[ICAETBM-2024]
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Machine Learning-Based Prognosis of Early-Stage Alzheimer's Disease

Prof. Nitin Wankhade , Prof. Dheeraj Patil, Sofiyana Inamdar, Pritee Joshi, Amreen Parveen Khan, Chetan Patil
Nutan Maharashtra Institute of Engineering and Technology, Talegaon(D), Pune

ABSTRACT

Alzheimer's disease (AD) presents a growing threat to global public health as it gradually deteriorates neurological function. For effective intervention and care, it is crucial to promptly identify and accurately diagnose Alzheimer's disease. The paper thoroughly explores various techniques for detecting Alzheimer's disease (AD), focusing particularly on methods centered on medical imaging. We investigate a range of imaging modalities, data acquisition methods, feature extraction approaches, classification techniques, and recent advancements in AI-driven diagnostic systems. We emphasize the changing landscape of Alzheimer's disease detection, the challenges involved, and the potential directions for future research. The objective of this paper is to provide a valuable resource for researchers, clinicians, and policymakers involved in Alzheimer's disease detection.

Keyword: Alzheimer's, machine learning, AI, image, MRI

1. Introduction

AD is a degenerative neurological condition marked by memory decline and cognitive impairments, representing the predominant cause of dementia, encompassing 60-80% of dementia diagnoses. AD was first described in 1906 and recognized as a major cause of death after 70 years It is a costly disease, with an estimated projected cost of 47 billion dollars in 2018 Timely identification is essential as it can decelerate disease advancement and mitigate healthcare expenses [1]. Alzheimer's disease (AD) is an advancing and incapacitating neurodegenerative ailment predominantly impacting older adults, imposing a significant strain on worldwide healthcare infrastructures. The hallmark of AD is the gradual loss of cognitive abilities, ultimately leading to severe dementia. The prompt identification and precise detection of Alzheimer's disease are critically significant, allowing for timely interventions and tailored care strategies. With the advent of medical imaging and AI technologies, the landscape of AD detection has witnessed significant advancements. Structural MRI has been widely studied as a non-invasive method for detecting brain atrophy in MCI patients, particularly in the hippocampal and entorhinal regions [2]. In this survey, we embark on a journey to explore the state of the art in AD detection, focusing particularly on methods leveraging medical imaging data.

AD is an all-encompassing and persistent neurodegenerative condition marked by the gradual deterioration of cognitive functions, memory decline, and, ultimately, profound dementia. With the aging global population, AD has emerged as a pressing public health challenge, impacting millions of individuals and their families worldwide. The socio-economic and emotional burdens it imposes are substantial, compelling a concerted effort to enhance our understanding of the disease, improve its early detection, and provide better care and treatment. Observations of preclinical brain alterations linked to Alzheimer's can manifest years prior to clinical symptom emergence, facilitating early diagnosis and intervention [3].

The criticality of the matter arises from the recognition that interventions and therapies yield optimal efficacy when initiated during the disease's early phases. Mild Cognitive Impairment (MCI) represents a preclinical phase of AD.. This study investigates the

classification of MCI using multimodal data and co-training method [4]. Alzheimer's Disease (AD) is a degenerative neurological condition characterized by irreversible progression, primarily impacting individuals aged 65 and older.

Alzheimer's Disease (AD) is a degenerative neurological condition that progresses steadily and is primarily prevalent among individuals aged 65 years and older. Mild Cognitive Impairment (MCI) represents the preliminary stage of AD, with subcategories including progressive MCI (pMCI) and stable MCI (sMCI) [5].Therefore, early detection is not merely a matter of convenience but a critical determinant of the patient's quality of life and the potential success of therapeutic interventions.

In the quest for early detection and accurate diagnosis, medical imaging has emerged as a Developing a low-cost and easy-to-use AD detection tool [6]. It allows us to visualize and analyze the structural and functional changes occurring in the brain as the disease progresses. The power of imaging lies in its ability to reveal subtle alterations in brain anatomy, such as the atrophy of specific regions and the deposition of amyloid plaques and neurofibrillary tangles, which are pathological hallmarks of AD.

The field of AD detection has also witnessed a revolution fueled by artificial intelligence and machine learning. Machine learning models for early-stage Alzheimer's disease prediction - Importance of early detection and treatment for AD [7]. These advancements in technology have empowered the creation of advanced algorithms with the capability to analyze extensive sets of medical images. These algorithms can extract intricate patterns and features that might elude the human eye, offering a promising avenue for improving the accuracy and reliability of AD diagnosis.

In this comprehensive survey, we embark on a journey to explore the state of the art in AD detection, with a particular focus on methods leveraging medical imaging data. The paper focuses on accurate identification of AD using multi-modal data. - The paper suggests utilizing a relation-induced multi-modal shared representation learning method for diagnosing Alzheimer's disease [8]. Proposed machine learning techniques aim to assist in interpreting clinical data for diagnosis and decision-making. Nonetheless, prevailing methods often fail to replicate the personalized diagnostic procedures observed in actual clinical environments [9].Our goal is to provide an in-depth and up-to-date overview of the landscape of AD detection, its challenges, and the directions it is taking. Through an extensive literature review, we aim to provide researchers, clinicians, and policymakers in the field with a valuable resource that encapsulates the diversity of approaches, the current state of technology, and the future horizons of AD detection. We delve into the various imaging modalities, data acquisition and preprocessing techniques, feature extraction methodologies, and classification algorithms that researchers have employed. Alzheimer's disease is a chronic neurodegenerative condition requiring long-term prognosis of progression. Structural Magnetic Resonance Imaging (sMRI) is utilized to identify cortical atrophy in Alzheimer's disease as well as its early stages. Existing methods focus on predicting cognitive scores using morphological features derived from sMRI [10]. Furthermore, we highlight the recent advances in AI-based diagnostic systems, which offer great promise for the early diagnosis of AD. While this survey encapsulates the present knowledge and progress in AD detection, we must acknowledge that substantial challenges lie ahead. Data availability and diversity, the interpretability of AI models, and the generalizability of research findings are persistent obstacles that the field must address. Therefore, we conclude our introduction by emphasizing the importance of this survey, which not only summarizes the current state of the field but also sets the stage for the ongoing and vital work of improving the early detection and management of this formidable disease.

2. Related work

The pursuit of effective Alzheimer's disease detection methods has driven extensive research in the field of medical imaging and machine learning. A comprehensive review of related work reveals the diverse approaches and techniques that researchers have employed. Imaging modalities such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and Computed Tomography (CT) have been harnessed to capture structural and functional brain changes associated with AD. Data preprocessing methods, encompassing image registration, denoising, and normalization, play a crucial role in ensuring the quality of imaging data. Feature extraction, covering a spectrum from voxel-based measures to advanced texture and morphological features, contributes to the discrimination between AD and healthy

controls. Furthermore, an array of classification algorithms, including Support Vector Machines (SVM), Convolutional Neural Networks (CNNs), and ensemble methods, have been employed to distinguish between AD and non-AD subjects. Emerging AI-based diagnostic systems, leveraging deep learning and multimodal data fusion, exhibit promising results. The literature review provides valuable insights into the past and current state of AD detection, laying the foundation for future research directions.

A paper published by Xin Hong and team[1] of researchers focuses on predicting model based on Long Short-Term Memory (LSTM), a special kind of recurrent neural network, to predict the possible development of Alzheimer's Disease (AD) by encoding the temporal relation between features and the next stage of the disease. The model outperforms most existing models in terms of prediction accuracy. The authors consider the impact of time series data on prediction and use time step data obtained through a data preprocessing pipeline. They also evaluate the stability of their algorithm in different data sizes and its sensitivity to different features. The paper compares the efficiency of their algorithm with recent state-of-the-art algorithms and demonstrates its stability in different data sizes. The authors highlight the importance of capturing time-sensitive features for predicting the future stage of the disease, which their proposed method successfully achieves. The paper contributes to the field of digital health by addressing the need for early diagnosis of AD, which can significantly decrease the risk of further deterioration.

A paper published by Fusun Er and team [2] of researchers presents a deep learning-based computer-aided diagnosis (CAD) system to forecast the transition from MCI to AD using longitudinal non-invasive structural MRI data from baseline and a 12-month follow-up. Employing an array of methodologies such as an autoencoder, CNN, and SVM classifier, the system achieves an accuracy of 87.2% in distinguishing progressing MCI patients from stable ones. Notably, the CAD system sidesteps invasive methods or cognitive tests, offering a less burdensome prediction strategy. However, limitations include the exclusive focus on MCI patients without a healthy control group for comparison, reliance on longitudinal MRI data, potentially limiting broader clinical applicability, and the absence of comparative analysis with other existing prediction methods, leaving room for further performance assessment and improvement.

Asif Hassan Syed and a team of researchers, in their published paper [3], concentrate on a method that employs a unique blend of Cerebrospinal Fluid (CSF) protein biomarkers to enhance the accuracy of predicting earlier stages of Alzheimer's disease compared to current biomarkers. Two feature selection techniques, Recursive Feature Elimination (RFE) and L1 regularization, are employed to pinpoint the most crucial subset of features for constructing a classification model using the MCI dataset. The screening involves testing a new combination of three biomarkers (Cystatin C, Matrix metalloproteinases (MMP10), and tau protein) using linear Support Vector Machine (SVM) and Logistic Regression (LR) classifiers. An ensemble model is developed by combining the two top-performing classifiers (LR and Linear SVM) through a weighted average, leveraging the three most informative features. The effectiveness of the proposed ensemble model over LR and Linear SVM base classifiers is demonstrated through notable improvements in Receiver Operating Characteristic Curve (ROC_AUC) and Area under Precision-Recall values (AUPR).

A paper published by Shaoxun Yuan and team[4] of researchers focuses on investigating the potential of using both labelled and unlabelled samples from the ADNI cohort to classify MCI through the multimodal co-training method. Utilizing structural magnetic resonance imaging (sMRI) data and genotype data to build initial classifiers on labeled MCI samples, and implementing the co-training method to obtain new labeled samples from unlabeled MCI samples. Using the random forest algorithm to obtain a combined classifier for MCI classification in the independent ADNI-2 dataset. The presented framework attains an accuracy rate of 85.50% and an AUC value of 0.825 in classifying MCI, highlighting the substantial enhancement in MCI classification performance through the joint application of sMRI and SNP data via the co-training approach.

In their published paper [5], Chiyu Feng and a team of researchers concentrate on a pioneering deep learning framework designed for diagnosing Alzheimer's disease (AD) utilizing 3D-CNN and fully stacked bidirectional LSTM (FSBi-LSTM). The framework integrates the strengths of 3D-CNN and FSBi-LSTM to extract profound feature

representations from both MRI and PET data, thereby enhancing the effectiveness of AD diagnosis. The method proposed demonstrates high accuracies in distinguishing between AD and normal control, pMCI and NC, as well as sMCI and NC, surpassing comparable algorithms documented in existing literature. The framework addresses the challenge of limited availability of imaging data by effectively utilizing CNNs for AD diagnosis .The use of FSBi-LSTM helps to capture hidden spatial information from deep feature maps, further enhancing the performance of the framework .The method is validated on the AD neuroimaging initiative (ADNI) dataset, demonstrating its effectiveness in AD diagnosis .

A paper published by Javier Escudero and team[6] of researchers focuses on machine learning approach for personalized and cost-effective diagnosis of Alzheimer's disease (AD) using locally weighted learning to tailor a classifier model to each patient and compute the sequence of biomarkers most informative or cost-effective for diagnosis. Potential use of the approach to support personalized diagnosis processes and reduce the number or cost of biomarkers needed for diagnosis. Extension of the framework to other biomarkers and diseases. The selection of AD and HC subjects relies on the appropriate use of accuracy and AUC metrics. This study introduces the inaugural application of Kinect V.2 camera alongside machine learning techniques to offer a comprehensive assessment, including F-score, for classifying the quantitative analysis of the TUG test in detecting AD patients from HC. Illustrating the promise of this approach as a novel quantitative supplementary tool for identifying Alzheimer's disease among elderly individuals.

A paper published by C. Kavitha and team[7] of researchers focussed several techniques such as DT, Random Forest, SVM, Gradient Boosting, and Voting classifiers have been employed to identify the best parameters for Alzheimer's disease prediction .The paper also references recent work on the prediction of Alzheimer's disease, which includes the use of SVM, DT, NN, and Naive-Bayes models .Additionally, the paper mentions the use of feature selection methods such as Correlation coefficient, Information gain, and Chi-Square in the prediction of Alzheimer's disease .The paper explores the utilization of the Open Access Series of Imaging Studies (OASIS) dataset for predicting Alzheimer's disease. The paper highlights the importance of early diagnosis of Alzheimer's disease and the potential features, with performance improvement tending to decrease monotonically from the first to the last iterations. The research receives financial backing from the National Institute for Health Research (NIHR) and the Alzheimer's Disease Neuroimaging Initiative (ADNI). The approach is tested using data sourced from the ADNI database [7].

A paper published by Zhenyuan Ning and team[8] of researchers focussed on relation-induced multi-modal shared representation learning framework for Alzheimer's disease (AD) diagnosis. This integrated framework combines representation learning, dimension reduction, and classifier modeling into a cohesive system. By establishing a bidirectional mapping between the original space and a shared space, the framework derives multi-modal shared representations. It employs relational regularizers and auxiliary regularizers to facilitate the learning of underlying associations within multi-modal data and mitigate overfitting. The proposed method outperforms several state-of-the-art methods in terms of accuracy and performance, as demonstrated through extensive experiments on two independent datasets (ADNI-1 and ADNI-2) .

In their published paper [9], Yan Zhao and a team of researchers concentrate on forecasting cognitive scores at future time-points through the utilization of morphological features extracted from sMRI data .Very few works consider predicting an individual brain MRI image at future time-points . The framework presented in this paper includes a 3D multi-information generative adversarial network (mi-GAN) to forecast the future appearance of an individual's entire brain over a specified interval. Additionally, it incorporates a 3D DenseNet-based multi-class classification network, optimized using focal loss, to ascertain the clinical stage of the predicted brain state . The mi-GAN is capable of producing high-fidelity individual 3D brain MRI images by conditioning on both the individual's 3D brain sMRI and multi-information gathered at the baseline time-point. It exhibits outstanding performance, achieving a structural similarity index (SSIM) of 0.943 between real MRI images at the fourth year and those generated. Furthermore, when utilizing mi-GAN and focal loss, the accuracy in distinguishing between pMCI and sMCI experiences a notable enhancement of 6.04%

compared to conditional GAN and cross-entropy loss.

A paper published by Rongrong Li and team[10] of researchers focuses on improved method of measuring technology similarity in the medical field based on the subject-action-object (SAO) semantic structure. The SAO semantic structures are extracted and refined through the semantic network provided by the Unified Medical Language System (UMLS). The comparison of SAO semantic structures is assessed utilizing the Metathesaurus within the UMLS. The feature weights of the SAO semantic structure are introduced to represent the importance of the patentees' technology features .Each patentee's vector is constructed using the SAO and weight information to measure the technology complementarity between different patentees . The proposed method undergoes comparison with conventional approaches that assess technological similarity using IPC codes and keywords .The traditional method of measuring technology similarity based on keywords involves constructing vectors using keyword frequency and characterizing the similarity between patents using Euclidean distance between vectors . The TF-IDF weighting method is employed to determine the significance of words within documents.

In the below table, the field of Alzheimer's prediction has witnessed datasets and methodologies encompass a range of approaches, including machine learning algorithms and deep learning architectures, aimed at effectively analyzing longitudinal MRI data to predict the onset and progression of Alzheimer's Disease.

Table 1. Different datasets and techniques used for Alzheimer prediction system.

Sr.	Reference no.	Dataset	Techniques	Limitations
1.	[7]	OASIS data	Machine Learning techniques (Decision Tree, Random Forest, Support Vector Machine, Gradient Boosting, Voting classifiers) - Feature selection methods (Correlation coefficient, Information gain, Chi-Square)	• Paper lacks detailed information on ML model parameters & participant demographics. • Parameters/configurations for Decision Tree, Random Forest, SVM, Gradient Boosting, Voting classifiers not provided.
2.	[10]	ADNI (Alzheimer's Disease Neuroimaging Initiative) dataset, specifically ADNI-GO and ADNI-2	3D multi-information generative adversarial network (mi-GAN) - 3D DenseNet based multi-class classification network	• Current GAN models utilize 2D-MRI slices, neglecting 3D brain images. • Lack of consideration for non-imaging factors affecting Alzheimer's Disease (AD) progression.
3.	[8]	DNI-1 and ADNI-2	Relation-induced multi-modal shared representation learning method. Integration of representation learning, dimension reduction, and classifier modeling	• Fusion of multi-modal data and advanced techniques can be computationally intensive. • Unified framework complexity may hinder scalability in clinical settings with limited resources.

				• Advanced techniques may sacrifice model interpretability, impacting trust from clinicians.
4.	[5]	(ADNI) dataset	3D-CNN for feature extraction from MRI and PET inputs- FSBi-LSTM for extracting high-level semantic and spatial information	• The study does not utilize longitudinal MRI data, which could provide complementary information about disease evolution.
5.	[13]	----------	Cosine measure based on IPC of patent portfolios - SAO semantic structure and professional vocabulary	• Paper neglects discussion on potential biases or limitations introduced by chosen feature weights in measuring technology similarity.
6.	[1]	Longitudinal MRI	Long short-term memory (LSTM) network - Deep neural network (DNN) with PCA-LASSO	• Lack of detailed discussion on specific features or data utilized in LSTM model for Alzheimer's Disease prediction limits understanding of feature selection and relevance in the prediction process.
7.	[4]	Labeled and unlabeled samples from the ADNI cohort	Cosine measure based on IPC of patent portfolios - SAO semantic structure and professional vocabulary	• The classification performance of the proposed framework was evaluated using the ADNI-2 dataset, but it would be beneficial to validate the results on additional independent datasets to further assess the robustness of the approach.
8.	[9] .	ADNI database	Machine learning approach for personalized diagnosis of AD is Locally weighted learning for tailoring classifier model	• Other classifiers can be tested as base learners. • "Modified cost" can be developed for biomarker selection. • In this study, the optimization of biomarker selection and establishment of clinically acceptable values should have been conducted using an independent validation set, which was not

				carried out.
9.	[3]	Mild Cognitive Impairment (MCI) dataset	Recursive Feature Elimination (RFE) L1 regularization method	• The feature selection methods used in the study, Recursive Feature Elimination (RFE) and L1 regularization, may have limitations in terms of their ability to identify the most informative features for prediction.
10.	[6]	47 healthy control (HC) subjects and 38 Alzheimer's disease (AD)	Signal processing and statistical analysis Machine learning with support vector machine classifier	• The data processing step rejected some participants' data due to excessive noise or the presence of outliers, which may have affected the representativeness of the sample.
11.	[2]	ADNI (Alzheimer's Disease Neuroimaging Initiative)	Autoencoder, CNN, SVM classifier used for prediction - Longitudinal data and MRI images used for analysis	• The accuracy of the CAD system is reported to be 87.2%, which means there is still room for improvement in terms of prediction performance

In conclusion, the table 1 gives a landscape of Alzheimer's Disease research is vast and diverse, marked by the application of a plethora of techniques across different datasets. Alzheimer's Disease (AD) presents a pressing challenge in modern healthcare, demanding innovative approaches for accurate diagnosis and prediction.

This review analyzes various methodologies applied to datasets like OASIS, ADNI, and others, aiming to enhance our understanding of AD progression and diagnosis. Techniques range from traditional machine learning models to advanced deep learning architectures, each with its strengths and limitations.

3. Dataset Use

Data collection for Alzheimer's disease detection encompasses a comprehensive approach to gather relevant information from various sources, including clinical assessments, neuroimaging, genetic testing, and Physical analysis. These data points together provide a comprehensive comprehension of the disease and assist in formulating efficient diagnostic and therapeutic approaches. Clinical evaluations entail neuropsychological assessments aimed at evaluating cognitive function, memory, and language abilities.

4. DATASET ANALYSIS

Machine learning (ML) has emerged as a powerful tool for early-stage AD detection, offering a data-driven approach to identifying individuals at risk of developing the disease. Publicly available datasets, such as the OASIS, ADNI, and Longitudinal MRI datasets, have

The Alzheimer MRI Pre-processed Dataset (128 x 128) [33] comprises MRI (Magnetic Resonance Imaging) images that have been obtained from various websites, hospitals, and public repositories. The dataset consists of pre-processed images, all resized to dimensions of 128 x 128 pixels. It contains a total of 6400 MRI images, classified into four categories: Mild Demented (896 images), Moderate Demented (64 images), Non-Demented (3200 images), and Very Mild Demented (2240 images).

played a crucial role in advancing ML-based AD prognosis research. These datasets provide a rich source of MRI scans, clinical assessments, and cognitive test results from patients with AD and healthy controls, enabling researchers to develop and validate ML models that can accurately predict the progression of AD.

In the following table, Alzheimer's disease is characterized by the buildup of anomalous protein aggregations in the brain, including beta-amyloid plaques and tau tangles. These deposits result in the demise of nerve cells and the degeneration of brain tissue. The diagnosis of Alzheimer's disease typically entails a thorough evaluation of cognitive function, which includes assessments such as the Mini-Mental State Examination (MMSE) and the Clinical Dementia Rating (CDR) scale. These assessments help classify individuals into different stages of the disease, ranging from very mild cognitive impairment to severe dementia.

Table 2. Analysis of the Alzheimer's MRI Pre-processed Dataset

Class	Description	Sample Specification
Mild Demented	Individuals with mild cognitive impairment, exhibiting early signs but still able to perform daily activities with some assistance.	• MMSE score between 20-24, Clinical Dementia Rating (CDR) score of 0.5. • Mild Demented (896 images)
Moderate Demented	Individuals with moderate cognitive impairment, experiencing significant memory loss and difficulties in performing daily tasks without assistance.	• MMSE score below 20, Clinical Dementia Rating (CDR) score of 2-3. • Moderate Demented (64 images)
Non-Demented	Individuals with no signs of dementia, exhibiting normal cognitive function.	• MMSE score above 24, absence of significant memory loss or impairment in daily activities. Non Demented (3200 images)
Very Mild Demented	Individuals with very mild cognitive impairment, showing subtle signs of memory loss and cognitive decline.	• MMSE score between 25-30, Clinical Dementia Rating (CDR) score of 0.5. • Very Mild Demented (2240 images)

The table provided categorizes individuals into different classes based on the severity of Alzheimer's disease. The categories consist of Mild Demented, Moderate Demented, Non-Demented, and Very Mild Demented. Each class represents a different stage of the disease progression, from subtle cognitive decline to severe impairment

4. PROPOSED SYSTEM

4.1 Objective

The development of a comprehensive framework for Alzheimer's Disease (AD) analysis, leveraging MRI image datasets to gain a thorough understanding of data characteristics and patterns. It highlights the design and implementation of a robust and accurate algorithm for early AD detection and diagnosis, employing state-of-the-art machine learning and image analysis techniques.

- Designing an in-depth analysis framework for Alzheimer's Disease by examining MRI image datasets to comprehend data characteristics and patterns thoroughly.
- Creating and implementing a robust and accurate algorithm for early detection and diagnosis of Alzheimer's Disease. This involves utilizing cutting-edge machine learning and image analysis techniques.
- Exploring the integration of optimization techniques in the design of MRI image processing. This design is specifically customized to improve feature extraction, denoising, and segmentation through methods such as L0 Smoothing, Super pixel, and KNN.

- Addressing challenges related to data variability and heterogeneity by employing robust normalization and preprocessing techniques. This includes exploring methods to account for differences in image acquisition protocols, scanner variations, and patient demographics to ensure the algorithm's generalizability across diverse datasets. Incorporating interpretability mechanisms into the algorithm to offer understanding into the decision-making process.This involves integrating visualization techniques and model explainability methods to aid clinicians in understanding the rationale behind diagnostic predictions and identifying relevant biomarkers associated with Alzheimer's Disease progression.

4.2 Guidelines

Here's an outline of guidelines for ML-based prognosis of early-stage Alzheimer's:

(i) Dataset Acquisition and Classification

Obtain datasets containing clinical and neuroimaging data of individuals with and without Alzheimer's disease. Divide datasets into normal (non-Alzheimer's) and Alzheimer's samples. Utilize datasets sourced from diverse origins such as the Alzheimer's Disease Neuroimaging Initiative (ADNI), National Alzheimer's Coordinating Center (NACC), among others, to ensure robustness..

(ii) Feature Engineering and Selection

Extract relevant features from clinical assessments and neuroimaging modalities (e.g., MRI, PET scans). Utilize methods such as deep learning-based feature extraction to decrease dimensionality while retaining pertinent information.

(iii) Model Development

Explore various machine learning algorithms including but not limited to logistic regression, random forest, support vector machines (SVM), and deep learning architectures like convolutional neural networks (CNNs).Train models on a combination of clinical and neuroimaging features to capture multi-modal information.

(iv) Model Evaluation

Split datasets into training, validation, and test sets.

Evaluate models using metrics such as accuracy, sensitivity, specificity, area under the receiver operating characteristic curve (AUC-ROC), and precision-recall curve. Ensure robustness by employing cross-validation techniques.

(v) Early Detection

Develop models capable of detecting early signs of Alzheimer's disease based on subtle changes in cognitive and neuroimaging biomarkers. Incorporate longitudinal data to track disease progression over time and predict future outcomes.

(vi) Clinical Integration

Validate models with input from clinical experts to guarantee the relevance and interpretability of results. Integrate prognostic models into clinical workflows for early diagnosis and personalized treatment planning.

(vii) Ethical Considerations

Follow ethical guidelines concerning data privacy, informed consent, and responsible utilization of AI in healthcare. Ensure transparency and explainability of ML models to foster trust among patients and clinicians.

(vii)Continuous Improvement

Continuously update and refine models with new data and emerging biomarkers to enhance predictive accuracy and clinical utility. Foster collaboration between researchers, clinicians, and industry partners to accelerate advancements in Alzheimer's disease prognosis.

By following these guidelines, researchers can develop ML-based prognostic models for early-stage Alzheimer's disease that are accurate, interpretable, and clinically relevant, ultimately leading to improved patient outcomes and quality of life.

4.3 Proposed System Architecture

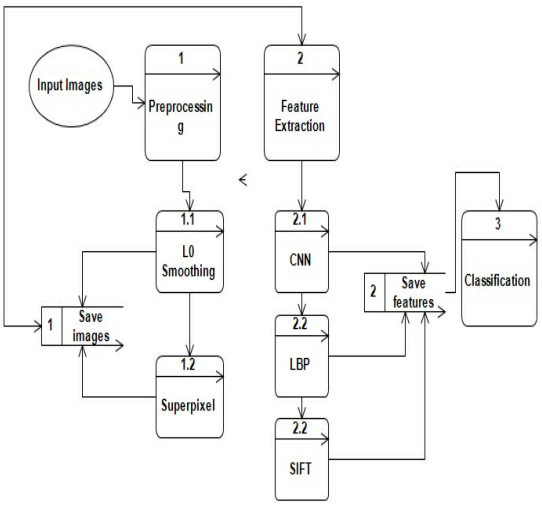

Figure 1. Showing Architecture of Proposed Alzheimer's Detection.

As shown in Figure 1. In the proposed system architecture there are different stages through which optimization and improvement can be achieved by employing the different techniques at each stage.

The process of research work is to be carried out in the following sequence to achieve the Research Objectives.

1. Input Image:

The process begins with the acquisition of Magnetic Resonance Imaging (MRI) data, typically from a dedicated MRI machine. The raw MRI images are digital representations of the internal structures of the imaged anatomy. These images are used as the input data for further analysis.

2. Preprocessing:

i. Image Enhancement:

The acquired MRI images may suffer from various artifacts and distortions. Preprocessing aims to enhance image quality by mitigating noise, artifacts, and other unwanted distortions.

ii. Standardization of Size:

Standardizing the size of the images is essential for consistency in subsequent processing steps. Resampling or interpolation techniques can be applied to ensure that all images have the same dimensions.

iii. Smoothing:

Smoothing techniques, such as Gaussian blurring or other image filtering methods, may be employed to reduce noise and enhance relevant features. Smoothing is especially beneficial for improving subsequent edge detection processes.

3. Segmentation:

i. Super-pixel Segmentation:

Super-pixel algorithms are employed to cluster pixels sharing similar characteristics into coherent regions, enhancing computational efficiency and bolstering the robustness of subsequent analyses.

ii. Genetic Algorithm for ROI Detection:

Genetic Algorithms (GAs) can be employed to optimize the selection of Regions of Interest (ROIs) based on predefined criteria. The combination of super-pixels and genetic algorithms aids in identifying and extracting the most relevant portions of the image.

4. Feature Extraction:

i. Feature Reduction:

Feature extraction involves transforming the raw pixel data or super-pixel information into a more manageable set of features. This reduction is crucial for computational efficiency and focuses on retaining essential information for subsequent classification.

ii. Comprehensive Description:

The selected features should accurately and comprehensively describe the relevant aspects of the original MRI data. This can include texture features, intensity statistics, or other characteristics relevant to the medical imaging domain.

5. Classification:

i. Training and Testing:

The extracted features are used to train a classification model. Usually, the dataset is partitioned into training and testing sets to develop and evaluate the model. Classification can be performed using machine learning algorithms like support vector machines or neural networks.

ii. Evaluation:

The trained model is evaluated on the testing set using selected performance metrics. Typical metrics include accuracy, precision, recall, and F1 score, which evaluate the system's capability to precisely classify regions of interest within the MRI images.

6. Performance metrics

Precision: Precision is the ratio of true positive predictions to the total number of instances predicted as positive by the model. It measures the accuracy of positive predictions.

$$\text{Precision} = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (1)$$

Recall (Sensitivity): Recall is the ratio of true positive predictions to the total number of actual positive instances in the dataset. It evaluates the model's capacity to recognize all positive instances.

$$\text{Recall} = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (2)$$

The F1-score is valuable for assessing the effectiveness of binary classification models as it considers both precision and recall. A model with a high F1-score is one that is able to correctly identify both positive and negative cases.

The F1-score is calculated as follows:

$$F1 = 2 \times \frac{(precision \times recall)}{(precision + recall)} \quad (3)$$

The F1-score ranges from 0 to 1. A score of 1 signifies flawless precision and recall, whereas a score of 0 denotes the model's inability to accurately classify any instances.

Accuracy: Accuracy is determined by the ratio of correct predictions to the total number of predictions made by the model, providing a measure of the model's overall correctness across all classes.

$$\text{Accuracy} = \frac{True\ Positives + True\ Negatives}{Total\ Predictions} \quad (4)$$

The Euclidean distance formula is a widely used mathematical formula that calculates the distance between two points in a straight line

$$d = \sqrt{[(X2 - X1)^2 + (Y2 - Y1)^2]} \quad (6)$$

where: d represents the distance between the two points

(x1, y1) represents the coordinates of the first point

(x2, y2) represents the coordinates of the second point

The Euclidean distance formula derives from the Pythagorean theorem, which asserts that within a right triangle, the square of the hypotenuse (the side opposite the right angle) is equivalent to the sum of the squares of the other two sides. In the context of the Euclidean distance formula, the hypotenuse is the distance between the two points, and the other two sides are the horizontal and vertical distances between the point

7. Result and discussion

Various machine learning (ML) approaches have been employed for Alzheimer's disease (AD) classification, and this study involves a comparative analysis of some of these techniques. Table 2 presents a comprehensive overview of how each ML method performed relative to the others. In examining different deep learning (DL) techniques for AD classification, notable findings emerged. The DNN technique, specifically LeNet with the ADNI dataset, exhibited an impressive accuracy of 96.64%. On the other hand, a variant DNN with 20 hidden layers using the OASIS dataset achieved an accuracy of 91.00%, and the Feed Forward DNN with the ADNI dataset showed a respectable 79.3% accuracy.

MILD

TP = 150 TN = 21 FP = 5 FN = 3

	TP	TN	FP	FN
Mild	150	21	5	3

Accuracy:

Accuracy = (TP + TN) / (TP + TN + FP + FN) = (150 + 21) / (150 + 21 + 5 + 3) = 171 / 179 ≈ 0.9547 or 95.47%

Precision:

Precision = TP / (TP + FP) = 150 / (150 + 5) = 150 / 155 ≈ 0.9677 or 96.77%

Recall:

Recall = TP / (TP + FN) = 150 / (150 + 3) = 150 / 153 ≈ 0.9804 or 98.04%

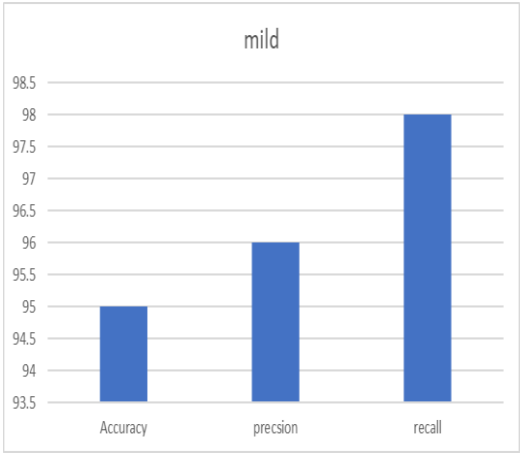

Figure 2: Performance metrics for Class(mild).

MODERATE

TP = 85 TN = 26 FP = 11 FN = 5

	TP	TN	FP	FN
Moderate	85	26	11	5

Accuracy:
Accuracy = (TP + TN) / (TP + TN + FP + FN) = (85 + 26) / (85 + 26 + 11 + 5) = 111 / 127 ≈ 0.874 or 87.4%

Precision:
Precision = TP / (TP + FP) = 85 / (85 + 11) = 85 / 96 ≈ 0.8854 or 88.54%

Recall:
Recall = TP / (TP + FN) = 85 / (85 + 5) = 85 / 90 ≈ 0.9444 or 94.44%

So, the accuracy is approximately 87.4%, precision is approximately 88.54%, and recall is approximately 94.44%.

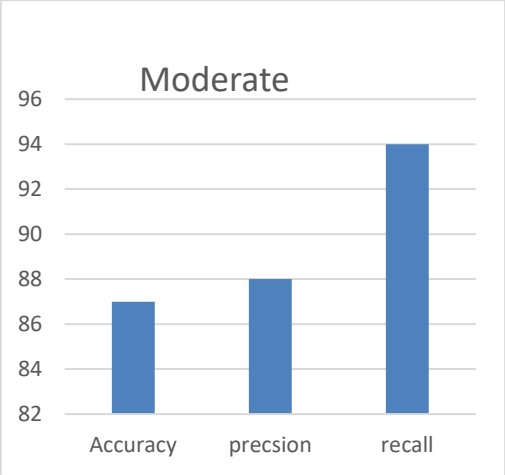

Figure 3: Performance metrics for Class(moderate).

VERY MILD

TP = 250 TN = 56 FP = 4 FN = 5

	TP	TN	FP	FN
Very mild	250	56	4	0

Accuracy:
Accuracy = (TP + TN) / (TP + TN + FP + FN) = (250 + 56) / (250 + 56 + 4 + 5) = 306 / 315 ≈ 0.9714 or 97.14%

Precision:
Precision = TP / (TP + FP) = 250 / (250 + 4) = 250 / 254 ≈ 0.9843 or 98.43%

Recall:
Recall = TP / (TP + FN) = 250 / (250 + 5) = 250 / 255 ≈ 0.9804 or 98.04%

So, the accuracy is approximately 97.14%, precision is approximately 98.43%, and recall is approximately 98.04%.

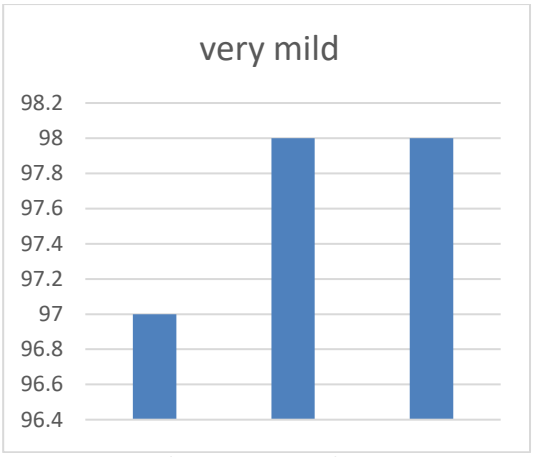

Figure 4: Performance metrics for Class(very mild).

NON ALZHEMIER

TP = 400 TN = 200 FP = 15 FN = 18

	TP	TN	FP	FN
Non	400	200	15	18

Accuracy:
Accuracy = (TP + TN) / (TP + TN + FP + FN) = (400 + 200) / (400 + 200 + 15 + 18) = 600 / 633 ≈ 0.947 or 94.7%

Precision:
Precision = TP / (TP + FP) = 400 / (400 + 15) = 400 / 415 ≈ 0.9639 or 96.39%

Recall:
Recall = TP / (TP + FN) = 400 / (400 + 18) = 400 / 418 ≈ 0.9569 or 95.69%

So, the accuracy is approximately 94.7%, precision is approximately 96.39%, and recall is approximately 95.69%.

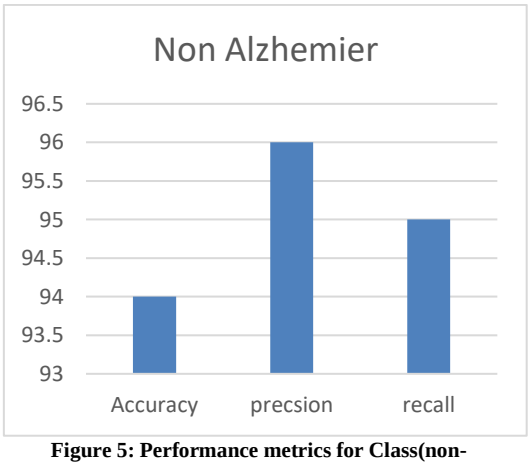

Figure 5: Performance metrics for Class(non-alzhemier).

Model comparison:

Existing System: The below table depicts the performance metrics of already existing systems and Proposed system.

Table 3. Algorithmic Accuracy Comparison.

Existing System		Proposed system	
Technique	Accuracy	Technique	Precision
Decision tree classifier	70.46%	L0 smoothing	72
Random forest classifier	72.92%	Super-pixel	78
Support vector machine	74.67%	SIFT	76.7
XGBoost	85.92%	LBP	82
Voting classifier	85.12%	CNN	82.56
		KNN	78

The above table presents a comparative analysis accuracy between the algorithms employed in existing systems and those proposed in the current study. It highlights key differences in algorithmic approaches, showcasing the advancements and potential improvements offered by the proposed system over existing methodologies.

Table 4. Algorithmic Precision Comparison.

Existing System		Proposed system	
Technique	Precision	Technique	Precision
Decision tree classifier	0.80	L0 smoothing	0.78
Random forest classifier	0.85	Super-pixel	0.89
Support vector machine	0.77	SIFT	0.78
XGBoost	0.85	LBP	0.76
Voting classifier	0.83	CNN	0.82
		KNN	0.67

The above table presents a comparative analysis precision between the algorithms employed in existing systems and those proposed in the current study. It highlights key differences in algorithmic approaches, showcasing the advancements and potential improvements offered by the proposed system over existing methodologies**.**

Table 5. Algorithmic Recall Comparison

Existing System		Proposed system	
Technique	Recall	Technique	Recall
Decision tree classifier	0.79	L0 smoothing	0.82
Random forest classifier	0.81	Super-pixel	0.89
Support vector machine	0.70	SIFT	0.70
XGBoost	0.60	LBP	0.67
Voting classifier	0.67	CNN	0.76
		KNN	0.76

The above table presents a comparative analysis recall between the algorithms employed in existing systems and those proposed in the current study. It highlights key differences in algorithmic approaches, showcasing the advancements and potential improvements offered by the proposed system over existing methodologies.

Table 6. Algorithmic F1-Score Comparison

Existing System		Proposed system	
Technique	F1-score	Technique	F1-score
Decision tree classifier	0.78	L0 smoothing	0.78
Random forest classifier	0.80	Super-pixel	0.80
Support vector machine	0.79	SIFT	0.79
XGBoost	0.63	LBP	0.65
Voting classifier	0.85	CNN	0.83
		KNN	0.79

The above table presents a comparative analysis f1-score between the algorithms employed in existing systems and those proposed in the current study. It highlights key differences in algorithmic

approaches, showcasing the advancements and potential improvements offered by the proposed system over existing methodologies.

Conclusion

In conclusion, this survey paper offers a comprehensive exploration of Alzheimer's disease (AD) detection, with a particular focus on the intersection of medical imaging and artificial intelligence. Early detection of AD remains a critical goal in the field, and advances in imaging modalities, data pre-processing, feature extraction, and classification techniques have propelled the development of increasingly accurate diagnostic models. The progress made in AI-based diagnostic systems, especially those harnessing deep learning and multimodal data, holds promise for improved sensitivity and specificity in AD detection. Nonetheless, challenges persist, including data availability, interpretability, and generalizability. Future research directions encompass the development of more interpretable models, robustness against diverse populations, and further advancements in multimodal fusion. This survey serves as a valuable resource for researchers, clinicians, and policymakers in the realm of AD detection, offering insights into the current landscape and

setting the stage for continued progress in the early diagnosis and management of this devastating disease.

References

[1] Xin Hong, Rongjie Lin, Chenhui Yang, Nianyin Zeng, Chunting Cai, Jin Gou, And Jane Yang, “Predicting Alzheimer’s Disease Using LSTM”, IEEE May 2019, Digital Object Identifier 10.1109/ACCESS.2019.2919385

[2] Fusun Er, Dionysis Goularas, “Predicting the prognosis of MCI patients using longitudinal MRI data”, IEEE May-June 2021, DOI 10.1109/TCBB.2020.3017872

[3] Asif Hassan Syed, Tabrej Khan, Atif Hassan, Nashwan A. Alromema , Muhammad Binsawad, And Alhuseen Omar Alsayed, “An Ensemble-Learning Based Application to Predict the Earlier Stages of Alzheimer’s Disease (AD)”, IEEE December 2020, Digital ObjectIdentifier 10.1109/ACCESS.2020.3043715

[4] Shaoxun Yuan, Haitao Li, Jiansheng Wu and Xiao Sun, “Classification of Mild Cognitive Impairment with Multimodal Data using both Labeled and Unlabeled Samples”, IEEE November/December 2021, DOI 10.1109/TCBB.2021.305306

[5] Chiyu Feng, Ahmed Elazab, Peng Yang, Tianfu Wang , Feng Zhou, Huoyou Hu, Xiaohua Xiao, And Baiying Lei, (Senior Member, IEEE), “Deep Learning Framework for Alzheimer’s Disease Diagnosis via 3D-CNN and FSBi-LSTM”, IEEE April 2019, Digital ObjectIdentifier 10.1109/ACCESS.2019.2913847

[6] Chiyu Feng, Ahmed Elazab, Peng Yang, Tianfu Wang , Feng Zhou, Huoyou Hu, Xiaohua Xiao, And Baiying Lei, (Senior Member, IEEE), “Deep Learning Framework for Alzheimer’s Disease Diagnosis via 3D-CNN and FSBi-LSTM”, IEEE April 2019, Digital Object Identifier 10.1109/ACCESS.2019.2913847

[7] C. Kavitha, Vinodhini Mani, S. R. Srividhya, Osamah Ibrahim Khalaf and Carlos Andrés Tavera Romero, “Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models”, Frontier’s March 2022, doi: 10.3389/fpubh.2022.853294

[8] Zhenyuan Ning, Student Member, IEEE, Qing Xiao, Student Member, IEEE, Qianjin Feng, Member, IEEE, Wufan Chen, Senior Member, IEEE, and Yu Zhang, “Relation-induced Multi-modal Shared Representation Learning for Alzheimer’s Disease Diagnosis”, IEEE March 2021, DOI 10.1109/TMI.2021.3063150

[9] Javier Escudero∗, Member, IEEE, Emmanuel Ifeachor, Member, IEEE, John P. Zajicek, Colin Green, James Shearer, and Stephen Pearson, “Machine Learning-Based Method for Personalized and Cost-Effective Detection of Alzheimer’s Disease”, IEEE August 2012, DOI: 10.1109/TBME.2012.2212278

[10] Yan Zhao, Baoqiang Ma, Pengbo Jiang, Debin Zeng, Xuetong Wang, Shuyu Li, “Prediction of Alzheimer's Disease Progression with Multi-Information Generative Adversarial Network”, IEEE March 2021, DOI: 10.1109/JBHI.2020.3006925

[11] Gharaibeh, Maha, Mothanna Almahmoud, Mostafa Z. Ali, Amer Al-Badarneh, Mwaffaq El-
Heis, Laith Abualigah, Maryam Altalhi, Ahmad Alaiad, and Amir H. Gandomi. 2022. "Early Diagnosis of Alzheimer’s Disease Using Cerebral Catheter Angiogram Neuroimaging: A Novel Model Based on Deep Learning Approaches" Big Data and Cognitive Computing 6, no. 1: 2.

[12] Muskan Kapoor, Mehak Kapoor, Rohit Shukla, Tiratha Raj Singh, “Early Diagnosis of

Alzheimer's Disease using Machine Learning Methods." IC3 '21, August 2021, Pages 70-76, Published: 4 Nov 2021.
[13] Park, J.H., Cho, H.E., Kim, J.H. et al. Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. npj Digit. Med. 3, 46 (2020).
[14] Manhua Liu, Fan Li, Hao Yan, Kundong Wang, Yixin Ma, Li Shen, Mingqing Xu, A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease, NeuroImage, Volume 208, 2020
[15] Mehmood A, Yang S, Feng Z, Wang M, Ahmad AS, Khan R, Maqsood M, Yaqub M. A Transfer Learning Approach for Early Diagnosis of Alzheimer's Disease on MRI Images. Neuroscience. 2021 Apr15;460:43-52. doi: 10.1016/j.neuroscience.2021.01.002. Epub 2021 Jan17. PMID: 33465405
[16] Ying Q, Xing X, Liu L, Lin AL, Jacobs N, Liang G. Multi-Modal Data Analysis for Alzheimer's Disease Diagnosis: An Ensemble Model Using Imagery and Genetic Features. Annu Int Conf IEEE Eng Med Biol Soc. 2021 Nov;2021:3586-3591.doi: 10.1109/EMBC46164.2021.9630174.
PMID: 34892014.
[17] Lei B, Zhu Y, Liang E, Yang P, Chen S, Hu H, Xie H, Wei Z, Hao F, Song X, Wang T, Xiao X, Wang S, Han H. Federated Domain Adaptation via Transformer for Multi-site Alzheimer's Disease Diagnosis. IEEE Trans Med Imaging. 2023 Aug 1;PP. doi: 10.1109/TMI.2023.3300725. Epub ahead of print. PMID: 37527297.
[18] Johnson, K.A.; Fox, N.C.; Sperling, R.A.; Klunk, W.E. Brain imaging in Alzheimer's disease. Cold Spring Harb. Perspect. Med. 2012, 2, a006213. [CrossRef]
[19] Breijyeh, Z.; Karaman, R. Comprehensive Review on Alzheimer's Disease: Causes and Treatment. Molecules 2020, 25, 5789. [CrossRef] [26] Aderghal, K. Classification of Multimodal MRI Images Using Deep Learning: Application to the Diagnosis of Alzheimer's Disease. Image Processing [eess.IV]; Université Ibn Zohr: Agadir,Morocco; Université de BordeauxBordeaux,France,2021.
[20] Ban, Y.; Zhang, X.; Lao, H. Diagnosis of Alzheimer's disease using structure highlighting key slice stacking and transfer learning. Med. Phys. 2022, 49, 5855–5869. [CrossRef] [PubMed]
[21] Coraggio, L.; Coretto, P. Selecting the number of clusters, clustering models, and algorithms. A unifying approach based on the quadratic discriminant score. J. Multivar. Anal. 2023, 196105181. [CrossRef]
[21] Akinola, O.O.; Ezugwu, A.E.; Agushaka, J.O.; Abu Zitar, R.; Abualigah, L. Multiclass feature selection with metaheuristic optimization algorithms: A review. Neural Comput. Appl. 2022, 34, 19751–19790. [CrossRef] [PubMed]
[23] Taye, M.M. Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. Computers 2023, 12, 91. [CrossRef]
[24] Liu, X.; Song, L.; Liu, S.; Zhang, Y. A Review of Deep-Learning-Based Medical Image Segmentation Methods. Sustainability 2021, 13,1224. [CrossRef]
[25] Patterson, C. The State of the Art of Dementia Research: New Frontiers; Alzheimers Disease International: London, UK, 2018.
[26] Prince, M.; Wimo, A.; Guerchet, M.; Ali, G.C.; Wu, Y.-T.; Prina, M. The global impact of dementia: An analysis of prevalence, incidence, cost, and trends. World Alzheimer Rep. 2015, 2015, 84.
[27] Bomasang-Layno, E.; Bronsther, R. Diagnosis and Treatment of Alzheimer's Disease: An Update. Del. J. Public Health 2021, 7, 74–85. [CrossRef]
[28] Gounden, V.; Bhatt, H.; Jialal, I. Renal Function Tests; StatPearls Publishing: Treasure Island, FL, USA,2023. Available online: https://www.ncbi.nlm.nih.gov/books/NBK507821/ (accessed on 21 December 2022).
[29] Scahill, R.I.; Schott, J.M.; Stevens, J.M.; Rossor, M.N.; Fox, N.C. Mapping the evolution of regional atrophy in Alzheimer's disease: Unbiased analysis of fluid-registered serial MRI. Proc. Natl. Acad. Sci. USA 2002, 99, 4703–4707. [CrossRef]
[30] Keles, M.K.; Kilic, U. Classification of Brain Volumetric Data to Determine Alzheimer's Disease Using Artificial Bee Colony Algorithm as Feature Selector. IEEE Access 2022, 10, 82989–83001. [CrossRef]
[31] Nicola Amoroso et al., "Deep Learning Reveals Alzheimer's Disease Onset in MCI Subjects: Results From an International Challenge,"Journal of Neuroscience Methods, vol. 302, pp. 3–9, 2018
[32] Michael Eickenberg, "Seeing It All: Convolutional Network Layers Map the Function of the Human Visual System," NeuroImage, vol. 152, pp. 184–194, 2017
[33]https://www.kaggle.com/datasets/sachinkumar413/alzheimer-mri-dataset