Original Article

A Novel Approach for Prediction of Human Disease using Symptoms by Multilayer Perceptron Algorithm to Improve the Accuracy and Compared with Gaussian Naïve Bayes Algorithm

S.Avinash Prabhu[1], V.Parthipan[2]

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, TamilNadu, India, 602105.
[2]Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, TamilNadu, India, 602105.

Abstract

Aim : The aim of this paper is to improve Accuracy in Disease prediction using symptoms by a novel multilayer perceptron classifier in comparison with the naive Bayes algorithm . **Materials and Methods :** Novel Multilayer perceptron classifier and naive bayes algorithm sample size (N=10) to predict the accuracy percentage of predicted disease. G-power is calculated for two different groups, alpha (0.05), power (80%). **Results:** Based on the measurement of data, Statistical Analysis and independent sample T-test, it shows that there is a statistically insignificant difference between the two study groups with value p=0.212 (p> 0.05). It was observed that the novel multilayer perceptron algorithm obtains an accuracy of 95%. It appears to have better accuracy than the naive Bayes algorithm (92%). **Conclusion:** The results prove that novel multilayer perceptron algorithm approaches with varied seed value have significant improvement in disease prediction using symptoms.

Keywords: Novel Multilayer Perceptron, Naive Bayes, Machine Learning, Symptoms, Disease, Decision.

DOI: 10.47750/pnr.2022.13.S04.099

INTRODUCTION

The goal of this study is to create a disease predictor system which enables users to enter their symptoms and by using various machine learning algorithms which are used to predict disease (Hamsa Gayatri and Vigneshwaran 2021). In today's world the challenges faced by people who seek health information online have increased and there is a huge demand for health treatment and recommendation systems. Advancement of machine learning has been effectively being used in the healthcare and biomedical field. To extract valuable bits of data from the predefined information in medical services networks (Rani and Tiwari 2020). In this project a user interface has been created with all labels and fields and users are expected to enter their symptoms and finally with the help of different machine learning algorithms diseases with the highest probability are chosen and displayed to the user (Kumar 2021). Database has been developed using sqlite3 it is used to store details of users like the name of the user , predicted diseases and which algorithm is used. Main advantage of using this database is to store medical history of users which helps them in future medical purposes (Harish and Gayathri 2019).
There are around 22 IEEE papers and 14 Google scholar papers which have been published over the past 5 years .(Goel et al. 2019) As the advancement of technology is rapidly increasing day by day machine learning approaches seem to have better results in healthcare and biomedical fields which analyze huge amounts of data and predict patterns. (Dahiwade, Patle, and Meshram 2019) Nowadays due to various climatic changes people are getting affected by various types of disease there which results in a huge amount of data being produced every year. Machine learning algorithms like CNN were used and got higher accuracy than the KNN algorithm. (Sneha.R, Monisha.S, and Nandini.S 2020) This paper deals with machine learning algorithms techniques to provide the most probable disease occurrence based on symptoms.And predicted disease is based on this dataset provided when data is more accurate and perfect results will be more effective. (Grampurohit and Sagarnal 2020) A disease prediction system has been developed and machine learning algorithms like decision tree, naive bayes

and random forest are used to predict diseases which have an accuracy of 95% .Our team has extensive knowledge and research experience that has translate into high quality publications(Bhansali et al. 2021; Jayanth et al. 2021; Sudhakar et al. 2021; Sathiyamoorthi et al. 2021; Deepanraj et al. 2021; Raju et al. 2021; Arun Prakash et al. 2020; Kamath et al. 2020; Shanmugam et al. 2021; Rajasekaran et al. 2020; Adhinarayanan et al. 2020; Rajesh et al. 2020; Aurtherson et al. 2021)
By our top to bottom examination, the results conclude that expectations done before didn't utilize a huge dataset . A huge dataset guarantees better expectation. Additionally what it needs is a suggestion framework (Shetty et al. 2017). Accuracy and predicted diseases are at least effective. Our proposed system has overcome these problems by providing better accuracy than the existing system by using the MLP algorithm.

Materials And Methods

This research work was performed in the OOAD Lab, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai. Basically it is considered that two groups of classifiers are used, namely novel multilayer perceptron classifier and decision tree. Group 1 is a novel multilayer perceptron algorithm with sample size of 10 and Guassian naive bayes is group 2 with sample size of 10 and they are compared for a more accurate score for choosing the best algorithm. Pre- test analysis has been prepared using clinical.com by having a G power of 80% and threshold 0.05% And a total of 20 samples in which standard deviation of MLP is .92660 and naive bayes is .67035 (Kumar 2021).

Multilayer Perceptron
MLP is a form of feed-forward neural network model that maps a set of input data to an appropriate result. It consists of several layers of nodes in the form of directed graph, where each layer is fully connected to the next one except for the input data where each node is a neuron which collides with a non-linear activation function. It uses supervised machine learning techniques. It is also called a standard linear perceptron.
Algorithm for MLP
Step 1: Load the dataset.
Step 2 : Inputs are pushed through MLP which takes a dot product between the input layer and hidden layer.
Step 3 : It utilizes activation functions which are calculated at each layer.
Step 4 : Calculated values are put together through any of the activation functions.
Step 5 : Move the other layer in the MLP and repeat Step 1.
Step 6 : Repeat steps 3 and 5 until the output layer is reached.
Step 7 : Once it reaches the output layer, the calculations use a back propagation method which corresponds to activation functions.
Step 8 : Finally decisions will be made based on the output.
Step 9 : End.

Gaussian Naive Bayes
The Naive Bayes algorithm is a set of algorithms that are dependent on the nave bayes theorem. They all follow the same principle: each pair of forecasts is independent of the other. This is one of the supervised machine learning approaches. It also assumes features which make an equivalent and independent impact to prediction. It includes a high-dimensional dataset which results in accurate results. In this project the naive Bayes algorithm predicts disease based on probability of disease in a training dataset. One of the important characteristics of naive bayes is that it can be used in real world problems and it has parallel performance with the decision tree algorithm.

Algorithm for naive bayes
Step 1: Analyze the training dataset T first.
Step 2: Calculate the mean and standard deviation of the predictor variables for each class.
Step 3: Using the gauss density equation, calculate the probability of fi in each class. Continue to
Step 4: Continue until all predictor variables (f1, f2, f3, f4, f5, f6, f7, f8, f9) have had their probabilities computed. And then do it again.
Step 5: Calculate the likelihood for each class.
Step 6: Make the most of your opportunities.
Step 7: Last but not least, the class with the highest probability must be chosen.

Data collection for this research has been taken from a study of the University of Columbia performed at New York Presbyterian Hospital during 2004. Testing setup for this research has all the components to do our test process. It has 2 types of configurations, Hardware configuration, and Software configuration. Hardware configurations include Intel core i3 5th generation processor, 8 GB RAM (Random Access Memory), 64-bit Windows OS. And software configuration includes Windows OS.

Statistical Analysis

IBM SPSS version 21 was used to conduct the analysis. It's comparable to a statistical tool for data analysis. 10 iterations with a maximum of samples obtained were used to compare the MLP and Naive Bayes algorithms, and the estimated accuracy for each iteration was documented for predicting correct accuracy. Symptoms are the dependent variables, while illness kinds are the independent variables. Finally, an independent sample T- test was used to determine the value acquired from these iterations and a graph was plotted to know the exact difference between MLP algorithm and naive bayes algorithm (Kumar 2021).

Results

Table 1 Shows statistical difference between novel multilayer perceptron algorithm and decision naive bayes algorithm. Due to this importance of uniformity of fluctuation, the likelihood esteem expresses that outcomes in this examination exertion are huge and connected with one another, this table exhibits distinction in exactness of both MLP and naive bayes. Accuracy comparison of MLP and Gaussian naive bayes algorithms is shown in Table 2. Results of this independent sample t-test are shown in Table 3. It is found that mean accuracy for MLP is 95.11 % and that of naive bayes is 92.12% with a standard deviation of .34464. Figure 1 shows graphical representation between the two algorithms Graphical representation of the bar graph is plotted using groupid as X-axis novel Multilayer perceptron and naive bayes. Y-Axis displaying the error bars with a mean accuracy of detection +/- 1 SD.

Discussion

Here is a framework which can anticipate infection based on manifestations given to it. Such a framework can diminish this surge at OPDs of emergency clinics and decrease some responsibility on clinical staff. (Islam et al. 2021). Normally accomplished an exactness of ~95%. Making this framework additionally added an approach to storing the information entered by users in this data set which can be utilized in future to help in making a superior rendition of such a framework. Our framework additionally has a simple to use interface. It additionally has a different visual portrayal of information gathered and the results accomplished using the sqlite3 database.

(Mohan and Jain 2020) It was observed that there are many possible ways to explore this research on disease prediction and future scope is to use several machine learning algorithms and artificial neural networks which can be used to improve accuracy. Using several machine learning algorithms to predict disease based on symptoms, there is demand in this current market where people worried about diseases can simply use this system and predict disease which helps them rushing to hospitals and waiting for doctor arrival. (S et al., 2020) As the emergence of big data in the healthcare and biomedical field has been rapidly increasing. An accurate study of data is required in order to give accurate results. As there are different symptoms for different diseases, prediction may result in inaccuracy. In order to provide efficient results machine learning algorithms are used and have an accuracy of 94% which is better than the existing algorithms (Kumar 2021). These results are used to predict the probability of diseases. This research deals with prediction of heart disease through machine learning modes which provide accurate results (Gupta, Sachdeva, and Dohare 2021). Comparative analysis has been done by using various machine learning, deep learning and artificial neural networks models where most of the prediction models are based on machine learning where accuracy of 98% is obtained in this model and data mining gives accuracy of 92.1% and fuzzy gives 93.27%. (Goel et al. 2019) Heart disease patients are increasing day by day there is a need for predicting such attacks. This ensemble model consists of SVM, ANN and decision tree to predict attack and can be used to get higher accuracy than the algorithms individually used (Wenxin 2020).

The limitations for this project is that the existing system currently deals with less data related to both symptoms and diseases which may result in partial disease prediction. The future scope of this project is to collect more information related to symptoms and their probable diseases which results in accurate prediction of diseases.

Conclusion

Prediction of disease using symptoms has been successfully developed. Current study focused on different machine learning algorithms like Multilayer perceptron algorithm and naive bayes algorithm where the outcome of this study proved that Multilayer perceptron algorithm has higher accuracy of 95% than naive bayes algorithm of 92%.

DECLARATIONS

Conflict of Interests

No conflict of interest

Authors Contribution

Author NNC was involved in data collection, data analysis, and manuscript writing. Author VP was involved in the Action process, Data verification and validation, and Critical review of the manuscript.

Acknowledgments

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this research work successfully.

Funding: We thank the following organization for providing financial support that enabled us to complete the study.

1. Softeon Pvt.Ltd,Chennai.
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

Reference

1. Dahiwade, Dhiraj, Gajanan Patle, and Ektaa Meshram. 2019. "Designing Disease Prediction Model Using Machine Learning Approach." *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*. https://doi.org/10.1109/iccmc.2019.8819782.
2. "DISEASE PREDICTION BASED ON SYMPTOMS USING CLASSIFICATION ALGORITHM." 2020. *JOURNAL OF XI'AN UNIVERSITY OF ARCHITECTURE & TECHNOLOGY*. https://doi.org/10.37896/jxat12.04/1055.
3. Goel, Sakshi, Abhinav Deep, Shilpa Srivastava, and Aprna Tripathi. 2019. "Comparative Analysis of Various Techniques for Heart Disease Prediction." *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*. https://doi.org/10.1109/iscon47742.2019.9036290.
4. Grampurohit, Sneha, and Chetan Sagarnal. 2020. "Disease Prediction Using Machine Learning Algorithms." *2020 International Conference for Emerging Technology (INCET)*. https://doi.org/10.1109/incet49848.2020.9154130.
5. Gupta, Vagisha, Shelly Sachdeva, and Neha Dohare. 2021. "Deep Similarity Learning for Disease Prediction." *Trends in Deep Learning Methodologies*. https://doi.org/10.1016/b978-0-12-822226-3.00008-8.
6. Hamsa Gayatri, P., and S. Vigneshwaran. 2021. "Symptoms Based Disease Prediction Using Machine Learning Techniques." *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICI)*. https://doi.org/10.1109/icicv50876.2021.9388603.
7. Harish, S., and K. S. Gayathri. 2019. "Smart Home Based Prediction of Symptoms of Alzheimer's Disease Using Machine Learning and Contextual Approach." *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*. https://doi.org/10.1109/iccids.2019.8862163.
8. Islam, Sheikh Raul, Rohit Sinha, Santi P. Maity, and Ajoy Kumar Ray. 2021. "Deep Learning on Symptoms in Disease Prediction." *Machine Learning for Healthcare Applications*. https://doi.org/10.1002/9781119792611.ch5.
9. Kumar, Anand. 2021. "Disease Prediction and Doctor Recommendation System Using Machine Learning Approaches." *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2021.36234.
10. Mohan, Narendra, and Vinod Jain. 2020. "Performance Analysis of Support Vector Machine in Diabetes Prediction." *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. https://doi.org/10.1109/iceca49313.2020.9297411.
11. Rani, Geeta, and Pradeep Kumar Tiwari. 2020. *Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning*. IGI Global.
12. Shetty, Deeraj, Kishor Rit, Sohail Shaikh, and Nikita Patil. 2017. "Diabetes Disease Prediction Using Data Mining." *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*. https://doi.org/10.1109/iciiecs.2017.8276012.
13. S. Vinitha, S. Vinitha, S. Sweetlin, H. Vinusha, and S. Sajini. n.d. "Disease Prediction Using Machine Learning Over Big Data." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3458775.
14. Wenxin, Xu. 2020. "Heart Disease Prediction Model Based on Model Ensemble." *2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD)*. https://doi.org/10.1109/icaibd49809.2020.9137483.

TABLES AND FIGURES

Table 1. Comparing accuracy of the values between the multilayer perceptron algorithm and the decision tree algorithm with sample size n=10. Accuracy of MLP (94.61) and naive bayes (91.31).

Sample(N)	Multilayer Perceptron	Naive Bayes
1	94	91
2	95	90
3	95	92
4	93	90
5	95	92
6	94	91
7	94	90
8	93	91
9	95	90
10	93	91

Table 2. Showing group statistics, the Multilayer perceptron algorithm with naive Bayes algorithm by grouping iterations with sample size 10, got a mean 94.6190, standard deviation .92660 and standard error mean .29302. Descriptive independent sample test of accuracy and precision is applied for this dataset which is in SPSS.And here it specifies Equal variances with and without assuming a T-Test Score of two groups with each sample size of 10.

Group	N	Mean	Std. Deviation	Std. Error Mean
MLP	10	94.6190	.92660	.29302
RF	10	91.3120	.67035	.21198

Table 3. Independent Samples T-test MLP shows significance value achieved is p=0.212 (p>0.05), which shows that two groups are statistically insignificant and mean difference is 3.30700 with standard error difference of .36166.

Dependent variables	Assumptions	F	sig.	t	df	sig.(2-tailed)	Mean Difference	Std.Error Difference	Lower	Upper

Accuracy	Equal variance assumed	1.677	.212	9.144	18	.000	3.30700	.36166	2.54719	4.06681
	Equal variances not assumed			9.144	16.395	.000	3.30700	.36166	2.54719	4.07218

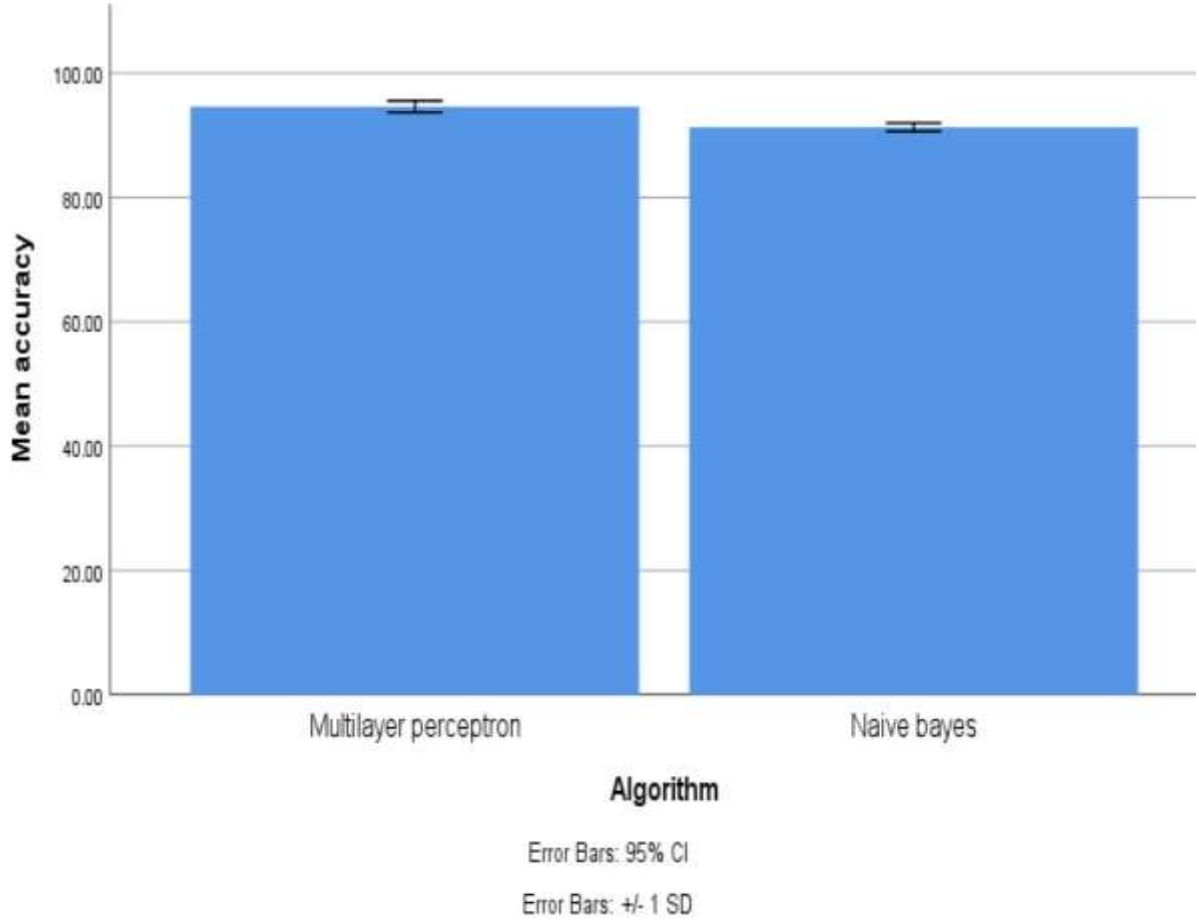

Fig. 1. Comparison of mean accuracy of multilayer perceptron algorithm and naive bayes it has been shown that there is slight difference between two algorithms and the multilayer perceptron has more accuracy than the naive bayes.Graphical representation of bar is plotted were group id represents X -axis labels and the mean accuracy in Y-axis displaying the error bars with a mean accuracy of detection +/- 1 SD.