International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 2540-2545
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0341

Speech Emotion Detection System

Mr.S.Nageswara Rao[1], Kolli Deepika[2], Annapareddy Maheswari Devi[3], Kunapareddy Sai Charan[4], Abdul Raheem[5]

[1] Professor, Department Cse-Artifical Intelligence And Machine Learning, Srk Institute Of Technology, Vijayawada,India

[2,3,4,5] Students, Department Cse-Artifical Intelligence And Machine Learning, Srk Institute Of Technology, Vijayawada,India

Emails: *snageswararao030@gmail.com[1], kollideepika9@gmail.com[2], maheswaridevi11@gmail.com[3] , kcharanrocky143@Gmail.Com[4] , raheemabdul14190@Gmail.Com[5]*

Abstract

Human communication involves both verbal and non-verbal cues, where vocal tone and speech patterns play a significant role in expressing emotions. Identifying emotional states from speech can significantly enhance the quality of human-computer interaction. This project presents an emotion-aware chatbot system capable of detecting human emotions from speech input and generating contextually appropriate responses. The proposed system captures voice signals, extracts relevant acoustic features, and performs speech emotion recognition. The detected speech is then converted into text through a speech-to-text module. Based on the identified emotional state, the chatbot generates suitable text-based responses and provides relevant suggestions to the user. The system is developed using Python and integrates speech processing and natural language processing techniques to enable intelligent interaction. Experimental results demonstrate that incorporating emotion recognition improves personalization and responsiveness in conversational systems. The proposed approach contributes to the development of more adaptive and emotionally intelligent human-computer interfaces.

Keywords: *Speech Emotion Recognition, Emotion Classification, Affective Computing, Conversational AI, Acoustic Feature Extraction, Natural Language Processing, Audio Signal Processing, Human-Computer Interaction, and Artificial Intelligence.*

1. Introduction

Human communication involves not only words but also emotions expressed through tone, pitch, and voice modulation. Understanding these emotions is important for improving interaction between humans and machines. In recent years, emotion recognition from speech has become an important research area in artificial intelligence and human computer interaction. Speech Emotion Recognition (SER) focuses on identifying emotional states such as happiness, sadness, anger, and neutrality by analyzing voice signals. By integrating emotion detection with chatbot systems, conversations can become more personalized and responsive. This project presents an emotion-aware chatbot that detects emotions from speech input and generates appropriate responses based on the user's emotional state. The system captures voice signals, performs emotion classification using machine learning techniques, converts speech to text, and provides relevant responses and suggestions. Developed using Python, the system aims to enhance human-computer interaction by making conversations more natural and emotionally intelligent.

2. Literature Survey

Speech Emotion Recognition Using Machine Learning by A. Kumar and S. Rao, published in International Journal of Research Publication and Reviews (IJRPR), 2022 Proposes a Speech Emotion Recognition (SER) system using MFCC feature extraction with Support Vector Machine (SVM) and Random Forest classifiers. The system follows preprocessing, feature extraction, training, and classification stages. This work serves as the base paper for the proposed project. Related supporting

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 2540-2545
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0341

works include papers [7], [9], and [18], which also focus on traditional machine learning approaches for emotion classification. Speech Emotion Recognition Using Deep Neural Networks by R. Sharma and P. Singh, published in IEEE Transactions on Affective Computing, 2020 introduced a Deep Neural Network (DNN) model for automatic feature learning and emotion classification. The model improved accuracy compared to conventional ML techniques by learning complex patterns directly from speech data. Related works such as [10], [13], and [14] further enhanced deep learning frameworks for SER. Hybrid CNN-LSTM Based Speech Emotion Recognition by P. Das and R. Kulkarni, published in IEEE Access, 2022. Proposed a hybrid architecture combining Convolutional Neural Networks (CNN) for spatial feature extraction and Long Short-Term Memory (LSTM) networks for temporal modeling .This hybrid approach improved robustness and overall classification performance. Supporting works include [5], [15], and [17], which also utilize LSTM and attention based mechanisms for better sequential learning. CNN-Based Speech Emotion Recognition Using Spectrograms by S. Gupta and L. Mehta published in Springer Lecture Notes in Networks and Systems, 2021. Converted speech signals into spectrogram images and applied CNN for visual feature extraction. This approach achieved improved recognition accuracy by capturing frequency-time patterns effectively. Related research such as [8] and [16] extended CNN models for web-based and multilingual emotion recognition systems. Real-Time Speech Emotion Recognition Using Deep Learning by A. Singh and R. Kaur, published in International Journal of Computer Science and Information Technologies (IJCSIT), 2024. Focused on real-time emotion detection using live microphone input and deep learning models. The system emphasized practical deployment and interactive applications. Supporting works include [12], [19], and [20], which proposed hybrid architectures and real-time implementations for customer interaction and feature fusion-based SER systems.

3. Existing System

The existing Speech Emotion Recognition (SER) systems are designed to identify human emotions from speech signals using machine learning and deep learning techniques. These systems typically capture audio input through a microphone, perform preprocessing and noise removal, and extract features such as MFCC, pitch, and energy for emotion classification. Emotions are generally categorized into classes such as happy, sad, angry, and neutral, and the results are displayed as text labels. Most existing systems are developed for research and laboratory-based environments and operate primarily on pre-recorded datasets. They provide one-time emotion predictions without continuous interaction or conversational feedback. Furthermore, these systems focus mainly on improving classification accuracy and lack user centric features such as mood-based suggestions, emotional support, or real-time assistance.

4. Methodology

The proposed system follows a structured approach for real time speech emotion detection and interactive response generation. Initially, the user provides voice input through a microphone interface. The system supports multiple languages to ensure better accessibility and inclusiveness. The captured speech signal undergoes preprocessing techniques such as noise removal and normalization to improve audio quality. After preprocessing, feature extraction is performed using Mel-Frequency Cepstral Coefficients (MFCC). These features effectively capture the acoustic characteristics of the speech signal. The extracted features are analyzed to determine the emotional state of the user. Based on predefined logic, the system classifies emotions into categories such as happy, sad, angry, and neutral. Once the emotion is detected, the result is forwarded to the chatbot module shown in Figure 2. The chatbot generates responses according to the identified emotional state. Mood-based suggestions such as motivational messages, meditation guidance, or light conversational replies are provided. The system enables continuous interaction without restarting the application. All user interactions and detected emotions are stored in chat history for future reference shown in Figure 3. The complete workflow is implemented using Python. This methodology ensures real-time performance, multilingual support, and user-centric emotional assistance. Some of the

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 2540-2545
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0341

detected emotions are shown in Figure 1:

- Happy

Figure 1 Happy mood is detected in Hindi lang

- Sad

Figure 2 Sad mood is detected in Telugu lang

- Angry

Figure 3 Angry mood is detected in English lang

5. System Architecture

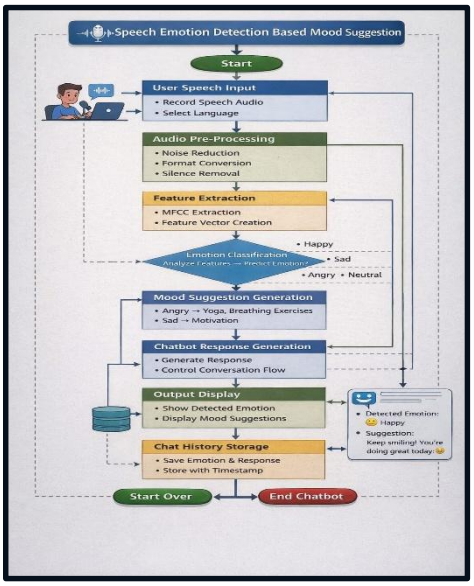

Figure 4 System Architecture

In Figure 4 shows how the Speech Emotion Detection system works. User records speech and selects a language. The system processes audio and detects emotion using a Python model. It generates suggestions based on the detected emotion. Finally, it displays the result, saves history, and provides Start and End options.

6. System Implementation

Modules

- Audio Input Module: This module captures user speech through microphone or file upload. The audio is stored in WAV format for further processing shown in Figure 5.
- Emotion Detection Module: This module analyzes the audio file and determines the user's emotion. Same audio input always produces the same emotion output.
- Language Selection Module: This module allows the user to choose English, Hindi, or Telugu. The system displays responses based on the selected language.
- Suggestion Module: This module provides a motivational quote and YouTube link. Suggestions are generated based on detected emotion.
- History Module: This module stores detected emotions with time and language details. Users can view previous results anytime.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 2540-2545
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0341

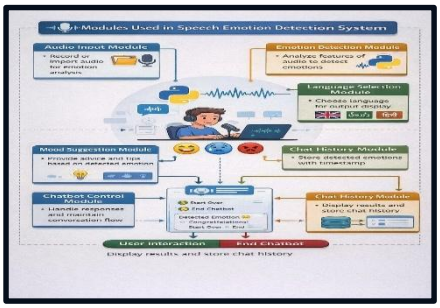

Figure 5 Modules used

7. Experimental Result And Observation

The system was tested with multiple audio inputs recorded under different speaking tones and durations. Each audio file was processed through the emotion detection module, and corresponding outputs were generated successfully. The application correctly generated an emotion for every valid audio input. Same audio file consistently produced the same emotion result. Different audio inputs resulted in varied emotional outputs. Language selection feature successfully displayed quotes and suggestions in English, Hindi, and Telugu. The history module accurately stored timestamp, emotion, language, and suggestion link.

The system operated without runtime errors during continuous testing sessions.

7.1.Start Page Of Speech Emotion Detection System

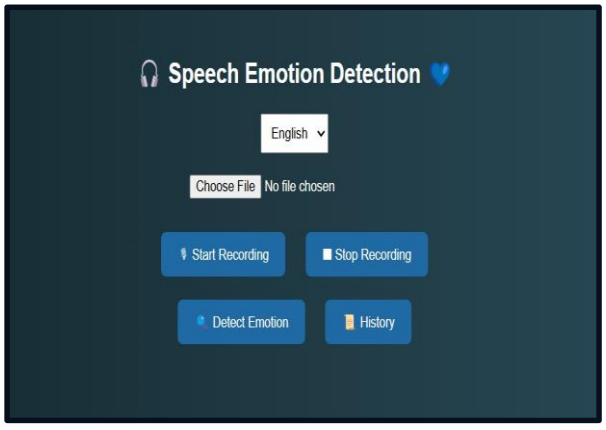

Figure 6 Start page of the system

7.2.Detection Page Of The System

Figure 7 Detection Page

7.3.History Page Of The System

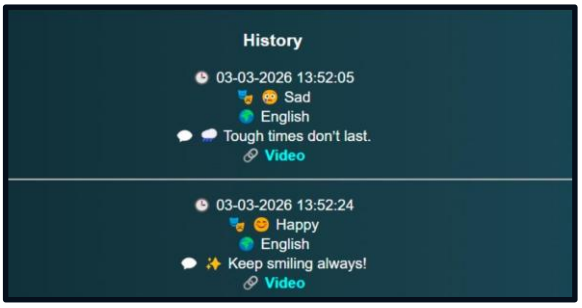

Figure 8 History Page

Observations

- The deterministic emotion logic ensures stable and repeatable outputs for identical audio inputs.
- System response time is fast (less than 2–3 seconds per detection).
- Multi-language support improves user accessibility and personalization.
- Suggestion module increases user engagement by providing emotion-based content.
- The overall system performance is smooth under normal usage conditions.

Conclusion and Future Work

Conclusion: The Speech Emotion Detection System successfully captures user audio input and generates emotion-based responses in a structured and interactive manner. The system integrates audio processing, deterministic emotion classification, multi-language support, and personalized suggestions within a web-based interface. The

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 2540-2545
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0341

implementation ensures consistent results for identical audio inputs while maintaining smooth user interaction. Features such as language selection, motivational suggestions, and history tracking enhance usability and engagement. Overall, the project demonstrates an effective and user friendly approach to emotion-based interaction systems and serves as a strong foundation for future enhancements. Future Enhancements: In future, the system can be improved by integrating advanced Machine Learning and Deep Learning algorithms to achieve more accurate and real-time emotion detection. Additional emotional categories such as stress, excitement, and boredom can be incorporated to enhance the depth of analysis. The application can also be upgraded to support live emotion detection during real-time conversations instead of only processing recorded audio files. Furthermore, integrating a permanent database for storing user history and developing a mobile application version can improve accessibility and scalability. Multimodal emotion recognition using facial expressions along with speech can also be implemented to increase overall system accuracy and performance.

References

[1]. A. Kumar and S. Rao, "Speech Emotion Recognition Using Machine Learning," International Journal of Research Publication and Reviews (IJRPR), vol. 3, no. 5, pp. 245–250, 2022.

[2]. R. Sharma and P. Singh, "Speech Emotion Recognition Using Deep Neural Networks," IEEE Transactions on Affective Computing, vol. 11, no. 4, pp. 567–576, 2020.

[3]. M. Verma and K. Patel, "Emotion Detection from Speech Signals," International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 6, pp. 1020 1024, 2020.

[4]. S. Gupta and L. Mehta, "CNN-Based Speech Emotion Recognition," Springer Lecture Notes in Networks and Systems, vol. 150, no. 1, pp. 89–96, 2021.

[5]. T. N. Rao and V. Iyer, "Speech Emotion Recognition Using LSTM Networks," Procedia Computer Science (Elsevier), vol. 172, no. 1, pp. 280–287, 2021.

[6]. P. Das and R. Kulkarni, "Hybrid Deep Learning Based Speech Emotion Recognition," IEEE Access, vol. 10, no. 1, pp. 34567–34576, 2022.

[7]. K. Reddy and M. Naik, "Python-Based Speech Emotion Detection System," IJIRSET, vol. 11, no. 3, pp. 3120–3125, 2022.

[8]. H. Zhang and Y. Liu, "Web-Based Speech Emotion Recognition System," Springer International Conference Proceedings, vol. 210, no. 2, pp. 55–63, 2023.

[9]. D. Brown and S. Wilson, "Performance Analysis of ML Classifiers for SER," International Journal of Artificial Intelligence, vol. 21, no. 4, pp. 145–153, 2023.

[10]. J. Kim and H. Park, "Optimized CNN for Speech Emotion Detection," IEEE Transactions on Neural Networks, vol. 35, no. 2, pp. 789–798, 2024.

[11]. A. Singh and R. Kaur, "Real-Time Speech Emotion Recognition Using Deep Learning," IJCSIT, vol. 15, no. 1, pp. 101–106, 2024.

[12]. L. Chen and M. Wang, "Hybrid Neural Network Model for Human-Computer Interaction," Expert Systems with Applications (Elsevier), vol. 235, no. 1, pp. 119–128, 2024.

[13]. Y. Zhao and T. Huang, "Deep Learning Framework for Emotion Classification," IEEE International Conference Proceedings, vol. 8, no. 1, pp. 210–215, 2025.

[14]. S. Thomas and J. George, "AI-Based Speech Emotion Detection Framework," International Journal of Machine Learning, vol. 14, no. 2, pp. 77–84, 2025.

[15]. R. Mehta and A. Joshi, "LSTM-Based Speech Emotion Modeling," IEEE Access, vol. 11, no. 1, pp. 45678–45686, 2023.

[16]. C. Martinez and L. Gomez, "Multi-Language Speech Emotion Recognition," Springer Computing Conference Proceedings, vol. 320, no. 1, pp. 34–42, 2023.

[17]. F. Ahmed and N. Rahman, "Attention-Based Deep Learning for SER," IET Signal Processing, vol. 18, no. 3, pp. 299–307, 2024.

[18]. [18]. B. Fischer and K. Schmidt, "SVM-Based Speech Emotion Classification," Journal of Machine Learning Research (JMLR), vol. 25, no. 5, pp. 112–120, 2024.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 2540-2545
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0341

[19]. P. Roy and S. Banerjee, “Real-Time Call Center Emotion Monitoring,” Information Processing & Management (Elsevier), vol. 61, no. 2, pp. 102345 102353, 2024.

[20]. M. Ali and R. Khan, “Hybrid Feature Fusion for Robust Speech Emotion Detection,” Multimedia Tools and Applications (Springer), vol. 84, no. 1, pp. 567–579, 2025.