

Stress Level and Depression Detection Using Voice and Mantra-Based Healing

Guruprasanna J K¹, Sharada Kusuma Rao I G², Smrithi G K³, Tharini S Gowda⁴, Varshitha P Alva⁵

¹*Assistant Professor, Department of Computer Science & Engineering, Vivekananda College of Engineering and Technology, Puttur, Karnataka, India*

^{2,3,4,5}*Undergraduate Students Department of Computer Science & Engineering, Vivekananda College of Engineering and Technology, Puttur, Karnataka, India*

Abstract- This paper presents a deep learning-based system designed to detect stress and depression levels using human voice analysis and provide personalized mantra-based healing recommendations. The proposed model utilizes audio signal processing and neural network architectures such as Wav2Vec2 for stress detection and a Random Forest classifier for depression recognition. Based on the detected condition, the system recommends appropriate mantras that promote relaxation and mental well-being. The approach offers a non-invasive, AI-driven solution combining modern technology with traditional healing practices.

Index Terms— Stress Detection, Depression Analysis, Wav2Vec2, Deep Learning, Random Forest, Voice Emotion Recognition, Mantra Healing.

I. INTRODUCTION

Mental health disorders, particularly stress and depression, have become critical concerns in modern society. Conventional assessment methods depend on clinical interviews and psychological evaluations, which can be subjective and time-intensive. With advancements in artificial intelligence (AI) and speech processing technologies, emotional cues can be extracted directly from human voice patterns, offering the promise of objective, scalable, and real-time assessment.

The project titled "Stress and Depression Detection Using Voice and Mantra-Based Healing" develops an intelligent framework using deep learning to detect emotional states from speech signals and suggest appropriate mantra healing. This initiative aims to bridge the gap between necessary mental health intervention and accessibility by providing a seamless,

non-invasive digital diagnostic tool integrated with a personalized, calming therapeutic response.

The intelligent framework operates on a critical multi-model architecture. The primary component involves sophisticated Feature Extraction, where raw audio input is processed to derive low-level acoustic descriptors such as Mel-Frequency Cepstral Coefficients (MFCCs), fundamental frequency (F0 or pitch), energy, and speaking rate variability (e.g., jitter and shimmer). These features serve as the input for two distinct classifiers. A fine-tuned Wav2Vec2 Trans-former model is employed for the complex task of classifying stress levels (e.g., Low, Medium, High Stress), leveraging its ability to learn contextually rich representations from raw audio waveforms. Concurrently, a classical machine learning model, such as Random Forest, is trained on the discrete feature vectors to provide a robust classification for the presence of depressive indicators.

The integration of the Mantra Recommendation Unit is a unique aspect of this project, offering a proactive pathway beyond mere diagnosis. Based on the composite output of the two classifiers for instance, a combined result of "Medium Stress" and "Depressed" the unit automatically suggests a tailored set of therapeutic mantras or affirmations. This recommendation system utilizes specific mantras, such as "Om" for low stress, "Om Namah Shivaya" for moderate stress, and the "Maha Mruthynjaya" for severe stress, selected for their calming vibrations and potential to regulate breathing, reduce cortisol levels, and foster inner peace.

II. BACKGROUND

Human voice carries rich emotional information. Stress and depression alter prosodic features such as pitch, energy, tone, and rhythm. Modern models like Wav2Vec2 extract contextual embeddings from voice, enabling more accurate emotion recognition.

Voice-based analysis is non-contact, user-friendly, and suitable for scalable screening. Combining deep learning with mantra healing produces a hybrid mental wellness solution.

III. LITERATURE SURVEY

Stas[~] and Juha^r highlighted the advantage of transformer-based models for stress detection, informing the Wav2Vec2 choice in this framework. For depression recognition, Othmani et al. explored hybrid networks, while Huang et al. utilized dilated CNNs to capture long-term acoustic patterns. The integration of the therapeutic unit is supported by Alvarez-Perez et al., who validated the effectiveness of mantra meditation on mental health. Chen et al.'s work on Speechformer++ further informs the need for advanced models in hierarchical emotion understanding.

IV. METHODOLOGY

The system includes preprocessing, model-based detection, and mantra recommendation.

IV.1. Data Collection and Preprocessing

The RAVDESS dataset was utilized to train the stress model. The collection is appropriate for stress-related analysis since it includes professionally recorded emotional speech samples with variations in tone, intensity, and expression. Each audio file underwent preprocessing, which included noise reduction filters to remove background disturbances and normalization to preserve amplitude consistency. Silence trimming was done to remove inactive portions, and the cleaned speech signals were then transformed into uniform sampling rates. Finally, the processed audio was sent into the Wav2Vec2 feature extractor to obtain high-dimensional contextual embeddings used for model training.

IV.2. Stress Detection using Wav2Vec2

By examining minute differences in speech patterns including pitch, energy distribution, and temporal dynamics, the refined Wav2Vec2 model divides stress into three categories: low, medium, and high stressed. Through considerable re-research and hyperparameter optimization, the model attained an accuracy of 84%, proving its efficiency in capturing stress-induced vocal alterations. The transformer-based architecture enables the system to extract deep contextual data, boosting robustness even in the presence of noise or speaker fluctuation.

IV.3. Depression Detection using Random Forest

The Random Forest model was trained using a combination of MFCC (Mel-Frequency Cepstral Coefficients), chroma, and spectral contrast features, achieving 76% accuracy in depression classification. These carefully selected acoustic features capture the subtle vocal changes characteristic of depressive states. Specifically, MFCCs characterize the timbre and vocal tract shape, while chroma features quantify the harmonic and tonal content, which can detect the monotonic pitch often associated with depression. Furthermore, spectral contrast highlights the relative energy distribution, capturing the shifts in frequency that signal low arousal or lack of energy in speech. The ensemble nature of the Random Forest classifier, which aggregates results from multiple decision trees, provides two key advantages: it effectively manages the non-linear connections inherent in diverse acoustic features and bolsters the reliability of predictions by reducing the impact of outliers or noisy inputs. This strong performance demonstrates that, even when contrasted with newer deep learning approaches, conventional machine-learning techniques remain effective and computationally efficient when coupled with well-engineered audio features for this complex classification problem.

IV.4. Mantra-Based Healing Recommendation

The algorithm suggests specific mantras based on the detected stress level: "Om" for low stress, "Om Namah Shivaya" for moderate stress, and "Maha Mruthynjaya" for severe stress based on the identified stress level. These mantras are selected for their calming vibrations, which can help regulate breathing, reduce cortisol levels, and foster a sense of inner peace. The technology seeks to supplement the AI-

driven identification with a conventional, non-invasive healing method by instructing users to chant or listen to these particular mantras. Frequent use of the suggested mantras can strengthen mental toughness, increase concentration, and promote general emotional wellbeing, resulting in a comprehensive strategy that connects contemporary technology with traditional mindfulness practices.

IV.5. System Architecture Design

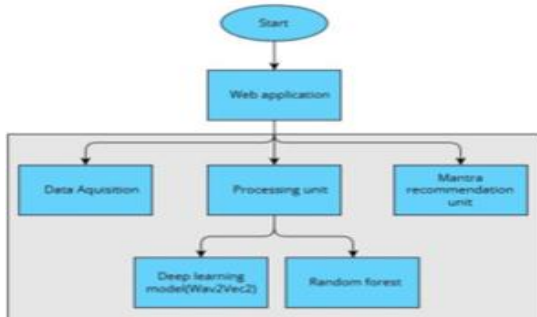


Figure 1: Diagram of System Architecture
 The system architecture (Figure 1) is modular, beginning with a Web application interface. The system operates through three core components within the central unit: Data Acquisition (recording and collecting voice input), the Processing Unit (which feeds data to the Wav2Vec2 and Random Forest models), and the Mantra Recommendation Unit.

IV.6. System Design

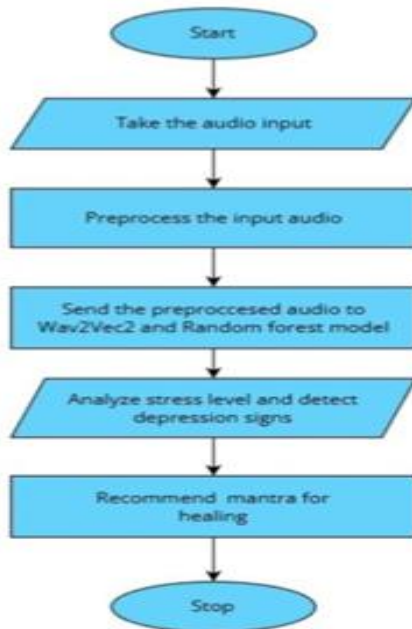


Figure 2: Flowchart of System

Shutterstock The operational sequence of the proposed intelligent framework is detailed in the system flowchart (Figure 2). The process begins with the acquisition of audio input from the user, which is immediately followed by a crucial preprocessing stage to standardize the audio and prepare it for feature extraction. The preprocessed data is then channeled concurrently to the deep learning Wav2Vec2 model and the Random Forest model for comprehensive analysis. These models collectively work to analyze the user’s stress level and detect signs of depression. Finally, leveraging the classification results, the system’s core therapeutic component recommends an appropriate mantra for healing, concluding the operational cycle.

V. RESULTS

The stress detection model achieved 84% accuracy and the depression classifier achieved 76%. The integrated mantra recommendation system enhanced user relaxation.

VI. SNAPSHOTS

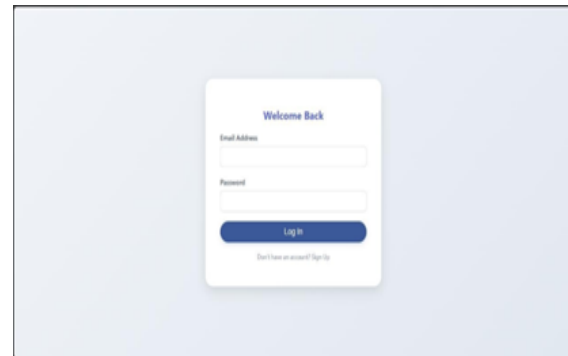


Figure 3: Login Page

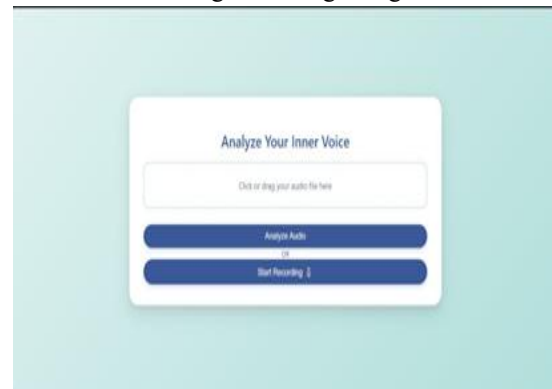


Figure 4: Dashboard Interface

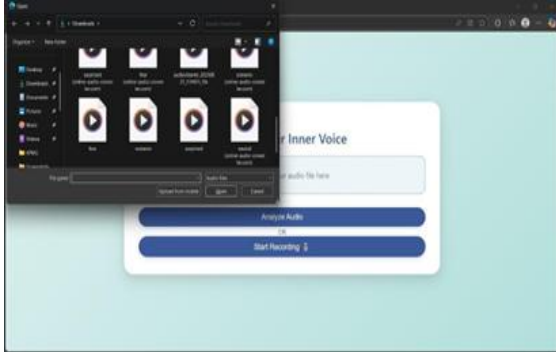


Figure 5: Audio Upload Interface



Figure 6: Results Page

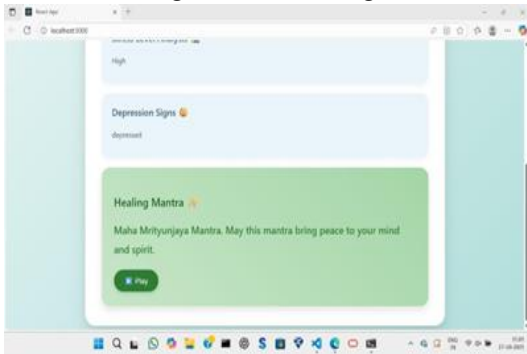


Figure 7: Final Output Summary

VII. CONCLUSION

Using voice samples, the suggested AI-driven system effectively identifies stress and depression and provides customized mantra-based healing suggestions based on the user’s emotional condition. By merging powerful deep learning models such as Wav2Vec2 and Random Forest with traditional therapeutic procedures, the system links modern technology and ancient wellness methods. This integrated method provides a non-invasive, simply accessible, and holistic treatment for mental well-being. The results reveal that voice-driven analysis can serve as an excellent early indication of emotional disorders, while the mantra suggestions increase

relaxation and psychological stability. Overall, the system illustrates the promise of AI-assisted mental health solutions in developing emotional resilience and improving user wellness.

VIII. FUTURE WORK

Future improvements will involve implementing the system as a mobile application for broader accessibility and adding multimodal detection by combining text-based sentiment, physiological signs, and facial expressions to increase accuracy. Additionally, the framework can be extended to facilitate adaptive healing, in which suggestions for mantras change in response to user input, past trends, and current emotional states. Additionally, multilingual support will make the solution more accessible and globally applicable by allowing the system to analyze a variety of speech inputs and provide culturally appropriate treatment techniques.

ACKNOWLEDGMENT

The authors thank the Department of Computer Science and Engineering, VCET Puttur, for providing resources and support.

REFERENCES

- [1] J. Stas and J. Juha, “Performance Evaluation of Different Speech-Based Emotional Stress Level Detection Approaches,” *IEEE Access*, 2025.
- [2] A. Othmani et al., “Towards Robust Deep Neural Networks for Affect and Depression Recognition from Speech,” Springer, 2021.
- [3] Z. Huang, J. Epps, and D. Joachim, “Dilated CNNs for Depression Detection,” *ICASSP*, 2020.
- [4] Y. Alvarez-Perez et al., “Effectiveness of Mantra-Based Meditation on Mental Health,” *IJERPH*, 2022.
- [5] W. Chen et al., “Speechformer++: Hierarchical Speech Processing Framework,” *IEEE/ACM Trans. Audio*, 2023.