International Journal of application or innovation in engineering and management

Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/910-915

ISSN: 2319-4847

Utilizing State-of-the-Art Machine Learning Methods for Alzheimer's Disease Prognosis

[1] B. Naresh, [2] P. Sravya,

[1]Assistant Professor, Megha Institute of Engineering & Technology for Women, Ghatkesar.
[2] MCA Student, Megha Institute of Engineering & Technology for Women, Ghatkesar.

Article Info

Received: 30-04-2025 ***Revised: 06 -06-2025*** ***Accepted: 17-06-2025*** ***Published:28/06/2025***

Abstract –

Machine Learning (ML) is the most well-known use of AI because of the way it is changing the face of research. The use of machine learning to diagnose Alzheimer's disease is the focus of this investigation. On a worldwide scale, Alzheimer's disease takes the lives of countless individuals. The use of machine learning methods allows for the determination of Alzheimer's disease by factoring in things like age, increased cholesterol levels, chest pain, etc. In this paper, four supervised ML algorithms are utilized: K-Nearest Neighbour, Random Forest, Artificial Neural Networks, and Logistic Regression. The accuracy of the predictions made by each of these algorithms ranges from 87% to 90%. The use of machine learning in healthcare, disease prediction, Alzheimer's disease, logistic regression, random forests, artificial neural networks, and KNN are all relevant terms.

I. INTRODUCTION

Predictions based on historical data are the focus of machine learning. The goal of machine learning, a subfield of artificial intelligence, is to create systems that learn from data and algorithms in a way that is consistent with human performance. The machine learning process consists of two steps: Testing and training must take precedence. For decades, machine learning software has struggled to diagnose conditions based on patient reports of symptoms and medical history. Machine learning technology is a game-changer when it comes to healthcare. Using the entire breadth of machine learning, we are tracking patients' health. We can create models that quickly clean, analyze, and provide results from data with the use of machine learning models. With this strategy, doctors will be able to make better diagnoses, which means patients will get better treatment and healthcare providers will be able to meet their patients' expectations for quality. Healthcare is the ideal setting for introducing ML to the medical field. This study will focus on structured data in an effort to enhance the precision of massive datasets. The present model will use linear algorithms, decision trees, and KNN for illness prediction. The existing method is able to predict the occurrence of illnesses that are unique to a particular demographic and region. This method is limited to the prediction of certain illnesses. To predict the likelihood of illnesses, this system employs CNN algorithms and Big Data. Naive Bayesian, Decision Tree, and K-NN are the machine learning methods used by the system when dealing with S type data. Classifier using Random Forest: The supervised learning method includes Random Forest as one of its components. Applications in machine learning including classification and regression are possible with its help. This model illustrates how to increase the expected accuracy of a dataset by averaging the results of many decision trees applied to different subsets of the input dataset. In order to provide more accurate predictions, the random forest approach takes into account the results from the majority of decision trees rather than depending on just one.

International journal of application or innovation in engineering and management

Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/910-915

ISSN: 2319-4847

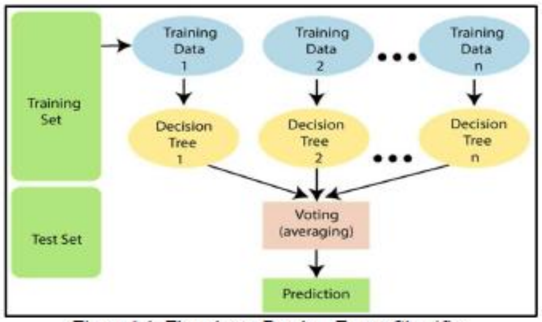

Figure 1.1: Flowchart - Random Forest Classifier.

Another machine learning method that builds on the supervised learning approach is K-Nearest Neighbor. The K-NN approach sorts new instances into groups according to how similar they are to existing ones, supposing that the new data and cases are equivalent.

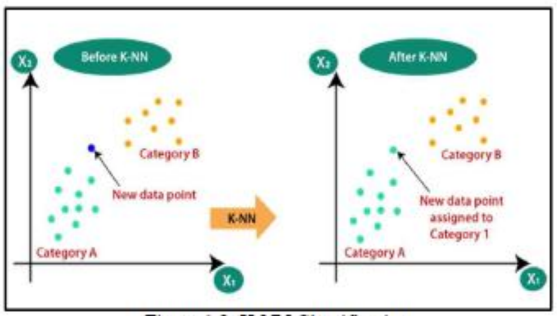

Figure 1.2: K-NN Classification.

"Artificial neural network" is a borrowing from the biological "neural networks" that form the basis of the human brain. An artificial neural network (ANN) functions similarly to a real brain in that it is composed of interconnected neurons. Nodes are the names given to these neurons. Dendrites are the building blocks of artificial neural networks, which in turn use cell nuclei as nodes, synapses as weights, and axons as their terminals.

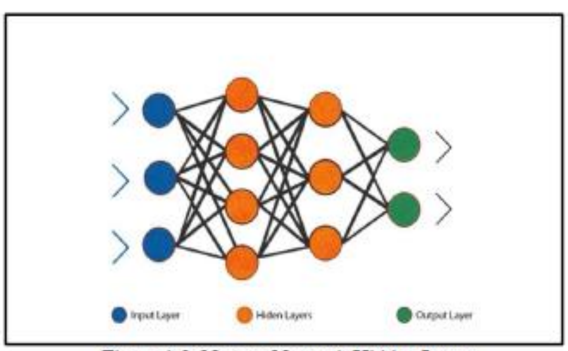

Figure 1.3: Neuron Network Hidden Layers.

One well-known supervised learning technique is logistic regression, which is used by machine learning algorithms. A categorical dependent variable may be predicted using it, with the help of a set of predetermined independent variables. If your dependent variable is categorical, you may use logistic regression to forecast its value. Results may only have a numeric or category quality. Instead of a precise integer between 0 and 1, it returns probabilistic values in the range of 0 and 1, indicating that it might be true or false.

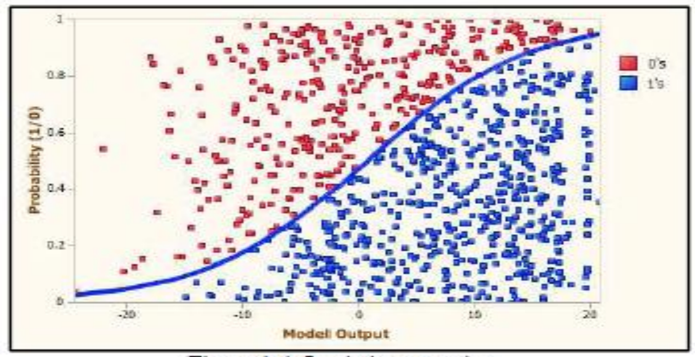

Figure 1.4: Logistic regression.

II. Literature Review

The use of ML approaches has been the main driver of the recent interest and innovation in the area of illness prediction and prevention [1]. An important study conducted by Gomathy (2021) establishes the groundwork for this area of inquiry,that illness prediction relies on computer learning. Following this, Mallela's contributions Published by IEEE in 2021 by Bhavani and Ankayarkanni examine in depth the many ML methods used for illness prediction [2]. Dear Chatterjee and the research of Roy (2020) on cancer prevention using ML approaches, especially in Chapter 7, provides valuable information for proactive cancer prevention [3]. Similarly, Tarigoppula, Rao, and Narayan (2013) investigate the feasibility of using machine learning algorithms for early Parkinson's disease prediction [4]. To improve the precision of cardiac illness prediction, Benjamin (2021) looks at ensemble learning techniques [5]. To demonstrate the use of ML in risk assessment, Wang, Chakraborty, and Chakraborty (2020) zero in on CKD risk prediction [6]. For a more comprehensive look of machine learning's role in illness pre-screening, see Kumar (2018) [7]. Introductory work on artificial neural networks for illness prediction was done by Rodvold, McLeod, and Murphy (2001) [8]. Additional works on the topic of symptom-based illness

International journal of application or innovation in engineering and management
Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/910-915
ISSN: 2319-4847

prediction include those of Pingale et al. (2019) [9] and Keniya et al. (2020) [10]. While Mohan et al. (2019) promote the use of hybrid ML approaches for the prediction of cardiac illness, Khourdifi and Bahaj (2018) improve ML strategies such as particle swarm modeling and ant colony optimization [11]. Taken as a whole, these citations demonstrate how machine learning is becoming more important for illness prediction and how flexible it is in dealing with many health-related issues.

Methodology

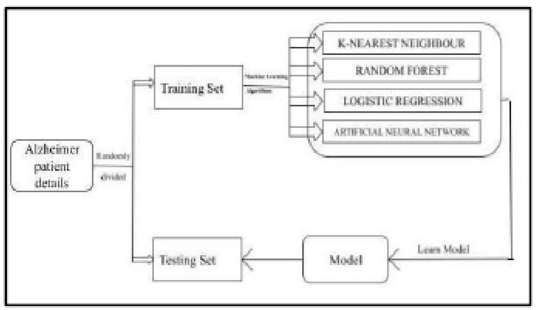

Figure 2.1: Methodology

First Step: Alzheimer's Database Getting a dataset on Alzheimer's illness was the initial stage in the technique. Information on demographics, health, cognitive abilities, and other pertinent elements is probably included in the dataset's many features or variables that characterize people and their traits. To train and test the machine learning models, this dataset is used. The Second Step: Dataset Segmentation At this stage, the dataset was partitioned into training and testing subsets. The purpose of this section is to build machine learning models that can use training data to identify links and patterns in data. In contrast, testing data will be used to evaluate the trained models' accuracy and performance. Third Step: Algorithms for Machine Learning To begin, we looked for trends in the dataset on Alzheimer's illness using four different machine learning methods [12]. The methods used were Logistic Regression, ANN, KNN, and Random Forest. Each algorithm takes a somewhat different approach to learning patterns in the training data and applying them to prediction. The fourth step is to test and calculate the accuracy. After the models were constructed using the training data, they were tested on the separate testing data subset. The purpose of this testing phase was to compare the models' ability to accurately predict outcomes or categories related to Alzheimer's disease. Each model's accuracy was assessed by comparing its predictions with the actual quantities in the testing data. Next, we divided the entire number of predictions by the total amount of accurate predictions to get the accuracy %. The research used this method to evaluate the KNN, random forest, ANN, and logistic regression machine learning algorithms' accuracy on the Alzheimer's dataset. This study is useful for understanding the performance of each algorithm and its ability to learn patterns from data in order to make accurate predictions about Alzheimer's disease.

Proposed Framework

To ensure correctness, we used many test scenarios. When it comes to picture datasets, random forest and ANN often outperform KNN and Logistic Regression [13]. In this case, we achieved random accuracies in random forest by raising the depth of the decision trees; in KNN, we raised the number of neighbors; in ANN, we progressively increased the epoch value; and in logistic regression, we increased the iterations to produce results.

Research Challenges

Despite the fact that machine learning algorithms can identify Alzheimer's disease, many research problems remain unsolved. One of the main issues is the lack of trustworthy data for training and testing algorithms.

As a complicated neurodegenerative illness, Alzheimer's makes it difficult to gather big datasets with varied demographic and clinical characteristics. Consequently, the algorithms' accuracy might be affected by the amount and quality of the data that is available for analysis. The need for algorithms to be transparent and interpretable presents an additional obstacle. It may get more challenging to comprehend the prediction process of more complex machine learning systems. This might be problematic when trying to convey the results to healthcare providers, patients, and caregivers. The accuracy of the algorithms might also be affected by the research population.

Factors such as age, sex, and race, for instance, may influence how precise the algorithms are. So, to make sure the algorithms can forecast Alzheimer's disease, it's important to test them on different populations. Last but not least, using machine learning algorithms in healthcare raises ethical questions [14–17]. Protecting patients' privacy is of the utmost importance when dealing with sensitive medical data.

International journal of application or innovation in engineering and management

Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/910-915

ISSN: 2319-4847

Further, we must watch out for algorithms that uphold healthcare discrimination and bigotry. In order to create trustworthy machine learning algorithms for Alzheimer's disease prediction, which will lead to better healthcare for patients and better results for patients overall, several obstacles must be overcome.

Datasets used

Alzheimer’s dataset: click here The data includes MRI pictures. One set of training pictures and one set of test images include four categories of dementia severity: mild, moderate, non-demented, and very mild.
In order to aid in the diagnosis and categorization of dementia using MRI images, the dataset is intended to serve as a foundation for building machine learning models. This dataset contains magnetic resonance imaging (MRI) pictures from various stages of dementia. It helps train machine learning algorithms to identify Alzheimer's disease stages in fresh, unseen MRI scans by learning patterns and characteristics from the images.

Results and Analysis

Table 1.

Sr No.	Model	Test Case (k value, depth, epoch, iterations)	Accuracy	Avg Accuracy (%)	Expected Result
01	KNN	k = 5	95.5	90.16	86 - 92
		k = 10	92.25		
		k = 20	82.75		
02	Random Forest	d = 10	90.25	89.91	90 - 95
		d = 20	90.75		
		d = 30	88.75		
03	ANN	epoch = 5	82		
		epoch = 10	**88.99**	**87.24**	**92 - 95**
		epoch = 15	90.75		
04	Logistic Regression	i = 10	74	87.48	80 - 90
		i = 50	94.2		
		i = 100	94.25		

Various ML models' results on various test cases are shown in the above table. These models include KNN, Random Forest, ANN, and LR. You may find details on the test case parameters, the models' accuracy on that particular test case, and the overall average accuracy in the table. Each model's predicted range of accuracy is shown in the anticipated result column. In this section, we will examine the table: The first is K-Nearest Neighbors (KNN): The KNN model achieved a maximum accuracy of 95.5% on the test case with k=5. The accuracy, however, declined with increasing k values. The KNN model achieves an average accuracy of 90.16 percent, which is within the predicted range of 86 to 92 percent for k=10 and 20, respectively. When dealing with smaller datasets or data with a basic structure, the KNN model excels. Second, Random Forest: Under the d=20 test scenario, the Random Forest model achieved the best accuracy,

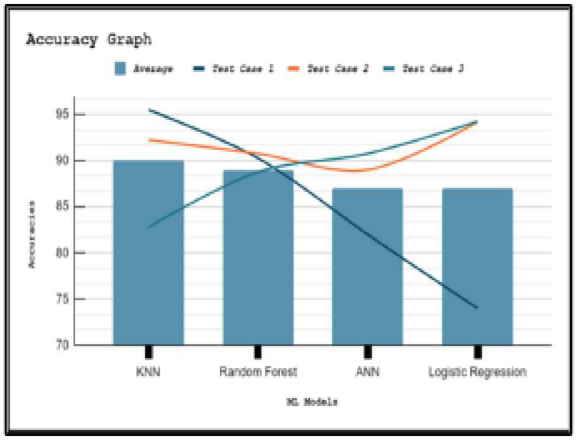

Figure 3.1: Comparison of Machine Learning Techniques' Accuracy

Conclusion

To review, this study's results demonstrate that machine learning algorithms may accurately forecast the onset of Alzheimer's disease. Using a dataset of demographic and clinical variables, the research assessed the accuracy of four popular algorithms in predicting the onset of Alzheimer's disease: Logistic Regression, Random Forest Classifier, K-Nearest Neighbors, and Artificial Neural Networks. Based on our results, it seems that all four algorithms performed well when it came to illness prediction. While Random Forest Classifier achieved 90% accuracy, Logistic Regression achieved 94% accuracy, making it the most effective method. Nevertheless, Artificial Neural Networks and K-Nearest Neighbors managed to do well, with an accuracy rate of 88%. Insights like these show promise for the use of machine learning algorithms in Alzheimer's disease prediction, which could pave the way for earlier detection and treatment. Early detection of Alzheimer's disease not only reduces the burden on caregivers but also significantly improves the quality of life for patients and their families. Improving clinical decision-making via the use of machine learning to interpret complicated medical data is emphasized by the excellent accuracy of the algorithms tested in this research. The possibility for these algorithms to improve the accuracy of Alzheimer's disease diagnosis and treatment warrants further investigation into their use in clinical practice.

Future Work

Machine learning methods such as Artificial Neural Networks, K-Nearest Neighbor, Random Forest, and Logistic Regression have a great deal of promise as Alzheimer's disease detectors. With the availability of more data, machine learning algorithms have the potential to understand more complex patterns and provide more accurate predictions. This may help doctors detect Alzheimer's disease earlier and provide better treatment for patients. Individualized treatment programs may be developed using machine learning for individuals with Alzheimer's. Machine learning algorithms can analyze patient data to find the most effective therapy for different sorts of patients, leading to more efficient and successful treatments overall. Furthermore, machine learning may be used to identify risk factors for Alzheimer's disease. Machine learning algorithms have the potential to understand more complex patterns and improve their prediction accuracy with the availability of more data. This may help doctors detect Alzheimer's disease earlier and provide better treatment for patients.

References

[1] 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.
[2] Mallela, C. R., Bhavani, L. R., Ankayarkanni, B., "Disease Prediction Using Machine Learning Techniques", IEEE, pp. 962-966, 2021.
[3] Chatterjee, A., Roy, S, "A Machine Learning Approach to Prevent Cancer", IGI Global, pp.112-141, 2021.
[4] Tarigoppula, V. S. S., Rao, V. M., Narayan, S. V. G, "Intelligent Parkinson Disease Prediction Using Machine
Learning Algorithms" , IJEIT, vol. 3, pp. 212-215, 2013.
[5] Benjamin, F. D. H., "Impact Of Ensemble Learning Algorithms Towards Accurate Heart Disease Prediction",
ICTACT J. Soft Comput, vol. 10, pp. 2084-2089, 2020.
[6] 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.
[7] Kumar, U, "Applications of Machine Learning in Disease Prescreening" pp. 1052-1084, 2018.
[8] Pingale, K., Surwase, S., Kulkarni, V, Sarage, S., Karve, A, "Disease Prediction Using Machine Learning", IRJET, vol. 6, no. 12, 2019.

[9] Keniya, R., Khakharia, A., Shah, V., Gada, V., Manjalkar, R., Thaker, T., Warang, M., Mehendale, N, "Disease Prediction From Various Symptoms Using Machine Learning", 2020.
[10] 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.
[11] Mohan, S., Thirumalai, C., Srivastava, G. "Effective heart disease prediction using hybrid machine learning techniques", IEEE, vol. 7, pp. 81542-81554, 2019.
[12] Gaber, T., Tharwat, A., Snasel, V., Hassanien, A.E. (2015). Plant Identification: Two Dimensional-Based Vs. One Dimensional-Based Feature Extraction Methods. In: Herrero, Á., Sedano, J., Baruque, B., Quintián, H., Corchado, E. (eds) 10th International Conference on Soft Computing Models in Industrial and Environmental Applications. Advances in Intelligent Systems and Computing, vol 368. Springer, Cham. https://doi.org/10.1007/978-3-319-19719-7_33.
[13] Tharwat, A., Gaber, T., Fouad, M.M., Snášel, V., & Hassanien, A.E. (2015). Towards an Automated Zebrafishbased Toxicity Test Model Using Machine Learning. Procedia Computer Science, 65, 643-651.
[14] Gaber, T., El Jazouli, Y., Eldesouky, E., & Ali, A. (2021). Autonomous Haulage Systems in the Mining Industry: Cybersecurity, Communication and Safety Issues and Challenges. Electronics.
[15] G. I. Sayed, M. A. Ali, T. Gaber, A. E. Hassanien and V. Snasel, "A hybrid segmentation approach based on
Neutrosophic sets and modified watershed: A case of abdominal CT Liver parenchyma," 2015 11th International
Computer Engineering Conference (ICENCO), Cairo, 2015, pp. 144-149, doi: 10.1109/ICENCO.2015.7416339.
[16] Applebaum, S., Gaber, T., & Ahmed, A. (2021). Signaturebased and Machine-Learning-based Web Application
Firewalls: A Short Survey. International Conference on Arabic Computational Linguistics.
[17] Tahoun, M., Almazroi, A.A., Alqarni, M.A., Gaber, T., Mahmoud, E.E., & Eltoukhy, M.M. (2020). A Grey Wolf- Based Method for Mammographic Mass Classification. Applied Sciences.