Analysis of machine learning and deep learning models for early Alzheimer's disease diagnosis

Vanshika Rathi

Abstract

Machine learning (ML) models and deep learning (DL) algorithms have become widely utilized tools in the healthcare sector. They are helpful in identifying patterns, analyzing imaging data, and utilizing various biomarkers to identify between a healthy and unhealthy individual. This technology contributes to confirming early diagnoses, particularly for Alzheimer's disease, and mild cognitive impairments (MCIs). Alzheimer's disease is a neurological disorder that causes memory loss and inhibits the individual's ability to function independently. This paper compares the ML models used to classify the disease, differentiating between how often they are used and their accuracy. The ML models include support vector machines (SVMs), random forest (RF), and logistic regression, while the DL algorithms utilized include convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Through the comparison, we find that the most effective ML models are SVMs due to their ability to handle large datasets and provide precise image classification, whereas the most effective DL models are CNNs, because of their multiple variations in structures and highest accuracy rates in analyzing patterns in medical images through in-depth analysis for early AD diagnosis.

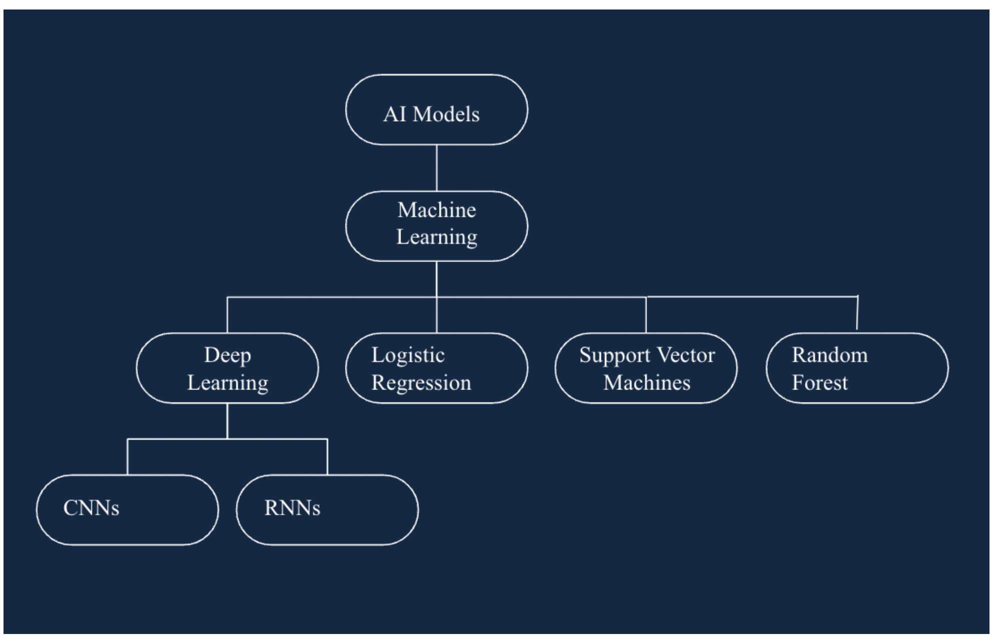
Introduction

Alzheimer's disease (AD) is a brain disorder that slowly impairs brain functioning. It initially starts by destroying memory and thinking skills but can ultimately lead to the inability to do even the simplest of tasks. By disrupting communication between neurons, also known as synaptic dysfunction, it can lead to cell death. The root of AD begins in the hippocampus, with evidence suggesting it is the region of the brain responsible for creating memories. It is the most common cause of dementia – the loss of cognitive functioning – amongst adults, and typically affects adults over the age of 65, but can begin as early as the 40s and 50s ((V. Rajak, A. Rajak, A. K. Shrivastava. Diagnosis of Alzheimer disease using machine learning approaches. *International Journal of Advanced Science and Technology*, (2019).)).

Often considered as the precursor to AD, Mild Cognitive Impairment (MCI) is a cognitive condition that is characterized by a decline in motor skills, such as a loss of balance, slower response times, and fine motor skills, where even buttoning a shirt can be difficult. Around 10-15% of patients with MCI develop AD, so identifying it earlier can allow for proper management of the disease's progression ((V. Rajak, A. Rajak, A. K. Shrivastava. Diagnosis of Alzheimer disease using machine learning approaches. *International Journal of Advanced Science and Technology*, (2019).)). Early prediction of AD has several benefits, as it allows for medical intervention with aid, clinical trials, and therapies for treatment. It can be determined through physical and neurological examinations, where patients are asked to perform tasks regarding eye movement, muscle strength, etc. Patients are also asked to fill out a Mini-Mental State Examination (MMSE) that evaluates memory and language, all of which help ascertain the level of cognitive impairment. Recently, machine learning (ML) and deep learning (DL) models have come to light as beneficial tools for recognizing symptoms and diagnosing early AD, because of good results with their performance and accuracy.

There are many key differences between ML and DL, but the most significant one would be that DL models are a subset of ML and use neural networks with multiple layers. They can automatically transform the data into a format they understand, while ML models typically need to be fed pre-set data (Figure 1.). ML uses smaller datasets with structured data, where the data is organized into a particular format (charts, tables, etc.), depending on the model ((M. M. Ahsan, S. A. Luna, Z. Siddique. Machine-learning-based disease diagnosis: A comprehensive review. *MDPI*, (2022).)). ML also has shorter training times, as ML models learn to recognize things more effectively if they have examples and more exposure to them. To do so, we can collect and label data, and then run and adjust the model multiple times so that it can execute its tasks better. Through shorter training times, ML models can adapt to a patient's history and be applied and are good at recognizing patterns in data and predicting treatment outcomes. On the other hand, DL is more complex, requiring larger datasets with unstructured data, which allows it to be particularly useful when analyzing images. It requires a lot more time to train due to its use of neural networks, but it can model more detailed patterns, which is favorable as AD is a complex disease. This study will examine three ML models commonly used in diagnosing early AD: logistic regression, random forest, and support vector machines. The other two models will be DL models, known as convolutional neural networks and recurrent neural networks, as both algorithms are advantageous at predicting disease progression and remain the most used models in healthcare.

AI Model Flowchart



Vanshika Rathi, ***Figure 1.*** Displays the category that each model belongs to and how they're all connected.

Results:

Logistic Regression

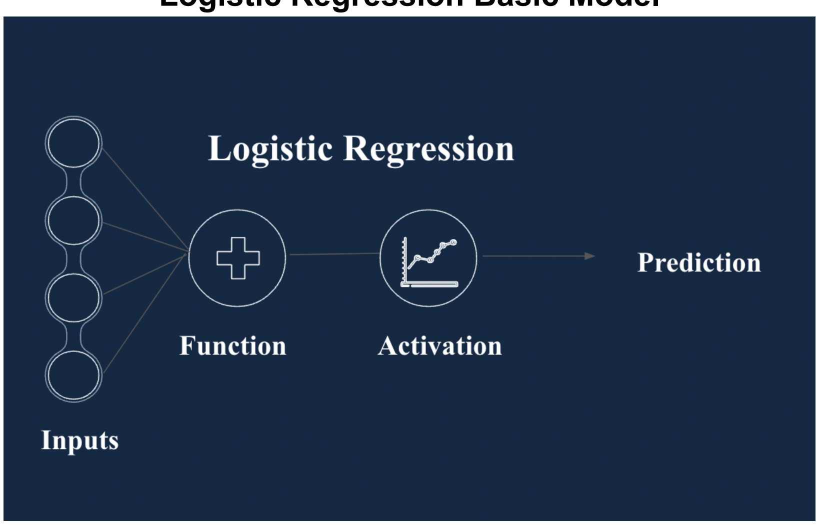
Logistic regression (LR) falls into the ML category, specifically in the supervised machine learning category. This means that both the input and output data are available, and it's the model's responsibility to distinguish a pattern between the two. LR is mainly used for binary outcomes, such as if a disease is present or not. After the model is trained, it uses data from the patient's history and records them as inputs into the model to get a probability score with projected values ranging from 0 to 1, and finally determine whether it falls in the positive or the negative category (Figure 2.) ((M. G. Alsubaie, S. Luo, K. Shaukat. Alzheimer's disease detection using deep learning on neuroimaging: A systematic review. *MDPI*, (2024).)).

LR was introduced to the healthcare sector in the 1970s because of its ability to find relationships and make accurate risk predictions with clinical data and was implemented for early AD identification research in the early 2000s. It evolved from using simple data such as age, demographics, and gender to more complex, multi-layered data and being integrated with other machine and deep learning models.

One study conducted in 2020 used the Oasis Longitudinal dataset, consisting of 373 patient records with features to classify AD, and discovered that LR demonstrates an accuracy of 78.95% ((A. D. Arya, S. S. Verma, P. Chakarabarti, T. Chakrabarti, A. A. Elngar, A.-M. Kamali, M. Nami. A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease. *Brain Informatics*, (2023).)). Jie Kuang et al. (Jie Kuang 2020) used 425 privately owned cases, with the goal of using LR to predict the transition from MCI to AD and found an accuracy of 89.52%. LR's accuracy varies depending on the forms of data, how organized it is, and the number of variables. LR is a straightforward model, but it holds a lot of potential to be developed considering its growth.

A limitation with the LR model is it assumes that the relationship between the input and output variables is linear, which does not always hold true and can lead to misinterpretations with the results. Therefore, it needs to be combined with other models to make up for the inconsistencies, which can be inconvenient. Overall, LR is a widely popular tool in early AD predictions as it's good for fast classification, with good accuracy and reliability and a lot of prospects for the future.

Logistic Regression Basic Model



Vanshika Rathi, ***Figure 2.*** There can be any number of inputs, the function places the values in a summation format, and the activation is the regression algorithm being applied. In simpler terms, it's the mathematical gate between the input and the prediction.

Support Vector Machines

Support vector machines (SVMs) are ML models and are mainly used for image classification. SVMs work by designating the best possible boundaries between two classes, such as the healthy versus the affected patients, and aim to maximize the distance between them. Simply put, that boundary helps SVMs make future predictions based on patient data, and the likelihood of them getting AD in the future. SVMs first gained traction in the late 1990s when the models were trained on distinguishing MCI in patients and predicting their likelihood of developing AD in the future. Researchers realized SVMs' potential in classification and began training them with MRI and PET scans, leading to their use in early AD diagnoses ((S. Sharma, A. Sharma. A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer disease using MRI scans. *Measurement: Sensors*, (2022).)).

J. Neelaveni et. al (J. Neelaveni 2020) conducted a study measuring the accuracy of patients with AD using an SVM model and achieved an accuracy of 85%. L.E. Collij et. al (L.E. Collij 2016) also used SVM for the detection of single-subject AD and MCI, reaching an accuracy of 82%. A study using the Oasis Longitudinal dataset used SVMs to classify AD and found an accuracy of 81.58%, all of which demonstrates the efficiency and consistency of SVMs ((A. D. Arya, S. S. Verma, P. Chakarabarti, T. Chakrabarti, A. A. Elngar, A.-M. Kamali, M. Nami. A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease. *Brain Informatics*, (2023).)).

SVMs tend to be computationally intensive, so with large datasets, the training time for SVMs increases. That leads to them requiring a lot of memory to hold the datasets while the model is in training, making them less suitable for processes that need fast turnaround. Furthermore, SVMs are fairly sensitive to outliers and noisy data, which makes it difficult to generalize the data and can lead to inaccurate predictions.

Random Forest

Another ML model, Random Forest (RF), is made up of multiple decision trees, which are supervised ML algorithms that represent a tree-like structure. They typically display multiple choices and their outcomes. RF models break the data into subsets, and each subset is trained using a decision tree model to make the best possible prediction, which then comes together to make a final decision.

RF models rose in popularity a bit later in comparison to other AI models and were applied in diagnosing AD in the early 2010s for simple classification. Researchers began using RF models for more clinical applications when they realized their effectiveness with multimodal data types, from genetic data to medical images in more recent years, developing more complex models.

A.R Vidushi et. al administered a study for the diagnosis of AD using machine learning approaches and the Oasis Longitudinal dataset, and utilized RF models, finding an accuracy of 84.21%. Emina Alickovic et. al (Emina Alickovic 2019) used RF models for the automatic detection of AD and found an accuracy of 85.77%.

Overall, RF models are fairly dependable, though they have their own set of challenges. RF models have shown bias towards the majority class – the data that shows up the most – and it falls into the hands of the people managing the model to manually balance the dataset, so the model doesn't give an inaccurate prediction. Furthermore, though this issue is not that common, RF models do tend to overfit the data, especially when there are outliers within the data. However, despite the limitations, RF models make powerful tools in the healthcare sector and diagnosing early AD. There are two more powerful tools that we will examine now, which are the DL models, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

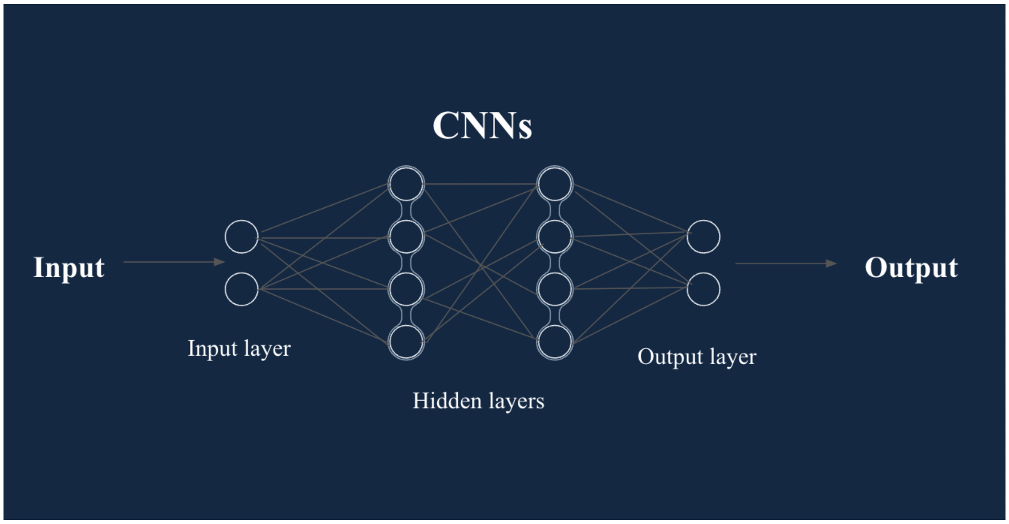
Convolutional Neural Networks

Convolutional neural networks (CNNs) fall in the DL category and are made to interpret visual data, so they're beneficial in interpreting medical images and in the case of AD, MRI and fMRI scans. They were introduced in the 1980s, initially to detect abnormalities in x-rays, but in the 2010s researchers began utilizing CNNs for neuroimaging analysis as their studies consistently produced reliable results ((G. Battineni, G. G. Sagaro, N. Chinatalapudi, F. Amenta. Applications of machine learning predictive models in the chronic disease diagnosis. *MDPI*, (2020).)).

In 2018, a landmark study by researchers from UCSF and UC Berkeley analyzed longitudinal MRI data using CNNs and demonstrated its accuracy when it identified patterns indicating early AD, even before the presence of clinical symptoms. This study showcased the link between CNNs and medical imaging, providing a foundation for future research.

CNNs have multiple architectures, but they all comprise of three parts, the input, hidden, and output layers, and the number of hidden layers depends on how complex the model is (Figure 3.). Some common but effective models include ResNet101, VGG16, and VoxCNN. VGG16 is a pre-trained CNN computer vision model with 13 hidden layers that's proven to be highly effective in early AD detection ((M. G. Alsubaie, S. Luo, K. Shaukat. Alzheimer's disease detection using deep learning on neuroimaging: A systematic review. *MDPI*, (2024).)).

CNN Basic Architecture



*Vanshika Rathi, **Figure 3.*** The input (typically an image) is added to the input layer, where it's analyzed for patterns through the hidden layers and output layer, before it comes to a prediction in the output.

For MRI data, a study conducted in 2019 achieved 95.73% accuracy, and another study done in 2022 in Berlin, Germany that was used to classify stages of AD achieved 89% accuracy ((V. Rajak, A. Rajak, A. K. Shrivastava. Diagnosis of Alzheimer disease using machine learning approaches. *International Journal of Advanced Science and Technology*, (2019).)). In terms of fMRI data and converting the data to extract specific features for early AD diagnosis, a study was tested using various parameters, and achieved over 99% accuracy, proving it to be highly reliable ((A. D. Arya, S. S. Verma, P. Chakarabarti, T. Chakrabarti, A. A. Elngar, A.-M. Kamali, M. Nami. A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease. *Brain Informatics*, (2023).)). Another study used 11 papers with fMRI data to identify the accuracy of the various CNN architectures listed above and found it to have an approximate accuracy of 92.81% ((M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. A. Mamun, M. Mahmud. Application of deep learning in detecting neurological disorders from magnetic resonance images: A survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia. *Brain Informatics*, (2020).)). Overall, there was not a significant difference in the accuracy between the models that were trained on MRI versus fMRI data.

However, a limitation with CNN models is that they require large sets of data for effective training and accurate results, and AD data are often limited in size ((A. Rahman, T. Debnath, D. Kundu, M. S. I. Khan, A. A. Aishi, S. Sazzad, M. Sayduzzaman, S. S. Band. Machine learning and deep learning-based approach in smart healthcare: Recent advances, applications, challenges and opportunities. *AIMS Public Health*, (2024).)). Furthermore, large data sets provide a lot of unwanted variability and outliers, and the models cannot generalize that data, making it difficult to interpret the data. Overall, CNNs are powerful tools in early AD detection and have seen significant progression, with an abundance of potential in the future as well.

Recurrent Neural Networks

Recurrent neural networks (RNNs) also fall in the DL category and are typically used for sequential data. This is mainly because unlike other neural networks, which assume the inputs to be independent of each other, RNNs maintain memory, which is helpful for time-series data. They were first introduced in the 1980s and held a lot of potential but didn't quite produce satisfactory results in the beginning, as older RNN data would get lost over time. However, with the help of Long Short Term-Memory (LSTM) networks, they became widely used in healthcare applications to model genetic sequences and predict patient health over a long period. In terms of AD, RNNs can predict the future health of the patient, along with the likelihood of patients with MCI getting AD based on the clinical records of patients at various time stamps.

Kai Lin et al (Kai Lin 2024) conducted a study in 2022 and used an RNN-LSTM framework to predict AD progression by measuring temporal dependencies such as cognitive tests and biomarker measurements with 563 fMRI images from 174 subjects with AD and reported an accuracy of 88.24%. Another study from 2023 combined two papers that utilized RNNs with sequential data from MRI and PET scans and found an average accuracy of 86.10%, overall demonstrating their steady applicability ((A. D. Arya, S. S. Verma, P. Chakarabarti, T. Chakrabarti, A. A. Elngar, A.-M. Kamali, M. Nami. A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease. *Brain Informatics*, (2023).)).

While RNN is effective in long-term predictions, a challenge is that it cannot do so with multiple types of data, and integrating multiple forms of data (images, lab results, records, and more) remains a limitation. Furthermore, RNNs are known as black-box models because RNN data is recorded in sequences, so it requires a lot of tracing back ((A. Rahman, T. Debnath, D. Kundu, M. S. I. Khan, A. A. Aishi, S. Sazzad, M. Sayduzzaman, S. S. Band. Machine learning and deep learning-based approach in smart healthcare: Recent advances, applications, challenges and opportunities. *AIMS Public Health*, (2024).)). Furthermore, since its memory updates per step, understanding how the model evolves is challenging. Since AD has a lot of complex data with many factors, interpreting RNN data is difficult. Ultimately, RNNs are a valuable DL tool in diagnosing early AD and should continue to be implemented.

Discussion:

ML and DL models have been more widely utilized in healthcare, specifically early AD diagnosis, due to their accurate predictions and assistance in making decisions. This study demonstrates how technology contributes to confirming early diagnoses, particularly for early AD diagnoses, though a meta-analysis of literature may be needed to increase confidence in the results. However, ML and DL models have only grown more advanced due to their ability to handle large volumes of different data types, from medical imaging such as MRI, fMRI, and PET scans to analyzing genetic data to detect trends and patterns.

CNNs in particular have shown the highest rates of accuracy and are also the most widely used. They have the most variations in structure and can analyze medical images effectively, which is suitable for early AD diagnosis.

SVMs are also convenient for image classification. While CNNs typically require large datasets to run, SVMs are beneficial when a quick diagnosis is needed, as once the model's trained, it can make predictions very quickly, and when the imaging data is smaller in size.

LR is one of the simpler models, as it only predicts binary outcomes, but is important in determining if a patient has early AD and classifying them as positive or negative. Their simplicity also allows them to be combined with other models, with LR commonly being used as one of the base models. Due to the binary nature of LR models, it narrows down the dataset significantly and allows other models to make more specific and customized predictions based on the patient.

RF is a more complex ML model as it comes up with many potential outcomes, depending on the number of subsets in the data, but is valuable as the more options it explores, the more likely it is to come up with the best possible prediction according to the patient's records.

RNNs' ability to store large amounts of data in their memory comes into play when monitoring patients over a longer period, as their records and data would all be sequential. For example, when monitoring patients who have been slowly developing symptoms, RNNs would be able to record and analyze the symptoms and find temporal patterns, leading to an accurate prediction for treatment outcomes.

ML and DL models all require a significant amount of training time, can be computationally intensive, and still have inaccuracies; however, they've grown exponentially and have demonstrated their potential as valuable tools in diagnosing early AD. The most effective ML models were SVMs as they can handle large datasets and provide definite image classification. Furthermore, the most effective DL models were CNNs, because of their various architectures and consistency with bringing forth the highest accuracy rates. In terms of future direction, ML and DL models can be utilized to analyze various forms of data, where it could even detect anomalies/early AD biomarkers through blood samples or be integrated into wearable technology so patients could be remotely monitored as well. ML and DL models are incredibly promising resources to enhance diagnosing and treatment options in the field of healthcare, and future studies should utilize and improve upon these technologies to make a profound impact in the early diagnosis of AD.

Methods

I used multiple databases, such as PubMed, IEEE Xplore, and Google Scholar, and conducted a systematic search for pertinent papers. Relevant keywords that I utilized included "Deep Learning," "Machine Learning," "Alzheimer's Disease," and "Early Diagnosis". The inclusion criteria required studies that addressed and focused on the application of deep learning or machine learning techniques for the early identification of Alzheimer's disease and were peer-reviewed and published within the last ten years.

Acknowledgments

I am extremely thankful to my mentor for the guidance, advice, and support throughout the writing process, and would also like to extend my sincere thanks to the writing editors. I'm very grateful to my family for their unwavering encouragement during the research project.

References:

1. V. Rajak, A. Rajak, A. K. Shrivastava. Diagnosis of Alzheimer disease using machine learning approaches. *International Journal of Advanced Science and Technology*, (2019). https://www.researchgate.net/profile/Akash-Rajak/publication/342847161_Diagnosis_of_Alzheimer_Disease_using_Machine_Learning_Approaches/links/5f0b37c6a6fdcc4ca463942e/Diagnosis-of-Alzheimer-Disease-using-Machine-Learning-Approaches.pdf

2. M. M. Ahsan, S. A. Luna, Z. Siddique. Machine-learning-based disease diagnosis: A comprehensive review. *MDPI*, (2022). https://www.mdpi.com/2227-9032/10/3/541

3. M. G. Alsubaie, S. Luo, K. Shaukat. Alzheimer's disease detection using deep learning on neuroimaging: A systematic review. *MDPI*, (2024). https://www.mdpi.com/2504-4990/6/1/24#:~:text=Deep%20learning%20models%20have%20emerged,from%20 large%2d Scale%20 imaging%20 datasets

4. A. D. Arya, S. S. Verma, P. Chakarabarti, T. Chakrabarti, A. A. Elngar, A.-M. Kamali, M. Nami. A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease. *Brain Informatics*, (2023). https://link.springer.com/article/10.1186/s40708-023-00195-7#Bib1

5. S. Sharma, A. Sharma. A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer disease using MRI scans. *Measurement: Sensors*, (2022). https://www.sciencedirect.com/science/article/pii/S2665917422001404

6. G. Battineni, G. G. Sagaro, N. Chinatalapudi, F. Amenta. Applications of machine learning predictive models in the chronic disease diagnosis. *MDPI*, (2020). https://www.mdpi.com/2075-4426/10/2/21

7. M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. A. Mamun, M. Mahmud. Application of deep learning in detecting neurological disorders from magnetic resonance images: survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia. *Brain Informatics*, (2020). https://braininformatics.springeropen.com/articles/10.1186/s40708-020-00112-2#Sec29

8. A. Rahman, T. Debnath, D. Kundu, M. S. I. Khan, A. A. Aishi, S. Sazzad, M. Sayduzzaman, S. S. Band. Machine learning and deep learning-based approach in smart healthcare: Recent advances, applications, challenges and opportunities. *AIMS Public Health*, (2024). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11007421/