IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Alzhimer's Disease Prediction Using Modern Machine Learning Techniques

R. Abinaya[1], M. Ramanamadhu[2], G. Mohanram[3], M. Manavalan[4]

[1]R. Abinaya Assistant Professor, CSE, Sri Ramakrishna Institute of Technology, Coimbatore, Tamil Nadu, India
[2]M. Ramanamadhu, CSE, Sri Ramakrishna Institute of Technology, Coimbatore, Tamil Nadu, India
[3]G. Mohanram, CSE, Sri Ramakrishna Institute of Technology, Coimbatore, Tamil Nadu, India
[4]M. Manavalan, CSE, Sri Ramakrishna Institute of Technology, Coimbatore, Tamil Nadu, India.

Abstract – Alzheimer's sickness ad is a identical all though model training we also use progressive neurodegenerative disease and a leading reason of dementia worldwide. Alzheimer's sickness is characterised with the aid of cognitive decline reminiscence impairment and behavioral adjustments that have an effect on the best of life of tens of millions of patients and their caregivers leading to work and healthcare prices because the populace a while the prevalence of alzheimer's disorder is predicted to growth similarly to treatment therefore early diagnosis and intervention are vital due to the fact they have the potential to slow the ailment improve patient effects and enhance remedy selections this observe investigates the use of machine learning algorithms for the prediction and early detection of alzheimer's ailment the use of scientific statistics and biological characteristics effectiveness of the K-Nearest Neighbors KNN, Logistic Regression, Random forests and ensemble mastering techniques to are expecting the possibility of alzheimer's disease primarily based on numerical facts. By combining scientific facts including affected person demographics cognitive exams and clinical records and biological features together with biomarker tiers and genetic information, we focus on styles and signs that can lead to early caution of the ailment to expand an awesome and accurate model we first use loads of information first the information is cleaned to get rid of missing values and eliminate contradictions that could have an effect on the have a look at then we normalize the profile so that each one features are similar which enables prevent bias and guarantees that all supports are a unique selection process to perceive the most vital and applicable information reduce the last statistics and reduce the danger of overfitting.

[1] INTRODUCTION

Alzheimer's disease (AD) is a chronic neurodegenerative disease that mostly affects the elderly, causing severe deterioration in memory, cognition and behavior. As one of the leading causes of dementia, AD accounts for 60-80% of dementia cases worldwide. This disease causes gradual but irreversible damage to brain cells, ultimately leading to severe intellectual deficits and loss of basic strength and life functions. It poses a serious challenge to the global health system, not only because of the increasing number of elderly people, but also because of its serious impact on patients, caregivers and all people. As the elderly population is expected to increase in the next few years, the incidence of Alzheimer's disease is expected to reach epidemic proportions, indicating an urgent need for improved diagnosis and treatment.

Despite years of research, the exact cause of Alzheimer's disease is still not understood, but it is believed that a combination of genetics, environment and lifestyle contribute to its development. The accumulation of beta-amyloid plaques and neurofibrillary tangles in the brain characterizes the disease. These pathological changes disrupt normal neuronal communication and lead to cell death, causing symptoms. However, when these symptoms appear, serious and often irreversible brain damage has already occurred. This emphasizes the importance of detecting Alzheimer's disease as early as possible, preferably in the preclinical stages or early symptoms when treatments have a chance of changing the course of the disease.

1.1. GENERAL INTRODUCTION

This project focuses on the development of machine learning-based prediction systems to improve the accuracy of disease diagnosis using ensemble learning techniques. Ensemble learning combines several algorithms such as K-Nearest Neighbors (KNN), Random Forest (RF), and Logistic Regression (LR) to improve prediction reliability. By analyzing clinical and biological data, the system ensures early detection and better decision-making. Additionally, a Django-based web framework is integrated to provide a user-friendly interface that allows members of health professions to enter patient data and visualize results in real-time. This approach provides a scalable, efficient and accurate solution for medical diagnosis. This model is optimized through data preprocessing, feature selection, and performance evaluation to ensure high accuracy and robustness. Future improvements include deep learning models and real-time adaptability, further improving diagnostic capabilities.

1.2. PROBLEM STATEMENT

Alzheimer's disease (AD) is a progressive and irreversible neurodegenerative disease that severely affects cognitive functions, including memory, thinking, and behavior. It is the most common form of dementia, accounting for 60–80% of patients worldwide and causing health, financial, and social problems. As the world's population ages, the incidence of Alzheimer's disease is expected to increase, with a major impact on the healthcare system. Early detection of Alzheimer's disease is important for timely intervention, as it can help reduce cognitive decline, improve patient outcomes, and provide better patient care and care.

However, current diagnostic methods often rely on observation, neuroimaging, and biomarker analysis, which can be costly, time-consuming, and disruptive. Treatment is most effective when the disease is diagnosed. Traditional approaches are also limited by their reliance on clinical studies, which may not capture the structural changes in the early stages of Alzheimer’s disease. In addition, the high-dimensional nature of clinical and biological data makes it difficult to identify meaningful predictive models using traditional methods. Invasive diagnostic tool to aid in early detection of Alzheimer’s disease. Machine learning (ML) offers the promise of using complex data to discover patterns and indicators that are not obvious from standard diagnostic methods.

1.3. Algorithm

1. Random Forest Classifier
2. Logistic Regression
3. KNN
4. Ensemble Learning

Random Forest Classifier: One component that makes up the supervised learning approach is Random Forest. It can be used for machine learning problems that involve regression and classification. It depicts a model that employs multiple decision trees on various input dataset subsets, averaging the outcomes to raise the dataset's anticipated accuracy. The random forest method employs forecasts from every decision tree rather than relying just on one, and it makes predictions on the basis of the responses received of most of those projections.

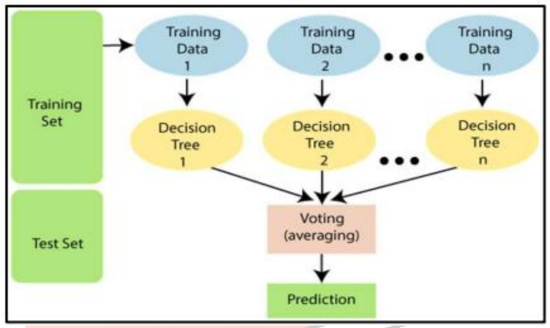


Figure 1.3.1: Flowchart - Random Forest Classifier.

Logistic Regression: Logistic regression is a popular machine learning algorithm that falls under the umbrella of supervised learning. It is utilised to forecast a categorical dependent variable by utilising a predefined collection of independent factors. Logistic regression is used to predict the output for a categorical dependent variable. The outcome must therefore be an individual or categorical value. It may be either false or true, however probabilistic values in the range of 0 and 1 are returned instead of an exact number between 0 and 1.

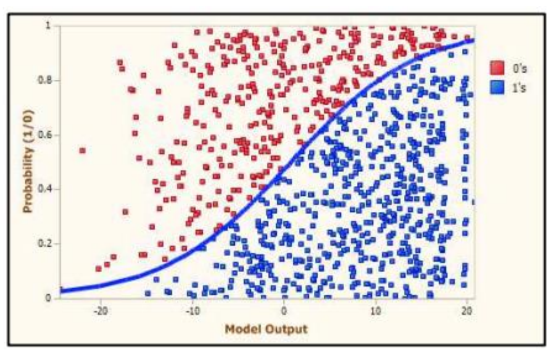


Figure 1.3.2: Logistic Regression

KNN Algorithm: K - Nearest Neighbors is another machine learning technique that uses the supervised learning approach as its foundation. Based on the supposition that the new case/data and the previous cases are comparable, the K-NN method assigns the new case to the group which is most similar to the current groups.

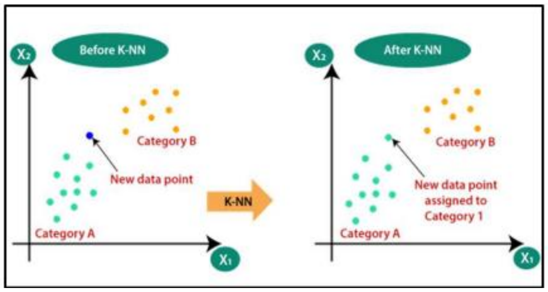


Figure 1.3.2: K-NN Classification.

Ensemble Learning: Ensemble learning is an advanced machine learning technique that improves predictability by combining several models rather than relying on a single algorithm. This approach uses the strengths of different models to improve reliability, reduce errors, and prevent over adaptation. The three main types of ensemble learning are bags, demand and stacks. Bagging, used in algorithms such as Randallswald, trains several models on different data subgroups to reduce variance. Boost, which can be seen in models such as Adaboost and Xgboost, improves weak models by fixing errors one after another. Stacking combines several models that use META models for final predictions to improve the total output. In this project, K-Nearest Neighbor (KNN), Random Forest (RF), and Logistic Regression (LR) are integrated using ensemble learning to improve Alzheimer's disease prediction. By combining these models, the system ensures higher accuracy, better generalization, and more reliable decision-making. This makes it a valuable tool for early diagnosis of healthcare.

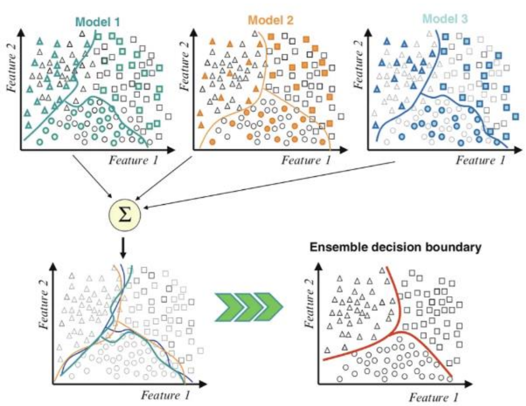


Figure 1.3.2: Ensemble learning

2. EXISTING METHODOLOGY

Existing methodologies for predicting Alzheimer's disease (AD) classify patients using clinical and biological data using support from four neighborhoods (KNNs), logistics regression, random forests, and vector machine (SVM). This process starts with data collection and preprocessing, dealing with the lack of values, removing unrelated features, and normalizing the data for consistency. Each algorithm is then trained using structured data records to improve performance using hyperparameter mood. Trained models are evaluated using accuracy, accuracy, recall, and F1 scores to determine their effectiveness in advertising classification. Cross-validation techniques are used to prevent over adaptation and to ensure that the model is well systematized for new data. Finally, we analyze the results to identify the most important clinical features that contribute to AD prediction. This methodology provides a systematic approach for early detection without the inclusion of ensemble learners or image-based analysis to focus on structured numerical data for accurate classification.

2.1. DISADVANTAGES

A drawback of existing methodologies is the trust in individual machine learning models that can limit prediction accuracy and robustness. Because all algorithms have their own advantages and disadvantages, usage can lead to inconsistent predictions without ensemble learning. Furthermore, the lack of image-based analysis means that no critical visual patterns from brain scans are used, which can provide valuable diagnostic knowledge. This limitation may lead to a decrease in model sensitivity to detection of Alzheimer's disease at an early stage, when subtle changes in brain structure play an important role.

3.PROPOSED METHODOLOGY

The proposed methodology improves the prediction of Alzheimer's disease (AD) by integrating ensemble learning techniques and image-based analysis to improve accuracy and reliability. In contrast to traditional approaches that rely on individual machine learning models, ensemble learning combines several algorithms such as K - Nearest Neighbors (KNN), logistics regression, random forests, and vector machines (SVM) support to create more robust and generalized predictive models. By aggregating the strengths of these classifiers, the ensemble model minimizes errors and improves diagnostic accuracy. Furthermore, the methodology includes image-based analysis with deep learning techniques for processing MRI or PET scans. The folding network (CNN) is used to extract important features from brain images and identify structural abnormalities related to AD progress. The combination of clinical data and imaging improves overall predictive power and allows for early detection with greater confidence.

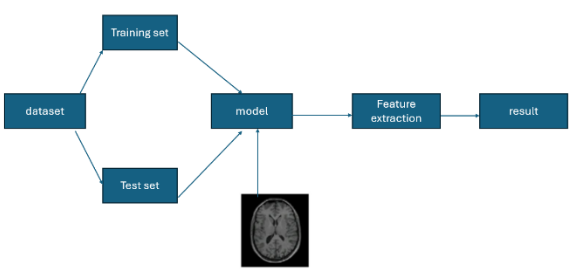
Follow a structured pipeline methodology that includes data preprocessing, feature selection, model training and performance evaluation to ensure efficiency. Date and enhancement techniques are used to handle missing values and improve the model of the model. The performance of the ensemble model is evaluated using metrics such as accuracy, accuracy, recall, and F1 scores to ensure reliability for real-world clinical applications.

By integrating structured clinical data into advanced, deep learning techniques, this methodology provides a comprehensive and scalable solution for Alzheimer's disease prediction, facilitating early intervention and improved patient care.

3.1. ADVANTAGES

The main advantage lies in the combination of ensemble learning and structured data analysis. By using several machine learning models such as K-Nearest Neighbors (KNN), Logistic Regression, Random Forests, and Ensemble learning, the system improves prediction accuracy and reduces individual models. The integration of image-based analyses with distinctive extraction techniques (no deep learning) allows for the identification of important patterns from medical images. This approach improves model interpretability and is more suitable for efficiency and reliability as well as clinical applications.

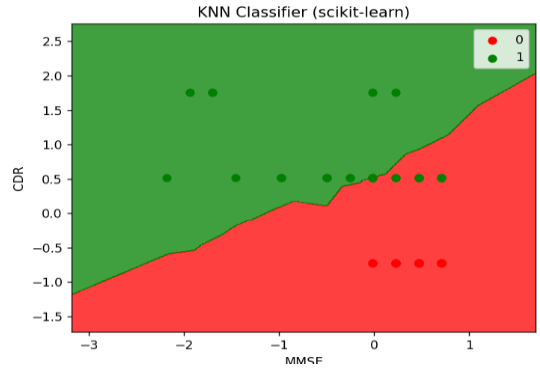
4.2. BLOCK DIAGRAM



MRI image

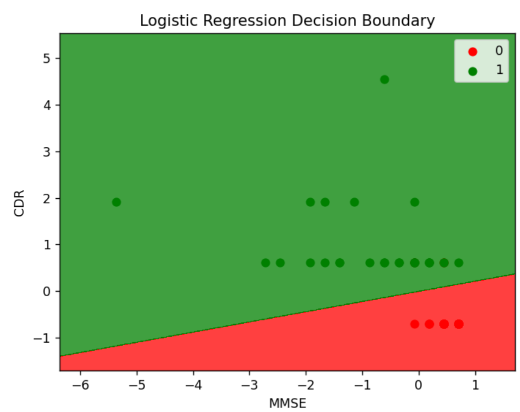
5.RESULTS

5.1 Results on each and every algorithm

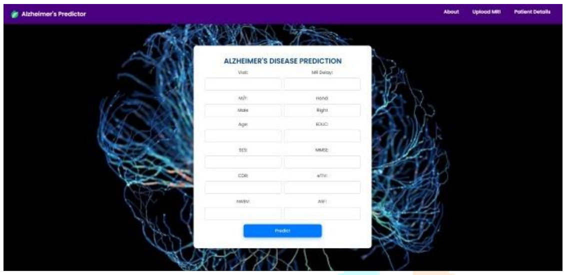


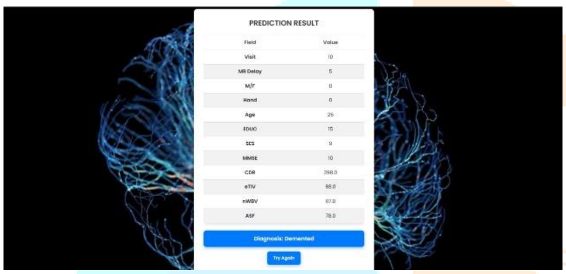
5.1.1 Decision boundary of KNN

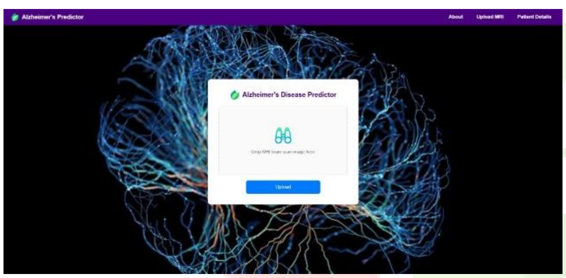
Decision boundary created by a K-Nearest Neighbors (KNN) classifier using the scikit-learn library. The plot depicts two features: "MMSE" (x-axis) and "CDR" (y-axis), where the classifier predicts two classes labeled `0` (red) and `1` (green). The colored regions represent the predicted class for any point in the feature space, while the points themselves are the actual data samples, with red circles for class `0` and green circles for class `1`. The decision boundary separates the regions where the KNN model predicts different classes.

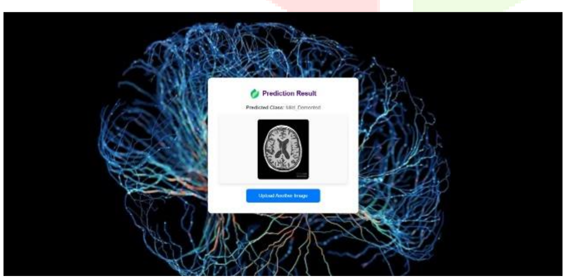


Decision boundary of a Logistic Regression classifier. It shows two features: "MMSE" (x-axis) and "CDR" (y-axis). The classifier predicts two classes, labeled `0` (red) and `1` (green). The green region indicates where the model predicts class `1`, and the red region indicates where it predicts class `0`. The points represent the data samples, with green circles for class `1` and red circles for class `0`. The decision boundary is linear, as expected from Logistic Regression, separating the two classes based on a linear combination of the features.









6.CONCLUSION

The Alzheimer's Prediction Project uses machine learning techniques, particularly K-Nearest Neighbors (KNN), Logistic Regression, and Random Forest, to analyze numerical data and identify underlying patterns associated with the disease. This project includes data preprocessing that resolves missing values and evaluates numerical features to ensure that the dataset is suitable for learning the model. Each algorithm has its own advantages; among them, KNN provides simple classification based on proximity, Logistic Regression provides prediction, and Random Forest improves accuracy by aggregating decision trees. Evaluation: Accuracy, precision, recall, and ROC-AUC scores that show their performance in predicting Alzheimer's disease. The program aims to help doctors with early diagnosis and intervention by creating a reliable prediction model, ultimately improving patient outcomes. Through the integration of data analysis and machine learning, the project not only aims to increase the accuracy of diagnosis, but also leads to a deeper understanding of the factors associated with Alzheimer's disease.

7.FUTURE WORK

Future work in the Alzheimer's Disease Prediction Project has the potential to improve the accuracy and generalizability of the model. One promising strategy is to expand the data to include a broader range of patients. Currently, datasets tend to focus on specific populations, which may limit the model's ability to generalize across populations. By combining data from different age groups, race, gender, and health history, the model can capture multiple aspects of Alzheimer's risk and improve the method. The method can be applied to patients of different ethnic backgrounds. In addition to expanding the population data, it is important to include other medical conditions. For example, genetic markers (like the APOE gene), lifestyle factors (like diet and physical activity), or advanced neuroimaging data can provide deeper insights into factors that increase the risk of Alzheimer's. These additional features will increase the predictive power of the model by integrating multiple variables known to influence the onset and progression of Alzheimer's disease, such as deep learning or integration.

8.REFERENCES

[1] Kumar, U, "Applications of Machine Learning in Disease Prescreening" pp. 1052 1084, 2018.

[2] Khourdifi, Y., Bahaj, M, "Heart Disease Prediction and classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization", IJIES, vol. 12, no. 1, 2019.

[3] Mohan, S., Thirumalai, C., Srivastava, G. "Effective heart disease prediction using hybrid machine learning techniques", IEEE, vol. 7, pp. 81542-81554, 2019.

[4] Pingale, K., Surwase, S., Kulkarni, V, Sarage, S., Karve, A, "Disease Prediction Using Machine Learning", IRJET, vol. 6, no. 12, 2019.

[5] Wang, W., Chakraborty, G., Chakraborty, B. "Predicting the risk of chronic kidney disease (ckd) using machine learning algorithm", MDPI Applied Sciences, vol. 11, no. 1, pp. 202,2020.

[6] Benjamin, F. D. H., "Impact Of Ensemble Learning Algorithms Towards Accurate Heart Disease Prediction", ICTACT J. Soft Compute, vol. 10, pp. 2084-2089, 2020.

[7] Chatterjee, A.Roy, S, "A Machine Learning Approach to Prevent Cancer", IGI Global, pp.112-141, 2021.

[8] Mallela, C. R., Bhavani, L. R., Ankayarkanni, B., "Disease Prediction Using Machine Learning Techniques", IEEE, pp. 962-966, 2021.

[9] Gomathy, CK and Naidu, Mr. A Rohith, "The Prediction Of Disease Using Machine Learning Techniques", International Journal of Scientific Research in Engineering and Management (IJSREM), Vol. 5, no. 7, 2021.

[10] Pragya Pranjal, Saahil Mallick, Ananya Das, Mohit Rajan, "Alzheimer's Disease Prediction using Modern Machine Learning Techniques", 2024