iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79283

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

AI-Based Acoustic Intelligence System for Mental Well-Being

Vulugundam Anitha[1], A. Shiva Tejaswini[2], B. Vaijayanthi[3], Bethi Akhila[4], Talveda. Sneha Reddy[5]
Electronics and Telematics Engineering, G. Narayanamma Institute of Technology and Science, Hyderabad, India

Abstract: ***The increasing prevalence of mental health challenges and the need for timely intervention have motivated the development of intelligent, emotion-aware support systems. This paper presents an AI-based Acoustic Intelligence System for Mental Well-Being that integrates Speech Emotion Recognition (SER), Natural Language Processing (NLP), and conversational AI to detect and respond to user emotions in real time. The proposed system processes speech input through pre-processing and feature extraction techniques such as Mel-Frequency Cepstral Coefficients (MFCC), pitch, and energy, followed by classification using machine learning and deep learning models. A hybrid approach combining acoustic features and textual sentiment analysis enhances emotion classification accuracy and robustness.***

The system is implemented using Python, Google Colab, and Streamlit, providing an interactive user interface with modules for memory analysis, entity detection, and contextual conversation management. Experimental results demonstrate reliable performance in recognizing emotions such as happy, sad, angry, and neutral across both speech and text inputs, even under moderate noise conditions. Additionally, the integration of a memory-based context retrieval mechanism enables personalized and context-aware responses. The proposed system highlights the effectiveness of combining SER and NLP for real-time emotional assistance and intelligent mental health monitoring. **Rapid identifier: ML, Google Colab, Streamlit**

I. INTRODUCTION

Mental health has become a major challenge in the modern society due to the high rates of lifestyle change, academic pressure, work pressure, and increased social pressure. Among the psychological problems that an increasing number of individuals have to face are stress, anxiety and depression, which, significantly, affect the emotional health, physical health, and overall quality of life. Even though there is heightened awareness on the issue of mental health problems, a good number of people face difficulties in seeking professional assistance because of absence of mental health professionals, prohibitive cost of treatment and psychological stigma associated with mental diseases [1]. These challenges make it clear that it is vital to implement smart technological changes that can assist people in monitoring their mental condition and enhancing it. The recent advances of the field of Artificial Intelligence (AI) have provided the opportunities to design intelligent systems able to read and comprehend the emotions of a human being. With the help of AI technologies, machines are able to process large amounts of data and make considerable trends in regard to emotional behaviour. Such systems are capable of handling both written and vocal messages to determine emotional states and mental health manifestations with the accompaniment of Natural Language Processing (NLP). Through these technologies, computers are able to learn human language, read feelings, and offer assistance to the users automatically with the assistance of interactive systems. Speech Emotion Recognition (SER) plays an extremely important role in emotion sensitive systems that are developed to observe mental health [2]. Human speech contains some critical emotional clues that have no connection with the verbal utterance. Acoustic features of pitch, tone, loudness, and rhythm of speech can be used to identify emotional condition of a person [3]. It is with these vocal characteristics that SER systems can classify the emotions as either positive, neutral or even negative states of emotion [4]. It is what allows smart systems to detect the signs of stress, anxiety, or emotional pain and respond in order to help its users. The emotion recognition systems have been in a position to maximize their performance with the aid of machine learning and deep learning algorithms [5]. More advanced frameworks such as recurrent neural networks and attention-based frameworks can be trained to learn extremely advanced emotional patterns based on speech and textual information [6]. These models work with the large datasets and are able to automatically pluck meaningful emotional features to augment the category emotional process. Therefore, in the modern world, AI systems are able to recognize emotions an increasing number, and act as aids to human-computer interaction in more intelligent and adaptive way [6]. Another important concept is affectionate computing which is also utilized in emotional conscious system as the idea is aimed at ensuring that the machines are capable of perceiving, analyzing and responding to the human feelings [7]. It is true that affective computing technologies have made conversational agents and intelligent assistants talk to their users in such a way that they are empathetic [8].

©IJRASET: All Rights are Reserved

These systems can provide supportive feedbacks, advice or suggestions based on the emotional condition that is determined about the communication by the user. Affective computing has been greatly applied in health care, education and personal assistants to improve the overall quality of human-computer interaction [7]. In addition to the emotional analysis, AI-based monitoring systems can also identify the occurrence of abnormal circumstances and potential safety threats in real-life scenarios in real-time [1]. Intelligent systems can learn the behavioural characteristics and report abnormal cases that may be requiring immediate treatment. AI technologies can prove effective in the area of mental health tracking and crisis response due to emotion recognition and automated monitoring and alert systems [9].

This kind of systems demonstrates the ways AI-based solutions may be employed to make the environment safer, provide relevant support, and promote the level of well-being. Generally, it can be concluded that the convergence of Artificial Intelligence, Natural Language Processing, and Speech Emotion Recognition is the foundation of the present-day intelligent mental health systems [1], [2], [10]. These technologies can be used to analyze the patterns of speech and communication in order to get emotional states and provide appropriate assistance automatically. Since the emotional recognition systems that are AI-based are still being developed and evolved further through increased research and advancement of technologies, they may become more accessible to mental health assistance and help an individual control their emotional condition more efficiently [9].

II. PROPOSED METHODOLOGY

The suggested Acoustic Intelligence System to Mental Well-Being is aimed at decoding human speech and recognizing their emotional state using the methods of Artificial Intelligence and machine learning. The system is a combination of Speech Emotion Recognition (SER), Natural Language Processing (NLP) and generation of conversational responses to provide emotional support to the user. Architecture of Acoustic Intelligence System for Mental Well-Being as shown in Figure 1.

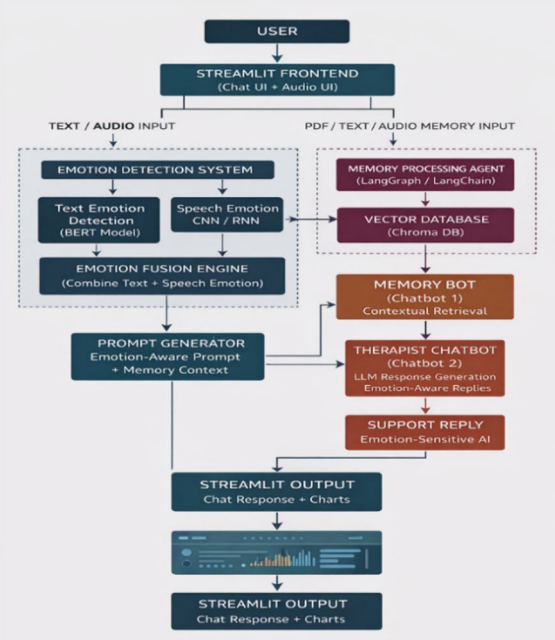


Fig. 1. Architecture of Acoustic Intelligence System for Mental Well-Being

The major aim of the system is to identify the emotional states namely stress, anxiety, happy, or sad off the speech and generate supportive messages in its response. The system is established on the platform of processing information and feature extraction, machine learning models, interaction modules between the conversation session and other modules, to generate an intelligent system that is emotion sensitive. The initial step in the system is acquisition of speech data whereby voice input of the user is captured by the use of a microphone or an audio recorder. This is the most important input in the system, a speech signal. However, the cues of raw speech may also incorporate some background or distortion in the form of noise, and it may also affect the accuracy of the emotion recognition. The audio obtained is therefore undergone to a pre-processing phase where noise reduction, signal normalization and signal segmentation procedures are implemented to refine the quality of the speech data and then it can be further analysed.

Pre-processing is used to ensure that the system extracts significant information of feeling within the speech signal. After the pre-processing of the system, feature extraction follows to convert the raw speech signals into the meaning numbers that can be used by the machine learning algorithms. The speech signals are used in deriving a number of acoustic characteristics which are Mel-Frequency Cepstral Coefficients (MFCC), pitch, energy and spectral characteristics. MFCC is also one of the most prevalent speech processing properties as it is a model of the human perception of the sound frequencies. These extracted features indicate meaningful emotional manifestations of the speech signal and these are inputs of emotion classification models. In the next step, the machine learning and deep learning models are utilized to classify the emotional states, depending on the extracted features.

The algorithms such as Support Vector Machines (SVM), Random Forest, and the deep learning models such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are usually used to solve emotion recognition tasks. They are paradigms that examine the characteristics extracted and prepare patterns that are associated with the different emotional states. The models are trained by using controlled learning in which labelled datasets that consist of speech samples with respect to specific emotions are used. The model can further perceiving user feelings by listening to the user with greater precision upon training. The system uses the Natural Language Processing (NLP) algorithms to derive the textual meaning of what the users are saying and give them suitable reply in the conversation. The NLP also allows the system to interpret the message and sentiment of the word typed or spoken by the user. This system will be capable of providing sympathetic responses, helpful suggestions, or advice to the user through speech emotion recognition combined with NLP-based dialogue processing.

III. RESULTS AND DISCUSSION

Proposed Acoustic Intelligence System of Mental Well-Being is tested to determine its effectiveness in the process of detecting emotions in the speech and textual interactions and responding to them with a supportive dialogue. Some of the technologies integrated into the system include speech recognition, emotion detection, Natural Language Processing (NLP) and conversational AI within a user interface developed in Streamlit. The Memory Analysis Interface, which is shown in Figure 2 helps us see what is stored about how people interact with things. This makes it easier to understand how people behave at times. The Entities Detected Module, shown in Figure 3 points out the things it finds in what people say. This shows that the methods used to understand language are good, at finding the important parts that mean something.

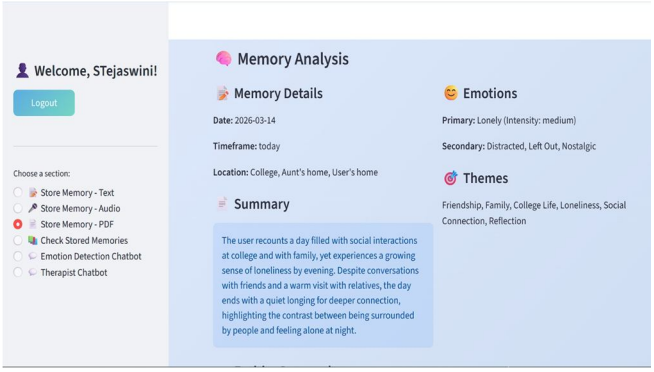


Fig. 2. Memory Analysis Interface

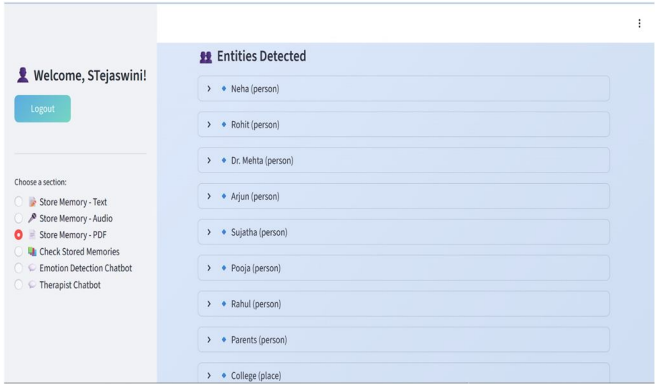


Fig. 3. Entities Detected Module

During testing the system worked well with both voice and text inputs. This made it easy for users to interact with it. The speech processing and NLP modules are combined. This creates an environment to analyze emotions and help mental well-being with smart responses. The Therapist Chatbot Emotion Analysis Interface, in Figure 4 Shows how emotions are understood. It is then used to create responses that make sense in the conversation.

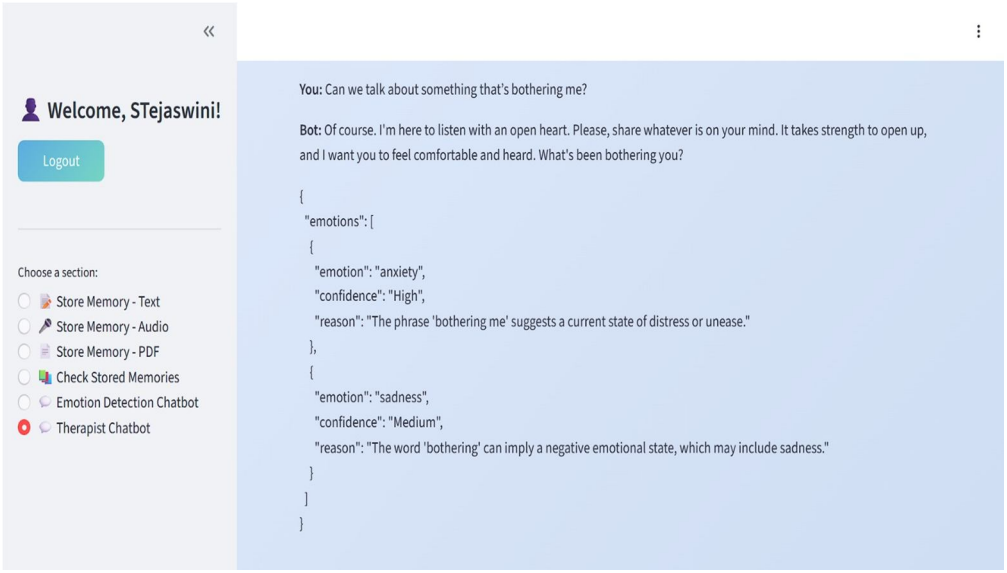


Fig. 4. Therapist Chatbot Emotion Analysis Interface

The emotion detecting facet was put to a test to know how far the system could go in the detection of the different emotional reactions when fed with user inputs. The system was able to recognize feelings of happy, sad, angry and neutral using acoustic features of pitch, tone and intensity and sentiment cues obtained in the text. The experiments presented the results to state that the efficiency of emotion recognition carried out by means of speech is higher in case in which the textual sentiment analysis is deployed to gain higher stability of the emotional labelling.

The hybridized approach will assist the system to be more concerned with the user emotions and generate responses more contextual and emotive. It was also revealed that the speech recognition module was a stable one because it was able to convert the voice inputs into a text data. The speech recognition model based on whispers could recognize speech of a user with a great variety of speaking speed, tone and accents. The Conversation History and Context Retrieval Interface, which is shown in Figure 5 and the Conversation History Module, which is shown in Figure 6 show that the system can store and get back conversations. The system has a memory part that uses vector embeddings. These vector embeddings are stored in a vector database. This helps the system remember things that were talked about before. The Conversation History and Context Retrieval Interface and the Conversation History Module use this memory to give responses. The responses are based on what the user said This is what the Conversation History and Context Retrieval Interface and the Conversation History Module do.

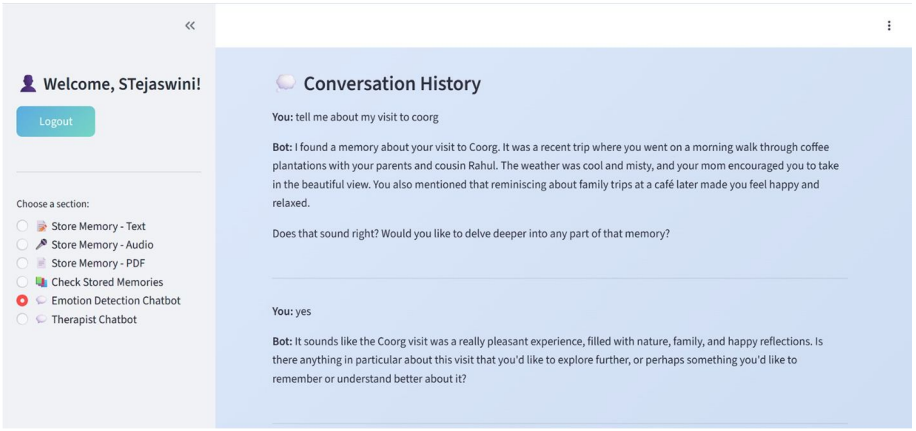


Fig. 5. Conversation History and Context Retrieval Interface

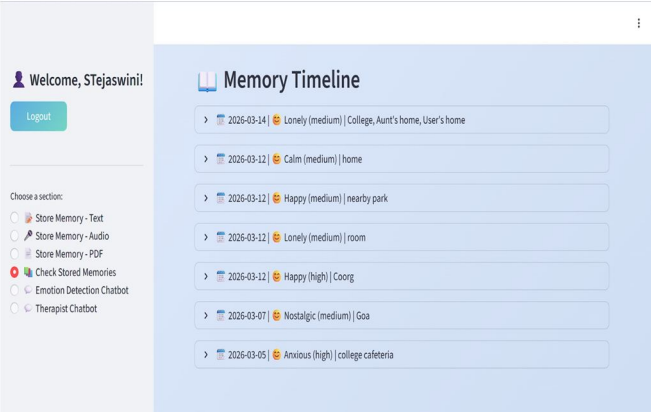


Fig. 6. Conversation History Module

The system still maintained a constant recognition performance and provide close real time transcription under moderate background noise conditions. The Distribution of Emotional States, in Figure 7 shows how emotions changed during user interactions. It highlights emotional trends.

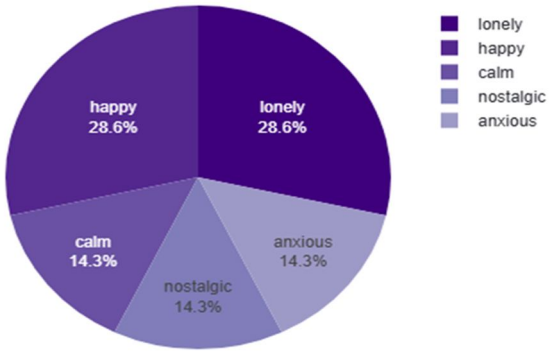


Fig. 7. Distribution of Emotional States

The ability to store the history of conversations and the contextual knowledge is the other salient feature of the system. The memory component recalls historical user interactions in the virtualization of the form of the vector embedding in a vector data bank. The system remembers the past interactions so as to provide more personalized and contextual responses to a new user query. The Count of Different Emotions shows how well we can analyze and visualize emotions (Fig. 8). This information helps us understand how users behave and how their emotions change over time. It also shows that systems that can understand emotions could be really helpful, in keeping an eye on peoples' health.

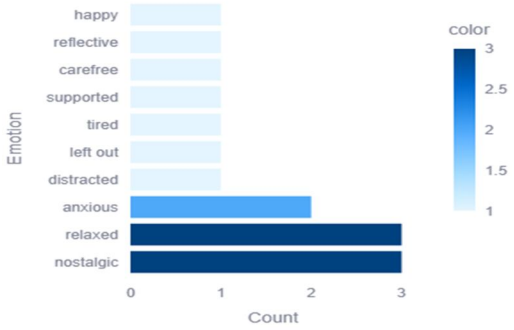


Fig. 8. Count of Different Emotions

The results of the project are also those that entail the visualization of emotional patterns as they were recognized in case of user interactions. Based on emotional distribution analysis, certain emotional states appear common when it comes to user communication behaviour. These emotional patterns may provide useful feedback on user behaviour, and can demonstrate how the ability to track mental health with time can be helped by emotional aware systems.

IV. CONCLUSION

This work presented an AI-based Acoustic Intelligence System for Mental Well-Being that integrates speech recognition, emotion detection, and NLP-driven conversational intelligence into a unified framework. The proposed methodology effectively combines acoustic feature analysis and textual sentiment understanding to improve the accuracy and stability of emotion classification.

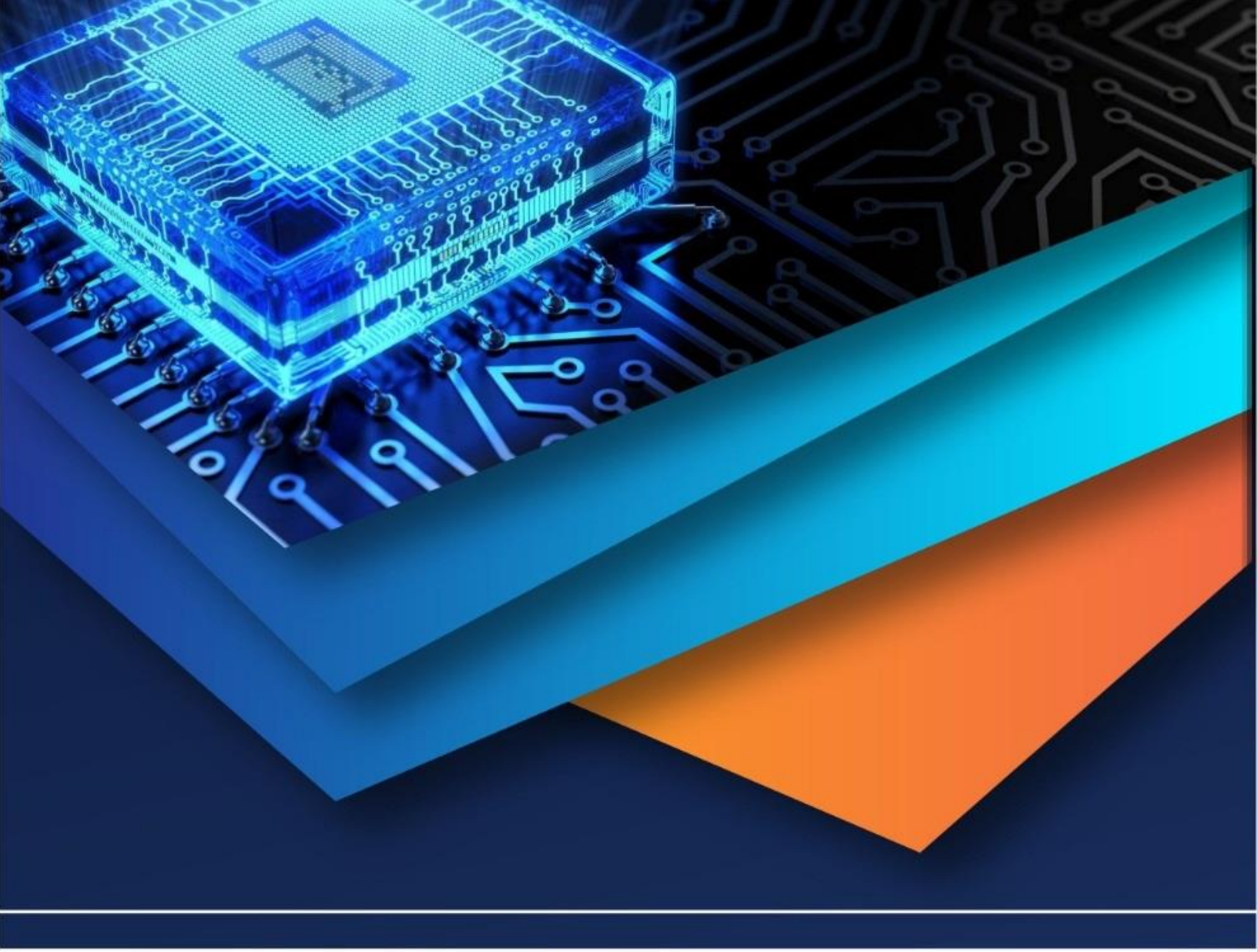
Experimental results confirm that the system can reliably recognize multiple emotional states from both speech and text inputs, with improved performance achieved through the hybrid SER–NLP approach. The implementation of a memory module using vector embeddings enables contextual awareness, allowing the system to generate more personalized and meaningful responses during user interactions. Furthermore, the system demonstrated stable real-time speech-to-text conversion and consistent performance under moderate background noise conditions.

Overall, the results validate that the proposed system is capable of providing intelligent emotional assistance and monitoring user well-being through adaptive and empathetic responses. Future work can focus on improving model generalization across diverse languages and accents, incorporating multimodal inputs such as facial expressions, and enhancing real-time deployment for scalable mental health support applications.

REFERENCES

[1] S. Latif, J. Qadir, A. Qayyum, M. Usama and S. Younis, “Speech Technology for Healthcare: Opportunities, Challenges, and State of the Art,” IEEE Reviews in Biomedical Engineering, vol. 14, pp. 342–356, 2021.
[2] G.-M. Li, N. Liu and J.-A. Zhang, “Speech Emotion Recognition Based on Modified Relief Feature Selection,” Sensors, vol. 22, no. 21, pp. 1–16, 2022.
[3] J. Singh, L. B. Saheer and O. Faust, “Speech Emotion Recognition Using Attention Model,” International Journal of Environmental Research and Public Health, vol. 20, no. 6, 2023.
[4] W. Zhu and X. Li, “Speech Emotion Recognition with Global-Aware Fusion on Multi-Scale Feature Representation,” IEEE Access, 2022.
[5] N. Elsayed, Z. ElSayed, N. Asadi Zanjani, M. Ozer, A. Abdel Gawad and M. Bayoumi, “Speech Emotion Recognition Using Deep Recurrent Systems for Mental Health Monitoring,” arXiv preprint arXiv:2208.12812, 2022.
[6] K. Huang, C. Wu, M. Su and C. Chou, “Mood Disorder Identification Using Deep Speech Features,” IEEE Signal Processing Conference, pp. 1–6, 2022.
[7] I. Gurowiec and N. Nissim, “Speech Emotion Recognition Systems and Their Security Aspects,” Artificial Intelligence Review, vol. 57, 2024.
[8] C. Barhoumi and Y. BenAyed, “Real-Time Speech Emotion Recognition Using Deep Learning and Data Augmentation,” Artificial Intelligence Review, vol. 58, 2025.
[9] J. H. Chowdhury, S. Ramanna and K. Kotecha, “Speech Emotion Recognition with Lightweight Deep Neural Ensemble Models,” Scientific Reports, vol. 15, 2025.
[10] E. Jordan, R. Terrisse, V. Lucarini, M. Alrahabi, M.-O. Krebs, J. Desclés and C. Lemey, “Speech Emotion Recognition in Mental Health: Systematic Review of Voice-Based Applications,” JMIR Mental Health, vol. 12, 2025.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)