ISSN: 2349-6002

Alzheimer's disease Prediction Using Machine Learning Techniques and Feature Selection

Mrs. Aparna T. Kulkarni[1], Dr. Anirudh K. Mangore[2]

[1]*Student, Dept. Computer Science and Engineering (Data Science), DY Patil Agriculture and Technical University, Talsande*

[2]*Associate Professor, Dept. Computer Science and Engineering, DY Patil Agriculture and Technical University, Talsande*

Abstract- One of the main reasons why older adults develop dementia is Alzheimer's disease (AD). Additionally, a sizable section of the global population has metabolic issues like diabetes and Alzheimer's disease. Degenerative effects of Alzheimer's disease are shown in the brain. This disease can lead to more people being inactive when the number of senior people increases since it affects their physical and mental abilities. This could affect their family members as well as the social, financial, and economic domains. Recently, researchers have looked into several deep learning and machine learning techniques to identify these illnesses early. Successful and minimally harmful recovery from AD is facilitated by early diagnosis and therapy. In order to predict Alzheimer's illness, this study suggests a machine learning model that combines GaussianNB, Decision Tree, Random Forest, XGBoost, Voting Classifier, and GradientBoost. The open access series of imaging studies (OASIS) dataset is used to train the model and assess its performance in terms of F1 score, recall, accuracy, and precision. Our results demonstrated that, for the AD dataset, the voting classifier achieved the greatest validation accuracy of 96%. Therefore, with precise identification, ML algorithms have the potential to significantly reduce the annual mortality rates from Alzheimer's disease.

Keywords- Alzheimer's disease (AD), Predict, Machine Learning, OASIS, etc

I. INTRODUCTION

A degenerative neurological condition that mainly affects those 65 and older, Alzheimer's disease (AD) causes severe cognitive deterioration and has a major influence on day-to-day functioning. It is the most common cause of dementia, a term that includes a variety of symptoms that impair thinking, memory, and social skills to the point where they become disruptive to day-to-day functioning. AD affects about 6.5 million people worldwide, and as the population ages, the number is expected to increase. Three stages of the disease—early, middle, and late—usually start with minor memory loss that gets worse with time. In the latter stages, people may suffer from severe cognitive impairment and need help with everyday activities. Memory loss, trouble thinking and solving problems, and behavioral and personality changes are some of the symptoms.

Alzheimer's disease is thought to be caused by a confluence of lifestyle, environmental, and genetic factors, while the precise etiology is yet unknown. Age is a major risk factor since the chance of having AD rises dramatically with age; genetics can influence susceptibility because of family history; and medical disorders such untreated depression, head trauma, and hypertension are associated with a higher risk. The disease is typified by the buildup of aberrant protein structures in the brain called tangles (made of tau protein) and plaques (formed of amyloid beta), which cause cell death and damage to neural connections, hence causing cognitive impairment. Alzheimer's disease has no known cure, although there are a number of therapies that try to control symptoms and reduce the illness's course.

These include behavioral and psychological therapies that help address emotional difficulties, drugs that can momentarily enhance memory and cognition-related symptoms, and lifestyle modifications like consistent exercise and a healthy diet that may promote cognitive health. Given its increasing prevalence, Alzheimer's disease poses a serious public health concern. Effective management of the illness depends on knowledge of its symptoms, causes, and potential treatments. There is optimism for discoveries that could eventually prevent or cure this crippling illness as research advances.

II. LITERATURE REVIEW

P. Khan, M.F. Kader, S.R. Islam, A.B. Rahman, M.S. Kamal, M.U. Toha, K.S. Kwak, “Machine learning and deep learning approaches for brain disease diagnosis: principles and recent advances” [1], The impact of brain illnesses on global health emphasizes how crucial early and accurate diagnosis is to improving treatment outcomes. In this study, the origins and advancements of machine learning (ML) and deep learning (DL) methods for brain disease diagnosis are investigated. The article provides an overview of the typical diagnostic challenges and explains how ML/DL technologies overcome them through precise, scalable, and automated analysis. The latest algorithms, feature extraction methods, and imaging modalities—such as MRI, CT, and PET—are thoroughly examined, with an emphasis on how they might be used to detect brain cancers, Parkinson's disease, and Alzheimer's disease. Along with discussing potential solutions and future research directions, the paper also addresses problems including interpretability, data scarcity, and the computing demands of ML/DL systems. This project aims to give academics and practitioners advice on how to use machine learning and deep learning to improve the diagnosis of brain disorders.

A. Mehmood, M. Maqsood, M. Bashir, Y. Shuyuan, “A deep Siamese convolution neural network for multi-class classification of Alzheimer disease” [2], To properly manage Alzheimer's disease (AD), a progressive neurological disorder, precise diagnostic methods are essential. This paper proposes a deep Siamese convolutional neural network (CNN) for the multi-class categorization of Alzheimer's disease stages, including mild cognitive impairment, severe AD, and healthy controls. To improve classification accuracy and address imbalanced data, the model makes advantage of paired input structures and advanced feature extraction techniques. Experiments on publicly available datasets demonstrate that the model is more accurate and efficient than traditional methods at identifying subtle differences between stages of sickness. With significant implications for early detection and customized treatment regimens, this study shows how deep learning techniques can enhance Alzheimer's disease diagnostic tools.

F.J. Martinez-Murcia, A. Ortiz, J.M. Gorriz, J. Ramirez, D.Castillo-Barnes, “Studying the manifold structure of Alzheimer’s disease: a deep learning approach using convolutional auto-encoders” [3], Many pathological features of Alzheimer's disease (AD) are hard to characterize using standard diagnostic methods. This study employs a deep learning technique that makes use of convolutional auto-encoders to examine the multifaceted structure of AD progression. Through the identification of latent features, the approach provides insight into the patterns and progression of disease. Experiments using publicly available datasets verify the method's capacity to detect AD stages and uncover underlying structures in the data. The study demonstrates how deep learning techniques can advance our understanding and the diagnosis of Alzheimer's disease.

H.A. Helaly, M. Badawy, A.Y. Haikal, “Deep learning approach for early detection of Alzheimer’s disease” [4], Early identification is essential for Alzheimer's disease (AD) management and effective treatments. This study presents a deep learning-based technique for early AD detection using neuroimaging data. The proposed methodology uses advanced neural network topologies to extract and analyze tiny indicators of cognitive decline. The technique also shows how automated diagnostic tools may be integrated into clinical workflows, offering a valuable tool for improving early diagnosis and treatment strategies for Alzheimer's disease.

C. Kavitha, V. Mani, S.R. Srividhya, O.I. Khalaf, C.A.T. Romero, “Early-stage Alzheimer’s disease prediction using machine learning models” [5], Initial identification is crucial for prompt diagnosis and treatment of Alzheimer's disease (AD), a progressive neurological illness. The robustness, classification accuracy, and feature selection efficacy of various machine learning methods are assessed. The work provides important insights for public health initiatives aimed at Alzheimer's disease and highlights the potential of ML-based systems in improving diagnosis accuracy and enabling early clinical decision-making.

T.M. Ghazal, G. Issa, “Alzheimer disease detection

ISSN: 2349-6002

empowered with transfer learning" [6], Accurate diagnosis of Alzheimer's disease (AD) is essential for effective patient care and management. This paper proposes a novel transfer learning approach for AD detection that utilizes pre-trained deep learning models to get over the drawbacks of tiny labeled datasets in medical imaging. Experimental results on benchmark datasets show that the transfer learning strategy works better than traditional methods in terms of precision, sensitivity, and resilience when it comes to identifying AD phases. This work illustrates the scalable and efficient application of transfer learning in Alzheimer's disease diagnosis.

R. Gaudiuso, E. Ewusi-Annan, W. Xia, N. Melikechi, "Diagnosis of Alzheimer's disease using laser-induced breakdown spectroscopy and machine learning" [7], Alzheimer's disease (AD) has a complex pathophysiology, making diagnosis challenging in clinical practice. In this work, we examine the potential of laser-induced breakdown spectroscopy (LIBS) in combination with machine learning (ML) techniques for the diagnosis of AD. LIBS analyzes the elemental composition of biological materials, classifies the data, and uses machine learning models to differentiate between AD and non-AD instances.

III. PROBLEM STATEMENT

To develop more robust, accurate, and efficient diagnostic tools for AD. This could lead to earlier interventions, better patient outcomes, and ultimately, improved quality of life for those affected by Alzheimer's disease.

IV. OBJECTIVE

1. To develop an ensemble machine learning model combining multiple algorithms.
2. To implement and compare various feature selection techniques to identify the most relevant biomarkers and neuroimaging features, reducing data redundancy and improving model efficiency.
3. To evaluate the proposed model's performance using the OASIS dataset, focusing on metrics such as F1 score, recall, accuracy, and precision for early-stage AD detection and multi-class stage classification.
4. To compare the proposed ensemble model's performance with existing approaches in AD diagnosis, including deep learning and transfer learning methods, to assess its effectiveness and potential clinical applicability.
5. To investigate the model's interpretability and explore ways to provide insights into the most influential features for AD prediction, enhancing its potential for integration into clinical decision-making processes.

V. DESIGN METHODOLOGIES

The machine learning algorithms uses to foresee Alzheimer's disease in three main steps: data gathering, preprocessing, and prediction. The original dataset is loaded using Pandas, and the necessary preprocessing libraries are imported. The accuracy of the machine learning algorithm is lowered by the input dataset's consistency and redundancy. In this study, superfluous values and attributes are removed from the data before it is fed into a machine learning system for the best results. After preprocessing, the data was sporadically poured into training and testing. Data is randomly divided in an 80:20 ratio, meaning that 20% of the data is used for testing and 80% is used to train the model. The system's procedure for early Alzheimer's disease prediction is shown in Figure 1.

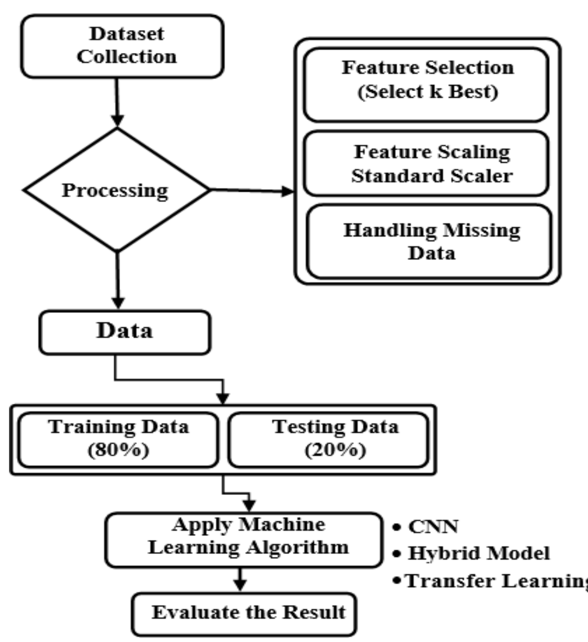

Figure 1: Working flowchart of the proposed system to predict AD

Model Type	Optimizer	Learning Rate	Validation Accuracy (%)	Test Accuracy (%)
CNN (Custom)	Adam	0.001	98.9	98.7
Hybrid Model	Adam	0.001	85.3	94.7
Transfer Learning	Adam	0.001	91.7	97.6
CNN (Custom)	SGD	0.01	96.2	96.4
Hybrid Model	SGD	0.01	90.5	95.6
Transfer Learning	SGD	0.01	93.4	97.5
CNN (Custom)	RMSprop	0.0001	97.5	98.2
Hybrid Model	RMSprop	0.0001	89.9	94.5
Transfer Learning	RMSprop	0.0001	92.2	97.2

1. Data Collection

The dataset for this project was sourced from publicly available medical imaging repositories, such as the Alzheimer's disease Neuroimaging Initiative (ADNI) and Kaggle MRI datasets. The dataset includes MRI scans labeled into four categories: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. The data was acquired in compliance with ethical guidelines, ensuring patient anonymity and adherence to medical data regulations.

2. Data Pre-processing Image Resizing

All MRI images were resized to a uniform dimension of 128x128 pixels to ensure compatibility with the deep learning model. Normalization: Pixel intensity values were normalized to a range of [0, 1] to improve model convergence. Data Augmentation: Techniques such as rotation, flipping, zooming, and contrast adjustment were applied to increase dataset diversity and mitigate overfitting. Noise Reduction: Filters were applied to enhance image clarity and remove artifacts that could distort predictions. Train-Test Split: The dataset was divided into training (80%), validation (10%), and test (10%) subsets. Data Generators: The Image Data Generator function was used for preprocessing and augmentation. train_datagen and val_datagen were initialized with rescaling to normalize pixel values. Train_generator and val_generator were set up to feed images into the model in batches of 32 with categorical labels.

3. Machine Learning Models

Traditional machine learning models such as Support Vector Machines (SVM), Random Forest, and K-Nearest Neighbors (KNN) were initially used for baseline classification. Feature extraction was performed using Principal Component Analysis (PCA) to reduce dimensionality and improve model efficiency. Although machine learning models provided reasonable performance, deep learning-based approaches significantly outperformed them.

4. Deep Learning Models

A Convolutional Neural Network (CNN) was designed for automated feature extraction and classification. The CNN architecture consisted of multiple convolutional layers, batch normalization, max-pooling, dropout layers, and a fully connected dense layer with a softmax activation function. The model was trained using the Adam optimizer with categorical cross-entropy loss.

5. Hybrid Models

A hybrid approach was explored by combining CNNs with traditional machine learning classifiers. Features extracted from the CNN were fed into a Random Forest or SVM classifier to assess potential improvements in prediction accuracy.

6. Transfer Learning Models

Pre-trained deep learning models such as VGG16, ResNet50, and EfficientNet-B0 were utilized for feature extraction. Transfer learning was employed to fine-tune these models on the Alzheimer's MRI dataset, leveraging their pre-trained weights from large-scale medical imaging datasets. Among these models, ResNet50 achieved the best balance of accuracy and computational efficiency.

7. Model Training & Evaluation Hyperparameter Tuning:

Learning rate, batch size, dropout rates, and activation functions were optimized. Optimizers Used: The following nine optimizers were tested for each model: Adam (learning rates: 0.01, 0.001, 0.0001) SGD with momentum (learning rates: 0.01, 0.001, 0.0001)

RMSprop (learning rates: 0.01, 0.001, 0.0001)

Evaluation Metrics: Accuracy, Precision, Recall, and F1-Score were computed to assess model performance. Confusion matrices were analyzed to evaluate misclassification patterns. Training Strategies: Early stopping was implemented to prevent overfitting. Data augmentation was leveraged to improve generalization

8. Matrix Analysis
Minimal misclassifications were found in the best-performing models, particularly in distinguishing No Impairment and Moderate Impairment. Hybrid models faced difficulties in correctly classifying Mild and Very Mild Impairment cases, leading to reduced overall performance.

VI. RESULTS

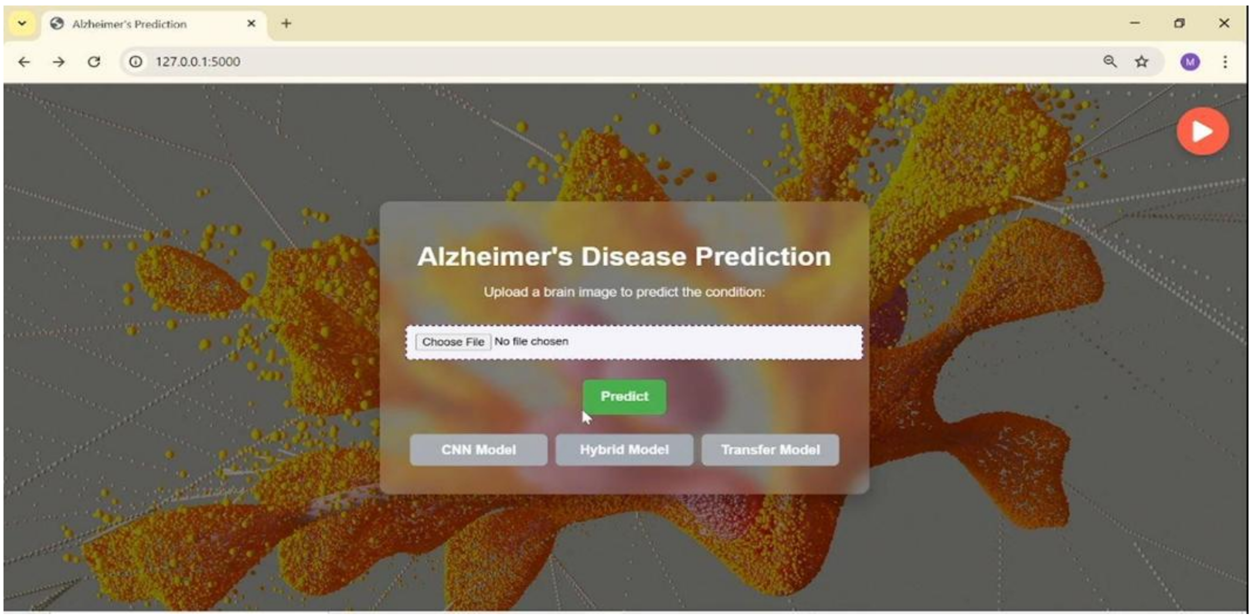

Figure 1: Dashboard

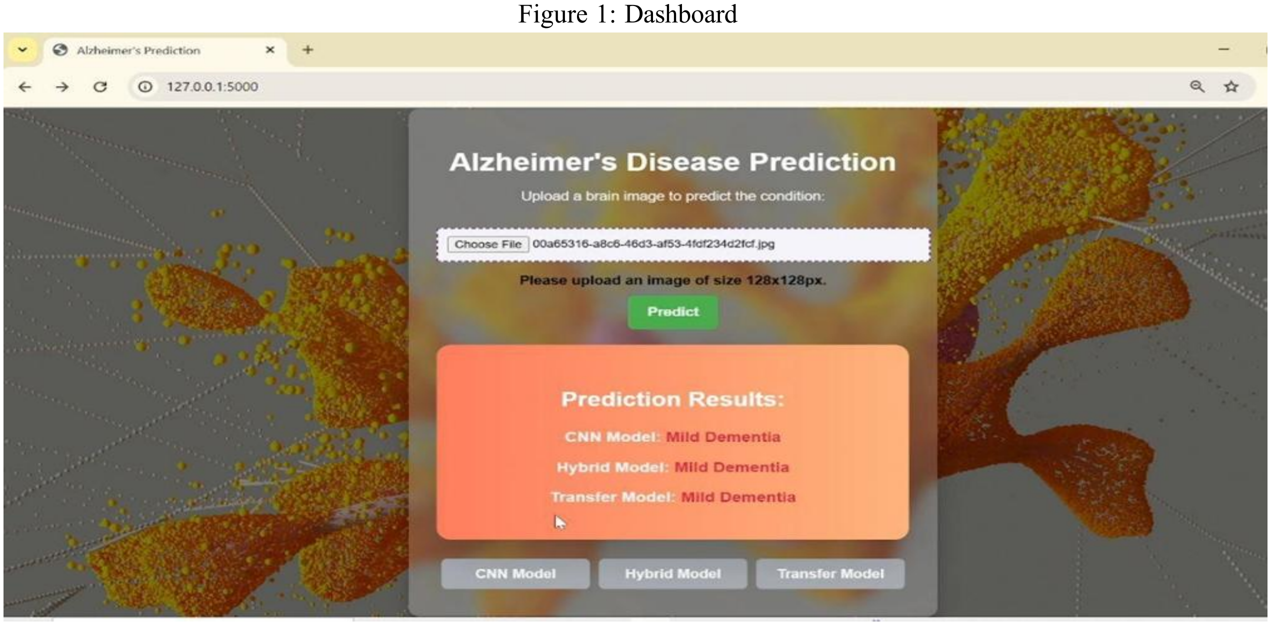

Figure 2: Prediction Results according to Condition

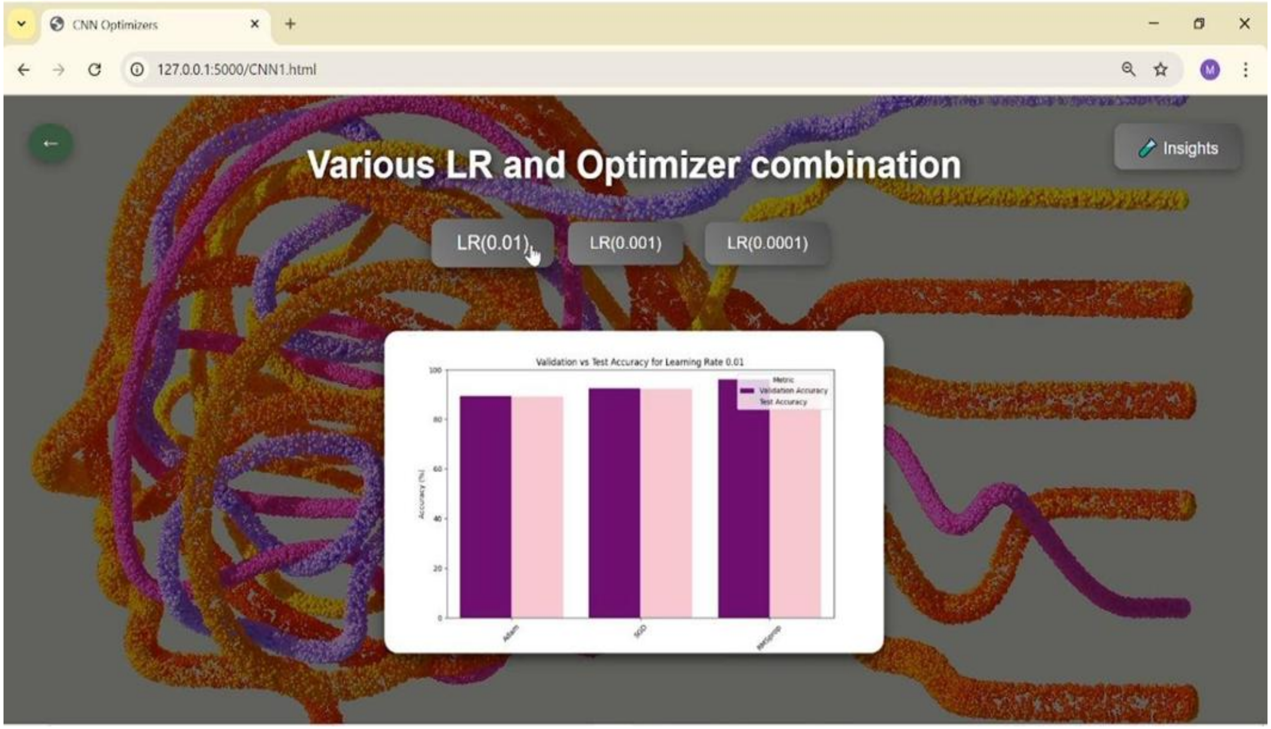

Figure 3: Combination of LR & Optimizer-1

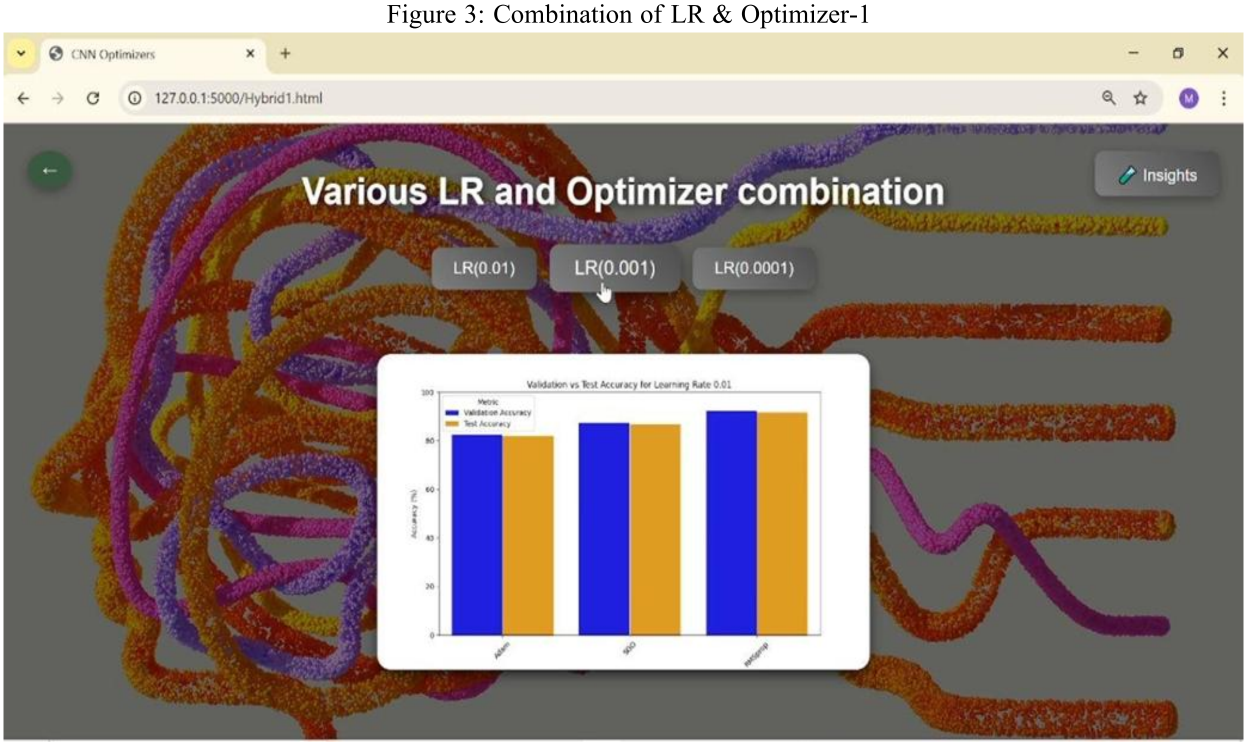

Figure 4: Combination of LR & Optimizer-2

VII. CONCLUSION

Numerous attempts have been made to identify Alzheimer's disease using a range of machine learning algorithms and micro-simulation techniques, according to the literature review; however, identifying pertinent traits that could potentially identify Alzheimer's disease early on remains a challenge. This study has employed several methods, including GaussianNB, Decision Tree, Random

Forest, XGBoost, Voting Classifier, and GradientBoost, to determine the most dependable component for Alzheimer's disease prediction. The accuracy of the machine learning algorithms is increased through the use of feature selection. The voting classifier has the highest validation accuracy of 96% on the AD test data, produced a more advantageous result.

REFERENCE

[1] P. Khan, M.F. Kader, S.R. Islam, A.B. Rahman, M.S. Kamal, M.U. Toha, K.S. Kwak, Machine learning and deep learning approaches for brain disease diagnosis: principles and recent advances. IEEE Access 9, 37622–37655 (2021)

[2] A. Mehmood, M. Maqsood, M. Bashir, Y. Shuyuan, A deep Siamese convolution neural network for multi-class classification of Alzheimer disease. Brain Sci. 10(2), 84 (2020)

[3] F.J. Martinez-Murcia, A. Ortiz, J.M. Gorriz, J. Ramirez, D.Castillo-Barnes, Studying the manifold structure of Alzheimer's disease: a deep learning approach using convolutional autoencoders. IEEE J. Biomed. Health Inform. 24(1), 17–26 (2019)

[4] H.A. Helaly, M. Badawy, A.Y. Haikal, Deep learning approach for early detection of Alzheimer's disease. Cogn. Comput. 14, 1711 (2021)

[5] C. Kavitha, V. Mani, S.R. Srividhya, O.I. Khalaf, C.A.T. Romero, Early-stage Alzheimer's disease prediction using machine learning models. Front. Public Health (2022). https:// doi. org/ 10. 3389/ fpubh. 2022. 853294

[6] T.M. Ghazal, G. Issa, Alzheimer disease detection empowered with transfer learning. Comput. Mater. Continua 70(3), 5005– 5019 (2022)

[7] R. Gaudiuso, E. Ewusi-Annan, W. Xia, N. Melikechi, Diagnosis of Alzheimer's disease using laserinduced breakdown spectroscopy and machine learning. Spectrochim. Acta Part B 171,105931 (2020)