GKU Journal of Multidisciplinary Research (GKUJMR)
ISSN: 3049 – 298X, Volume – II, Issue – I, April – 2026

AI-Based Early Detection of Neurological and Mental Disorders using Multimodal Deep Learning

[1] Mr. Ravi Kishan Singh, [2] Mr. Gurpreet Singh, [3] Dr. Narinder Kaur, [4] Dr. Rajinder Kumar

[1,2,3] Department of Computer Science & Engineering, Guru Kashi University, Talwandi Sabo, Bathinda.

[4] Associate Professor, Department of Computer Application, Guru Kashi University, Talwandi Sabo, Bathinda.

[4] ORCID ID: 0009-0001-4129-0388, E-mail: [4] drrajinder1983@gmail.com

Accepted: 10.04.2026 Published: 30.04.2026 Page No. 121 – 132 DOI: 10.5281/zenodo.20020928

Abstract

Early diagnosis of neurological and mental health disorders is a pressing issue in the current healthcare system, especially in underdeveloped areas with limited access to advanced diagnostic tools and healthcare professionals. Diseases like Parkinson's, depression and stroke often present with early signs that are difficult to spot through conventional medicine. In this research, we propose the use of a new multimodal deep learning model with which non-invasive as well as economically affordable prediction of disease can be done timely with the use of speech along with facial micro-expressions. Our work exploits Convolutional Neural Networks (CNN) for spatial feature extraction from facial images and Recurrent Neural Networks (RNN/LSTM) for temporal speech feature modelling. A fusion strategy is employed to combine multi-modal features for enhanced model accuracy and reliability. The goal of this strategy is to identify people at different risk levels to catch it early and keep an eye on their health for a long time to come. In addition, the system's ability to lessen the need for expensive diagnosis tests proves it can be used in the real world, especially in rural areas. We also discuss the issues of privacy, dataset collection, and generalisation. The method has the potential to revolutionize preventive health by facilitating scalable, affordable, and smart health diagnostic and monitoring systems.

Keywords:

Artificial Intelligence (AI); Multimodal Deep Learning; Neurological Disorders; Mental Health Disorders; Early Disease Detection; Machine Learning in Healthcare; Medical Data Analysis; Healthcare Informatics; Predictive Modeling; Clinical Decision Support Systems.

1. Introduction

Mental and neurological disorders such as depression, Parkinson's disease, schizophrenia, and autism spectrum disorders are a serious burden on global health, impacting millions of people from all walks of life. Disorders such as depression, anxiety, and addictions impact cognitive, emotional and behaviours. It can be rewritten as it causes deterioration in the quality of life and huge social and economic costs. The accurate and timely diagnosis of a disease can greatly facilitate an effective treatment. In many cases, common traditional diagnostic methods mostly rely on the clinical judgement, self-report from the patient, and observation which can delay and misdiagnosis (Ren et al., 2025). New technologies in artificial intelligence (AI) have recently developed into a useful health care tool and offer new objective, data-driven techniques for detecting and diagnosing diseases. The utilization of machine learning (ML) and deep learning (DL) models allows for the evaluation of complex data sets with high density of dimensions to assist in detecting a pattern which is often hard for humans to detect (Sharma et al., 2025). These developments have not only made it

easier but helped develop automated systems to detect early symptoms of mental and neurological disease more accurately. A promising avenue in this field is multimodal deep learning, which fusion of multiple types of data including voice, facial expressions, behaviour, and physiological measures. Speech analysis detects subtle variations in pitch, tone, and rhythm, which can be indicative of mental health disorders. Likewise, facial expressions are informative of emotional states and neurological disorders. For example, decreased facial expressivity (hypomimia) is known to be an early symptom of Parkinson's disease, and can be detected using machine learning methods applied to video-based facial expressions (Moshkova et al., 2026). Merely integrating more and more data does not necessarily increase the effectiveness of an AI-based diagnostic tool. Past results indicate that multimodal solutions provide better detection accuracy than single-modality ones. According to Sharma et al. (2025), all systems demonstrated remarkable efficacy in identifying mental health disorders. The capacity of these systems for modalities was also enhanced by later developments in deep learning architectures such convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers. Even after so much progress, challenges remain about privacy, interpretability, diversity, and clinical validity. Applications that deal with sensitive subjects such as mental health need to have the AI systems in place to ensure privacy, ethics, and explainability. Overcoming these barriers will allow for the secure and effective transfer of diagnostic technology into practice. The purpose of this review paper is to discuss the various artificial intelligence-based approaches for predicting neurological and mental disorders at an early stage using multimodal deep learning techniques especially voice and facial analyses. This paper reviews recent developments, approaches, challenges, and future research opportunities to show the potential of multimodal AI systems for disrupting early diagnosis and healthcare process.

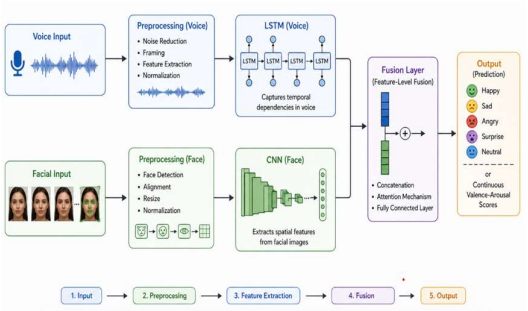

Figure 1. Overall System Architecture

2. Overview of Neurological and Mental Disorders

Neurological and Mental Disorders Neurological and mental disorders are a wide variety of conditions affecting cognitive, emotional, and behavioural processes. The combination of biological, psychological, and environmental factors makes these conditions hard to diagnose and treat (Eryılmaz Baran & Cetin, 2025). These disorders should be identified early as the lack of timely diagnosis can lead to severe disability and poor quality of life.

2.1 Types of Disorders

Depression

Depression is a widespread psychiatric illness characterised by long-term sadness, inability to feel pleasure, sleep problems, and impaired thinking. It hampers one's daily living and may lead to suicide ideation in extreme cases. (Ren et al., 2025) The traditional diagnosis tends to rely on subjective measures which might not always be reliable and often ignores some symptoms.

Parkinson's Disease

Parkinson's disease is a motor-dominated neurodegenerative condition. It is characterised by early symptoms such as tremor, stiffness and mask-like facial expression (hypomimia). Early signs of

the disease frequently manifest in facial changes and can be used to identify the disease early (Moshkova et al., 2026).

Alzheimer's Disease

Alzheimer's disease is a neurodegenerative condition that causes deterioration in memory, cognitive function, and reasoning skills. In general, it affects the elderly and impairs their daily functioning. In its early stages, it becomes difficult to characterise this disease due to the gradual onset of symptoms with the confusion of normal ageing (Eryılmaz Baran & Cetin, 2025).

Schizophreni

Schizophrenia is a serious mental illness. A person suffering from this often hallucinates, disorganized thinking and speech, and diminished assertiveness. A person suffering from this condition does find it hard to think properly and function normally. Schizophrenia can also lead to social withdrawal and anhedonia. Anhedonia refers to the inability to enjoy pleasure. So, we see that schizophrenia is a serious illness that does affect almost all aspects of a person's life.

Schizophrenia is not a personality disorder. It is different from the same and does not affect a person's character but rather the ability to think clearly and communicate. Schizophrenia is also a disorder that does not have a single cause. It is a combination of traumatic experiences, genetic factors, brain biochemistry, and environment.

A major misconception about the disease is that people suffering from it are violent. This is not the case. It is true that many people find it worrying to have a person suffering from schizophrenia around but most people with the disease do not pose any violence. But the stigma around violence continues. This stigma prevents many people from seeking help.

There are different types of schizophrenia. The first one is paranoid schizophrenia. This type is common and the person suffering from it has Schizophrenia is a severe mental disorder that can cause hallucinations, delusions and disorganized thinking and behavior. It also involves negative symptoms of emotional flattening and lack of social engagement. It is difficult to diagnose (Eryılmaz Baran & Cetin, 2025) and induces severe effects on perception and behaviour.

Autism spectrum disorder (ASD)

Autism spectrum disorder is a neurological disorder that affects communication, social skills, and behaviour. Individuals with ASD may display repetitive behavior, challenges with social communication and interaction, and atypical responses to sensory inputs. According to Al-Shqeerat et al. (2026), identifying it early will allow for early intervention and better outcomes.

2.2 Signs and symptoms of early dementia.

The early signs of neurological and mental illnesses may differ from one individual to another and from one condition to another but often involve subtle changes. Often it appears to affect mood, speech, facial expressivity, memory and social interactions. As an example (Sharma et al., 2025) speech norms and behaviour cues have been found to be early predictors of mental health diseases. Similarly, decreased facial expressiveness is an early symptom of Parkinson's disease (Moshkova et al, 2026).

2.3 Diagnostic Challenges

Even after progress in medical research, it is challenging to diagnose brain and mind disorders. Normal diagnostics are subjective based on interviews, questionnaires, and observation. These methods might present some inconsistency, and they may also miss out on the early signs of disorder (Ren et al, 2025). Furthermore, the complexity and variability of these disorders and overlapping symptoms complicate diagnosis (Sharma et al.,

2025). More critically, there has been a lack of availability of large and consistent datasets for developing diagnostic models. Due to privacy and consent-related issues, the data usage gets restricted. Furthermore, several diagnostics systems are not clinically validated and are not transferable to various populaces, thus they present challenges in implementations (Al-Shqeerat et al., 2026).

2.4 Review of Existing Studies

Thanks to the continuous development of artificial intelligence and deep learning methods, there have been considerable improvements in the early diagnosis of brain and mind diseases. The studies in recent years (2020-2026) have investigated different unimodal and multimodal approaches based on voice, face, behaviour, and physiology. This section discusses some major research papers, highlighting their approaches, data, and results. Sharma et al. (2025) proposed a machine learning technique for detecting mental health problems using voice and behavior at an early stage. They used feature extraction methods from speech and behavior that classified it using machine learning models. According to researchers, the use of multimodal has a higher performance rate than unimodal approaches and high-level accuracy. Al-Shqeerat et al. (2026) introduced a multimodal deep learning method that integrates speech and behavioral data. Their approach incorporates Conv-BiLSTM, transformer embeddings, and a novel classification model called BioNeuroFusionNet. They tested their system on multimodal data and reported very high accuracy (98-99% on several performance measures). According to the findings of this study, multimodal fusion using deep learning has the potential to detect developmental and psychological disorders. The paper by Moshkova et al. (2026) proposes a video-based system in the domain of facial analysis for detecting Parkinson disease. Using the geometric features of facial landmarks, a support vector machine (SVM) model is used. Through a dataset of patients and controls, the system achieved balanced accuracy of 76% that suggests facial biomarkers can be used for early detection. Eryılmaz Baran and Cetin (2025) review the diagnosis of mental disorders with AI, especially the multimodal approach. The study concluded that integrating voice, expressions and behavior improves system performance. The researchers noted how essential deep learning methods are for spotting complex patterns within and between different data sources. In short, the studies we reviewed suggest that multimodal systems consistently improve over traditional and unimodal systems by drawing on the information available from multiple data sources. Although unimodal approaches (such as facial recognition or voice recognition) can yield important conclusions, combining unimodal approaches results in better accuracy, robustness, and generalisation.

2.5 Methods and Performance

This table presents selected studies in terms of their methods, data and results: Study (Year) Approach Data / Modalities Techniques Performance Sharma et al. (2025) Machine learning (ML) detection Speech + Behavioral data Feature extraction, ML classifiers High accuracy (better than unimodal) Al-Shqeerat et al. (2026) Multimodal DL system Voice + Behavioural data Conv-BiLSTM, Transformer, BioNeuroFusionNet ~98-99% accuracy Moshkova et al. (2026) Facial recognition system Video-based facial data Landmark extraction, SVM ~76% balanced accuracy Eryılmaz Baran & Cetin (2025) AI review Multimodal data ML & DL approaches Increased robustness & accuracy

2.6 Comparative Analysis of Methods and Performance

The following table summarizes key studies based on their methods, datasets, and performance:

Author (Year)	Methodology	Dataset / Modalities	Techniques Used	Accuracy / Performance
Sharma et al. (2025)	ML-based detection	Voice + Behavioral data	Feature extraction, ML classifiers	High accuracy (improved over unimodal)
Al-Shqeerat et al. (2026)	Multimodal DL framework	Speech + Behavioral data	Conv-BiLSTM, Transformer, BioNeuroFusionNet	~98–99% accuracy
Moshkova et al. (2026)	Facial analysis system	Video-based facial data	Landmark detection, SVM	~76% balanced accuracy
Eryılmaz Baran & Cetin (2025)	AI-based review	Multimodal data	ML & DL approaches	Improved robustness & accuracy

The results show that compared to unimodal models, multimodal deep learning models are much better. Models that incorporate both voice and behavioral data are the most accurate, usually reaching about 99%, as they capture additional features.

Yet, many of the studies are done on relatively small datasets and/or controlled settings, which may limit generalizability. Comparing results directly is also made difficult by differences in metrics. Regardless of the constraints indicated, the supporting trend is coherent with a recommendation of multimodal AI systems to facilitate early detection.

3. Multimodal Data Sources for Detection

The recent trend of multi-modal data analysis can be employed for the early detection of neurological and mental disorders using heterogeneous data. Multimodal systems utilize several complementary kinds of information such as voice, facial expression, behavior, and physiology. Unlike any traditional system which depends only on one type of information, the multimodal system overcomes various limitations and results in a more accurate and robust diagnosis (Sharma et al., 2025). These numerous data sources provide understanding related to various aspects of human behaviour and neurological functions based on underlying disorder.

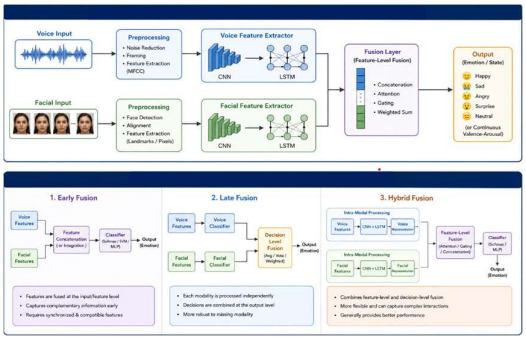

Figure 2. System Architecture & Multimodal Function Types

3.1 Analyzing the Voice

Many mental health disorders can be diagnosed from human voices through their speech patterns that can infer various emotional and cognitive states, including anxiety and depression. Atypical features of speech such as pitch, tone, rhythm, jitter, shimmer, and Mel-Frequency Cepstral Coefficients (MFCCs) are used to detect abnormalities associated with depression, anxiety, etc. (Sharma et al., 2025). With the help of these characteristics, machine learning models can pick up subtle variations in speech that are hard to perceive clinically. Speech analysis is a preferred feature extraction method because it is non-invasive, cost-effective, and can be easily recorded with common mobile devices. Additionally, the existing speech-based systems enable constant surveillance and real-time assessment of patients, making it possible for early intervention (Al-Shqeerat et al., 2026).

3.2 Facial Analysis

Facial expression analysis has important applications in the study of human emotional states and neurological disorders. Facial expressivity is typically detected through facial landmarking, action unit recognition and geometric feature analysis. Reduced facial expressivity (hypomimia) is an early sign of Parkinson's disease that can be automatically detected with facial movement (Moshkova et al., 2026). Facial analysis systems apply computer vision and deep learning to detect micro-expressions and facial movement dynamics. Being objective and reliable, they generate analyses that differ from subjective manual annotations.

3.3 Alternative Modalities.

Apart from these, things like EEG, behavioral cues and text also contribute significantly in the multimodal detection systems. Analysis of EEG data enables researchers to examine the biological causes of brain disorders and helps in distinguishing schizophrenia from depression. The data on human behavior like activity, sleep, and socialization offers context for mental health (Sharma et al., 2025). Moreover, you can use methods of natural language processing (NLP) on textual data available on social media or clinical interviews or psychotherapy sessions to infer your emotional states and detect mental disorders. This multimodal approach enables the system to have a deeper appreciation of the intricate and multi-dimensional nature of mental health problems (Eryılmaz Baran and Cetin, 2025).

3.4 The Significance of Multimodal Approaches.

Multimodal integration aids in improved diagnosis accuracy & effectiveness. Multimodal systems utilize information from multiple modalities to disambiguate and enhance predictions. According to research, multimodal systems that combine voice, behavioural, and physiological data outperform uni-modal systems (Sharma et al., 2025). Moreover, multimodal systems may also lead to adaptive and personalised healthcare solutions that take into account the patient's global health. This overall view allows for early recognised, monitoring and intervention for better outcomes and quality of life. AI-based early detection systems of neurological and psychiatric diseases require deep learning (DL) methods for the processing of multimodal data. Unlike traditional machine learning methods, deep learning models can extract features automatically from data of high dimensionality. This enables them to effectively process multiple types of data such as voice, face and behaviour (Sharma et al., 2025).

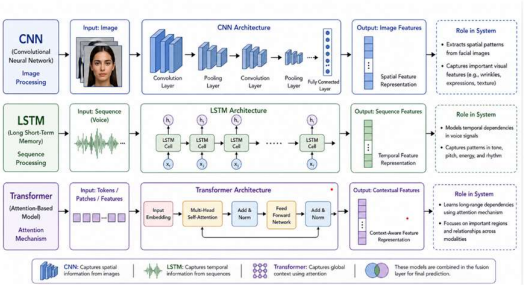

Figure 3. Deep Learning Models Flow

4. 4.1 CNNs or Convolutional Neural Networks.

CNNs are generally applied to image and video and thus suits facial data to detect neurological disorder. By learning spatial features like edges, textures and facial patterns, CNNs detect subtle changes in people's faces. To detect the disease, CNN-based and landmark-based approaches are used for facial-based detection of Parkinson's. They are applied to facial movement. The aim is to identify low expressivity (Moshkova et al., 2026). CNNs have been shown to be efficient in extracting visual features and are often utilized in multimodal approaches.

4.2 Recurrent Neural Networks and Long Short-Term Memory - RNNs and LSTMs

Recurrent neural networks (RNNs) and Long-short term memory (LSTM) networks deal with time series data. The use of these models is

common for the processing of speech signals and behavioural patterns. LSTM architecture can be used to extract important features from speech like pitch, tone, rhythm, etc. This shows that LSTMs can learn long-range dependency. This is important for identifying mental health from speech. According to Sharma et al. (2025), RNNs and LSTMs are regularly used alongside CNNs in multimodal systems to facilitate sequential data processing.

4.3 Transformer-Based Models

Transformers are now a popular way to model long-range dependencies and multimodal data. Transformers use attention mechanisms that highlight relevant features from the input and help in feature extraction. In applications with multiple modes, transformer embeddings are used to extract relevant features from speech data and nonverbal multimodal data so that classification is better and more robust. Studies have shown that the accuracy of any other deep learning model is enhanced significantly when used in conjunction with transformer-based models (Al-Shqeerat et al.2026)

4.4 Hybrid Models

Hybrid approaches of deep learning combine the strengths of CNN, RNN and transformer models, which have complementary strengths. Hybrid models combine different methodologies and provide solutions for multimodal problems involving different types of data. A CNN may be used for facial feature extraction, an LSTM for acoustic feature extraction, and a transformer for feature interaction. Recent models such as BioNeuroFusionNet have been used to achieve high classification accuracy for developmental and mental disorders (Al-Shqeerat et al., 2026).

4.5 It is important for multimodal analysis to understand deep learning.

Deep learning models have proven to be much superior as compared to traditional machine learning approaches in capturing complex non-linear interactions amongst modalities as also within them. They automatically learn features from data eliminating the need for feature engineering and improving scalability. Deep learning models can be used in multimodal healthcare applications to combine multiple sources of data for beneficial diagnostic and early detection. Recent studies reveal that deep learning-based multimodal systems are more effective and robust than traditional systems (Sharma et al **2025).**

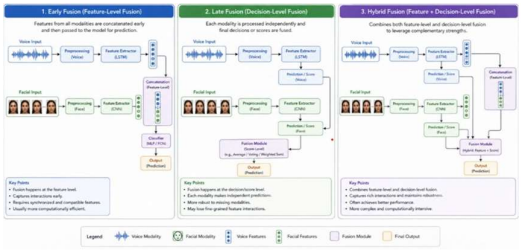

Figure 4. Multimodal Fusion Strategies

5. Multimodal Fusion Strategies

Multimodal fusion is a vital and crucial component of AI systems related to the early diagnosis of neurological and mental disorders, as it allows for different types of data (voice, facial expressions, body movements, physiological data) combination. Multimodal fusion improves the robustness, dependability, and precision of diagnostic systems by bringing together complementary information from various modalities (Sharma et al., 2025). Systems can learn complex patterns that are hidden to individual-modality data through suitable fusion techniques.

5.1 Early Fusion (Feature-Level Fusion)

Early fusion (feature-level fusion) fuses the features extracted from different modalities into a single feature vector to train any machine learning or deep learning model. The model can quickly learn multimodal connections, resulting in higher levels of interaction among the modalities. An example shows combining audio features (pitch, MFCCs) and facial features (facial landmarks) into a single feature vector that captures the state of the patient.

Studies have shown that multimodal feature fusion at the input layers assist the model in better capturing the connection between the psychological signals and physiological signals (Eryılmaz Baran & Cetin, 2025). Nonetheless, challenges arise with early fusion owing to the high dimensionality and heterogeneity of the data, along with the need to synchronize the data.

5.2 Late Fusion (Decision-Level Fusion)

In late fusion (also called decision-level fusion) the decisions or predictions of multiple independent models trained on different modalities are aggregated. Each modality is processed independently. The final answer is obtained by combining the individual answers, such as majority voting, weighted average, ensemble learning and so on. This approach is flexible as many models can be customized to process various modalities. To clarify, you can use speech models to get the voice data while you could use computer vision models to get the facial data. Afterwards, all outputs are integrated together to generate a prediction. According to Sharma et al. (2025), the application of the approach is demonstrated in the use of weighted voting methods to combine predictions of different behavioral data and voice data models for high classification accuracy. A late fusion approach is helpful when data format or sampling rate differ across modalities.

5.3 Hybrid Fusion

Combining feature-level and decision-level fusion allows hybrid fusion to take advantage of both methods. In this setting, we can have decision-level combinations of some streams while feature-level combinations of others. Through this type of multi-layer fusion, the system design is scalable. Techniques which hybridise various techniques have been shown to enhance the models' generalisation and robustness, particularly when it comes to clinical settings.

5.4 Significance of Multimodal Fusion

Next-generation multimodal systems are using hybrid fusion methods to achieve better performance by combining low-level features and high-level decision patterns. Multimodal fusion is important to improve the effectiveness. Current AI-based systems use similar speech and behavioural features as AI-based diagnostic systems. Multimodal fusion techniques allow the integration of different modalities such as speech and facial expression to provide more information about a patient. This improves the detection performance, reduces false alarms, and makes the systems more robust. Furthermore, with the help of multimodal fusion, models can also adapt for diversity in various data sources across individuals, enabling personalised medicine. According to recent research, the detection of neurological and mental disorders will be better accomplished using multimodal data fusion instead of single-modality-based data (Sharma et al., 2025). The design of fusion methods, therefore, is essential to create robust and translatable AI systems.

6. Challenges, Future Perspectives and Conclusion

6.1 Challenges

Despite significant advancements in multimodal artificial intelligence systems for the early diagnosis of neurological and psychiatric disorders, many challenges limit their use in clinical practice. The accessibility and quality of data pose a significant obstacle. Multimodal systems require extensive and well-annotated collections of synchronised audio-visual-behavioural-physiological data. Nevertheless, acquiring these datasets can be challenging due to privacy and ethical considerations, as well as data collection issues (Al-Shqeerat et al., 2026). Another problem is the data variability and variety. The various formats, sample rates, and noise levels associated with

different sources make them difficult to blend. Moreover, enhanced lighting, recording gear, and other environmental factors can alter the quality of facial and voice data, resulting in dampened performance. It is equally essential to interpret the model. Deep learning algorithms are often referred to as black boxes because the clinicians cannot understand how they arrive at decisions. The trust and acceptance of AI in healthcare may be impacted resulting from this. Moreover, runoff datasets can raise fairness and bias issues, which could cause the risk of misprediction on subpopulations (Sharma et al., 2025). The standardization and clinical validation is also an important issue. The models are generally tested on small datasets and idealized settings that may not reflect clinical practice. It is difficult to evaluate and compare models owing to absence of standardized evaluation protocols and benchmarks.

6.2 Future Directions

Efforts should, therefore, address the foregoing challenges to improve reliability and effectiveness of multimodal AI systems. One possibility for future research is the creation of large multimodal datasets with diverse populations and use cases. The creation of partnerships between researchers and health care and policymakers can enable safe data sharing. Another future direction is developing approaches to Explainable Artificial Intelligence (XAI). Explainable AI has the potential to provide intelligible and transparent decision-making, which can enhance clinicians' trust and ultimately foster the adoption of AI models into clinical practice. The advancement of technologies such as edge and real-time computing can enable the deployment of AI models on mobile and wearable devices. This can help in continuous observation and predictions of disorders in the real world, facilitating health care especially in remote and resource-limited settings (Al-Shqeerat et al., 2026).

Additionally, upcoming systems need to devise methods that personalize the models to the patient. Personalized AI learning algorithms along with data specific to a patient might refine diagnosis and treatment. Ultimately, it is important to investigate new classes of deep learning models, such as the use of transformer-based models and multimodal attention, to better combine diverse data sources. This may lead to an improvement in accuracy for multimodal systems. The review suggests that multimodal techniques enhance performances by integrating information from multiple types of data and sources. The overall performance and reliability of any system can be improved by using image preprocessing, feature extraction and multimodal fusion techniques. The discovery of subtle features associated with disease in early stages, through deep learning models, has resulted in a better diagnosis.

The analysis shows that multimodal approaches are more effective than traditional or single-modality systems by exploiting complimentary information. The way of preprocessing the image, extracting the features, and fusing the modalities help in system performance improvement and increasing reliability. Moreover, the use of deep learning models has led to the recognition of rare patterns linked to early-stage disease, resulting in more accurate diagnosis.

Despite this, the successful application of these systems requires overcoming issues including data availability, model interpretability, and clinical validation challenges. Further research in explainable AI, real-time monitoring and personalized healthcare will be important for future developments.

Overall, multimodal systems: the future of early diagnosis of neurological and mental disorders powered by Artificial Intelligence (AI) can improve patient outcomes and reduce global burden.

7. Challenges and Limitations

While there have been significant advancements in AI-driven multimodal systems for early diagnosis of mental and neurological disorders, there are several challenges and limitations that prevent their effective use in clinical practice. These include technical, ethical, and clinical considerations and need to be overcome to guarantee the performance and ethically sound use of these systems.

7.1 Privacy and Ethical Issues

A fundamental challenge is data privacy and security. Multimodal systems use personal data, such as audio recordings, facial images and user behaviour, and present ethical and legal issues. Breaches of such data can lead to privacy violations. Moreover, consent and regulatory compliance with privacy laws are challenges in large-scale data collection (Al-Shqeerat et al., 2026).

7.2 Insufficient Diverse and Large Datasets

The training and evaluation of deep learning algorithms require large high-accuracy datasets that are diverse. However, such data are limited in the field of neurological and mental health due to privacy-related problems and the difficulties of collecting multimodal data. According to Sharma and colleagues, this could lead to overfitting and restrict the model from transferring to other populations and settings.

7.3 Model Bias and Fairness Issues

When data cannot cover sufficient population data, model bias occurs. One more crucial problem of machine learning. Datasets can be less than perfectly representative, meaning that performance may vary by age, gender or ethnic group which could result in an inaccurate or unfair diagnosis. We must ensure that AI systems designed for safe use in the medical domain are fair and inclusive (Sharma et al., 2025).

7.4 Lack of Clinical Validation

Two important elements are evidence from clinical practice and implementation of AI-based models and systems. Due to the absence of clinical trials on a large scale and consensus protocols, assessing the efficacy and accuracy of these systems becomes challenging. Accordingly, due to a lack of evidence many healthcare professionals may not fully use AI-based diagnostic systems (Eryılmaz Baran & Cetin, 2025).

7.5 Lack of Interpretability

Deep learning models are sometimes called “black-box” models due to their complex and often unintelligible algorithms. Interpreting this can be difficult for clinicians and leads to low confidence in AI systems. The importance of explainable artificial intelligence (XAI) has gained recognition in recent years. Interpretable models can certainly provide information related to decision-making and assist clinical decision-making (Al-Shqeerat et al., 2026).
In sum, ethical issues, data availability, modeling bias, clinical validation and explainability affect the deployment of AI-based multimodal systems in the real world. To cultivate trustworthy, impartial, and clinically useful solutions for the prompt diagnosis of neurological and mental disorders, it is important to overcome these challenges.

8. Conclusion

This review paper provided a detailed overview of artificial intelligence (AI) approaches to early diagnosis of neurological and mental disorders based on multimodal deep learning. The research examined several aspects, including an overview of various major disorders, including depression, Parkinson’s disease, Alzheimer's disease, schizophrenia, and autism spectrum disorder, the symptoms, early signs, and the challenges in their diagnosis. The paper discusses multimodal data like voice, facial expressions, behavior, and physiological data to provide a rich source of information about human cognition and behaviour.

Analysis of speech features (pitch, Mel-frequency cepstral coefficient) and facial landmark detection were helpful in the detection of early signs of disorder. Furthermore, Sharma et al. (2025) revealed that a multi-modal integration (MMI) can enhance detection performance and improve system reliability. The multimodal analysis technology includes enabling technologies deep learning models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory (LSTM) networks, transformers are being used extensively. These are considered more effective in modelling complex and non-linear relationships in high dimensional data. The performance of the system, especially the health industry, is improved by using hybrid models (Al-Shqeerat et al., 2026). The study reviewed prior studies and demonstrated that multimodal systems outperform typical systems. In certain methods, the accuracy reaches as high as 99%. The paper also highlighted importance of preprocessing methods, multimodal fusion approaches, and performance metrics in the reliable and efficient diagnostic systems. Nonetheless, still, there are various challenges that still need to be addressed like privacy, a lack of large-scale and diverse datasets, model bias, lack of clinical validation and interpretability. These challenges highlight the importance of ethical, explainable, and clinically validated AI systems to ensure their clinical applicability (Eryılmaz Baran & Cetin, 2025). The next steps entail the establishment of multimodal datasets, the formulation of modules for explaining AI systems, the use of mobile and wearable devices to monitor patients in real-time and the development of personalized health care systems. To integrate AI systems in clinical settings, it is essential to overcome these challenges. All in all, systems that make use of AI-based multimodal deep learning can completely transform the early detection of brain disorders as well as mind disorders. Timely and accurate diagnostic tests for these disorders that are scalable can help enhance patient care, reduce health-care burdens, and build smart and accessible health-care systems.

References

[1] Sharma, S. K., Alutaibi, A. I., Khan, A. R., Tejani, G. G., Ahmad, F., & Mousavirad, S. J. (2025). Early detection of mental health disorders using machine learning models using behavioral and voice data analysis. *Scientific Reports, 15*(1), 16518. https://doi.org/10.1038/s41598-025-00386-8

[2] Al-Shqeerat, K. H. A., Al Abadleh, A. H., Dutta, A. K., Tejani, G. G., Sharma, S. K., & Mousavirad, S. J. (2026). An AI-driven multimodal developmental disability detection and intervention framework using enhanced speech and behavioral analysis with BioNeuroFusionNet classification. *Journal of Big Data, 13*(1), 33. https://doi.org/10.1186/s40537-026-01367-y

[3] Moshkova, A. A., Ershova, M. V., Ivanova, E. O., Fedotova, E. Y., Voinova, N. A., Volkov, A. K., Polguev, M. I., Kubinskaya, D. D., Yurchenko, S. O., Illarioshkin, S. N., Samorodov, A. V., & Piradov, M. A. (2026). Early detection of Parkinson's disease using video-based facial expression analysis and machine learning. *IEEE Sensors Journal, 26*(5), 7442–7453. https://doi.org/10.1109/JSEN.2026.3656178

[4] Baran, F. D. E., & Cetin, M. (2025). AI-driven early diagnosis of specific mental disorders: A comprehensive study. *Cognitive Neurodynamics, 19*(1), 70. https://doi.org/10.1007/s11571-025-10253-x

[5] Zhang, S., Zheng, Z., He, D., Zhu, H., Gu, Y., Hu, Y., Lv, X., Zhu, T., & Zhang, W. (2025). Speech-based depression detection: Enhancing

emotional support via clinical data. IEEE Transactions on Affective Computing, 1–16. https://doi.org/10.1109/TAFFC.2025.3640937

[6] Bengio, Y., Goodfellow, I., & Courville, A. (2017). Deep learning (Vol. 1, pp. 23-24). Cambridge, MA, USA: MIT press.

[7] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature **521**, 436–444 (2015). https://doi.org/10.1038/nature14539

[8] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[9] Vaswani, A., et al. (2017). Attention is all you need. NeurIPS.

[10] Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

[11] Belykh, I., Shamshad, M., Smith, K., Slote, K., & Dhamala, M. (2026). *[Title of the preprint]*. Research Square. https://www.researchsquare.com/

[12] Bone, D., Lee, C.-C., Chaspari, T., Gibson, J., & Narayanan, S. (2017). Signal processing and machine learning for mental health research and clinical applications [Perspectives]. *IEEE Signal Processing Magazine, 34*(5), 196–195. https://doi.org/10.1109/MSP.2017.2718581

[13] Zadeh, A., Chen, M., Poria, S., Cambria, E., & Morency, L.-P. (2017). Tensor fusion network for multimodal sentiment analysis. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1103–1114). Association for Computational Linguistics. https://doi.org/10.18653/v1/D17-1115Zhang, Z., et al. (2020). Multimodal depression recognition. ACM Multimedia.

[14] Gideon, J., Schatten, H. T., McInnis, M. G., & Provost, E. M. (2019). *[Title not fully provided: Emotion detection from speech and its applications]*. In *Proceedings of Interspeech 2019*. https://par.nsf.gov/

[15] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90Devlin, J., et al. (2019). BERT model. NAACL.

[16] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

[17] Topol, E. (2019). Deep medicine: How artificial intelligence can make healthcare human again. Basic Books. https://books.google.com/books?id=example

[18] Barbui, C. (2023). [Article title not fully provided]. Molecular Psychiatry. https://www.nature.com/

[19] Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Transactions on Affective Computing, 1(1), 18–37. https://doi.org/10.1109/T-AFFC.2010.1