SIES Graduate School of Technology

Department of Computer Engineering

Study of Alzheimer's Disease through Machine Learning Techniques

By

Anshul Choudhari 122A1016

Sanjana Yadav 122A1080

Shreeya Bhalerao 122A1092

Vinita Narayanan 122A1123

Guided by: Prof. Ujwala Ravale

Introduction

- Alzheimer’s disease (AD) is a neurodegenerative disorder that progressively impairs memory and cognitive function, making early detection critical.

- This project explores machine learning techniques for diagnosing and predicting the progression of Alzheimer's. We implemented models such as K-Nearest Neighbors (KNN), Naive Bayes, Logistic Regression, Decision to analyze clinical data and cognitive assessments.

- By comparing the performance of these models in terms of accuracy and precision, the study aims to identify the most effective method for early diagnosis. The goal is to improve diagnostic accuracy and provide insights that can help healthcare professionals offer more personalized care for Alzheimer’s patients.

Problem Statement

- Alzheimer's disease (AD) is the most **common cause of dementia** in older adults, leading to severe cognitive decline. However, AD is often **difficult to diagnose in its early stages**, where intervention can significantly slow progression and reduce long-term damage. Early diagnosis allows for more effective treatments, which can improve patient outcomes.

- Machine learning (ML) offers a promising approach for predicting Alzheimer's disease in its early stages by analyzing patterns in medical data that may not be easily detectable through traditional methods. This project aims to utilize various ML models to predict AD, assisting healthcare professionals in identifying at-risk individuals. Early detection and intervention can significantly enhance the prevention and management of Alzheimer's disease.

Objective

The objective of this project is to utilize machine learning techniques to enhance the early detection of Alzheimer's Disease, which is crucial for timely intervention and treatment. Our specific goals are to:

- Analyze clinical data from patients at risk of developing Alzheimer's to uncover patterns that may be missed by traditional diagnostic methods.
- Implement and evaluate a variety of machine learning classification algorithms, including methods such as decision trees, support vector machines, and neural networks, to predict Alzheimer's in its early stages.
- Compare the performance of these algorithms using metrics like accuracy, precision, recall, and F1-score to determine the most effective model for predicting Alzheimer's.
- Provide insights into which machine learning models are best suited for clinical applications in the early diagnosis of Alzheimer's, contributing to improved patient outcomes.

Scope

This project explores the application of machine learning techniques in the early diagnosis of Alzheimer's Disease, focusing on:

- Data Analysis: Utilizing clinical data from patients at risk of Alzheimer's to identify key patterns and features relevant to early detection.
- Model Implementation: Implementing a range of machine learning algorithms, including decision trees, support vector machines, neural networks, and others, to predict the onset of Alzheimer's.
- Performance Evaluation: Comparing the accuracy, precision, recall and F1-Score of each algorithm to assess their effectiveness in detecting Alzheimer's at an early stage.
- Model Optimization: Fine-tuning models to improve their performance, ensuring they are suitable for real-world clinical use.
- Clinical Impact: Providing insights that can contribute to the development of better diagnostic tools for healthcare professionals, with the potential for more timely and effective treatment of patients.

Literature Survey

Sr no	Title of the Paper	Algorithm and Dataset used	Conclusion	Limitations
1	Comparison of Machine Learning-based Approaches to Predict the Conversion to Alzheimer's Disease from Mild Cognitive Impairment March 2023	• **Algorithms**: Random Forest, Gradient Boosting, XGBoost • **Dataset**: ADNI (Alzheimer's Disease Neuroimaging Initiative)	• Achieved 90% accuracy using Random Forest on the ADNI dataset. • Limited analysis due to the use of only three models. • Psychological and AD-related biomarkers reach the highest accuracy in AD prediction.	• Only used 3 classification models for comparisons, for a holistic and varied study more than 3 would be ideal. • Highest accuracy reached on Random Forest is only 90%.
2	Classification of Alzheimer's Disease using Machine Learning Techniques	• **Algorithms**: KNN, Decision Tree, Rule Induction, Naive Bayes, Generalized Linear Model , Deep Learning • **Dataset**: ADNI	• The Generalized Linear Model (GLM) achieved the highest accuracy of 88.24%, outperforming other classifiers. • This accuracy is insufficient for reliable Alzheimer's detection.	• Limitations include a maximum accuracy of 88.24%,lack of biomarker correlation analysis, and unclear feature importance, limiting clinical utility and interpretability.

Literature Survey

Sr no	Title of the Paper	Algorithm and Dataset used	Conclusion	Limitations
3	Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models March 2022	• **Algorithms**: Decision Tree, Random Forest, SVM, XGBoost and Voting classifiers. • **Dataset**: Open Access Series of Imaging Studies (OASIS)	• Random Forest (86.92%) and XGBoost (85.92%) were found to be the most accurate models. • 5-fold cross-validation was applied to ensure robust and reliable model performance.	• The dataset was limited in size potentially affecting the model's generalizability. • The correlations between biomarkers and the target variable were not fully analyzed, limiting insights into feature interactions.
4	Early prediction of Alzheimer's disease using convolutional neural network: a review. November 2022	• **Algorithms**: Convolutional neural network (CNN) • **Dataset**: ADNI	• Two CNN architectures are compared: an 18-layer CNN and a 3D CNN. • The 18-layer CNN achieved 98% accuracy in detecting AD, outperforming the 3D CNN's 88%.	• Highly layered CNNs are complex in terms of computation time and resource requirements. • CNNs face complexity in identifying affected areas in older adults, as MRIs do not account for memory loss.

Literature Survey

Sr no	Title of the Paper	Algorithm and Dataset used	Conclusion	Limitations
5	A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease. March 2021	• **Algorithms:** Support Vector Machine (SVM), Logistic Regression, Decision Tree, Random Forest • **Dataset**: OASIS	• The study compared SVM, Logistic Regression, Decision Tree and Random Forest on the OASIS dataset. • SVM had the highest accuracy, highlighting the potential for improvement with larger datasets and more algorithms.	• Small OASIS dataset limits generalization. • Overfitting in Decision Tree and Random Forest. • Limited to four algorithms, missing advanced techniques. • Findings may not apply globally.
6	A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease July 2023	• **Algorithms**: Logistic Regression, SVM, CNN, RNN • **Dataset**: ADNI, OASIS	• CNN achieved 98.6% accuracy,SVM 85.7%, RNN 91.2%, and CNN ensemble 99.4% for NC/AD and MCI prediction. • Machine learning techniques like CNN and RNN detect Alzheimer's with high accuracy.	• Focused only on MRI/PET scans, excluding other diagnostic methods. • Exclusion of incomplete data may bias results.

Research Gaps

- Some studies utilized only three classification models, limiting the diversity of their analyses.
- In one study, Random Forest achieved a maximum accuracy of just 90%, while overall accuracy was reported at 88.24%.
- Several studies faced issues with small datasets, hindering the models' generalizability.
- Insufficient analysis of biomarker correlations in certain studies restricts insights into feature relationships.
- Overfitting was a common concern in Decision Tree and Random Forest models across various studies.
- The exclusion of advanced algorithms in some research reduces comprehensiveness and potential for improved outcomes.

Proposed System: Tabular Data

Data Pre-processing

- Unwanted attributes are removed, and NULL values are handled to ensure a clean, complete dataset.
- Standardization or normalization is applied where necessary.

Feature Engineering

- Key attributes are selected, and the dataset is split into training and testing sets for model evaluation.
- Chi-square correlation test done

Model Training

- Various classification algorithms (KNN, Naive Bayes, Logistic regression, Random Forest, Decision Tree, SVM) are trained.
- Hyperparameters are tuned using cross-validation to optimize performance.

Performance Evaluation

- Metrics like accuracy, precision, and recall are used to evaluate and compare the models to identify the most effective one for Alzheimer's detection.

Performance Measures

- Accuracy: Accuracy is the proportion of all classifications that were correct, whether positive or negative. It is mathematically defined as:

$$Accuracy = \frac{Correct\ classifications}{Total\ classifications} = \frac{\begin{matrix} True\ Negative \\ +True\ Positive \end{matrix}}{\begin{matrix} True\ Negative + False\ Positive \\ + True\ Positive + False\ Negative \end{matrix}}$$

- Precision: Precision is the proportion of all the model's positive classifications that are actually positive. It is mathematically defined as:

$$Precision = \frac{Correctly\ classified\ actual\ positives}{Everything\ classified\ as\ positive} = \frac{TP}{TP + FP}$$

Performance Measures

- Recall: The proportion of all actual positives that were classified correctly as positives, is also known as recall. It is mathematically defined as:

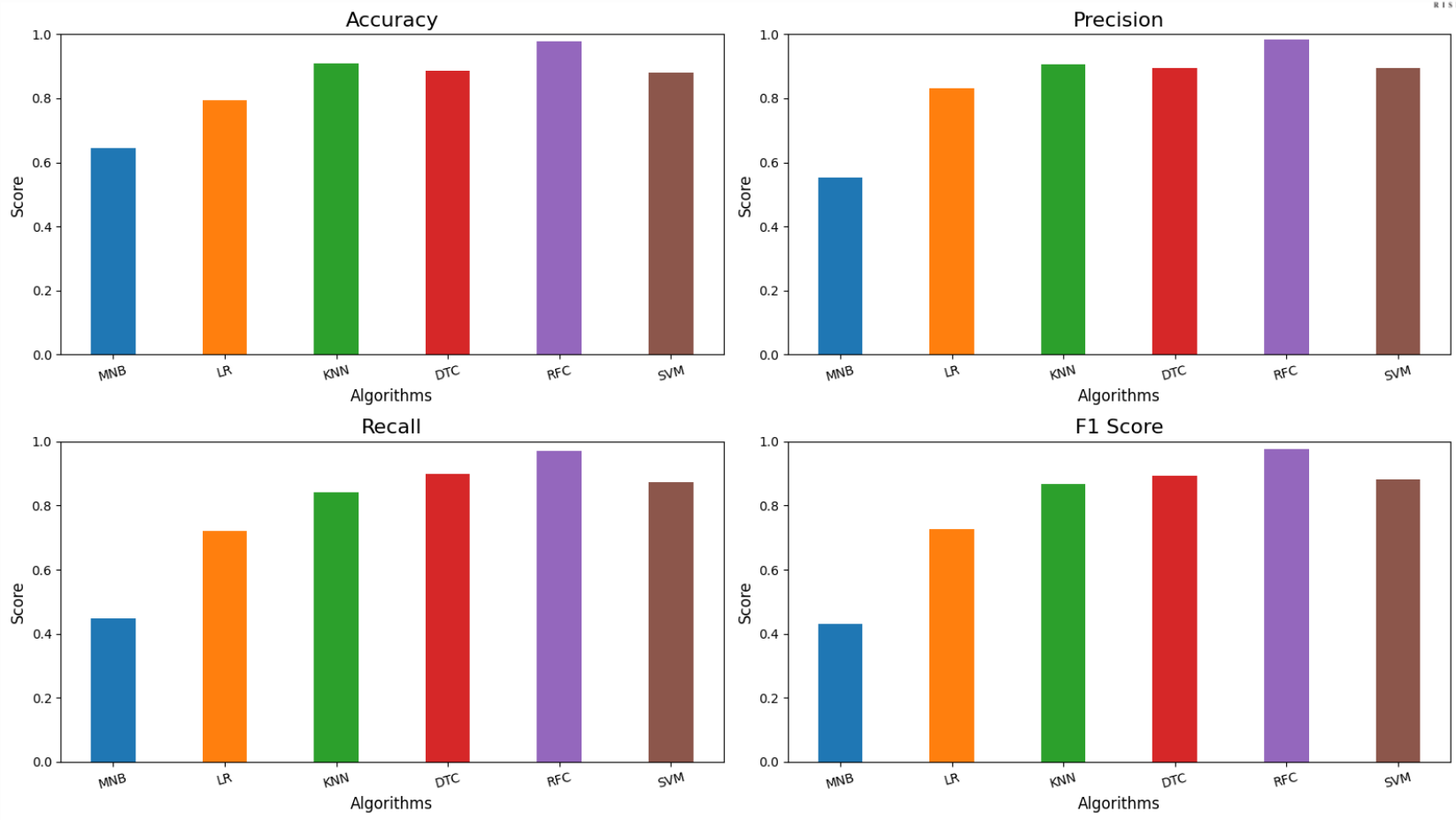
$$Recall = \frac{Correctly\ classified\ actual\ positives}{All\ actual\ positives} = \frac{TP}{TP + FN}$$

- F1 Score: F1 Score is a metric that balances precision and recall. It is calculated as the harmonic mean of precision and recall. It is mathematically defined as:

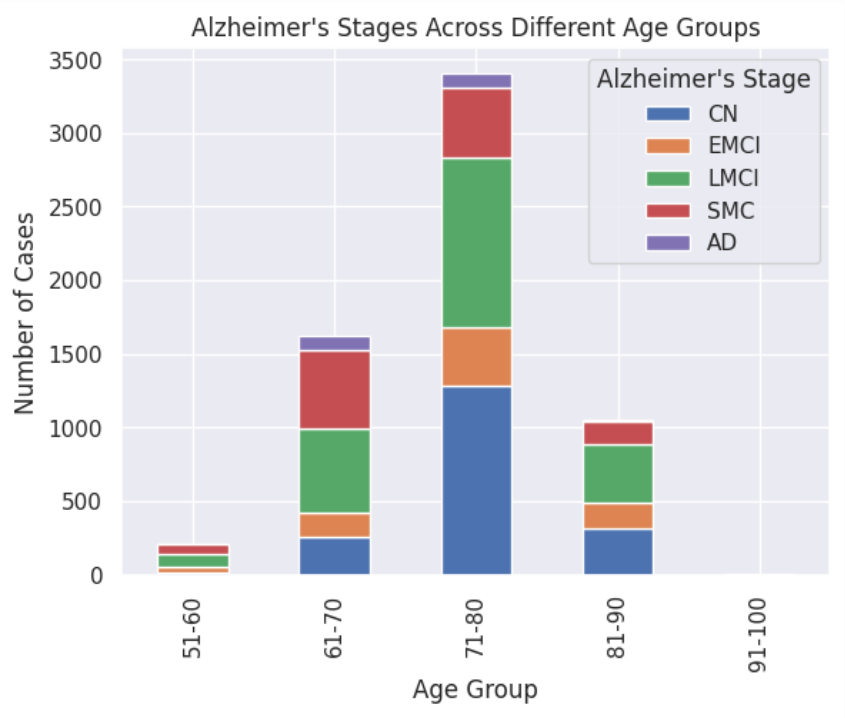
$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

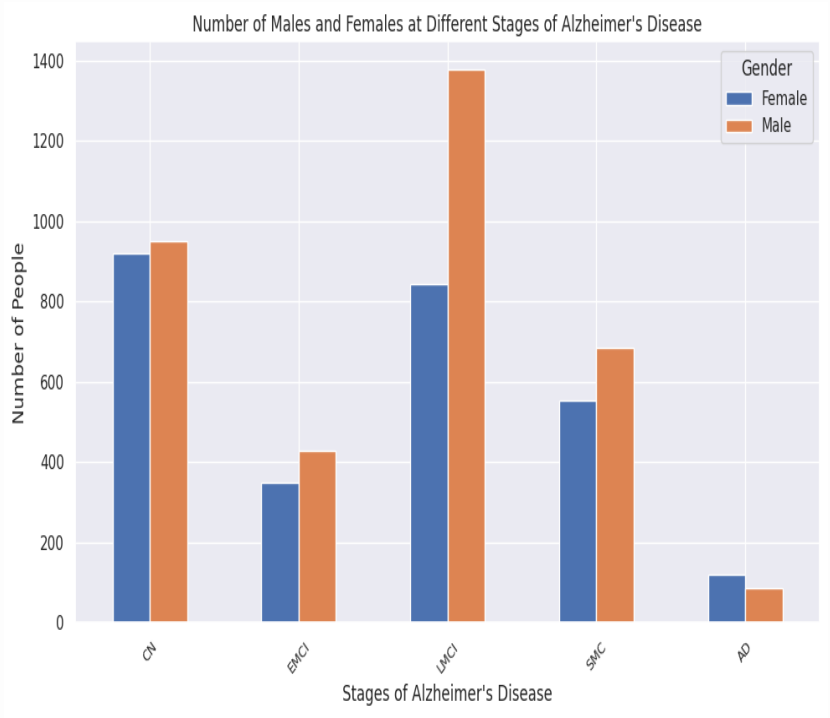
Result

Algorithm	Accuracy	Precision	Recall	F1-Score
Multinomial Naïve Bayes	0.644	0.552	0.449	0.430
Logistic Regression	0.795	0.831	0.720	0.728
K Nearest Neighbours	0.910	0.907	0.842	0.867
Decision Tree Classifier	0.885	0.896	0.899	0.894
Random Forest Classifier	0.979	0.983	0.970	0.976
Support Vector Machine	0.881	0.895	0.872	0.881

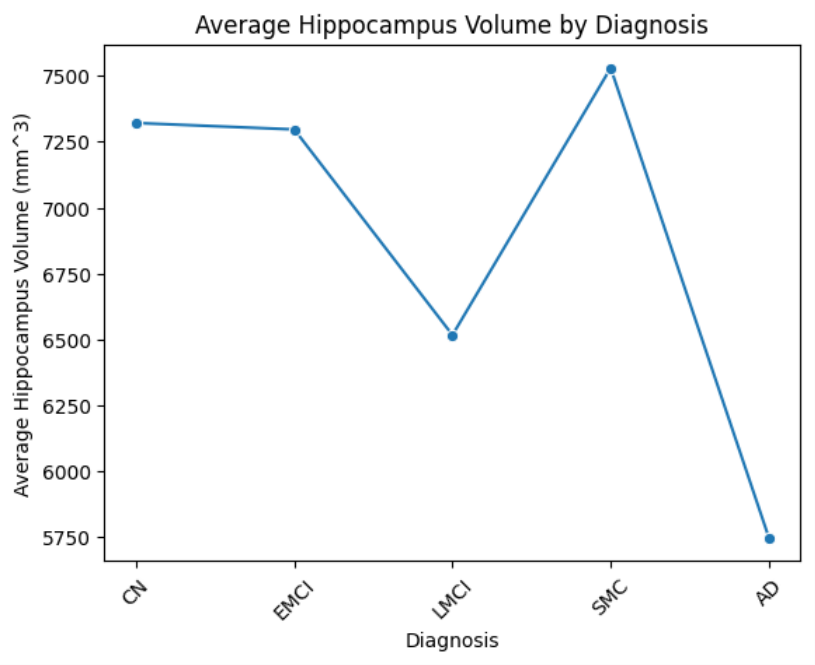

Accuracy
Precision
Recall
F1 Score
Score
Algorithms
MNB
LR
KNN
DTC
RFC
SVM
0.0
0.2
0.4
0.6
0.8
1.0

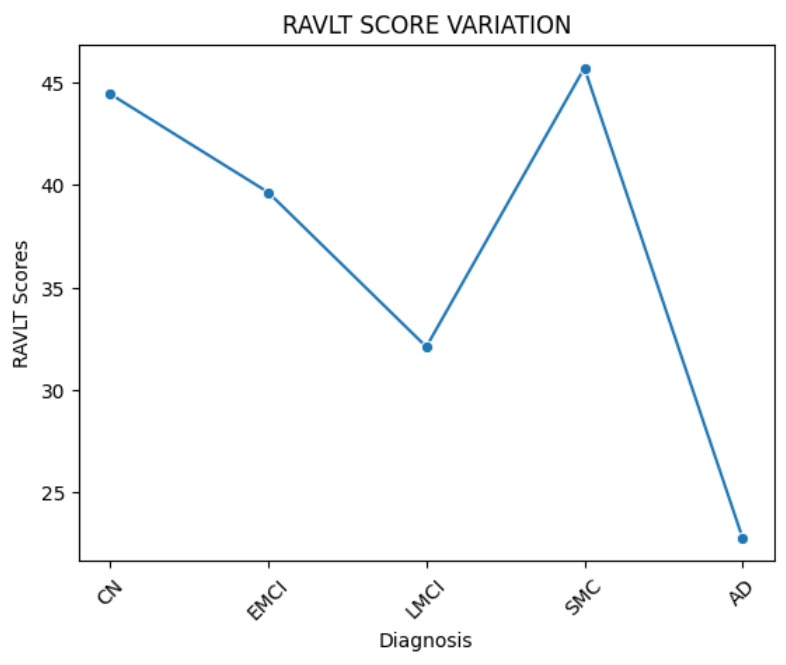
Attribute Correlation





Attribute Correlation





Summary

Our project focuses on utilizing machine learning (ML) techniques for the early diagnosis and prediction of Alzheimer's disease (AD). By analyzing clinical data, we implemented various ML models, including K-Nearest Neighbors (KNN), Naive Bayes, Logistic Regression, Decision Tree, and SVM. The objective is to compare the performance of these models in terms of accuracy and precision, aiming to identify the most effective method for early detection. Early diagnosis plays a crucial role in improving patient outcomes, and this project seeks to assist healthcare professionals by providing more accurate diagnostic tools. Key tasks included data collection, preprocessing, feature extraction, model training, and evaluation, with performance metrics such as accuracy, precision, recall, and F1-score used for comparison.

References

- R. Franciotti et al., “Comparison of Machine Learning-based Approaches to Predict the Conversion to Alzheimer’s Disease from Mild Cognitive Impairment,” Neuroscience, vol. 514, pp. 143–152, Mar. 2023, doi: https://doi.org/10.1016/j.neuroscience.2023.01.029.
- C. Kavitha, V. Mani, S. R. Srividhya, O. I. Khalaf, and C. A. Tavera Romero, “Early-Stage Alzheimer’s Disease Prediction Using Machine Learning Models,” Frontiers in Public Health, vol. 10, Mar. 2022, doi: https://doi.org/10.3389/fpubh.2022.853294.
- V. Patil, M. Madgi, and A. Kiran, “Early prediction of Alzheimer’s disease using convolutional neural network: a review,” The Egyptian Journal of Neurology, Psychiatry and Neurosurgery, vol. 58, no. 1, Nov. 2022, doi: https://doi.org/10.1186/s41983-022-00571-w.
- M. Bari Antor et al., “A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer’s Disease,” Journal of Healthcare Engineering, vol. 2021, p. e9917919, Jul. 2021, doi: https://doi.org/10.1155/2021/9917919.

References

- Akhilesh Deep Arya et al., “A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer’s disease,” Brain Informatics, vol. 10, no. 1, Jul. 2023, doi: https://doi.org/10.1186/s40708-023-00195-7.
- https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall
- https://klu.ai/glossary/accuracy-precision-recall-f1

Thank You