



# Deep Learning Models To Detect Brain Disorders: A Review

<sup>1</sup>Sankalp Shukla, <sup>2</sup>Swapnendu Let

Student

Master Of Computer Applications

School of CSET, Bennett University (The Times Group), Gr. Noida, Uttar Pradesh, India

**Abstract:** Neurological and behavioral symptoms arising from medical conditions can significantly impact daily life due to their influence on brain structure and function. Early and accurate detection of these conditions is crucial, prompting significant interest in the application of computer-aided algorithms for diagnosis. Recent advancements in artificial intelligence (AI), particularly deep learning, offer novel solutions. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are being explored for their potential to analyze medical imaging data, such as Magnetic Resonance Imaging (MRI) scans, in the identification of these conditions. Deep learning algorithms demonstrate promising results with high accuracy. The accuracy rates for certain architectures, such as 18-layer CNN models, have been reported to exceed 98% in studies utilizing the powerful computer systems equipped with CUDA-enabled GPUs and large data sets of MRI scans from initiatives like Alzheimer's Disease Neuroimaging Initiative (ADNI). These results indicate that deep learning-based methods have the potential to improve and expedite medical diagnoses, paving the way for advancements in patient care.

**Index Terms** – Deep Learning, Machine Learning, ADNI, MRI, Alzheimer's, CNN, Review

## I. INTRODUCTION

As we age, our bodies undergo a universal biological process that leads to a decline in physiological function across different organ systems, including the central nervous system (CNS). This decline in cognitive abilities often speeds up after the age of 60. While some degree of age-related cognitive decline is normal, a portion of the population experiences a more pronounced decline that can develop into neurological disorders. These disorders, such as Alzheimer's disease (AD), ataxia, Bell's palsy, and brain tumors, can be highly debilitating and even life-threatening. A recent study published in "Lancet Neurology" estimated that nearly 47 million people globally, primarily in mid-to-late life, are affected by such neurological disorders, significantly impacting their daily function and healthcare resource utilization.

Given the vast spectrum of neurological disorders, this study focuses on AD and dementia due to their prevalence and significant healthcare burden. AD is a progressive neurodegenerative disorder characterized by the accumulation of amyloid plaques and neurofibrillary tangles in the brain, leading to cognitive decline and functional impairment. Early diagnosis of AD is crucial, as it allows for earlier intervention and potentially better patient outcomes. The primary pathological hallmark of AD is the abnormal accumulation of proteins, forming "plaque-tau tangles" composed of neurofibrillary tangles within the brain. While this protein buildup is the main driver of AD and dementia, its presence allows for medical intervention through advanced neuroimaging techniques. Early diagnosis relies on identifying the characteristic structural changes in the brain associated with AD progression. Mild cognitive impairment (MCI) is a pre-symptomatic stage of AD, and identifying individuals with MCI at an early stage is critical, as it offers the best window for intervention and potentially delaying progression to a more severe stage of AD. Research suggests that MCI may or may not progress to full-blown AD within 1.5 years of initial symptoms. Structural magnetic resonance imaging

(sMRI), functional magnetic resonance imaging (fMRI), and fluorodeoxyglucose positron emission tomography (FDG-PET) are key imaging techniques used to detect functional alterations, individual metabolism, and neuronal connectivity related to AD [12]. These advanced imaging techniques provide valuable data for clinicians in diagnosing AD and monitoring disease progression.

Medical research has established a predictable pattern of progression for AD. This pattern, utilizing a patient's electronic health record (EHR) data, which comprises lab tests, vital signs, medications, and treatment history, is a valuable and effective method for assessing the accuracy of the individual's risk for AD dementia. The widespread adoption of EHRs and readily available computing power have paved the way for utilizing artificial intelligence (AI), including machine learning (ML) and deep learning (DL) algorithms, for analyzing vast amounts of data. This approach helps reduce human error in identifying patterns and improves early detection accuracy. This study will explore the application of AI-based techniques, particularly ML and DL algorithms, for early AD detection and prediction of disease progression. We will examine the use of predictive modeling techniques in ML, such as decision trees, support vector machines (SVM), and k-means clustering, which leverage statistical data for risk assessment. Additionally, we will investigate the role of DL in leveraging high-computational resources for analyzing medical images. Techniques like convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep neural networks (DNNs) will be explored in the context of AD diagnosis. Our study will conduct a comprehensive analysis of relevant research that utilizes the techniques. The information will be sourced through extensive searches on reputable databases such as PubMed and Google Scholar to ensure the most current and accurate data is obtained. We believe that this research will provide valuable insights and benefit those interested in deep learning techniques for early AD diagnosis and progression prediction.

## II. RELATED WORKS

### 2.1 Past Studies

Neurological disorders like Alzheimer's Disease (AD) pose significant healthcare challenges due to their progressive nature and increasing prevalence. Early diagnosis is crucial for timely intervention and improved patient outcomes. This review investigates the application of advanced machine learning (ML) and deep learning (DL) techniques for the early detection of neurological disorders, with a particular focus on AD. Recent advancements in ML and DL offer promising avenues for developing robust models for early AD detection. We explore "state-of-the-art" approaches that leverage diverse datasets to address the challenges associated with this task (refer to Table 2 for a summary of these challenges).

The effectiveness of ML and DL models heavily relies on the quality and type of data used for training and validation. This review examines studies published within the past decade, focusing on the variety of data sources employed in these studies (detailed in Table 3). These sources can be categorized as Publicly available open datasets: The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset is a widely used resource due to its comprehensive range of data modalities, including neuroimaging data and clinical information. Custom datasets: Some studies create customized datasets through image augmentation techniques to expand the available data and improve model generalizability. Patient-driven datasets are an important source of real-world data that significantly enhance the clinical relevance and generalizability of the model. These datasets are collated from various healthcare institutions and contain a wealth of information that can be used to improve patient outcomes. By analyzing patient data in real-world settings, researchers can gain a better understanding of the effectiveness of treatments and interventions, as well as identify new patterns and trends in patient care. As a result, patient-derived datasets play a crucial role in advancing healthcare research and improving patient care. Convolutional Neural Networks (CNNs) have emerged as a prominent technique for analyzing medical images due to their ability to extract relevant features automatically. One recent study by Mehmood et al. (2021) [1] employed a multi-layered CNN model with custom convolutions on ADNI data. This approach aimed to enhance the prediction accuracy of AD. The authors trained their model on a moderate-sized dataset of 300 samples divided into four classes: AD, normal control (NC), early mild cognitive impairment (EMCI), and late mild cognitive impairment (LMCI). The model achieved an accuracy score of approximately 80%, demonstrating multi-layered CNNs' potential for AD diagnosis. Other studies, such as those by Odusami et al. (2021) [24] and Venugopalan et al. (2021) [33], explored variations of CNN architectures like 18-layered CNNs and 3D CNNs on comparable datasets. While these studies introduced architectural novelties, a detailed evaluation of their performance and limitations is required to assess their effectiveness. This review will delve deeper into such comparisons and highlight areas for further research [32].

Table 1: Reference studies

S. no.	Reference	Year	Study	Evaluation	datasets	Issues/ Challenges	Future Perspective
1	[35]	2020	Zhang et al.	✓	✓	✓	✓
2	[30]	2020	Segato et al.	✓	×	✓	×
3	[23]	2020	Noor et al.	✓	✓	✓	✓
4	[19]	2022	Lima et al.	✓	✓	✓	✓
5	[2]	2020	Altinkaya et al.	✓	✓	×	×
6	[5]	2023	Arya et al.	✓	✓	✓	✓
7	[34]	2023	Warren et al.	✓	✓	✓	✓
7	This Paper	2024	Sankalp et at.	✓	✓	✓	✓

Table 2: AD identification approaches studied

Feature	Measures/Factors
Categories	
Affecting Clinical Notes summary	Discharge
Laboratory tests. Common demographics education, race	CSF, vitals, medications, medical history, other lab age, gender,
Neuroimaging	Genetic
Cognitive assessments	APOE $\epsilon$ 4, family history
	MRI (structural, functional, unspecified), PET (FDG, amyloid)
	MMSE, ADAS-Cog, others (CDR, FAQ)

## 2.2 Extraction of features for Visible Anomalies

This study investigates the potential of hippocampal atrophy as a biomarker for early AD diagnosis. According to recent research, the combination of magnetic resonance imaging (MRI) scans and powerful deep learning algorithms shows promising results in identifying the shrinkage in the hippocampus as well as distinguishing accurately between various stages of Alzheimer's Disease (AD). This advanced imaging technique has significantly impacted the early detection and motoring of AD, which can lead to more effective treatment and management of the disease [14]. Furthermore, Hill et al. (2014) suggest MRI as the only currently qualified biomarker for diagnosing AD and other neurological disorders [10].

When examining neurological disorders, researchers often focus on the hippocampi using various imaging modalities like functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and standard MRI (sMRI) scans. These techniques have revealed that, in addition to dementia, reduced fluorodeoxyglucose (F-FDG) uptake in the hippocampi and cortex is another indicator of AD [21][22][31].

### III. BACKGROUND

#### 3.1 Understanding Alzheimer's Disease and Neurological Disorder

Alzheimer's disease (AD) represents a significant global health challenge, with current estimates exceeding 47 million cases worldwide. This neurodegenerative disorder profoundly impacts the lives of patients and their families, causing substantial stress and disruption to daily routines. While the high prevalence of AD is concerning, attributing it solely to medical negligence is inaccurate. Early and accurate diagnosis can be challenging due to the often subtle and non-specific nature of AD symptoms. These can include memory lapses, diminished focus, and difficulties with language – as highlighted by Akash Shah et al. (2020) [29]. Unlike physical ailments, these symptoms may be initially misinterpreted as a natural consequence of aging. AD is the most common form of dementia, accounting for 60-80% of dementia cases globally in older adults [12]. Other dementia subtypes, such as vascular dementia and Lewy body dementia, contribute a smaller but still significant portion, with hundreds of thousands of individuals affected worldwide.

Table 3: Past studies referred to in this study

Research Studies	Model	Data Sets
Atif Mehmood and colleagues (2021)	The CNN of multi-layers	Data sets from ADNI
Modupe Odusami and Colleagues (2021)	18-layered CNN	Data sets from ADNI
Janani Venugopalan and Colleagues (2021)	3D CNN	Data sets from ADNI
Amnaya Pradhan and Colleagues (2021)	DenseNet169, VGG19	Kaggle Data set
Akash Shah and colleagues (2020)	SVM Classifier based on voting Xboost Random forest classifier	Yearwise collected dataset
Huanhuan Ji and colleagues (2019)	ConvNet	Data sets from ADNI
Firouzeh Razavi and colleagues (2019)	Softmax Regression	Data sets from ADNI
Jyoti Islam and colleagues (2018)	Deep CNN	Data sets from ADNI
Priyanka Thakare and colleagues (2016)	SVM	Smt. Kashibai Navale Medical Hospital Pune

Diagnosing the specific type of dementia can be challenging, as many symptoms overlap with the general effects of aging. Memory loss, impaired thinking and reasoning, language difficulties, poor judgment, behavioral changes, and personality alterations are all commonly observed. The lack of readily distinguishable symptoms between various dementia types underscores the importance of advanced computational algorithms in modern diagnostic approaches. Researchers and medical professionals are increasingly turning to these tools for improved early detection and differentiation of dementia subtypes.

Unfortunately, the prognosis for AD is currently limited. The disease is progressive, meaning symptoms worsen over time. While there is no definitive cure to date, ongoing research efforts are actively exploring potential treatment options.

#### 3.2 Image and Dataset

This research leverages data collected by the Alzheimer's Disease Neuroimaging Initiative (ADNI) a public-private partnership established in 2003. Led by principal investigator Michael W. Weiner, MD, ADNI's primary objective was to assess the utility of serial magnetic resonance imaging (sMRI), positron emission tomography (PET), and other biomarkers, along with neuropsychological assessments, in studying the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD) [14].

The ADNI database offers a rich resource for researchers investigating AD. Studies have primarily focused on T1-weighted MRI (T1-MRI) and fluorodeoxyglucose (FDG)-PET images due to their distinct properties and the extensive collection available on the ADNI platform. Detailed information regarding data

acquisition protocols can be found in Jack et al. (2008) [14]. Data segmentation refers to the process of classifying images into distinct labels. In the context of AD research, segmented datasets often categorize images into four primary classes: normal control (NC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and AD. The number of samples per class can vary significantly, ranging from approximately 300, as observed in Mehmood et al. (2021) [1], to over 6,000 as reported in Pradhan et al. (2021) [28] (refer to Table 1 for details).

Multi-atlas label propagation with expectation-maximization (MALP-EM) is a frequently employed framework for automated segmentation of brain MRI images. This approach leverages label fusion from multiple atlases, along with label refinement techniques that incorporate the expectation-maximization (EM) algorithm to achieve optimal results. Studies referenced in Table 1, including Ledig et al. (2015) [6], demonstrate the effectiveness of the MALP-EM framework in achieving desired segmentation accuracy. Notably, within the MCI class, further sub-segmentation into EMCI and LMCI categories is often performed based on DSM-5 criteria, enhancing the dataset's utility for detection and prediction tasks.

Table 4: Visualizing dataset

Research Studies	Model	Total samples	NC	EMCI	LMCI	AD
Atif Mehmood and colleagues (2021)	The CNN of multi-layers	300	85	70	70	75
Modupe Odusami and Colleagues (2021)	18-layered CNN	138	25	25	25	25
Janani Venugopalan and Colleagues (2021)	3D CNN	503	132	104		266
Amnaya Pradhan and Colleagues (2021)	DenseNet169, VGG19	6000				
Akash Shah and colleagues (2020)	SVM Classifier based Xboost Random forest classifier	437	72			14
Huanhuan Ji and colleagues (2019)	ConvNet	615	182	254		179
Firouzeh Razavi and colleagues (2019)	Softmax Regression	99	52	43	56	51
Jyoti Islam and colleagues (2018)	Deep CNN	416	73	75	62	33
Priyanka Thakare and colleagues (2016)	SVM	988	400	661		169

## IV. DEEP LEARNING METHODOLOGIES IN STUDY

### 4.1 Data Pre-Processing

Pre-processing of medical images is a crucial initial step in machine learning (ML) and deep learning (DL) pipelines for computer-aided diagnosis (CAD) of Alzheimer's disease (AD). Tools like Roboflow and zxttools streamline this process. A common pre-processing technique involves image resampling to a resolution compatible with the chosen model. Square-shaped samples with sizes ranging from 221x221 to 416x416 pixels are frequently used. Padding of 1-2 pixels is applied to maximize feature extraction during the convolution process, often employing a 3x3 kernel filter [7]. For paired datasets containing both MRI and PET scans, rigid registration aligns each PET image to its corresponding MR image within the ADNI database [37].

Selecting a region of interest (ROI) is crucial for feeding relevant information to the models. The hippocampus is a particularly important ROI due to its well-established role in AD diagnosis [15]. However, computational limitations may necessitate image cropping to reduce image size. Recent studies have explored using a larger 5x5 kernel size to compensate for this downscaling. In datasets with pre-segmented ROIs, such as the hippocampus region, the ROI coordinates are readily available. For paired MRI-PET datasets, a random

sample of the initial MR image serves as a template for affine registration [13]. This process calculates center points across all MR images, generating an affine matrix for each image to define the central region. This approach fosters robust dataset creation suitable for various CAD algorithms.

Finally, for PET images that have been rigidly registered to their corresponding MR counterparts for hippocampal analysis, isotropic resampling is performed to achieve a uniform spacing of 1mm.

## 4.2 Methodologies

Convolutional neural networks (CNNs) are a powerful type of deep learning architecture inspired by the structure and function of the visual cortex [17]. These feedforward neural networks consist of multiple layers of interconnected artificial neurons, enabling them to excel at processing large-scale images. Traditional machine learning (ML) approaches often rely on handcrafted feature extraction techniques for medical image analysis. In contrast, CNNs leverage convolutional layers to automatically learn relevant features directly from the image data. This automated feature learning capability is a significant advantage of CNNs compared to traditional ML methods [34][4][3].

The main component of a CNN is the convolutional layer. Convolutional filters, also called kernels, are used on the input image to extract localized features by performing element wise multiplication and summation. These filters are shared across different parts of the image, which helps in reducing computational complexity and promoting parameter efficiency [25]. Pooling layers further downsize the feature maps while preserving essential information, leading to a more robust representation of the input data. Fully connected layers at the end of the network integrate the extracted features and perform classification or regression tasks. Compared to other feed-forward neural networks, CNNs require a smaller number of connections and parameters due to weight sharing within convolutional layers. This characteristic makes them computationally more efficient and well-suited for large-scale image recognition tasks. Additionally, CNNs benefit from faster training times due to the reduced number of parameters and efficient processing of image data. These advantages have cemented their position as a cornerstone of modern deep-learning architectures.

## 4.3 Evolution Metrics and Feature Selection Techniques

The performance of machine learning models is typically evaluated using various metrics. Accuracy, a common metric, reflects the proportion of correctly classified instances. Libraries like sci-kit-learn provide tools for calculating such metrics and understanding model performance [27]. Our investigation explored various ML algorithms, including Support Vector Machines (SVM), XGBoost, and others listed in Table 3. These algorithms were carefully chosen based on their established performance in medical image classification tasks. To ensure a robust evaluation, a sequential forward feature selection process was implemented using libraries like scikit-learn and extending. This approach iteratively selects features that contribute most significantly to model performance [9]. Additionally, 10-fold cross-validation was employed to mitigate overfitting and provide a more reliable estimate of model generalizability [16].

The selection of the “random state” parameter is crucial during model training. This parameter controls the random seed used for splitting data into training and validation sets. Setting the random state to 0 ensures consistent image selection across training runs, facilitating model reproducibility [28]. While the random state is typically set to 0 for consistency in traditional machine learning, CNNs often benefit from using a random seed during training to introduce variability and potentially improve model generalizability.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (1)$$

Artificial Neural Networks (ANNs) are a class of machine learning models inspired by the structure and function of the human brain. Similar to the biological brain, ANNs consist of interconnected layers of artificial neurons that process and analyze information. These networks typically comprise input, hidden, and output layers, and their learning algorithms enable them to improve performance iteratively. Gorji et al. (2022) [8] proposed a novel approach using pseudo-Zernike moments (PZMs) for differentiating Alzheimer’s disease (AD) from control subjects (CN) within mild cognitive impairment (MCI) datasets. This method leverages ANNs to achieve accurate classification.

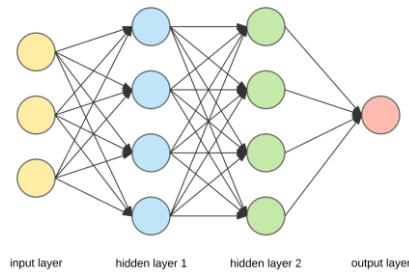


Figure 1: artificial neural network architecture

Table 5: ML techniques and performance

Machine Learning Methods	Accuracy Obtained (%)
Voting Classifier	87.45
Decision Tree Classifier	82.86
SVM – support vector machine	82.87
XgBoost	86.02

Pattern recognition neural networks (PRNNs) often employ a feed-forward architecture for information processing. Additionally, learning vector quantization (LVQ) is a prominent technique for data classification tasks in AD research [25]. Studies have reported promising results using these approaches, with accuracies exceeding 98% achievable with deeper architectures like 18-layered CNNs [26]. A significant advantage of deep learning models lies in their ability to learn hierarchical feature representations. Lower layers within the network typically capture low-level image features, while higher layers progressively extract more complex and abstract representations [36]. This hierarchical learning capability is particularly valuable for analyzing medical images, where subtle features can hold significant diagnostic potential.

Stacked autoencoders (SAEs) represent another type of deep learning architecture that can be employed for feature extraction in neuroimaging analysis. SAEs are known for their ability to learn latent representations of data, uncovering hidden patterns within the original neuroimaging and biological features [11]. The dataset utilized in our research originated from the ADNI database. It comprised baseline magnetic resonance imaging (MRI) scans, along with 18F-fluorodeoxyglucose (FDG) PET images, providing valuable biological information.

#### 4.4 Recurrent Neural Network for AD

Recurrent neural networks (RNNs) have emerged as a promising tool for modeling the progressive nature of Alzheimer's disease (AD). Unlike traditional feed-forward networks, RNNs can capture temporal dependencies within data sequences, making them well-suited for analyzing longitudinal medical records. Li et al. (2019) proposed a prognostic model for early-stage MCI and AD prediction. Their approach utilized a Long Short-Term Memory (LSTM) autoencoder to generate deep MRI image representations [20]. These representations, along with hand-crafted features such as hippocampal volume and demographic data, were fed into a regression model ("Cog regression model") to predict clinical scores for early AD detection. This method offers a valuable tool for clinicians by automating the prediction process and reducing human effort.

Fouladvand et al. (2008) also demonstrated the effectiveness of LSTMs for analyzing MRI data in the context of AD diagnosis [33]. However, their study highlighted the limitations associated with incomplete or unprocessed data, emphasizing the importance of robust data pre-processing pipelines. Building upon these prior works, Lee et al. (2019) introduced the concept of a Multimodal Recurrent Neural Network (MRNN) for predicting MCI-to-AD conversion [18]. MRNNs represent a variant of RNNs specifically designed to handle multimodal data. This approach integrates information from diverse sources, including demographics,

neural imaging phenotypes, cognitive performance, and cerebrospinal fluid (CSF) measurements, all of which are considered relevant for AD progression modeling. The proposed MRNN architecture utilizes a Gated Recurrent Unit (GRU) to process the time-series data from each modality. The individual outputs are then concatenated and fed into a final layer for prediction. This approach leverages the inherent temporal information within the data alongside the rich variety of features provided by the multimodal approach. Consequently, the MRNN can potentially achieve more accurate predictions compared to models relying solely on individual data modalities.

**Table 6: Deep learning architectures and performance**

used)	CNN network (types Accuracy Obtained (%)
Convolutional neural network (CNN)	80
18-layered CNN	98
3D CNN	88

## V. STATISTICAL ANALYSIS AND EVALUATION

### 5.1 Performance Measures

The performance of machine learning (ML) and deep learning (DL) models is typically assessed using a variety of metrics. Three of the most commonly employed metrics include accuracy, precision, and recall. Before delving into the specific calculations of these metrics, it's crucial to establish a clear understanding of their underlying concepts.

Precision quantifies the proportion of true positive predictions among all predicted positive cases. Mathematically, it is defined as: (Equation 2 Here, TP (True Positives) represents the number of instances correctly classified as positive, and FP (False Positives) denotes the number of instances incorrectly classified as positive. A high precision value indicates that the model is effective at minimizing false positive predictions. In the context of medical diagnosis, high precision is particularly important as it reduces the likelihood of misclassifying healthy individuals as having a disease.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

Sensitivity, also known as recall, represents the ability to capture true positive predictions among all actual positive cases. Mathematically, it is defined as: (Equation 3) Here, TP (True Positives) represents the number of instances correctly classified as positive, and FN (False Negatives) denotes the number of instances incorrectly classified as negative (when they actually belong to the positive class). A high recall value indicates that the model is effective at identifying all true positive cases. In the context of medical diagnosis, high recall is crucial as it minimizes the risk of missing true disease cases. The trade-off between precision and recall is an essential consideration when evaluating model performance.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

While accuracy is a commonly used metric, it can be misleading in imbalanced datasets, where one class significantly outnumbers the others. Mathematically, it is defined as: (Equation 1) Here, TP (True Positives) represents the number of instances correctly classified as positive, TN (True Negatives) denotes the number of instances correctly classified as negative, FP (False Positives) refers to the number of instances incorrectly classified as positive, and FN (False Negatives) represents the number of instances incorrectly classified as negative. Accuracy reflects the overall proportion of correctly classified instances. However, it is essential to consider precision and recall alongside accuracy for a more comprehensive evaluation, especially in imbalanced classification tasks.

Table 7: Precision, Accuracy and Recall

Classifier	Accuracy (%)	Precision (%)	Recall (%)
SVM	95.96	94.70	97.78
K-NN	95.53	95.81	96.67
AdaBoos t	95.00	96.00	94.44
XgBoost	94.38	94.18	95.56

Scikit-learn and matplotlib are popular Python libraries that provide tools for generating informative visualizations of model performance metrics. One such visualization is the confusion matrix. The confusion matrix is a tabular representation that summarizes the performance of a classification model. It includes the counts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions for each class.

1. True Positive (TP): These represent instances correctly classified as the positive class (e.g., AD in this case).
2. False Positive (FP): These represent instances incorrectly classified as the positive class (e.g., non-AD samples mistakenly classified as AD).
3. True Negative (TN): These represent instances correctly classified as the negative class (e.g., non-AD samples correctly identified as such).
4. False Negative (FN): These represent instances incorrectly classified as the negative class (e.g., AD samples mistakenly classified as non-AD).

Analyzing the confusion matrix allows researchers to gain insights into various aspects of model performance, including:

**Overall accuracy:** This can be calculated as the sum of TP and TN divided by the total number of samples. **Specificity (True Negative Rate):** This is the proportion of true negative predictions among all actual negative cases ( $TN / (TN + FP)$ ). **Sensitivity (Recall):** This is the proportion of true positive predictions among all actual positive cases ( $TP / (TP + FN)$ ). Utilizing the confusion matrix in conjunction with other metrics like precision and recall provides a comprehensive understanding of model strengths and weaknesses, facilitating informed decision-making when evaluating model suitability for a specific classification task.

## 5.2 Accuracy, Recall and Precision

The receiver operating characteristic (ROC) curve is a graphical representation of a classification model's performance at all possible classification thresholds. It depicts the trade-off between sensitivity (recall) and specificity (true negative rate) for a binary classification task. ROC curves are constructed by plotting the true positive rate (TPR) on the y-axis against the false positive rate (FPR) on the x-axis. The TPR is equivalent to recall, while the FPR is calculated as  $1 - \text{specificity}$ .

Models with superior performance exhibit ROC curves that approach the upper left corner of the graph. This indicates a high true positive rate (correctly identifying positive cases) alongside a low false positive rate (minimizing misclassification of negative cases). The F1-score is a harmonic mean of precision and recall, providing a single metric to summarize a model's performance. Mathematically, it is defined as: (Equation 4). Here, precision and recall are calculated as described previously. The F1-score ranges between 0 and 1, with a value of 1 indicating perfect balance between precision and recall. This metric is particularly useful in imbalanced datasets, where accuracy alone can be misleading. A high F1-score signifies a model's effectiveness in correctly classifying both positive and negative cases.

$$f1 - score = \frac{2(TP)}{2(TP) + FP + FN} \quad (4)$$

Table 8: ML approach analysis; Acc. (%); Sensitivity (%); Specificity (%); Precision (%); Recall (%); F1-score (%)

ML Algorithm	Data Set	Acc.	Sensitivity	Specificity	Precision	Recall	F1-Score
SVM	FMRI	100	80	70			
K-means clustering	PET	76.56					
Logistic Regression (LR)	ADNI	98.12	90	95			
Random Forest (RF)	FMRI	98.42	85	95			
Logistic regression CV (LRCV)	MRI	86.1			87.5	91.3	89.4
Multinomial LR	FMRI	89					
AdaBoost	Over Sampled	91.16			92.32	92.22	
XgBoost	Over Sampled	91.6			91.42	93.33	
K-NN	Over Sampled	95.53			95.81	96.67	

### 5.3 Hyper-Parameter Optimization

Hyperparameters are critical parameters of a machine learning model that control its learning process and influence its final performance. Selecting optimal hyperparameter values is crucial for maximizing model accuracy and generalizability. Grid search cross-validation (GridSearchCV) is a widely used technique for hyperparameter optimization. This method systematically evaluates a predefined grid of hyperparameter values and selects the combination that yields the best performance on a validation dataset. Stratified 10-fold cross-validation can be employed within GridSearchCV to mitigate the impact of imbalanced class distributions.

SVMs are a common machine learning algorithm used for classification tasks. Hyperparameter tuning plays a significant role in optimizing SVM performance. An example grid search for an SVM classifier might involve exploring different values for three hyperparameters: cost, gamma, and kernel function. The cost (C) parameter plays a crucial role in balancing training error and model complexity. A higher cost value puts greater emphasis on reducing training errors, but it may also increase the risk of overfitting the model. Conversely, a lower cost value may result in underfitting. Common cost values to explore might range from 1 to 30. Gamma (Y): This parameter controls the influence of each training instance on the decision boundary. Higher gamma values correspond to stronger influence from closer training points, while lower gamma values lead to smoother decision boundaries. Exploring a range of gamma values such as [1, 0.1, 0.01, 0.001, 0.0001] can be beneficial. Kernel Function: The kernel function defines the similarity measure between data points. Common kernel functions for SVMs include Radial Basis Function (RBF), Sigmoid, and Linear kernels. Grid search can be used to evaluate the performance of the SVM with different kernel functions on data obtained from both original and oversampled datasets.

By systematically evaluating different hyperparameter combinations using GridSearchCV, researchers can identify the configuration that optimizes model performance for a specific dataset and task. This optimization process can significantly improve model accuracy, precision, and other performance metrics.

## VI. ISSUES, CHALLENGES AND FUTURE PERSPECTIVES

Even after having achieved high-performance measures, still the detection of AD using DL and ML leaves researchers with unsolved errors. Just like the CNN or convolutional neural network limiting the complete detection of AD in the early stages i.e. initiation of the symptoms in hippocampi. A major issue is faced at the time of detection in the brain MRIs of old aged patients as the multi-layered CNN becomes even more complex while extracting the required features from scans.

- Complexity of multi-layered CNN while extracting features
- Loss of data while feature extraction
- CNNs limitation to data loss
- The limited availability of labeled data for supervised architectures
- Adversarial noise in the scans makes it difficult to detect disease
- Image acquisition in a non-standardized manner makes the dataset misleading or unfit for usage

The studies mention the reusability of produced models for early detection of diseases as grave as dementia and Alzheimer's. Although the limitations of the datasets and related DL and ML models have a major impact on the present problem, overcoming similar issues is fairly possible in the future. The scope of DL models such as CNN, RNN, GNN, and AE supported with transfer learning methods fetched highly varying results in terms of performance creating the look-through hole for future perspective where this goal surely deems achievable. Suggestions for the researchers continuing the topic in the future include the usage of semi-supervised and unsupervised learning algorithms to extract the maximum benefits from both labeled and unlabeled datasets.

Table 9: DL approach analysis; Acc. (%); Sensitivity (%); Specificity (%); Precision (%); Recall (%); F1-score (%)

DL Model	Data Set	Acc.	Sensitivity	Specificity	Precision	Recall	F1-Score
DASE	ADNI	91.95	89.49	93.82			
DBM	OASIS	91.76	90.59	92.96			
2D CNN	OASIS +						
	Local data	90	86	94			
AlexNet	OASIS	92.85					
ADVIAN	Oasis +						
	Local data	98.25	97.65	97.86			
3D-DenseNets	ADNI	98.83					
ADGNET	KACD						
	(6400 subjects)		99.69				99.61
ResNet-50	ADNI	97.10			95.5	95.3	
Inception V4	OASIS	96.25					
DenseCNN2	ADNI	92.5	88.2	94.9			
3D CNN	ADNI +						
	MRIAD	91.4	92.32	92.42			

Table 10: Literature review summary for limitations in prior studies

Author	Technique	Limitations
Janghel and Rathore	SVM, K-means clustering	Imbalance dataset
Gao et al.	Novel 3DMgNet Architecture	Low model sensitivity
Memon et al.	LR, DT, SVM	Low sensitivity
Dinu and Manju	Random forest and tree bagger	Low sensitivity
Salehi et al.	CNN	Imbalance dataset
Eke et al.	Support vector machine	Improvement in acc.
Leong and Abdullah	DNN, RF, GBM, LR	Low sensitivity
Wang et al.	CNN	High computational cost
Liu et al.	Linear SVC, LRCV, bagging, MLP	Acc. needs improvement
Revathi et al.	RF, multinomial LR	Low acc.
Goenka and Tiwari	3D-CNN	High computational cost

## VII. CONCLUSION

Diagnosis and early detection of neurological disorders such as Alzheimer's as majorly discussed in this article, brain tumors, etc has always been one of the most challenging tasks and will surely be deemed the same in the future as well, although computer engineers are constantly putting in hours to build the required machines to help patients and doctors by advancing the early detection process and reducing the scope for human error. For classification problems of identifying the stage and type of neurological disorder, CNN are most widely preferred and used model by researchers to provide strong outputs. Performance of known ML techniques including SVM, K-NN, AdaBoost, etc, confirmed the competitive performance with over 95% accuracy achieved, yet overfitting of the model is one major problem that requires addressing.

A noticeable scope for the future can be focusing on the major classification phase as so far in the aforementioned studies, the point of focus was majorly on feature extraction and formatting the datasets. With a crisp focus on the classification details, novel models can be brought to light with new insights into neurological disorder diagnosis and detection.

## VIII. REFERENCES

- [1] Ahed Abugabah, Atif Mehmood, Sultan Almotairi, and Ahmad AL Smadi. Health care intelligent system: A neural network based method for early diagnosis of alzheimer's disease using mri images. *Expert Systems*, 39(9):e13003, 2022.
- [2] Emre Altinkaya, Kemal Polat, and Burhan Barakli. Detection of alzheimer's disease and dementia states based on deep learning from mri images: a comprehensive review. *Journal of the Institute of Electronics and Computer*, 1(1):39–53, 2020.
- [3] Khalid Anwar, Aasim Zafar, and Arshad Iqbal. An efficient approach for improving the predictive accuracy of multi-criteria recommender system. *International Journal of Information Technology*, 16(2):809–816, 2024.
- [4] Khalid Anwar, Aasim Zafar, Arshad Iqbal, Shahab Saquib Sohail, Amir Hussain, Yeliz Karaca, Mohammad Hijji, Abdul Khader Jilani Saudagar, and Khan Muhammad. Artificial intelligence-driven approach to identify and recommend the winner in a tied event in sports surveillance. *Fractals*, 31(10), 2023.
- [5] Akhilesh Deep Arya, Sourabh Singh Verma, Prasun Chakarabarti, Tulika Chakarabarti, Ahmed A Elngar, Ali-Mohammad Kamali, and Mohammad Nami. A systematic review on machine learning and deep learning techniques in the effective diagnosis of alzheimer's disease. *Brain Informatics*, 10(1):17, 2023.
- [6] Esther E Bron, Marion Smits, Wiesje M Van Der Flier, Hugo Vrenken, Frederik Barkhof, Philip Scheltens, Janne M Papma, Rebecca ME Steketee, Carolina M'endez Orellana, Rozanna Meijboom, et al.

Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural mri: the caddementia challenge. *NeuroImage*, 111:562–579, 2015.

[7] Huanhuan Feng, Tingting Zheng, Mingyu Li, Junwei Wu, Hongjun Ji, Jiaheng Zhang, Weiwei Zhao, and Jinhong Guo. Droplet-based microfluidics systems in biomedical applications. *Electrophoresis*, 40(11):1580–1590, 2019.

[8] Maysam B Gorji, Alix de Pannemaecker, and Samuel Spevack. Machine learning predicts fretting and fatigue key mechanical properties. *International Journal of Mechanical Sciences*, 215:106949, 2022.

[9] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[10] James M Hill, Christian Clement, Aileen I Pogue, Surjyadipta Bhattacharjee, Yuhai Zhao, and Walter J Lukiw. Pathogenic microbes, the microbiome, and alzheimer's disease (ad). *Frontiers in aging neuroscience*, 6:97417, 2014.

[11] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[12] Xue Hua, Alex D Leow, Neelroop Parikshak, Suh Lee, Ming-Chang Chiang, Arthur W Toga, Clifford R Jack Jr, Michael W Weiner, Paul M Thompson, Alzheimer's Disease Neuroimaging Initiative, et al. Tensor-based morphometry as a neuroimaging biomarker for alzheimer's disease: an mri study of 676 ad, mci, and normal subjects. *Neuroimage*, 43(3):458–469, 2008.

[13] Yechong Huang, Jiahang Xu, Yuncheng Zhou, Tong Tong, Xiahai Zhuang, and Alzheimer's Disease Neuroimaging Initiative (ADNI). Diagnosis of alzheimer's disease via multi-modality 3d convolutional neural network. *Frontiers in neuroscience*, 13:509, 2019.

[14] Clifford R Jack Jr, Marilyn S Albert, David S Knopman, Guy M McKhann, Reisa A Sperling, Maria C Carrillo, Bill Thies, and Creighton H Phelps. Introduction to the recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):257–262, 2011.

[15] Huanhuan Ji, Zhenbing Liu, Wei Qi Yan, and Reinhard Klette. Early diagnosis of alzheimer's disease using deep learning. In *Proceedings of the 2nd international conference on control and computer vision*, pages 87–91, 2019.

[16] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada, 1995.

[17] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

[18] Garam Lee, Kwangsik Nho, Byungkon Kang, Kyung-Ah Sohn, and Dokyoon Kim. Predicting alzheimer's disease progression using multi-modal deep learning approach. *Scientific reports*, 9(1):1952, 2019.

[19] Aklima Akter Lima, M Firoz Mridha, Sujoy Chandra Das, Muhammad Mohsin Kabir, Md Rashedul Islam, and Yutaka Watanobe. A comprehensive survey on the detection, classification, and challenges of neurological disorders. *Biology*, 11(3):469, 2022.

[20] Jinping Liu, Lirong Chang, Yizhi Song, Hui Li, and Yan Wu. The role of nmda receptors in alzheimer's disease. *Frontiers in neuroscience*, 13:43, 2019.

[21] Lisa Mosconi, Sandro Sorbi, Mony J de Leon, Yi Li, Benedetta Nacmias, Paul S Myoung, Wai Tsui, Andrea Ginestroni, Valentina Bessi, Mozghan Fayyazz, et al. Hypometabolism exceeds atrophy in

presymptomatic early-onset familial alzheimer's disease. *Journal of Nuclear Medicine*, 47(11):1778–1786, 2006.

[22] Lisa Mosconi, Wai H Tsui, Karl Herholz, Alberto Pupi, Alexander Drzezga, Giovanni Lucignani, Eric M Reiman, Vjera Holthoff, Elke Kalbe, Sandro Sorbi, et al. Multicenter standardized 18f-fdg pet diagnosis of mild cognitive impairment, alzheimer's disease, and other dementias. *Journal of nuclear medicine*, 49(3):390–398, 2008.

[23] Manan Binth Taj Noor, Nusrat Zerine Zenia, M Shamim Kaiser, Shamim Al Mamun, and Mufti 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*, 7:1–21, 2020.

[24] Modupe Odusami, Rytis Maskeliūnas, and Robertas Damaševičius. Optimized convolutional fusion for multimodal neuroimaging in alzheimer's disease diagnosis: Enhancing data integration and feature extraction. *Journal of personalized medicine*, 13(10):1496, 2023.

[25] Sunday O Olatunji, Aisha Alansari, Heba Alkhorasani, Meelaf Alsubaii, Rasha Saklouna, Reem Alzahrani, Yasmeen Alsaleem, Reem Alassaf, Mehwash Farooqui, Mohammed Imran Basheer Ahmed, et al. Research article preemptive diagnosis of alzheimer's disease in the eastern province of saudi arabia using computational intelligence techniques. 2022.

[26] Vijeeta Patil, Manohar Madgi, and Ajmeera Kiran. Early prediction of alzheimer's disease using convolutional neural network: a review. *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, 58(1):130, 2022.

[27] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[28] Amnaya Pradhan, Jerin Gige, and M Eliazer. Detection of alzheimer's disease (ad) in mri images using deep learning. *Int. J. Eng. Res. Technol.*, 10(3):580–585, 2021.

[29] Joseph H Schwab and Akash A Shah. Spinal epidural abscess: diagnosis, management, and outcomes. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons*, 28(21):e929–e938, 2020.

[30] Alice Segato, Aldo Marzullo, Francesco Calimeri, and Elena De Momi. Artificial intelligence for brain diseases: A systematic review. *APL bioengineering*, 4(4), 2020.

[31] Reisa A Sperling, Clifford R Jack Jr, Sandra E Black, Matthew P Frosch, Steven M Greenberg, Bradley T Hyman, Philip Scheltens, Maria C Carrillo, William Thies, Martin M Bednar, et al. Amyloid-related imaging abnormalities in amyloid-modifying therapeutic trials: recommendations from the alzheimer's association research roundtable workgroup. *Alzheimer's & Dementia*, 7(4):367–385, 2011.

[32] Muhammad Tanveer, Bharat Richhariya, Riyaj Uddin Khan, Ashraf Haroon Rashid, Pritee Khanna, Mukesh Prasad, and Chin-Teng Lin. Machine learning techniques for the diagnosis of alzheimer's disease: A review. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 16(1s):1–35, 2020.

[33] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[34] Samuel L Warren and Ahmed A Moustafa. Functional magnetic resonance imaging, deep learning, and alzheimer's disease: A systematic review. *Journal of Neuroimaging*, 33(1):5–18, 2023.

- [35] Li Zhang, Mingliang Wang, Mingxia Liu, and Daoqiang Zhang. A survey on deep learning for neuroimaging-based brain disorder analysis. *Frontiers in neuroscience*, 14:560709, 2020.
- [36] Guoqiang Zhong, Li-Na Wang, Xiao Ling, and Junyu Dong. An overview on data representation learning: From traditional feature learning to recent deep learning. *The Journal of Finance and Data Science*, 2(4):265–278, 2016.
- [37] Xiaoying Zhuang, Xiaoyu Xiang, William Grizzle, Dongmei Sun, Shuangqin Zhang, Robert C Axtell, Songwen Ju, Jiangyao Mu, Lifeng Zhang, Lawrence Steinman, et al. Treatment of brain inflammatory diseases by delivering exosome encapsulated anti-inflammatory drugs from the nasal region to the brain. *Molecular Therapy*, 19(10):1769–1779, 2011.

