ORIGINAL RESEARCH

Submitted: 3.10.2022; Accepted: 1.11.2022; Published online: 25.12.2022

Heart Plaque Detection with Improved Accuracy using Naive Bayes and comparing with Least Squares Support Vector Machine

Vankamaddi Sunil Kumar, K Vidhya

Department of Electronic and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai,Tamilnadu. India. Pincode:602105.

ABSTRACT

Aim: The main aim of this research is to detect heart plaque using the Naive Bayes algorithm with improved accuracy and comparing it with Least Squares Support Vector Machine. **Materials and Methods:** Naive Bayes algorithm and Least squares Support Vector Machine algorithms are two groups compared in this study. In the Kaggle dataset on Heart Plaque Disease, there were a total of 20 samples. Clincalc is used to calculate sample G power of 0.08 with 95% confidence interval. The training dataset (n = 489 (70 %)) and the test dataset (n = 277 (30 %)) are divided into two groups. **Result:** The accuracy of the Naive Bayes algorithm and the Least Squares Support Vector Machine algorithm is assessed. The Naive Bayes method was 78% accurate, whereas the Least Squares Support Vector Machine method was only 67.3% correct.**Conclusion:** In this work, the Naive Bayes algorithm outperformed the Least Squares Support Vector Machine algorithm in detecting heart plaque disease in the dataset under consideration.

Keywords

Heart Plaque disease, Novel intensity feature, Naive Bayes algorithm, Least Squares Support Vector Machine, Prediction, Machine learning.

Imprint

Vankamaddi Sunil Kumar, K Vidhya. Heart Plaque Detection with Improved Accuracy using Naive Bayes and comparing with Least Squares Support Vector Machine. Cardiometry; Special issue No. 25; December 2022; p. 1595-1599; DOI: 10.18137/cardiometry.2022.25.15951599; Available from: http://www.cardiometry.net/issues/no25-december-2022/naive-bayes-comparing

INTRODUCTION

The approaches for detecting, diagnosing, and self-managing Heart Plaque Disease are examined in this study. The detection and identification of cardiac plaques, as well as the detection and self-management options for Coronary Disease, were all comprehensively explored [1]. The goal of the study is to develop a machine learning-based prediction system and identify the best classifier for achieving the best results when compared to clinical outcomes. The proposed strategy, which is based on predictive analysis, aims to discover traits that can help in the early detection of heart plaque formation [2]. These techniques yielded a wide variety of accuracy results. As a result, scientists have been experimenting with new classifiers or combining many classifiers to increase the quality of their models. The goal of this research is to see which algorithm, depending on the patient's diagnostic data, gives the best outcomes in terms of detecting a current disease and forecasting the chance of developing one in the future.

Using machine learning approaches, a lot of research has recently been done on a number of Heart Plaque disease diagnoses. Over a five-year period, 7 research articles were published in scientific journals for the diagnosis of heart plaque disease, while 547 publications were located in Google scholar. In a recent study, they projected an accuracy of 72 percent using the Wavelet Transform approach [3] in a paper called Analysis of Heart Plaque for Early Prediction Using Wavelet Transform algorithm. The author [4] used the best attributes from the Heart Plaque disease patient to detect the disease with an accuracy of 73%. For the detection of Heart Plaque disease, the researchers [5] and [6] employed several algorithms such as CWT and huygens and achieved a 80 percent accuracy. Principal component analysis (PCA), minimum redundancy and maximum relevance (mRMR), and five cross validation were proposed by the author [7] for analyzing the models for deducting dimensionalities accuracy was reached by 67 percent. Researchers [8] used wavelet transform algorithms to diagnose Heart Plaque disease with a 75.7 percent accuracy.Our team has extensive knowledge and research experience that has translate into high quality publications[9]–[20]

Since the existing technique is sensitive to outliers, it leads to incorrect results. The goal is to increase the accuracy of Heart Plaque disease detection by using a Naive Bayes algorithm approach rather than a Least Squares Support Vector Machine.

MATERIALS AND METHODS

This study was conducted in the Image Processing Lab, Department of Electronics and Communication Engineering at Saveetha School of Engineering, SIMATS, Chennai. The number of groups taken to collect the samples for statistical analysis is 2.The Sample preparation of group 1 in Naive Bayes is one of the well-known methods that are used to predict the tumor cell from pneumonia images. The proposed technique exhibits improved accuracy outcomes, according to the simulation findings. The specified sample analysis is completed using the G power statistical tool with a probability of 80 %. A display with a resolution of 1920x1080 pixels (2nd gen, Ryzen 5 series, 8GB RAM, 512 GB SSD) and a Matlab program with suitable library and tool capabilities are required to train these datasets. The output is obtained using MATLAB software [21].

In group 1, sample preparation is completed by downloading a kaggle dataset. Import the data into Google Colab. Calculate the precision using various iterations. For each group, 20 samples are taken into account to calculate the accuracy score. The Sample preparation of group 1 is for classification and regression analysis, the Naive Bayes and supervised learning technique are utilized. It's a non-parametric test. It worked well with [7] both category and continuous output variables. The Novel intensity feature based Naive Bayes is a two-stage classification process with a learning phase and a prediction step [4]. In the learning stage, the model is trained using the provided training data, and in the prediction stage, it is used to anticipate the response for the given testing data.

In group 2, sample preparation is completed by downloading a kaggle dataset. Import the data into Google Colab. Calculate the precision using various iterations. For each group, 20 samples are taken into account to calculate the accuracy score. The Sample preparation of group 2 is Least-squares versions of support-vector machines, which are a set of related supervised learning methods that examine data and recognise patterns and are used for classification and regression analysis, and are used in statistics and statistical modeling. Instead of addressing a convex quadratic programming (QP) problem, this version finds the solution by solving a set of linear equations. Suykens and Vandewalle proposed least-squares SVM classifiers. Kernel-based learning methods such as LS-SVMs are a subset of kernel-based learning methods.

The Heart Plaque disease data collection was obtained via kaggle. The dataset is now ready for training and testing after being processed. Missing data should be deleted, null values should be replaced with mean or median values, and data should be standardised throughout data processing. The K-Nearest Neighbor and Least Squares Support Vector Machine methods are used to process a preprocessed dataset using pictures as input. To achieve the most accurate result and detection, 70% of the data from the full sample size is used for training, while the remaining 30% is used for testing.

For this Heart Plaque disease data collection, a total of 659 patient records were gathered using kaggle. There are 427 healthy persons samples and 232 patients with cardiac plaque disease samples in this study. The database is divided into 12 columns and 659 rows. They include details on 659 people, such as pregnancies, family history, blood pressure, cholesterol, age, smoking, ECG, and blood sugar levels, as well as their outcomes.

Statistical Analysis

IBM SPSS version 21 was used for the analysis. It's a type of statistical software that's used to analyze data. For both proposed and current algorithms, 10 iterations with a maximum of 20 samples were performed, with the expected accuracy documented for each iteration. Independent sample t-tests' significant values are determined. Pregnancy, family history, blood pressure, cholesterol, age, smoking, ECG, Blood Sugar, and outcome are all independent variables, whereas accuracy is the dependent variable. These values have been subjected to a thorough examination in order to predict cardiac disease [22].

RESULTS

Figure 1 compares Naive Bayes algorithm and the Least Squares Support Vector Machine in terms of accuracy. The two methods are compared using an independent t-test, and the mean accuracy value shows a statistically significant difference. The Novel intensity feature based Naive Bayes algorithm technique outperforms the Least Squares Support Vector Machine algorithm by 78 %.

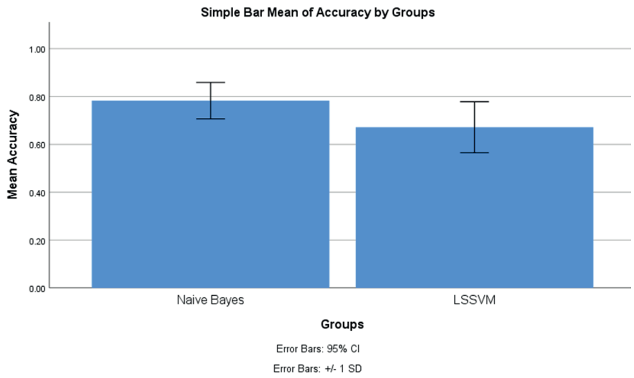

Fig. 1. Comparison of Naive Bayes algorithm algorithm and Least Squares Support Vector Machine algorithm in terms of mean accuracy. The mean accuracy of the Naive Bayes algorithm is better than the Least Squares Support Vector Machine algorithm and the standard deviation of the Naive Bayes algorithm is slightly better than the Least Squares Support Vector Machine algorithm. X-axis: (GROUPS) Naive Bayes algorithm Vs Least Squares Support Vector Machine Classifier and Y-axis: Mean accuracy of Prediction ±1 SD.

For the comparison of two algorithms, an independent t-test was utilized, and a statistically significant difference ($p<0.05$) was found. The Novel intensity feature based Naive Bayes algorithm has a 78 percent accuracy. In terms of accuracy, the Naive Bayes method surpasses the Least Squares Support Vector Machine technique, and the standard deviation of FWT is somewhat better than the Least Squares Support Vector Machine algorithm.

Table 1 summarizes the results of the Naive Bayes algorithm and the Least Squares Support Vector Machine for the Heart Plaque disease data set. The Naive Bayes algorithm has a detection accuracy of 78 %, whereas the Least Squares Support Vector Machine approach has a detection accuracy of 67 %. Table 2 shows the accuracy statistics for the Naive Bayes algorithm and the Least Squares Support Vector Machine methods. The average for the Least Squares Support Vector Machine algorithm is 67 %, whereas the average for the Naive Bayes algorithm is 78 %. The standard deviation of the Novel intensity feature based Naive Bayes algorithm is 0.194, while the standard deviation of the Least Squares Support Vector Machine algorithm is 0.152. The standard error mean for the Naive Bayes algorithm is 0.043, while the standard error mean for the Least Squares Support Vector Machine algorithm is 0.034. When compared to the Least Squares Support Vector Machine algorithm, the Naive Bayes algorithm has a higher mean accuracy. The Novel intensity feature based Naive Bayes algorithm outperformed the Least Squares Support Vector Machine algorithm, according to the data. Independent t-tests with a significance threshold of 0.05 are shown in Table 3.

Table 1

Samples,features and classes from dataset. In the given data set 659 samples are taken. The data set contains 2 classes (with Heart Plaque disease and without Heart Plaque disease).

Data set	No of patients	Features	Classes
Heart Plaque	659	20	2

Table 2

Comparison of accuracy between Naive Bayes algorithm and Least Squares Support Vector Machine algorithm.

Dataset	Naive Bayes algorithm		Least Squares Support Vector Machine algorithm	
Heart Plaque	Accuracy	78 %	Accuracy	67.3 %

Table 3

Statistical analysis of Naive Bayes algorithm and Least Squares Support Vector Machine algorithm. Mean accuracy value, Standard deviation and Standard Error Mean for Naive Bayes algorithm and Least Squares Support Vector Machine algorithm are obtained for 10 iterations.

	Group	N	Mean	Std Deviation	Std Error Mean
Accuracy	Naive Bayes algorithm	20	0.7843	0.1941	0.043
	Least Squares Support Vector Machine	20	0.6732	0.1529	0.3419

Table 4 shows The values of accuracy are classified into two categories using Levene's test: when equality of variance is assumed and when equality of variance is not assumed. Because the significance value is less than 0.05, the hypothesis holds true.

DISCUSSION

In this study, the Novel intensity feature based Naive Bayes algorithm appears to be more accurate in identifying Heart Plaque disease than the Least Squares Support Vector Machine approach ($p=0.01$, Independent sample Test). The Naive Bayes algorithm has a greater accuracy (mean accuracy= 78) than the Least Squares Support Vector Machine technique (mean accuracy= 67). The collection includes several qualities that indicate the disease condition, as well as normal and abnormal human circumstances. The Naive Bayes Transform algorithm outperforms the Least Squares Support Vector Machine.

Table 4

Independent Sample t-Test for significance and standard error determination. P value is less than 0.05 considered to be statistically significant and 95% confidence intervals were calculated.

		Levene's test for Equality of Variances		T-test for Equality of Means						
		F	sig.	t	df	Sig. (2-tailed	Mean diff	Std. Error diff	Lower	Upper
Accuracy	Equal variances assumed	3.027	0.02	1.517	38	0.000	0.838	0.552	-0.02	0.195
	Not equal variances assumed			1.517	36.02	0.000	0.838	0.552	-0.02	0.195

Boosting was discovered by researchers [23] to increase the accuracy of the Naive Bayes method and the Least Squares Support Vector Machine algorithm. The accuracy of these two methods, as well as the accuracy of other algorithms, can be improved. The Naive Bayes algorithm and the Least Squares Support Vector Machine algorithm, both of which included and did not include bagging, yielded identical results with a 67 % accuracy rate. It is apparent that algorithms perform better when updated bagging methods are used. To compare classification algorithms, the researcher [24] used a variety of performance indicators, including accuracy, sensitivity, recall, and specificity. The researchers discovered that boosting can improve the accuracy of the Novel intensity feature based Naive Bayes algorithm by 75%. [25]. They took the best sample and compared the findings, which clearly show that the Naive Bayes algorithm technique is more efficient and performs better than the Least Squares Support Vector Machine approach by 69 percent. The author [26] evaluated the classifying algorithms with the DenseNET and multiple machine learning methods, finding that multiple machine learning algorithms outperformed the algorithms by 87 %.

The resulting model is difficult to comprehend and analyze, making integration of our business logic unfeasible. The model is challenging to comprehend and analyze because of the various weights and individual effects that we can't accomplish in fewer calibrations to the model. In the future, more data can be collected to improve this.

The biggest disadvantage is the long training time necessary for large datasets. The accuracy of the Naive Bayes methodology can be enhanced in the future by including additional data in the training sets, employing a multiclass Naive Bayes approach.

CONCLUSION

Naive Bayes algorithm and the Naive Bayes algorithm were utilized in this study to detect Heart Plaque disease. In terms of accuracy, the Naive Bayes methodology (78 % is significantly better than Least Squares Support Vector Machine technique (67.3 %).

REFERENCES

1. S. S. A and S. A. Sushma, "Comparative Study of Naive Bayes, Gaussian Naive Bayes Classifier and Decision Tree Algorithms for Prediction of Heart Diseases," International Journal for Research in Applied Science and Engineering Technology, vol. 9, no. 3. pp. 475–486, 2021. doi: 10.22214/ijraset.2021.33228.
2. J. Bohacik and M. Zabovsky, "Naive Bayes for statlog heart database with consideration of data specifics," 2017 IEEE 14th International Scientific Conference on Informatics. 2017. doi: 10.1109/informatics.2017.8327218.
3. U. N. Dulhare, "Prediction system for heart disease using Naive Bayes and particle swarm optimization," Biomedical Research, vol. 29, no. 12. 2018. doi: 10.4066/biomedicalresearch.29-18-620.
4. P. Priyanga and N. C. Naveen, "Web Analytics Support System for Prediction of Heart Disease Using Naive Bayes Weighted Approach (NBwa)," 2017 Asia Modelling Symposium (AMS). 2017. doi: 10.1109/ams.2017.12.
5. S. Pattekari, Intelligent Heart Disease Prediction System Using Naive Bayes: Heart Attack Prediction System Using Data Mining Technique. LAP Lambert Academic Publishing, 2014.
6. M. Gasiewski, Comparing Naïve Bayes and Perceptron Machine Learning Techniques in Heart Disease Diagnostics. 2016.
7. M. D. Gondkar, Detection Of Heart Attack Type Along With Suggestion Of Treatment. LAP Lambert Academic Publishing, 2015.
8. M. W. Berry and M. Browne, Lecture Notes in Data Mining. World Scientific, 2006.
9. L. R. Chellapa, S. Rajeshkumar, M. I. Arumugham, and S. R. Samuel, "Biogenic Nanoselenium Synthesis

and Evaluation of its antimicrobial, Antioxidant Activity and Toxicity," Bioinspired Biomim. Nanobiomaterials, pp. 1–6, Jul. 2020.

10. M. Lavanya, P. M. Kannan, and M. Arivalagan, "Lung cancer diagnosis and staging using firefly algorithm fuzzy C-means segmentation and support vector machine classification of lung nodules," Int. J. Biomed. Eng. Technol., vol. 37, no. 2, p. 185, 2021.

11. K. Raj R, E. D, and R. S, "β-Sitosterol-assisted silver nanoparticles activates Nrf2 and triggers mitochondrial apoptosis via oxidative stress in human hepatocellular cancer cell line," J. Biomed. Mater. Res. A, vol. 108, no. 9, pp. 1899–1908, Sep. 2020.

12. D. P. Shilpa-Jain, J. Krithikadatta, D. Kowsky, and V. Natanasabapathy, "Effect of cervical lesion centered access cavity restored with short glass fibre reinforced resin composites on fracture resistance in human mandibular premolars- an in vitro study," J. Mech. Behav. Biomed. Mater., vol. 122, p. 104654, Oct. 2021.

13. S. S, K. R, and S. P, "Quantification of sweat urea in diabetes using electro-optical technique," Physiol. Meas., vol. 42, no. 9, Sep. 2021, doi: 10.1088/1361-6579/ac1d3a.

14. R. Ramadoss, R. Padmanaban, and B. Subramanian, "Role of bioglass in enamel remineralization: Existing strategies and future prospects-A narrative review," J. Biomed. Mater. Res. B Appl. Biomater., vol. 110, no. 1, pp. 45–66, Jan. 2022.

15. S. Wu, S. Rajeshkumar, M. Madasamy, and V. Mahendran, "Green synthesis of copper nanoparticles using Cissus vitiginea and its antioxidant and antibacterial activity against urinary tract infection pathogens," Artif. Cells Nanomed. Biotechnol., vol. 48, no. 1, pp. 1153–1158, Dec. 2020.

16. R. Kalidoss, S. Umapathy, and U. Rani Thirunavukkarasu, "A breathalyzer for the assessment of chronic kidney disease patients' breathprint: Breath flow dynamic simulation on the measurement chamber and experimental investigation," Biomed. Signal Process. Control, vol. 70, p. 103060, Sep. 2021.

17. R. Kaja et al., "Biofeedback flutter device for managing the symptoms of patients with COPD," Technol. Health Care, vol. 28, no. 5, pp. 477–485, 2020.

18. C. H. Antink et al., "Fast body part segmentation and tracking of neonatal video data using deep learning," Med. Biol. Eng. Comput., vol. 58, no. 12, pp. 3049–3061, Dec. 2020.

19. M. Paul et al., "Non-contact sensing of neonatal pulse rate using camera-based imaging: a clinical feasibility study," Physiol. Meas., vol. 41, no. 2, p. 024001, Mar. 2020.

20. H. Malaikolundhan et al., "Anticarcinogenic effect of gold nanoparticles synthesized from Albizia lebbeck on HCT-116 colon cancer cell lines," Artif. Cells Nanomed. Biotechnol., vol. 48, no. 1, pp. 1206–1213, Dec. 2020.

21. A. Subasi, Practical Guide for Biomedical Signals Analysis Using Machine Learning Techniques: A MATLAB Based Approach. Academic Press, 2019.

22. B. P. O'connor, "SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test," Behavior Research Methods, Instruments, & Computers, vol. 32, no. 3. pp. 396–402, 2000. doi: 10.3758/bf03200807.

23. S. Bilgin and A. C. Guzeler, "Naive Bayes classification of neurodegenerative diseases by using discrete wavelet transform," 2015 19th National Biomedical Engineering Meeting (BIYOMUT). 2015. doi: 10.1109/biyomut.2015.7369440.

24. Y. Ji, S. Yu, and Y. Zhang, "A novel Naive Bayes model: Packaged Hidden Naive Bayes," 2011 6th IEEE Joint International Information Technology and Artificial Intelligence Conference. 2011. doi: 10.1109/itaic.2011.6030379.

25. R. Alizadehsani et al., "Machine learning-based coronary artery disease diagnosis: A comprehensive review," Comput. Biol. Med., vol. 111, p. 103346, Aug. 2019.

26. G. Litjens et al., "State-of-the-Art Deep Learning in Cardiovascular Image Analysis," JACC Cardiovasc. Imaging, vol. 12, no. 8 Pt 1, pp. 1549–1565, Aug. 2019.