

Prediction of Mental Problem Based on Deep Learning Models

Ziqi Huang^{1,a,*}

¹*Department of Mathematics and Data Science, University of California Santa Barbara, Goleta, United State*

a. ziqihung@gmail.com

**corresponding author*

Abstract: The rising prevalence of mental health issues identifies the urgent need for accurate, scalable, and timely prediction systems. Deep learning, a subset of a machine learning inspired by human's neuron structure, has offered an opportunity for innovative solutions for mental health diagnosis. The main idea of this paper is analyzing the application of deep learning in diagnosing mental disorders, including but not limited to Alzheimer, Parkinson and Schizophrenia. An enormous number of techniques will be put into real life application while dealing with diagnosis of mental health issues, but our focus will be convolutional neural networks (CNN), recurrent neural networks (RNN) and deep neural networks (DNN) strategies this time. On top of that, we'll also discuss integrating multimodal data, combining neuroimaging with behavioral and textual analysis which is able to enhance diagnostic accuracy. Techniques like CNN-DNN hybrids provide another comprehensive view of mental health issues. Despite these successes, challenges such as data privacy, ethical considerations, and security issues remain significant obstacles, which is our topic as well. Future prospects include federated learning, blockchain technology, and privacy-preserving methods to enhance security and maintain patient trust. All in all, this paper aims at giving a comprehensive understanding of deep learning methods through the application of the mental health diagnosis area.

Keywords: Deep learning, Mental health diagnosis, CNN, DNN, RNN.

1. Introduction

Mental health disorders have emerged as significant public health concerns globally, impacting both individuals and healthcare systems. Traditional mental health diagnostics are excessively dependent on clinical interviews, self-reported symptoms, and standardized questionnaires, which can often be subjective, time-consuming, and inconsistent. For instance, individuals may underreport their symptoms due to stigma, or clinicians might interpret data differently based on experience. These limitations emphasize that mental assessment is in dire need of change to make it more innovative, scalable, and objective.

In recent years, Artificial Intelligence has shown promise in dealing with some of these challenges by automating mental health analysis and providing relatively more accurate predictions. However, integrating AI into mental health diagnostics carries its own challenges as well, particularly in data collection and analysis. Mental health data is often heterogeneous, including textual information

(social media posts, therapy transcripts), physiological signals, and neuroimaging data. This enormous diversity makes the consolidation of datasets into a unified format suitable for machine learning models become struggling. Additionally, data privacy and ethical considerations bring significant barriers to obtaining high-quality datasets.

While traditional machine learning models have been used to analyze mental health data, they are limited by their overreliance on structured data and non-negligible defects of the ineffective management of unstructured, multimodal inputs. Deep learning methods, on the other hand, are capable of analyzing large-scale, complex datasets and uncovering non-linear relationships within mental health indicators, which compensates for the flaws of traditional methods. In Figure 1, Natural Language Processing (NLP) has proven effective in analyzing sentimental data, such as social media texts, to detect early signs of depression or suicidal tendencies [1]. This shift from traditional approaches to deep learning offers new opportunities to identify mental health patterns, enhance diagnostic accuracy, and facilitate early intervention.

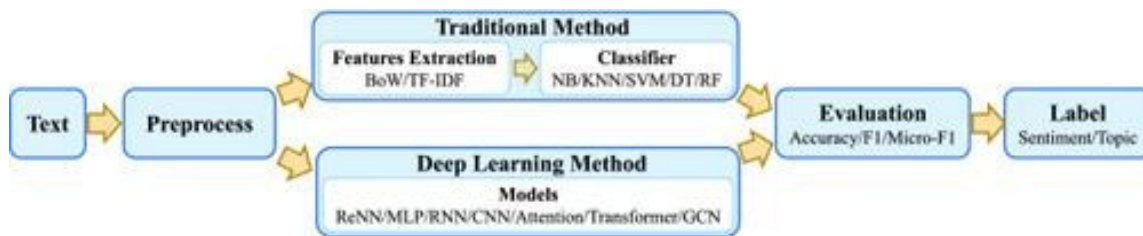


Figure 1: Flowchart of the text classification with classic methods in each module [1].

Given these opportunities and challenges, this paper aims to explore the application of deep learning algorithms in mental health prediction. Specifically, the research will address the following questions:

Out of all those advanced deep learning models, which one is the most suitable and effective one for various mental disorders diagnosis and predictions?

What are the key challenges in implementing AI for mental health diagnosis, particularly concerning data privacy, model interpretability, and performance scalability?

The remainder of this paper is structured as follows: Section 2 introduces the 3 main mental disorders our paper will cover with some proven deep learning models' unique effects on the assessment of mental health. On top of that, the presentation of the research results of each model's performance on each corresponding disease will be the emphasis of this part. The next part will be the discussion of challenges and future exploring, which covers the limitations we've faced so far and the potential areas for future innovation or improvement identified during the research and experimentation process.

This study contributes to the growing body of knowledge on AI in mental health by providing a comprehensive overview of deep learning applications, addressing critical challenges, and proposing future pathways. The findings aim to support researchers, clinicians, and policymakers in leveraging deep learning techniques to improve mental health diagnostics, foster early intervention strategies, and enhance patient outcomes.

2. Applications for deep learning in mental health diagnosis

Deep Learning, being a machine learning method, can be used to build models which learn high dimensional features from data, and has revolutionized data-driven problem-solving across numerous domains, including healthcare. Utilizing neural network architectures inspired by the human brain, deep learning models are designed to recognize patterns and extract meaningful insights from complex datasets. In mental health, these models have demonstrated transformative potential in

diagnosing and understanding mental disorders. Techniques like convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers have been applied to analyze diverse data modalities, such as neuroimaging, electroencephalograms (EEGs), and textual data from patient records or social media. Figure 2 represents a summary of different MRI techniques and concerned DL methods.

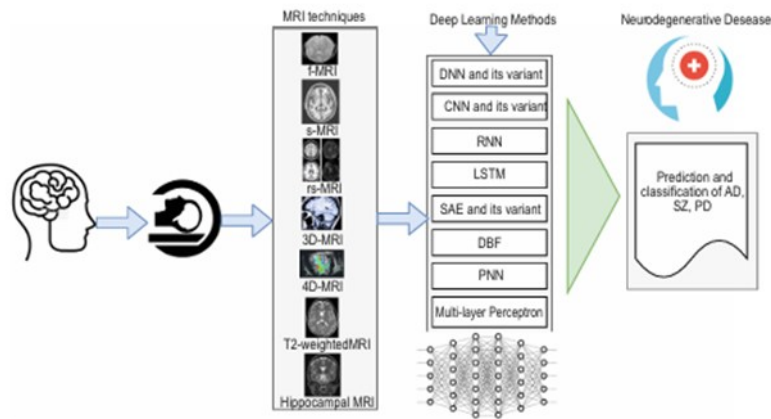


Figure 2: Applications of DL to classify and predict neurodegenerative disease based on different variants of MRI [2]

Recent advancements in mental health diagnosis using deep learning include automated detection of depression, anxiety, bipolar disorder, and schizophrenia. For example, CNNs have been employed to interpret neuroimaging data, identifying structural and functional anomalies in the brain with remarkable accuracy. RNNs and NLP models like BERT and CNN have been instrumental in analyzing textual data, uncovering linguistic patterns indicative of mental health conditions. Additionally, deep reinforcement learning is emerging as a tool for designing personalized therapeutic interventions.

2.1. Neuroimaging techniques in the diagnosis of neurodegenerative diseases

Neuroimaging data, such as magnetic resonance imaging (MRI) and positron emission tomography (PET) scans, provides precisely such a modality, offering a direct view into the structural and functional changes within the brain. Unlike behavioral data, which may only hint at the presence of neurological abnormalities, neuroimaging can pinpoint definitive markers of diseases like Alzheimer's or Parkinson's. For instance, while a behavioral anomaly detected by an autoencoder may signal cognitive decline, an MRI scan can confirm the presence of hippocampal atrophy, a hallmark of Alzheimer's disease. This integration of indirect behavioral insights with direct biological markers exemplifies the potential of multi-modal diagnostic frameworks.

The relationship between behavioral and neuroimaging data also opens avenues for targeted diagnostics. Patterns identified through autoencoder analysis can guide subsequent neuroimaging investigations, focusing attention on individuals or populations exhibiting high-risk features. For example, anomalies in activity patterns or linguistic expressions, flagged by behavioral data analysis, could prioritize individuals for neuroimaging studies, reducing the resource burden and enhancing the efficiency of diagnostic workflows. Furthermore, the latent features extracted from behavioral data can complement neuroimaging models. By integrating these features into deep learning architectures designed for imaging analysis, such as convolutional neural networks (CNNs), it is possible to build richer predictive models that capture both external manifestations and internal pathologies of neurodegenerative diseases. This layered approach not only improves diagnostic

accuracy but also facilitates a deeper understanding of the interplay between behavior and brain structure. Therefore, we delve into the application of deep learning in neuroimaging next, exploring how advanced architectures enhance the detection and characterization of neurodegenerative diseases, building on insights gained from behavioral data analysis.

2.1.1. Alzheimer's Disease

Alzheimer's disease (AD) is marked by progressive cognitive decline typically associated with aging, resulting from the deterioration of specific brain regions. Extensive research efforts have focused on identifying the underlying mechanisms of this neurodegeneration and developing automated methods for detecting degeneration patterns in neuroimaging data.

Deep CNNs have been widely utilized in Alzheimer's disease (AD) detection. For instance, the authors in [3] employed the LeNet CNN architecture to separate HC from AD cases. In contrast, Amoroso et al. took a purely machine learning-based approach, leveraging Random Forest for feature selection and using a deep neural network (DNN) for classification to facilitate early AD detection [4]. Ullah et al. proposed a CNN framework designed for detecting Alzheimer's disease (AD) and dementia from 3D MRI images, incorporating manual feature extraction [5]. Similarly, Dolph et al. introduced a pioneering model that utilized a Stacked Auto-Encoder (SAE) combined with a DNN-based multi-classifier. This model demonstrated the ability to learn intricate nonlinear atrophy patterns, effectively classifying AD, mild cognitive impairment (MCI), and cognitively normal (NC) individuals using datasets from both in-house sources and the standardized public-domain CADDementia framework [6].

2.1.2. Parkinson

Parkinson's disease (PD), a movement disorder, has also seen applications of deep learning (DL) for identifying its signs and detecting it from neuroimaging data. Kollias et al. developed a deep neural network combining CNNs for extracting detailed internal representations from input data with B-LSTM/GRU RNNs to analyze temporal progression, enabling final predictions [7]. Similarly, Shinde et al. introduced a fully automated CNN with a discriminative localization architecture to distinguish PD from healthy controls (HC) and to create prognostic and diagnostic biomarkers from neuromelanin-sensitive MRI [8].

Additionally, Kollia et al. applied a convolutional-RNN approach to predict PD by extracting latent variable information from trained deep neural networks, utilizing both MRI and DaT Scan data [9]. Esmaeilzadeh et al., on the other hand, employed a 3D Convolutional Neural Network with a voxel-based segmentation approach, leveraging data augmentation to expand the training set and classify PD and HC [10].

2.1.3. Schizophrenia

Schizophrenia (SZ), a significant psychiatric disorder linked to structural and functional brain abnormalities, progressively impairs cognition, emotion, and behavior. Recent advancements have focused on developing automated diagnostic techniques for SZ using deep learning (DL) and MRI data. Qureshi et al. introduced a 3D-CNN-based DL classifier, incorporating independent component analysis (ICA) to semi-automatically filter noise and artifacts from rs-fMRI data, enabling differentiation between SZ patients and healthy controls (HC) [11].

Many existing methods have utilized DNN-based classification techniques for SZ diagnosis [12]. Srinivasagopalan et al. demonstrated that DL could revolutionize SZ detection by comparing DNN classification performance with traditional machine learning methods such as support vector machines (SVM) and logistic regression [12]. Matsubara et al. proposed a deep generative model

(DGM) for rs-fMRI data, combining it with a DNN to calculate the contribution weight of various brain regions in SZ diagnosis using Bayes' rule [13].

In contrast, Han et al. proposed leveraging the advantages of backpropagation combined with feed-forward processing on rs-fMRI features to effectively differentiate individuals with schizophrenia (SZ) [14].

Furthermore, recent studies have explored auto-encoders for extracting functional connectivity (FC) features [12, 14, 15]. These features are subsequently employed in SVM classifiers [16] or DNN classifiers for automatic SZ diagnosis [12, 14]. Table 1 provides a comprehensive summary of the deep learning (DL) applications discussed, highlighting key details such as the type of MRI utilized, the datasets employed, and the reported classification accuracy.

2.2. Deep Learning-Based Study of AD, SZ and PD Prediction and Classification

2.2.1. Dataset for Studies of AD, SZ and PD

The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset has been widely used in studies on Alzheimer's disease and mild cognitive impairment (MCI). It provides demographic information, raw neuroimaging scans, APOE genotype data, CSF measurements, neuropsychological test results, and diagnostic information. Similarly, the OASIS dataset, which offers longitudinal neuroimaging, clinical, cognitive, and biomarker data, has been applied to studies focusing on normal aging and Alzheimer's disease.

In the study of Parkinson disease, the NTUA Parkinson dataset is used which is composed of MRT, DatScans of 55 patients suffering from Parkinson and 23 subjects with Parkinson related syndromes. There are also some other studies which have used Parkinson's Progression Markers Initiative (PPMI) public domain database in order to detecting the biomarkers of Parkinson progression.

For schizophrenia research, the Center for Biomedical Research Excellence dataset is frequently used. It includes data from 72 subjects aged 18–65 across healthy control and SZ groups. The OpenfMRI database has also been utilized, providing information on 50 SZ patients, 49 bipolar disorder patients, and 122 HC. Additionally, the function biomedical informatics research network (fBIRN) dataset is often referenced, containing data from 135 SZ patients (including those with schizoaffective disorder) and 169 HC. Furthermore, several studies have collected data from multiple sites or hospitals to ensure the robustness and validation of their proposed models.

Table 1: Summary of DL based studies for prediction and classification of AD, SZ and PD

Reference	Data Type	Dataset	DL technique	Performance
Alzheimer				
[3]	fMRI	ADNI	CNN	96.86%
[4]	MRI	ADNI	DNN	38.8%
[5]	3DMRI	OASIS	CNN	80.25%
[6]	sMRI	ADNI	SAE-DNN	(51% ~ 58%)
Parkinson				
[7]	MRI, DATScan	NTUA	CNN-RNN	98%
[8]	sMRI	NIMHANS	CNN	85.7%
[9]	sMRI, DaTScan	NTUA	CNN-RNN	98%
[10]	MRI	PPMI	3D-CNN	100%
Schizophrenia				
[11]	rs-fMRI	COBRE	3D-CNN	98.09%

Table 1: (continued).

[12]	fMRI, sMRI	COBRE	DNN	94.4%
[13]	rs-fMRI	OpenfMRI	DNN	76.6%
[14]	rs- fMRI	Hospital	DNN	79.3%
[15]	rs-fMRI	COBRE	SAE-DNN	
[16]	fMRI	COBRE	SAE	92%
[17]	fc-MRI	Multisite, COBRE	DDAE	(81 ~ 85)%

2.2.2. Performance Analysis

The studies referenced in this paper encompass various aspects of research, including Alzheimer's disease (AD) prediction, conversion from mild cognitive impairment (MCI) to AD, and multiclass AD classification. Notably, the authors in [4] achieved the highest accuracy of approximately 100% using a 3D-CNN for distinguishing AD from healthy controls (HC). In contrast, an accuracy of 38.8% was reported in a scientific challenge, where the model placed third among 19 participating teams in classifying AD, which is relatively lower compared to most other studies [10].

Using a 3D-CNN, the study in [10] achieved 100% accuracy on both validation and test sets for Parkinson's disease (PD) diagnosis. Similarly, research in [8] reported an accuracy of 85.7% in differentiating PD from typical parkinsonian syndromes.

In the case of schizophrenia (SZ) detection, the highest accuracy of 98.09% was observed in [11], which utilized 3D-CNN-based classification. Studies in [12] and [16] also reported accuracies exceeding 90%. Meanwhile, other studies demonstrated accuracy ranges between 70% and 85%.

3. Challenge and limitations

The impressive performance of deep learning (DL) models in neuroimaging-based disease diagnosis has ushered in a new era of healthcare solutions. However, alongside these advancements, data privacy and security concerns emerge as critical challenges that must be addressed to sustain trust and ensure the ethical deployment of such technologies. With the increasing reliance on sensitive patient data, including MRI scans, genetic profiles, and clinical records, protecting this information against breaches and misuse becomes paramount.

One of the primary challenges revolves around the inherent vulnerability of centralized data storage systems, which are often targeted by cyberattacks. For instance, a well-documented case involved a breach at a prominent health research institution where hackers accessed patient data stored in an unsecured cloud repository [18]. Such incidents highlight the risk of entrusting large datasets to centralized platforms without robust security measures. These breaches not only compromise patient confidentiality but also expose institutions to legal and financial repercussions, threatening the broader adoption of DL technologies in clinical settings.

Another pressing issue is the ethical management of patient consent in data collection and sharing. Many DL applications rely on datasets aggregated from multiple hospitals and research facilities, raising questions about whether patients have provided informed consent for their data to be used in secondary analyses. A recent case in Europe revealed that a hospital shared anonymized MRI data with a research lab, only for the lab to inadvertently re-identify some patients due to advanced data linkage techniques [19]. This incident underscores the need for stringent policies to govern data anonymization and ensure compliance with privacy laws such as the General Data Protection Regulation (GDPR).

Future perspectives in addressing these challenges emphasize integrating federated learning and blockchain technologies. Federated learning allows DL models to be trained across decentralized datasets without transferring data to a central repository. This method ensures data privacy while still enabling collaborative research across institutions. Additionally, blockchain technology can enhance security by providing a tamper-proof ledger for tracking data access and usage. By implementing these technologies, researchers can mitigate the risks associated with centralized storage and unauthorized data access.

Another promising avenue lies in advancing privacy-preserving techniques, such as differential privacy and homomorphic encryption. Differential privacy introduces statistical noise into datasets, ensuring that individual patient data cannot be extracted from the aggregated results. Homomorphic encryption, on the other hand, allows computations to be performed directly on encrypted data without decrypting it, preserving confidentiality throughout the analysis process. These methods, though computationally intensive, hold significant potential for ensuring the privacy of sensitive medical data.

Moreover, fostering a culture of ethical data stewardship is essential. Healthcare institutions and researchers must prioritize transparency in their data handling practices, informing patients about how their data will be used and ensuring compliance with legal frameworks. Establishing independent oversight committees to review and approve data-sharing agreements can further enhance accountability.

4. Conclusion

In conclusion, this paper has explored the transformative potential of deep learning algorithms in diagnosing mental health disorders, including Alzheimer's disease, Parkinson's disease, and schizophrenia, leveraging diverse neuroimaging techniques and multimodal data analysis. By highlighting the success of CNNs, RNNs, and advanced architectures in capturing intricate patterns in complex datasets, the study underscores the role of AI in revolutionizing mental health diagnostics and early intervention strategies. However, the challenges of data heterogeneity, model interpretability, and privacy concerns remain significant barriers to full-scale clinical adoption. Addressing these issues through federated learning, privacy-preserving techniques, and ethical data stewardship is essential to build trust and ensure sustainable implementation.

Future research should focus on improving the scalability of models, integrating real-time diagnostic capabilities, and addressing the ethical complexities of data usage in healthcare. By fostering collaboration among researchers, clinicians, and policymakers, the integration of deep learning into mental health diagnostics can significantly enhance early detection, personalize treatment pathways, and improve patient outcomes. This study contributes to the growing body of knowledge by offering a comprehensive analysis of current applications, identifying critical challenges, and proposing future directions for leveraging AI in mental health care.

References

- [1] Li, Qian and Peng, Hao and Li, Jianxin and Xia, Congying and Yang, Renyu and Sun, Lichao and Yu, Philip S. and He, Lifang. (2022) *A Survey on Text Classification: From Traditional to Deep Learning*, 2157-6904
- [2] Noor, M.B.T., Zenia, N.Z., Kaiser, M.S., Mahmud, M., Al Mamun, S. (2019). *Detecting Neurodegenerative Disease from MRI: A Brief Review on a Deep Learning Perspective*. In: Liang, P., Goel, V., Shan, C. (eds) *Brain Informatics. BI 2019. Lecture Notes in Computer Science()*, vol 11976. Springer, Cham.
- [3] Sarraf, S., Tofighi, G.: *Classification of Alzheimer's disease using fMRI data and deep learning CNNs*. *CoRR abs/1603.08631* (2016)
- [4] Amoroso, N., et al.: *Deep learning reveals Alzheimer's disease onset in MCI subjects: results from an international challenge*. *J. Neurosci. Methods* 302, 3–9 (2018)

- [5] Ullah, H.M.T., et al.: Alzheimer's disease and dementia detection from 3D brain MRI data using deep CNNs. In: *Proceedings of the I2CT 2018*, pp. 1–3 (2018)
- [6] Dolph, C.V., Alam, M., Shboul, Z., Samad, M.D., Iftikharuddin, K.M.: Deep learning of texture and structural features for multiclass Alzheimer's disease classification. In: *Proceedings of the IJCNN 2017*, pp. 2259–2266 (2017)
- [7] Kollias, D., et al.: Deep neural architectures for prediction in healthcare. *Complex Intell. Syst.* 4(2), 119–131 (2018)
- [8] Shinde, S., et al.: Predictive markers for Parkinson's disease using deep neural nets on neuromelanin sensitive MRI. *NeuroImage: Clin.* 22, 101748 (2019)
- [9] Kollia, I., Stafylopatis, A., Kollias, S.D.: Predicting Parkinson's disease using latent information extracted from deep neural networks. *CoRR abs/1901.07822* (2019)
- [10] Esmaeilzadeh, S., Yang, Y., Adeli, E.: End-to-end Parkinson disease diagnosis using brain MR-images by 3D-CNN. *CoRR abs/1806.05233* (2018)
- [11] Qureshi, M.N.I., Oh, J., Lee, B.: 3D-CNN based discrimination of schizophrenia using resting-state fMRI. *Artif. Intell. Med.* 98, 10–17 (2019)
- [12] Srinivasagopalan, S., et al.: A deep learning approach for diagnosing schizophrenic patients. *J.Exp.Theoret.Artif. Intell.* 31, 1–14 (2019)
- [13] Matsubara, T., et al.: Deep neural generative model of functional MRI images for psychiatric disorder diagnosis. *IEEE Trans. Biomed. Eng.* 66(10), 2768–79 (2019)
- [14] Han, S., et al.: Recognition of early-onset schizophrenia using deep-learning method. *Appl. Inform.* 4(1), 16 (2017)
- [15] Kim, J., et al.: Deep NN with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *NeuroImage* 124, 127–146 (2015)
- [16] Patel, P., Aggarwal, P., Gupta, A.: Classification of schizophrenia versus normal subjects using deep learning. In: *Proceedings of the ICVGIP, India*, pp. 281–286 (2016)
- [17] Zeng, L.L., et al.: Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI. *EBioMedicine* 30, 74–85 (2018)
- [18] Smith, J., & Brown, R. (2020). "Cybersecurity breaches in health research institutions: Challenges and lessons." *Journal of Health Informatics*, 35(3), 234–246.
- [19] Johnson, A., et al. (2021). "The risks of re-identification in anonymized medical datasets." *European Data Protection Review*, 14(2), 112–124.