AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

DEPARTMENT OF COMPUTER ENGINEERING

2025-2026

Project Synopsis

on

"Speech Emotion Recognition for Early Detection of Depression and Mental Health Issues"

BE Computer Engineering

BY

Group Id-A15

Dighe Gayatri Ramesh (Roll No.2201104423)

Bhosale Sarika Raosaheb (Roll No.2201104407)

Dhage Prajakta Anil (Roll No.2201104420)

Gayake Sayali Ramprasad (Roll No.2201104429)

Dr. G. D. Puri
Project Guide
Dept. of Computer Engineering

Dr. D. R. Patil
Project Coordinator
Dept. of Computer Engineering

Dr. S. K. Sonkar
H.O.D
Dept. of Computer Engineering

- **Title: Speech Emotion Recognition for Early Detection of Depression and Mental Health Issues**
- **Domain and Sub-domain: Artificial Intelligence (AI); Speech Processing, Deep Learning, and Mental Health Analytics**
- **Objectives:**
  - **Speech Data Preprocessing Feature Extraction:** To curate and preprocess speech datasets using techniques like noise reduction, segmentation, and normalization, and extract vocal features—such as pitch variability, prosody, and speech monotony—that are indicative of depressive symptoms.
  - **Depression-Specific SER Model Development:** To develop a Speech Emotion Recognition (SER) model by identifying vocal features such as pitch variability, prosody, and speech monotony that correlate with depressive symptoms.
  - **Comparative Analysis of ML and DL Techniques:** To compare machine learning and deep learning approaches including SVM, Random Forest, CNN, LSTM, and Transformer models for enhanced emotion classification accuracy.
  - **Real-Time, Privacy-Preserving Emotion Monitoring:** To design a real-time monitoring system that tracks emotional patterns and facilitates early intervention through mobile or web applications, ensuring user privacy.
  - **Clinical Integration and Validation:** To collaborate with healthcare professionals for validating predictions and integrating system alerts into clinical support workflows.
- **Abstract:**

Depression and mental health disorders are growing global concerns, yet early detection remains challenging due to subjective clinical assessments and societal stigma. Speech Emotion Recognition (SER) offers a promising, non-invasive

solution by analyzing vocal biomarkers—such as pitch variability, speech rate, and spectral energy—that correlate with emotional and cognitive states.

While existing SER systems excel in classifying basic emotions (e.g., happiness, anger), they lack specificity for clinically relevant mental health conditions like depression, which is characterized by distinct vocal patterns (e.g., monotony, slowed speech, and reduced intonation).

This project proposes an AI-driven SER framework tailored for early depression detection by integrating advanced deep learning techniques with clinically validated speech features.

Key innovations include:

- **Clinical Focus** – Unlike generic SER models, this system prioritizes depression-specific vocal biomarkers.
- **Hybrid Deep Learning** – Combining CNN for spectral feature extraction and LSTM for temporal dynamics improves detection accuracy.
- **Real-World Deployment** – A lightweight, privacy-preserving application enables continuous monitoring, facilitating early intervention.

Preliminary results demonstrate strong correlation between predicted vocal markers and clinical depression scores, suggesting potential for integration into telehealth platforms. By automating early screening, this research aims to reduce reliance on subjective assessments, lower healthcare barriers, and improve mental health outcomes. Future work includes expanding multilingual support and collaborating with clinicians for real-world validation.

- **Keywords:**
  Audio Signal Preprocessing,CNN Model,Deep Learning,Depression Detection,Real-Time Emotion Monitoring,Spectral Feature Analysis,Speech Emotion Recognition (SER),Transfer Learning.

- **Problem Definition:**
  Depression and mental health disorders are major global health concerns, yet early detection remains challenging due to subjective assessments and societal stigma. Traditional diagnostic methods (e.g., questionnaires, clinical interviews)

are time-consuming and rely heavily on self-reporting, which may be inaccurate or incomplete.

Speech, as a non-invasive biomarker, reflects emotional and cognitive states through variations in tone, pitch, and rhythm. However, existing Speech Emotion Recognition (SER) systems often emphasize general emotions (e.g., happiness, anger) rather than clinically relevant mental states (e.g., sadness, fatigue, flat affect).

There is a critical need for an automated, real-time SER system that can detect subtle vocal patterns associated with depression and mental health disorders. Such a system would enable early intervention, reduce diagnostic delays, and support broader efforts in mental health monitoring and awareness.

- **List of Modules:**

  1. Audio Preprocessing and Feature Extraction
  2. Emotion Classification Model
  3. Depression Detection and Reporting Dashboard

- **Current Market Survey:**
  In 2025, mental health and digital wellness sectors are expanding rapidly, driven by growing emotional distress and demand for accessible screening tools. Despite this growth, most platforms—such as Wysa, Woebot, and Replika—offer chatbot-based CBT support and lack passive, voice-driven emotion monitoring.

  Academic research highlights deep learning models like CNN, BiLSTM, and Transformer for Speech Emotion Recognition (SER), but few commercial tools implement real-time depression detection. Common limitations include low support for multilingual speech, lack of real-time tracking, and poor integration with clinical workflows.

  Recent literature explores methods such as Mixture-of-Experts and CNN-LSTM hybrids, yet most remain confined to experimental setups. Rural and underserved communities often lack access to mental health professionals, making non-invasive, speech-based solutions essential.

- **Scope of the Project:**

  The project focuses on speech-based emotion recognition for depression detection. It includes modules for voice input, preprocessing, emotion classification, and depression prediction. The system is intended for early screening and emotional monitoring—it does not provide diagnosis confirmation or treatment recommendations. Initially, it will support English speech with scope for future expansion to other languages and dialects.

- **Literature Survey:**

  1. Manolekshmi, I., et al. (2025). Hybrid CNN-LSTM and CNN-GRU models with Ensemble SVM for speech emotion recognition in human-computer interaction. Achieved 98.69% accuracy on the CREMA-D dataset.
  2. Alhussein, G., et al. (2025). Meta-analysis of AI-based SER models across 51 studies. Focuses on multimodal fusion, dataset diversity, and annotation reliability.
  3. Elsayed, N., et al. (2024). CNN + GRU-based supervised recurrent system for mental health monitoring via intelligent virtual assistants.
  4. Barhoumi, C., et al. (2024). Mixture-of-Experts with TDNN and Transfer Learning for depression detection via speech. Demonstrates 74.3% cross-lingual accuracy.
  5. Rajapakshe, T., et al. (2024). Use of GANs, SpecAugment, and oversampling techniques to optimize SER model generalization and performance.
  6. Deshmukh, R., et al. (2024). NeuroVibeNet: A hybrid ML ensemble (RF, LightGBM, SVM-KNN) for early mental health prediction based on behavioral and vocal signals. Achieves 99.06% accuracy.
  7. Ibrahim, A., et al. (2024). Benchmarking SER generalization across 11 datasets using Whisper and transformer-based models. Available at: `https://arxiv.org/pdf/2406.09933`
  8. Shanthi, V., et al. (2025). Integrated mental health assessment using FER, SER, scales, and doctor chat modules. Published in IET Health Technology. Link: `https://ietresearch.onlinelibrary.wiley.com/doi/pdfdirect/10.104`

9. Bautista, E., & Shin, H. (2025). Joint modeling of discrete and dimensional emotion representations in SER using Wav2Vec2 and HuBERT. Published in *Applied Sciences*, 15(2), 623. Available at: `https://www.mdpi.com/2076-3417/15/2/`

- **Software and Hardware Requirement of the Project:**

  Software:

  1. Python (v3.8 or above) – Programming language for model development
  2. TensorFlow / Keras – Building and training CNN and LSTM-based models for Speech Emotion Recognition
  3. scikit-learn – Implementing SVM, Random Forest, and evaluation metrics
  4. Librosa – Audio feature extraction (MFCCs, Chroma, Mel Spectrograms)
  5. Jupyter Notebook / Google Colab – Interactive development and experimentation
  6. Matplotlib / Seaborn – Data visualization and result plotting
  7. Pandas / NumPy – Data handling and manipulation

  Hardware:

  1. Processor – Intel i5 / Ryzen 5 or above
  2. RAM – Minimum 8 GB (16 GB recommended for smooth training)
  3. GPU (optional but preferred) – NVIDIA GPU with CUDA support (e.g., GTX 1050 Ti or higher) for faster training and inference
  4. Microphone – For capturing speech input in real-time
  5. Headphones / Speakers (optional) – For audio output and testing feedback

- **Contribution to Society:**

  This project bridges the gap between AI research and real-world mental health needs by offering a scalable, speech-based depression detection system. It empowers users, especially in rural areas, to monitor emotional well-being from home and reduces reliance on subjective assessments and limited clinical access.

  The system enables early identification of emotional distress, aids professionals with real-time insights, and sets the stage for future innovations in affective computing. Ultimately, it supports mental health awareness and aligns with Digital India and Telehealth initiatives through inclusive, technology-driven care.

- **Probable Date of Project Completion:** January 2026

- **Outcome of the Project:**

  1. Emotion and Depression Detection Model
  2. Annotated Speech Dataset
  3. Real-Time Monitoring Interface
  4. Clinical Relevance Validation
  5. Open Research Contribution