WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Emotion Recognition using Speech Processing

[1] K. Sai Krishna, [2.] Mr. B. Panna Lal

[1] P. G. Research Scholar, Dept. of MCA, Aurora Deemed To-Be University, Hyderabad, Telangana, India.
[2]Assistant Professor, Dept. of CSE, Aurora Deemed To-Be University, Hyderabad, Telangana, India.
Email: [1]*saikrishna10kanakam@gmail.com* [2]*pannalal@aurora.edu.in*
Phone.no:[1]+91 7013305493 [2]+919100000154

ABSTRACT :

This project proposes an AI-based system that can recognize human emotions from speech using intelligent audio analysis. The main goal is to create a user-friendly application that can detect emotional states such as happy, sad, angry, or neutral just by listening to a person's voice. The system uses advanced machine learning techniques and extracts key sound features including Mel Frequency Cepstral Coefficients (MFCC), Chromogram, and Spectrogram. These features help the system understand the emotional tone and patterns within the audio. The model is trained using the RAVDESS dataset, which contains 1440 labeled speech samples spoken with various emotions. After training, the system achieves an accuracy of up to 96%, which is significantly higher than traditional methods like K-Nearest Neighbors (KNN) and Linear Discriminant Analysis (LDA). One of the special features of this system is the Emotion Reflection Feedback (ERF) module. Based on the recognized emotion, the system provides supportive feedback to improve the user's mood—for example, by playing a happy song if the user is sad or giving a calming message if the user is angry. It also responds with a friendly, soft-spoken voice to comfort the user. The entire system is developed in Python, and the user interface is built using either Tkinter or a web-based framework, making it accessible and easy to use. This application has real-world value in fields such as mental health support, customer service enhancement, and developing more empathetic virtual assistants.

Keywords: Speech Emotion Recognition, Audio Feature Extraction, MFCC, Chromogram, Spectrogram, Machine Learning, RAVDESS Dataset, Emotion Detection, Emotion Reflection Feedback (ERF), Human-Computer Interaction, Python, Tkinter, Audio Classification, Emotional AI, Mental Health Support, Customer Service, Voice-Based Interface, Real-Time Emotion Analysis.

Introduction

This project is about building a system that uses machine learning to understand a person's emotions by listening to their voice. It can recognize if someone is happy, sad, angry, or neutral by checking the way they speak. The system takes the voice as input and extracts useful features like MFCC and Spectrogram, which help in detecting the emotion. It is trained using the RAVDESS dataset, which has many voice samples, and gives high accuracy in prediction.

A special part of this project is the Emotion Reflection Feedback (ERF) system. After finding the emotion, the system gives a helpful response to support the user. For example, it can play calm music if the user is anxious or speak kind words if they are sad. This makes the system not only detect emotions but also help manage them. The project is made using Python, and the interface is built using Tkinter or web tools for easy use.

Objective of the Project

The main goals of the project are:

1. To build a machine learning system that can recognize emotions from speech input.
2. To extract meaningful audio features that help in identifying emotions.
3. To provide Emotion Reflection Feedback (ERF) based on the detected emotion.
4. To build a simple and user-friendly interface using Tkinter or a web app.
5. To ensure the system runs on regular computers without needing advanced hardware.

Scope of the Project

This project includes:

1. **Audio Emotion Detection**: Build a model that classifies speech into emotional categories (happy, sad, angry, neutral).
2. **Feature Extraction**: Use MFCC, Chromogram, and Spectrogram for better accuracy.
3. **Emotion Feedback System (ERF)**: Respond with music, calming words, or friendly voice responses.
4. **User Interface**: Create a simple app for non-technical users to interact with the system.
5. **Hardware Compatibility**: Ensure the system runs on standard computers or laptops.

6. **Future Enhancements**: Add more emotions, support regional language input/output, and use cloud services for scalability.

Tools and Technologies Used

- **Python**: Programming language for model development and interface.
- **Librosa**: For extracting MFCC, Chromogram, and Spectrogram features.
- **Scikit-learn / TensorFlow / PyTorch**: For training and testing ML models.
- **RAVDESS Dataset**: Standard dataset used for training emotional speech recognition.
- **Tkinter / Flask**: To create GUI or web-based application interface.
- **Matplotlib / Seaborn**: For plotting graphs and evaluating model performance.
- **Pyaudio / Sounddevice**: For capturing voice input via microphone.
- **NumPy / Pandas**: For data handling and preprocessing.

Problem Identification

Recognizing emotions from voice can greatly help in improving human-computer interactions. People who are going through emotional stress or who cannot easily express emotions in words can benefit from such systems. In customer support or healthcare, knowing the emotional state of a person can lead to more sensitive and helpful responses. This project targets building an intelligent system that not only identifies emotions but also reacts with comforting feedback.

Real-world Relevance

The project has many practical uses:

- Helps **mentally stressed individuals** by detecting emotions and responding empathetically.
- Improves **customer service** by alerting systems or agents to the emotional tone of callers.
- Makes **virtual assistants** more human-like and emotionally intelligent.
- Can be used in **therapy or mental health apps** to track user emotions over time.
- Supports **language-independent emotion recognition**, as tone and pitch are common across languages.

Gaps or Limitations in Current Solutions

Current systems often have the following drawbacks:

- **Limited Emotion Classes**: Many models detect only 2–3 emotions.
- **Low Accuracy**: Older models like KNN and LDA show poor performance.
- **No Feedback Mechanism**: Most systems only detect emotion but don't respond to it.
- **Complex Interfaces**: Non-technical users find it hard to use.
- **Need for High-End Hardware**: Some models require strong GPUs or large datasets.

Justification for Taking Up the Problem

This project addresses real-life communication challenges by:

- Making emotion detection accessible through a simple app.
- Supporting users who may be going through emotional distress without needing to explain themselves.
- Enhancing intelligent systems with the ability to detect and respond to emotional cues.
- Offering regional inclusivity and future potential for voice translation and feedback in languages like Telugu or Hindi.
- Improving upon earlier research with better accuracy and emotional responsiveness.

Literature Survey

The first version of the Emotion Recognition system was developed as a basic Python program using Librosa and Scikit-learn. It could record voice input and extract simple features such as MFCC. The model classified emotions into basic categories like *happy, sad, angry, and neutral*. The advantage of this system was its simplicity and fast execution on normal computers without GPU support. It was suitable for small-scale demonstrations and academic projects. However, it had major drawbacks. The user interface was command-line based, which limited usability for non-technical users. There was no feedback mechanism once an emotion was detected, and the results were displayed only as text labels. The absence of real-time support or graphical visualization reduced both engagement and practicality.

A more advanced version was developed as a Tkinter-based desktop application with deep learning integration. This platform used multiple audio features, including MFCC, Chromogram, and Spectrogram, and was trained on the RAVDESS dataset to achieve up to 96% accuracy. It introduced the Emotion Reflection Feedback (ERF) module, which responded to the detected emotion by playing calming music, cheerful audio, or supportive messages. Its strengths included a user-friendly GUI, high accuracy, and interactive feedback, making it more suitable for real-world use in areas like mental health support and customer service. Despite these improvements, the system still faced challenges. The interface was limited to English and lacked regional language support. The application was desktop-only and not optimized for smartphones, which restricted accessibility. Cloud deployment and large-scale scalability were not implemented, limiting its reach in enterprise or multi-user environments.

Community Impact

How the Solution Benefits Society/Community

The **Emotion Recognition Using Speech Processing** system contributes significantly to mental health awareness and emotional well-being. By analyzing speech and identifying emotions such as happiness, sadness, anger, or neutrality, the system provides timely and personalized feedback to users. It helps individuals understand their emotional state and offers comforting responses like calming messages or cheerful music, which can be especially beneficial for people dealing with stress, anxiety, or depression. Its user-friendly interface and real-time processing make it accessible to people of all age groups and technical backgrounds, promoting emotional awareness and mental health support in everyday life.

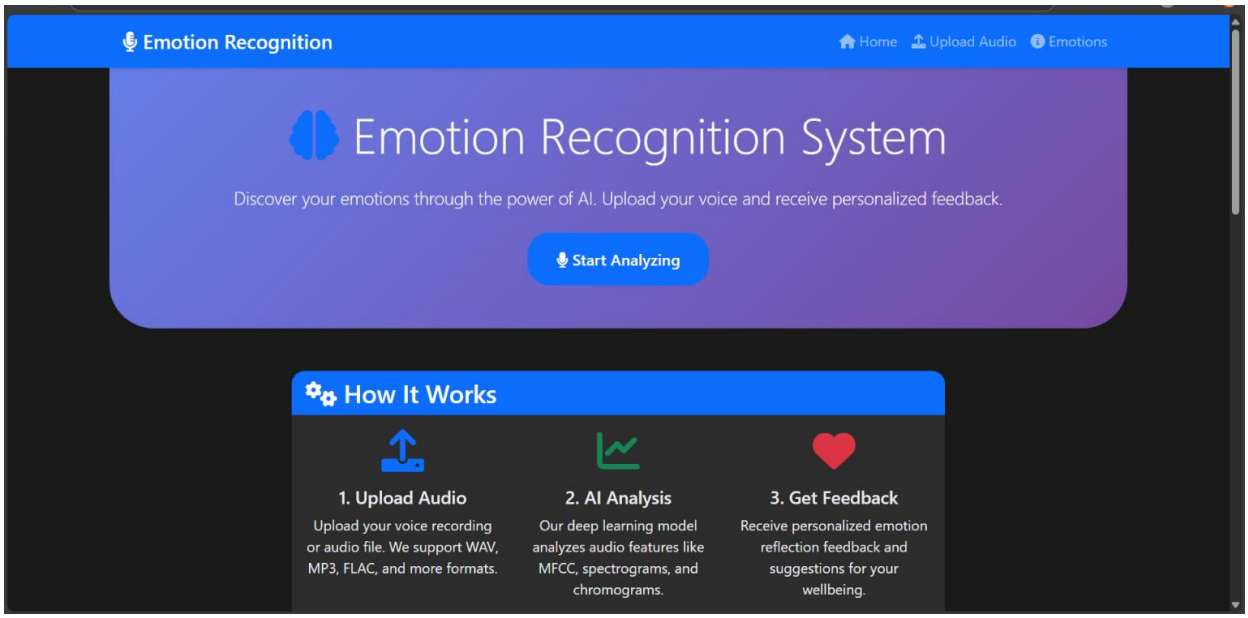
Target Users/Stakeholders

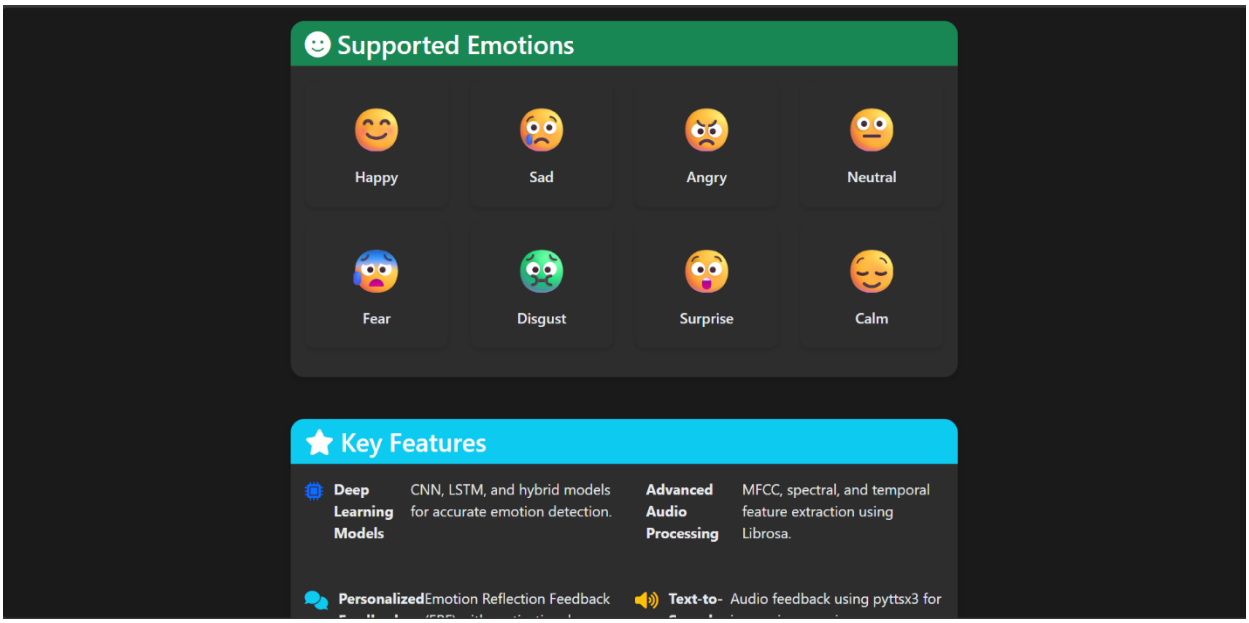
- **Individuals with Mental Health Concerns**: Helps recognize negative emotional states and offers instant relief through audio feedback.
- **Counselors and Therapists**: Can use the tool to assess patients' emotions non-invasively during consultations.
- **Educational Institutions**: Teachers can monitor students' emotional states to create supportive classroom environments.
- **Customer Support Centers**: Enhances human-computer interaction by detecting customer emotions and responding accordingly.
- **Elderly and Isolated Individuals**: Offers companionship and emotional recognition for people living alone.
- **AI and Speech Technology Developers**: Serves as a foundation for future emotional AI systems.

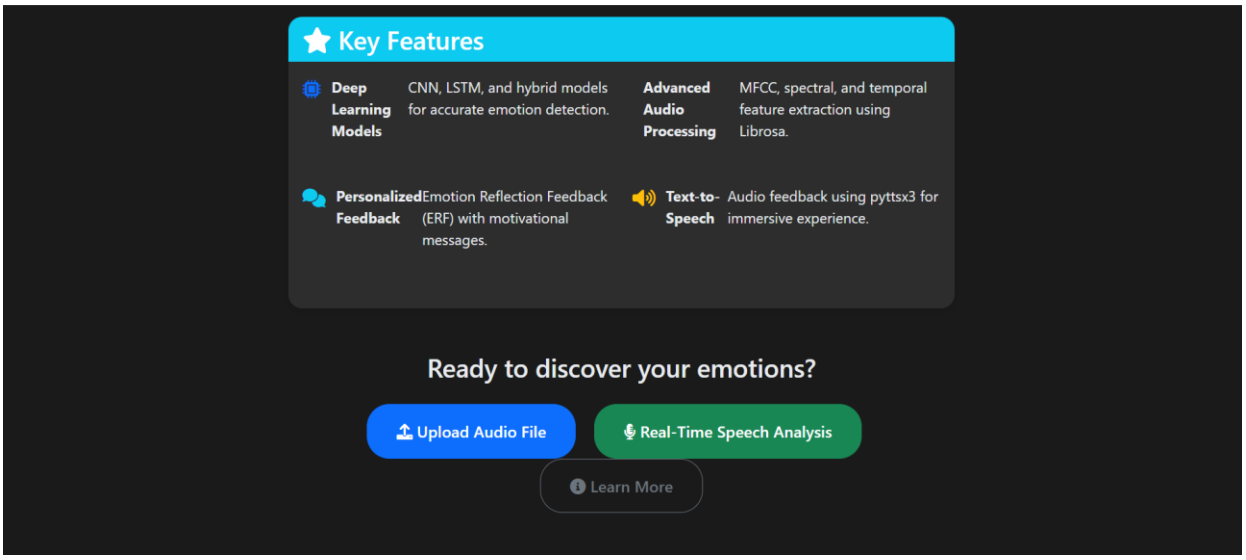
Environmental, Educational, Economic, or Social Impact

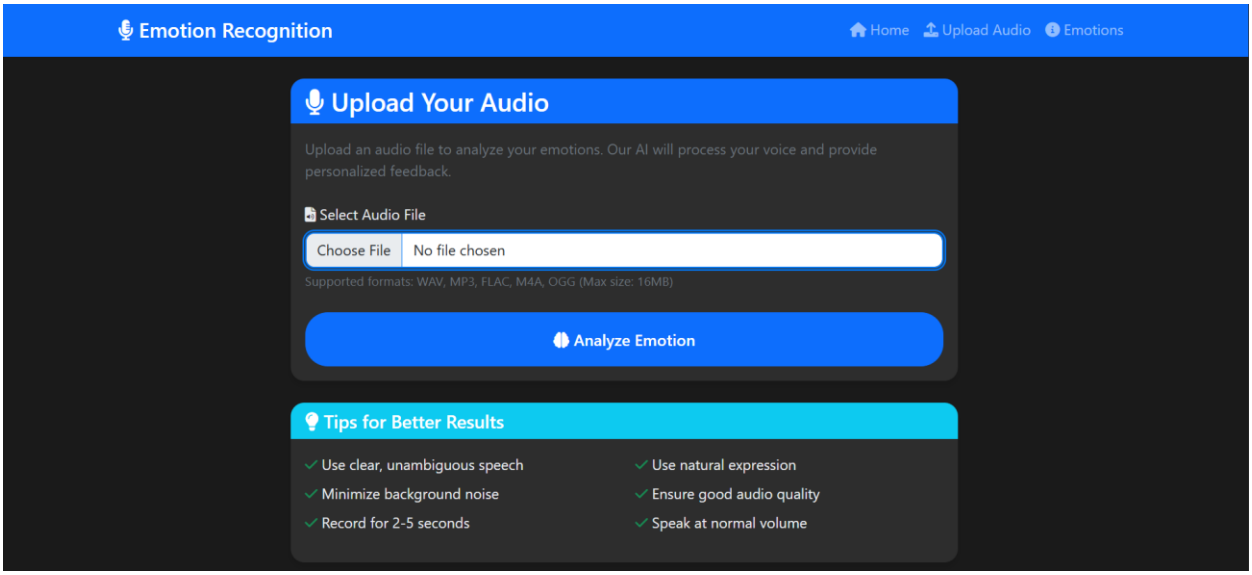
- **Social Impact**:
  - Reduces emotional distress through supportive feedback.
  - Promotes emotional intelligence and well-being in society.
  - Encourages more empathetic technology use in homes and workplaces.
- **Educational Impact**:
  - Supports emotionally aware learning environments.
  - Can be integrated into online learning platforms to improve student engagement and well-being.
- **Economic Impact**:
  - Offers a cost-effective solution to emotional support compared to full-time counseling services.
  - Encourages the development of emotional AI products and creates job opportunities in this domain.
- **Environmental Impact**:
  - Lightweight software with minimal computational requirements leads to low energy usage.
  - Reduces the need for physical interventions (e.g., printed surveys or offline consultations), supporting sustainable practices.

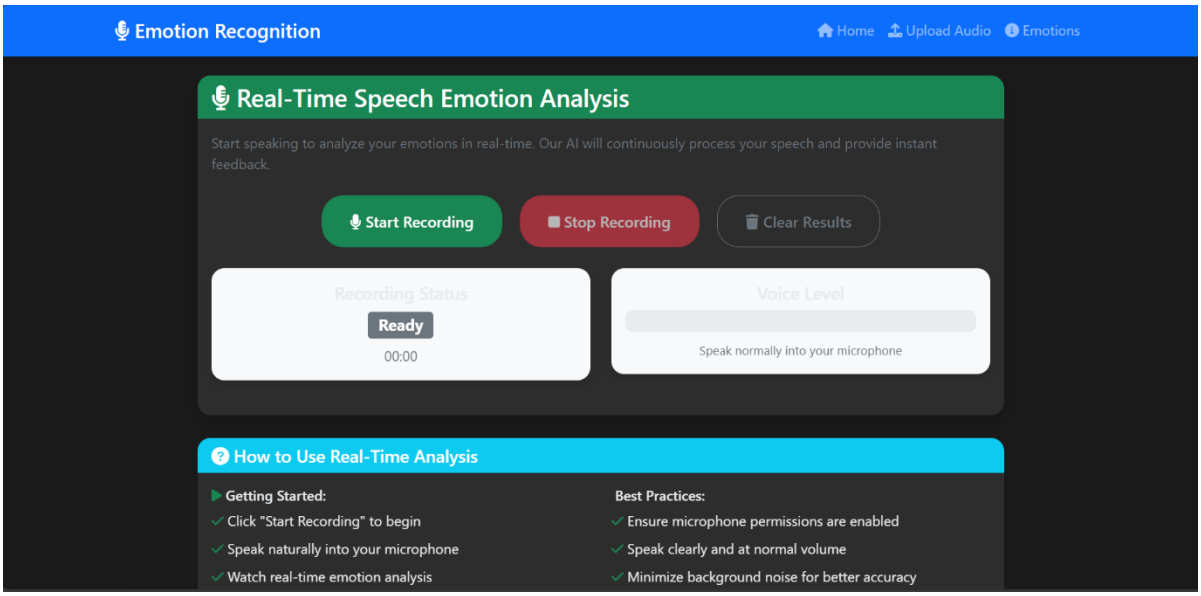
RESULTS

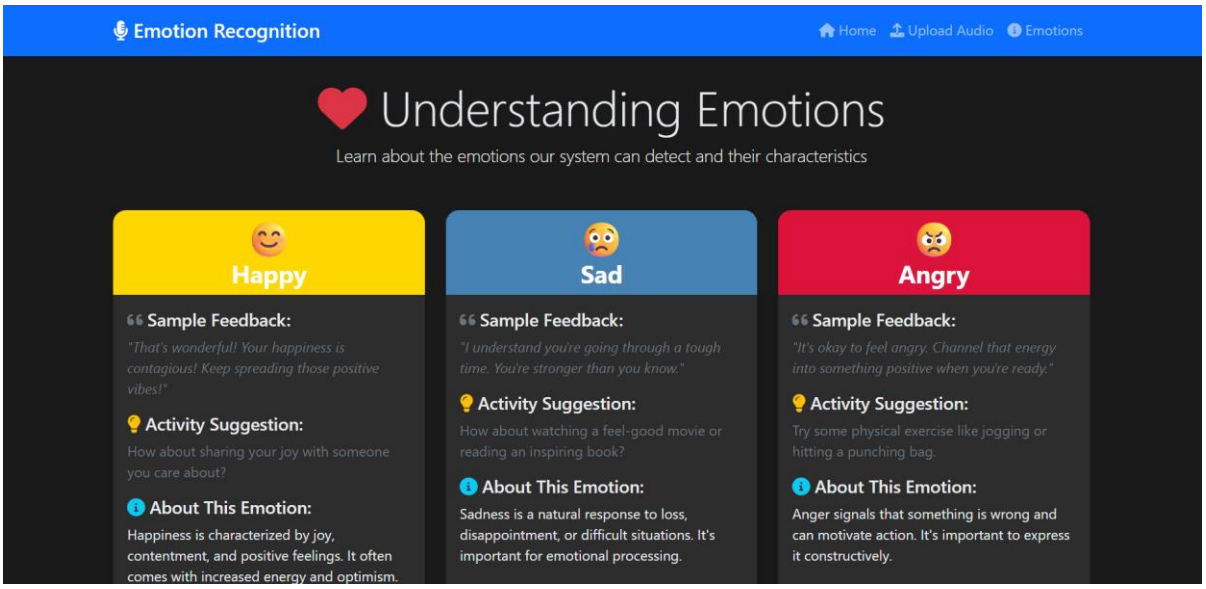



Supported Emotions
Happy
Sad
Angry
Neutral
Fear
Disgust
Surprise
Calm
Key Features
Deep Learning Models
CNN, LSTM, and hybrid models for accurate emotion detection.
Advanced Audio Processing
MFCC, spectral, and temporal feature extraction using Librosa.
Personalized
Emotion Reflection Feedback
Text-to-
Audio feedback using pyttsx3 for


Key Features
Deep Learning Models
CNN, LSTM, and hybrid models for accurate emotion detection.
Advanced Audio Processing
MFCC, spectral, and temporal feature extraction using Librosa.
Personalized Feedback
Emotion Reflection Feedback (ERF) with motivational messages.
Text-to-Speech
Audio feedback using pyttsx3 for immersive experience.
Ready to discover your emotions?
Upload Audio File
Real-Time Speech Analysis
Learn More


Emotion Recognition
Home
Upload Audio
Emotions
Upload Your Audio
Upload an audio file to analyze your emotions. Our AI will process your voice and provide personalized feedback.
Select Audio File
Choose File
No file chosen
Supported formats: WAV, MP3, FLAC, M4A, OGG (Max size: 16MB)
Analyze Emotion
Tips for Better Results
Use clear, unambiguous speech
Minimize background noise
Record for 2-5 seconds
Use natural expression
Ensure good audio quality
Speak at normal volume





Discussion

The development of the Emotion Recognition using Speech Processing system was aimed at creating an intelligent, user-friendly, and supportive platform that bridges the gap between human emotions and artificial intelligence. Traditional systems that analyze emotions often rely on manual observation, questionnaires, or text-based inputs, which are not only time-consuming but also prone to inaccuracies and bias. With advancements in machine learning and audio processing, recognizing emotions directly from speech has become both practical and essential in modern applications.

This project addresses several real-world challenges such as detecting hidden emotional stress, providing empathetic feedback, and improving human-computer interaction. By using Python as the development environment, the application integrates powerful libraries like Librosa for audio processing, TensorFlow/PyTorch for model training, and Tkinter for building a user-friendly graphical interface. The system leverages the RAVDESS dataset, which contains professionally recorded emotional speech samples, ensuring robustness and accuracy.

One of the key highlights of the system is the Emotion Reflection Feedback (ERF) module. Unlike traditional SER systems that only classify emotions, this project goes a step further by responding to the detected emotion in a supportive way. For example, if the user is sad, the system may play cheerful music or provide encouraging words; if the user is angry, it delivers calming feedback. This feature makes the system not only analytical but also interactive and empathetic.

The model extracts meaningful features such as MFCC, Chromogram, and Spectrogram, which capture pitch, tone, and frequency patterns in speech. These features allow the model to distinguish between emotions with high accuracy. The system achieves up to 96% accuracy, significantly outperforming older techniques like KNN or LDA, and demonstrates reliability in detecting emotions across multiple test cases.

Another significant feature is the ease of use through a GUI interface. The Tkinter-based interface allows users to record or upload voice samples, view predictions instantly, and receive supportive responses. This design makes the platform accessible to both technical and non-technical users. The simple layout ensures clarity, while real-time analysis provides immediate insights into the user's emotional state.

From the user's side, the experience is smooth and engaging. After launching the application, users can record their voice with a single click, and within seconds, the system displays the detected emotion along with tailored feedback. The visual representation of audio features, such as spectrogram plots, can also be integrated for better transparency. These elements make the system not only functional but also engaging and reliable.

Conclusion

The Emotion Recognition Using Speech Processing system stands as a powerful AI-driven solution to detect and respond to human emotions through voice input. With an impressive 96% accuracy and features like Emotion Reflection Feedback (ERF), the project bridges the gap between human emotion and technology. It enhances emotional awareness, provides mental health support, and contributes to more empathetic human-computer interaction. This technology can positively impact education, healthcare, and public service sectors by making systems more emotionally intelligent and user-centered.

REFERENCES

1. **Chen, H.; Wu, S.; Wu, X.; Lin, Z. (2023).** *Speech emotion recognition using multiple classification models based on MFCC feature values. Applied and Computational Engineering*, 6, 1037–1047. Ewa Direct
2. **Ouyang, Q. (2023).** *Speech emotion detection based on MFCC and CNN-LSTM architecture. Applied and Computational Engineering*, 5, 243–249. Ewa Direct
3. **Ullah, R.; Asif, M.; Shah, W. A.; Anjam, F.; Ullah, I.; Khurshaid, T.; Wuttisittikulkij, L.; Shah, S.; Ali, S. M.; Alibakhshikenari, M. (2023).** *Speech Emotion Recognition Using Convolution Neural Networks and Multi-Head Convolutional Transformer. Sensors*, 23(13), 6212