INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

MindCare - AI Based Mental Wellness Voice Based Companion

Rohan Tirkhunde; Vaishnavi Patole; Sneha Yamgar; Nayana S. Thombare; Nikita Shinde

Department of Computer Engineering, K J College of Engineering & Management Research Pune, India

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150600035

Received: 08 June 2026; Accepted: 13 June 2026; Published: 03 July 2026

ABSTRACT

Mental health disorders such as stress, anxiety, and depression are increasing rapidly across all age groups. Many individuals do not receive timely support due to limited access to professional mental healthcare services. Artificial intelligence–based voice assistants provide a new opportunity to offer emotional support and continuous mental health monitoring. This paper surveys voice-based artificial intelligence mental health companions integrated with emotion tracking systems. The study reviews existing technologies, emotion detection techniques, system architectures, and practical applications. It also discusses major challenges including privacy, ethical responsibility, and emotional accuracy. The survey highlights that combining voice interaction with emotion tracking can significantly improve accessibility and user engagement in mental healthcare systems. This paper aims to provide a comprehensive overview for researchers and developers working in the field of intelligent mental health support systems.

Keywords: Artificial Intelligence, Mental Health, Voice Assistant, Emotion Tracking, Natural Language Processing, Human Computer Interaction

INTRODUCTION

Mental health is an essential component of overall human wellbeing. It changes how people think, feel, and act every day. Emotional stability plays a major role in personal growth and productivity. Over the past decade, emotional health issues have increased significantly. Stress, anxiety, loneliness, and sadness are commonly reported problems.

Students face academic pressure and career uncertainty.

Working professionals experience workload stress and job insecurity. Elderly individuals often suffer from loneliness and social isolation. These factors collectively contribute to emotional imbalance.

Despite the increasing need, access to mental healthcare remains limited. A lot of people are afraid to get help from a professional. Social stigma is one of the biggest problems. Fear of being judged prevents emotional expression.

Financial constraints also restrict access to therapy.

In some regions, trained mental health professionals are scarce. As a result, emotional problems often remain untreated. Untreated emotional stress can lead to serious mental disorders. Early emotional support is therefore highly important.

Advancements in technology have created new opportunities in healthcare. Artificial intelligence has shown promising results in many domains. Healthcare systems increasingly rely on intelligent technologies. AI enables automation, personalization, and continuous support. In mental health, AI can provide basic emotional assistance. Chatbots are one such AI based technology. They allow users to interact using natural language. Chatbots are available at any time without human intervention. They offer privacy and non-judgmental communication.

Most existing chatbots rely on text-based interaction.

Text-based communication has certain limitations.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

Typing may not be comfortable for all users. Text interaction limits emotional expression. Users may find it difficult to describe feelings in text. Voice-based communication is more natural for humans. Speaking allows better emotional expression than typing. Speech contains emotional cues such as tone and intensity. These cues help in understanding user emotions.

Voice-based chatbots can improve user engagement. They provide a more human-like interaction experience. When combined with emotion detection, effectiveness increases. Emotion detection allows systems to understand user feelings. Detected emotions can guide response generation. Supportive responses improve user comfort. Such systems can help users feel understood. They can reduce feelings of loneliness. They can provide emotional reassurance.

Emotion detection can be performed using text and speech analysis. Natural language processing helps analyze emotional content. Speech processing captures vocal emotional features. Simple emotion categories include happy, sad, and angry. Accurate emotion detection remains a challenging task. However, even basic detection can be useful. For prototype systems, lightweight models are sufficient. These models ensure real-time interaction. They also reduce computational complexity.

This study primarily investigates a voice-based emotional chatbot. The chatbot allows users to speak freely. It detects emotions from user input. Based on emotion, it provides supportive responses. The system is made to be user friendly. It is intended for emotional support, not medical diagnosis. It does not replace human therapists. Instead, it acts as a supportive companion. It provides understanding and care during emotional situations.

A prototype implementation is developed to evaluate the feasibility of the proposed system. The prototype validates the core functionality, including voice-based interaction and emotion detection. Experimental observations are analyzed to assess system performance and usability. System limitations and ethical considerations are also identified. The results indicate that the proposed system has strong potential to contribute to accessible emotional support solutions, with scope for further improvement in future work.

Problem Statement

Mental health challenges such as stress, anxiety, and emotional instability are increasing steadily across different age groups. Nevertheless, access to timely emotional support remains limited for many individuals due to social stigma, financial constraints, and the shortage of trained mental health professionals. As a result, many people hesitate to seek professional help, allowing emotional difficulties to remain unaddressed.

Most existing digital mental health solutions rely primarily on text-based interaction, which restricts emotional expression and reduces user engagement. Text-based systems are often unable to capture important vocal emotional cues such as tone, pitch, and intensity, which play a crucial role in understanding an individual’s emotional state. Additionally, many existing solutions lack real time emotion aware response mechanisms, limiting their effectiveness in providing meaningful emotional support.

LITERATURE SURVEY

The application of artificial intelligence (AI) and conversational agents in mental health support has gained significant attention in recent years. Initial efforts concentrated on text-based chatbots, which utilized natural language processing (NLP) and sentiment analysis to recognize user emotions and respond accordingly. While these systems allowed users to express basic emotions, they were limited in conveying emotional nuance, as they could not detect voice tones, speech patterns, or prosody.

Emotion detection from text commonly relies on lexicon-based approaches, machine learning classifiers, or neural network models to categorize emotional states such as happiness, sadness, anger, and anxiety. More recent techniques leverage transformer-based architectures to better understand contextual meaning. Nevertheless, text-only systems often struggle with sarcasm, mixed emotional states, and ambiguous expressions, which can compromise their accuracy in real-world applications.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

To overcome the limitations of text-based systems, researchers have investigated speech emotion recognition (SER). SER systems extract acoustic characteristics such as pitch, intensity, tone, speech rate, and spectral properties to classify emotions. Traditional machine learning algorithms, including support vector machines and neural networks, have demonstrated effectiveness in controlled settings. However, most of these systems focus exclusively on emotion classification and do not integrate conversational capabilities, which restricts their usefulness in practical emotional support scenarios.

In recent years, multimodal approaches that combine text and speech have been developed to improve emotion detection. By integrating linguistic and paralinguistic cues, deep learning models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can capture both temporal and contextual aspects of emotion. Despite the improved accuracy, these approaches demand large datasets, substantial computational power, and careful preprocessing**,** making them less feasible for lightweight or real-time applications.

Several commercial mental health chatbots, including Woebot, Wysa, and Replika, offer emotional support through scripted text responses. While these platforms provide **24/7** access and non-judgmental communication, they generally lack voice interaction, transparent emotion tracking, and rigorous validation, and they often fail to adequately address privacy and ethical concerns associated with sensitive user data.

Research indicates that combining voice-based interaction with real-time emotion recognition can enhance user engagement and improve emotional understanding. Voice cues allow users to express feelings more naturally, and adaptive responses based on detected emotions can increase perceived support and satisfaction**.**

Despite these developments, there remains a clear research gap**.** Current systems either focus on emotion detection without conversational support or provide chatbots without emotion awareness. There is a need for a user-friendly voice-based chatbot that can simultaneously detect emotions in real time, respond empathetically, and adhere to ethical guidelines. The proposed system aims to address this gap by integrating voice processing, emotion recognition, and intelligent conversational responses into a unified framework, offering an accessible and supportive tool for mental health assistance.

METHODOLOGY

System Design Approach

The proposed system is designed using a modular architecture that integrates speech processing, emotion analysis, natural language understanding, and response generation. Each module operates independently while maintaining seamless communication with other components. This design approach allows flexibility in improving individual modules without affecting the overall system performance.

Voice Input Acquisition

User interaction with the system begins through voice input captured using a microphone-enabled interface. Voice-based interaction is chosen to provide a natural and intuitive communication experience, especially during emotional distress when users may find typing inconvenient. The recorded speech signals are forwarded to the speech processing module for further analysis.

Speech-to-Text Conversion

The captured voice input is converted into textual form using an automatic speech recognition (ASR) system. Speech-to-text conversion enables the system to extract linguistic information required for natural language processing. The transcribed text serves as the primary input for sentiment analysis and intent detection.

Emotion Recognition

Emotion recognition is performed using both speech-based and text-based cues. Acoustic features such as pitch, tone, speech intensity, and rhythm are analyzed to identify emotional patterns from the voice signal.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

Simultaneously, sentiment analysis techniques are applied to the transcribed text to determine the emotional polarity and context. The fusion of vocal and textual emotion analysis improves the accuracy of emotional state detection.

Natural Language Processing

Natural language processing techniques are employed to understand the user's intent and conversational context. Tokenization, lemmatization, and semantic analysis are used to interpret user input meaningfully. This step enables the system to generate relevant and empathetic responses rather than generic replies.

Response Generation

Based on the detected emotional state and user intent, the system generates supportive responses using predefined therapeutic prompts and AI-based conversational models. The response generation module prioritizes empathy, reassurance, and emotional validation while avoiding medical diagnosis. The generated text responses aim to encourage emotional expression and self-reflection.

Text-to-Speech Synthesis

The generated textual response is converted back into speech using a text-to-speech (TTS) engine. Voice output ensures a continuous conversational flow and enhances emotional engagement. Natural-sounding voice synthesis is used to maintain a calm and supportive interaction environment.

Data Storage and Privacy

User interaction data, such as emotional trends and session logs, are securely stored for analysis and system improvement. Privacy and confidentiality are maintained by anonymizing user data and ensuring secure storage practices. The system is designed to comply with ethical considerations related to mental health data handling.

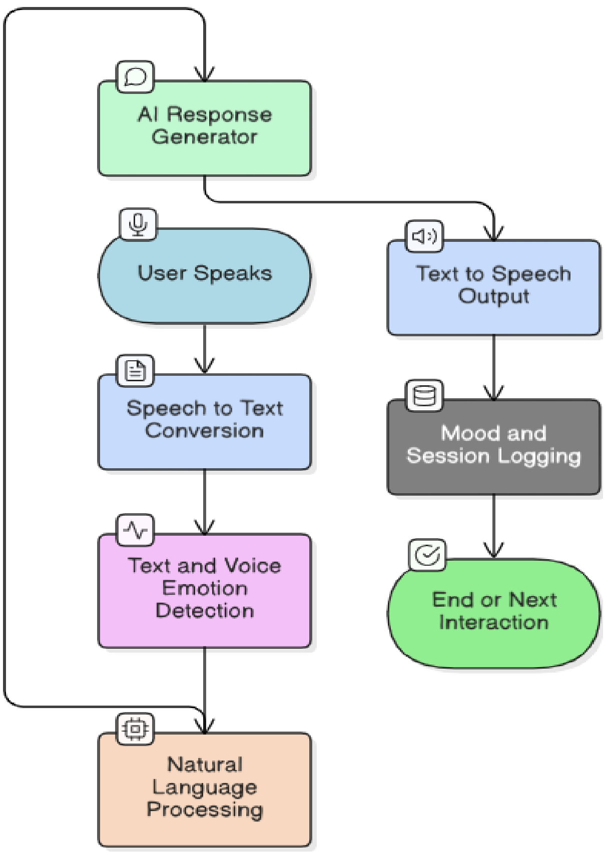

Fig. Flowchart of the proposed emotion-aware voice-based AI interaction system

INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)
ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

RESULTS

A preliminary experiment was conducted to evaluate the emotion detection component of the proposed system. A dataset containing 1,200 labeled emotional text samples across five categories (Happy, Sad, Stress, Anxiety, and Neutral) was used. The dataset was divided into 80% training data and 20% testing data. Standard preprocessing techniques such as tokenization and TF-IDF vectorization were applied.

A Support Vector Machine (SVM) classifier was used for emotion classification. The model achieved a training accuracy of 78.2% and a validation accuracy of 75.1%. The results indicate that the system is capable of identifying emotional patterns with moderate accuracy. The model performed better for clearly expressed emotions such as happiness and sadness, while some overlap was observed between stress and anxiety categories.

SVM MODEL PERFORMANCE

Parameter	Value
Classes	5 emotions
Train/Test Split	80% / 20%
Model	SVM
Training Accuracy	78.2%
Validation Accuracy	75.1%

Future Possibilities for Study

The proposed AI-powered voice-based mental health companion provides several opportunities for future research and enhancement. Future work can focus on improving emotion recognition accuracy by using larger and more diverse speech datasets, including multilingual and accent-inclusive data. The use of advanced deep learning models, such as transformer-based architectures, may further enhance the system's ability to detect subtle emotional variations in speech.

Another important area of research is personalization. By analyzing long-term user interaction patterns, the system could generate adaptive responses and provide personalized emotional feedback. This approach could improve user engagement and make the support system more effective for different individuals.

Additionally, future systems may integrate multimodal emotion analysis by combining voice data with facial expressions, text input, and physiological signals to improve emotion detection accuracy. The system could also include real-time crisis detection to identify high-risk emotional states and connect users with professional mental health services, supported by user studies and collaboration with mental health professionals.

CONCLUSION

This paper presented an AI-powered voice-based mental health companion designed to provide early emotional support through natural speech interaction. The proposed framework integrates speech-to-text conversion, emotion recognition, natural language processing, and text-to-speech synthesis to enable empathetic and context-aware conversations. By utilizing voice-based communication, the system improves emotional expression and user engagement compared to traditional text-based assistants. The modular architecture ensures flexibility and scalability for future improvements. Overall, the study demonstrates the potential of artificial intelligence and affective computing in developing accessible and supportive mental health assistance systems.

ACKNOWLEDGMENT

The authors gratefully acknowledge the guidance of **Prof. Nayana S. Thombare** and the Department of Computer Engineering, K. J. College of Engineering and Management Research, Pune, for their invaluable support.

REFERENCES

1. World Health Organization, “Mental health: Strengthening our response,” WHO, Geneva, Switzerland, 2022.
2. A. Miner, L. Milstein, S. Schueller, R. Hegde, C. Mangurian, and E. Linos, “Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health,” JAMA Internal Medicine, vol. 176, no. 5, pp. 619–625, 2016.
3. T. B. Fitzpatrick, A. Darcy, and M. Vierhile, “Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent,” JMIR Mental Health, vol. 4, no. 2, pp. 1–11, 2017.
4. R. W. Picard, Affective Computing. Cambridge, MA, USA: MIT Press, 1997.
5. S. G. Koolagudi and K. S. Rao, “Emotion recognition from speech: A review,” International Journal of Speech Technology, vol. 15, no. 2, pp. 99–117, 2012.
6. Z. Zhang, J. Han, E. Coutinho, and B. Schuller, “Deep learning for emotion recognition: A survey,” IEEE Transactions on Affective Computing, vol. 11, no. 1, pp. 3–23, Jan.–Mar. 2020.
7. D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed. Upper Saddle River, NJ, USA: Pearson, 2021.
8. A. V. Vasilateanu, D. Gifu, and M. N. Dascalu, “Natural language processing techniques for sentiment analysis in mental health applications,” in Proc. IEEE Int. Conf. on Artificial Intelligence, 2019, pp. 45–50.
9. B. Schuller, S. Steidl, and A. Batliner, “The INTERSPEECH 2009 emotion challenge,” in Proc. INTERSPEECH, Brighton, U.K., 2009, pp. 312–315.