iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78427

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 III Mar 2026- Available at www.ijraset.com

A Survey Paper on 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 growing incidence of psychological stress and safety risks in emergency environments necessitates intelligent systems capable of providing timely and context-aware assistance. This work investigates an emotion-aware conversational framework that leverages Speech Emotion Recognition (SER) to analyze vocal characteristics associated with stress, anxiety, and emotional distress. Acoustic features extracted from speech signals are utilized to classify emotional states into positive, neutral, and negative categories, enabling the generation of empathetic and adaptive conversational responses that support user well-being. The framework is further extended to aquatic safety scenarios, where behavioral patterns and panic-related vocal cues are analyzed to identify potential drowning risks. Upon detecting high-risk conditions, the system provides calming guidance and simultaneously triggers real-time alerts to caregivers or lifeguards for rapid intervention. The integration of emotion recognition with safety monitoring demonstrates the potential of AI-driven conversational systems to enhance mental health support and strengthen emergency response mechanisms in critical situations.
Rapid identifier: ML, Google Colab, Streamlit

I. INTRODUCTION

Mental health has become an important concern in modern society. Due to rapid lifestyle changes, academic pressure, work stress, and social expectations, many individuals experience mental health problems such as stress, anxiety, and depression[14]**.** These conditions affect a person's emotional well-being, physical health, and overall quality of life. Despite increasing awareness, many people still face difficulties in accessing mental health support because of the limited availability of professionals, high treatment costs, and the social stigma associated with mental health disorders[14], [17]**.**

Recent advances inArtificial Intelligence (AI) and Natural Language Processing (NLP) have enabled the development of intelligent systems capable of understanding human emotions and providing automated assistance [11], [13]**.**These technologies allow machines to analyze text, speech, and behavioral patterns to detect emotional states and mental health indicators [12], [18]**.**

Speech Emotion Recognition (SER) has emerged as an important technique for identifying emotions from speech signals by analyzing acoustic and prosodic features such as pitch, tone, and intensity [12], [15]**.** Deep learning models, including recurrent neural networks and attention-based architectures, have significantly improved the accuracy of emotion detection in speech and conversational systems [2], [18]**.**

Furthermore, affective computing enables systems to recognize, interpret, and respond to human emotions, allowing conversational agents to provide empathetic interactions and emotional support [3], [6], [13]**.** Such systems have been widely applied in healthcare, education, and intelligent assistants to enhance human-computer interaction [3], [6]**.**

In addition to emotional analysis, AI-based monitoring systems can also detect abnormal situations and safety risks in real-time environments. Previous research has demonstrated the use of AI techniques for monitoring behavioral patterns, detecting emergencies, and generating alerts using intelligent systems [10], [16], [19]**.**

II. BACKGROUND

Recent advancements in Artificial Intelligence (AI) and affective computing have significantly improved the ability of machines to understand and respond to human emotions. Affective computing focuses on developing intelligent systems that can recognize, interpret, and simulate human emotions through various modalities such as speech, facial expressions, and textual data [3], [13]**.** These systems are widely applied in healthcare, education, and conversational agents to enhance human–computer interaction and provide personalized support.

One of the key techniques used in emotion-aware systems is Speech Emotion Recognition (SER)**.** SER analyzes acoustic features such as pitch, tone, intensity, and prosody to identify emotional states from speech signals [12], [15]**.**

Early studies focused on feature extraction and statistical modeling of emotional speech, while recent approaches employ deep learning techniques such as recurrent neural networks and attention mechanisms to improve emotion classification accuracy [2], [18].

In addition to speech analysis, Natural Language Processing (NLP) has played a vital role in understanding emotional content from textual communication. Deep learning-based NLP models have demonstrated significant success in analyzing sentiment, detecting abusive language, and identifying psychological conditions from textual data [7], [11]. These techniques are also used to analyze social media posts and conversational interactions to detect emotional distress and behavioral patterns [10].

Furthermore, AI-driven emotional intelligence systems are increasingly being explored formental health monitoring and depression detection. Studies have shown that vocal characteristics, speech patterns, and behavioral signals can be used to detect early signs of depression and emotional distress [14], [17]. Such systems provide opportunities for early intervention and continuous monitoring of mental well-being.

In recent years, advanced technologies such as digital twins and smart healthcare ecosystems have also been introduced to improve health monitoring and well-being management. These frameworks integrate data from multiple sources to provide intelligent insights and real-time assistance for individuals in various environments [16], [19]. By combining speech emotion recognition, natural language processing, and intelligent monitoring, AI-based systems can support both emotional well-being and safety applications.

III. CLASSIFICATION OF MENTAL HEALTH AI SYSTEMS

Mental health AI systems can be categorized into several types based on the methods they use to collect, analyze, and interpret user data. These systems aim to detect emotional states, monitor mental well-being, and provide support through intelligent analysis of different forms of communication. With the development of Artificial Intelligence, machine learning, and data analytics, various approaches have been introduced to identify mental health conditions more effectively[11],[14]. The main classifications of mental health AI systems include text-based systems, speech-based systems, chatbot-based systems, and multimodal systems[13], [14].

Text-based mental health AI systems focus on analyzing written communication to detect emotional states and possible mental health conditions. These systems examine data from sources such as social media posts, online forums, emails, text messages, and chat conversations. Natural Language Processing (NLP) techniques are commonly used to analyze the structure, meaning, and sentiment of the text[11], [13]. By identifying specific words, phrases, and patterns in language, these systems can detect emotions such as sadness, stress, anger, or anxiety. Text-based analysis is widely used in applications that monitor social media platforms or provide automated mental health screening tools. It helps researchers and healthcare professionals identify potential warning signs and emotional trends in large groups of users[7], [9], [11], [13].

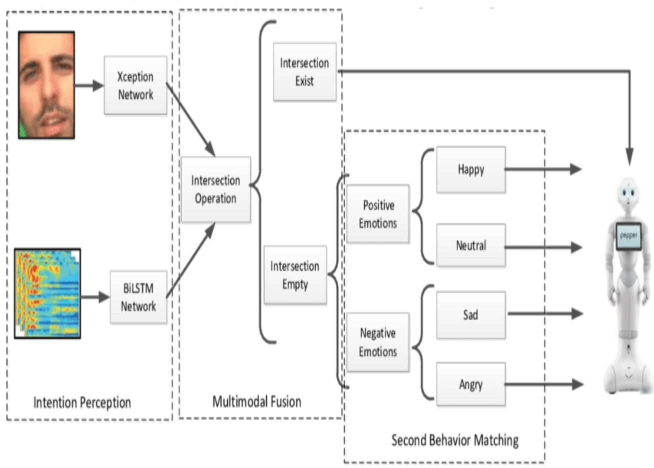

Fig: Multimodal emotion recognition framework integrating speech, text, facial expressions, and physiological signals

Speech-based mental health AI systems analyze vocal signals to detect emotional and psychological states. Human speech contains valuable emotional information beyond the words being spoken. Characteristics such as tone, pitch, speech rate, loudness, and rhythm can reflect a person's emotional condition[12], [15]. Speech Emotion Recognition (SER) techniques are used to extract and analyze acoustic features from voice recordings.

Machine learning models then classify these features to identify emotions such as happiness, sadness, anger, or stress[2], [18]. Speech-based systems are particularly useful in voice assistants, counseling support systems, and call center analysis, where emotional cues from speech can provide important insights into a user's mental state [2], [12], [18].

Chatbots represent another category of mental health AI systems. These conversational agents interact with users in real time using text or speech communication.

Chatbots are designed to simulate human conversation and provide basic emotional support, coping strategies, and mental health guidance[3], [6], [8]. They can ask questions, listen to user responses, and provide helpful suggestions or resources. In many cases, chatbots offer a safe and private environment where users feel comfortable sharing their thoughts and emotions. Although chatbots do not replace professional mental health services, they can serve as an accessible support tool for individuals who may not have immediate access to therapists or counsellors[8].

Multimodal mental health AI systems combine multiple sources of data to gain a more comprehensive understanding of a user's emotional and mental state. These systems may analyze text, speech, facial expressions, body language, and physiological signals such as heart rate or skin conductance. By integrating different types of data, multimodal systems can improve the accuracy of emotion detection and mental health assessment[13], [17]. For example, a system may analyze both the tone of a user's voice and their facial expressions to better understand their emotional condition. This approach provides a more holistic view of mental health and helps create more reliable and intelligent support systems[13], [17].

Overall, the classification of mental health AI systems highlights the diverse technological approaches used to detect and monitor emotional well-being. Each type of system offers unique advantages, and combining these approaches can lead to more effective and comprehensive mental health support solutions [14], [17].

IV. DATASETS AND FEATURE EXTRACTION

Datasets play a fundamental role in the development and training of Artificial Intelligence models used for mental health analysis and emotion recognition. In speech emotion recognition systems, datasets typically contain audio recordings of speech samples that are labeled with specific emotional categories such as happiness, sadness, anger, fear, or neutrality[12], [18]. These labeled datasets allow machine learning and deep learning models to learn patterns associated with different emotional states. The quality, size, and diversity of the dataset greatly influence the performance and accuracy of the trained models[12]. Well-structured datasets help systems generalize better and perform effectively in real-world environments.

Speech emotion datasets are usually collected from different sources such as controlled laboratory recordings, acted emotional speech, and real-life conversations[12], [4]. In many cases, professional actors are asked to speak certain sentences while expressing specific emotions so that the dataset contains clear emotional expressions. Some datasets also include spontaneous speech collected from natural interactions, which helps models learn more realistic emotional patterns [12]. Before the data is used for training, it goes through several preprocessing steps to remove noise, normalize the audio signals, and prepare the data for analysis [2], [18].

Feature extraction is an important step in speech emotion recognition because raw audio signals are complex and difficult for machine learning models to interpret directly [2] [18]. Feature extraction techniques are used to transform speech signals into meaningful numerical representations that capture emotional characteristics. These features help models identify patterns and differences between various emotional states.

One of the most commonly used features in speech processing is the Mel-Frequency Cepstral Coefficient (MFCC). MFCCs represent the short-term power spectrum of speech and are designed to reflect how humans perceive sound [12]. They are widely used in both speech recognition and emotion recognition tasks because they effectively capture important spectral characteristics of speech [18]. In addition to MFCCs, several other acoustic features are used to represent emotional information [15].

Pitch is an important feature that reflects the frequency of a person's voice. Changes in pitch often indicate different emotional states, such as excitement, happiness, or stress [15]. Energy or intensity represents the loudness of speech and can vary depending on emotional expression. Speech rate is another useful feature that measures how fast or slow a person speaks, which may change when someone is nervous, sad, or excited [12], [15]. Spectral features, including spectral centroid and spectral flux, help describe the distribution of energy across different frequency components of speech [15].

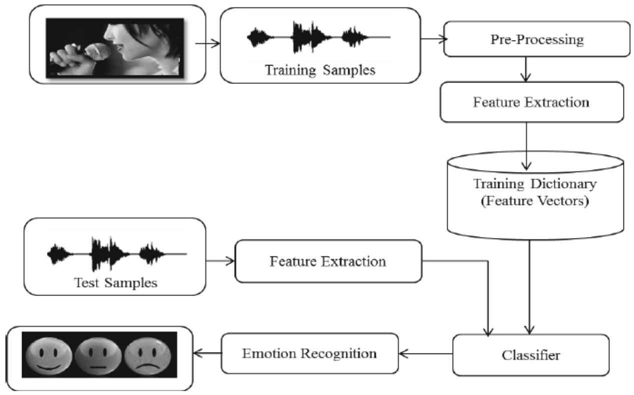

Fig: General workflow of a Speech Emotion Recognition system showing stages such as speech input, preprocessing, feature extraction, and emotion classification.

By extracting and combining these acoustic features, machine learning models can analyze speech signals more effectively and classify emotions with higher accuracy [2], [18]. These features provide important information about vocal characteristics and emotional cues present in speech [12], [18]. As a result, datasets and feature extraction techniques together form the foundation for building reliable and efficient speech emotion recognition systems used in mental health monitoring and human–computer interaction.

V. MACHINE LEARNING MODELS

Machine learning and deep learning models play an important role in analyzing emotional and mental health data. These models are designed to learn patterns from large datasets and make predictions based on the information they process [11], [12]. In the field of speech emotion recognition and mental health monitoring, machine learning models help identify emotional states by analyzing features extracted from speech, text, or other behavioral data [12], [18]. By learning from labeled datasets, these models can classify emotions such as happiness, sadness, anger, stress, or neutrality [12].

Traditional machine learning algorithms have been widely used in early emotion recognition systems. Some of the commonly used algorithms include Support Vector Machines (SVM), Decision Trees, and Random Forests [12], [13]. Support Vector Machines are effective classification models that work by finding an optimal boundary between different classes of data. They are widely used in speech emotion recognition because they perform well even with limited datasets [12]. Decision Trees are another type of algorithm that makes predictions by creating a tree-like structure of decisions based on different features. Random Forest is an extension of decision trees that combines multiple trees to improve prediction accuracy and reduce overfitting [13]. These traditional models are relatively simple and efficient for many classification tasks.

In recent years, deep learning techniques have gained significant attention due to their ability to learn complex patterns directly from raw data [11], [18]. Deep learning models use artificial neural networks with multiple layers to automatically extract important features from input data. Convolutional Neural Networks (CNNs) are commonly used for analyzing speech spectrograms and text patterns because they are effective at detecting local patterns in data [11]. Recurrent Neural Networks (RNNs) are designed to process sequential data such as speech and text, making them suitable for analyzing time-dependent information [2], [18].

Long Short-Term Memory (LSTM) networks are a special type of recurrent neural network that can learn long-term dependencies in sequential data [2]. LSTM models are particularly useful in speech emotion recognition because they can capture changes in emotional expression over time [2], [18]. By combining these advanced machine learning and deep learning techniques, emotion recognition systems can achieve higher accuracy and improved performance in detecting emotional states and supporting mental health applications [11], [18].

VI. COMPARATIVE ANALYSIS OF EXISTING WORKS

The evolution of AI mental health chatbots has evolved since the rule-based system to the intelligent conversational agent based on NLP, deep learning, and emotion recognition. They combine therapeutic measures such as CBT in order to support them in a humane manner, minimizing stress, and making them more accessible. The studies are aimed at the quality of conversations, privacy, clinical efficacy, and consistent emotional comprehension.

In [1], researchers have suggested an AI-based system to diagnose mental illnesses through machine learning and deep learning methods on social media text. The system took linguistic patterns and emotional clues and processed them with the help of the state-of-the-art NLP models, including BERT, RoBERTa, XLNet, LSTM, GRU, CNN, and Random Forest classifiers. Such models went through user-generated content and differentiated between early psychological distress and normal communication. Transformer based architectures, in particular, XLNet, had the highest accuracy of almost 97 percent because they had better contextual learning capacity. The paper has demonstrated that artificial intelligence offers scalable, non-invasive, and efficient approach to early mental health detection to implement timely intervention and enhance digital mental health support.

The article in [2], presented the AI-based dialogue system to assist in the early identification of mental problems in the form of a Digital Twin model. The system combines artificial intelligence, natural language processing, and clinical knowledge to determine the mental condition of users and provide them with feedback. An Rasa-based chatbot trained on a fine-tuned BERT model (E-DAIC dataset) identified the severity of depression and had a detection accuracy of 69% and high usability of 84.75. This study indicates that AI-based dialogue systems are capable of enhancing accessibility and can be used in scalable early mental health screening which is non-invasive.

In [3], important elements of Speech Emotion Recognition (SER) systems have been studied like popular databases like EMO-DB, RAVDESS and SAVEE. They used significant acoustic features, such as MFCC, LPC, and pitch, and the classifiers, such as SVM, Random Forest, CNN, and RNN. They found that deep learning models are more effective in capturing emotional patterns in speech, which results in a higher recognition and better performance of SER system.

In [4], AI and Natural Language Processing are examined in respect of offering emotional support to stressed, anxious, and depressed individuals. Conversational inputs reflected the feelings of users based on the linguistic cues and sentiment patterns. Emotion recognition techniques were used to process the data and classify the mental states and provide adaptive responses. The Artificial Intelligence-based solutions were based on supportive dialogue and therapeutic approaches to propose mindfulness exercises and assist patients in controlling emotions. The findings showed that this system made psychological support more accessible and more engaging to the user, which is a non-invasive and personalized solution to mental health support that is more effective in comparison to the conventional chatbot-based methods of mental health support.

In reference [5], which is based on a similar study, the authors examined the AI-based mental health prediction based on a Mixture-of-Experts model that utilized textual and behavioral social media data. Anxiety and depression were identified and responses were provided to these conditions using NLP and emotion recognition techniques. The AI system provided feedback in CBT form, mood monitoring, and stress-reduction. The results demonstrated the enhanced effectiveness of prediction and personalized care thus improving access to digital mental health care and compassion.

[6] A mental health chatbot named BlissBot was created in and intended to help people in a state of distress. The system also uses Artificial Intelligence and Natural Language Processing to interpret the queries of users, recognize emotional nuances, and provide them with targeted therapeutic advice. It includes Cognitive Behavioral Therapy interventions, behavioral activation plans, and ongoing mood checking as some ways to enable the users to grapple with anxiety and depression. Safe and normal conversation enhances the user to share their emotions without the need to be stigmatized. It was shown that the system had better emotional well-being, user acceptance, and better accessibility due to 24/7 availability, which provides evidence that empathetic and confidential chatbot support can adequately supplement traditional mental health care.

The article [7] involves the creation of an AI-based mental health chatbot, namely Wellness Buddy, that is capable of helping university students with anxiety and depression. Natural Language Processing, Deep Learning, and Transfer Learning are used in the system to recognize the emotional cues and give self-help instructions, psychoeducation, and supportive interaction. Developed in TensorFlow and released as an Android application, with AWS, it guarantees the privacy of the user, storing the information on the device. Methods of L2 regularization were used to overcome overfitting and enhance performance. The chatbot successfully worked on the emotional level of the students and was a cost-effective, accessible, and versatile digital mental health intervention that addressed low-resource settings.

In [8], a talkative agent that was emotion aware acoustically was created to identify and react to the emotions of the users in speech. The system studied the vocal characteristics of tone, pitch, and energy in order to categorize emotions like happy, sad, and angry. Comforting and supportive empathetic responses of dialogue were then created. The results of the testing on 75 respondents indicated a higher level of emotional satisfaction, which proves that the combination of the Speech Emotion Recognition with the empathetic response contributes to the increased human-likeness and effectiveness of chatbots in mental health care.

A stress, anxiety, and depression support chatbot named SAAC [9] was created as an AI-based mental health chatbot designed to offer customized emotional assistance to people. It can receive both text and voice interactions, with the Natural Language Processing and a deep learning model based on Keras to analyze emotions and produce an empathetic response. Other applications like Text-to-Speech, voice recognition and chat history boost constant and convenient interaction. The results indicated that no mental health care was more accessible, private and supportive, with user experience and emotional well-being improving with the use of SAAC.

[10] the growing popularity of mental health apps using AI was examined as a reaction to the absence of access to conventional psychological services. These emotional support programs are fast, convenient and anonymous and are chatbot-based. The research however generated some anxiety about the lack of human interaction, privacy issues, algorithmic discrimination, ambiguity about legal responsibility and lack of cultural sensitivity. The authors stressed that whereas digital mental health applications can be used to supplement care in resource-intensive environments, human practitioners need to be at the center of the provision of safe, ethical, and equitable mental health care.

In [11], a dialogue system was created using AI that helps to identify mental health disorders, e.g. depression, at early stages through conversation analysis of the user. The system uses Digital Twin to construct an online version of the mental state of a user to allow personalized feedback and monitoring. It applies a BERT-based Natural Language Processing model that is trained on the E-DAIC dataset to categorise depression by severity into several levels. The chatbot was developed based on the Rasa framework and tested in terms of usability and performance. The paper demonstrates the opportunities of conversational AI to promote access, lower stigma, and early intervention in mental health care, even though it has weak emotional intelligence and inadequate data sets.

In [12] Affective computing for emotional support is discussing the role of Artificial Intelligence in developing emotion-aware conversational agents that can recognize and respond to human emotions. It focuses on integrating technologies such as Natural Language Processing (NLP) and Speech Emotion Recognition (SER) to analyze both spoken language and vocal characteristics. The system extracts acoustic features from speech and applies machine learning models to classify emotional states. Based on the detected emotions, the conversational agent generates empathetic responses to improve user interaction. The study highlights the potential of AI-based systems in applications such as mental health support, virtual assistants, and human–computer interaction, while also discussing challenges related to accuracy, data quality, and privacy.

[12] A survey of textual emotion recognition and its challenges presents an overview of Artificial Intelligence–based approaches used for mental health support and emotion recognition. It explains how technologies such as Natural Language Processing (NLP) and Speech Emotion Recognition (SER) help analyze human emotions from text and speech. The study discusses different types of AI systems, datasets, feature extraction techniques, and machine learning models used for detecting emotional states. It also highlights challenges such as data limitations, privacy concerns, and variability in emotional expression. The paper concludes that integrating advanced AI techniques can improve emotion-aware systems and make mental health support more accessible and effective in real-world applications.

In [13] reviews the use of Artificial Intelligence technologies for emotion recognition and mental health support. It explains how AI techniques such as Natural Language Processing (NLP), Speech Emotion Recognition (SER), and machine learning models can analyze emotional information from text and speech data. The study discusses the importance of datasets, feature extraction methods, and different machine learning algorithms used for detecting emotional states. It also highlights various challenges including data diversity, noise in speech signals, and privacy concerns related to sensitive mental health information. The paper concludes that advanced AI methods and multimodal approaches can improve emotion detection and enhance intelligent mental health support systems.

[14] proposes a deep learning approach for speech emotion recognition (SER)usingRecurrent Neural Networks (RNNs) with a local attention mechanism**.** The model automatically learns emotional patterns from speech signals instead of relying only on handcrafted features. It captures both **short**-term acoustic features and long-term temporal dependencies in speech. The local attention layer helps the model focus on emotionally important segments of the audio signal, improving recognition performance. Experiments conducted on benchmark emotional speech datasets demonstrate that the proposed RNN with attention achieves better accuracy than traditional machine learning methods**,** highlighting the effectiveness of deep learning for emotion recognition in speech.

[15] proposed the relationship between vocal prosody and depression severity using speech recordings from clinical interviews. The study involved 57 participants diagnosed with Major Depressive Disorder, whose depression levels were assessed using the Hamilton Rating Scale for Depression (HRSD)**.** Bothperceptual evaluation by listeners and quantitative analysis of vocal features such as pitch, timing, and speech pauses were conducted.

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 III Mar 2026- Available at www.ijraset.com

The results showed that variations in vocal prosody strongly correlate with changes in depression severity over time. The analysis could explain about 60% of the variation in depression scores and classify severity levels with reasonable accuracy, demonstrating the potential of speech-based analysis for automated depression monitoring and screening systems.

[16]Speech emotion recognition (SER) is vital for human-computer interaction, but current small-scale datasets often cause model overfitting. To address this, Fan et al. introduced LSSED, a large-scale English dataset containing 147,025 sentences (over 206 hours) from 820 subjects. LSSED simulates real-world distributions, significantly outperforming existing databases in volume and diversity. The researchers also developed PyResNet, which uses pyramid convolutions to capture multi-scale acoustic information. Experiments show that LSSED-trained models exhibit superior generalization and effectively transfer to downstream tasks like depression detection.

[17] Fan et al. present LSSED, a massive English speech emotion dataset designed to overcome the overfitting issues prevalent in smaller, traditional databases like IEMOCAP and MELD. Containing 147,025 utterances from 820 subjects, LSSED offers over 206hours of natural, spontaneous speech, simulating real-world distributions across various ages and genders. The authors introduced PyResNet, which enhances the ResNet architecture with pyramid convolutions to capture multi-scale acoustic information. Benchmarking experiments reveal that while existing algorithms struggle with this large-scale data, PyResNet achieves superior performance with a weighted accuracy of 0.624. Furthermore, pre-training on LSSED proves highly effective for downstream tasks such as depression detection, outperforming models trained on speech recognition (ASR) by focusing on acoustic rather than semantic features. This work establishes a challenging new benchmark for advancing robust, generalized human-computer interaction.

[18] speech emotion dataset designed to overcome the overfitting issues prevalent in smaller, traditional databases like IEMOCAP and MELD. Containing 147,025 utterances from 820 subjects, LSSED offers over 206hours of natural, spontaneous speech, simulating real-world distributions across various ages and genders. The authors also proposed PyResNet, which enhances the ResNet architecture with pyramid convolutions to capture multi-scale acoustic information. Benchmarking reveals that while many existing algorithms struggle with this large-scale data, PyResNet achieves superior performance with a weighted accuracy of 0.624. Furthermore, pre-training on LSSED proves highly effective for downstream tasks such as depression detection, outperforming models trained on automatic speech recognition (ASR) by prioritizing acoustic features over semantic ones. This work establishes a challenging new benchmark for advancing robust human-computer interaction.

[19] Artificial Intelligence (AI) is revolutionizing mental healthcare by addressing the disparity between the high global demand for services and the limited availability of traditional in-person care. By integrating AI, the field is evolving through early detection tools, personalized treatment plans, and virtual therapists that provide 24/7, stigma-free support. AI technologies, such as Natural Language Processing (NLP) and sentiment analysis, enable the early identification of disorders by analyzing speech, text, and facial cues. Furthermore, predictive modeling and data-driven insights optimize treatment outcomes by tailoring interventions to individual needs, including genetic profiles and behavioral patterns. However, these advancements necessitate responsible implementation to address ethical challenges regarding privacy, bias mitigation, and maintaining the essential human element in therapeutic relationships.

[20] The ethical integration of sentiment and emotion analysis into depression-focused chatbots presents significant opportunities and challenges. While these technologies enable chatbots to offer empathetic, tailored support by identifying emotional states and suggesting appropriate interventions, they introduce complex ethical risks. Key concerns identified include the technical difficulty of accurately recognizing emotions in individuals with depression and the potential for biased training data leading to discriminatory outcomes. Furthermore, misinterpretation of user data can have severe consequences, particularly if risk-detection systems, such as those for suicide, fail. Privacy remains paramount, as sensitive mental health information is collected. Consequently, these systems require careful development and supervision by health professionals to ensure they enhance rather than diminish user autonomy, balancing innovation with patient safety and accountability

VII. CHALLENGES AND LIMITATIONS

Despite significant advancements in Artificial Intelligence–based mental health systems, several challenges and limitations still exist. One of the main challenges is the complexity of human emotions. Emotional expressions can vary greatly from person to person depending on cultural background, personality, and personal experiences [13], [14]. The same emotional state may be expressed differently by different individuals, making it difficult for AI systems to accurately detect and classify emotions [12], [13].

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 III Mar 2026- Available at www.ijraset.com

This variability can reduce the reliability of emotion recognition models, especially when the systems are applied in real-world environments [12].

Another important challenge is related to the quality of speech data used in speech emotion recognition systems. Speech signals can be affected by background noise, poor recording quality, or interruptions during conversations [2], [18]. These factors can interfere with the extraction of important acoustic features such as pitch, tone, and energy [12], [15]. As a result, the performance of machine learning models may decrease when they are used outside controlled environments [18]. Additionally, language differences, regional accents, and variations in speaking styles can also affect the accuracy of emotion detection systems [12]. Models trained on specific datasets may not perform well when applied to speakers from different linguistic or cultural backgrounds [12], [18].

Privacy and ethical concerns are also major limitations in the use of AI for mental health monitoring. Mental health data is highly sensitive and personal, and improper handling of such information can lead to serious privacy risks [1], [10]. Users may feel uncomfortable sharing personal thoughts, emotions, or speech recordings if they are unsure about how the data will be stored, processed, or protected [1]. Therefore, strong data protection policies and ethical guidelines are necessary to ensure user confidentiality and responsible use of AI technologies [10], [16].

Another limitation is that AI systems still lack the deep emotional understanding and empathy that human professionals provide. Although AI-based conversational agents can generate supportive responses, they cannot fully replace the experience, judgment, and emotional sensitivity of trained psychologists or therapists [3], [6]. As a result, AI systems should be considered as supportive tools rather than complete replacements for professional mental health care [14]. Continuous research and development are required to improve the accuracy, reliability, and ethical use of these systems [13], [17].

VIII. RESEARCH GAPS

Although significant progress has been made in the development of Artificial Intelligence–based systems for emotion recognition and mental health support, several research gaps still remain. One of the major gaps is the limited availability of diverse and large-scale datasets. Many existing studies rely on small or controlled datasets that contain speech or text samples from a limited group of participants [12], [18]. Such datasets may not represent the diversity of real-world populations in terms of language, culture, accent, age, and gender [12], [13]. As a result, AI models trained on these datasets may not perform accurately when applied to broader and more diverse user groups [12], [18].

Another important research gap is the reliance on single-modal data in many existing systems. Some systems focus only on analyzing text data using Natural Language Processing, while others concentrate solely on speech signals through speech emotion recognition techniques [11], [12]. However, human emotions are complex and are usually expressed through multiple channels, including speech, facial expressions, body language, and physiological signals [13], [17]. Systems that rely on only one type of data may fail to capture the complete emotional state of the user. Therefore, there is a need for research that integrates multiple data sources to improve the accuracy and reliability of emotion detection [13], [17].

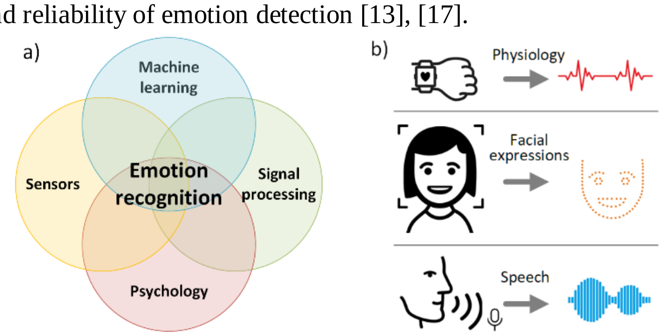

Fig: Taxonomy of emotion recognition approaches showing different research directions including speech-based, text-based, and multimodal emotion analysis techniques.

Additionally, many emotion recognition models are developed and tested in controlled environments with clean data and minimal noise [2], [18]. In real-world situations, speech signals may contain background noise, interruptions, and variations in communication style. These factors can significantly affect system performance [12], [18]. More research is needed to design robust models that can function effectively in real-life environments [18]. Addressing these research gaps will help improve the effectiveness, reliability, and practical application of AI-based mental health systems [14], [17].

IX. FUTURE RESEARCH DIRECTIONS

Future research in AI-based mental health systems is expected to focus on improving the accuracy, reliability, and usability of emotion recognition technologies [13], [14]. One promising direction is the development of advanced multimodal systems that combine different types of data to better understand human emotions. Instead of relying only on text or speech, future systems may integrate speech signals, textual communication, facial expressions, body language, and physiological signals such as heart rate or skin conductance [13], [17]. By combining these different sources of information, multimodal systems can provide a more comprehensive and accurate understanding of a user's emotional state [17].

Another important area for future research is the improvement of deep learning techniques used for emotion recognition. Advanced models such as transformer-based architectures, improved recurrent neural networks, and hybrid deep learning approaches can help detect complex emotional patterns more effectively [11], [18]. Researchers are also exploring the use of larger and more diverse datasets to train AI models. The availability of high-quality datasets that represent different languages, accents, cultures, and age groups will help improve the generalization capability of emotion recognition systems [12], [18].

Privacy and ethical considerations will also play a major role in future research. Since mental health data is highly sensitive, researchers must develop privacy-preserving techniques that protect user information while still allowing systems to analyze emotional data effectively [1], [10]. Techniques such as secure data storage, anonymization, and federated learning may help address these concerns. Additionally, integrating AI systems with professional mental health services could enhance their practical usefulness [14]. AI-based systems could assist mental health professionals by providing early detection, monitoring emotional changes, and offering supportive resources, ultimately improving access to mental health care [17].

X. CONCLUSION

Artificial Intelligence has become an important technology for supporting mental health monitoring and emotion analysis. This survey reviewed recent developments in AI-based systems designed to detect and analyze human emotions, particularly through speech emotion recognition, natural language processing, and conversational agents. By examining both acoustic and linguistic features present in human communication, these systems are capable of identifying emotional states such as stress, anxiety, happiness, and sadness. Such capabilities enable the development of intelligent systems that can assist individuals by providing emotional insights and supportive responses.

The study also discussed the key components involved in building emotion recognition systems, including datasets, feature extraction methods, and machine learning models. Features such as Mel-Frequency Cepstral Coefficients, pitch, energy, and spectral characteristics play a vital role in representing emotional information from speech signals. Traditional machine learning techniques along with modern deep learning models, including Convolutional Neural Networks and Long Short-Term Memory networks, have significantly improved the performance of emotion classification systems. These approaches allow models to learn complex emotional patterns from large datasets and provide more accurate predictions.

Despite these advancements, several challenges still exist in the practical deployment of AI-based mental health systems. Variability in emotional expression, limited dataset diversity, background noise in speech recordings, and privacy concerns related to sensitive mental health data can affect system reliability. Furthermore, AI systems cannot replace the empathy and professional expertise of trained mental health practitioners.

Future research should focus on developing multimodal emotion recognition systems and improving model robustness, enabling AI technologies to play a supportive role in enhancing mental health awareness and early emotional distress detection.

REFERENCES

[1] Y. Kasanneni, A. Duggal, R. Sathyaraj, and S. P. Raja, "Effective analysis of machine and deep learning methods for diagnosing mental health using social media conversations," IEEE Trans. Comput. Soc. Syst., vol. 12, no. 1, pp. 274–290, Feb. 2025, doi: 10.1109/TCSS.2024.3487168.

[2] A. Abilkaiyrkyzy, F. Laamarti, M. Hamdi, and A. El Saddik, "Dialogue system for early mental illness detection: Toward a digital twin solution," IEEE Access, vol. 12, pp. 1-12, 2024, doi: 10.1109/ACCESS.2023.3348783.

[3] G. H. Mohmad Dar and R. Delhibabu, "Speech databases, speech features, and classifiers in speech emotion recognition: A review," IEEE Access, vol. 12, pp. 151122–151146, 2024, doi: 10.1109/ACCESS.2024.3476960.

[4] K. Denecke, S. Vaaheesan, and A. Arulnathan, "A mental health chatbot for regulating emotions (SERMO) - Concept and usability test," IEEE Trans. Emerg. Top. Comput., vol. 9, no. 3, pp. 1172–1183, Jul.-Sep. 2021, doi: 10.1109/TETC.2020.2974478.

[5] S. Sawhney, H. Joshi, S. Gandhi, R. Shah, and R. Mahata, *"Mixture of experts for depression and anxiety disorder prediction from textual and non-textual social media data,"* in Proc. 2021 IEEE International Conference on Big Data (Big Data), 2021, pp. 1–10.

[6] S. Jaybhaye, S. Deogade, S. Sanap, T. Mali, and U. Shendge, "BlissBot: Mental Health Chatbot," in Proc. 2024 IEEE Region 10 Symp. (TENSYMP), 2024, pp. 1–5, doi: 10.1109/TENSYMP61132.2024.10752315.

[7] J. Deng and F. Ren, “A survey of textual emotion recognition and its challenges,” IEEE Transactions on Affective Computing, vol. 12, no. 2, pp. 286–299, Apr.–Jun. 2021, doi: 10.1109/TAFFC.2018.2874999.

[8] M. L. Joshi, “Depression detection using emotional artificial intelligence: A survey,” Materials Today: Proceedings, vol. 56, pp. 1916–1922, 2022, doi: 10.1016/j.matpr.2021.11.531.

[9] S. Mirsamadi, E. Barsoum, and C. Zhang, "Automatic speech emotion recognition using recurrent neural networks with local attention," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), 2017, pp. 2225–2229, doi: 10.1109/ICASSP.2017.7952504.

[10] W. Fan, X. Xu, X. Xing, W. Chen, and D. Huang, "LSSED: A large-scale dataset and benchmark for speech emotion recognition," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), 2021, pp. 6384–6388, doi: 10.1109/ICASSP39728.2021.9414542.

[11] Y. Yang, C. Fairbairn, and J. F. Cohn, “Detecting depression severity from vocal prosody,” IEEE Transactions on Affective Computing, vol. 12, no. 4, pp. 988–1001, Oct.–Dec. 2021, doi: 10.1109/TAFFC.2019.2926968.

[12] K. Denecke, S. Vaaheesan, and A. Arulnathan, "A mental health chatbot for regulating emotions (SERMO) - Concept and usability test," IEEE Trans. Emerg. Top. Comput., vol. 9, no. 3, pp. 1172–1183, Jul.-Sep. 2021, doi: 10.1109/TETC.2020.2974478.

[13] F. Laamarti, H. F. Badawi, Y. Ding, F. Arafsha, B. Hafidh, and A. E. Saddik, ‘‘An ISO/IEEE 11073 standardized digital twin framework for health and well-being in smart cities,’’ IEEE Access, vol. 8, pp. 105950–105961, 2020.

[14] M. Ogamba, J. Olukuru, J. Gitonga, J. Sevilla, and B. Muriithi, "Wellness Buddy: An AI Mental Health Chatbot for Kenyan University Students," in Proc. 2023 First Int. Conf. Adv. Artif. Intell. African Context (AAIAC), 2023, pp. 1–6, doi: 10.1109/AAIAC60008.2023.10465291.

[15] K. Hegde and H. Jayalath, "Emotions in the Loop: A Survey of Affective Computing for Emotional Support," arXiv preprint arXiv:2505.01542, 2025.

[16] A. Ghandeharioun, D. McDuff, M. Czerwinski, and K. Rowan, "EMMA: An emotion-aware wellbeing chatbot," arXiv preprint arXiv:1812.11423, 2019.

[17] K. Denecke and E. Gabarron, "The ethical aspects of integrating sentiment and emotion analysis in chatbots for depression intervention," Front. Psychiatry, vol. 15, no. 1462083, Nov. 2024, doi: 10.3389/fpsyt.2024.1462083.

[18] D. B. Olawade, O. Z. Wada, A. Odetayo, A. C. David-Olawade, F. Asaolu, and J. Eberhardt, "Enhancing mental health with Artificial Intelligence: Current trends and future prospects," J. Med. Surg. Public Health, vol. 3, p. 100099, 2024, doi: 10.1016/j.glmedi.2024.100099.

[19] J. Hu et al., "The acoustically emotion-aware conversational agent with speech emotion recognition and empathetic responses," IEEE Trans. Affect. Comput., vol. 14, no. 1, pp. 17–30, Jan.-Mar. 2023.

[20] R. Jayabhaduri, G. Ceralaathan, A. Vijayaraghavan, A. K. R, and S. S. S, "AI powered chatbot for mental health treatment," in Proc. 2024 First Int. Conf. Technol. Innov. Adv. Comput. (TIACOMP), Bengaluru, India, 2024, pp. 168–172, doi: 10.1109/TIACOMP64125.2024.00037.