Analysis of Machine Learning Models to Automate the Early Detection of Alzheimer Disease

Sardar Un Nisa[1], Azaz Ahmed Kiani [1*], Maria Hilal[1], Saim Amjid[1], Moeed Ahmed[1], Shahzaib Ishtiaq[1]

[1] National University of Modern Languages, Rawalpindi, Pakistan

*** Correspondence:** azaz.kiani@numl.edu.pk.

Citation | Nisa. S. U, Kiani. A. A, Hilal. M, Amjid. S, Ahmed. M, Shahzaib. I, "Analysis of Machine Learning Models to Automate the Early Detection of Alzheimer's Disease", IJIST, Vol. 7 Issue. 1 pp 322-335, Feb 2025

Received | Jan 10, 2025 **Revised |** Feb 14, 2025 **Accepted |** Feb 15, 2025 **Published |** Feb 18, 2025.

Alzheimer's disease is an advanced neurological illness that primarily affects those over 65. It is characterized by memory loss and cognitive deterioration. Although there isn't a known cure, early intervention can greatly delay the disease's progression, which emphasizes how crucial a prompt and precise diagnosis is. Early-stage identification is still a difficult and time-consuming procedure, though. This study uses machine learning (ML) to improve and speed up Alzheimer's disease detection. The National Alzheimer's Coordinating Center (NACC) dataset, which consists of clinical and genomic data, was subjected to three ML algorithms: Elastic Net Classifier (ENC), Random Forest (RF), and Artificial Neural Network (ANN). Unlike established methodologies that largely rely on Magnetic Resonance Imaging (MRI) paired with other modalities, this research highlights the utilization of limited datasets and comparatively underexplored clinical-genomic data. The models were trained and assessed using the Scikit-learn and TensorFlow frameworks. With an accuracy, F1 score, and recall of 92%, ANN outperformed the other models, indicating its potential for early Alzheimer's identification. This study demonstrates the feasibility of addressing difficulties in early-stage Alzheimer's diagnosis by combining clinical and genomic data with machine learning algorithms.

Keywords: Alzheimer's Disease, Machine Learning, Classification Algorithms, Artificial Intelligence in Healthcare

Introduction

Alzheimer's disease (AD), is a progressive neurodegenerative illness that affects the brain and the risk increases with age. People over the age of 65 are at higher risk of developing AD. The disease is characterized by the gradual degeneration of brain cells, leading to memory loss, cognitive decline, and difficulty performing daily activities. AD is the most common cause of dementia, a broader term that describes a decline in cognitive function severe enough to interfere with daily life. As life expectancy rises due to advancements in healthcare, dementia is becoming more and more common worldwide. Current research estimates that dementia affects over 55 million people worldwide, and studies predict this number will rise significantly in line with projected population growth rates. This increases existing concerns for patients, their caregivers, and healthcare systems.

One of the symptoms of dementia is the loss of thinking, memory, and decision-making skills. Alzheimer's disease stands as the most prevalent form of dementia. Although there is currently no cure for Alzheimer's disease, research suggests that its development is influenced by a combination of environmental factors, lifestyle choices, and genetics. Early intervention has been demonstrated as the best strategy to control symptoms and slow the progression of the disease. Cognitive training, exercise, and a healthy diet can improve the quality of life. The recommended treatments aim to slow the progression of cognitive impairment, while early diagnosis allows patients and their families to develop proactive plans and access necessary resources. However, diagnosing AD is challenging due to its gradual onset and overlapping symptoms with other neurological conditions [1]. A comprehensive clinical evaluation for Alzheimer's disease involves extensive neurological and physical testing, which can be both time-consuming and costly. The analysis of Magnetic Resonance Imaging (MRI) scans forms the basis of current diagnostic practices. For instance, the Multi-Slice Multi-Echo (MSME) score assessment requires manual evaluation of thousands of brain tissue slides. However, this procedure is expensive, time-consuming, and frequently unavailable to many patients. As a result, novel strategies to expedite the diagnostic procedure without sacrificing accuracy are urgently needed [2][3].

Advancements in technology present new opportunities for improving AD diagnosis. ML is one of the quickest ways to expedite and enhance Alzheimer's disease detection. Recent developments in ML have shown promise in analyzing complex data, finding patterns, and distinguishing between AD and normal cognition [4]. However, despite these advantages, there are certain drawbacks as well. Clinical applications of ML frequently prioritize interpretable statistical models, like linear regression, over more complex approaches because of worries about explainability and transparency. Although deep learning and other advanced ML models can provide higher accuracy, their opaque decision-making processes prevent widespread clinical adoption. Closing this gap between interpretability and accuracy is essential to achieve the peak performance of ML in Alzheimer's diagnosis along with care [5]. Therefore, the primary goal of this study is to overcome these obstacles by utilizing cutting-edge ML algorithms to identify Alzheimer's disease with clinical and genomic data from NACC. This study aims to enhance early detection, boost diagnostic effectiveness, and add to the expanding corpus delving into Alzheimer's disease by utilizing ML techniques like ANN, RF, and ENC. The results show that ANN, with an accuracy, F1 score, and recall of 92%, outperformed the other models, indicating its potential for early Alzheimer's identification.

Objectives:

The primary objectives of this study are as follows:

- To explore how Machine Learning methods identify early signs of Alzheimer's disease when interventions may be more effective.

- To conduct a literature review of existing research about Alzheimer's detection using ML techniques.
- To compare results obtained by different algorithms on clinical and genomic data and determine the most effective approach.
- To evaluate the feasibility of integrating ML-based diagnostic methods into clinical workflows to assist healthcare professionals in early AD detection

Novelty:

By utilizing clinical and genomic data from the underutilized NACC dataset, this paper presents a novel, data-efficient method of detecting Alzheimer's disease that departs from the traditional dependence on MRI-based diagnostics. This work addresses the accessibility and cost constraints of traditional methods by showing that high diagnostic accuracy can be attained through non-imaging modalities, in contrast to earlier research that mainly relies on large-scale imaging datasets. By systematically applying and comparing three ML algorithms ENC, RF, and ANN, this study not only supports the predictive power of clinical-genomic data but also shows ANN's better performance, attaining a 92% accuracy, F1 score, and recall. A crucial feature of this study resides in its capacity to generate strong results utilizing a restricted dataset, a typical constraint in Alzheimer's research. This effectiveness highlights how scalable and useful the model is in actual clinical settings, where access to huge imaging datasets may be limited. Furthermore, a more comprehensive diagnostic framework is provided by the incorporation of genomic data, which provides deeper insights into hereditary risk factors. By showing that ML combined with clinical-genomic data can compete with, and even outperform, conventional imaging-based methods, our work establishes a new benchmark for Alzheimer's detection and provides a quicker, more affordable, and just as accurate option for early-stage diagnosis.

Literature Review:

Advancements in computational methods have significantly contributed to the early detection and diagnosis of Alzheimer's disease. Researchers have explored machine learning (ML) approaches that leverage large datasets to improve predictive accuracy and efficiency. ML models can process diverse data sources, including clinical records, genomic information, and neuroimaging scans, enabling automated feature extraction and classification. Venugopalan et al. [6] discussed a deep learning-based multi-modal analysis to improve Alzheimer's Disease prediction results. The model received training data from associations of clinical information and genomic data along with Magnetic Resonance Imaging (MRI) inputs. This research employed Denoising Auto-Encoders to extract features from both clinical data with genetic components, while simultaneously employing 3D Convolutional Neural Networks (CNNs) to process MRI data. The research was based on data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Deep models demonstrated superior performance through studies that compared their success against shallow methods including SVMs, Decision Trees, Random Forests, etc. The work also demonstrated that multi-modal approaches performed better than unimodal approaches. Their findings outperformed shallow models despite facing limitations due to scarce research data. The research by Shagun et al. [7] utilized 6400 and 6330 MRI images published on Kaggle. Feature extraction was performed using VGG16, while a Neural Network conducted the predictive analysis. The evaluation metrics consisted of accuracy, precision, recall, and AUC together with the F1-score. The proposed model demonstrated an accuracy of 90.4% on dataset 1 and achieved 71.1% accuracy on dataset 2. Murugan et al. [8] proposed a custom model named DEMNET. The MRI dataset used in their study was obtained from Kaggle. To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was employed for oversampling. The researchers distributed their available data into three sections consisting of 80% training data and 10% allocated for validation purposes along with another 10% dedicated for testing. They selected a Convolutional Neural Network (CNN) as their model

architecture. Performance evaluation included accuracy, precision, recall, F1 score, AUC, and Cohen's kappa. Without SMOTE, the framework achieved 96% training accuracy but only 78% validation accuracy. SMOTE techniques lead to a model reaching 99% accuracy while producing 94% validation accuracy. Test results from the DEMNET model on ADNI data produced an accuracy rate of 84.83%.

The research conducted by Diogo et al. [9] focuses on building an MRI-based multi-diagnostic classification biomarker through a taxonomy that utilizes multiple classifiers together with voting procedures. Both datasets for this study were sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI) alongside the Open Access Series of Imaging Studies (OASIS). A combined dataset from ADNI and OASIS enabled the model to deliver a 90.6% balanced accuracy in binary classification tasks. The model reached a balanced accuracy of 62.1% for multi-class diagnosis when analyzing data from the ADNI database. Cognitive scores demonstrated the potential to enhance prediction accuracy levels. The research of T. M. Ghazal et al. [10] applied transfer learning to discover Alzheimer's disease at an early stage. The authors named the proposed model, ADDTLA. The model uses MRI scans as data for training and testing. The dataset containing MRI images was obtained from Kaggle. The MRI images are passed through the preprocessing layer which changes their dimensions. After preprocessing the images are transferred to the model for training in the application layer. The application layer consists of a modified version of AlexNet. For transfer learning, all layers of AlexNet, except for the last three, are extracted. The assessment incorporates sensitivity, specificity, precision, and accuracy alongside False negative rate (FNR), False positive rate (FPR), miss rate, F1 Score, Likelihood Ratio Positive (LRP), and Likelihood ratio negative (LRN). The most accurate results were achieved at 91.7% through the completion of 40 training epochs. Transfer learning with feature freezing allows Saeeda Naz et al. [11] to detect Alzheimer's disease. Alzheimer's disease research uses MRI scans that derive their data from the ADNI database. The researchers utilized 11 pre-trained neural network models including AlexNet, GoogLeNet and VGG16/19, ResNet-18/50/101, MobileNetV2, InceptionV3, Inception-ResNet-V2, and DenseNet201. They used three types of classification: MCI-AD, AD-CN and MCI-CN. Different layers from VGGNet FC6 produced 99.26% accuracy while AlexNet's conv5 layer achieved the lowest performance at 71.48%. Future work will involve merging different network layers, whenever feasible, to enhance performance in Alzheimer's Disease (AD) classification. The authors plan to fuse various layers to improve AD classification in future studies. Esther E. Bron et al. [12] assess the generalizability of Alzheimer's disease prediction drawn from MRI using a CNN and a Support Vector Machine. The data undergoes two stages of pre-processing operations: minimal and extensive. Data for 1715 patients was obtained from ADNI while 557 patients were comprised from Health-RI Parelsnoer Neurodegenerative Diseases Biobank (PND). AUC and accuracy were used as performance measures. The AUC for SVM on the PND dataset was 0.896, while for CNN it was 0.876. SVM and CNN used different brain regions in the classification process, both could be combined to make a hybrid model for better performance. Scientists Taeho Jo et al. [13] combined CNN with Long Short-Term Memory (LSTM) and attention mechanisms in an LSTM-CNN model for Alzheimer's disease classification using genomic data analysis. The dataset was sourced from ADNI. To determine optimal fragment size researchers split genomic sections into non-overlapping chunks that ranged from 10 to 200 Single Nucleotide Polymorphisms (SNPs). The researchers measured performance by using accuracy scores. Results from the CNN model demonstrated the best accuracy at 75%. Modupe Odusami et al. [14] used MRI and Positron Emission Tomography (PET) images which were retrieved from the ADNI dataset. The classification process employs a Vision Transformer (ViT). The integrated model reached 98.5% accuracy when analyzing fusions between brain images. Performance testing on single source data remained unavailable while the available dataset reached low capacity. The prediction of Alzheimer's disease employs classifiers including

SVM Random Forest (RF) and Elastic Net through Abhibhav Sharma et al. [15]. The RF and Lasso Regression approach serves for the feature selection process. Genomic data from NCBI-GEO, a functional genomics data repository, served as the foundation for this investigation. Data integration was also performed to increase the sample size. Accuracy, SEN, SPE, Precision, and Mathew's Correlation were used as performance measures. Elastic Net performed the best in most tests. With the Lasso feature selection method and a focus on the prefrontal cortex region of the brain, Elastic Net achieved an accuracy of 100%. Table 1 provides a condensed overview of research findings from the literature review process.

Table 1 Related Literature Review Summary

Sr. No	Reference	Date of Publication	Strengths/Features/Purpose	Limitations
1	[6]	2021	Used a Multi-modality DL model to improve AD detection accuracy Developed a method of interpreting DL models	Limited dataset size
2	[7]	2022	Early identification of Alzheimer's Disease using MRI images. Neural network model with VGG16 feature extractor. Outperformed state-of-the-art models.	More datasets could be collected for better training.
3	[8]	2021	Used a CNN made from scratch SMOTE was used to address class imbalance ReLU was used as the activation function for the CNN	Class imbalance in the dataset
4	[9]	2022	MRI scans were used Two independent data collections serve as training and testing grounds for the model (ADNI and OASIS) Several classifiers are used Potential for clinical applicability is evaluated	The predictive capacity could have been improved through MCI Cognitive scores but this data type existed exclusively in one of the sources.
5	[10]	2021	Used transfer learning for detection MRI scans were used for training and testing Multiple epochs were used to improve the accuracy	Convolutional layers could be fine-tuned Other datasets could be used
6	[11]	2021	AlexNet, GoogLeNet, VGG16/19, ResNet-18/50/101, MobileNetV2, InceptionV3, Inception-ResNet-V2 and DenseNet201 were used as pre trained models MRI images were used which were obtained from ADNI	A fusion of layers could be applied for more robustness

			Three types of binary classifications were performed Accuracy was used as the performance measure	
7	[12]	2021	Use structural MRI data for training and testing Use CNN and SVM as models The dataset used was obtained from ADNI Two types of preprocessing were used: minimal and extensive ACC and AUC were used as performance measures	Models could be combined to form a hybrid model Some patients may have been misdiagnosed
8	[13]	2022	Genotypes were used as data obtained from ADNI CNN, LSTM+CNN, Attention, and LSTM were used as models Accuracy, AUC, and Standard Deviation were used as performance measures	Did not target the early stages of Alzheimer's disease
9	[14]	2023	Fused MRI and PET images. ADNI dataset is used. Used Visual Transformer model for classification	Limited dataset. Fusion parameters could be further optimized. Model performance was not evaluated on just MRI and PET images without fusion.
10	[15]	2021	SVM, RF, and Elastic Net classifiers were used as models. RF and Lasso Regression methods are used for feature selection. Data from the functional genomics data repository 'NCBI-GEO' served as the research basis for this study. Accuracy, SEN, SPE, Precision, and Mathew's Correlation were used as performance measures. New biomarkers were identified	The new biomarkers identified in this study need to be tested

Literature review reveals that existing models suffer from lack of data [6] [7] [14] or only use a single type of data [7] [8] [9] [10] [11] [12] [13] [14] which limits their performance. Additionally, comparatively less research has been done with clinical and genomic data. The proposed model will work with a large amount of data and multi-modality with clinical and genomic data to overcome these problems.

Material and Methods:

The principal data source for this study is the NACC dataset [16], which consists of 180,004 data cases from various clinical settings in the United States and contains a wide range

of clinical and genomic data, including patient demographics, cognitive tests, and genetic markers; each patient's outcome is classified into one of three classes: AD, cognitively normal (meaning no AD), or cognitive impairment not classified as AD; the dataset contains both numerical and categorical variables; important clinical characteristics include Age, Years of Education, Gender, Clinical Dementia Rating (CDR), and Mini-Mental State Examination (MMSE) scores; the genomic component is the APOE Genotype, a known genetic risk factor for Alzheimer's disease. To simplify the model and lower dimensionality, other factors like medical history, dementia in the family, and results from other cognitive tests were first taken into consideration but later eliminated throughout the feature selection process.

Using a domain-driven methodology, feature selection was done by hand with an emphasis on factors that are clinically important in the development and diagnosis of Alzheimer's disease. This required looking through the body of research, clinical recommendations, and professional judgment to determine which characteristics were most closely linked to the development and course of Alzheimer's disease. Because they didn't improve model performance in early tests, redundant and irrelevant features were eliminated. Age, Years of Education, Gender, CDR, MMSE, and APOE Genotype were selected as the final feature set because they balance clinical relevance and predictive power, keeping the model effective and interpretable.

To guarantee compliance with machine learning methods, the data pretreatment procedures involved addressing missing values, encoding category variables, and normalizing numerical characteristics. To properly assess model performance and resolve any potential imbalances in the data, the dataset was then divided into training and testing sets while preserving the integrity of the class distributions. Figure 1 depicts the research process.

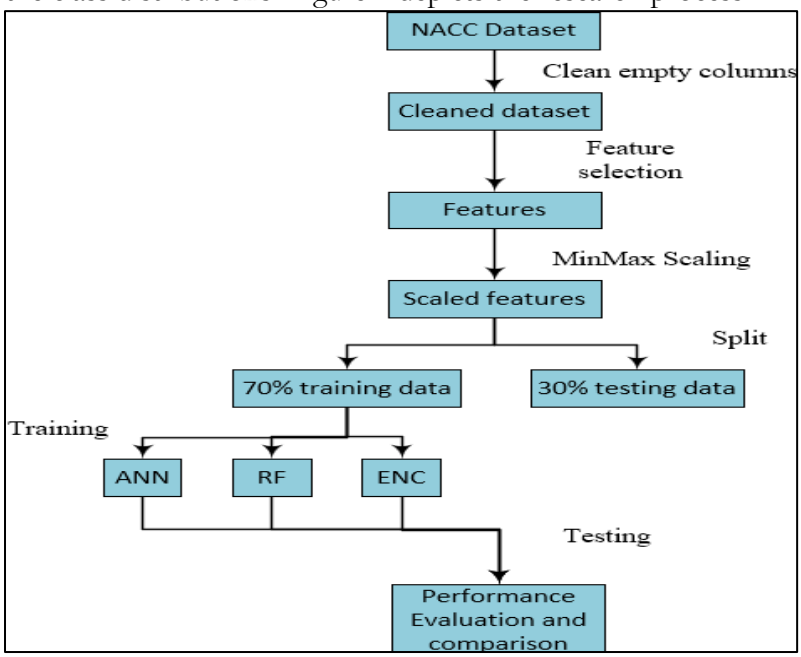

Figure 1. Flow of research

After splitting the dataset, the testing and training subsets receive their assigned portions: 70% training data and 30% testing data for the analysis. Before model input, all inputs are normalized through MinMaxScaler scaling. A computer training phase with Artificial Neural Network, Random Forest, and ElasticNet Classifier followed by testing occurs on the training data set. Tensorflow is used in the Python programming language for model training of ANN while Scikit Learn employs RF and ElasticNet. ElasticNet is a form of linear or logistic regression that combines L1 (Lasso) and L2 (Ridge) regularization to penalize the model. The formula for ElasticNet is given as:

$$\hat{\beta} = \arg_{\beta}\min(\| y - X\beta \|^2 + \lambda 2 \| \beta \|^2 + \lambda 1 \| \beta \|_1)$$

Where:

- y is the response vector,
- X is the design matrix,
- β is the coefficient vector,
- λ1 and λ2 are regularization parameters,
- $\| y - X\beta \|^2$ represents the squared error between the observed and predicted values,
- $\| \beta \|^2$ is the L2-norm (ridge regularization term),
- $\| \beta \|_1$ is the L1-norm (lasso regularization term)

For the implementation of ElasticNet, the Logistic Regression model from Scikit-Learn was utilized, with the penalty set to "elasticnet." The L1 and L2 ratios were both set to 0.5, and the saga solver was employed. RF uses multiple Decision Trees and uses the majority vote of the trees to determine the output. Scikit-learn's RandomForest model was used with default parameters for training and testing.

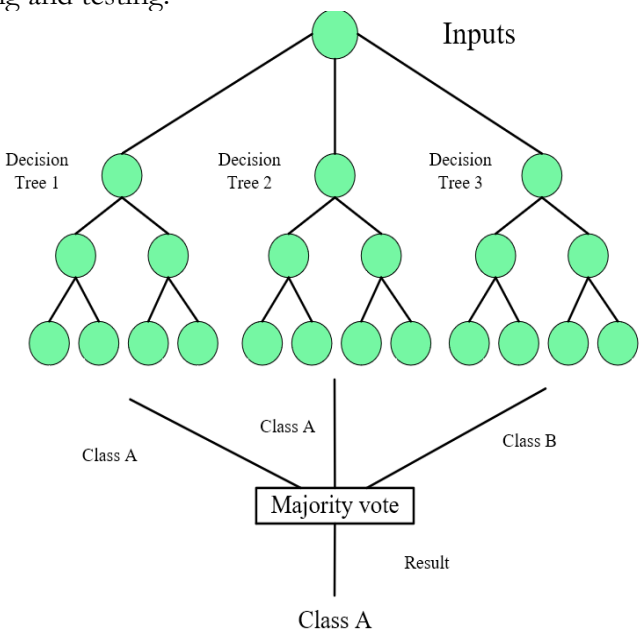

Figure 2. Random Forest architecture

ANN consists of units called neurons connected. It contains an input layer, followed by layers of neurons called hidden layers. The neurons in these layers can vary. The network concludes with an output layer that provides the result of the computation. In addition to neurons, Neural Networks also have weights associated with the layers as well as activation functions. The developers used Tensorflow to build their Neural Network program. A neural network structure includes 3 hidden layers with ReLU activation functions utilizing 6 then 50 and finally 250 neurons for each hidden layer. Three output neurons consist of the classification categories which include Alzheimer's Disease together with Cognitive impairment and Normal. The activation function Softmax applies to the design. The loss function included categorical cross entropy since it deals with multi-class classification problems. The system executed 40 cycles during training. The structure of the ANN used in this research is given in Figure 3.

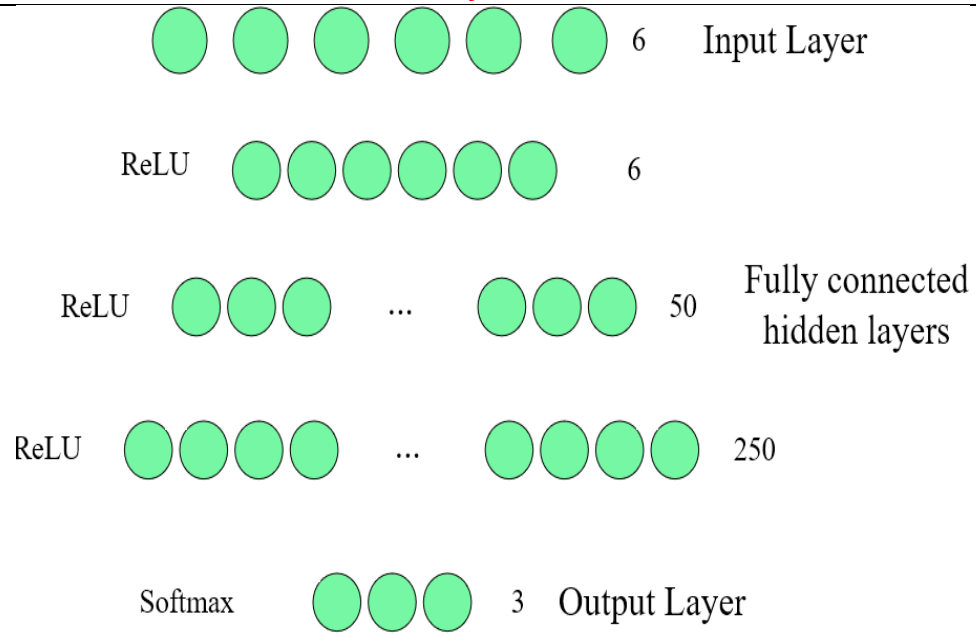

Figure 3. Artificial Neural Network Architecture

Accuracy, recall, and F1 score were among the performance measures that were calculated using Scikit-learns algorithms. Details of performance metrics are as follows.

Accuracy: This statistic calculates the proportion of accurate predictions (including true positives and true negatives) to total predictions, giving a general idea of how effective the model is. The relationship between exact matches and the entire set of forecasts is demonstrated by accurate prediction counts. The model's accuracy shows how many of its forecasts were accurate out of all of its predictions. The following is a formula for accuracy:

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

Recall: Recall gauges the model's capacity to accurately discover positive cases, which is essential for determining which individuals genuinely have Alzheimer's disease. A strong recall reduces the possibility of false negatives by demonstrating the model's sensitivity in identifying the illness. In medical diagnostics, reducing false negatives is crucial, as this study's focus on recall makes clear. Failing to diagnose a patient who has Alzheimer's disease (false negative) can have major repercussions, such as postponing intervention and treatment. Following is a formula for recall:

$$\text{Recall} = TP / (TP+FN)$$

F1 Score: When working with unbalanced datasets, such as those commonly used in medical diagnostics, the F1 Score provides a more thorough assessment by striking a compromise between precision and recall. In essence, the F1 Score is the harmonic mean of recall and precision. This classification methodology is useful when dealing with imbalanced datasets because measurements like accuracy could not provide enough information in those situations. A higher F1 score means that the model does a good job of avoiding false positives and recognizing the disease. The formula for the F1 score is:

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

Results and Discussion:

The NACC dataset, which was preprocessed to remove inconsistencies and select relevant features for Alzheimer's disease identification, was used in the experiment. To guarantee interoperability with ML models, preprocessing procedures included encoding categorical variables, standardizing numerical features, and addressing missing data. A k-fold cross-validation approach was implemented during training, where the dataset was divided into a 70:30

ratio, and models were trained iteratively on different splits. This technique ensured a more reliable estimation of model performance by reducing dependency on a single train-test split.

Python was used for the implementation, and popular ML technologies were used. NumPy and Pandas were utilized for data manipulation and numerical calculations, respectively, while Matplotlib and Seaborn were employed for visualization. Scikit-learn, which included tools for data preprocessing, model implementation, and performance evaluation, was used for both model training and evaluation. To split data, scale features, and compute metrics, functions such as train_test_split, StandardScaler, classification_report, and confusion_matrix were used. Hyperparameter adjustment was used to improve the models' performance. Standard criteria, such as Accuracy, Recall, and F1 Score, were used to evaluate the performance of three distinct models: ENC, RF, and ANN. The True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values that the models predicted were used to compute these metrics. Grid search was used to optimize the parameters and architecture of the ANN model. Hyperparameters were adjusted for RF and ENC using Scikit-learn's GridSearchCV to achieve the best results. The training set was used to train all models, and the test set was used to assess them. Accuracy, recall, and F1 score were among the performance measures that were calculated using Scikit-learn's algorithms.

The models achieved performance improvements with the implementation of hyperparameter adjustment methods. To identify its ideal design and adjust its fundamental parameters, the ANN implementation used a grid search technique. Three hidden layers with 6, 50, and 250 neurons each made up the final architecture. All hidden layers used the ReLU activation function, and the output layer used the Softmax activation function. The model was trained using a batch size of 32, and the Adam optimizer was chosen with a learning rate of 0.001. The model was trained for 40 epochs with a dropout rate of 0.2 to avoid overfitting. Other regularization strategies, such as L2 weight decay, were examined to guarantee stable performance but were shown to be superfluous because of the effective dropout rate. Hyperparameters for the RF and ENC were optimized for optimal performance using Scikit-learn's GridSearchCV tool. To avoid overfitting while maintaining adequate model complexity, the RF model was built up with n_estimators set to 200 and max_depth limited to 20. To maximize tree growth and generalization, other parameters like min_samples_split and min_samples_leaf were set to 2 and 1, respectively. To control model complexity and avoid overfitting, the ENC was adjusted with an L1 ratio of 0.5 and an alpha value of 0.1, striking a balance between L1 and L2 regularization. To reduce bias and preserve generalization, the regularization strength (alpha) was carefully chosen. Furthermore, the iterative fitting method implicitly controlled the learning rate for ENC optimization, guaranteeing convergence to an ideal solution. To balance model accuracy and computational efficiency, the tolerance for halting criterion was chosen at 1e-4.

These hyper-parameter techniques enhanced the generalization of the model. Overfitting was avoided by the ANN's dropout (0.2) and early halting, and effective convergence was guaranteed by the Adam optimizer (learning rate 0.001). With strong findings, RF's ensembling (n_estimators=200, max_depth=20) naturally reduced overfitting. By balancing bias and variance, ElasticNet's L1 ratio (0.5) and alpha (0.1) successfully managed model complexity. The models were evaluated using the Accuracy measure, F1 score, and Recall evaluation criteria. The evaluation metrics were calculated using the TP, TN, FP, and FN values that the models produced. All models' training processes ran on the training data, and the test data was used for testing.

The findings indicate that the ANN model was highly effective in learning from the dataset and accurately identifying both positive and negative cases of Alzheimer's disease. The results show that (as shown in Table 2) ANN achieved 92% Accuracy, 92% F1 Score, and 92% Recall, outperforming the other models.

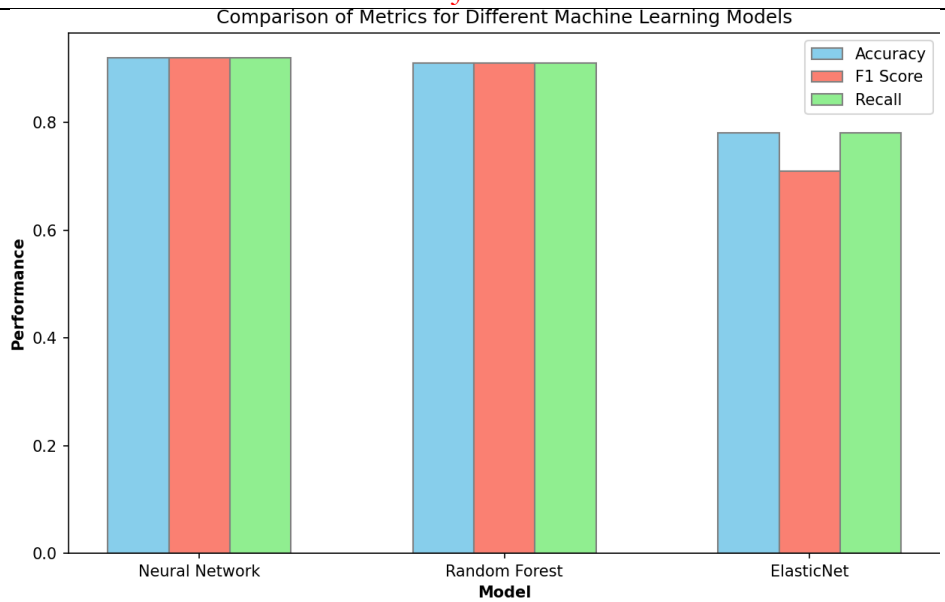

Figure 4. Performance Comparison

Table 2 Performance Metric

Model	Accuracy	F1 Score	Recall
Artificial Neural Network (ANN)	92%	92%	92%
Random Forest (RF)	91%	91%	91%
ElasticNet Classifier (ENC)	78%	71%	78%

The RF model also performed well, achieving 91% accuracy, 91% F1 score, and 91% recall. While it slightly lagged behind the ANN, it still produced reliable results.

Of the three models, the ENC had the lowest performance (78% Accuracy, 71% F1 Score, and 78% Recall). The intricacies of Alzheimer's disease detection were difficult for ENC to handle, despite its effectiveness with simpler linear correlations. Its lower performance suggests that more advanced models, such as ANN and RF, are better suited for identifying the underlying patterns in the data. Each model's performance is shown in Figure 4.

In this study, the dataset exhibited an imbalance between classes, with a lower proportion of Alzheimer's patients compared to the healthy controls. To mitigate the effects of imbalance, several techniques were employed such as Class Weighting, Synthetic Oversampling (SMOTE), and Threshold Tuning improving recall without drastically lowering precision.

This imbalance created a bias toward predicting the majority class would yield high accuracy without truly capturing the minority class. RF and ANN achieved high accuracy (91-92%), but accuracy alone was insufficient to determine their reliability. On the other hand, ENC had a lower accuracy (78%), likely due to its linear nature being less effective at handling imbalance.

Recall is a more appropriate metric for assessing performance in imbalanced datasets, as it measures the proportion of correctly identified positive cases. ANN and RF maintained a high recall (90-92%), indicating their ability to detect the minority class effectively. This was due to proper hyperparameter tuning and techniques like class weighting. EC had a lower recall (~71%), meaning it misclassified more Alzheimer's cases, likely due to its reliance on linear relationships.

The study's findings support the idea that ML models, specifically ANN and RF, have the potential to be incorporated into clinical workflows to aid in early diagnosis. Enabling

prompt treatment, has important ramifications for enhancing patient care and results. Therefore, these models could be used to help clinicians diagnose patients in actual clinical situations.

Discussion:

When compared to previous studies in the field of AD detection, the results of our investigation provide important new information. Unlike the more widely utilized MRI and Positron Emission Tomography (PET) data used in previous studies, our method makes use of clinical and genetic data from the NACC dataset. This change highlights the potential of underutilized clinical-genomic datasets in the early identification of AD in addition to addressing the constraints related to the availability and high cost of imaging data. Using MRI scans and VGG16 for feature extraction and a neural network for prediction, Shagun et al. [7] reached 90.4% accuracy on one dataset and 71.1% on another. Although their model performed well on certain datasets, its reliance on MRI restricts its scalability because imaging requires a lot of resources. On the other hand, our ANN model, which was trained using clinical and genomic data, outperformed Shagun et al.'s best results and showed that non-imaging data can effectively predict AD with an accuracy, F1 score, and recall of 92%. To overcome class imbalance, Murugan et al. [8] developed the DEMNET model utilizing MRI data and SMOTE. Their CNN has an 84.83% test accuracy on ADNI and a 94% validation accuracy after SMOTE. The ANN model in our study produced a greater overall accuracy without the requirement for oversampling approaches like SMOTE, even though their method successfully managed data imbalance and obtained high validation accuracy.

Taeho Jo et al. [13] achieved a 75% peak accuracy in their analysis of genomic data by combining CNN with LSTM and attention mechanisms. Although our study and their focus on genomic data are similar, our model performs noticeably better than theirs, demonstrating the benefits of our feature selection procedure and the combination of clinical and genomic data. The significance of multimodal data in improving predicted accuracy is highlighted by this comparison. With a remarkable accuracy of 98.5%, Modupe Odusami et al. [14] applied a Vision Transformer (ViT) model on merged MRI and PET images. Their reliance on multi-modal imaging data brings higher costs and accessibility difficulties, even though their model performs better than ours in terms of raw accuracy. Our model is a more viable option for widespread early identification of AD, especially in areas with low resources, because it relies on easily accessible clinical and genetic data.

In conclusion, this comparison shows that our study advances the field by showing that genomic and clinical data can be effective inputs for machine learning models in the early identification of Alzheimer's disease. The potential for scalable, affordable diagnostic tools is highlighted by the ANN model's better performance when compared to models that use more complicated or resource-intensive data. To further improve prediction accuracy while preserving accessibility and cost-effectiveness, future studies could investigate the merging of imaging and non-imaging data.

Conclusion and Future Work:

The common form of dementia known as Alzheimer's disease results in memory deterioration leading to patient death. The cause is still unknown. Diagnosing this disease in its early stages is difficult but can be aided by Machine Learning. The paper evaluates three different ML algorithms ANN, RF, and ENC for their ability to identify Alzheimer's disease at its earliest stages through analysis of NACC data. ANN performs the best with 92% accuracy. The group ratio of the dataset creates difficulty for model performance prediction. Future research must develop techniques to overcome classification bias while achieving better distribution between classes through data augmentation along with resampling methods and additional datasets. In the future, further improvements can be made by incorporating automatic feature selection techniques to refine model performance. Additionally, expanding the dataset with more diverse modalities and validating the models on external datasets could enhance their generalizability.

Once these models are refined and validated, they have the potential to be integrated into clinical workflows, reducing the time and cost associated with early Alzheimer's detection while assisting medical professionals in decision-making

References:

[1] A. Dementia, "Alzheimer's Disease Facts and Figures," *Alzheimer's Assoc.*, 2019, [Online]. Available: https://www.alz.org/alzheimers-dementia/facts-figures

[2] R. R. Janghel and Y. K. Rathore, "Deep Convolution Neural Network Based System for Early Diagnosis of Alzheimer's Disease," *IRBM*, vol. 42, no. 4, pp. 258–267, Aug. 2021, doi: 10.1016/J.IRBM.2020.06.006.

[3] U. Kumar and U. Kumar, "Applications of Machine Learning in Disease Pre-screening," *https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-5225-7131-5.ch010*, pp. 278–320, Jan. 1AD, doi: 10.4018/978-1-5225-7131-5.CH010.

[4] S. D. Nonie Alexander, Daniel C. Alexander, Frederik Barkhof, "Identifying and evaluating clinical subtypes of Alzheimer's disease in care electronic health records using unsupervised machine learning," *BMC Med. Informatics Decis. Mak. Vol.*, vol. 21, no. 343, 2021, doi: https://doi.org/10.1186/s12911-021-01693-6.

[5] D. A.-J. Abbas Saad Alatrany, Wasiq Khan, Abir Hussain, Hoshang Kolivand, "An explainable machine learning approach for Alzheimer's disease classification," *Sci. Rep.*, vol. 14, no. 2637, 2024, doi: https://doi.org/10.1038/s41598-024-51985-w.

[6] M. D. W. Janani Venugopalan, Li Tong, Hamid Reza Hassanzadeh, "Multimodal deep learning models for early detection of Alzheimer's disease stage," *Sci. Rep.*, vol. 11, no. 3254, 2021, doi: https://doi.org/10.1038/s41598-020-74399-w.

[7] S. K. Shagun Sharma, Kalpna Guleria, Sunita Tiwari, "A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans," *Meas. Sensors*, vol. 24, p. 100506, 2022, doi: https://doi.org/10.1016/j.measen.2022.100506.

[8] S. Murugan *et al.*, "DEMNET: A Deep Learning Model for Early Diagnosis of Alzheimer Diseases and Dementia from MR Images," *IEEE Access*, vol. 9, pp. 90319–90329, 2021, doi: 10.1109/ACCESS.2021.3090474.

[9] D. P. Vasco Sá Diogo, Hugo Alexandre Ferreira, "Early diagnosis of Alzheimer's disease using machine learning: a multi-diagnostic, generalizable approach," *Alzheimers. Res. Ther.*, vol. 14, no. 107, 2022, doi: https://doi.org/10.1186/s13195-022-01047-y.

[10] T. M. Ghazal *et al.*, "Alzheimer Disease Detection Empowered with Transfer Learning," *Comput. Mater. Contin.*, vol. 70, no. 3, pp. 5005–5019, Oct. 2021, doi: 10.32604/CMC.2022.020866.

[11] S. Naz, A. Ashraf, and A. Zaib, "Transfer learning using freeze features for Alzheimer neurological disorder detection using ADNI dataset," *Multimed. Syst.*, vol. 28, no. 1, pp. 85–94, Feb. 2022, doi: 10.1007/S00530-021-00797-3/METRICS.

[12] A. Bron, E. E., Klein, S., Papma, J. M., Jiskoot, L. C., Venkatraghavan, V., Linders, J., Aalten, P., De Deyn, P. P., Biessels, G. J., Claassen, J. A., Middelkoop, H. a. M., Smits, M., Niessen, W. J., Van Swieten, J. C., Van Der Flier, W. M., Ramakers, I. H., "Cross-cohort generalizability of deep and conventional machine learning for MRI-based diagnosis and prediction of Alzheimer's disease," *NeuroImage Clin.*, vol. 31, p. 102712, 2021, doi: https://doi.org/10.1016/j.nicl.2021.102712.

[13] A. J. Jo, T., Nho, K., Bice, P. J., & Saykin, "Deep learning-based identification of genetic variants: Application to Alzheimer's disease classification," *Brief. Bioinform.*, 2022, doi: https://doi.org/10.1101/2021.07.19.21260789.

[14] R. Odusami, M., Maskeliūnas, R.,Damaševičius, "Pixel-Level Fusion Approach with Vision Transformer for Early Detection of Alzheimer's Disease," *Electronics*, vol. 12, no. 5, p. 1218, 2023, doi: https://doi.org/10.3390/electronics12051218.

[15] P. D. Abhibhav Sharma, "A machine learning approach to unmask novel gene signatures and prediction of Alzheimer's disease within different brain regions," *Genomics*, vol. 113, no. 4, pp. 1778–1789, 2021, doi: https://doi.org/10.1016/j.ygeno.2021.04.028.

[16] Research & Funding, "National Alzheimer's Coordinating Center," *Div. Neurosci.*, [Online]. Available: https://www.nia.nih.gov/research/dn/national-alzheimers-coordinating-center-nacc