International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(9s) 2026 pp. 1069-1080
https://doi.org/10.70917/ijcisim-2026-3519
Article

A COMPREHENSIVE REVIEW OF MULTIMODAL ARTIFICIAL INTELLIGENCE SYSTEMS FOR MENTAL HEALTH ANALYSIS USING IMAGE, SPEECH, AND IOT-BASED PHYSIOLOGICAL DATA

Mayur S. Katore[1] , S. M. Deshmukh[2] , Prashant V. Ingole[3]

Department of Electronics and Telecommunication, Prof. Ram Meghe Institute of Technology and Research, Amravati, Maharashtra, 444701, India
Email: mskatore1990@gmail.com , smdeshmukh001@yahoo.com ,pvingole@gmail.com

Abstract: With the rising rates of depression, anxiety, stress, and other mental ailments, mental health disorders are now a significant global health concern. While mental health monitoring is ongoing, usual approaches to counselling and diagnosis have been found to have some drawbacks, such as subjective diagnosis, high cost, scarcity of mental health workers and inaccessibility in remote areas. With the new developments in Artificial Intelligence (AI) technology, intelligent and automated healthcare production solutions have been developed which will enhance mental health assessment and monitoring. A multimodal AI-based system based on image processing, speech processing and physiological sensor monitoring for mental health analysis is examined in this review paper. In this proposed multimodal system, facial expression analysis, speech emotion recognition and biomedical sensors like heart rate, respiratory rate, blood pressure, and skin temperature are all combined to provide more accurate assessment of psychological conditions. Behavioral and emotional patterns that are linked to mental disorders are identified using machine learning and deep learning algorithms. Moreover, the integration of the Internet of Things (IoT) enabled wearable devices improves system functionality by enabling real-time health monitoring and facilitating remote counselling application. This paper underscores the need for using multimodal data fusion over the traditional single-modality systems to increase the accuracy of predictions and decrease the number of false diagnoses, respectively. The review also emphasizes the role of AI in remote healthcare systems for delivering scalable, cost-effective, and continuous mental health assessment capabilities in today's healthcare landscape.

Keywords: Mental Health Analysis, Artificial Intelligence, Machine Learning, Deep Learning, Speech Processing, Image Processing, Internet of Things, Remote Counselling

1. Introduction

The development of mental health problems, such as anxiety, depression, stress and post-traumatic stress disorder (PTSD) has been a significant health burden across the world. Psychological well-being has been seriously impacted by urbanization, work stress, social isolation, economic insecurity and an overreliance on technology. COVID-19 worsened mental health as there was an uptick in fear, loneliness, distress and uncertainty across the globe (Chen *et al.* 2024). The extended periods of lockdown, isolation, and joblessness brought about a significant psychological crisis in the wake of the pandemic, highlighting the importance of robust psychological health care systems. A timely diagnosis of psychological disorders is crucial, as early detection will decrease the severity of the disease, and help with treatment. In many cases, however, access to traditional mental health services is not easily available and is inept.

The typical approach to mental health diagnosis is the clinical interview, questionnaire, psychological testing of behaviour and actions carried out by psychiatrists and psychologists (Sangeetha *et al.* 2024). These approaches are generally commercially common, and require a lot of human knowledge and subjectivity. A lot of people don't like to express their emotions in the open due to the stigma and judgment within the society. Moreover, traditional counseling methods require precious time, money and, in rural settings, are inaccessible. A lack of mental healthcare professionals adds to the need for automated and intelligent diagnostic systems.

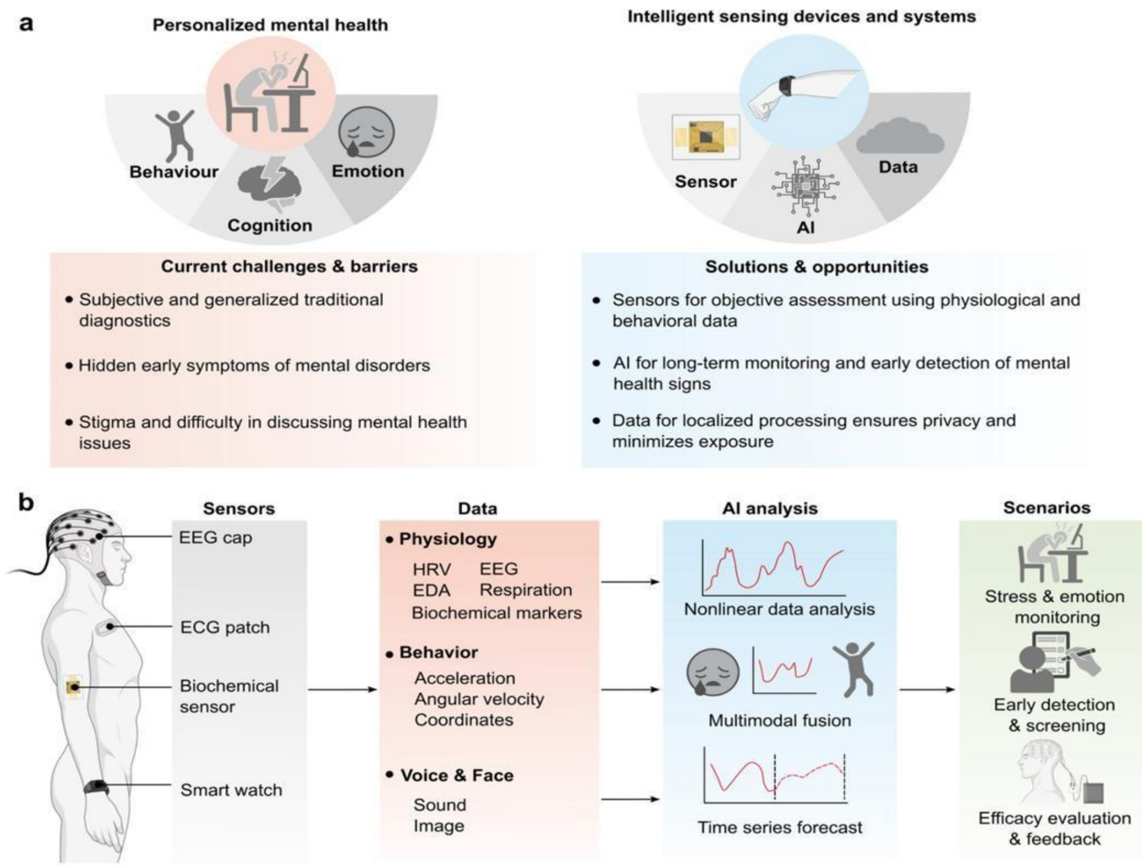

Figure 1: "Architecture of Multimodal AI-Based Mental Health Analysis System"

(Source: Sangeetha *et al.* 2024)

Healthcare Artificial Intelligence (AI) is a key technology affecting the industry that can help equip doctors with tools for automated diagnosis, predictive analysis and intelligent monitoring of patients. Facial expressions, speech and physiological data could be used to better detect mental health conditions using machine learning and deep learning techniques. But a single data source based system will tend to provide partial satisfactions as psychological events are very complicated. Multimodal AI systems overcome this limitation by integrating visual, vocal, and physiological data to improve diagnostic accuracy and reliability. Therefore, this review paper evaluates existing multimodal AI techniques for mental health analysis employing image processing, speech processing, machine learning, deep learning, and IoT-enabled biomedical sensing systems.

2. Mental Health and Need for AI Systems

2.1 Common Mental Health Disorders

Mental health disorders impact emotional, psychological, and social health, and can affect one's thoughts, feelings, and actions. Depression is a very common mental health condition; it's one of the more disabling among all mental health conditions in the world and is one that includes low spirits, lack of interest, and a lack of motivation. Anxiety disorders consist of excessive fear, nervousness and emotional instability which disrupt a person's daily functioning (Elfouly and Alouani, 2025). Another serious illness marked by variations in high and low mood swings is bipolar disorder.

Due to the increasing academic pressure, increasing competition in the workplace, insecurity about money and society, stress related disorders are becoming more prevalent by the day. Post-traumatic stress disorder (PTSD) is a form of psychological trauma caused by unusual events like disasters, violence or accidents, and can lead to emotional symptoms, issues with sleeping and flashbacks.

2.2 Psychological and Behavioral Indicators

The changes in behavior, emotions, and physiology can be indicators of a mental health problem. Facial expressions give a good indication of sadness, anger, fear and anxiety. Likewise emotion in speech patterns, pitch range, speech rhythm and voice tone can indicate the level of stress and depression (Paniagua-Gómez and Fernandez-Carmona, 2025). People with psychological diseases will often communicate abnormally and with emotional instability in speech interaction.

Occupational Physiology is also an important area of mental health analysis. Commonly associated emotional stress/Anxiety conditions include Heart Rate Variability, Blood Pressure fluctuations, Respiratory rate, and Skin temperature.

2.3 Importance of Early Detection

An early diagnosis of mental health disorders is critical to ensuring that the debilitating mental health problems can be avoided, and that treatment is more effective. Earlier diagnosis can minimize the chance of inducing harm and suicide to the child, and help health workers offer early counseling and therapy (Mohanraj *et al.* 2026). Early intervention also helps with

emotional recovery and promotes overall quality of life. In today's healthcare landscape, the need for remote accessibility has come to the forefront, particularly for patients in rural or underserved areas.

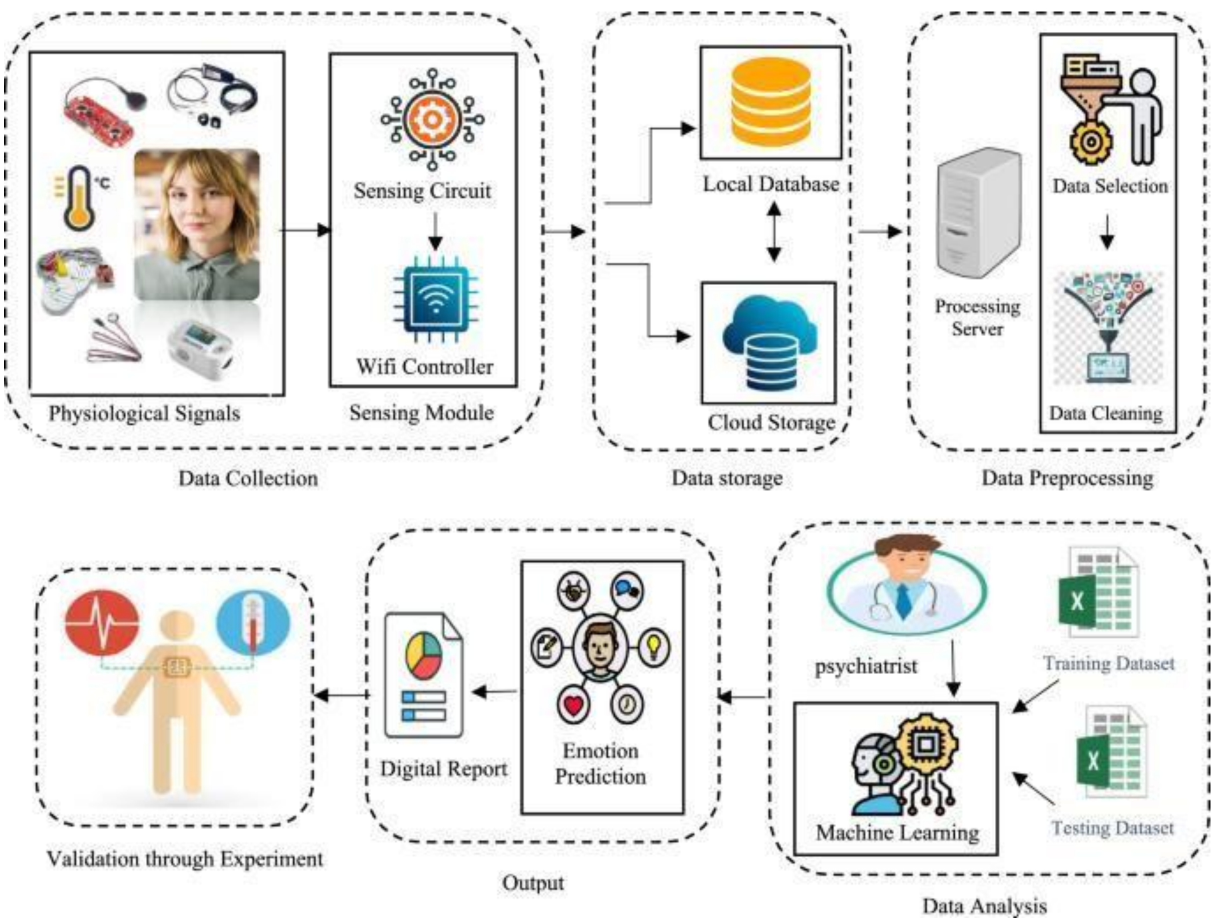

Figure 2: "Facial Expression and Emotion Recognition for Mental Health Monitoring"

(Source: Mohanraj *et al.* 2026)

3. Multimodal Data Acquisition for Mental Health Analysis

3.1 Overview of Multimodal Systems

Multimodal Artificial Intelligence Systems are intelligent systems that combine various forms of input data to facilitate analysis and decision-making. Multimodal systems are used in mental health applications, a combination of visual data, speech signals, and data collected from physiological sensors is used to provide full mental evaluation (Bouchouras and Kotis, 2025). Unlike more conventional, mono-modal systems, multimodal AI captures emotional, behavioral, and physiological traits at the same time, which helps to boost diagnostic reliability and lower predictive error rates.

Multiple data sources are integrated, so that various aspects of human behavior can be analyzed. Emotional reactions are communicated by facial expression, emotional tone in speech by the volume and/or pitch of voice, and physical responses to anxiety or depression through physiology.

3.2 Visual Data Acquisition

Visual Data acquisition identifies and records faces or this can also be in video through camera or video-enabled devices. One of the most important cues of emotional and psychological states is facial expression. The use of artificial intelligence (AI) systems can measure facial movements, behaviors, and expressions to identify stress, anxiety, depression, and emotional changes (Gan *et al.* 2022).

CNNs are widely used in modern image processing systems to process facial images and to recognize the nature of the emotion. CNN-based models can detect subtle changes in the face and include such facial changes related to psychological states. Facial Action Coding System (FACS) is also a system that is commonly applied to analyze patterns of facial muscle activity and emotions. Facial micro-expressions and eye movement tracking along with the rate of blinks give clues to other cognitive and emotional states.

3.3 Vocal/Audio Data Acquisition

The importance of speech and voice analysis in monitoring mental health is discussed. Vocal data collection involves recording voice data through a microphone or sound recorder during a conversation or counseling session. Speech patterns may be impacted by emotional stress, depression and anxiety, like tone, pitch, speaking rate and vocal energy.

These acoustic features can be used in Speech Emotion Recognition (SER) systems to detect emotions. Pitch variation, frequency components, speech level and tonal variations are all important audio characteristics. The linguistic features revolve around the actual words and what they mean in the speech, whereas the paralinguistic features relate to aspects of the speech such as pauses, hesitations or emotions (Javed *et al.* 2023). These features are given to machine learning algorithms which identify the emotional and psychological states.

3.4 Physiological Sensor Data Collection

Physiological data gathering is critical in the comprehension of emotional stress since emotional stress may bring about bodily responses. Parameters like heart rate, pulse rate, blood pressure, respiratory rate, skin temperature and oxygen saturation are monitored with biomedical sensors (Shivarkar *et al.* 2026). These physiological indicators are objective evidence of emotions and mental states.

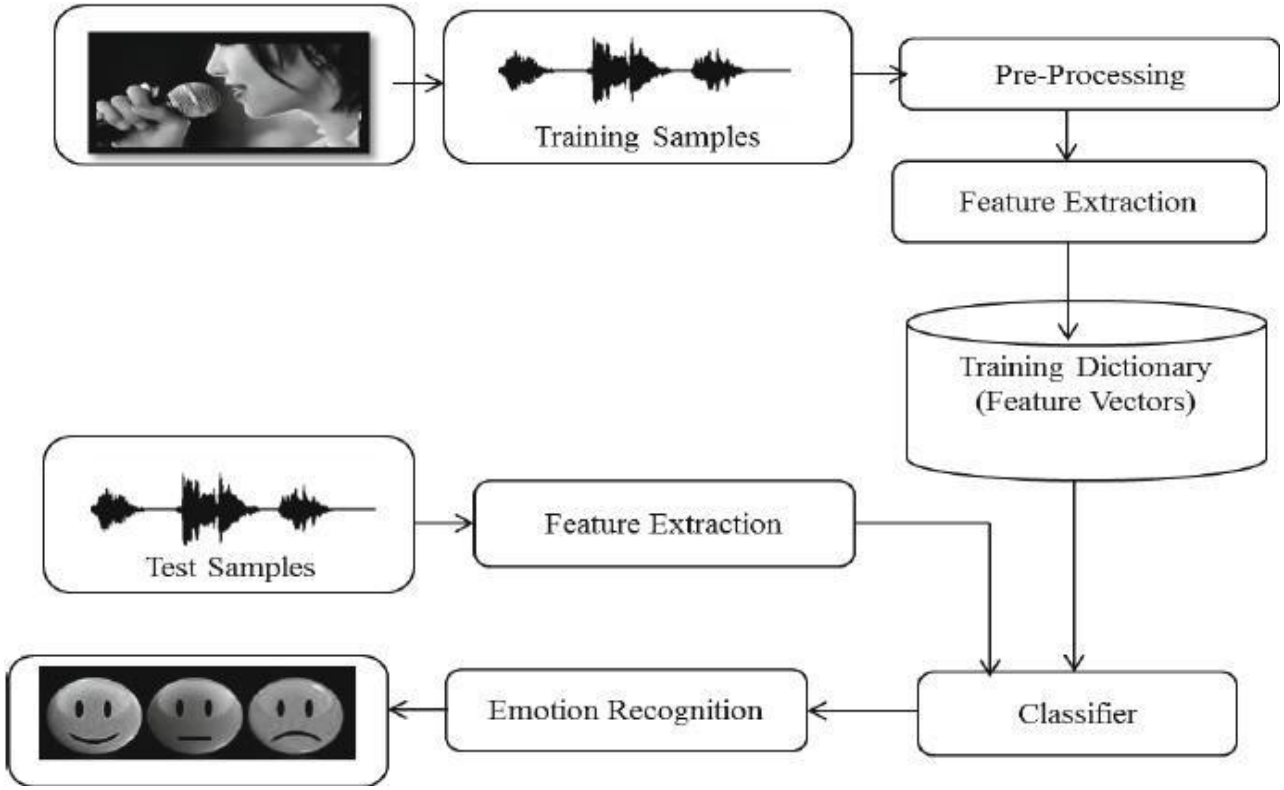

Figure 3: "Speech Emotion Recognition and Audio Signal Processing Framework"

(Source: Shivarkar *et al.* 2026)

Some examples include higher heart rates and patterns of irregular breathing, which often go hand-in-hand with anxiety and stress disorders. Flooding of emotions, for example, can also result in variations in blood pressure and in lowered oxygen saturation. Continuous monitoring of these parameters enables real-time assessment of mental health conditions and supports intelligent decision-making systems.

3.5 IoT-Based Biomedical Devices

Internet of Things (IoT) - based biomedical devices enable real-time monitoring and communication between healthcare systems and patients. Smart watches, fitness bands and

portable biomedical sensors gather physiological information constantly, and share it with a cloud-based healthcare platform. These technologies can help in remote monitoring and limit the necessity to visit the hospital frequently.

Healthcare systems with IoT capabilities are used to continuously track health, alerting systems or reports automatically and as soon as they happen. Implementing IoT and AI for analysis-based services increases accessibility in health care and enables patient services for patient counselling services in geographically remote areas.

4. Feature Extraction Techniques

4.1 Importance of Feature Extraction

Feature extraction is an important component in artificial intelligence systems since a raw data such as image, speech signal and biomedical sensor data to be fed to machine learning algorithms are multimodal in nature. The main goal of feature extraction is to transform complex raw data into representative features and significant patterns to accurately reflect human emotional and psychological states (Alshehri and Muhammad, 2020). An optimized feature extraction increases the learning ability of machine learning systems, decreases the computational complexity of the system, and increases classification accuracy.

Extracted features are the emotional expressions, speech stress expressions and physiological expressions related to mental disorders in mental health analysis. These traits enable machine learning systems to recognize subtle behavioral trends linked to anxiety, depression, and stress and other mental health problems. Feature extraction also removes redundant information and improves the intelligent healthcare systems performance.

4.2 Image Feature Extraction

The main objective of image feature extraction is extracting visual patterns from images and videos of the faces. Facial expressions and expressions of emotions play an important role in human communication and can be used to assess mental health disorders. Eye position, eyebrow movements, lip shape and the movement of facial muscles are often used to identify emotions.

Along with this, some minor changes in the face which accompany depression and stress are captured by the application of texture analysis technique. By analyzing facial expressions, AI

systems can be used to identify emotions like sadness, anger, fear, and happiness. Convolutional Neural Networks (CNNs) are widely applied in facial image analysis because of their ability to automatically learn complex visual features from large datasets.

Face shape detection is performed using Histogram of Oriented Gradients (HOG) and local edge orientation while edge characteristics and micro-expressions detection is done by Local Binary Pattern (LBP) methods. Together, these techniques enable facial emotion recognition and allow to appropriately analyze the behavior.

4.3 Audio Feature Extraction

In audio feature extraction, emotion and mental state information are extracted by examining the shape of voices in speech. Speech signals are preprocessed before feature extraction to enhance the quality of the speech signal such as noise removal, normalization and silence elimination (Bhaganagare *et al.* 2023). Spectral analysis is then applied to analyze the distribution of frequencies and acoustics speech signals.

Mel Frequency Cepstral Coefficients (MFCCs) are among the most commonly used features in an Audio based Emotion Recognition System as it provides a good representation for human auditory perception of sound. Other features like pitch, speech energy, and Zero Crossing Rate (ZCR) are also utilized to recognize variations of voice related to stress, anxiety and depression. These give machine learning models signals of emotional instability and abnormal communication behavior.

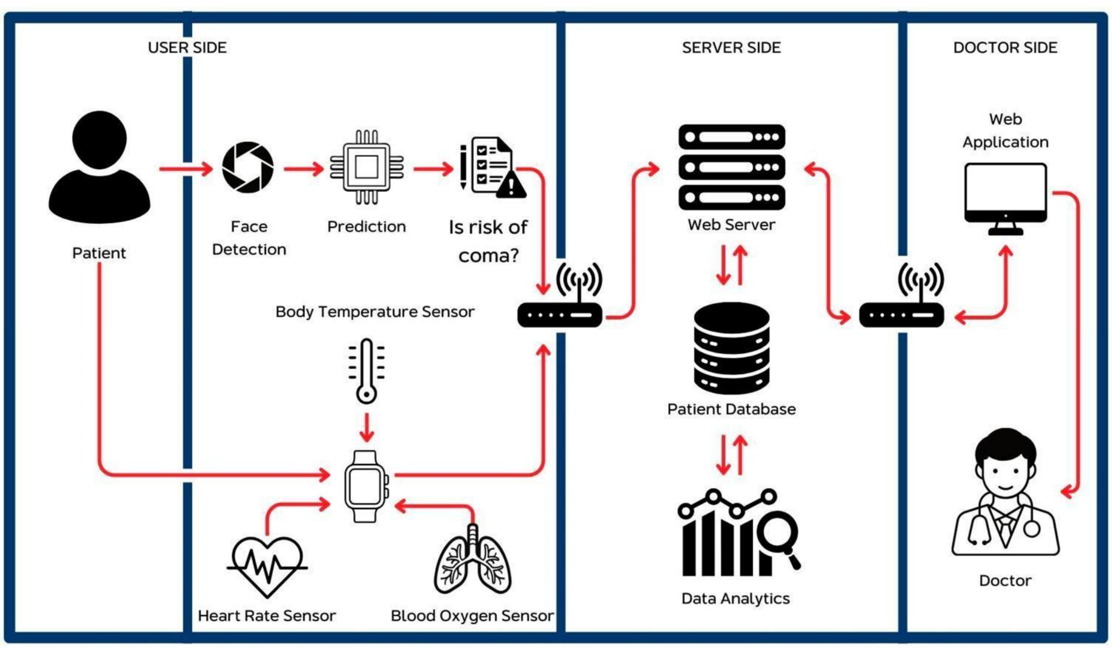

Figure 4: "IoT-Based Physiological Monitoring for Remote Mental Healthcare"

(Source: Bhaganagare *et al.* 2023)

4.4 Sensor Feature Extraction

Physiological sensor features extraction emphasizes on the study of the biomedical signal which is associated with mental stress and emotional conditions. Stress and anxiety can affect the cardiovascular response and one of the most significant indicators of stress and anxiety is heart rate variability (HRV). Additional stress parameters like blood pressure changes, breathing rates and temperature of the skin are also recorded to evaluate psychological conditions.

Wearable devices also show data on sleep and activity levels, which are important behavioral insights of emotional well-being. Irregular sleeping patterns and inactivity are strictly related to depression and stress-related disorders.

4.5 Feature Selection and Optimization

To optimize mental health AI system's efficiency, accuracy, feature selection and optimization techniques are crucial. Stacking relevant information is very helpful, yet frequently, the datasets we obtain are huge, and contain noisy or irrelevant data that can complicate computation and cause the models to be less effective. Dimensionality reduction is a widely

employed technique to reduce the number of features and components of the high-dimensional data and convert them to a lower-dimensional form, the common technique used for this is called Principal Component Analysis or PCA (Yadav *et al.* 2025).

The optimization method also reduces noise and speeds up computation during the training regime of the model. Choosing the most appropriate features can improve the prediction accuracy while keeping computational time short, assisting real-time intelligent healthcare systems.

5. Machine Learning and Deep Learning Models

Machine learning - driven behavioral prediction systems aid in early diagnosis and ongoing monitoring of mental health issues as it enables system to identify behavioral patterns and forecast psychological states using multimodal data including facial expressions, speech signals, and physiological signals. The subsequent section discusses various Machine Learning and Deep Learning models employed in this domain, along with the key insights derived from their application.

5.1 Traditional Machine Learning Algorithms

Several traditional machine learning algorithms have been extensively employed in mental health analysis systems, each offering distinct advantages for behavioral and emotional classification tasks. Support Vector Machines (SVM) excel at emotion classification by establishing optimal decision boundaries across complex data patterns, while Random Forest algorithms combine multiple decision trees to enhance prediction accuracy and mitigate overfitting. K-Nearest Neighbor (KNN) algorithms classify emotional states based on distance metrics between data samples, Decision Trees provide interpretable classification structures suited for behavioral analysis, and Naïve Bayes offers computationally efficient, probabilistic classification. Despite these advantages, traditional algorithms face notable limitations when applied to large-scale multimodal datasets, including high computational demands, complex feature engineering requirements, and reduced sensitivity to nuanced emotional expressions.

Following table summarizes these algorithms key advantages and limitations:

Algorithm	Advantages	Limitations
SVM	High classification accuracy	Computationally expensive
Random Forest	Reduces overfitting	Complex model structure
KNN	Simple implementation	High memory requirement
Decision Tree	Easy interpretation	Overfitting risk
Naïve Bayes	Fast computation	Assumes feature independence

5.2 Deep Learning Approaches

The ability to automatically extract hierarchical features from multimodal datasets through DL models has greatly advanced the mental health analysis field. The most common uses of CNNs are facial emotion recognition and behavioural analysis with an image. Such networks are able to detect facial patterns and emotional expressions of mental health disorders efficiently.

Predictive model training is useful for analyzing sequential data, like speech processing and behavior monitoring, such as recurrent neural networks (RNNs). The advanced network type of RNN as long short-term memory (LSTM) networks has proven to be effective in representing the long-term dependency in physiological signals and speech (Yu and Zhou, 2021). Lazy coughing Saint models are extensively utilized for the wide range of emotions detection, speech categorization, and also tension forecasting.

Hybrid CNN-LSTM networks can be used to integrate visual and sequential learning, leading to an enhanced multimodal behavioural analysis. These hybrid systems can result in greater accuracy in the diagnosis of the emotional and psychological state.

5.3 Multimodal Fusion Techniques

Multimodal fusion techniques combine information from visual, auditory and physiological modalities to enhance the accuracy of the prediction of mental health. Examples of early fusion are fusion of features from several data sources prior to the classification step and examples of late fusion are fusion of the prediction results generated independently by several models.

To enable the benefits of better performance and robustness, hybrid fusion approaches which are fused by mixing the early and the late fusion methods are performed. Multimodal fusion provides a tremendous increase of classification accuracy as it catches emotional, behavioral and physiological features (N. Saeedi *et al*., 2026).

5.4 Training and Optimization

To create precise and dependable mental health systems powered by Artificial Intelligence, it is necessary to undergo training and optimization. Labeled data that includes emotional and behavioral information are used to build machine learning, deep learning models. Using suitable dataset preparation and pre-processing, the efficiency of model learning is enhanced.

The Hyper-parameters are tuned using techniques to optimize the model to yield lesser prediction error. To assess the reliability of the model, cross validation methods are used, to avoid overfitting. Sustained and continuous optimization is essential to enhance the accuracy of classification and development of real-time intelligent counseling systems.

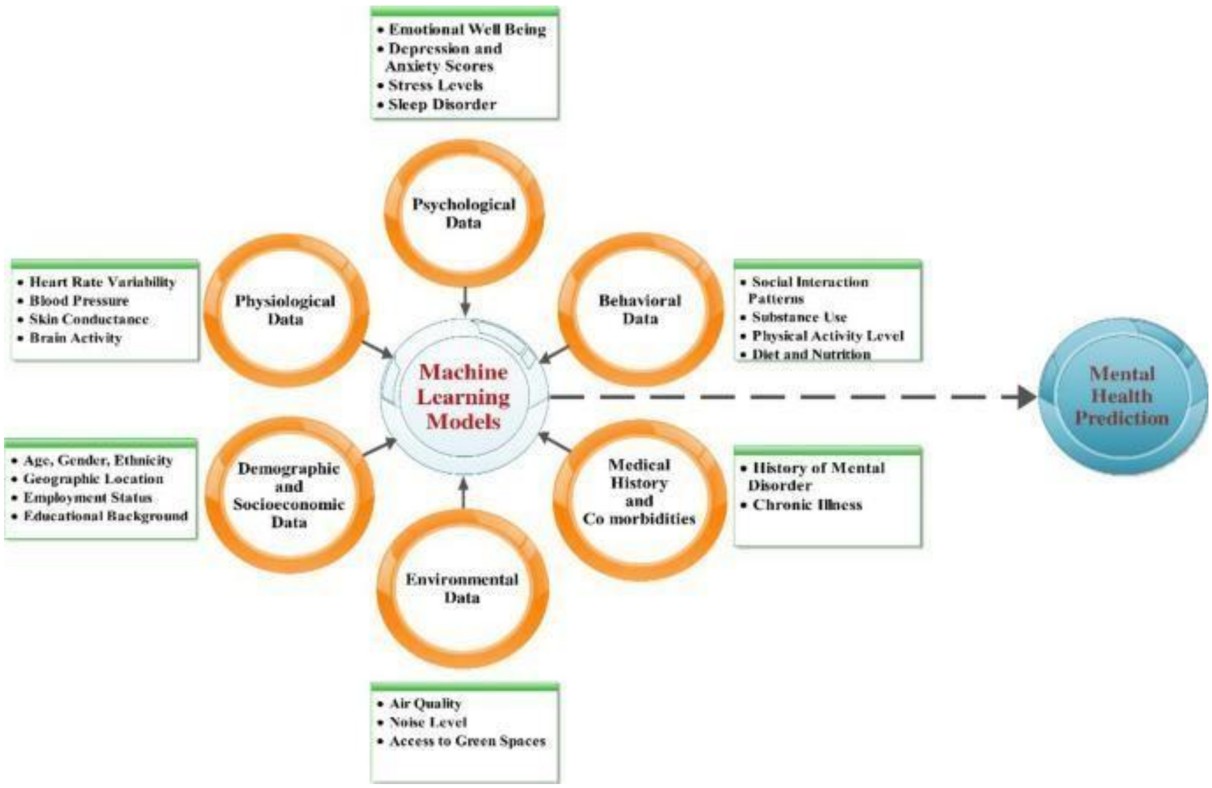

Figure 5: "Machine Learning and Deep Learning Workflow for Mental Disorder Prediction"

(Source: Yu and Zhou, 2021)

6. Classification and Evaluation Techniques

6.1 Mental Health Classification Models

The basis of mental health categorization models is that, by using multimodal data on behavioral and physiological characteristics, they categorize psychological conditions. These models are largely applied to determine depression, classify anxiety and predict stress. Machine learning and deep learning algorithms are used to classify a person's emotions from their physical and verbal cues and sensor data, for greater accuracy in detecting mental health-related diseases (Bout *et al.* 2025).

Depression detection systems are based on recognition of emotional sadness, energy in speech and physiological abnormalities. Anxiety classification models are able to measure the vocal and cardiac response to stress, and stress prediction systems track the emotional instability and behavioral irregularities.

6.2 Performance Evaluation Metrics

There are a number of evaluative tools to measure the effectiveness of mental health analysis systems. The overall correctness of the classification results is called accuracy, the correctness of the predicted positive cases is called precision. Recall is how well the system can successfully identify the actual positive cases.

Balance between precision and recall in F1 score.

$$F1=2\times Precision\times Recall \,/\, Precision+Recall$$

To measure classification performance at various thresholds, another receiver operating characteristic (ROC) analysis – Area Under Curve (AUC) is reported.

Metric	Purpose
Accuracy	Overall prediction correctness

Precision	Correct positive predictions
Recall	Detection capability
F1-score	Balanced evaluation metric
ROC-AUC	Classification discrimination ability

6.3 Comparative Analysis of Existing Systems

The current mental health analysis systems can be listed as single modal and multimodal. Single-modality systems rely on single communication modality (e.g., speech, facial expression) to generate their predictions, resulting in fairly limited accuracy and reliability with restricted scope of improvement. Multimodal system, in contrast, uses combination of voice, image, physiological and other signals (multi-modal) to deliver comprehensive behavioral assessment (Bin Heyat *et al.* 2025).

Moreover, AI analysis diagnosis systems exhibit some benefits over regular mental health evaluation techniques. Automated systems offer continuous monitoring, much quicker prediction, and lesser reliance on human intervention as compared with clinical assessment in a manual method.

6.4 Importance of Real-Time Evaluation

A continuous monitoring of mental health and dynamic counseling systems requires the implementation of real-time evaluation. AI can process the emotional and physical shifts continuously and alert to trigger in extreme psychological cases. Real-time monitoring means enhanced treatment responsiveness and the ability for voices of the treatment team to respond to immediate needs via remote counseling and intervention.

7. IoT and Remote Counseling Framework

7.1 Concept of Remote AI Counselling

AI hybrid counseling systems integrate artificial intelligence, remote counseling, and intelligent health solutions to deliver psychological assessment and assistance services remotely without physical clinic visits. These help patients connect with healthcare providers via digital platforms and allow AI models to continually assess the patients' emotional and physical states (Morales *et al.* 2026). Telemedicine counseling enhances the availability of mental health care services, particularly for those who reside in rural areas, where access to psychiatric care is scarce.

AI- or machine-assisted smart healthcare systems, for example, can automatically track people's emotional behavior, speech features and vital signs in real time. The architecture proposed is a multi-modal combination of image processing, speech analysis and biomedical sensor monitoring which can support mental health diagnosis and counseling with intelligence. This approach enhances the efficiency of health care, and decreases reliance on delivering face-to-face services.

7.2 IoT Integration in Mental Healthcare

The Internet of Things (IoT) is a key component in a remote mental healthcare system, which helps communicate between biomedical sensors, wearable devices and healthcare platforms. Sensor networks are continuously used to capture physiological information (heart rate, blood pressure, respiratory rate, skin temperature, etc.) to evaluate the patients' mental health.

The collected data can then be sent in real-time to cloud-based healthcare technology platforms where machine learning algorithms can analyze the data. Cloud storage can enable access to data from remote locations, constant monitoring and secure management of data for healthcare staff (Xu *et al.* 2024). Implementing IoT in Mental Healthcare help in the effective and real-time monitoring aspect of mental health.

7.3 Smartphone-Based Mental Health Systems

Smartphone Based Mental Health System makes the counseling users easy to use and prevent with a portable system that is available on their phones and with their wearable devices. The camera and speaker enabled smartphones with internet connectivity can take photos of patients' faces, record their speech and gather physiological information for later analysis by AI.

The feature of wearability with smart-watches and fitness bands also aids in monitoring activities in real-time. Portable counseling systems enhance patient ease of use, and allow consistent assessment of emotional condition which occurs outside the hospital setting.

8. Challenges, Limitations and Research Gap

8.1 Technical Challenges

Although there has been substantial progress in the field of AI-based mental health care systems, there are some technical obstacles to overcome. An important problem stems from the heterogeneity of the data, with image data often being structured in a different way from speech data and/or from physiological data, and the need for different processing for each type of data (Bhaganagare *et al.* 2023). The speech recordings and image data may contain noise which may degrade the quality with which features can be extracted, and may impair the accuracy of prediction. Furthermore, biomedical sensors can also have a limit in their ability to measure and give off inconsistent or inaccurate readings under different environmental conditions or due to the limitations of the device.

8.2 Ethical and Privacy Issues

In mental healthcare systems, ethical matters are of great concern since patient information is highly sensitive. Keeping patient data and ensuring that they are stored safely are challenging issues. To ensure that information cannot be accessed by unauthorized players, data security mechanisms and types of encryptions need to be employed. For the selection of patients who are collecting or analyzing data, they need to be informed about data collection and analysis processes, which also requires consent management.

8.3 Limitations of Existing Systems

Current mental health monitoring systems are mostly based on either facial analysis or speech recognition, and deliver poor diagnostic reliability. Numerous studies also do not have an impact test, and are based on small-scale experimental datasets. It creates obstacles in delivering reliable and scalable AI-driven counseling services.

8.4 Research Gap

There is growing need for integration of multi-modal systems which combine data from physiological sensors, image and speech within a single framework for holistic analysis. However, most of the existing studies do not combine AI and IoT, let alone for remote counseling applications together (Islam, 2025). Also, little research exists on the real-time adaptive system that can constantly monitor and analyze real-time mental health problems adaptively.

9. Future Scope and Recommendations

With the evolution of mental healthcare systems, advanced technologies are projected to be incorporated for more precise and personalized diagnostics using artificial intelligence. Explainable AI (XAI): Enhances the transparency of the systems by making them understandable to the healthcare experts. Designed to improve data privacy, federated learning can be used to train models on decentralized data networks without the disclosure of sensitive patient information.

Continuous emotional monitoring will be further enhanced by real-time emotion wearable analytics and individual emotional health prediction systems. Seamless connection with smart hospitals and cloud AI can enhance the management and remote assessment at a large scale. Use of edge computing technologies could help to lower the processing lag that occurs when analyzing healthcare data locally.

Multimodal deep learning-based emotion recognition systems are also believed to enhance the accuracy in the behavioral analysis. Building intelligence, scalability and privacy protection frameworks for mental health care in future should be the priority.

10. Conclusion

The growing prevalence of mental health disorders is a significant worldwide health issue, requiring the development of smart and user-friendly systems for diagnosis. In this review paper, the role of Multi Modal artificial intelligence techniques in Mental health analysis employing Image processing, Speech Analysis, Machine learning, Deep learning, and Biomedical Sensing systems which are enabled by IoT have been analyzed. The survey was conducted to explore different feature extraction methods, machine learning methods, multiform fusion methods, and remote time counseling systems for psychological assessment.

Combining facial expression analysis, speech emotion recognition, and monitoring from physiological sensors are more accurate and reliable in terms of diagnosing mental health status than the traditional use of a single modality. This is realized by being able to predict behaviour, classify emotion and monitor continuously, all of which can be achieved with a set of machine learning and deep learning techniques. In addition, smart wearables and mobile phone technologies facilitate engagement in real-time of remote health care systems, and enhance contact with counselling services.

While privacy issues, data from multiple sources, and a lack of real-world validation are potential hurdles, the multimodal AI systems have a remarkable potential to revolutionize modern mental healthcare. The possibilities of multimodal AI enabled mental healthcare systems are endless in the field of remote psychological diagnosis, offering intelligent, continuous, and readily available psychological health assessment solutions.

Reference

1. Alshehri, F. and Muhammad, G., 2020. A comprehensive survey of the Internet of Things (IoT) and AI-based smart healthcare. *IEEE access*, *9*, pp.3660-3678.
2. Bhaganagare, S., Chavan, S., Gavali, S. and Godase, V., Mood Mate: Advancing Real-Time Mental Health Monitoring with Multimodal AI.
3. Bhaganagare, S., Chavan, S., Gavali, S. and Godase, V., Mood Mate: A Review of Multimodal AI in Real-Time Mental Health Monitoring.
4. N. Saeedi et al., "Depression Detection Using Cross-Attention Multimodal Fusion on Audio-Visual Data," in Artificial Intelligence for Neuroscience, Mental Health, and Neurodegenerative Disorders, IWINAC 2026, Lecture Notes in Computer Science, vol. 16574, Springer, Cham, 2026.
5. Bin Heyat, M.B., Adhikari, D., Akhtar, F., Parveen, S., Zeeshan, H.M., Ullah, H., Chen, Y.H., Wang, L. and Sawan, M., 2025. Intelligent internet of medical things for depression: current advancements, challenges, and trends. *International Journal of Intelligent Systems*, *2025*(1), p.6801530.
6. Bouchouras, G. and Kotis, K., 2025. Integrating artificial intelligence, internet of things, and sensor-based technologies: a systematic review of methodologies in autism Spectrum disorder detection. *Algorithms*, *18*(1), p.34.
7. Bout, N., Moukhliss, G., Belhadaoui, H., Afifi, N. and Abik, M., 2025. Integrating emotional AI, IoT, and robotics for patient-centered healthcare: challenges and future directions. *Discover Internet of Things*, *5*(1), p.75.
8. Chen, X., Xie, H., Tao, X., Wang, F.L., Leng, M. and Lei, B., 2024. Artificial intelligence and multimodal data fusion for smart healthcare: topic modeling and bibliometrics. *Artificial Intelligence Review*, *57*(4).
9. Elfouly, T. and Alouani, A., 2025. A comprehensive survey on wearable computing for mental and physical health monitoring. *Electronics*, *14*(17), p.3443.
10. Gan, W., Dao, M.S., Zettsu, K. and Sun, Y., 2022, June. IoT-based multimodal analysis for smart education: Current status, challenges and opportunities. In *Proceedings of the 3rd ACM Workshop on Intelligent Cross-Data Analysis and Retrieval* (pp. 32-40).
11. Islam, M.M., 2025. AI-driven multi-purpose platform for smart healthcare systems.
12. Javed, A.R., Saadia, A., Mughal, H., Gadekallu, T.R., Rizwan, M., Maddikunta, P.K.R., Mahmud, M., Liyanage, M. and Hussain, A., 2023. Artificial intelligence for cognitive health assessment: state-of-the-art, open challenges and future directions. *Cognitive Computation*, *15*(6), pp.1767-1812.
13. Mohanraj, P., Mathan, M., Yuvaraj, S. and Prasanth, R., 2026. An Al-Driven Multimodal Voice, Text, and Emotion Analysis Framework for Early Mental Health Disorder Detection. *International Journal of Engineering & Extended Technologies Research (IJEETR)*, *8*(2), pp.4457-4462.
14. Morales, A.S., Reis, T.D.L., Panisson, A.R., Ourique, F. and Sene Jr, I.G., 2026. Affective Intelligent Systems in Healthcare: A Systematic Review. *Technologies*, *14*(3), p.188.
15. Paniagua-Gómez, M. and Fernandez-Carmona, M., 2025. Trends and challenges in real-time stress detection and modulation: The role of the iot and artificial intelligence. *Electronics*, *14*(13), p.2581.
16. Sangeetha, S.K.B., Immanuel, R.R., Mathivanan, S.K., Cho, J. and Easwaramoorthy, S.V., 2024. An empirical analysis of multimodal affective computing approaches for advancing emotional intelligence in artificial intelligence for healthcare. *IEEE Access*, *12*, pp.114416-114434.
17. Shivarkar, S.A., Nirgude, V.N., Bhutad, S., Lokare, R., Bachhav, G.V. and Malvankar, R., 2026. Intelligent IoT-Based Mental Health Prediction Framework for Smart Cities Using Deep Learning: Integrating Facial Emotions and Questionnaire. *Cognitive Computation*, *18*(1), p.45.
18. Xu, X., Fu, C., Camacho, D., Park, J.H. and Chen, J., 2024. Internet of things for emotion care: Advances, applications, and challenges. *Cognitive Computation*, *16*(6), pp.2812-2832.
19. Yadav, G., Bokhari, M.U., Alzahrani, S.I., Alam, S. and Shuaib, M., 2025. Emotion-aware ensemble learning (EAEL): Revolutionizing mental health diagnosis of corporate professionals via intelligent integration of multi-modal data sources and ensemble techniques. *IEEE Access*, *13*, pp.11494-11516.
20. Yu, H. and Zhou, Z., 2021. Optimization of IoT-based artificial intelligence assisted telemedicine health analysis system. *IEEE access*, *9*, pp.85034-85048.