ORIGINAL RESEARCH

Submitted: 4.11.2022; Accepted: 22.11.2022; Published online: 25.12.2022

Heart Plaque detection with improved accuracy using Logistic Regression 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 major goal of this study is to compare the effectiveness of the Logistic Regression classifier with the Least Squares Support Vector Machine classifier in detecting plaque in the heart with high accuracy. **Materials and Methods:** In this work, the Logistic Regression and least squares Support Vector Machine methods are compared. There were a total of 20 samples in the Kaggle dataset on Heart Plaque disease. To calculate sample G power of 0.08 with 95% confidence interval, Clincalc is utilized. There are two groups in the training dataset (n = 489 (70 percent)) and the test dataset (n = 277 (30 percent)). **Results:** The accuracy of both the Logistic Regression and Least Squares Support Vector Machine algorithms is evaluated. The Least Squares Support Vector Machine approach was only 67.3 % accurate, while the Logistic Regression method was 96 % accurate. Since p(2-tailed) < 0.05, in SPSS statistical analysis, a significant difference exists between the two groups. **Conclusion:** The Logistic Regression algorithm is significantly better than Least Squares Support Vector Machine algorithm in this study in detecting cardiac plaque disease in the dataset.

Keywords

Heart Plaque, Novel texture feature, Logistic Regression algorithm, Least Squares Support Vector Machine, Prediction, Machine learning.

Imprint

Vankamaddi Sunil Kumar, K Vidhya. **Heart Plaque detection with improved accuracy using Logistic Regression and comparing with Least Squares Support Vector Machine.** Cardiometry; Special issue No. 25; December 2022; p. 1600-1604; DOI: 10.18137/cardiometry.2022.25.16001604; Available from: http://www.cardiometry.net/issues/no25-december-2022/logistic-regression-comparing

INTRODUCTION

Cholesterol deposition is the sole cause of heart plaque, which, if left untreated, can lead to coronary artery disease. Early-stage heart plaque disease diagnosis and risk prediction are currently a substantial real-world challenge. This study looks at the methods for detecting, diagnosing, and self-managing Heart Plaque Disease. The detection and identification of cardiac plaques, as well as the diagnosis and self-management alternatives for Coronary Disease, were all thoroughly investigated [1]. The study's purpose is to create a machine learning-based prediction system and determine which classifier produces the best results when compared to clinical outcomes. The proposed technique, which is based on predictive analysis, intends to find qualities that can aid in the early diagnosis of the formation of cardiac plaque [2]. These methods produced a wide range of accuracy outcomes. As a result, scientists are experimenting with novel classifiers or combining many classifiers in order to improve the quality of the models. As a result of their remarkable outcomes in disease diagnosis and prediction, machine learning technologies are becoming more prominent in medical research applications [3].

A lot of study has been done on a number of Heart Plaque disease diagnostics using machine learning methodologies. Over a five-year period, 7 research articles for the diagnosis of heart plaque disease were published in scientific journals, whereas 547 publications were found in Google scholar. In a recent study, they projected an accuracy of 96 percent using the Logistic Regression [4] in a paper called Analysis of Heart Plaque for Early Prediction Using Logistic Regression algorithm. The author [5] used the best attributes from the Heart Plaque disease patient to detect the disease with an accuracy of 67 %. For the detection of Heart Plaque disease, the researchers [6] and [7] 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 [8] for analyzing the models for deducting dimensionalities accuracy was reached by 67 percent. Researchers [9] used logistic regression to diagnose Heart Plaque disease with a 96 percent accuracy.Our team has extensive knowledge and research experience that has translate into high quality publications[10]–[21]

According to current studies, recognising Heart Plaque disease is difficult due to a lack of accuracy, sensitivity, and specificity. The purpose is to use a Logistic Regression algorithm method rather than the Least Squares Support Vector Machine algorithm strategy to improve the accuracy of Heart Plaque disease identification.

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 Logistic Regression 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 software with suitable library and tool capabilities are required [22].

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, [23] the Logistic Regression and supervised learning technique are utilized. A Novel texture feature based Logistic Regression is a two-stage classification process with a learning phase and a prediction step. The model is trained using the training data, and in the prediction stage, it is used to predict 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 (SVM) [24], 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.

Using kaggle, a data set for Heart Plaque disease is collected. The dataset is now ready to be trained and tested after it has been processed. Data should be normalized and missing data should be deleted. Null values should be replaced with mean or median values. The K-Nearest Neighbor and Least Squares Support Vector Machine methods are used to the preprocessed dataset including images as input. To acquire the most accurate result and detection, 70% of the data from the full sample size is used for training, while 30% is used for testing.

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

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 (O'connor 2000). 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.

RESULTS

Figure 1 compares Logistic Regression algorithm and the Least Squares Support Vector Machine technique in detecting Heart plaque. The two methods are compared using an independent t-test, and the mean accuracy value shows a statistically significant difference. The Novel texture feature based Logistic Regression algorithm technique outperforms the Least Squares Support Vector Machine algorithm.

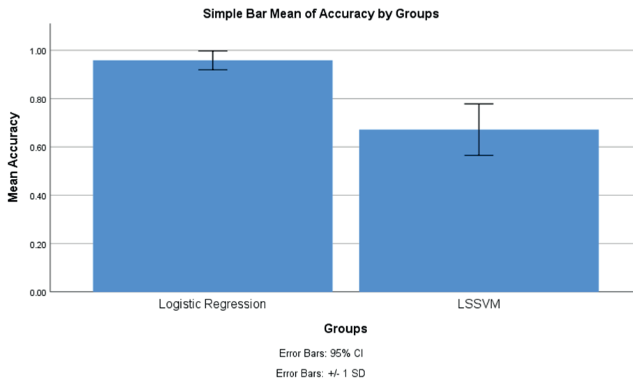

Fig. 2. Comparison of Logistic Regression algorithm algorithm and Least Squares Support Vector Machine algorithm in terms of mean accuracy. The mean accuracy of the Logistic Regression is better than the Least Squares Support Vector Machine algorithm and the standard deviation of the Logistic Regression algorithm is slightly better than the Least Squares Support Vector Machine. X-axis: (GROUPS) Logistic Regression 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 Logistic Regression algorithm has a 96 % accuracy. As shown in Fig 2 the Logistic Regression method surpasses the Least Squares Support Vector Machine technique, and the Logistic Regression is comparatively better than the Least Squares Support Vector Machine algorithm due to the Novel texture feature extraction.

Table 1 summarizes the results of the Logistic Regression algorithm and the Least Squares Support Vector Machine for the Heart Plaque disease data set. The Logistic Regression algorithm has a detection accuracy of 96 percent, whereas the Least Squares Support Vector Machine approach has a detection accuracy of 67.3 percent. Table 2. shows the accuracy statistics for the Logistic Regression algorithm and the Least Squares Support Vector Machine methods. Results of Group Statistics The average for the Least Squares Support Vector Machine algorithm is 67 percent, whereas the average for the Logistic Regression algorithm is 96 percent. The standard deviation of the Logistic Regression algorithm is 0.1589, while the standard deviation of the Least Squares Support Vector Machine algorithm is 0.1529. The standard error mean for the Logistic Regression algorithm is 0.035, while the standard error mean for the Least Squares Support Vector Machine is 0.034. The Logistic Regression algorithm has a higher mean accuracy as Novel texture features are extracted and utilized. The Logistic Regression algorithm outperformed the LSSVM, according to the data. The results of an independent t-test with a significance threshold of 0.05 are shown in Table 3. The Values of Accuracy are classified into two categories in this table 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, our hypothesis holds true. Independent Sample t-Test for significance and standard error determination is shown in Table 4 and as p value is less than 0.05, there is significant difference between two groups.

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 Logistic Regression algorithm and Least Squares Support Vector Machine. The accuracy value obtained for Logistic Regression and Least Squares Support Vector Machine algorithm is 96 % and 67.3 % respectively.

Dataset	Logistic Regression algorithm		Least Squares Support Vector Machine algorithm	
Heart Plaque	Accuracy	96 %	Accuracy	67.3 %

Table 3

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

	Group	N	Mean	Std Deviation	Std Error Mean
Accuracy	Logistic Regression algorithm	20	0.9602	0.1589	0.355
	Discrete Wavelet Transform	20	0.6732	0.1529	0.3419

DISCUSSION

The Logistic Regression technique utilizing Novel texture feature appears to be more accurate than the Least Squares Support Vector Machine strategy in detecting Heart Plaque disease. The accuracy of the Logistic Regression algorithm (96 %) is higher than that of the Least Squares Support Vector Machine technique (67.3 %). In the collection are several attributes that define the disease condition, as well as normal and abnormal human circumstances. The Least Squares

Table 4

Independent Sample t-Test for significance and standard error determination. P value less than 0.05 is 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	0.86	0.010	5.529	38	0.000	0.2727	.04932	0.172	0.372
	Not equal variances assumed			5.529	37.94	0.000	0.2727	.04932	0.172	0.372

Support Vector Machine approach is outperformed by the Logistic Regression algorithm.

The accuracy of the Least Squares Support Vector Machine algorithm and the Logistic Regression method [25]. It is possible to increase the accuracy of these two approaches, as well as the accuracy of other algorithms. Logistic Regression algorithm and the Least Squares Support Vector Machine algorithm, both of which contained and did not include bagging, produced identical results. When new bagging approaches are utilized, it is clear that algorithms perform better. To compare algorithms for classification, the researcher [26] used a variety of performance indicators, including accuracy, sensitivity, recall, and specificity. The accuracy of the Logistic Regression algorithm is 75% [27] [28]. The classifying algorithms [29] with the DenseNET and multiple machine learning methods, proved that multiple machine learning algorithms outperformed the algorithms by 92 percent.

As logistic regression has a linear decision surface, it cannot tackle nonlinear issues. As a result, non linear features must be transformed, which can be done by increasing the number of features such that the data can be separated linearly in higher dimensions. In future the features of a dataset are linearly separated, to enhance the effectiveness of Logistic Regression.

CONCLUSION

The methods utilized in this work are Logistic Regression and Least Squares Support Vector Machine technique. It is observed that Logistic Regression is significantly better than Least Squares Support Vector Machine technique with accuracy values of 96 % and 67.3 % respectively.

REFERENCES

1. A. H. M. Yahaya, Multiple Logistic Regression Approach to Coronary Heart Disease (CHD). 2012.
2. K. Mala, V. Sadasivam, and S. Alagappan, "Neural network based texture analysis of CT images for fatty and cirrhosis liver classification," Appl. Soft Comput., vol. 32, pp. 80–86, Jul. 2015.
3. 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.
4. P. Kvam and J. S. Sokol, "A logistic regression/ Markov chain model for NCAA basketball," Naval Research Logistics, vol. 53, no. 8. pp. 788–803, 2006. doi: 10.1002/nav.20170.
5. W. R. Thompson, Variable Selection of Correlated Predictors in Logistic Regression: Investigating the Diet-heart Hypothesis. 2009.
6. F. E. Harrell, Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer Science & Business Media, 2013.
7. C. E. Ford, Polychotomous Logistic Regression Analysis. 1986.
8. 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.
9. Y. Liu et al., "Relationship between Coronary VH-IVUS Plaque Characteristics and CTRP9, SAA, and Hcy in Patients with Coronary Heart Disease," J. Healthc. Eng., vol. 2022, p. 1635446, Mar. 2022.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. A. Vidales, Econometric Models with MATLAB: Generalized Linear Models, Poisson Regression, Logistic Regression, Decision Trees and Discriminant Analysis. Independently Published, 2019.
23. Y. Huang, H.-Y. Wang, W. Jian, Z.-J. Yang, and C. Gui, "Development and validation of a nomogram to predict the risk of death within 1 year in patients with non-ischemic dilated cardiomyopathy: a retrospective cohort study," Sci. Rep., vol. 12, no. 1, p. 8513, May 2022.
24. J. A. K. Suykens, T. Van Gestel, and J. De Brabanter, Least Squares Support Vector Machines. World Scientific, 2002.
25. R. D. Bawol, Measurement Error in Logistic Regression by Discriminant Analysis with Applications to the Epidemiology of Coronary Heart Disease. 1979.
26. S. Jiang, Heart Disease Prediction Using Machine Learning Algorithms. 2020.
27. U. Neisius, G. Zhou, R. E. Ward, R. C. Ellison, J. M. Gaziano, and L. Djoussé, "Dairy product consumption and calcified atherosclerotic plaques in the coronary arteries: The NHLBI Family Heart Study," Clin Nutr ESPEN, vol. 49, pp. 517–521, Jun. 2022.
28. L. Boix-Palop et al., "Risk of cardiac device-related infection in patients with late-onset bloodstream infection. Analysis on a National Cohort," J. Infect., May 2022, doi: 10.1016/j.jinf.2022.05.022.
29. T. N. K. Mäkelä, T.-P. Tuomainen, S. Hantunen, and J. K. Virtanen, "Associations of serum n-3 and n-6 polyunsaturated fatty acids with prevalence and incidence of non-alcoholic fatty liver disease," Am. J. Clin. Nutr., Jun. 2022, doi: 10.1093/ajcn/nqac150.