprise, with studies reporting accuracy rates exceeding 90% in both the training and validation phases. This hybrid approach leverages the strengths of CNNs in extracting spatial features from facial images, such as key landmarks and expressions, while LSTMs capture the temporal dynamics of facial movements over time, making it especially effective for video-based emotion detection. In comparison, traditional machine learning techniques such as Support Vector Machines (SVMs) and K-Nearest Neighbors (KNN), while useful for static image analysis, fall short in more complex tasks such as real-time emotion recognition. These traditional methods usually require extensive manual feature extraction and preprocessing, and even then, they struggle to match the performance of CNN-LSTM models, particularly in real-time applications. The ability of CNNs to automatically learn hierarchical features, combined with the capacity of LSTMs to handle sequential data, makes this hybrid model more efficient and accurate for emotion detection in real-world scenarios.

However, challenges remain, particularly with generalizing models to new datasets, which is crucial for ensuring reliability across diverse populations and conditions.

Conclusion

This study presents an effective method for emotion detection using CNN and LSTM, focusing on accurately identifying human emotions through facial expressions. The approach leverages the potential of CNNs for spatial feature extraction and LSTMs for capturing temporal dependencies, achieving high accuracy in both image and video-based emotion recognition tasks. Future work will concentrate on enhancing both speed and accuracy in emotion detection by leveraging advancements in modern deep learning techniques. Real-time emotion detection is also explored, ensuring the model can efficiently process emotions in dynamic environments, such as live video streams.

Future efforts will aim to improve emotion intensity classification, providing more granular insights, such as the degree of happiness or anger. Further research will address challenges such as the need for larger, more diverse datasets and better generalization across different demographics, lighting conditions, and facial variations. Advancements in neural networks, computer vision, and sensor technologies present viable pathways for enhancing system accuracy, robustness, and real-time performance.

REFERENCES

[1]. **Al Osman, H., & Falk, T. H. (2017).** Multimodal affect recognition: Current approaches and challenges. *Emotion and Attention Recognition Based on Biological Signals and Images* (pp. 59-86).

https://doi.org/10.5772/65683

[2]. **Aouani, H., & Ayed, Y. B. (2020).** Speech emotion recognition with deep learning. *Procedia Computer Science,* 176, 251-260.

https://doi.org/10.1016/j.procs.2020.08.027

[3]. **Batliner, A., Schuller, B., Seppi, D., Steidl, S., Devillers, L., Vidrascu, L., & Amir, N. (2011).** *The Automatic Recognition of Emotions in Speech* (pp. 71-99). Springer Berlin Heidelberg.

https://doi.org/10.1007/978-3-642-15184-2_6

[4]. **Canedo, D., & Neves, A. J. (2019).** Facial expression recognition using computer vision: A systematic review. *Applied Sciences,* 9(21), 4678.

https://doi.org/10.3390/app9214678

[5]. **Dhavale, M., & Bhandari, S. (2022, August).** Speech emotion recognition using CNN and LSTM. In *2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA)* (pp. 1-3). IEEE.

https://doi.org/10.1109/ICCUBEA54992.2022.10010751

[6]. **Han, K., Yu, D., & Tashev, I. (2014, September).** Speech emotion recognition using deep neural network and extreme learning machine. In *Interspeech 2014* (pp. 223-227).

https://doi.org/10.21437/Interspeech.2014-57

[7]. **Khalil, R. A., Jones, E., Babar, M. I., Jan, T., Zafar, M. H., & Alhussain, T. (2019).** Speech emotion recognition using deep learning techniques: A review. *IEEE Access,* 7, 117327-117345.

https://doi.org/10.1109/ACCESS.2019.2936124

[8]. **Liu, G., Cai, S., & Wang, C. (2023).** Speech emotion recognition based on emotion perception. *EURASIP Journal on Audio, Speech, and Music Processing,* 2023(1), 22.

https://doi.org/10.1186/s13636-023-00289-4

[9]. **Mountzouris, K., Perikos, I., & Hatzilygeroudis, I. (2023).** Speech emotion recognition using convolutional neural networks with attention mechanism. *Electronics,* 12(20), 4376.

https://doi.org/10.3390/electronics12204376

[10]. **Pantic, M., & Rothkrantz, L. J. M. (2000).** Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 22(12), 1424-1445.

https://doi.org/10.1109/34.895976

[11]. **Pervaiz, M., & Khan, T. A. (2016).** Emotion recognition from speech using prosodic and linguistic features. *International Journal of Advanced Computer Science and Applications,* 7(8), 84-90.

[12]. Sarmah, K., Gogoi, S., Das, H. C., Patir, B., & Sarma, M. J. (2024). A state-of-arts review of deep learning techniques for speech emotion recognition. *Journal of Electrical Systems,* 20(7s), 1638-1652.

[13]. Shinde, A. S., & Patil, V. V. (2021, May). Speech emotion recognition system: A review. In *Proceedings of the 4th International Conference on Advances in Science & Technology (ICAST2021).*

[14]. Tripathi, S., Kumar, A., Ramesh, A., Singh, C., & Yenigalla, P. (2019). Deep learning based emotion recognition system using speech features and transcriptions. *arXiv preprint arXiv:1906.05681.*

https://doi.org/10.48550/arXiv.1906.05681

[15]. Ullah, R., Asif, M., Shah, W. A., Anjam, F., Ullah, I., Khurshaid, T., & Alibakhshikenari, M. (2023). Speech emotion recognition using convolution neural networks and multi-head convolutional transformer. *Sensors,* 23(13), 6212.

https://doi.org/10.3390/s23136212

[16]. Zhao, S., Yang, Y., Cohen, I., & Zhang, L. (2021, August). Speech emotion recognition using auditory spectrogram and cepstral features. In *2021 29th European Signal Processing Conference (EUSIPCO)* (pp. 136-140). IEEE.

https://doi.org/10.23919/EUSIPCO54536.2021.9616144

[17]. Zisad, S. N., Hossain, M. S., & Andersson, K. (2020, September). Speech emotion recognition in neurological disorders using convolutional neural network. In *International Conference on Brain Informatics* (pp. 287-296). Cham: Springer International Publishing.

https://doi.org/10.1007/978-3-030-59277-6_26

ABOUT THE AUTHORS

Dr. Manish Goswami is an Associate Professor in the Department of Computer Science and Engineering at SB Jain Institute of Technology, Management, and Research, Nagpur, Maharashtra, India. With over 24 years of experience in academia, he specializes in Compiler Design, Theory of Computation, Machine Learning, and Data Science. Holding a Ph.D. in Computer Science and Engineering, he has made significant contributions to these fields and continues to drive research and innovation while imparting his knowledge to students.

Aditya Parate is currently working as an Assistant Professor in the Computer Science Engineering Department at S.B. Jain Institute of Technology, Management, and Research, Nagpur, Maharashtra, India. He is pursuing an M.Tech from Kalinga University, Chhattisgarh, India, and has completed a B.Tech and BCCA. He has experience in Linux and System Administration, Database Development, Web Design, and WordPress Development. He has successfully published two books: Operating System Insights with Linux and The Beginning: Shell Scripting.

Nisarga Kapde is a B.Tech student specializing in Computer Science and Engineering at SB Jain Institute of Technology Management and Research, Nagpur, Maharashtra, India. Aspiring to become a Data Analyst, she has developed a strong foundation in Python, SQL, Machine Learning, Data Analysis, Power BI, and Pandas. With certifications in Data Analytics, Python Data Analytics, Tableau, and Enterprise Data Science, she has worked on projects including the development of a MediPredict system, a Speech Emotion Detection system using voice and facial expressions, and creating interactive dashboards for Olympic and sales data.

Shashwat Singh is a B.Tech Student specializing in Computer Science and Engineering at SB Jain Institute of Technology Management and Research, Nagpur, Maharashtra, India, with a strong foundation in research-oriented data analysis and backend development. His experience includes internships where he conducted exploratory data analysis, improved machine learning model accuracy, and optimized backend systems. Proficient in tools like Spring Boot, Python, and Statistical Modeling, he has developed projects focused on enhancing operational efficiency and data precision.

Nitiksha Gupta is a B.Tech Student specializing in Computer Science and Engineering at SB Jain Institute of Technology, Management, and Research, Nagpur, Maharashtra, India. Aspiring to become a Full-Stack Developer, she has gained valuable experience through internships as a Web Development Intern at Codemate IT Services and as a Data Science Intern at LGPS Pvt. Ltd., Nagpur, India. Skilled in Python, HTML, CSS, and SQL, she has worked on diverse projects, including real-time object detection using YOLO and a speech emotion detection system that leverages both voice and facial expressions.

Meena Surjuse is a B.Tech student specializing in Computer Science and Engineering at SB Jain Institute of Technology Management and Research, Nagpur, Maharashtra, India. Aspiring to become a Full Stack Developer, she has gained experience as a Full Stack Developer Intern at TechVegan Hardware and Software, a Front-End Web Developer at Bharat Intern, and a Data Analyst at LGPS Hybrid Energy Pvt. Ltd., Nagpur, India. She possesses strong technical skills in HTML, CSS, C++, SQL, Python, Machine Learning, and Data Analysis.