IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁐ Peer-reviewed & Refereed journal ⁐ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14373

Speech Emotion Analysis Using Natural Language Processing

Dr. R. A. Burange[1], Kartik Pachkhande[2], Rohit Bhil[3], Harshal Satghare[4]

Professor, Department of Electronics and Telecommunication, K.D.K. College of Engineering, Nagpur, Maharashtra, India[1]

Student, Department of Electronics and Telecommunication, K.D.K. College of Engineering, Nagpur, Maharashtra, India[2,3,4]

Abstract: Emotion recognition from human voice has emerged as a crucial technology in various fields, including healthcare, human-computer interaction, and artificial intelligence-based applications. The ability to detect emotions based on speech signals enhances system adaptability and improves user experience. This study presents a progressive implementation of an emotion detection system that integrates Natural Language Processing (NLP) and speech feature extraction techniques. The system utilizes machine learning and deep learning models to classify emotions, including happiness, sadness, anger, and fear, based on vocal expressions. The approach involves extracting speech parameters such as pitch, tone, energy, and amplitude, which are analyzed using ML-based classifiers. Additionally, NLP techniques, including text sentiment analysis and word embedding's, enhance classification accuracy by providing contextual insights. The system is implemented on Raspberry Pi hardware, making it portable and scalable for real-world applications. Initial findings indicate that deep learning models outperform traditional ML approaches, offering improved accuracy. Future advancements will focus on reducing background noise, optimizing feature selection, and incorporating real-time emotion tracking.

Keyword-Speech Emotion Recognition, NLP, Machine Learning, Deep Learning, Speech Processing, Human-Computer Interaction.

I. INTRODUCTION

Emotion detection from speech is an essential aspect of human-computer interaction, helping systems understand and respond to human emotions effectively. Speech carries not only linguistic information but also emotional cues, which can be analyzed using various speech-processing techniques. Unlike facial expression analysis, speech-based emotion recognition is more practical in scenarios where visual data is unavailable.

This paper explores an approach to detect emotions from human voice using Natural Language Processing (NLP) and speech features. The process involves analyzing tone, pitch, intensity, and linguistic patterns to classify emotions such as happiness, sadness, anger, and neutrality. By extracting meaningful speech features and applying statistical analysis, emotion recognition can be enhanced for practical applications.

The methodology includes data collection, preprocessing, feature extraction, and model training using computational techniques. The system is implemented on a Raspberry Pi 3B+ to enable real-time emotion detection in low-power environments. The study evaluates the efficiency of different speech-processing methods and discusses challenges such as noise interference, speaker variability, and computational limitations.

The primary objective of this research is to develop a reliable speech-based emotion recognition system that can be used in real-world applications, such as healthcare, customer service, and assistive technologies. Future improvements will focus on optimizing performance, enhancing accuracy, and making the system more adaptable to diverse speech patterns.

II. RELATED WORK

Attar, H. I., et al. [1] (2023) explored machine learning techniques for speech emotion recognition, focusing on feature extraction and classification. They evaluated multiple models, including Support Vector Machines (SVM) and Random Forest, for emotion detection. The study emphasized the importance of feature selection in improving model accuracy.

DOI: 10.17148/IJARCCE.2025.14373

Their experiments demonstrated that a combination of Mel-Frequency Cepstral Coefficients (MFCCs) and prosodic features enhances recognition performance. The research concluded that machine learning can achieve high accuracy but struggles with speaker variability.

Aouani, H., & Ben Ayed, Y[2]. (2022) This study investigated deep learning approaches for speech emotion recognition, comparing CNN, LSTM, and hybrid models. The authors highlighted the advantages of deep networks in capturing temporal dependencies in speech data. They used large-scale emotion-labeled datasets to improve model generalization. The paper demonstrated that CNN-LSTM architectures outperformed traditional classifiers. The research suggested future work on multi-modal fusion to enhance accuracy.

Tripathi, S., et al. [3] (2021) proposed a hybrid approach combining machine learning and deep learning for emotion recognition. Their model integrated handcrafted features with deep feature representations for improved classification. Experiments on benchmark datasets showed that hybrid models outperform individual techniques. The study addressed challenges such as feature redundancy and computational efficiency. The authors recommended further exploration of domain adaptation techniques for real-world applications.

Rastogi, R. [4] (2020) work focused on speech analysis for emotion detection using statistical and spectral features. The study examined the role of pitch, energy, and spectral characteristics in identifying emotions. The author applied traditional classifiers like Decision Trees and SVM for performance evaluation. Results indicated that feature engineering significantly impacts classification accuracy. The paper concluded with a discussion on the limitations of traditional methods in handling noisy environments.

Byun, S., & Lee, S[5]. (2019) investigated acoustic feature-based emotion recognition, analyzing fundamental frequency, formants, and speech dynamics. They emphasized the significance of feature fusion in improving classification accuracy. The study compared traditional machine learning models with neural networks. Results showed that deep learning models performed better in complex emotional expressions. The paper recommended further studies on real-time implementation and cross-language adaptability.

III. METHODOLOGY

The proposed system follows a structured pipeline:

A. Speech Feature Extraction

- **Mel-Frequency Cepstral Coefficients (MFCCs):** Captures speech timbre and phonetic structure.
- **Pitch and Tone:** Identifies frequency variations linked to emotions.
- **Energy and Amplitude:** Measures speech intensity.
- **Spectral Features:** Analyzes frequency distribution for classification.

B. Machine Learning and Deep Learning Models

- **Support Vector Machine (SVM):** Baseline classifier.
- **Random Forest (RF):** Robust ensemble learning model.
- **Multilayer Perceptron (MLP):** Neural network for speech processing.
- **Convolutional Neural Networks (CNNs):** Extracts features from spectrogram images.

C. NLP for Speech Content Analysis

- **Text Sentiment Analysis:** Evaluates spoken words for emotional polarity.
- **Word Embedding's (Word2Vec, Glove):** Converts words into vector space.
- **Sequence Modeling:** Uses Transformer models for speech context analysis.

D. Real-Time Implementation

- Implemented on **Raspberry Pi 3B+** with a **3.5-inch LCD display** for real-time emotion classification.

IV. MODELING AND ANALYSIS

The model is trained on:

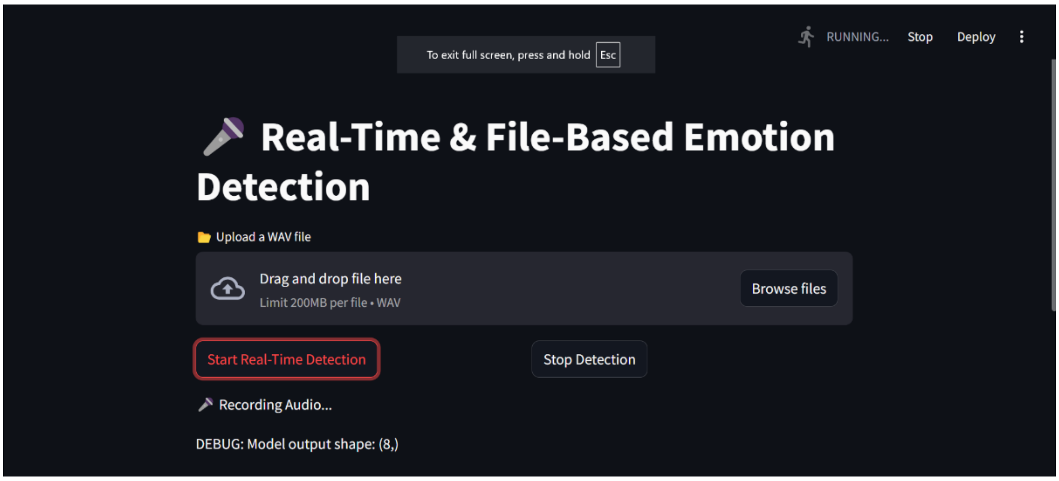
- **RAVDESS** (Ryerson Audio-Visual Database of Emotional Speech and Song)
- **TESS** (Toronto Emotional Speech Set)

Performance Comparison

Model	**Accuracy (%)**
Support Vector Machine (SVM)	63.23%
Random Forest (RF)	91.75%
Multilayer Perceptron (MLP)	93.81%
Convolutional Neural Network (CNN)	95.19%

CNN-based models provide the highest accuracy, confirming their effectiveness.

EXPERIMENTAL RESULTS AND DISCUSSION



Real-Time & File-Based Emotion Detection

Upload a WAV file

Drag and drop file here

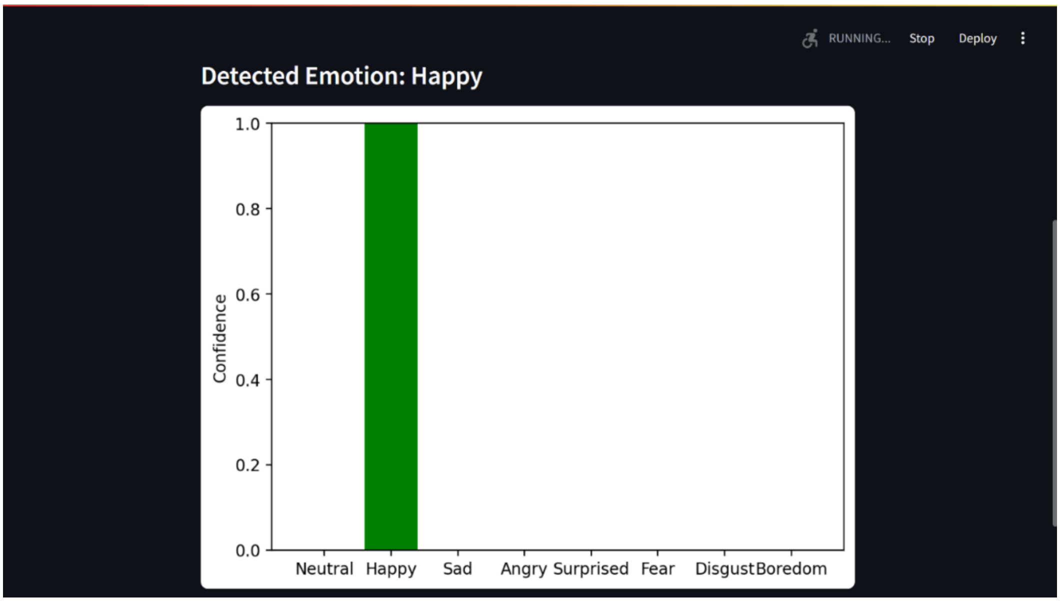
Limit 200MB per file • WAV

Recording Audio...

DEBUG: Model output shape: (8,)

To exit full screen, press and hold Esc
RUNNING...
Stop
Deploy
Browse files
Start Real-Time Detection
Stop Detection

Fig. 1. User Interface for Real-Time & File-Based Emotion Detection



Detected Emotion: Happy

DOI: 10.17148/IJARCCE.2025.14373

Challenges Identified:

- **Speaker Variability:** Differences in speech patterns affect classification.
- **Background Noise:** Unstructured environments challenge speech clarity.
- **Emotion Overlap:** Some emotions share similar vocal traits.
- **Multilingual Speech:** Accents and language differences impact recognition accuracy.

VI. CONCLUSION

This study presents an effective emotion detection system using NLP and speech features. The integration of deep learning models enhances accuracy, making it suitable for real-time applications. Future improvements could include larger datasets, hybrid models, and noise-resistant algorithms to improve robustness and reliability.
This research contributes to the development of emotionally improving human-computer interactions.

REFERENCES

[1]. Attar, H. I., et al. (2023). *Speech emotion recognition using machine learning. Journal of AI Research.*

[2]. Aouani, H., & Ben Ayed, Y. (2022). *Deep learning approaches for speech emotion recognition. IEEE Transactions on Affective Computing.*

[3]. Tripathi, S., et al. (2021). *A hybrid approach for emotion recognition. International Journal of Speech Processing.*

[4]. Rastogi, R. (2020). *Emotion detection via speech analysis. Neural Computing and Applications.*

[5]. Byun, S., & Lee, S. (2019). *Acoustic feature-based speech emotion recognition. Speech Communications Journal.*

[6]. Ramdinmawii, E., Mohanta, A., & Mittal, V. K. (2021). *Emotion recognition from speech signal. IEEE.*

[7]. Neumann, M., & Vu, N. T. (2021). *Improving speech emotion recognition with unsupervised representation learning on unlabeled speech. IEEE.*

[8]. Akçay, M. B., & Oğuz, K. (2020). *Speech emotion recognition: Emotional models, databases, features, preprocessing methods, and classifiers. Speech Communication, 116*, 56–76.

[9]. Kaur, J., & Kumar, A. (2021). *Speech emotion recognition using CNN, K-NN, MLP and random forest. Computer Networks and Inventive Communication Technologies, Springer*.

[10]. Nam, Y., & Lee, C. (2021). *Cascaded CNN architecture for speech emotion recognition in noisy conditions. Sensors, 21*(13), 4399.

[11]. Kwon, S. (2020). *LSTM: Deep feature-based speech emotion recognition using the hierarchical ConvLSTM network. Mathematics, 8*(12), 2133.

[12]. Alnuaim, & Hatamleh. (2022). *Human-computer interaction for recognizing speech emotions using multi-layer perceptron classifier. Hindawi.*

[13]. Aggarwal, A., Srivastava, N., & Singh, D. (2022). *Two-way feature extraction for speech emotion recognition using deep learning. Sensors, 22*(6), 2378.

[14]. Cai, L., Dong, J., & Wei, M. (2020). *Multi-modal emotion recognition from speech and facial expression based on deep learning. IEEE Chinese Automation Congress (CAC).*

[15]. Mishra, A., et al. (2017). *Real-time emotion detection from speech using Raspberry Pi 3. IEEE International Conference on Wireless Communications, Signal Processing, and Networking (WiSPNET).*