www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

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

[1]Mr. K. Lakshmi Narayana, [2]Vadlamudi Srinivasulu,

[1]Assistant Professor, Department of Master of Computer Applications, Rajamahendri Institute of Engineering & Technology.
Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.
[2]Student, Department of Master of Computer Applications, Rajamahendri Institute of Engineering & Technology.
Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.

Abstract

Most people are familiar with ML, or Machine Learning, the most prominent use of AI because of the way it is changing the face of scientific inquiry. This research looks at the use of machine learning to diagnose Alzheimer's disease. Worldwide, Alzheimer's disease takes the lives of countless individuals. It is possible to use machine learning algorithms to diagnose Alzheimer's by considering symptoms such as chest pain, high cholesterol, age, etc. This paper utilizes four supervised ML algorithms: K-Nearest Neighbour (90%) with an 89% prediction accuracy rate, Random Forest (89%) with an 87% prediction accuracy rate, Artificial Neural Networks (87% accuracy rate), and Logistic Regression (87% accuracy rate). Logistic

Keywords — Machine Learning, KNN, ANN, Random Forest, Regression, Alzheimer's, Disease Prediction, Machine Learning in Healthcare

INTRODUCTION

Using historical data to generate new predictions is the focus of machine learning research. Machine learning is a subfield of artificial intelligence that aims to imitate human learning by gradually improving a system's accuracy via the use of data and algorithms. In machine learning, there are two steps: Priorities should be training and testing. For decades, machine learning software has struggled to diagnose a patient's illness based on their history and symptoms alone. Machine learning technology is a game-changer when it comes to healthcare. We are using the whole spectrum of machine learning to monitor patients' well-being. We can create models that quickly clean, analyze, and provide results from data with the use of machine learning models. When doctors use this strategy, they'll be able to make better diagnoses, which means patients will get better treatment and healthcare providers will provide higher-quality services overall. Within the framework of healthcare, ML may be most effectively used to the medical field. This study will focus on structured data in an effort to enhance the precision of massive datasets. Decision trees, KNN, and linear algorithms will be used in the present model for illness prediction. Diseases unique to a particular demographic and region are predicted by the existing system. Only some illnesses may be predicted using this method. For illness risk forecasting, this system employs Big Data and CNN algorithms. Machine learning techniques K-NN, Decision Tree, and Naive Bayesian are used by the system while dealing with S type data. The present system's accuracy percentage is 94.8%. Classifier based on random forests: Random Forest is a part of the supervised learning method. Machine learning tasks including classification and regression may make advantage of it. It shows a model that averages the results of many decision trees applied to different parts of the input dataset in order to increase the expected accuracy of the dataset. Instead of depending on a single decision tree, the random forest approach uses all of them to generate predictions, and it bases its predictions on the results of the majority of those predictions.

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

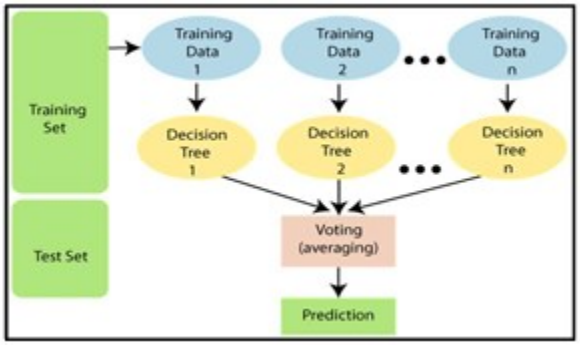


Figure 1.1: Flowchart - Random Forest Classifier.

K-Nearest Neighbor, or the K-NN Algorithm, is yet another ML method with supervised learning at its core. Assigning new instances to groups that are most similar to existing ones is how the K-NN approach works, predicated on the idea that the new cases/data are comparable to earlier cases.

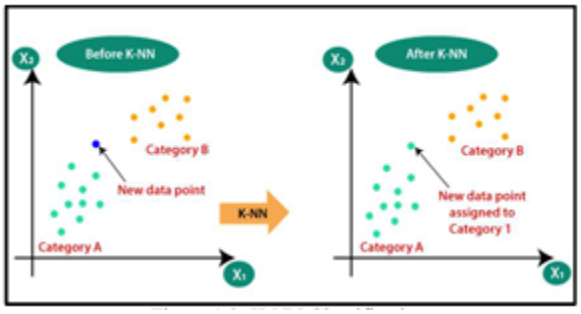


Figure 1.2: K-NN Classification. Artificial Neural Networks :

"Artificial neural network" is borrowed from the biological "neural networks" that comprise the brain's architecture. 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. Artificial neural networks use these dendrites as input;

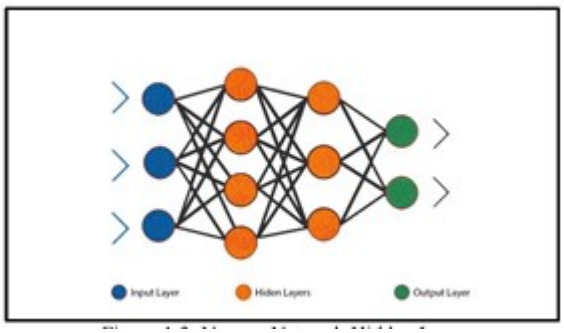


Figure 1.3: Neuron Network Hidden Layers.

Logistic Regression :

One well-known supervised learning method is logistic regression, which is used in machine learning. Using a set of predetermined independent variables, it predicts the future value of a categorical dependent variable. To forecast the value of a dependent variable that is categorical, logistic regression is used. This means that a numerical or category value must be the end result. It might be true or false, but instead of a precise number between 0 and 1, it returns a probability value in that range.

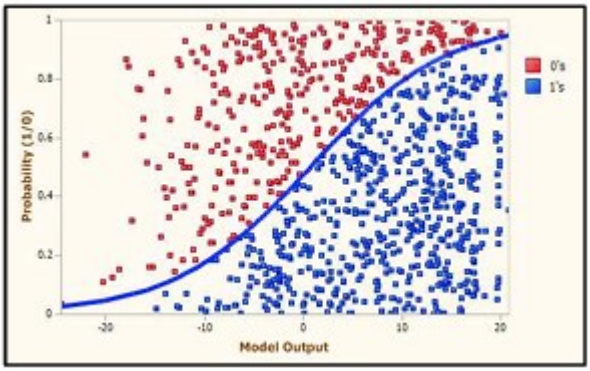


Figure 1.4: Logistic regression.

Literature Review

The use of ML approaches has been the primary driver of the recent interest and innovation in illness prediction and prevention [1]. The groundwork for this area of study is laid by Gomathy's (2021) seminal work, which highlights the importance of machine

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

learning in illness prediction. Following this, Mallela's contributions

An article by Bhavani and Ankayarkanni (2021) published by IEEE delves further into the specifics of ML methods used for illness prediction [2]. Insights into proactive cancer prevention may be found in Chapter 7 of Chatterjee and Roy's (2020) study on cancer prevention by machine learning approaches [3]. At the same time, in order to demonstrate the promise of machine learning algorithms, Tarigoppula, Rao, and Narayan (2013) investigate the possibility of early Parkinson's disease prediction [4]. The use of ensemble learning techniques to improve the precision of cardiac illness prediction is explored by Benjamin (2021) [5]. Wang, Chakraborty, and Chakraborty (2020) highlight the use of ML in risk assessment by focusing on CKD risk prediction [6]. The wider range of machine learning uses in illness pre-screening is investigated by Kumar (2018) [7]. Artificial neural networks for illness prediction were introduced to medical practitioners by Rodvold, McLeod, and Murphy (2001) [8]. Additional works on the topic of symptom-based illness prediction include those of Pingale et al. (2019) [9] and Keniya et al. (2020) [10]. When it comes to predicting the occurrence of cardiac disease, Khourdifi and Bahaj (2018) recommend using ML methods such as particle swarm modeling and ant colony optimization [11], while Mohan et al. (2019) suggest using hybrid ML approaches [11]. Taken together, these citations demonstrate how machine learning is rapidly becoming an integral part of illness prediction and how adaptable it is in dealing with many health-related problems.

Methodology

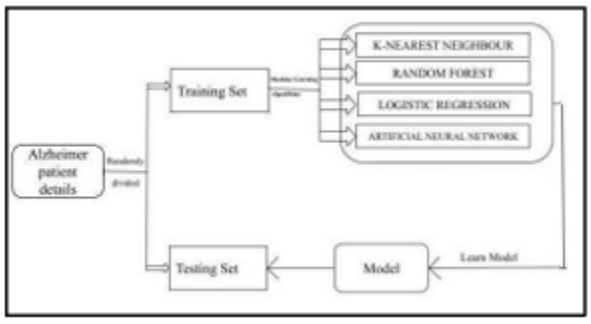


Figure 2.1: Methodology Step 1: Alzheimer's Dataset

Acquiring an Alzheimer's disease dataset was the first stage of the process. The dataset probably includes a lot of information on people, including their demographics, health history, cognitive test results, and maybe other pertinent aspects and variables. To train and test machine learning models, this dataset is used.
Third Step: Algorithms for Machine Learning This stage included using four separate machine learning approaches to detect patterns in the dataset related to Alzheimer's disease [12]. The methods used included Logistic Regression, ANN, KNN, and Random Forest. While making predictions, each algorithm takes a somewhat different approach to learning from the training data. Fourth Step: Testing and Calculating Accuracy Following model development using the training data, the separate testing data subset was used to evaluate the models. The purpose of this testing phase was to compare the models' ability to accurately predict outcomes or categories related to Alzheimer's disease. The accuracy of each model was assessed by comparing its predictions with the actual values found in the testing data. Afterwards, the accuracy percentage was determined by dividing the total number of predictions by the total number of right predictions. The research used this method to evaluate the KNN, random forest, ANN, and Logistic Regression machine learning algorithms' accuracy on the Alzheimer's dataset. To better understand how each algorithm learns patterns from the data and makes accurate predictions regarding Alzheimer's disease, this analysis is helpful. Systematic Approach For accuracy, we used many test instances. When it comes to picture datasets, random forest and ANN often outperform KNN and Logistic Regression [13]. Here, we got results for logistic regression by increasing the iterations, for ANN by gradually raising the epoch value, for KNN by increasing the number of neighbors, and for random forest by increasing the depth of the decision trees to reach random accuracies. At this stage, the data was partitioned into training and testing subsets. This section's primary objective is to prepare machine learning models to identify links and patterns in data via the use of training data. In contrast, testing data will be used to evaluate the trained models' accuracy and performance.

Machine learning algorithms can already identify Alzheimer's disease, yet many concerns about the

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

disease remain unsolved. One of the main issues is the lack of trustworthy data for training and testing algorithms. As a complicated neurodegenerative condition, Alzheimer's disease makes it difficult to gather big datasets with varied demographic and clinical characteristics. This means that the algorithms' precision is sensitive to the amount and quality of the data used for analysis. Making sure the algorithms are transparent and easy to understand is another obstacle. As machine learning algorithms get more complex, it may become increasingly challenging to understand how they produce predictions. This might be a problem when trying to explain the results to those who are caring for patients or professionals. An further factor that could affect the algorithms' accuracy is the research population. For instance, the algorithms' precision might be affected by factors like age, sex, and race. Hence, to guarantee the algorithms' efficacy in predicting Alzheimer's disease, it is crucial to verify them across varied populations. Last but not least, using ML algorithms to healthcare settings raises ethical questions [14–17]. Strict privacy and security protocols are necessary for the usage of sensitive medical data in order to safeguard patient confidentiality. Another important consideration is whether or not the algorithms contribute to healthcare-related prejudices and discrimination. In order to enhance patient outcomes and healthcare delivery, it is essential to address these problems and build accurate and reliable machine learning algorithms for predicting Alzheimer's disease.

Files used Retrieved from: Alzheimers_dataset_ The data includes MRI pictures. The data set contains four picture classifications: mildly demented, moderately demented, non-demented, and very mildly demented. Two sets of photos are used for training and testing purposes. 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. Machine learning algorithms may be trained to correctly identify and forecast the stage of Alzheimer's disease in fresh, unseen MRI scans by using this dataset, which contains MRI pictures from various stages of dementia.

Results and Analysis

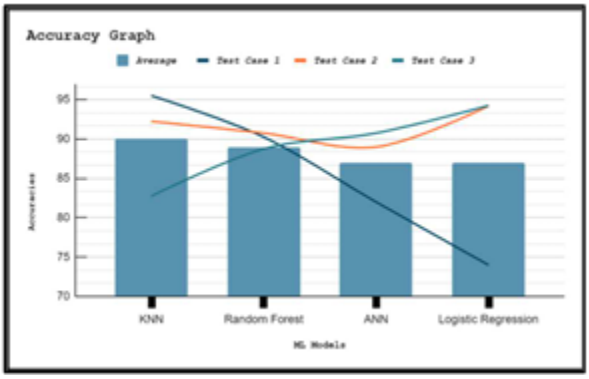
Table 1.

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. The settings used to run the test, the accuracy each model got on that test, and the overall accuracy are all detailed in the table. You can see an approximation of the accuracy range for each model in the predicted result column. Here is the breakdown of the table: First, there is K-Nearest Neighbors (KNN): A maximum accuracy of 95.5% was achieved by the KNN model on the test case with k=5. Having said that, the precision decreased with increasing k. For k=10 and 20, the predicted range of accuracy is 86-22%, hence the average accuracy attained by the KNN model is 90.16%. On smaller datasets with basic structures, the KNN model tends to perform better. Second, Random Forest: With a d-value of 20, the Random Forest model had the best accuracy of all the models tested (90.75%). As d became larger, the precision diminished. At d=10 and d=30, the Random Forest model is projected to reach an accuracy of 90–95%, which is 89.91% on average. Large datasets with complicated structures often provide better results when using the Random Forest model. 3. ANN: Achieving a high accuracy of 90.75% on the test case with epoch=15 is what the ANN model is all about. As the number of epochs increased, so did

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

the precision. With an estimated range of 92% to 95% for epochs 5 and 15, the ANN model achieves an average accuracy of 87.24 percent. It is well-known that ANN models excel on datasets with complicated structures and a large number of features. 4. Logistic Regression: An accuracy of 94.25% was achieved by the Logistic Regression model on the test case with i=100. The accuracy improved in direct proportion to the iteration count. Logistic Regression models typically have an accuracy rate of 87.48%. This is within the expected range of 80–90% for i = 10–50. Logistic Regression is well-known to perform optimally on linearly structured data.



Conclusion

In conclusion, this study's results demonstrate that machine learning algorithms can accurately forecast the onset of Alzheimer's disease. Using a dataset of demographic and clinical variables, the research compared 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. Our results indicate that the illness prediction abilities of all four algorithms were very successful. With a 94% success rate, Logistic Regression was the most accurate method, but Random Forest Classifier only managed a 90% success rate. Still, with 88% accuracy, K-Nearest Neighbors and Artificial Neural Networks did a good job. These findings highlight the promise of machine learning algorithms as a tool for Alzheimer's disease prediction, which might pave the way for earlier detection and treatment. Early detection of Alzheimer's disease improves the quality of life for patients and their families while reducing the burden on caregivers. The study's algorithms' excellent accuracy further emphasizes the significance of using machine learning to examine complicated medical data in order to enhance clinical decision-making. These algorithms have the ability to improve the accuracy of Alzheimer's disease diagnosis and treatment; future studies should investigate how to incorporate them into clinical practice. Looking Ahead Logistic Regression, Random Forest, K-Nearest Neighbor, and Artificial Neural Network are some of the machine learning techniques that show great promise in the detection of Alzheimer's disease. With the availability of more data, machine learning algorithms have the potential to understand more intricate patterns and provide more accurate predictions. This may help doctors detect Alzheimer's disease earlier and provide better treatment for patients. to Individualized treatment programs may be developed for individuals with Alzheimer's disease by using machine learning. More efficient and effective treatments are possible thanks to algorithms based on machine learning that analyze patient data to find out which medicines work best for different sorts of patients. And machine learning may also help identify Alzheimer's disease risk factors. With the availability of more data, machine learning algorithms have the potential to understand more intricate patterns and provide more accurate predictions. 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 Pre screening" 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., Snásel, V., & Hassanien, A.E. (2015). Towards an Automated Zebrafish based 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). Signature based 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 Classification. Applied Sciences.