International Journal of
Applied Data Science in Engineering and Health
https://ijadseh.com

Deep Learning Applications for Mental Health Disorder Diagnosis Using Medical Imaging

Ahmad Razavitabar[1], Faezeh Norouzi[2], Tahereh Afrasyabi[1]

[1] *Islamic Azad University, South Tehran Branch, Iran*

[2] *Psychiatry and Behavioral Science, Isfahan University of Medical Science, Isfahan, Iran*

Received date: October 12, 2025; Accepted date: November 9, 2025

Abstract

Mental health disorders such as depression, schizophrenia, and bipolar disorder remain difficult to diagnose objectively. Medical imaging techniques—such as MRI, fMRI, PET, and EEG—offer valuable insights into brain structure and function, revealing biomarkers linked to psychiatric conditions. Deep learning has recently transformed the analysis of these complex data by automatically extracting meaningful features from high-dimensional images. This review summarizes recent advances in applying deep learning models, including convolutional, recurrent, and graph neural networks, to mental health diagnosis. It highlights key imaging modalities, representative applications, and current limitations such as small datasets and limited interpretability. Emerging directions, including multimodal fusion and explainable AI, promise to enhance clinical reliability and understanding. Deep learning thus holds strong potential for improving early detection and personalized treatment in mental health care.

Keywords*:* Mental Health; Deep Learning; Depression; Anxiety.

Introduction

Mental health disorders such as depression, schizophrenia, bipolar disorder, and autism spectrum disorder represent a significant global health burden, affecting hundreds of millions of individuals worldwide. These conditions often lead to considerable personal suffering, social impairment, and economic costs. Despite their prevalence, the diagnosis of mental health disorders remains largely dependent on clinical interviews, behavioral assessments, and subjective symptom reporting. Such approaches are prone to variability between clinicians and may fail to capture the underlying neurobiological mechanisms of these disorders [1]. Recent advances in medical imaging have opened new possibilities for objective and quantitative assessment of brain structure and function. Techniques such as magnetic resonance imaging (MRI), functional MRI (fMRI), positron emission tomography (PET), and electroencephalography (EEG) provide detailed insights into the brain's anatomy, activity, and connectivity patterns. These imaging modalities have revealed subtle but consistent neural alterations associated with various psychiatric conditions, offering the potential for early diagnosis and personalized treatment planning [2].

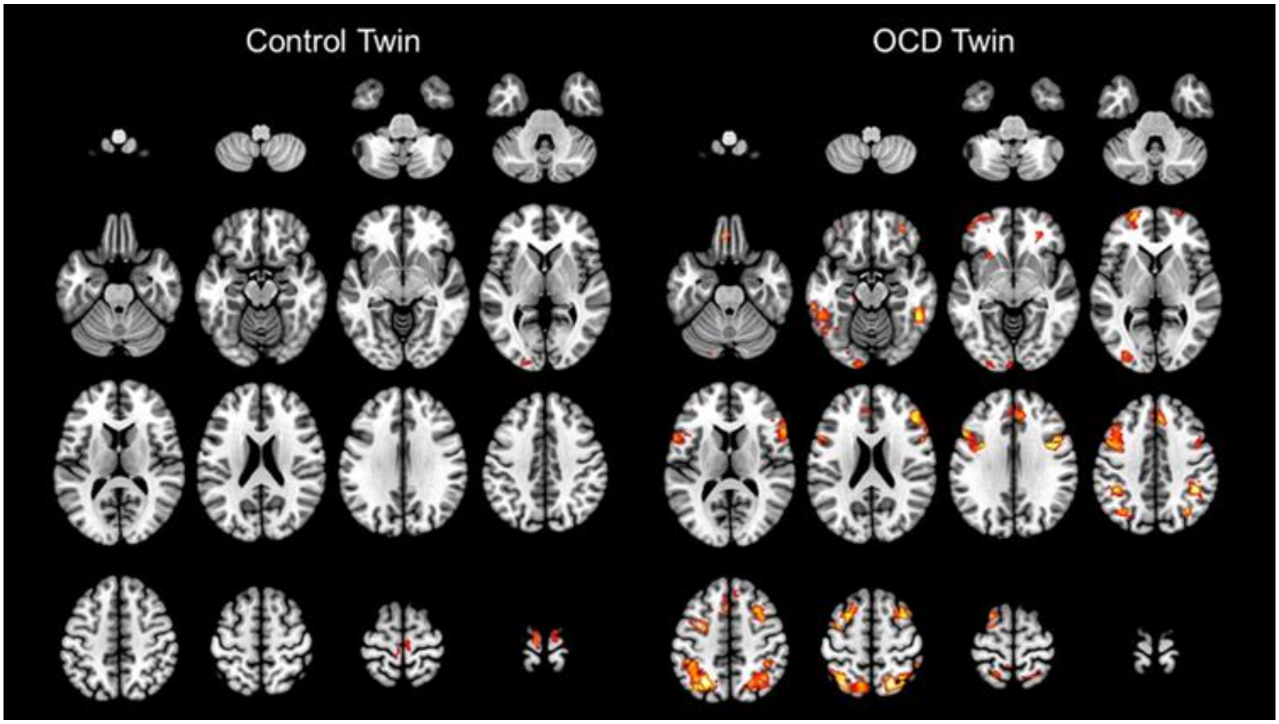

Figure 1MRI of the brains of a pair of identical twins. The one on the right has OCD, while the one on the left does not [3].

However, analyzing these high-dimensional and complex imaging data presents significant challenges. Traditional statistical and machine learning methods often require manual feature extraction and struggle to model the nonlinear relationships inherent in neuroimaging data. In contrast, deep learning has emerged as a transformative approach capable of automatically learning hierarchical and discriminative representations from raw images or signals. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and graph neural networks (GNNs) have shown remarkable success in pattern recognition tasks across diverse imaging modalities [4-7]. In the context of mental health, deep learning models have demonstrated promising results in detecting structural and functional abnormalities, identifying biomarkers, and differentiating between patient subgroups. These techniques can uncover complex brain signatures linked to psychiatric symptoms, advancing our understanding of mental illnesses beyond traditional diagnostic frameworks. Despite this progress, several barriers—such as limited dataset sizes, inter-site variability, and lack of model interpretability—still hinder clinical deployment.

This short review aims to summarize recent developments in the application of deep learning techniques to medical imaging for mental health disorder diagnosis. It highlights commonly used imaging modalities, key deep learning architectures, representative studies on major psychiatric conditions, and current challenges and future research directions in this rapidly evolving field.

Common Imaging Modalities Used in Mental Health Diagnosis

Medical imaging plays a pivotal role in investigating the structural, functional, and metabolic abnormalities associated with mental health disorders. Each modality provides unique insights into the brain's organization and activity, supporting data-driven diagnostic and predictive modeling. The following are the most commonly used imaging techniques in deep learning–based psychiatric research.

Structural Magnetic Resonance Imaging (sMRI)

Structural MRI is widely employed to examine brain anatomy and detect morphological abnormalities in psychiatric disorders. It provides high-resolution images of gray and white matter, allowing the measurement of cortical thickness, surface area, and volumetric changes in specific brain regions. Studies have shown that patients with depression or schizophrenia often exhibit reduced gray matter volume in the prefrontal cortex and hippocampus. Deep learning methods, particularly 3D convolutional neural networks (3D CNNs), have been used to automatically extract structural features from sMRI for the classification of mental disorders. These models outperform traditional feature-based approaches by learning spatially distributed biomarkers directly from imaging data [8,9].

Functional Magnetic Resonance Imaging (fMRI)

Functional MRI measures blood-oxygen-level-dependent (BOLD) signals to assess brain activity and functional connectivity patterns. It is particularly useful for identifying disrupted brain networks in mental illnesses such as major depressive disorder, schizophrenia, and autism spectrum disorder. Deep learning models applied to fMRI data often focus on learning spatiotemporal patterns that represent brain dynamics. Recurrent neural networks (RNNs) and graph convolutional networks (GCNs) have been used to model these complex functional interactions. For instance, GNN-based frameworks can represent brain regions as nodes and their connectivity as edges, enabling the discovery of altered communication pathways in patients with psychiatric conditions [10].

Positron Emission Tomography (PET) and Single-Photon Emission Computed Tomography (SPECT)

PET and SPECT imaging techniques provide valuable information about cerebral metabolism, neurotransmitter activity, and receptor binding. These modalities are especially informative for disorders involving neurochemical imbalances, such as depression, Alzheimer’s disease, and schizophrenia. Deep learning approaches, including CNNs and autoencoders, have been utilized to identify metabolic signatures and predict disease progression. Moreover, hybrid PET/MRI systems enable multimodal deep learning models to jointly analyze metabolic and structural information, improving diagnostic accuracy [11].

Electroencephalography (EEG) and Magnetoencephalography (MEG)

EEG and MEG record brain activity with high temporal resolution, making them suitable for studying dynamic neural oscillations in psychiatric disorders such as attention-deficit/hyperactivity disorder (ADHD), depression, and anxiety. Deep learning models, particularly convolutional and recurrent architectures, have been applied to EEG signals for emotion recognition, stress detection, and early diagnosis of mental health abnormalities. By automatically extracting temporal and spectral features, these models have demonstrated improved classification accuracy compared to conventional signal-processing methods [12].

Multimodal Imaging

Combining multiple imaging modalities provides a more comprehensive understanding of the brain's structure and function. For example, integrating sMRI and fMRI allows simultaneous analysis of both anatomical and functional abnormalities, while fusion with EEG or PET data enhances sensitivity to neurochemical and electrophysiological changes. Multimodal deep learning frameworks—such as hybrid CNN-RNN or attention-based fusion models—have shown great potential in improving diagnostic performance and interpretability in mental health research [13].

Deep Learning Methods for Mental Health Diagnosis

Deep learning has emerged as a powerful subset of machine learning capable of extracting complex patterns from large, high-dimensional data such as medical images and neurophysiological recordings. Unlike conventional machine learning approaches that rely on handcrafted features, deep learning models automatically learn hierarchical representations directly from raw input data, enabling them to capture subtle and nonlinear relationships in brain imaging. In the context of mental health, these methods have been successfully applied across a range of modalities—MRI, fMRI, PET, and EEG—to aid in diagnosis, prognosis, and biomarker discovery.

Convolutional Neural Networks (CNNs)

Convolutional Neural Networks are the most widely used architecture for image-based applications. CNNs employ convolutional and pooling layers to hierarchically extract spatial features, making them particularly effective for analyzing 2D and 3D medical images. In psychiatric neuroimaging, CNNs have been applied to structural MRI for identifying morphological abnormalities in schizophrenia and depression, as well as to fMRI for detecting altered functional connectivity patterns. For example, 3D CNN models trained on MRI scans have achieved high classification accuracy in differentiating individuals with major depressive disorder from healthy controls, outperforming traditional voxel-based morphometry methods [14,15]

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) Networks

RNNs are designed to handle sequential data by maintaining temporal dependencies across time steps. Their extension, Long Short-Term Memory (LSTM) networks, overcomes the vanishing gradient problem and is well suited for time-series neuroimaging and electrophysiological data such as EEG and fMRI. In mental health research, RNNs and LSTMs have been used to model temporal brain activity fluctuations, enabling the prediction of dynamic functional connectivity and disease states over time. These approaches are particularly valuable for disorders characterized by altered temporal synchronization, such as schizophrenia and bipolar disorder [17].

Autoencoders and Variational Autoencoders (VAEs)

Autoencoders are unsupervised neural networks that learn efficient low-dimensional representations of input data by reconstructing it through an encoder-decoder framework. Variational autoencoders (VAEs) extend this concept by introducing probabilistic modeling, which helps capture latent features useful for anomaly detection and feature compression. In mental health imaging, autoencoders have been utilized for denoising MRI scans, reducing dimensionality of fMRI data, and discovering latent biomarkers that differentiate patients from healthy subjects. Such models enable unsupervised learning from limited labeled data, a common constraint in psychiatric datasets [17].

Transformer Models

Transformers, originally developed for natural language processing, have recently gained attention in medical imaging due to their ability to model long-range dependencies using self-attention mechanisms. Vision Transformers (ViTs) and hybrid CNN-Transformer architectures have shown strong performance in analyzing MRI and EEG data for neurological and psychiatric disorders. Transformers can effectively capture both global and local patterns in brain images and are increasingly being explored for multimodal fusion—combining imaging data with clinical or genetic information to improve diagnostic robustness [18].

Applications for Specific Mental (and Related Neuropsychiatric) Disorders

Deep learning has been deployed across structural MRI (sMRI), functional MRI (fMRI), PET/SPECT, and EEG/MEG to detect disorder-specific brain signatures, stratify subtypes, and support prognosis. Below, we summarize representative applications and lessons learned.

Major Depressive Disorder (MDD)

CNNs on sMRI and resting-state fMRI have been used to capture morphology and connectivity alterations in cortico-limbic circuits (e.g., hippocampus, subgenual cingulate, medial prefrontal regions). 3D CNNs trained on T1-weighted MRI reported competitive diagnostic accuracy versus classical voxel-based pipelines, while fMRI models that learn graph or spatiotemporal representations reveal disrupted default-mode and salience network interactions. Despite good cross-validation results, cross-site generalization remains challenging due to scanner and cohort heterogeneity [19].

Schizophrenia

For schizophrenia, deep models frequently use connectivity graphs derived from fMRI and diffusion imaging, analyzed with GNNs/GCNs to identify dysconnectivity across fronto-temporal and default-mode systems. CNNs on sMRI also capture cortical thinning and subcortical volume differences. Reviews consistently report AUCs in the 0.70–0.90 range within site, with performance drops across sites; interpretability (e.g., relevance propagation, saliency) helps localize contributory regions and edges [20].

Bipolar Disorder

Bipolar disorder applications often fuse multimodal information—sMRI morphology, resting-state connectivity, and sometimes diffusion or spectroscopy—to separate bipolar from schizophrenia and unipolar depression. Hybrid CNN–RNN and attention-based models leverage both spatial and temporal patterns; nevertheless, label noise (episodic states, medication effects) and small cohorts hinder robustness. Domain adaptation and harmonization are active needs [21].

Autism Spectrum Disorder (ASD)

Large multi-site initiatives like ABIDE have enabled deep learning at scale. CNNs and autoencoders trained on sMRI/fMRI detect ASD-related alterations in default-mode, salience, and social cognition networks. Notably, site variability and age effects require careful confound control; nevertheless, end-to-end and representation-learning approaches (e.g., denoising autoencoders) have shown reproducible gains over handcrafted features [22].

Alzheimer's Disease / Mild Cognitive Impairment (MCI)

Although neurodegenerative rather than strictly psychiatric, AD/MCI studies are methodologically relevant. 3D CNNs and Transformer-based models on sMRI and PET learn atrophy and hypometabolism patterns. Multimodal fusion (PET+MRI) and transfer/self-supervised learning boost early-stage detection and MCI-to-AD conversion prediction, illustrating how multimodal pipelines can translate into psychiatric contexts [23].

Conclusion

Deep learning has become a transformative tool in the analysis of medical imaging for mental health disorder diagnosis. By automatically learning hierarchical and complex representations from high-dimensional neuroimaging data, deep learning models have achieved significant progress in detecting structural and functional brain abnormalities associated with disorders such as depression, schizophrenia, bipolar disorder, autism spectrum disorder, and Alzheimer's disease. Across modalities—from MRI and fMRI to PET and EEG—these approaches have demonstrated their ability to reveal hidden neural signatures that traditional methods often fail to capture. Despite the impressive advancements, several challenges remain before deep learning can be fully integrated into clinical psychiatry. Limited and imbalanced datasets, differences in imaging protocols across sites, and the "black-box" nature of many deep models hinder generalizability and clinical trust. Moreover, ethical concerns surrounding data privacy, bias, and interpretability emphasize the need for transparent and explainable AI systems. Addressing these issues requires larger, standardized datasets, advanced harmonization techniques, and collaboration between computer scientists, clinicians, and neuroscientists.

Looking forward, the future of deep learning in mental health imaging lies in multimodal integration, self-supervised learning, and explainable AI. Combining imaging data with genetic, behavioral, and clinical information could lead to more holistic models of psychiatric illness. As these technologies mature, deep learning has the potential not only to enhance diagnostic accuracy but also to reshape our understanding of the biological foundations of mental health, leading toward truly personalized and data-driven psychiatry.

Conflict of interest

The authors declared no conflict of interest.

References

[1] Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. The Lancet Psychiatry. 2016 Feb 1;3(2):171-8.

[2] Greicius MD, Flores BH, Menon V, Glover GH, Solvason HB, Kenna H, Reiss AL, Schatzberg AF. Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. Biological psychiatry. 2007 Sep 1;62(5):429-37.

[3] Diwadkar VA, Burgess A, Hong E, Rix C, Arnold PD, Hanna GL, Rosenberg DR. Dysfunctional activation and brain network profiles in youth with obsessive-compulsive disorder: a focus on the dorsal anterior cingulate during working memory. Frontiers in human neuroscience. 2015 Mar 17;9:149.

[4] Chen J, Liu Y, Wei S, Bian Z, Subramanian S, Carass A, Prince JL, Du Y. A survey on deep learning in medical image registration: New technologies, uncertainty, evaluation metrics, and beyond. Medical Image Analysis. 2025 Feb 1;100:103385.

[5] Rohani, M. M., Sharifi, S., & Durson, S. (2025). Deep learning in medical imaging for disease diagnosis. World Journal of Advanced Research and Reviews, 25(2), 2522–2526. https://doi.org/10.30574/wjarr.2025.25.2.0558

[6] Wang J, Wang S, Zhang Y. Deep learning on medical image analysis. CAAI Transactions on Intelligence Technology. 2025 Feb;10(1):1-35.

[7] Hosseinabadi F, Ghasemi Y. Artificial Intelligence in Radiology: Concepts and Applications. International Journal of Applied Data Science in Engineering and Health. 2025 Jun 18;1(3).

[8] Halkiopoulos C, Gkintoni E, Aroutzidis A, Antonopoulou H. Advances in neuroimaging and deep learning for emotion detection: A systematic review of cognitive neuroscience and algorithmic innovations. Diagnostics. 2025 Feb 13;15(4):456.

[9] Westlin C, Guthrie AJ, Paredes-Echeverri S, Maggio J, Finkelstein S, Godena E, Millstein D, MacLean J, Ranford J, Freeburn J, Adams C. Machine learning classification of functional neurological disorder using structural brain MRI features. Journal of Neurology, Neurosurgery & Psychiatry. 2025 Mar 1;96(3):249-57.

[10] Ktena SI, Parisot S, Ferrante E, Rajchl M, Lee M, Glocker B, Rueckert D. Metric learning with spectral graph convolutions on brain connectivity networks. NeuroImage. 2018 Apr 1;169:431-42.

[11] Lu D, Popuri K, Ding GW, Balachandar R, Beg MF. Alzheimer's Disease Neuroimaging I. Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. Sci Rep. 2018;8(1):5697.

[12] Craik A, He Y, Contreras-Vidal JL. Deep learning for electroencephalogram (EEG) classification tasks: a review. Journal of neural engineering. 2019 Apr 9;16(3):031001.

[13] Skaramagkas V, Pentari A, Kefalopoulou Z, Tsiknakis M. Multi-modal deep learning diagnosis of Parkinson's disease—a systematic review. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2023 May 18;31:2399-423.

[14] Vieira S, Pinaya WH, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. Neuroscience & Biobehavioral Reviews. 2017 Mar 1;74:58-75.

[15] Sarraf S, Tofighi G. Alzheimer's Disease Neuroimaging InitiativDeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. BioRxiv. 2016;4:070441.

[16] Alves CL, Thaise GD, Porto JA, de Carvalho Aguiar PM, de Sena EP, Rodrigues FA, Pineda AM, Thielemann C. Analysis of functional connectivity using machine learning and deep learning in different data modalities from individuals with schizophrenia. Journal of neural engineering. 2023 Sep 28;20(5):056025.

[17] Suk HI, Shen D. Deep learning-based feature representation for AD/MCI classification. InInternational conference on medical image computing and computer-assisted intervention 2013 Sep 22 (pp. 583-590). Berlin, Heidelberg: Springer Berlin Heidelberg.

[18] Bravo-Ortiz, M.A., Holguin-Garcia, S.A., Quiñones-Arredondo, S., Mora-Rubio, A., Guevara-Navarro, E., Arteaga-Arteaga, H.B., Ruz, G.A. and Tabares-Soto, R., 2024. A systematic review of vision transformers and convolutional neural networks for Alzheimer's disease classification using 3D MRI images. Neural Computing and Applications, 36(35), pp.21985-22012.

[19] Mwangi B, Hasan KM, Soares JC. Prediction of individual subject's age across the human lifespan using diffusion tensor imaging: a machine learning approach. Neuroimage. 2013 Jul 15;75:58-67.

[20] Udayakumar P, Subhashini R. Connectome-based schizophrenia prediction using structural connectivity-Deep Graph Neural Network (sc-DGNN). Journal of X-Ray Science and Technology. 2024 Jul;32(4):1041-59.

[21] Abrol A, Damaraju E, Miller RL, Stephen JM, Claus ED, Mayer AR, Calhoun VD. Replicability of time-varying connectivity patterns in large resting state fMRI samples. Neuroimage. 2017 Dec 1;163:160-76.

[22] Heinsfeld AS, Franco AR, Craddock RC, Buchweitz A, Meneguzzi F. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. NeuroImage: clinical. 2018 Jan 1;17:16-23.

[23] Lu D, Popuri K, Ding GW, Balachandar R, Beg MF. Alzheimer's Disease Neuroimaging I. Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. Sci Rep. 2018;8(1):5697.