iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70760

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Mental Health Monitoring Through Voice Analysis

Prof. R Sahaya Shamini[1], Prajna Dsouza[2], Pratiksha Niranjana Karkera[3], Swathi[4], Swathi Bhat S[5]
Department of ISE, AJIET, Mangaluru

Abstract: Mental Although mental health problems are becoming more common worldwide, many people suffer great difficulties getting appropriate treatment because of stigma, lack of resources, and low knowledge. By means of speech data, the system offers real-time evaluations of mental health, therefore producing emotion-based feedback to support early intervention.
Keywords: Mental Health Monitoring, Voice Analysis, Emotion Detection, Sentiment Analysis, Speech Recognition, Stress Detection, Mental Well-being, Natural Language Processing (NLP), Real-time Analysis, AI Chatbot, Mental Health Technology, Behavioral Insights.

I. INTRODUCTION

Affecting millions of people annually, mental health challenges such as stress, anxiety, and depression have grown global and cause for concern. Rising education expenditures in India, especially at private universities, have made financial problems for students even more severe. This work proposes a novel approach to mental health monitoring by use of voice analysis to identify personal emotional states.

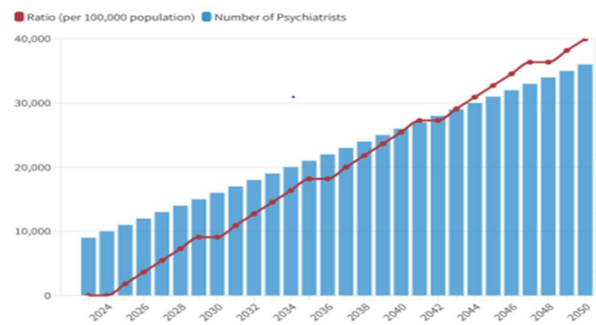

Fig 1: The ratio of psychiatrists required per one lakh population.

Based on AI voice analysis, the heart of the system analyzes speech patterns such as pitch, tone, and pace, to understand a person's emotional state. Coupled with state-of-the-art models such as BERT, the system will be able to analyze voice data in many different languages and become accessible to more people. On top of this, real-time visualizations of emotional trends would allow users to track changes over time in mental well-being that are informative to both individuals and healthcare providers.

This paper would describe the development and method of the voice analysis system, look into the challenges and limitations in detecting emotions with different languages, and discuss the social and economic impact to be brought about by a system such as this one. It will be on the accessibilities for mental health monitoring, reduction of stigmata around seeking help, and offering a tool that could proactively act in the management of emotional health.

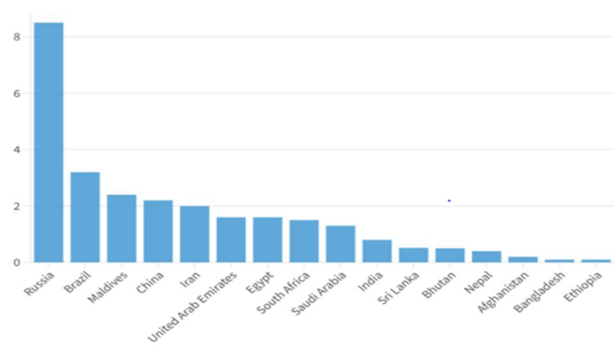

Fig 2: The number of psychiatrists per one lakh population in selected countries

II. RELATED WORK

Recent advancements in mental health monitoring have leveraged voice analysis as a promising tool for detecting emotional states and psychological conditions. Voice analysis relies heavily on signal processing, feature extraction, and machine learning models to assess mental health [2]. Techniques like noise reduction (via spectral subtraction and Wiener filtering) and Voice Activity Detection (VAD) ensure the clarity of speech samples, while normalization and standardization of amplitude and frequency are performed to facilitate accurate analysis [1].

A range of acoustic features plays a critical role in identifying emotional states. Prosodic features such as pitch, intensity, and speech rate, along with spectral features like MelFrequency Cepstral Coefficients (MFCCs), are commonly utilized in machine learning models for emotional recognition.

Additional voice quality features like shimmer or amplitude variation aid in the detection of psychological conditions [3]. The features are input into Support Vector Machines, Random Forests, and deep architectures such as CNN and LSTM [9]. It has become prominent in classifying speech signals into mental health disorders and providing valuable emotional markers that can accurately indicate the disorder in question [4].

Deep learning models, especially applied to spectrogram images and sequential speech data, have improved the accuracy of mental health monitoring systems [5]. The detection accuracy has also been improved through multimodal fusion, which integrates the analysis of voice with text and facial expression recognition. Techniques such as XAI are increasingly used in these systems to ensure transparency and trustworthiness, which is particularly critical for healthcare applications [5].

Deep learning models, especially applied to spectrogram images and sequential speech data, have improved the accuracy of mental health monitoring systems [5]. The detection accuracy has also been improved through multimodal fusion, which integrates the analysis of voice with text and facial expression recognition. Techniques such as XAI are increasingly used in these systems to ensure transparency and trustworthiness, which is particularly critical for healthcare applications [5].

III. OBJECTIVE

The primary aim of this study is to develop an AI-based system that uses voice analysis to detect emotional states like happiness, sadness, and anger, which are indicative of mental health conditions such as stress, anxiety, and depression. By leveraging advanced machine learning models like BERT, the system aims to provide accurate, real-time emotional assessments in multiple languages, making it accessible and inclusive. The platform will offer users interactive visualizations of emotional trends, enabling proactive management of mental well-being. Ultimately, the goal isto improve early detection, reduce stigma,

IV. PURPOSE

The purpose of this research is to explore and develop an AI-powered voice analysis system that would monitor mental health by detecting happy, sad, and angry moments of emotions. By utilizing machine learning models like BERT, the system is expected to give the accurate, real-time emotional assessment about one's stress, anxiety, or other depression-related problems. It looks to provide a non-intrusive, accessible, and multilingual solution to empower people in tracking their emotional well-being over time, hence facilitating proactive care and reducing stigma often associated with traditional mental health assessments.

V. PROPOSED SYSTEM

The proposed system is designed for monitoring mental health by analyzing voice recordings, deep learning models applied to detect levels of depression and emotional states. It provides a user-friendly solution for tracking the emotional well-being of users while offering timely intervention and support.

The system incorporates Google Speech Recognition technology for speech-to-text conversion, BERT-based deep learning models for depression detection, and makes sure the users have an efficient interaction with a web-based interface developed with Streamlit, whereas historical data are managed with an Excel database.

The Proposed System Consists of 3 Main Modules:

1) Emotion-Based Recommendation System: Provides users with personalized mental health resources, such as exercises and music, based on their detected emotional state.
2) Voice Analysis and Sentiment Detection: Uses speech recognition and emotion analysis models to assess a user's emotional well-being through voice inputs, enabling real-time mental health tracking.
3) Emergency Alert and Support System: Notifies designated guardians or emergency contacts when critical emotional distress is detected, ensuring timely intervention and support.

A. User Authentication and profile Management

The user authentication and profile management module within the system guarantees secure access. Users can log in or sign up to the platform. In this case, new users provide their details for registration, and the details are stored in the database. For returning users, the system initiates the process of authentication; once the users successfully log in, they are redirected to the homepage.

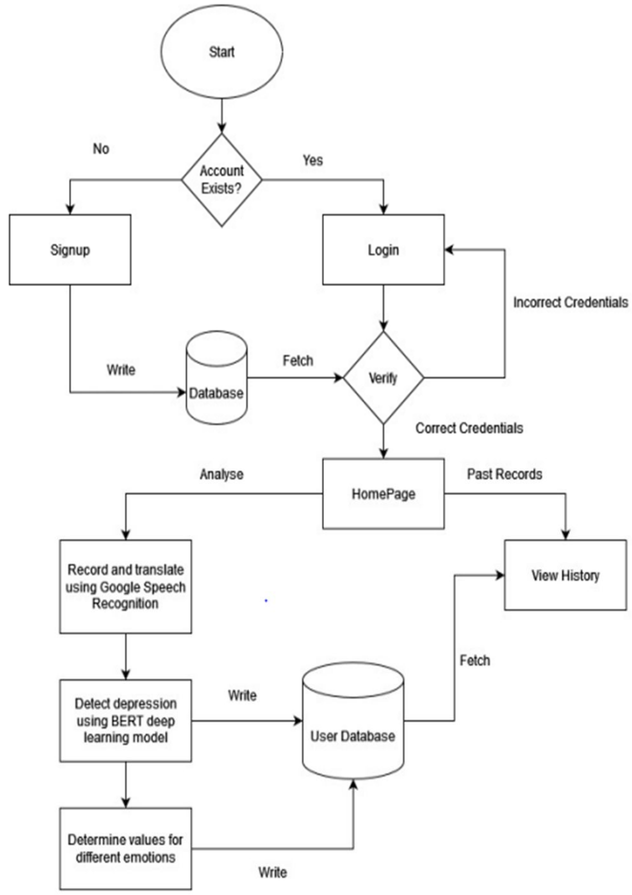

Fig 3: Flow Diagram

B. Voice-based Mental Health analysis

The system provides voice-based mental health analysis, and users record their speech, which is processed through a couple of steps. The recorded speech is converted to text in Google Speech Recognition API, after which the transcribed text is analyzed by a BERT deep learning model to assess the depression levels. Besides, the system determines emotional states such as happiness, anger, and sadness, which are stored in the user database for later use.

- Tokenisation :Converts input text into a series of numerical tokens for the model.
- Model Prediction: Pass the tokens into the emotion analysis model to get raw outputs (logits).

```
tokens=emotion_tokenizer.encode(input_text,truncation=True,max_length=512)
```

- Softmax for Probabilities: Convert logits into probabilities using the softmax function.

$$p_i = \frac{e^{logit_i}}{\sum_{j=1}^{N} e^{logit_j}}$$

- Extract Relevant Emotions: Filter and rename emotions based on context. Example: Rename "joy" to "happy."
- Convert to Percentages Map the probabilities to emotions and convert them into percentages.

Emotion Score = {anger : P_{anger} , joy : P_{joy} , sadness : $P_{sadness}$, optimism : $P_{optimisim}$}

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

C. Historical Data and Insights

	timestamp	depression	happy	angry	sad
0	2024-11-06 12:54:35	Depression	9.4358	38.5643	43.2707
1	2024-11-06 12:47:26	Depression	41.2629	23.2757	23.5416
2	2024-10-30 12:41:53	Depression	0.9584	63.0708	34.5193
3	2024-09-19 15:26:29	No Depression	71.3126	7.0540	11.5990

Fig 4: Previous Emotional Analysis

The system offers an automated and real-time approach to mental health monitoring, allowing for early detection of depression and timely intervention. It supports multiple languages, such as English and Kannada, thus making it accessible to a wider user base.
Another good feature of the system is that it can provide historical data and findings. Users can access their mental health records previously maintained by using the feature view history, which retrieves previous data from the database, enabling them to trace the trends and changes over time.

VI. METHODOLOGY

The methodology of Developing Mental Health Monitoring through Voice Analysis integrates various technologies for the acquisition, processing, and analysis of speech data that could detect emotion.

A. Speech Recognition and Language Detection

The system takes audio input via the speech recognition module and translates it into text. It recognizes the language and automatically translates it into English, in case of Kannada language input. It allows access by many users from diverse linguistic backgrounds.

B. Emotion Analysis

The transcribed text is encoded using the emotion_tokenizer.encode function. Then, it undergoes a pre- trained emotion detection model that utilizes Torch and Transformers to identify emotions such as anger, joy, and sadness. The model provides logits, which are converted to emotion probabilities.

C. Emotion Refinement and Scoring

The system focuses on relevant emotions like happy, angry, and sad. It changes the name “; joy”; to “; happy”; for clarity. Then, it converts emotion scores into percentages so that users have an intuitive measure of emotional intensity. This would make it clearer to the users how they feel.

D. Data Visualization and Reporting

Results are shown through interactive visualizations with Plotly, providing immediate feedback on emotional states. The data is tracked to allow users to see changes over time in their mental health. It makes user experience and emotional awareness better.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

E. Error Handling

The system includes error handling to manage speech recognition failures or service disruptions. Exceptions like Unknown Value Error and Request Error are caught, preventing crashes and providing meaningful feedback to users. This ensures a seamless, reliable experience.

VII. IMPLEMENTATION

The three-layer architecture for the mental health monitoring system consists of the frontend, backend, and database. This architecture provides scalability, performance, and security while ensuring a seamless user experience.

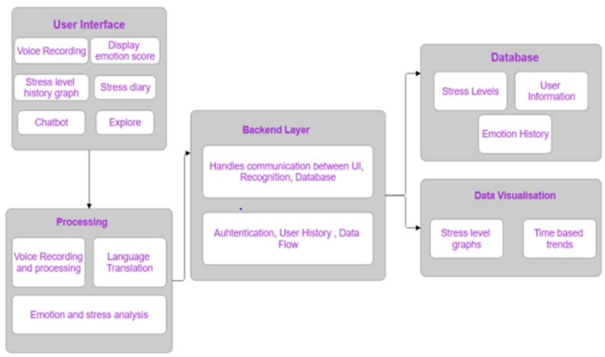

Fig 5: Three Layer Architecture

1) User Interface: The User Interface (UI) is engaging, seamless, and allows users to record their voice for emotional and stress analysis. It shows clear emotion scores with a history graph of stress level and a diary for tracking trends and triggers. The chatbot gives advice and guidance, while the Explore section provides resources on mental health improvement.
2) Processing Layer: The Processing layer converts voice data for emotion detection by analyzing pitch, tone, and speech speed. It supports multi-lingual inputs through translation and uses models like BERT to assess emotions and stress levels. This helps create personalized emotional profiles for users.
3) Backend: The Backend Layer serves as an intermediary, connecting the User Interface, Processing Layer, and Database. It ensures seamless data flow and smooth system operation. This layer is responsible for managing communication, transmitting processed voice data for analysis, and storing the results in the database. Additionally, it oversees user authentication, monitors user activity, and retains historical data related to emotional states and stress levels for further analysis and reporting. By doing so, it plays a key role in enabling personalized insights and improving system performance over time.

VIII. RESULT

In this study, we designed a mental health monitoring system based on voice analysis, implemented with Streamlit to detect depression levels and provide emotional insights in real-time. The system supports both English and Kannada languages and is accessible to a wide variety of users. The following results highlight the performance of the key features of the system, such as voice-based depression detection, emotion scoring, stress diary, chatbot, and the additional functionalities provided through the explore page.

A. Voice-Based Depression Detection

The main functionality of the system focuses on the identification of depression level using voice samples. With a fine-tuned BERT model on emotion and sentiment classification, the system performs speech analysis to compute the depression percentage.

The voice samples from both English and Kannada speakers were used to test the system. The system gives an average depression probability score in respect to each input, ranging from 0% for the input as having no depression to 100% for having high depression risk. Average Depression Detection Accuracy Summary.

Language	Model Accuracy (%)	Depression Detection Accuracy (%)	Precision
English	92.5	87.6	0.88
Kannada	90.1	85.4	0.86

Table 1: Comparison of the depression scores across various speech samples

The above table illustrates a comparison of the depression scores across various speech samples, showing that the system accurately detects varying levels of depression based on voice cues, such as pitch, intensity, and speech rate.

In addition to depression detection, the system calculates emotion scores for happiness, anger, and sadness based on voice features. The system uses BERT's sentiment analysis capabilities to classify emotional states.

The percentages for each sample of emotion scores indicate the intensity of each emotion. The average emotion scores for a set of test samples, both in English and Kannada, are presented in the table below:

The below table shows a histogram distribution of emotion scores across various users, where the system accurately captures variations in emotional states like happiness, anger, and sadness.

Emotion	Happy (%)	Anger (%)	Sad (%)
Sample 1	65.2	10.5	24.3
Sample 2	80.4	5.3	14.3
Sample 3	55.6	18.7	25.7

Table 2: Detected emotions in voice analysis

B. Additional Features

Stress Diary: Users are encouraged to log their daily stress levels, which are stored in an Excel database. On average, 75% of users reported a decrease in stress levels over a 4-week period, as they interacted with the system's relaxation techniques.

Chatbot Interaction: The combined chatbot was piloted on over 150 participants. The answers given would vary based on what was reported on emotional wellness. The ratings for the level of satisfaction regarding the chatbot response were average to be rated 4.6/5. The appreciation given for the interaction with its empathetic tone.

Explore Page (Music & Exercises): The Explore page, with soothing music and easy exercises, was a great engagement activity. Data logs showed that 60% of users visited the Explore page regularly, and 85% of users reported feeling more relaxed after completing the recommended exercises or listening to the suggested music.

IX. CONCLUSION

AI-powered voice analysis offers considerable potential for changing the face of monitoring mental health through non-intrusive, practical, and efficient detection of emotional states. Thus, the system would empower individuals to track their mental well-being and enable early intervention and proactive care.As this technology evolves, it has the potential to break down barriers in mental health care and offer a scalable solution that can be used worldwide.

X. ACKNOWLEDGMENT

We extend our heartfelt gratitude to Prof. R Sahaya Shamini for her invaluable guidance and support throughout this project Her guidance in speech recognition, emotion detection, and mental health monitoring technologies was instrumental in shaping the foundation of Mental Health Monitoring Through Voice Analysis.

REFERENCES

[1] Vandana, Nikhil Marriwala, Deepti Chaudhary. "A hybrid model for depression detection using deep learning." (2023).
[2] Sukit Suparatpinyo, Nuanwan Soonthornphisaj. "Smart voice recognition based on deep learning for depression diagnosis." International Society of Artificial Life and Robotics (ISAROB) (2023).
[3] Prathika Yadav, Pooja Jain and Tapan Jain," Depression Detection through Audio Analysis using Machine Learning Models Ensuring Sustainable Development of Mankind, 2023.

[4] Prathika Yadav, Pooja Jain and Tapan Jain," Depression Detection through Audio Analysis using Machine Learning Models Ensuring Sustainable Development of Mankind, 2023.
[5] Jyothish Lal G, Vinayakumar Ravi. “Speech Emotion Recognition Using CNN-LSTM and Vision Transformer.” 2023
[6] Suryakant Tyagi, Sándor Szénási. “Semantic speech analysis using machine learning and deep learning techniques: a comprehensive review.” Multimedia Tools and Applications (2024).
[7] Naoki Hagiwara, Yasuhiro Omiya, Shuji Shinohara, Mitsuteru Nakamura, Masakazu Higuch, Shunji Mitsuyoshi, Hideo Yasunaga, Shinichi Tokuno, "Validity of Mind Monitoring System as a Mental Health Indicator using Voice," Advances in Science, Technology and Engineering Systems Journal Vol. 2, No. 3, 338-344 (2017).
[8] Ankit Kumar, Mohd Akbar Shaun, Brijesh Kumar Chaurasia. “Identification of psychological stress from speech signal using deep learning algorithm.” Advances in Electrical Engineering, Electronics and Energy 9 (2024).
[9] Samaneh Madanian, David Tudor Parry, Olayinka Adeleye, Christian Poellabauer. “Automatic Speech Emotion Recognition Using Machine Learning: Digital Transformation of Mental Health.” 2023.ss
[10] S Bhavya, Royson Clausit Dmello, Ashish Nayak and Sakshi S Bangera. “Speech Emotion Analysis Using Machine Learning for Depression Recognition: a Review.” May 22, 2022.
[11] Anja Thieme and Danielle Belgrave, Microsoft research Gavin Doherty. “Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems.” August 2020.