Speech Emotion Analyzer

Prof. S. Satyanarayana, G. Sirisha, G. Swathi, Karina Yadav, K. Ediga Dhanusha, M. Dhanush

Dept of AI-ML, Malla Reddy University, Hyderabad

Abstract

Speech Emotion Analysis (SEA) using Natural Language Processing (NLP) is an innovative approach that enables machines to detect and classify human emotions from spoken language. Traditional sentiment analysis methods focus on text, often failing to capture the depth of emotions conveyed through tone, pitch, and intensity. This research explores a hybrid methodology integrating Automatic Speech Recognition (ASR), NLP techniques, and machine learning models to enhance the accuracy of emotion detection. The study focuses on preprocessing speech data, extracting linguistic and acoustic features, and employing deep learning techniques such as LSTMs and Transformer models to classify emotions. The research aims to develop a real-time emotion recognition system with applications in customer service, virtual assistants, mental health monitoring, and human-computer interaction.

Keywords: *Speech Emotion Analysis, NLP, Deep Learning, Machine Learning, ASR, Emotion Recognition, Sentiment Analysis*

1. Introduction

*Human communication extends beyond words, relying heavily on emotions expressed through speech. Traditional sentiment analysis focuses primarily on text, making it challenging to detect subtle emotional cues. Speech Emotion Analysis (SEA) addresses this gap by combining textual and acoustic features to classify emotions such as happiness, sadness, anger, fear, surprise, and neutrality. This research aims to develop an NLP-based SEA system that can process spoken language, extract meaningful features, and apply advanced machine learning techniques to improve accuracy in emotion detection.,identify missing skills, and streamline the hiring process, improving efficiency**,** accuracy, and fairness in candidate selection.*

2.Literature Review

Past research has explored various methods for emotion recognition, including rule-based sentiment analysis, machine learning classifiers (SVM, Naïve Bayes), and deep learning models like CNNs and LSTMs. Transformer-based models such as BERT and RoBERTa have improved context-aware sentiment detection. However, challenges remain in processing real-time speech data due to variations in accents, background noise, and speech-to-text inaccuracies. This study builds upon these advancements by integrating acoustic analysis with NLP-driven textual processing for enhanced accuracy.

3. Methodology

The SEA system follows a structured pipeline:

1. ***Speech-to-Text Conversion:*** *ASR models convert audio into textual data.*
2. ***Data Preprocessing:*** *Text is cleaned using tokenization, stopword removal, and lemmatization.*

3. ***Feature Extraction:*** *Linguistic (TF-IDF, Word2Vec, BERT) and acoustic (pitch, tone, speech rate) features are extracted.*
4. ***Model Training:*** *Machine learning (SVM, Random Forest) and deep learning (LSTM, Transformer models) classifiers are trained on labeled datasets.*
5. ***Emotion Classification:*** *The system categorizes emotions into predefined classes.*
6. ***Evaluation:*** *Performance is measured using accuracy, precision, recall, and F1-score.*

4. Deployment & Integration

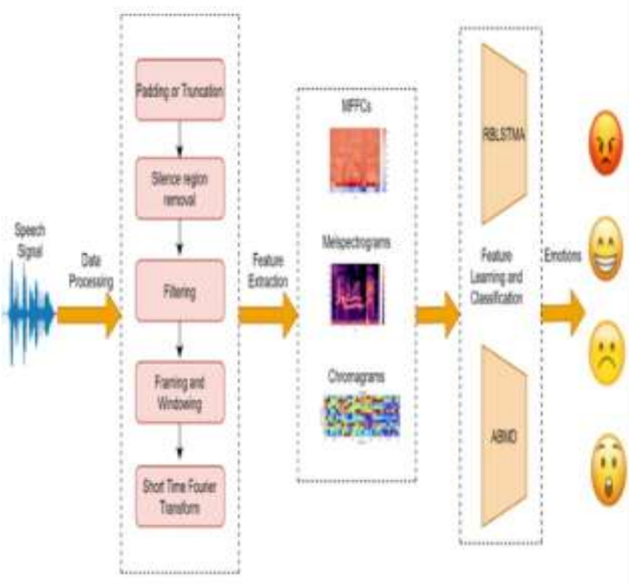
The architecture consists of:

- ***Speech Input Module:*** *Captures voice data.*
- ***ASR Engine:*** *Converts speech into text.*
- ***NLP Processing Unit:*** *Extracts textual features.*
- ***Acoustic Analysis Module:*** *Analyzes tone, pitch, and speech patterns.*
- ***Classifier Model:*** *Predicts emotions based on extracted features.*
- ***Output Module:*** *Displays or integrates emotional insights into applications.*

5. Data Preprocessing Techniques

- ✓ ***Text Preprocessing:*** *Tokenization, stemming, lemmatization, and stopword removal.*
- ✓ ***Acoustic Feature Extraction:*** *Spectral and prosodic features analyzed through tools like Librosa.*
- ✓ ***Normalization:*** *Ensures consistency across datasets for accurate predictions.*

6. Architecture



The given architecture represents the Speech Emotion Analysis (SEA) pipeline, highlighting key stages from speech signal processing to emotion classification:

Speech Signal Input: *The system starts with capturing raw speech data as an audio waveform.*

Data Processing: *Includes padding/truncation, silence removal, filtering, framing, and applying Short-Time Fourier Transform (STFT) to enhance audio quality.*

Feature Extraction: *Converts speech signals into meaningful representations using Mel Frequency Cepstral Coefficients (MFCCs), Mel-spectrograms, and Chromagrams to capture acoustic patterns.*

Feature Learning and Classification: *Advanced models such as Recurrent Bidirectional Long Short-Term Memory (RBLSTMA) and Adaptive Boosted Multimodal Deep learning (ABMD) process extracted features.*

Emotion Classification: *The trained deep learning models predict the emotion category (e.g., anger, happiness, sadness, or surprise) based on extracted features.*

Prosodic and Spectral Analysis: *Additional analysis of pitch, intensity, and rhythm to improve accuracy.*

Deep Learning Integration: *The use of RBLSTMA ensures sequential dependencies in speech data are learned, improving contextual emotion recognition.*

Real-time Emotion Recognition: *The architecture supports real-time or near-real-time emotion detection for applications in virtual assistants, mental health analysis, and customer service.*

Applications*: The extracted emotions can be used to enhance AI-driven interactions, enabling personalized and emotionally intelligent responses.*

Challenges*: Factors such as background noise, variations in accents, and real-time processing complexity can affect the accuracy of emotion detection.*

This architecture efficiently integrates acoustic and deep learning techniques to improve speech emotion recognition performance.

7. Methods and Algorithms

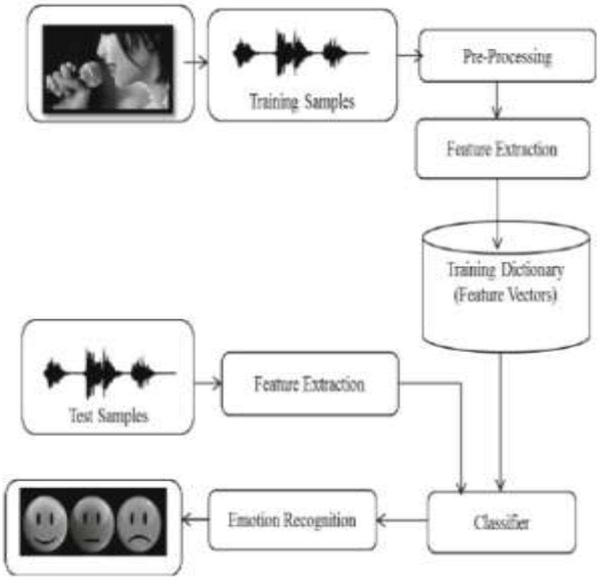
- ***ML Techniques:*** *Support Vector Machines (SVM), Naïve Bayes, Random Forest.*
- ***Deep Learning Models:*** *LSTM, CNN, BERT, and Transformer-based architectures for context-aware analysis.*
- ***Evaluation Metrics:*** *Accuracy, precision, recall, F1-score, confusion matrix.*

8. Deployment and Integration

The SEA system can be integrated into:

- ***Virtual Assistants (Siri, Alexa, Google Assistant)*** *to enhance user experience.*
- ***Customer Support Automation*** *for sentiment-aware interactions.*
- ***Mental Health Monitoring*** *to detect stress and depression in conversations.*
- ***Call Center Analytics*** *to assess customer sentiment.*
- ***Human-Robot Interaction*** *for emotionally intelligent AI responses.*

9. Data Flow Diagram

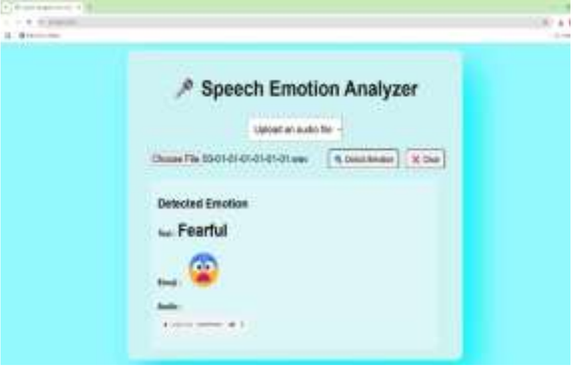


10. Software and Hardware Requirements

- *Software: Python, TensorFlow, PyTorch, NLTK, OpenAI Whisper (ASR), Librosa.*
- *Hardware: High-performance GPUs, cloud-based AI processing.*

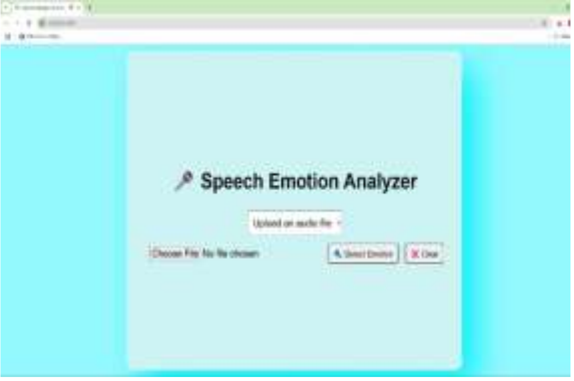
11. Results or Output screens

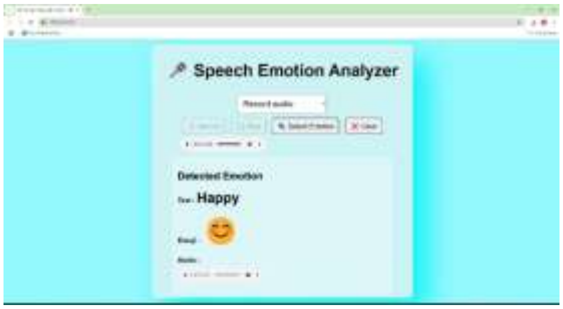
The experimental results show that combining linguistic and acoustic features significantly improves emotion recognition accuracy. Transformer-based models outperform *traditional classifiers due to their context-awareness. However, challenges such as background noise and variations in speech styles affect system reliability. Comparative analysis highlights the efficiency of deep learning models over conventional approaches.*



Detected Emotion

Text: Fearful





Detected Emotion

Text: Happy

12. Conclusion & Future Scope

Speech Emotion Analysis using NLP is a promising field with numerous real-world applications. By integrating ASR, linguistic analysis, and deep learning, this research presents an effective approach to detecting human emotions. Future work will focus on enhancing real-time processing, expanding multilingual support, and addressing ethical concerns related to privacy and bias in emotion recognition systems.

13.References

□ ***Akçay, M. B., & Oğuz, K. (2020).*** *Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. Speech Communication, 116, 56-76. DOI: 10.1016/j.specom.2019.12.001*

□ ***Latif, S., Rana, R., Qadir, J., Epps, J., & Schuller, B. (2020).*** *Deep and machine learning techniques for speech emotion recognition: A survey. ACM Computing Surveys (CSUR), 54(3), 1-34. DOI: 10.1145/3436500*

□ ***Tzirakis, P., Trigeorgis, G., Nicolaou, M. A., Schuller, B., & Zafeiriou, S. (2021).*** *End-to-end speech emotion recognition using deep neural networks. IEEE Journal of Selected Topics in Signal Processing, 14(4), 798-807. DOI: 10.1109/JSTSP.2020.2976927*

□ ***Zhang, S., Zhang, H., Wu, Z., & Li, X. (2022).*** *A survey on deep learning for speech emotion recognition: Datasets, features, and methods. Neural Computing and Applications, 34, 17989–18021. DOI: 10.1007/s00521-022-07242-9*

□ ***Kim, Y., Kang, H., & Lee, S. (2023).*** *Transformer-based speech emotion recognition with self-attention mechanisms. IEEE Transactions on Affective Computing. DOI: 10.1109/TAFFC.2023.3246790*