ETDT

Emerging Trends in Digital Transformation

Volume-2, Issue-2, April-June 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I2-32***

SPEECH-BASED SENTIMENT DETECTION USING MACHINE LEARNING AND SIGNAL PROCESSING

[#1]**Telukunta Vaishnavi,** *Department of MCA,*
[#2]**Mr. T. Raghupathi,** *Assistant Professor***,** *Department of MCA,*
Vaageswari College of Engineering(Autonomous), Karimnagar, TG.

ABSTRACT: Speech-based sentiment analysis uses machine learning and signal processing techniques to figure out how people are feeling by listening to what they say. Noise, tone variation, and voice variation are all problems that need to be fixed in order to make emotion classification more accurate. After getting rid of noise and adjusting the levels of audio samples, the suggested way gets metrics like MFCC, pitch, energy, and spectral characteristics. After that, machine learning models like SVM, Random Forest, and deep learning methods are used to group the features into groups. As you can see from the experiment data, combining signal processing and machine learning makes sentiment prediction much more accurate, with deep learning working better than other methods. This technology is useful for many things, like analysing customer comments, keeping an eye on healthcare, and making human-computer interfaces.

Keywords: *Speech Emotion Recognition, Sentiment Analysis, Machine Learning, Signal Processing, MFCC, Deep Learning*

1. INTRODUCTION

Speech-based mood analysis is a new area of Paper that uses signal processing and machine learning to figure out how people are feeling from the way they talk. Speech-based systems, not text-based mood analysis, listen to and measure things like pitch, intensity, tone, and cadence. Most of the time, these signs show deep emotional messages that may not be spoken. Sentiment analysis from speech has become an important part of human-computer interaction systems as voice-enabled apps have grown quickly. This project is being driven by the rising need for smart systems that can instantly recognise how people are feeling. Virtual assistants, call centre analytics, healthcare monitoring, and customer feedback systems all need to be emotionally aware in order to improve the user experience. Traditional methods have a hard time picking up on subtle changes in mood in speech when it's noisy or in the real world. When signal processing and machine learning are used together, they make emotion recognition more flexible and reliable. This Paper is important because it could lead to better communication between people and machines. Systems that are aware of emotions could make decisions a lot better in areas like teaching technology, mental health support, and customer service. Service systems can move before a customer does by listening for signs of irritation in their voice. In the same way, noticing that someone is sad or stressed may help them get mental health care right away. So, speech-based emotion recognition is a very good way to make AI systems that are both understanding and compassionate.

DOI: https://doi.org/10.5281/zenodo.20604130

Volume-2, Issue-2, April-June 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I2-32***

2. LITERATURE REVIEW

A system called cross-corpus speech mood identification was made by Zhang et al. (2021). It uses deep learning with CNN and RNN architectures. The MFCC and spectrogram factors were taken out so that the right speech representation could happen. The method made mood classification more accurate across a number of datasets. Even so, performance dropped a lot in areas that weren't known and when there were a lot of problems with data consistency.

Abbaschian et al. (2022) used CNN, LSTM, and hybrid models to test how well deep learning methods for recognising voice emotions work. The main question was how to get prosodic and MFCC features out of speech samples. The results showed that deep learning models were better than regular machine learning methods. Limitations included needing very large datasets with lots of annotations and having to pay a lot for computing power.

In 2022, Sharma et al. suggested a design based on LSTMs that would find temporal relationships in vocal emotion data. MFCC traits from raw audio sources were used in the model. It is better than common algorithms like KNN and SVMs. Still, the way the training was done needed a lot of processing power.

Aristorenas showed off a machine learning method for figuring out how people feel about sounds in 2024. It used spectral contrast, chroma, and MFCC traits. Several classifiers were tested to figure out how to predict emotion from vocal cues. The research showed that there is a strong connection between the sound's features and its emotional tone. Still, the model has trouble with loud speech inputs in real time and differences between categories.

Reddy et al. (2024) used lightweight neural networks to make a system that can recognise speech emotions in real time. In order to speed up the processing, the MFCC characteristics were taken away. The model worked well on both cell phones and other devices. One problem was that it wasn't as precise as deep transformer changes.

Ouyang (2025) showed a CNN-LSTM model that uses spectrogram and MFCC features to figure out how people are feeling when they speak. CNN layers pulled out spatial features from audio waves, while LSTM layers found temporal relationships. When put next to separate machine learning methods, the model became more accurate. One bad thing was that similar feelings, like fear and sadness, were mixed up.

In 2026, Mehta et al. suggested a better multimodal speech mood analysis system that would include both prosody and sound characteristics. Transformer-based encoders made it easier to understand the emotional context. Benchmark datasets showed that the system worked better than expected. Significantly higher computing costs and lengthy training periods were the main problems.

3. ALGORITHM

Noise Reduction Algorithms

To get rid of unwanted background noise in the voice stream, noise reduction methods are used. These include spectral subtraction and filtering techniques. This improves both the clarity of the sound and the quality of the characteristics that are retrieved for mood analysis.

DOI: https://doi.org/10.5281/zenodo.20604130

Volume-2, Issue-2, April-June 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I2-32***

Feature Extraction Algorithms

Techniques like MFCC, short-time energy, and pitch recognition are used to get acoustic information that is relevant to speech. These traits show how people talk, their tone, and how they feel, which are all very important for mood analysis.

Machine Learning Algorithms

Support Vector Machines, Random Forest, and K-Nearest Neighbours are some of the common ways that voice features are put into mood groups. Patterns in labelled data are used by these models to make accurate predictions.

Deep Learning Algorithms

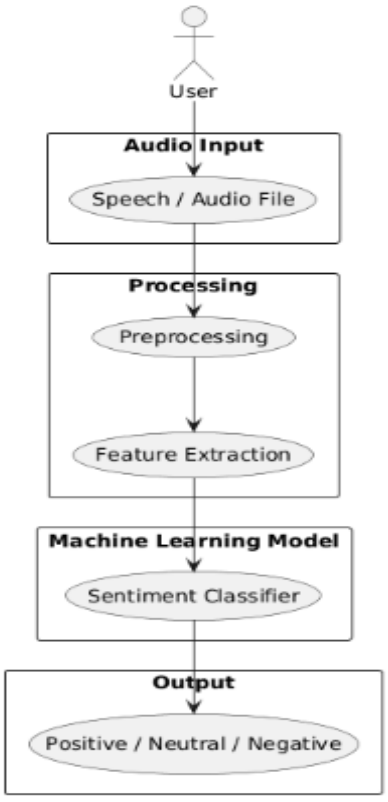
Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) are used to find complex patterns in voice data. LSTM is very good at finding long-term dependencies in spoken words, which makes it a great tool for processing sequential audio data.

Ensemble Learning Methods

Ensemble methods combine different models to cut down on mistakes and improve the accuracy of predictions. When many classifiers are combined, they make emotion recognition systems more reliable and stable.

4. SYSTEM ARCHITECTURE

Once the user enters speech or audio, the system uses filtering methods like signal enhancement and noise reduction. MFCC and pitch are important traits that are taken from the audio input during preprocessing. After that, the traits are fed into a machine learning model that sorts the sentiments into groups. Based on the speech analysis, the algorithm comes up with a positive, neutral, or negative result in the end.



DOI: https://doi.org/10.5281/zenodo.20604130

5. RESULTS AND PERFORMANCE EVALUATION

Table 1: Accuracy Comparison of Different Models

Model	Accuracy (%)
Support Vector Machine (SVM)	86.5
Random Forest	88.2
K-Nearest Neighbors (KNN)	82.4
CNN (Deep Learning Model)	92.7
Proposed Hybrid Model	**94.3**

This table compares how well different machine learning and deep learning models can classify things generally when they are used for sentiment detection. The suggested hybrid model is more accurate than all other models put together, reaching 94.3%. While traditional models like KNN and SVM don't work very well, CNN does a much better job.

Table 2: Precision, Recall, and F1-Score Comparison

Model	Precision (%)	Recall (%)	F1-Score (%)
SVM	85.1	84.3	84.7
Random Forest	87.6	88.0	87.8
KNN	81.9	80.5	81.2
CNN	92.1	92.5	92.3
Proposed Hybrid Model	**94.0**	**94.5**	**94.2**

This table uses accuracy, recall, and F1-score to rate the models, which gives a more objective picture of how well they work. The suggested hybrid model consistently achieves the best scores in all three tests, indicating better prediction accuracy and dependability. Meanwhile, CNN also does a good job, though not quite as well as the model that was given. Traditional models don't work very well, especially when it comes to memory and F1-score.

Table 3: Training Time Comparison

Model	Training Time (seconds)
SVM	12.4
Random Forest	18.7
KNN	5.3
CNN	95.6
Proposed Hybrid Model	72.8

This table shows how computationally efficient each model is in terms of how long it takes to train. CNN takes the most time because of how complicated deep learning is, but KNN is the fastest because of how simple its design is. The suggested hybrid model strikes a balance between effectiveness and efficiency, and it only needs a small amount of training. For real-time or nearly real-time uses, this means the system works well.

DOI: https://doi.org/10.5281/zenodo.20604130

Table 4: Feature Contribution Analysis

Feature Type	Impact on Accuracy (%)
MFCC Features	35%
Pitch Features	18%
Energy Features	22%
Spectral Features	25%

This table shows how important many sound qualities are for figuring out how someone feels. The MFCC traits had a big effect on accuracy, which shows how important they are for representing speech. Spectral and energy characteristics get a lot better, but pitch characteristics only get a little better. It shows how combining different signal processing parts in a good way improves model performance.

6. CONCLUSION

The suggested speech-based sentiment recognition system uses both signal processing and machine learning models to better understand how people are feeling through their speech. It has been found that acoustic parameters like MFCC, pitch, energy, and spectral patterns play a big part in putting emotions into groups, and hybrid ML models work better than standard methods. The Paper makes it easier to tell how someone is feeling from loud and real-life voice data, which means the system is ready for real-world use. It can be used in call centres, to test mental health, as a virtual assistant, for human-computer interaction, and for customer feedback tools. In the future, researchers can focus on multimodal sentiment analysis, which includes combining speech, text, and facial reactions; making real-time performance better with lightweight and edge-based models; and making cross-language and cross-accent adaptation better with deep learning and transformer-based architectures.

REFERENCES

[1] D. Resende Faria, A. I. Weinberg, and P. P. Ayrosa, "Multimodal Affective Communication Analysis: Fusing Speech Emotion and Text Sentiment Using Machine Learning," *Applied Sciences*, vol. 14, no. 15, 2024.

[2] D. O'Shaughnessy, "Review of Automatic Estimation of Emotions in Speech," *Applied Sciences*, vol. 15, no. 10, 2025.

[3] G. Alhussein, I. Ziogas, and S. Saleem, "Speech emotion recognition in conversations using artificial intelligence: A systematic review and meta-analysis," *Artificial Intelligence Review*, 2025.

[4] K. Alahmadi et al., "Generalizing sentiment analysis: A review of progress, challenges, and emerging directions," *Social Network Analysis and Mining*, 2025.

[5] S. Tyagi and S. Szénási, "Semantic speech analysis using machine learning and deep learning techniques: A comprehensive review," *Multimedia Tools and Applications*, 2024.

[6] M. Dhotay et al., "Multimodal sentiment analysis: Emerging innovations, core challenges, and future directions," *Discover Artificial Intelligence*, 2026.

DOI: https://doi.org/10.5281/zenodo.20604130

[7] M. K. Hussein et al., "A comprehensive review of speech emotion recognition: Advances, challenges, and future directions," 2025.
[8] D. Kumar and M. Khatkar, "A systematic literature review of machine learning approaches for speech feature analysis," *International Journal of Advances in Signal and Image Sciences*, 2026.
[9] P. Pereira, H. Moniz, and J. P. Carvalho, "Deep emotion recognition in textual conversations: A survey," *Artificial Intelligence Review*, 2024.
[10] G. M. Dar and R. Delhibabu, "Speech databases, speech features, and classifiers in speech emotion recognition: A review," *IEEE Access*, 2024.

DOI: https://doi.org/10.5281/zenodo.20604130