International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Predicting the Transition from Mild Cognitive Impairment to Alzheimer's Disease using Cognitive Tests and MRI Measures of Demographic Data with an Ensemble Model

Pothala Ramya[1], Dr. Ch. Ramesh[2], Dr. O. Srinivasa Rao[3]

Submitted: 17/09/2023 **Revised**: 17/11/2023 **Accepted**: 29/11/2023

Abstract: Dementia, one of the most dreaded illnesses, has an enormous annual impact on health and social care expenses worldwide than cancer and chronic heart disease. Despite the lack of a treatment or standardized clinical test, using machine learning techniques to identify people at risk of developing Dementia could represent a new step toward proactive management. Cerebrospinal fluid (CSF), positron emission tomography (PET) scans, magnetic resonance imaging (MRI), biological markers (biomarkers), clinical scans, and neuropsychological therapy are all integrated to track the development of early Alzheimer's disease (AD) and moderate cognitive impairment (MCI). Early detection of Alzheimer's disease (AD) is crucial for controlling the illness, assistance, and the accessibility of healthcare resources. This study focused on the detection rate and false positive rate of a disease determined from ADNI-ADNIMERGE demographic data using a variety of machine learning techniques, including KNN, SVM, RF, NB, LOGISTIC, and Ensembled: LOGISTIC-PCA, SVM, KNN as the final algorithm with feature selection and hyper-tuning parameter optimization. Performed a comparison analysis between machine learning methods and ensemble model. Ensemble model showed best results with change in biomarker and baseline biomarker of disease detection rate, false positive rate and test accuracy 92%, 90% of AD respectively.

Keywords: *Machine Learning, Ensemble, ADNIMERGE, Feature selection, principal component analysis, and Alzheimer's disease (AD)*

1. Introduction

This paper's primary focus is proposing a study on various data mining techniques to identify degenerative brain diseases. Alois Alzheimer coined AD in 1906 after studying human brains to understand why memory loss persisted and how biological marker plaques and tangles formed in neurons. The primary causes of AD, a neurodegenerative disorder, are two aggregates called amyloid plaques and microtubule-associated protein tau that are joined to form neurofibrillary tangles inside dying neurons.

Alzheimer's disease is a common cause of Dementia that interferes with essential functions like communication, the ability to carry out everyday tasks, identification, and reasoning. It also causes memory loss and the repair of neurons and their networks. The primary brain region, the hippocampus, which aids in memory formation, was severely damaged by AD and tissue loss.

Before 2010, there were 24 million cases of Dementia worldwide; by 2016, there were about 50 million cases. With population growth and aging, it was predicted that by 2050, there would be more than four times as many people. Because AD is a chronic

condition that may last the entire life, it is essential to prescribe medication appropriately to prevent severe brain damage. Early diagnosis required the application of sophisticated algorithms, which was a time-consuming and challenging process.

The main goal is to predict degenerative brain disease (Alzheimer's disease) more accurately, but detection is also crucial. For this reason, we need solid, cutting-edge, non-traditional methodologies like machine learning techniques. To effectively and accurately identify the affected humans in this study, we applied sociodemographic ADNIMERGE data gathered from ADNI using different machine learning techniques like KNN, SVM, RF, NB, LOGISTIC, and Ensembled: LOGISTIC-PCA, SVM, KNN as a final algorithm with feature selection and hyper-tuning parameter optimization to fine who are developed with AD and who are not at risk.

The rest of the covers; Sect. 2, related research and analysis of prior work on supervisor learning methods for Alzheimer's detection. Sect. 3. Methodology offers details on applying machine learning techniques. Sect. 4 discusses the findings of the experiment. The conclusion is at Sect. 5.

[1]*Research Scholar, Department of Computer Science and Engineering, JNTUK University, Kakinada, Andhra Pradesh*
[1]*Assistant Professor, Department of IT, MVGR College of Engineering, India*
[2]*Professor(CSE) & Dean (IQAC), Aditya Institute of Technology and Management, Tekkali Srikakulam(Dt), Andhra Pradesh, India. Email id:chappa_ramesh01@yahoo.co.in*
[3]*Professor, Head of Computer Science and Engineering Department, JNTUK University, Kakinada, Andhra Pradesh, India*
** Corresponding Author Email:pothalaramya.583@gmail.com*

2. Literature Review

There is no curable treatment of Alzheimer's Disease at all the only solution is early detection and monitoring. So, to track and manage with this problem, up to now so many methodologies have been explored. In this section we meticulously examine literature survey with machine and deep learning algorithms, ensemble learning.

Fadi Thabtah et al. [1], a data-driven methodology based on feature selection and classification, was used to study the Functional Activities Questionnaire (FAQ) components, an early screening instrument. The findings of using demographic data from FAQ with machine learning techniques reveal models with accuracy, sensitivity, and specificity all-surpassing 90%. The progression class also showed correlations with FAQ elements such as Administration and Shopping attributes; using ADNIMERGE, FAQ datasets were combined to create a new dataset called ADNIFAQ and applied on NB, LR, and DT->C4.5, which achieved the highest accuracy, sensitivity, and specificity of 92, 93, and 92%, respectively.

Erik D. Huckvale et al. [2] employed 793,600 extracted MRI features and 49,288 biomarkers to analyze feature correlation in the ADNI dataset. Based on our Bonferroni corrected analysis (p-value 1.40754 1013), we discovered that 100% of MRI features, 92.549% of gene expression levels, and 93.457% of biomarkers were strongly linked with at least one other component in ADNI. MRI, gene expression, and ADNIMERGE domains were combined into a single dataset for correlation analysis. The total number of features in the combined dataset for 743 people was 842,888, consisting of 1131 ADNIMERGE features, 48,157 gene features, and 793,600 MRI features.

Massimiliano Grassi et al. [3] The proposed ensemble-based machine learning system makes use of a variety of factors, such as sociodemographic traits, clinical data, and neuropsychological measurements, to predict the progression from Mild Cognitive Impairment (MCI) to Alzheimer's Disease (AD). The algorithm makes use of a wide range of factors, such as sociodemographic information (such as age, gender, and educational level), clinical data (such as medical history and family history of AD), and neuropsychological measures (such as results from cognitive tests and memory evaluations). Data from diverse sources are combined in the proposed technique, representing the complexity of MCI and AD. This holistic approach makes a more thorough understanding of the risk factors and potential indicators of illness conversion possible. On removed test data, the final ensemble method displayed an AUROC of 0.88, a sensitivity of 77.7%, and a specificity of 79.9%. For 100% sensitivity, the algorithm's specificity was 40.2%.

Jayant Prakash et al. [4] used a dataset of clinical data from a cohort of Alzheimer's patients, including patient demographics, medical history, cognitive test results, neuroimaging data, and genetic data. For capturing the variation within the AD population, this data is crucial. They have used a variety of unsupervised machine learning approaches on ADNI Dataset, including dimensionality reduction methods like Principal Component Analysis and clustering algorithms like K-means, hierarchical clustering, and DBSCAN. Four clinical subpopulations of AD were found using between-cluster mean fold changes, with C1–4 representing the least, most, and mild severity, respectively. The four found clusters offer quantifiable, data-enabled support for protocols to categorize subpopulations of AD patients using conventional, easily accessible clinical criteria. Consistent sub-population classification of AD patients may lessen patient heterogeneity, which otherwise muddles clinical trial assessments of AD therapy success.

Muhammad Irfan et al. [5] implemented a machine-learning strategy to deal with this problem, utilizing cognitive and neuroimaging characteristics for building predictive models. This study brought attention to the usefulness of cognitive test results in identifying Dementia straightforward process. The AdaBoost Ensemble model demonstrated strong performance with an accuracy rate of about 83% after being trained on cognitive characteristics. Benchmark models, including the Artificial Neural Network, Support Vector Machine, and Naive Bayes, are outperformed by this model.

Afreen Khan et al. [6] Generated psychometric test results using a cognitive-based, three-tiered machine learning (ML) system that uses baseline traits to forecast AD or moderate cognitive impairment (MCI). Current methods for diagnosing AD using machine learning employ a binary or multinomial classification algorithm. It relies on creating a robust hybrid cognitive ML algorithm that uses demographic information from ADNI to forecast the disease accurately and precisely—devised a method for stacking 2-layer models. Six ML classifier combinations—Logistic Regression, Naive Bayes, Support Vector Machine, Decision Trees, Random Forest, and eXtreme Gradient Boosting—were outperformed by model stacking. Tier1 XGB, Random Forest, and SVM all had an Accuracy of 89.63%, whereas Random Forest had an accuracy of 93.90%. Tier 2 enhanced categorization and overall prediction performance. With experiment 1 providing 90.24% accuracy and experiment 2 yielding 95.12% accuracy, tier 3 hybrid modeling accuracy significantly increased.

Tanveer M et al. [7] SVM, ANN, and DL, three crucial supervisor learning methods, were used to diagnose brain dementia. They have also researched additional learning

techniques like transfer, ensemble, and multi-kernel learning. SVM-based algorithms have detected Alzheimer's illness numerous times, proving its dependability. This is because SVM does not suffer from the drawbacks of local minima, unlike methods like ANN. SVM was widely used because it is simpler to understand than black-box models like neural networks. Future work on this problem should focus on the clinical interpretability of deep learning models. Additionally, it has been noted that researchers have paid more attention to the feature extraction stage than the categorization stage.

Esther E. Bron et al. [8] studied two groups: the first group included 1715 persons from the ADNI, while the second group was drawn from the Parelsnoer Neurodegenerative Disease Biobank (PND) at health-RI. The data is imaged using structural MRI T1w at 1.5T or 3T. The photos are uncoiled utilizing several channels, including 8-channel, 16-channel, and 40-channel (N=1). The images are corrected using the N4 algorithm and then translated into MNI space using brain masks that exhibit comparable transformations. The two methods utilized to build the model were SVM and CNN. The C parameter in SVM is used five times on the training set. The CNN network comprises seven models, including filters such as a 3D convolution layer, dropout, batch normalization (BN), and ReLu as an activation function. The findings demonstrated that, compared to the AUC curve modulated with T1w pictures, the GM maps-based AUC curve had a more considerable delay. Similar results were observed with CNN, with GM maps generating more accuracy than the T1w pictures.

Jun Pyo Kima et al. [9] Frontal Temporal Dementia (FTD) was the most common form of early-onset Dementia. A subject categorization model for each subject would be considerably more beneficial than a group analysis. The main objective of this research was to categorize every patient into a specific category utilizing a machine learning-based classification algorithm and surface-based cortical thickness data. PET images were used to organize participants into AD or CN groups based on their labels. According to the classification results, each subject was accurately classified into one of five clinical categories with an accuracy rate of 75.8%. To classify FTD clinical symptoms differently, we developed an artificial classifier. Using a fully automated classifier, cortical thickness data alone could classify FTD clinical subgroups and AD with good to outstanding accuracy.

Mingxia Liuy et al. [10] Convolutional neural networks (CNNs) were part of a deep learning model. They proposed a method for disease prediction using autonomous landmark-based deep feature learning (LDFL). The Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD) and Alzheimer's Disease Neuroimaging Initiative (ADNI) datasets were used. LDFL lays the way for discriminative biomarkers in morphological analysis of MR images and computer-aided diagnosis of Alzheimer's disease.

Binny Naik et al. [11] Alzheimer's disease (AD) is a neurological condition that causes persistent memory loss and cognitive decline. ML techniques are used to determine whether or not a disorder has affected complicated neuroimaging data in various neurological diseases. Many experiments were run to evaluate the effectiveness of utilizing a multi-classifier because the main objective of SVM is to provide the optimum hyper-plane that separates data points of various classes that were not formed. Adding the PET/SPECT/CSF modality to MRI increased classification accuracy over MRI alone, which helped the SVM classifier's performance rates. When several modalities are added, the classifier's accuracy improves.

Samaneh Abolpour Mofrad et al. [12] The proposed model includes cortical parcellation and sub-cortical segment extraction from TW1 images using FreeSurfer v.6.0, followed by the selection of 3D regions that exhibit a propensity for AD and fitting into a model for prediction. Sub-cortical segments are subcortical structures divided into segments that make it easy for neuro-imaging data and can be used for multiple analyses. For CN and MCI, the accuracy is roughly 73% and 78%. With the same methodologies as its base reference model, which had previously only achieved 64% efficiency among 224 candidates, the primary goal of this research is to increase accuracy and performance. Additionally, it has suggested algorithmic instability that results in poor model quality. The use of FreeSurfer v.6.0 can be held responsible for part of the variation being reduced (which subsequently improves the base model's forecast accuracy), although instability still exists.

F.J. Martinez-Murcia et al. [13] FJ CNN with MLP (multilayer perception) was first introduced using quicker and more precise models. A system of encoder-decoder has been developed that extracts features and lowers the reconstruction error. The Z-layer features, which also comprise MLP, NR (Neural Regression), and SVM, were hampered by an encoder and decoder interference. GAP (Global Average Pooling), which aids in overcoming the issue of over-fitting and enhances convergence while significantly reducing parameters, was used to substitute the encoder's output. ReLU is frequently employed in CNN, although the Z-layer employs a linear activation function for the Z-manifold. Batch normalization served as a regularizer and was used to speed up convergence. An MLP and two hidden layers of 64 neurons were used to develop a prediction model. The CAE space is explained with a visual aid. The Z-features and other forms of data

are correlated using regression analysis. Utilizing tissue maps and clinical data, each neuron was mapped. With a correlation coefficient of 0.63, the tissue map with the highest score was identified as GM (classification as GM, WM, and norm). The case of regression is where GM and WM diverge. It noted that AD first affects GM before moving on to WM. In the end, they have attained 84% accuracy.

Suhad Al-Shoukry et al. [14] congenital observations were a limitation of machine learning techniques. Deep learning is frequently used to identify AD. The body of research on the history of AD and the value of applying deep learning for diagnosis is comprehensive. The National Health And Ageing Trends Study, Open Access Series of Imaging Studies(OASIS), Max Planck Institute Leipzig Mind-Body Dataset-Lemon, and ADNI 1.5T imaging data were also used to investigate the AD MRI, PET, and SPECT results.

Behnaz Ghoraani et al. [15] A method to build a precise and enduring methodology for diagnosing MCI and AD was proposed. This study establishes an automated process for analyzing a mental decline in MCI and AD patients using only gait data and identifies essential gait factors for machine learning-based categorization. Using gait as a cognitive impairment screen may prompt medical professionals to schedule additional testing to identify MCI and AD. They recorded the gaits of 78 elderly individuals as they walked in various single- and dual-task environments. The 108 gait features from each individual were extracted, and significant uncorrelated components were found. After that, they were depending on the given gait factors, a machine learning technique was used to obtain the clinical diagnosis. The method produces 25 significant uncorrelated gait factors for differentiating between healthy and MCI, healthy and AD, and MCI and AD, as well as 13 for MCI and AD. The five-fold classification accuracy was 78 percent using the given gait parameters, a little under 83 percent.

Xia-an Bi et al. [16] Researched multimodal Alzheimer's disease data fusion. Correlation analysis is used to explore associations between genes and brain regions. Second, using the CERF to assess "brain region-gene pairings" and eliminate the traits that set AD and HC apart is advised. The CERF is also included in an AD diagnostic framework that uses categorization, fusion feature synthesis, and feature selection approaches to consider relevant factors. This approach identified abnormal brain regions and AD-causing genes, such as the thalamus, precuneus, insula, and the DAB1 and LRP1B genes. Future research should, however, concentrate on "brain region-gene combinations" and validation using big datasets.

Ruhul Amin Hazarika et al. [17] tried to develop a trustworthy and affordable method for categorizing brain illness using MRI data. Performance comparisons, virtues and demerits, and thorough observations are only a few of the classification methods presented. According to the outcome comparison (about 93.19 percent), the ANN-based categorization technique generates the most compelling conclusions. Obtaining adequate data points from diverse data sources is one of the main challenges. Because the brain's anatomy is so complex, it was difficult to identify the changes. Because of this, efficient preprocessing procedures like removing the brain's skull and segmenting its numerous components are essential but challenging.

K.R. Kruthika et al. [18] created Content-based image retrieval (CBIR) systems to boost prediction rates. These systems combine self-regulating image classification with radiologist expertise. The stages of the disease are categorized using machine learning models such as the Naive Bayes classifier, Support Vector Machine (SVM), and K-nearest Neighbour (KNN) and used PSO, a form of swarm intelligence, and feature selection techniques to portray the structural abnormalities in the brain connected to Alzheimer's disease while it was developing clinically. Cortical and volume thickness features are created as part of a feature set by looking at the feature selection process to include significant features. The AD classification system's developed picture retrieval approach also produced successful results.

Jyoti Islam et al. [19] proposed a deep convolution neural network to classify binary output and detect different stages of AD disease. Comparing the ensemble model and baseline deep CNN models to find the new classification model showed that these models achieve encouraging success when used to diagnose AD using MRI data. The proposed model has been tested on various AD datasets, but the technique has great promise for applying CNN to other fields with sparse data.

Naimul Mefraz Khan et al. [20] used transfer learning to overcome issues like early AD detection. The cutting-edge VGG paradigm is pre-trained using weights from enormous original image sets in this method. They postulate that a robust and illustrated architecture for authentic images combined with applying transfer learning on learned data may increase a model's accuracy while reducing dependency on a sizable training set. Finally, a Class Activation Maps (CAM) that shows a constructed model focused on discriminative image regions associated with neuropathology is supplied. This might be helpful for a doctor's decision-making process.

Zhao Fan et al. [21] Principal component analysis (PCA) was utilized to extract current methodologies. To categorize and identify AD, magnetic resonance imaging (MRI) imaging data, face recognition systems additionally used principal component analysis along with linear discriminant analysis (LDA) and support vector machine

(SVM). The model demonstrated that SVM can retrieve Alzheimer's disorder processes and may also be utilized to investigate ambiguous information in the cloud.

P. Kishore et al. [22] cerebral scans cannot be relied upon as the only indicator of a person's experience level. The proposed structure shows an extensive processing method from a data mining standpoint. This study uses classifiers and various machine learning techniques to prepare the rate and characteristics of Alzheimer's disease. According to earlier research, the Support Vector Machine classifier could only accurately diagnose Alzheimer's disease to a shallow degree. It is necessary to increase precision in light of this. They have used different data categorization techniques to boost the effectiveness of diagnosing the disease, as mentioned earlier, showing that the Support Vector Machine with linear kernel model offers more accuracy than other models.

Jack Albright et al. [23] conducted a study to understand better the use of past and present clinical data to predict a patient's future cognitive status by developing machine learning models that can correlate clinical data gathered from patients at a one-time point with the progression of AD in the future. Numerous machine learning models successfully predicted AD's course in cognitively healthy individuals and people with MCI. Since one of the major causes leading to the frequent failure of AD clinical trials is the inability to identify individuals early, these strategies may help increase the likelihood of finding a therapy for AD.

Manan Binth Taj Noor et al. [24] created a computer-aided brain diagnosis (CABD) system to identify Alzheimer's disease. A variety of feature extraction techniques are used in the process to categorize magnetic resonance imaging (MRI) data. In hospitals, the non-invasive MRI technique is routinely used to check for abnormalities in cognition. Images are acquired using the T2 imaging sequence. Filtering, feature extraction, feature selection based on Student's t-test, and classification based on k-nearest Neighbour (KNN) are some of the quantitative approaches included in the paradigm. Additional feature extraction methods covered in the literature are utilized to conduct comparative research. Our findings suggest that the Shearlet Transform (ST) feature extraction strategy improves Alzheimer's diagnosis compared to other methods, enhancing the CABD system's efficacy.

Yousry AbdulAzeem et al. [25] SVM classification used to forecast the specifics of Alzheimer's. The SVM method only does clustering, separating people who mostly have Alzheimer's disease. This endeavor started with basic, well-known techniques such as Logistic Regression, Decision Tree (DT), Random Forest (RF), Naive Bayes (NB), and three different Support Vector Machines (SVMs) to predict Alzheimer's disease.

U. Rajendra Acharya et al. [26] created a CNN-based end-to-end architecture. The first of the framework's five levels is acquiring MRI data. The training datasets are enhanced in the second layer using the adaptive thresholding and data augmentation approaches. The third layer of the CNN is trained using the cross-validation method. Cross-validation establishes the ideal values for the training parameters to prevent overfitting. The fourth layer makes use of the CNN model. Three convolutional layers comprise the CNN design, with max pooling done after each layer. Two completely coupled layers follow the convolutional layers. Several algorithms are used to carry out the categorization process in the fifth layer.

3. Methodology

3.1 ADNI

We Considered Demographic data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) into account in this investigation. It is a publicly accessible warehouse started in 2003; the investigator is Michael W. Weiner, MD. The ADNI can track the course of MCI and early AD using MRI, PET, other biomarkers, clinical tests, and cognitive evaluations. ADNI Warehouse gathered Data from 50 US and Canadian locations, including participants with mental normality, MCI, and AD. They conducted Follow-up exams every six months.

3.2 ARCHITECTURE

They have applied the suggested methodology in this study to determine whether the patient has a condition accurately. The healthcare expert entered the feature values based on the patient's health report. The information is incorporated into a model that forecasts the likelihood of developing the disease. It becomes challenging to manage large amounts of data to produce the desired outcomes; therefore, creating a model enforced by an algorithm becomes vital. Even if we compare the models, it is critical to understand how each algorithm works to select the method with the highest performance. The whole process is depicted in Figure 1.

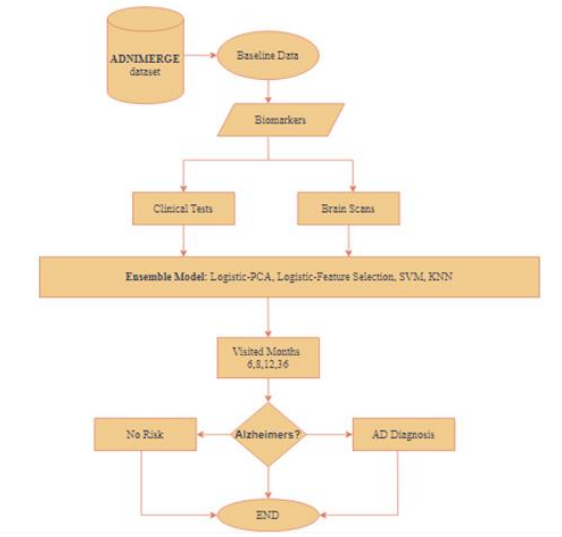

Fig 1: Architecture of AD prediction using ML

Steps involved in the above fig:

Step1: *data collection*- In the first stage, acquired information from the ADNI-> Study-Info-> The ADNIMERGE dataset, which combines and integrates data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) project, is helpful for academics and clinicians researching Alzheimer's disease and related diseases.

Data Types

A wide variety of data types, including clinical, cognitive, neuroimaging, genetic, and biomarker data, are included in the dataset.

- Clinical data often includes personal characteristics, medical background, and clinical evaluations.
- Results from various neuropsychological tests used to evaluate cognitive function are included in the category of cognitive data.
- Neuroimaging data consists of brain scans, such as MRI and PET scans, which reveal information on the structure and function of the brain.
- Genetic information may contain genetic **variants** linked to an increased risk of Alzheimer's disease.
- Data on blood-based and cerebrospinal fluid (CSF) biomarkers pertinent to AD are included in biomarker data.
- ADNIMAGE 113 columns and 14036 rows are present.

S.No	Feature Name	Description
1	RID	Participant roster ID
2	EXAMDATE	date of the clinical examination
3	Month	Months since baseline
4	PTGENDER	Gender: male, Female
5	PTEDUCAT	total years of education
6	AGE	age at baseline visit
Main measures to be predicted		
1	DX	Diagnosis
2	ADAS13	Alzheimer’s Disease Assessment Scale with 13 items, score range 0 to 85
3	Ventricles	
Cognitive test		
1	CDRSB	Sum of Boxes score of the Clinical Dementia Rating Scale.
2	ADAS11	Alzheimer’s Disease Assessment Scale with 11 items, score range 0 to 70
3	MMSE	Mini-Mental State Examination. Max score 30 <25: normal >24: abnormal.

4	RAVLT_immediate	Rey Auditory Verbal Learning Test
MRI Measures		
1	Hippocampus	medial temporal lobe of the brain
2	WholeBrain	total brain size
3	Entorhinal	Adjacent to hippocampus
4	MidTemp	Part of cerebral cortex

Table1: Change in Biomarkers

S.No	Feature Names	Description
1	EXAMDATE.bl	date of the clinical examination
Main measures to be predicted for Baseline		
1	DX_bl	diagnosis
2	ADAS13.bl	Alzheimer's Disease Assessment Scale
3	Ventricles.bl	
Cognitive test for Baseline		
1	CDRSB.bl	Sum of Boxes score of the Clinical Dementia Rating Scale for Baseline
2	ADAS11.bl	Alzheimer's Disease Assessment Scale
3	MMSE.bl	Mini-Mental State Examination.
4	RAVLT.immediate.bl	Rey Auditory Verbal Learning Test
MRI Measures for Baseline		
1	Hippocampus.bl	medial temporal lobe of the brain
2	WholeBrain.bl	total brain size
3	Entorhinal.bl	Adjacent to hippocampus
4	MidTemp.bl	Part of cerebral cortex

Table 2: Baseline biomarkers

Participants are subjected to ADAS-Cog 13 at baseline, 6, 12, and continued annually for CN and MCI subjects for neuropsychological testing. We employed ADAS-cog 13, which comprises activities like number cancellation or mazes, verbal memory, nonverbal memory, praxis, delayed word recall, planning, and executive function assessments—scores on the ADAS-cog 13 range from 0 to 85. In people with AD, the ADAS-cog 13 is more responsive to disease progression than the ADAS-cog 11, while in subjects with pre-dementia symptoms, it is similar to or slightly more responsive [27]. The ADAS-Cog shows superior diagnostic performance in AD patients and predictive solid validity as a screening tool for both MCI and AD [28].

Step 2: Preprocessing- Preprocessing or data exploration. The three phases are as follows:

Data Cleaning: Remove duplicate or incorrect data via data cleaning. Missing values and noisy data prevent ML

classifiers from processing NaN values directly. Therefore, they must transform into numerical values and then substituted for the column's mean. The data cleaning process required this step.

Data Transformation: Combining unstructured and structured data is known as data transformation. Hidden patterns are more straightforward to find.

Data Reduction: Reduces enormous original data sets into more manageable datasets.

Step 3: Splitting- The data is split into two halves.

The test set with 20% of the data used and 80% of the train set validated. Data preprocessing was a crucial step in the process because it enhanced our model's performance in terms of accuracy achieved by hyper-tuning parameters and cross-validation, feature importance.

3.3 HYPERPARAMETER TUNING AND CROSS-VALIDATION

The hyper-parameters used in machine learning techniques typically allow variable algorithm customization during training. To achieve the best performance when applied to examples outside of the training set, varying values of these hyper-parameters result in algorithms with various prediction capabilities. As a performance parameter, the Area Under the Receiving Operating Curve (AUROC) was targeted for improvement. All of the machine learning algorithms created for this study produce a continuous prediction score (range: 0–1; the closer to 1, the subject's estimated conversion risk is).

We employed the frequently used 10-fold cross-validation process to get a reliable performance estimate, repeated ten times. The percentage of converters and non-converters in each fold was stratified (i.e., balanced) during the formation of the folds at random. To obtain a final point estimate of the generalized performance, the 100 algorithm performance estimates provided for each hyper-parameter configuration were averaged. Each machine learning method's hyper-parameter structure that showed the best average cross-validated AUROC was kept.

The optimal hyperparameters for k-NN (number of neighbors), SVM (C and gamma), Random Forest (number of trees and maximum number of features), and AdaBoost (number of estimators) were chosen using cross-validation.

3.4 MODEL VALIDATION

Model validation minimizes the overfitting problem. The ML model is trained through cross-validation, which is also used to assess the model's correctness. It is challenging to develop a noise-free ML model. Cross-validation is helpful for the noise-free model, which separates the entire dataset into n equal sections. To train, the ML model divides each iteration into n-1 parts. The efficiency of the procedure is assessed using the mean of all n-folds.

3.6 FEATURE IMPORTANCE

We used the same five-train/test split protocol to iteratively create logistic regression models with just one feature in the train subsets, and these models were used to generate the continuous prediction scores in the five test subsamples. This allowed us to provide a general ranking of the importance of the predictors used in this study. Finally, the test subsample scores were combined to determine each predictor's AUROC for the entire sample test. As a result, each predictor has an importance meter distinct from the machine learning method applied, and every other predictor is added to the algorithm.

Step4: Classification

Training data were analyzed during the classification phase to determine the new category observations. After completing the abovementioned stages, various machine-learning techniques were used to refine the illness. These included SVM, KNN, Logistic Regression, Naïve Bayes, Random Forest, and Ensemble method.

3.5 CLASSIFIERS

1. K-Nearest Neighbors

It is a simple technique that relies on the notion that data points with similar characteristics will likely belong to the same class. It determines the K-nearest data points in the training dataset to the new data point to classify it. You must provide a predetermined value for the hyperparameter "K." Based on a distance metric, frequently Euclidean distance, these closest neighbors are selected. The new data point is then given the class that appears the most often among these K nearest neighbors. A tiny K could cause forecasts to be noisy, whereas a big K could lead to over-smoothing. Cross-validation or other techniques should be used to find the ideal value of K.

2. Support Vector Machine

The SVM model can be utilized for prediction once it has been trained. A new data point is categorized according to which side of the hyperplane. SVM predicts the target value for regression based on the separation from the hyperplane. SVM can effectively handle non-linear data thanks to the kernel method.

3. Naive Bayes

Naive Bayes determines the likelihood that a new data point will belong to each class when given features. The Bayes theorem and the conditional probabilities computed during training are used to do this. The data point is given the class with the highest posterior probability.

4. Logistic

For situations involving binary and several classes, it is frequently utilized. The logistic function, often known as the sigmoid function, is used in logistic regression to represent the likelihood that an input belongs to a specific class. Any real-valued number can be converted to a number between 0 and 1 using the sigmoid function. Once trained, the model can forecast the likelihood that a new data point will belong to one of the classes. A probability threshold defines a decision boundary (often 0.5). The data point is classified into one class if the anticipated probability is greater than or equal to the threshold; otherwise, it is classified into the other class.

5. Ensemble:

In machine learning ensemble approaches, predictions from various models are combined to produce a more reliable and accurate predictive model. The assumption is that the ensemble can perform better than individual models by maximizing the strengths of several models and minimizing their flaws.

1. Bootstrap Aggregating or bagging

Training numerous instances of the same base model on various portions of the training data is known as bagging. Through bootstrap sampling (sampling with replacement), each subset is obtained. For regression or classification, the final prediction is often the average or mode of the forecasts from each model. A well-known technique called Random Forest bases its models on bagging and decision trees.

- Random Forest

Its a part of supervised learning, outperforms the predecessor Decision Tree by allowing a majority vote across all the trees. A group of decision trees is referred to as a random forest. Both classification and regression problems can be handled by it.

The decision tree's center node is chosen based on the Gini index value, and from there, more child nodes are added to the tree. The decision tree works as a subset of the random tree in this way. A decision is made using the stated constraint or a test set once the nodes have fully developed into a tree. The tree is typically processed using various techniques, including breadth-first search (BFS), depth-first search (DFS), and BFS. The results of a random forest are based on the individual trees; frequently, maximum voting is used to choose the consequences.

2. Boosting:

By sequentially training weak models and giving greater weight to mistakenly predicted instances, boosting creates a robust model. Every new model builds on the mistakes made by the prior one, increasing the accuracy of predictions. Popular boosting algorithms include Gradient Boosting Machines (GBM) and AdaBoost (Adaptive Boosting).

- Gradient Boosting

Extreme Gradient Boosting is the meaning behind the acronym XGBoost. For the fastest and most effective results, gradient-boosted decision trees are used. Gradient boosting machines must be more scalable since model training must be done sequentially. The focus of XGBoost is on performance and speed.

4. Experimentation

Python was used for the implementation, which was done in a "Jupyter Notebook." Following the uploading and cleaning of the data, it was divided into two categories: train and test.

There is a sizable group in the demographic data, with roughly 650 men and about 520 women. The adni_bl dataset, which only contains the initial visit for each patient, indicates 1170 patients in the whole dataset. A baseline diagnosis of cognitively normal (CN) was given to 329 people, subjective memory complaints (SMC) was given to 48 people, early mild cognitive impairment (EMCI) was given to 224 people, and late mild cognitive impairment was given to 395 people. Alzheimer's disease (AD) was given to 174 people on their first visit.

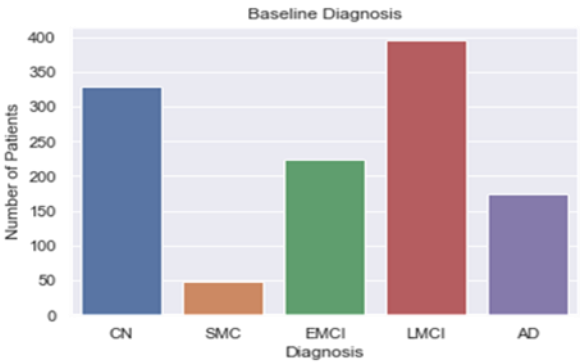

Fig 2: shows count of baseline biomarker diagnosis

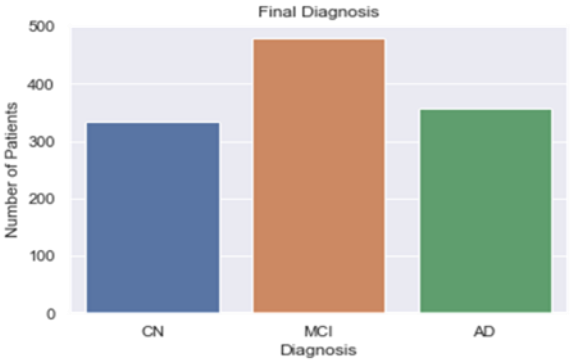

Fig 3: shows count of final diagnosis

Figure 3 shows the merging of SMC with CN, EMCI and LMCI into MCI after a few further visits. The last three diagnoses were CN, MCI, and AD.

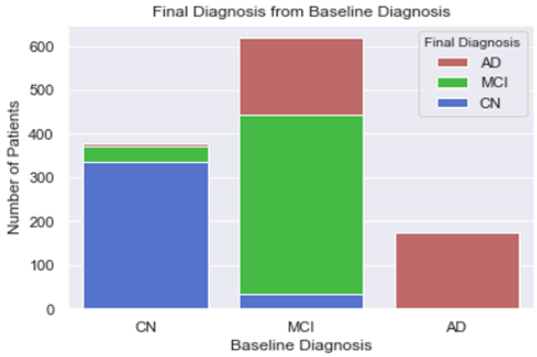

Fig 4: shows comparisons between the first and final diagnoses of the precise change. The majority of the patients (916/1170) did not experience a change in diagnosis. Only 254 patients underwent a diagnostic change, and 36 of those patients saw an improvement.

The number of patients with the baseline diagnosis given on the x-axis (377 as CN, 619 as MCI, and 174 as AD) is represented by the total height of each bar.

• The study began with a CN diagnosis: 37 ended with an MCI diagnosis, only five were diagnosed with AD, while the remaining 335 were completed with no overall change in diagnosis.

• Of the patients initially diagnosed as MCI, 34 ended the study with a diagnosis of CN. Most (409) of baseline MCI patients experienced no overall change in diagnosis, while 176 were diagnosed with AD at their last visit.

• Of the original 174 patients diagnosed with AD, two were diagnosed as MCI at their last visit, with the remaining 172 still diagnosed with AD.

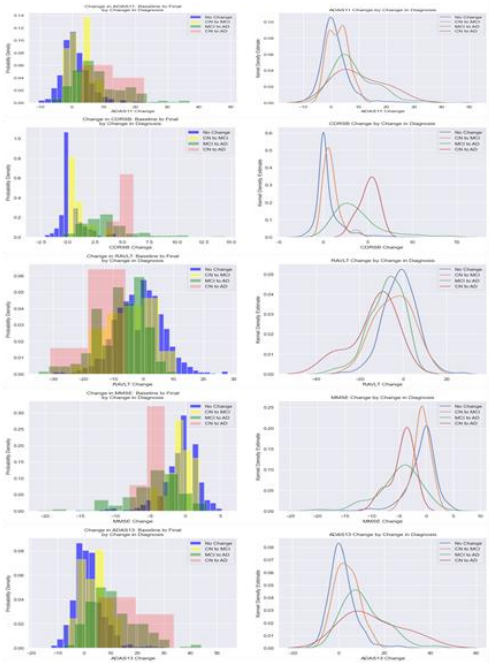

Fig 5: shows histogram plot of change in cognitive functions from initial visit to final examinations by change in diagnosis.

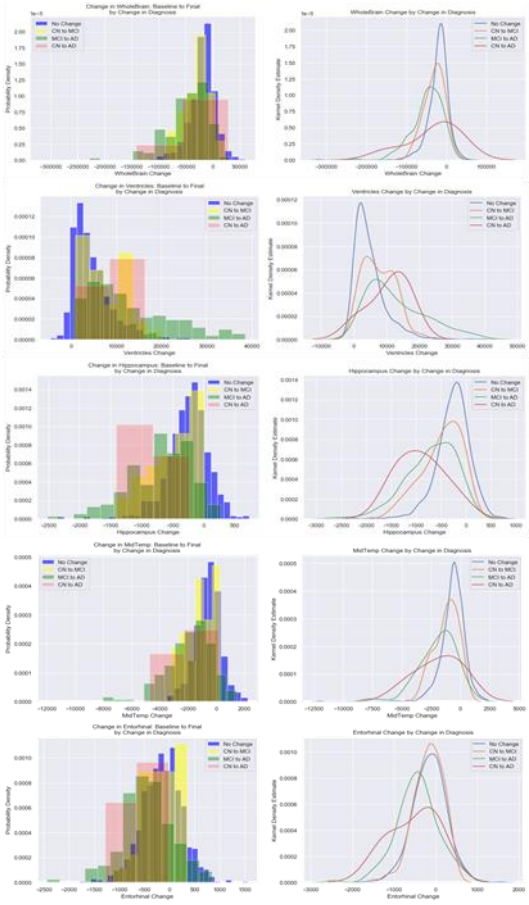

Fig 6: shows histogram plot of change in biological measures from initial visit to final examination by change in diagnosis. Only about 15.8% of patients with a ventricle change diagnosed as CN and 22 patients had changes in WholeBrain, no change in diagnosis and changes in hippocampus.

Several Biomarkers Have Been Discovered as Potential Alzheimer's Disease Predictors. Several biomarkers have been found that may predict AD, even though only 218 patients had a diagnosis that AD was on the horizon throughout the study (with 172 already having AD). As mentioned earlier, the histogram and distribution plots revealed that most of the biomarkers probably had some degree of change, which is cause for alarm. The best cognitive tests that seemed helpful were ADAS11 and ADAS13, and measurements of the hippocampal and middle temporal gyrus also had threshold levels that would indicate a patient is at risk for developing AD.

4.1 STATISTICAL DISTRIBUTION

All biomarkers were identified from the data as mentioned earlier analysis as being excellent candidates for statistical analysis;

• Clinical examinations: MMSE, RAVLT_immediate, ADAS11,

ADAS13, CDRSB

• Hippocampus, ventricles, whole-brain, entorhinal, and mid

temporal lobe scans

Statistical Tests

Given the small amount of data collected, bootstrapping will produce a distribution for the change in each biomarker for patients whose diagnoses remained stable throughout the research.

1. Permutation tests to determine whether or not to split data by gender.

- Null hypothesis: No distinction in biomarker distributions between males and females exists.

- Alt hypothesis: The distributions of one or more biomarkers differ between males and females.

2. Using bootstrapping to calculate the thresholds for biomarker changes related to AD progression.

- The null hypothesis states that all patient groups with different diagnoses (CN to MCI, MCI to AD, CN to AD) will have the same distribution as the group that had no change at the end of the trial. The influence of patients with non-CN diagnoses (MCI to MCI and AD to AD) and no change in diagnosis on the outcomes may need to be investigated.

- The alternative hypothesis is that the distributions for each group will differ sufficiently to allow for the identification of threshold values to denote the need for early MCI/AD treatment or to signal concerns about AD progression. For each diagnosis group, this study aims to give confidence intervals. This will be used to determine how much change warrants worry. Obtaining p-values is less emphasized in this analysis.

3. Bootstrapping calculates thresholds for baseline biomarkers that indicate which research participants will be diagnosed with AD.

- Null hypothesis: Patients with an AD diagnosis won't have baseline biomarkers above non-AD levels.

- Alt hypothesis: a threshold value can be used to separate individuals into groups that received an AD diagnosis by the conclusion of the trial from those who did not.

S.No	**Variables**	**p-value**	**Threshold**	**Progressing CN to MCI %**	**Progressing MCI to AD %**	**Progressing CN to AD %**
Cognitive Functions						
1	CDRSB_DELTA	0.4391	0.60	46	95	100
2	ADAS11_DELTA (Male)	0.0382	1.68	57	81	100
	ADAS11_DELTA (female)		2.05	50	85	100
3	ADAS13_DELTA (male)	0.077	2.14	52	80	100
	ADAS13_DELTA (female)		2.39	63	89	100
4	MMSE_DELTA	0.121	-1.01	27	84	100
5	RAVLT_DELTA	0.166	-1.91	59	74	100
Biological Measures						
1	HIPPOCAMPUS	0.2203	-281.22	59	77	80
2	VENTRICLES (male)	0.0007	5776.6	67	79	100

	VENTRICLES (female)		4675.2	56	84	67
3	WHOLEBRAIN	0.2017	-21287.7	54	73	40
4	ENTORHINAL	0.1248	-176.4	41	73	40
5	MIDTEMP (male)	0.0373	-686.10	67	73	50
	MIDTEMP (female)		-774.2	63	82	67

Table 3: Statistical analysis to explore what amount of change is associated with a change in diagnosis. Some cognitive functions and biological measures in patients defined as CN to MCI, MCI to AD, CN to AD.

Two bootstrapping values were used to determine the threshold values for baseline biomarkers: The 75th quantile for patients without an AD diagnosis:

- Given that only 25% of non-AD patients should have readings higher than this cutoff, this should translate to an average false positive rate of 25%.
- For patients with an AD diagnosis, the 25th quantile. Since 75% of individuals with an AD diagnosis would fall outside of this range, this should translate to an average detection rate of 75%.

S.No	**Variables**	**Male**			**Female**		
		Threshold Range	**DR %**	**FPR %**	**Threshold Range**	**DR %**	**FPR %**
1	CDRSB	1.5	83	15	1.03 to 1.91	86	11.3
2	ADAS11	10.4-11.3	82	21	9-12	90	10
3	ADAS13	17-19.1	86	15	14-20	93	8
4	MMSE	27.2	80	26	26-28	88	10
5	RAVLT	29.1	78	22.4	32.3 to 37.3	91	14
1	HIPPOCAMPUS	6673-6973	66	35	6241-6390	80	21
2	VENTRICLES	33458-50797	45	56	24362-37916	40	53
3	WHOLEBRAIN	1025651-1120009	47	67	928765-979330	56	46
4	ENTORHINAL	3474-3655	67	34	3144-3152	76	25
5	MIDTEMP	19176-20683	57	45	17617-18028	69	30

Table 4: Statistical analysis shows which baseline biomarkers is more predictive power to predict AD of male and female patients based on threshold range identifies detection rate (DR) and false positive rate (FPR).

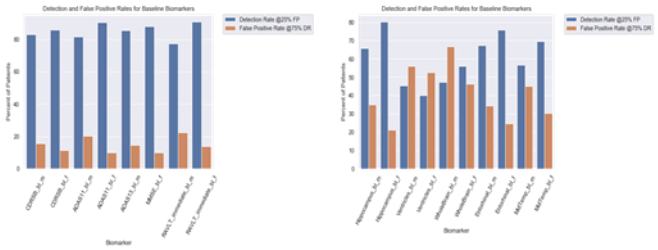

Fig 7: shows how well each biomarker performed at the calculated thresholds. The clinical exams outperformed the brain scans for being able to predict a final diagnosis of AD. The tall blue bar indicates high detection power at a 25% false positive rate and a short red bar indicates low false positive rate at a 75% detection rate.

The starting points appeared promising for predicting AD diagnosis in patients on their initial visit. When taken together, these thresholds—which frequently worked well on their own—will likely have a very high power to detect AD and a low percentage of false positives. While lowering the number of false positives, the threshold values for females demonstrated more predictive potential. To estimate the outer borders of each distribution, the thresholds were obtained by bootstrapping the extreme values from the data. Clinical examinations served as the best baseline biomarkers.

4.2 PERFORMANCE INDICATORS

Performance metrics are the standards used to assess the effectiveness of the machine learning algorithm. Metrics are necessary when comparing the effectiveness of various algorithms. Additionally, it influences how the results are shown. Confusion matrix, accuracy metrics, precision, F1 score, and other performance metrics are assessed, and 5-fold validation is carried out.

Confusion matrix:

The confusion matrix can be generated, the simplest method for evaluating an algorithm's performance. Making a 2X2 table with four components—True Positive, False Positive, True Negative, and False Negative—is how it is done. The actual class is on the x-axis, and the predictive class is on the y-axis in binary classification.

Pre-dicted	Actual 1	Actual 0
1 0		
1 1		
2	True Positive	False Negative
3	False Negative	True Positive

Let's understand the terms: -

1) True positives (TP) – When the output of the actual class is Dementia (1), and the predicted class is Dementia (1), known as true positive (correctly predict positives).

2) True Negatives (TN) - When the output of the actual class is Non-Dementia(0), and the predicted class is Non-Dementia (0), it is known as a true negative (correctly predicts negatives).

3) False Positive (FP) – When the value of the actual class is non-dementia (0), and the predicted class is Dementia (1), known as False Positive (incorrectly predict positives).

4) False Negative (FN) – when the value of the actual class is Dementia (1) and the predicted class is Non-dementia (0), known as False negative (incorrectly predict negatives).

(i)Accuracy: -

It is one of the main important measures of performance that can be determined as the ratio of the number of correct predictions to the sum of all predictions.

Accuracy = $TP+TN/(TP+TN+FP+FN)$

(ii)Precision: -

It can be determined as the ratio of correctly predicted positives to the sum of correct and false predictions.

Precision = $TP/(TP+FP)$

(iii)Recall or sensitivity: -

It can be determined as the ratio of correctly predicted positives to the sum of correctly predicted positives and incorrectly predicted negatives.

Recall = $TP/(TP+FN)$

(iv)Specificity: -

In contrast to the sensitivity, specificity is the number of negatives the ML models return.

Specificity= $TN/(TN+FP)$

(v)F1 Score: -

It can be calculated using the harmonic mean of precision and recall. The value that the F1 score can take is (1,0), with one being the best and 0 being the worst score.

F1= 2* (precision * recall) / (precision + recall)

Improving the Model Performance:

1. It used dimension reduction to decrease model complexity and remove background noise that might hinder model performance.

PCA and feature selection

2. We use ensemble techniques in conjunction with dimension reduction, combining multiple models to improve predictions.

In the below table the principal component analysis for the change in biomarker data shows that 11 principal components are required to explain most of the variance in the data. About 95% of the variance can be explained using 11 principal components, so the models will be run using that number of components. This does not reduce the dimensionality of the data very much and could still be prone to overfitting on further unseen data. The next approach to reducing the dimension in the change in biomarkers data will be to do a feature selection analysis

S.No	model	hyper_params	train_acc	test_acc	Auc	tp	fn	tn	fp	precision	recall	fpr	neg_f1	AD_f1
0	knn	k: 17	0.9	0.9	0.89	28	26	279	8	0.78	0.52	0.03	0.94	0.62
1	svm	C: 0.6, gamma: 0.0835	0.87	0.85	0.91	46	8	244	43	0.52	0.85	0.15	0.91	0.64
2	RF	trees: 151, max_feats: 11	1	0.87	0.88	29	25	268	19	0.6	0.54	0.07	0.92	0.57
3	AdaBoost	num_estimators: 51	0.94	0.88	0.88	32	22	269	18	0.64	0.59	0.06	0.93	0.62
4	logreg	None	0.86	0.85	0.92	45	9	246	41	0.52	0.83	0.14	0.91	0.64
5	bayes	None	0.84	0.88	0.89	32	22	267	20	0.62	0.59	0.07	0.93	0.6

Table 5: Change in Biomarkers Dimension Reduction: Principal Components Analysis

S. No	model	hyper_params	train_acc	test_acc	auc	tp	fn	tn	fp	precision	Recall	fpr	neg_f1	AD_f1
0	knn	k: 33	0.88	0.88	0.95	80	26	230	15	0.84	0.75	0.06	0.92	0.8
1	svm	C: 0.7, gamma: 0.0847	0.89	0.86	0.94	96	10	205	40	0.71	0.91	0.16	0.89	0.79
2	RF	trees: 101, max_feats: 11	1	0.88	0.94	86	20	223	22	0.8	0.81	0.09	0.91	0.8
3	AdaBoost	num_estimators: 81	0.96	0.86	0.91	83	23	219	26	0.76	0.78	0.11	0.9	0.77
4	logreg	None	0.87	0.88	0.96	98	8	212	33	0.75	0.92	0.13	0.91	0.83
5	bayes	None	0.7	0.87	0.94	75	31	230	15	0.83	0.71	0.06	0.91	0.77

Table 6: Baseline Biomarkers Dimension Reduction: Principal Components Analysis

The baseline models with 11 principal components performed relatively well, with SVM and logistic models achieving 91% detection rates, with 16% and 13% false positive rates, respectively. This was not a significant improvement on the model with fewer features, though the logistic model appears slightly better.

Feature Selection:

This tool will be used to search for three circumstances to consider excluding features:

1. Co-linearity: If found, linear regression will quantify the relationship and determine whether or not to exclude certain features.

2. Zero Importance Features: This tool uses Gradient Boosting to search for zero-importance features, which do not contribute to predicting the target.

3. Low Importance Features: The Gradient Boosting method will assign normalized importance values based on the amount of variance in the target explained by each feature.

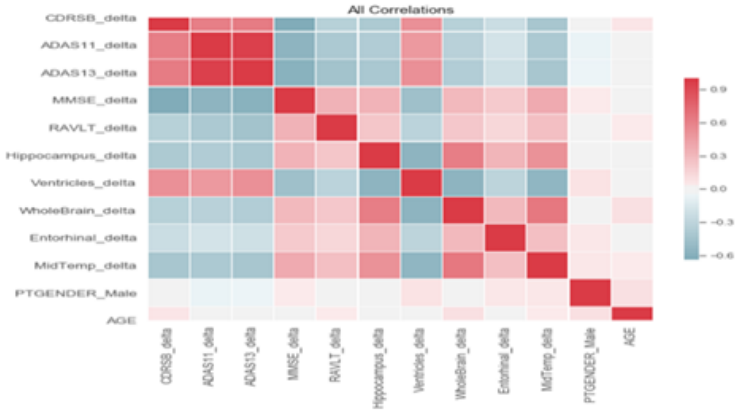

Fig 9: shows correlations between clinical tests and brain scans. ADAS13_delta is collinear with ADAS11_delta. So, ADAS13_delta will be removed from the model.

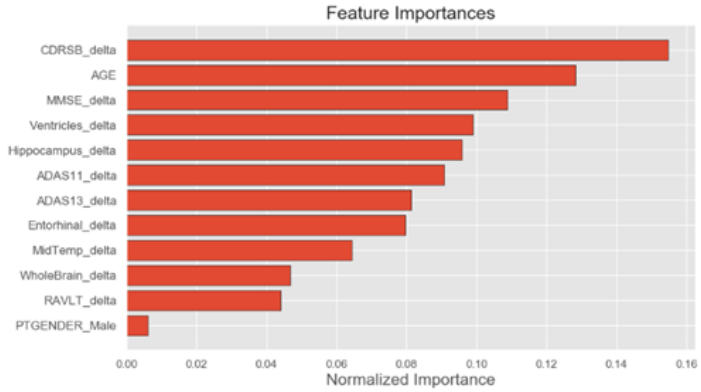

Fig 10: shows feature importance between clinical tests and brain scans. PTGENDER has a very low significance with the model. As statistical analysis suggested, gender impacted specific biomarkers and did not influence whether or not a patient was more/less likely to be diagnosed with Alzheimer's during this study.

PCA improved the models a little at the expense of loss of interpretability in the features. Feature selection identified some features that should be removed from the machine learning analysis because they do not contribute significantly to the model performance. For both the change in biomarkers and the baseline biomarkers, ADAS13 was removed due to high collinearity with ADAS11. PTGENDER was removed because although it appears to contribute to the magnitude of values for the features, it did not factor in determining whether or not a patient would be diagnosed with Alzheimer's disease in this study.

S.No	model	hyper_params	train_acc	test_acc	auc	tp	fn	tn	fp	precision	Recall	fpr	neg_f1	AD_f1
0	knn	k: 18	0.89	0.89	0.9	26	28	278	9	0.74	0.48	0.03	0.94	0.58
1	svm	C: 1.0, gamma: 0.0999	0.89	0.85	0.91	45	9	245	42	0.52	0.83	0.15	0.91	0.64

2	RF	trees: 131, max_feats: 1	1	0.89	0.92	28	26	275	12	0.7	0.52	0.04	0.94	0.6
3	AdaBoost	num_estimators: 31	0.93	0.89	0.92	38	16	265	22	0.63	0.7	0.08	0.93	0.67
4	logreg	None	0.86	0.86	0.91	47	7	247	40	0.54	0.87	0.14	0.91	0.67
5	bayes	None	0.84	0.86	0.89	34	20	260	27	0.56	0.63	0.09	0.92	0.59

Table 7: Change in Biomarkers Dimension Reduction: Feature Selection Class.

The reduced models performed very similarly to the full feature models. The best models were again SVM and logistic regression with detection rates (recall) of 83% and 87%, respectively. The false positive rates were also comparable at 15% and 14% respectively.

S. No	**model**	**hyper_params**	**train_acc**	**test_acc**	**auc**	**tp**	**fn**	**tn**	**fp**	**precision**	**Recall**	**fpr**	**neg_f1**	**AD_f1**
0	knn	k: 7	0.89	0.87	0.94	82	24	225	20	0.8	0.77	0.08	0.91	0.79
1	svm	C: 1.25, gamma: 0.1011	0.91	0.85	0.94	95	11	204	41	0.7	0.9	0.17	0.89	0.79
2	RF	trees: 171, max_feats: 10	1	0.89	0.94	82	24	230	15	0.85	0.77	0.06	0.92	0.81
3	AdaBoost	num_estimators: 81	0.92	0.88	0.91	83	23	225	20	0.81	0.78	0.08	0.91	0.79
4	logreg	None	0.87	0.88	0.96	97	9	213	32	0.75	0.92	0.13	0.91	0.83
5	bayes	None	0.77	0.87	0.95	91	15	214	31	0.75	0.86	0.13	0.9	0.8

Table 8: Baseline Biomarkers Dimension Reduction: Feature Selection Class.

Performance was on par with the full-featured models. The SVM model scored 90% detection, and the logistic model achieved 92%. These models had 17% and 13% false favorable rates, respectively. The logistic model stands out here with a high detection rate and relatively low false positive rate.

This model aims to create an ensemble with the best-performing models to get the highest detection rate possible. A bonus would be to decrease the false positive rate. Still, the primary goal is identifying patients at higher risk for Alzheimer's, so this section will focus on optimizing the detection rate (recall).

Grand Ensemble Models:

S.No	**model**	**train_acc**	**test_acc**	**tp**	**fn**	**tn**	**Fp**	**recall**	**Fpr**
1	Ensemble_deltas	0.95	0.92	44	10	238	49	0.9	0.11
2	Ensemble_bl	0.94	0.9	92	14	213	32	0.89	0.12

Table 9: Grand Ensemble Model for Change in Biomarkers and Baseline.

Fig 11: Change in Biomarkers Analysis

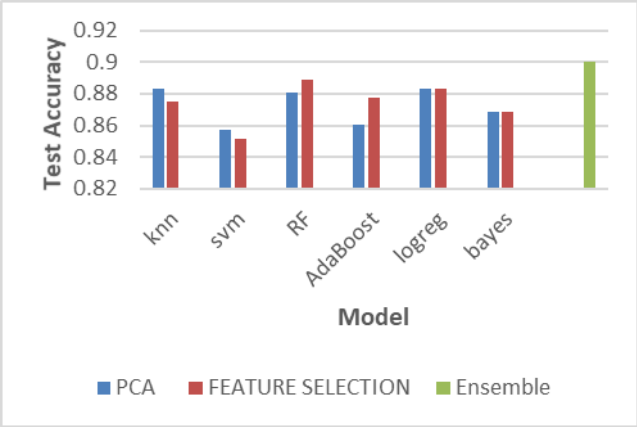

Fig 12: Baseline Biomarkers Analysis

5. Conclusion and Future Scope

The accuracy generated is used to compare the performances of the various Ensemble ML models. They were standardized to ensure the values fit into the ML models without difficulty. This study focused on the detection rate and false positive rate of a disease determined from ADNI-ADNIMERGE demographic data using a variety of machine learning techniques, including KNN, SVM, RF, NB, LOGISTIC, and Ensembled: LOGISTIC-PCA, SVM, KNN as the final algorithm with feature selection and hyper-tuning parameter optimization. Performed a comparison analysis between machine learning methods and ensemble model. Ensemble model showed best results with change in biomarker and baseline biomarker of disease detection rate, false positive rate and test accuracy 92%, 90% of AD respectively. Future improvements to the framework model could be made by using larger datasets, hybrid models (ML & CNN), and better performances.

References

[1] Fadi Thabtah, Swan Ong and David Peebles (2022). Detection of Dementia Progression from Functional Activities Data using Machine Learning Techniques. Pure.hud.ac.uk.

[2] Erik D. Huckvale, Matthew W. Hodgman, Brianna B. Greenwood (2021). Pairwise Correlation Analysis of the Alzheimer's Disease Neuroimaging Initiative (ADNI) Dataset Reveals Significant Feature Correlation. Genes (Basel). 2021 Nov; 12(11): 1661. Published online 2021 Oct 21. doi: 10.3390/genes12111661, PMCID: PMC8619902, PMID: 34828267

[3] Massimiliano Grassi, Nadine Rouleaux, Daniela Caldirola (2019). A Novel Ensemble-Based Machine Learning Algorithm to Predict the Conversion From Mild Cognitive Impairment to Alzheimer's Disease Using Socio-Demographic Characteristics, Clinical Information, and Neuropsychological Measures. Front Neurol. 2019; 10: 756. Published online 2019 Jul 16. doi: 10.3389/fneur.2019.00756. PMCID: PMC6646724. PMID: 31379711.

[4] Jayant Prakash, Velda Wang, Robert E. Quinn (2021). Unsupervised Machine Learning to Identify Separable Clinical Alzheimer's Disease Sub-

Populations. Journals Brain Sciences Volume 11 Issue 8 10.3390/brainsci11080977.

[5] Muhammad Irfan, Seyed Shahrestani and Mahmoud Elkhodr (2023). Enhancing Early Dementia Detection: A Machine Learning Approach Leveraging Cognitive and Neuroimaging Features for Optimal Predictive Performance. Appl. Sci. 2023, 13, 10470. https://doi.org/10.3390/app131810470 https://www.mdpi.com/journal/applsci.

[6] Afreen Khan, Swaleha Zubair (2022). Development of a three tiered cognitive hybrid machine learning algorithm for effective diagnosis of Alzheimer's disease. Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 10, Part A, November 2022, Pages 8000-8018.

[7] M. Tanveer, B. Richhariya and R. U. Khan (2020). Machine learning techniques for the diagnosis of Alzheimer's disease. ACM Transactions on Multimedia Computing, Communications, and Applications, 16(1s), 1–35.

[8] Esther E. Bron, Stefan Klein, Janne M. Papma (2021). Cross-cohort generalizability of deep and conventional machine learning for MRI-based diagnosis and prediction of Alzheimer's disease. NeuroImage: Clinical, 31, 102712.

[9] Jun Pyo Kima, Jeonghun Kimb, Yu Hyun Parka, Seong Beom Parka (2019). Machine learning based hierarchical classification of frontotemporal dementia and Alzheimer's disease. NeuroImage: Clinical, 23, 101811.

[10] Mingxia Liuy, Jun Zhangy, Dong Nie, Pew-Thian Yap, Dinggang Shen, Fellow (2018). Anatomical landmark based deep feature representation for MR images in brain disease diagnosis. IEEE Journal of Biomedical and Health Informatics, 22(5), 1476–1485.

[11] Binny Naik, Ashir Mehta and Manan Shah (2020). Denouements of Machine Learning and Multimodal Diagnostic Classification of Alzheimer's disease. Visual Computing for Industry, Biomedicine, and Art, 3(1)

[12] Samaneh Abolpour Mofrad, Arvid Lundervold, Alexander Selvikvåg Lundervold, (2021). A predictive framework based on brain volume trajectories enabling early detection of Alzheimer's disease. Computerized Medical Imaging and Graphics, 90, 101910.

[13] F.J. Martinez-Murcia, A. Ortiz, J.M. Gorriz, J. Ramirez and D. Castillo-Barnes (2020). Studying the manifold structure of Alzheimer's disease: A deep learning approach using convolutional autoencoders. IEEE Journal of Biomedical and Health Informatics, 24(1), 17–26.

[14] Suhad Al-Shoukry, Taha H. Rassem and Nasrin M. Makbol (2020). Alzheimer's diseases detection by using Deep Learning Algorithms: A mini-review. IEEE Access, 8, 77131–77141.

[15] Behnaz Ghoraani, Lillian N. Boettcher, Murtadha D. Hssayeni (2021). Detection of mild cognitive impairment and Alzheimer's disease using dual-task gait assessments and machine learning. Biomedical Signal Processing and Control, 64, 102249.

[16] Xia-an Bi, Xi Hu, Hao Wu, and Yang Wang (2020). Multimodal Data Analysis of Alzheimer's disease based on Clustering Evolutionary Random Forest. IEEE Journal of Biomedical and Health Informatics, 24(10), 2973–2983.

[17] Ruhul Amin Hazarika, Ajith Abraham, Samarendra Nath Sur, Arnab Kumar Maji & Debdatta Kandar (2021). Different techniques for Alzheimer's disease classification using brain images: A study. International Journal of Multimedia Information Retrieval, 10(4), 199–218.

[18] K.R. Kruthika, Rajeswari, H.D. Maheshappa (2019). Multistage classifier-based approach for Alzheimer's disease prediction and retrieval. Informatics in Medicine Unlocked, 14, 34–42.

[19] Jyoti Islam and Yanqing Zhang (2018). Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional Neural Networks. Brain Informatics, 5(2).

[20] Naimul Mefraz Khan, Nabila Abraham, Marcia Hon (2019). Transfer learning with intelligent training data selection for prediction of Alzheimer's disease. IEEE Access, 7, 72726–72735.

[21] Zhao Fan, Fanyu Xu, Xuedan Qi, Cai Li, Lili Yao (2019). Classification of Alzheimer's disease based on Brain MRI and machine learning. Neural Computing and Applications, 32(7), 1927–1936.

[22] P. Kishore, Ch. Usha Kumari, M.N.V.S.S. Kumar, T. Pavani (2021). Detection and analysis of alzheimer's disease using various machine learning algorithms. Materials Today: Proceedings, 45, 1502–1508.

[23] Jack Albright. (2019). Forecasting the progression of alzheimer's disease using neural networks and a novel preprocessing algorithm. Alzheimer's & Dementia: Translational Research & Clinical Interventions, 5(1), 483–491.

[24] Manan Binth Taj Noor, Nusrat Zerin Zenia, M Shamim Kaiser, Shamim A Mamun and Mufti Mahmud (2020). 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).

[25] Yousry AbdulAzeem, Waleed M. Bahgat, Mahmoud Badawy (2021). A CNN based framework for classification of alzheimer's disease. Neural Computing and Applications, 33(16), 10415–10428.

[26] U. Rajendra Acharya, Steven Lawrence Fernandes, Joel En WeiKoh & Edward J. Ciaccio (2019). Automated detection of alzheimer's disease using brain MRI images– a study with various feature extraction techniques. Journal of Medical Systems, 43(9).

[27] Kueper, J. K., Speechley, M., & Montero-Odasso, M. (2018). The Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog): Modifications and responsiveness in pre-dementia populations. A narrative review. Journal of Alzheimer's Disease, 63(2), 423–444. https://doi.org/10.3233/JAD-170991

[28] Park, S.-H., & Han, K. S. (2022). Is the Alzheimer's Disease Assessment Scale-Cognitive Subscale useful in screening for mild cognitive impairment and Alzheimer's disease? A systematic review. Current Alzheimer Research, 19(3), 202–211. https://doi.org/10.2174/1567205019666220404104854