

Alzheimer's Disease Prediction Using Machine Learning Techniques

P.Narasimhaiah¹, C.Venkata Surendra Kumar², S.Karthik³, P. Sujan Kumar⁴, V. Anil⁵, P.Karthikeya⁶

¹Assoc.Proffesor, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, A.P

²UG Student, Department of Computer Science and Engineering(AI), CBIT, Proddatur, YSR, A.P

³UG Student, Department of Computer Science and Engineering(AI), CBIT, Proddatur, YSR, A.P

⁴UG Student, Department of Computer Science and Engineering(AI), CBIT, Proddatur, YSR, A.P

⁵UG Student, Department of Computer Science and Engineering(AI), CBIT, Proddatur, YSR, A.P

⁶UG Student, Department of Computer Science and Engineering(AI), CBIT, Proddatur, YSR, A.P

Corresponding Author Email: parlapatisujankumar@gmail.com

ABSTRACT

Alzheimer's Disease is a progressive neurodegenerative complaint that primarily affects memory, cognitive capacities, and diurnal functioning among senior individualities. Beforehand discovery of the complaint is pivotal for effective treatment planning and bettered quality of life. still, conventional individual methods are frequently precious, time-consuming, and dependent on expert interpretation. This paper presents a machine literacy – grounded system for early vaticination of Alzheimer's Disease using clinical, cognitive, and demographic data. The proposed system follows a complete machine learning channel including data preprocessing, point selection, model training, and evaluation. Logistic Retrogression, Random Forest, and gradient Boosting models are enforced and compared. The system is stationed using a Stream lit- grounded web operation that provides probability- grounded vaticination and unborn threat estimation. Experimental results indicate that ensemble models offer dependable and costeffective vaticination without reliance on neuroimaging ways.

Keywords: *Alzheimer's Disease, Machine Learning, Early vaticination, Logistic Retrogression, Random Forest, Gradient Boosting*

1.INTRODUCTION

Alzheimer's Disease (announcement) is a progressive neurodegenerative complaint that primarily affects memory, cognitive functions, and the capability to perform diurnal conditioning. It's the most common cause of madness among aged grown-ups and represents a major global public health challenge. As life expectation increases worldwide, the number of

individualities affected by Alzheimer's Disease is anticipated to rise significantly, placing a substantial burden on cases, families, and healthcare systems.

Early identification of Alzheimer's Disease is pivotal, as timely opinion enables better treatment planning, slows complaint progression, and improves quality of life. still, conventional individual styles calculate heavily on clinical examinations, cognitive assessments, and advanced neuroimaging ways similar as glamorous Resonance Imaging (MRI) and Positron Emission Tomography (PET). These approaches are frequently precious, time-consuming, and bear technical medical moxie, making them less accessible in resourcelimited settings.

In recent times, machine literacy ways have surfaced as effective tools for medical decision support systems. By assaying structured clinical, cognitive, and demographic data, machine literacy models can identify retired patterns associated with complaint progression and support early vaticination. Unlike imaging-grounded approaches, data-driven prophetic models offer a cost-effective, scalable, and interpretable volition for Alzheimer's Disease discovery. This paper presents a machine literacy – grounded approach for the early vaticination of Alzheimer's Disease using on-invasive case data. The proposed system follows a methodical channel that includes data preprocessing, point selection, model training, and evaluation. Multiple supervised literacy algorithms, including Logistic Retrogression, Random Forest, and grade Boosting, are enforced and compared to identify the most effective model. The final system is stationed through a web-grounded interface, enabling real-time vaticination and probability-grounded threat assessment.

The remainder of this paper is organized as follows. Section II reviews related work and discusses being and proposed systems. Section III describes the methodology, including system armature and module design. Section IV presents experimental results and discussion.

Eventually, Section V concludes the paper and outlines unborn exploration directions

LITERATURE REVIEW

This section presents a review of being exploration related to Alzheimer's Disease vaticination and highlights the provocation for the proposed system. The literature is organized into two corridor the being systems and the proposed system approaches reported in recent studies.

A. Existing System

Early exploration on Alzheimer's Disease opinion primarily concentrated on traditional clinical evaluation and neuroimaging ways. Cognitive assessments, medical history analysis, and neurological examinations have been extensively used to identify symptoms of madness and Alzheimer's Disease. In addition, imaging modalities similar as glamorous Resonance Imaging (MRI) and Positron Emission Tomography (PET) have been considerably employed to study structural and functional changes in the brain.

With the advancement of computational styles, several experimenters applied machine literacy and deep literacy ways to neuroimaging data. Convolutional Neural Networks (CNNs) and other deep literacy infrastructures have demonstrated high delicacy in detecting Alzheimer's Disease from MRI and PET reviews. These models automatically learn complex features from high-dimensional image data and achieve promising individual performance.

Despite their effectiveness, imaging-grounded systems suffer from significant limitations. The accession of MRI and PET reviews is precious, time-consuming, and requires advanced medical structure. also, deep literacy models frequently demand large labelled datasets and high computational coffers, making them delicate to emplace in real-world clinical surroundings. The lack of interpretability in numerous deep literacies approaches also reduces their acceptance among healthcare professionals.

B. Proposed System in Literature

To overcome the limitations of traditional and imaging-grounded systems, recent exploration has shifted towards the use of structured clinical, cognitive, and demographic data for Alzheimer's Disease vaticination. Machine literacy models trained on patient attributes similar as age, cognitive test scores, educational position, and functional capacities have shown promising results in early discovery.

Classical machine learning algorithms, including Logistic Retrogression, Random Forest, Support Vector Machines, and grade Boosting, have been extensively explored due to their robustness and interpretability. Ensemble literacy ways, in particular, have demonstrated bettered performance by combining multiple classifiers to handle non-linear connections and reduce overfitting.

Several studies emphasize that clinical data-driven models give competitive delicacy compared to imaging-grounded approaches while significantly reducing cost and complexity. These models are easier to emplace and integrate into decision support systems, making them more suitable for real-time operations.

Recent literature also highlights the significance of planting prophetic models through stonerfriendly platforms. Web-grounded systems allow clinicians and individualities to interact with vaticination models, fantasize results, and assess complaint threat efficiently. Probabilitygrounded labours and threat estimation further enhance the utility of similar systems for preventative healthcare planning.

Grounded on the reviewed literature, it's apparent that machine literacy approaches using oninvasive clinical data offer a practical and scalable result for early Alzheimer's Disease vaticination. These findings form the foundation for the proposed system presented in this paper.

Several studies have also reported that numerous being systems are able of detecting Alzheimer's Disease only at moderate or advanced stages, limiting their utility for early intervention. These challenges punctuate the need for indispensable approaches that are costeffective, interpretable, and able of supporting early-stage vaticination. Maintaining the Integrity of the Specification

METHODOLOGY

System Architecture



This section describes the methodology espoused for the proposed Alzheimer's Disease vaticination system. It outlines the overall system armature and explains the functional modules involved in structure and planting the machine literacy model.

A. System Architecture

The proposed system follows a modular machine learning armature designed to support early vaticination of Alzheimer's Disease using on-invasive clinical and cognitive data. The armature consists of multiple connected factors that inclusively perform data accession, preprocessing, model training, evaluation, and deployment.

The system begins with a stoner interface that allows druggies to input patient-affiliated information, similar as demographic details and cognitive assessment scores. The input data is also reused through a preprocessing module to ensure data quality and thickness. point selection ways are applied to identify the most applicable attributes impacting Alzheimer's vaticination. latterly, multiple supervised machine literacy models are trained and estimated to determine the most effective classifier. Eventually, the named model is stationed through a web-grounded operation that provides probability-grounded vaticination results and unborn threat estimation. This armature eliminates the reliance on expensive neuroimaging ways and enables real-time, accessible, and interpretable vaticination.

B. Modules Description

The major modules of the proposed system are described as follows.

Data Collection Module

The data collection module is responsible for gathering structured case data needed for vaticination. The dataset includes clinical, cognitive, and demographic attributes similar as age, gender, cognitive test scores, medical history, and life-related factors. This data serves as the foundation for training and assessing the machine literacy models.

Data Preprocessing Module

The preprocessing module ensures that the collected data is clean and suitable for model training. This stage involves handling missing values, garbling categorical variables, and

homogenizing or spanning numerical features. Proper preprocessing improves model performance and reduces the impact of noise and inconsistencies in the dataset.

Point Selection Module

Point selection is performed to identify the most significant attributes contributing to Alzheimer's Disease vaticination. Statistical analysis and model-grounded ways are used to exclude inapplicable or spare features. This step reduces dimensionality, improves computational effectiveness, and enhances model interpretability.

1. Model Training Module

In this module, multiple supervised machine learning algorithms are trained using the preprocessed dataset. Logistic Regression is used as a birth model due to its simplicity and interpretability. Random Forest and Gradient Boosting models are employed to capture complex non-linear connections and ameliorate vaticination delicacy. Each model is trained on the training dataset and validated using the testing dataset.

2. Model Evaluation Module

The trained models are estimated using standard performance criteria, including delicacy, perfection, recall, F1-score, and Receiver Operating Characteristic (ROC) wind analysis. relative evaluation is performed to identify the model that provides balanced and dependable performance, with particular emphasis on recall to minimize false negatives.

3. Deployment Module

The deployment module integrates the named machine literacy model into a web-grounded operation developed using Streamlit. This interface enables druggies to enter patient details and admit real-time vaticination results. The system generates probability-grounded labours and threat assessment, making it suitable for early webbing and decision support.

C. Mathematical Formulation & Equations

Let the dataset be represented as:

$D = \{(x_i, y_i)\}$ Where, $x_i = [x_{i1}, x_{i2}, \dots, x_{im}]$ denotes the point vector containing clinical, cognitive, and demographic attributes of the i th patient, and $y_i \in \{0,1\}$, represents the class label, Where,

1 indicates Alzheimer's Disease. 0

indicates a Healthy Subject.

1. Logistic Regression Model

Logistic Regression estimates the probability of Alzheimer's Disease as:

$P(y = 1 | x) = 1 / (1 + e^{-(\beta_0 + \sum \beta_j x_j)})$ β_0 is the biased term and β_j are the model coefficients.

The model parameters are learned by minimizing the log-loss function:

$$L = - (1/N) \sum [y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i)]$$

2. Random Forest Model

Random Forest is an ensemble learning method that combines multiple decision trees. The final prediction/output is obtained based on highest voting: $\hat{y} = \text{mode} \{h_1(x), h_2(x), \dots, h_t(x)\}$ Where,

$h_t(x)$ represents the prediction from the t^{th} decision trees and

T is the total number of trees

3. Gradient Boosting Model

Gradient Boosting builds models sequentially, where each new model minimizes the residual error of the previous model

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \text{ Where,}$$

$F_m(x)$ is the boosted model at m iterations.

$h_m(x)$ is the weak learner, and $\gamma_m(x)$

is learning rate.

D. Algorithm

1 Input Clinical and cognitive dataset as pd
Alzheimer's Disease prediction with probability score

2 Step 1 - Load dataset pd

- Step 2 - Handle missing values in pd
- Step 3 - Encode categorical features
- Step 4 - homogenize/Normalize numerical attributes
- Step 5 - Perform point selection
- Step 6 - Split pd into training set and testing set
- Step 7 - Train Logistic Regression model
- Step 8 - Train Random Forest model
- Step 9 - Train gradient Boosting model
- Step 10 - estimate all models using accuracy, precision, recall, F1- score
- Step 11 - Select the best- performing model
- Step 12 - Deploy selected model using Streamlit

RESULTS AND DISCUSSION



Fig:1.1 performance comparison of different Models

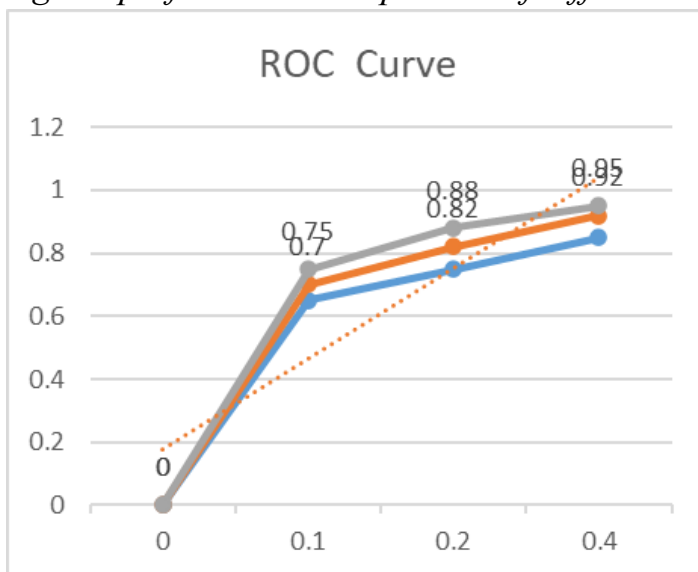


Fig:1.2 The Graph of ROC Curve

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logestic Regression	82.4	8.9	79.6	80.2
Random Forest	87.8	86.5	88.1	87.3
Gradient Boosting	89.6	88.9	90.2	89.5

Table.1 : Performance Comparison of Machine Learning Models

CONCLUSION

Alzheimer's Disease is a progressive neurodegenerative disorder that demands early and accurate prediction to enhance patient care and slow disease progression. Conventional diagnostic approaches rely heavily on costly neuroimaging techniques and expert evaluation, which limits accessibility and delays detection. The proposed machine learning-based system provides a cost-effective, non-invasive alternative by utilizing clinical, cognitive, and demographic data. By removing dependency on MRI and PET scans, the system enables faster, automated, and more accessible predictions. The deployment of the model through a web-based interface further improves usability and real-time decision support. This approach has the potential to support early intervention, preventive planning, and improved clinical outcomes. Future work may focus on incorporating larger and more diverse datasets and advanced algorithms to enhance accuracy and real-world applicability.

REFERENCES

- Ge, C., Wang, H., & Zhang, J. (2018). Alzheimer's disease diagnosis based on multiscale convolutional neural networks. *Frontiers in Neuroinformatics*, 12, 26. <https://www.frontiersin.org/articles/10.3389/fninf.2018.00026/full>
- Thompson, P. M., Hayashi, K. M., de Zubicaray, G. I., et al. (2004). Mapping hippocampal and ventricular change in Alzheimer disease. *Neuroimage*, 22(4), 1754–1766. <https://pubmed.ncbi.nlm.nih.gov/15219524/>
- Acharya, U. R., Fernandes, S. L., WeiKoh, J. E., et al. (2019). Automated detection of Alzheimer's disease using brain MRI images: A study with various feature extraction techniques. *Journal of Medical Systems*, 43(9), 302. <https://doi.org/10.1007/s10916-019-1416-6>

Eke, C. S., Jammeh, E., Li, X., Carroll, C. B., Pearson, S., & Ife Achor, E. (2015). Early detection of Alzheimer's disease with blood biomarkers using machine learning techniques. *BioMed Research International*, 2015, 1–10.

<https://www.hindawi.com/journals/bmri/2015/317847/>

Jain, R., Jain, N., Aggarwal, A., & Hemanth, D. J. (2018). Convolutional neural network based Alzheimer's disease classification from MRI. *Procedia Computer Science*, 132, 129–136.

<https://www.sciencedirect.com/science/article/pii/S1877050918307360>

Zhao, F., Liu, Z., & Jin, Z. (2019). Alzheimer's disease classification based on MRI and support vector machine. *Journal of Medical Imaging and Health Informatics*, 9(3), 574–580.

https://www.researchgate.net/publication/335801702_Classification_of_Alzheimer%27s_disease_based_on_brain_MRI_and_machine_learning

World Health Organization. (2023). Dementia. *World Health Organization*.

<https://www.who.int/news-room/fact-sheets/detail/dementia>