ORIGINAL RESEARCH

Submitted: 2.10.2022; Accepted: 1.11.2022; Published online: 25.12.2022

Heart Plaque Detection with Improved Accuracy using Decision tree in comparison 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 Decision Tree algorithm with improved accuracy and comparing it with Least Squares Support Vector Machine. **Materials and Methods:** Decision tree and Least squares Support Vector Machine algorithms are two groups compared in this study. Each group has 20 samples and calculations utilized pretest power of 0.08 with 95% confidence interval. The G power is estimated for samples using clincalc, which has two groups: alpha, power, and enrollment ratio. These samples are split into two groups: training dataset (n = 489 [70%]) and test dataset (n = 277 [30%]). **Results:** The accuracy obtained for Decision Tree was 68.13 % and 67.3 % for the Least Squares Support Vector Machine technique. Since p (2-tailed) < 0.05, in SPSS statistical analysis, a significant difference exists between the two groups. **Conclusion:** It is found that the Decision Tree algorithm is significantly better than the Least Squares Support Vector Machine algorithm in Heart plaque disease detection for the dataset considered.

Keywords

Heart Plaque, Novel statistical feature, Decision Tree, Least Squares Support Vector Machine, Prediction, Machine learning.

Imprint

Vankamaddi Sunil Kumar, K Vidhya. Heart Plaque Detection with Improved Accuracy using Decision tree in comparison with Least Squares Support Vector Machine. Cardiometry; Special issue No. 25; December 2022; p. 1584-1589; DOI: 10.18137/cardiometry.2022.25.15841589; Available from: http://www.cardiometry.net/issues/no25-december-2022/heart-plaque-detection

INTRODUCTION

Heart plaque is entirely created by cholesterol deposition, and if it begins to build up, it will result in coronary artery disease. The approaches for detecting, diagnosing, and self-managing Heart Plaque Disease are investigated in this study. Automatic heart plaque detection and diagnosis techniques were thoroughly examined, as well as detection and self-management options for Coronary Disease [1]; [2]. A machine learning-based prediction system finds the optimal classifier to get the best results when compared to clinical outcomes. The suggested technique, which is based on predictive analysis, focuses on selecting variables that aid in the early detection of Heart Plaque deposition [3]. These techniques yielded a wide variety of accuracy rates. This has motivated academics to experiment with new classifiers or combining many classifiers to increase the quality of their models [4]. Because of its greater accuracy and precision, the suggested pipeline may be utilized to construct mobile applications that can assist medical practitioners in early heart diseases [5]. As a result of its superior performance even in non-linear classification problems, it has been implemented in a diverse spectrum of research fields, ranging from text classification, face recognition, financial application, brain-computer interface, bio-medicine, human action recognition, horse race odds prediction and multiple instance learning [6].

Using machine learning techniques, a lot of research has recently been done on a variety of Heart Plaque disease diagnosis. In a five-year period, 7000 research articles were published in scientific publications and 547 articles were identified in Google scholar for the diagnosis of heart plaque disease. In a recent study, they predicted an accuracy of 57.7 % using the Support Vector Machine algorithm [7] in a paper called Analysis of Heart Plaque for Early Prediction Using K-Means Clustering. The attributes used [8] for heart Plaque disease patients results with an accuracy of 57.7%. For the detection of heart plaque disease, the researchers [9] and [10] employed several algorithms such as Support Vector Machine and k-nearest neighbor achieved a 60 percent accuracy. Principal component analysis (PCA), minimum redundancy and maximum relevance, and five cross validation [11] for deducting dimensionalities achieved an accuracy of 54 percent. Researchers [12], [13][14] used ResNet al-

gorithms 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[15]–[26]

The presence of outliers from the dataset is the main reason for the existing model's poor results. The aim of this work is to detect plaque in the heart using the Decision Tree algorithm with improved accuracy and comparing it with 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 analysis is 2. Group 1 contains Decision Tree and group 2 was the Least Squares Support Vector Machine. The total sample size obtained is 20 (i.e for each group 10). The specified sample analysis is completed using the G power statistical tool with a probability of 80 % [27]. A display with a resolution of 1920x1080 pixels (2nd gen, Ryzen 5 series, 8 GB RAM, 512 GB SSD) and a Matlab program with suitable library and tool capabilities are required to train these datasets. The simulation is done using MATLAB [28] software.

In Group 1, sample preparation is completed by downloading a kaggle dataset. Import the data into Google Colab and 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 decision tree and supervised learning technique are utilized. It's a non-parametric test and worked well with [29] category and continuous output variables. A decision tree is a two-stage classification process with a learning phase and a prediction step [30]. 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 problem, this version finds the solution by solving a set of linear equations. Suykens and Vandewalle proposed Least-squares versions of support-vector machines classifiers. Kernel-based learning methods such as Least-squares versions of support-vector machines are a subset of kernel-based learning methods.

The data set for Heart Plaque Disease was obtained via kaggle. After being processed, the dataset is now ready for training and testing. In data processing, missing data should be removed, null values should be replaced with mean or median values, and data should be standardized. The preprocessed dataset containing photos as input is provided with the Decision Tree and Least Squares Support Vector Machine algorithms. 70% of the data from the entire sample size is utilized for training, while the remaining 30% is used for testing in order to obtain the most accurate result and detection.

A total of 659 patient records were retrieved from kaggle for this Heart Plaque disease data collection. There are 427 samples of healthy people and 232 samples of patients with cardiac plaque disease. There are 12 columns and 659 rows in the database. They contain information on 659 people, including pregnancies, family history, blood pressure, cholesterol, age, smoking, ECG, and blood sugar levels, as well as their outcomes.

Statistical Analysis

The analysis was done using IBM SPSS version 21. It is a statistical software tool used for data analysis. For both proposed and existing algorithms 10 iterations were done with a maximum of 20 samples and for each iteration the predicted accuracy was noted for analyzing accuracy [31]. 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. A detailed analysis has been done on these values for predicting heart diseases.

RESULTS

Figure 1 shows the input image that has been used for analyzing the detection of heart plaque disease and

preprocessed using filters. Fig. 2 compares Decision Tree algorithm and Least Squares Support Vector Machine algorithm in terms of mean accuracy. The two methods are compared using an independent t-test, and the mean accuracy value shows a statistically significant difference. The Decision Tree technique outperforms the Least Squares Support Vector Machine algorithm.

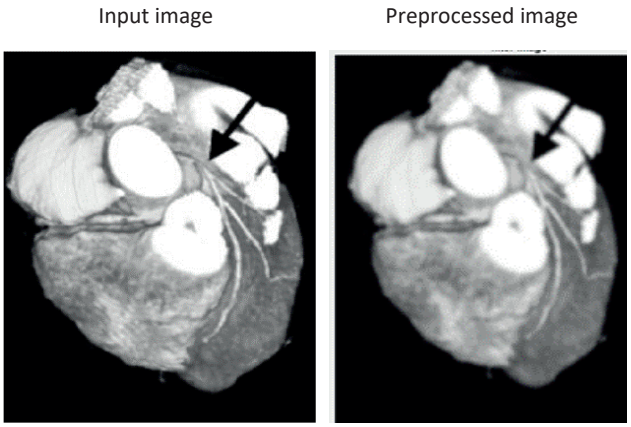

Fig. 1. Input image and Preprocessed image. Preprocessing the input image to remove noise.

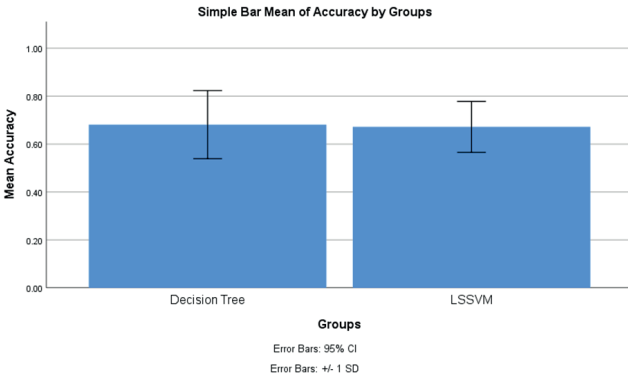

Fig. 2. Comparison of Decision Tree algorithm and Least Squares Support Vector Machine algorithm in terms of mean accuracy. The mean accuracy of Decision Tree is better than Least Squares Support Vector Machine algorithm classifier. X-axis: (Groups) Decision Tree Vs Least Squares Support Vector Machine 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 accuracy of the Novel statistical feature based Decision Tree is 68.13 %. As shown in Fig. 2 the Decision Tree algorithm outperforms the Least Squares Support Vector Machine algorithm in terms of accuracy, and the standard deviation of Decision Tree is better than the Least Squares Support Vector Machine algorithm.

The findings of the Decision Tree and the Least Squares Support Vector Machine for heart plaque disease are shown in Table 1. The detection accuracy for the Decision Tree algorithm is 68.13 %, whereas the detection accuracy for the Least Squares Support Vector Machine approach is 57.3 %. The accuracy statistics for the Decision Tree and the Least Squares Support Vector Machine approaches are shown in Table 2. The Least Squares Support Vector Machine algorithm has a mean of 57.3 %, whereas the Decision Tree algorithm has a mean of 68.13 %. The Decision Tree has a standard deviation of 0.18, while the Least Squares Support Vector Machine approach has a standard deviation of 0.16. The Novel statistical feature based Decision Tree has a standard error mean of 0.04, while the Least Squares Support Vector Machine approach has a standard error mean of 0.03. The Decision Tree has a greater mean accuracy when compared to the Least Squares Support Vector Machine (Table 3). According on the findings, the Novel statistical feature based Decision Tree approach outperformed the Least Squares Support Vector Machine. Table 4 displays the results of an independent t-test, with a significance level of 0.05. The values of accuracy are shown in this table using

Table 1

Samples, features and classes from the 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 Decision Tree and Least Squares Support Vector Machine algorithm. The accuracy value obtained for the Decision Tree and Least Squares Support Vector Machine algorithm is 68.13 % and 67.3 % respectively.

Dataset	Decision Tree		Least Squares Support Vector Machine algorithm	
Heart Plaque	Accuracy	68.13 %	Accuracy	67.3%

Table 3

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

	Group	N	Mean	Std Deviation	Std Error Mean
Accuracy	Decision Tree	20	0.6813	0.0245	0.00549
	Least Squares Support Vector Machine	20	0.6732	0.2750	0.00615

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	0.687	0.41	2.831	38	0.000	0.159	0.05	.0453	.2728
	Not equal variances assumed			2.831	37.596	0.000	0.159	0.05	.0452	0.2729

Levene's test, and they are divided into two categories: when equality of variance is assumed and when equality of variance is not assumed. The hypothesis holds true because the significance value is less than 0.05. Output values (dependent variables) alter in response to changes in input values (independent variables).

DISCUSSION

The Decision Tree achieved a higher accuracy in detecting Heart Plaque disease than the Least Squares Support Vector Machine method (p(2-tailed) = 0.000, Independent sample Test). The Decision Tree has a higher accuracy (mean accuracy= 68.13) than the Least Squares Support Vector Machine technique (mean accuracy = 67.3). The collection contains several attributes that define the disease condition, as well as normal and abnormal human circumstances. The Novel statistical feature based Decision Tree algorithm outperforms the Least Squares Support Vector Machine approach.

The researchers [32] discovered that the accuracy of algorithms may be improved by selection of features. To compare classification algorithms, the researcher [33] used a variety of performance indicators, including accuracy, sensitivity, recall, and specificity. The researchers [34] discovered that feature selection can improve the accuracy of the Decision Tree is 70% [2], which clearly show that the Novel statistical feature based Decision Tree technique is more efficient and performs better than the Least Squares Support Vector Machine approach. The author [35] evaluated the classifying algorithms with the Rsnet101 and multiple machine learning methods, finding that multiple machine learning algorithms outperformed the algorithms by 86 percent [36]. With varying weights and individual impacts that we can't perform in less calibrations to the model, the resulting model is difficult to grasp and analyze, making it challenging to include our business logic. This can be improved in the future by collecting more data.

When the target classes are overlapping, the transform fails. When the dataset size is large, it takes a long time to train and the approach is not computationally efficient. The accuracy of the Decision Tree can be enhanced in the future by including additional data in the training sets, employing multiclass novel statistical features, and integrating it with other techniques. To improve accuracy, multidimensional data can also be transformed to binary data.

CONCLUSION

The Decision Tree Algorithm and Least Squares Support Vector Machine algorithm are two machine learning methods utilized in this work to detect heart plaque disease. In terms of accuracy, the Decision Tree (68.13 %) is significantly better than the Least Squares Support Vector Machine algorithm (67.3%).

REFERENCES

1. D. Hanbay, "An expert system based on least square support vector machines for diagnosis of the valvular heart disease," Expert Syst. Appl., vol. 36, no. 3, pp. 4232–4238, Apr. 2009.
2. M. M. Ghiasi, S. Zendehboudi, and A. A. Mohsenipour, "Decision tree-based diagnosis of coronary artery disease: CART model," Comput. Methods Programs Biomed., vol. 192, p. 105400, Aug. 2020.
3. R. Kumar, A. Jain, and P. Rani, "Coronary artery disease diagnosis using extra tree-support vector machine: ET-SVMRBF," International Journal of Computer Applications in Technology, vol. 66, no. 2. p. 209, 2021. doi: 10.1504/ijcat.2021.10043464.
4. S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?," PLoS One, vol. 12, no. 4, p. e0174944, Apr. 2017.

5. K. S. Nugroho, A. Y. Sukmadewa, A. Vidianto, and W. F. Mahmudy, "Effective predictive modelling for coronary artery diseases using support vector machine," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 11, no. 1. p. 345, 2022. doi: 10.11591/ijai.v11.i1.pp345-355.
6. P. Anooj, "Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules and decision tree rules," Open Computer Science, vol. 1, no. 4. pp. 482–498, 2011. doi: 10.2478/s13537-011-0032-y.
7. B. Alizadeh Savareh et al., "A machine learning approach identified a diagnostic model for pancreatic cancer through using circulating microRNA signatures," Pancreatology, vol. 20, no. 6, pp. 1195–1204, Sep. 2020.
8. N. Sneha and T. Gangil, "Analysis of diabetes mellitus for early prediction using optimal features selection," Journal of Big Data, vol. 6, no. 1. 2019. doi: 10.1186/s40537-019-0175-6.
9. M. Alehegn, R. R. Joshi, and P. Mulay, "Diabetes analysis and prediction using random forest KNN Naïve Bayes and J48: An ensemble approach," Int. J. Sci. Technol. Res, vol. 8, no. 9, pp. 1346–1354, 2019.
10. F. Özyurt, E. Sert, and D. Avcı, "An expert system for brain tumor detection: Fuzzy C-means with super resolution and convolutional neural network with extreme learning machine," Med. Hypotheses, vol. 134, p. 109433, Jan. 2020.
11. E. V. Carrera, A. Gonzalez, and R. Carrera, "Automated detection of diabetic retinopathy using SVM," 2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON). 2017. doi: 10.1109/intercon.2017.8079692.
12. T. Mahboob Alam et al., "A model for early prediction of diabetes," Informatics in Medicine Unlocked, vol. 16, p. 100204, Jan. 2019.
13. A. K. Gárate-Escamila, A. Hajjam El Hassani, and E. Andrès, "Classification models for heart disease prediction using feature selection and PCA," Inform. Med. Unlocked, vol. 19, no. 100330, p. 100330, 2020.
14. D. Indumathy, K. Ramesh, G. Senthilkumar, and S. Sudha, "Investigations on coronary artery plaque detection and subclassification using machine learning classifier," J. Xray Sci. Technol., vol. 30, no. 3, pp. 513–529, 2022.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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.
25. 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.

26. 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.
27. H. M. Johnson and J. H. Stein, "Carotid Intima-Media Thickness Measurement and Plaque Detection for Cardiovascular Disease Risk Prediction," Preventive Cardiology: Companion to Braunwald's Heart Disease. pp. 503–511, 2011. doi: 10.1016/b978-1-4377-1366-4.00030-5.
28. O. Demirkaya, M. H. Asyali, and P. K. Sahoo, Image Processing with MATLAB: Applications in Medicine and Biology. CRC Press, 2008.
29. C. Qin, D. Yao, Y. Shi, and Z. Song, "Computer-aided detection in chest radiography based on artificial intelligence: a survey," Biomed. Eng. Online, vol. 17, no. 1, p. 113, Aug. 2018.
30. A. Narin, "Accurate detection of COVID-19 using deep features based on X-Ray images and feature selection methods," Comput. Biol. Med., vol. 137, p. 104771, Oct. 2021.
31. 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.
32. B. Xiao et al., "Follow the sound of children's heart: A deep-learning-based computer-aided pediatric CHDs diagnosis system," IEEE Internet Things J., vol. 7, no. 3, pp. 1994–2004, Mar. 2020.
33. Z. F. M. Apandi, R. Ikeura, and S. Hayakawa, "Arrhythmia detection using MIT-BIH dataset: A review," in 2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA), Kuching, Aug. 2018. doi: 10.1109/icassda.2018.8477620.
34. T. Chen, S. Zhao, S. Shao, and S. Zheng, "Non-invasive diagnosis methods of coronary disease based on wavelet denoising and sound analyzing," Saudi J. Biol. Sci., vol. 24, no. 3, pp. 526–536, Mar. 2017.
35. 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.
36. R. K. Tripathy, A. Bhattacharyya, and R. B. Pachori, "Localization of Myocardial Infarction From Multi-Lead ECG Signals Using Multiscale Analysis and Convolutional Neural Network," IEEE Sensors Journal, vol. 19, no. 23. pp. 11437–11448, 2019. doi: 10.1109/jsen.2019.2935552.