International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Special Issue: 1 Volume No.7 April – 2023
DOI:10.46647/ijetms.2023.v07si01.092 ISSN: 2581-4621

PREDICTION OF HEART DISEASE USING MACHINE LEARNING

S. Athinarayanan[1], SP Maithily[2], K Manikanta[3], T Mounika[4], T Praveen Kumar[5]
Professor, Student, Department of CSE, AITS, Tirupati

Abstract
Selected characteristics are used to compare machine learning algorithms and ensemble learning techniques. Models are evaluated using a variety of techniques, including accuracy, memory. The outcomes demonstrate that the decision tree-based technique produced the best results.
Keywords: Machine Learning, Support Vector Machine, Naive Bias, Hybrid model

Introduction

17.7 million deaths worldwide in 2016 were on average caused by heart disease, or around 31% [20]. Reveals that low- and middle-income nations account for 82% of the cases, 17 million people under the age of 70 are at risk for non-infectious diseases, 6.7 million experience strokes, and 7.4 million have heart disease. [19]. As well as for one-third of all fatalities worldwide. Cardiovascular disease affects not only the health of individuals but also the economics and costs of nations. The majority of heart illnesses, including stroke and cardiac problems, are of microvascular origin. The main causes of cardiac disease are being overweight, having diabetes, having a family history of smoking, and having high cholesterol[18].The specific issues that must be examined in order to examine the heart disease mischance are those that relate to actions.

Related Work

1.Material and Methods:

1.1 Dataset:

Pham et al. [7] employed K-mean, a genetic algorithm, and a recurrent neural network (RNN) to predict cardiac disorders.The outcome demonstrates that, when compared to other methods, ensemble learning has enhanced the performance of heart illness prediction. and Win [4] utilised sequential minimising feature subset selection (CFS) and optimising (SMO) to the prognosis of heart disease.

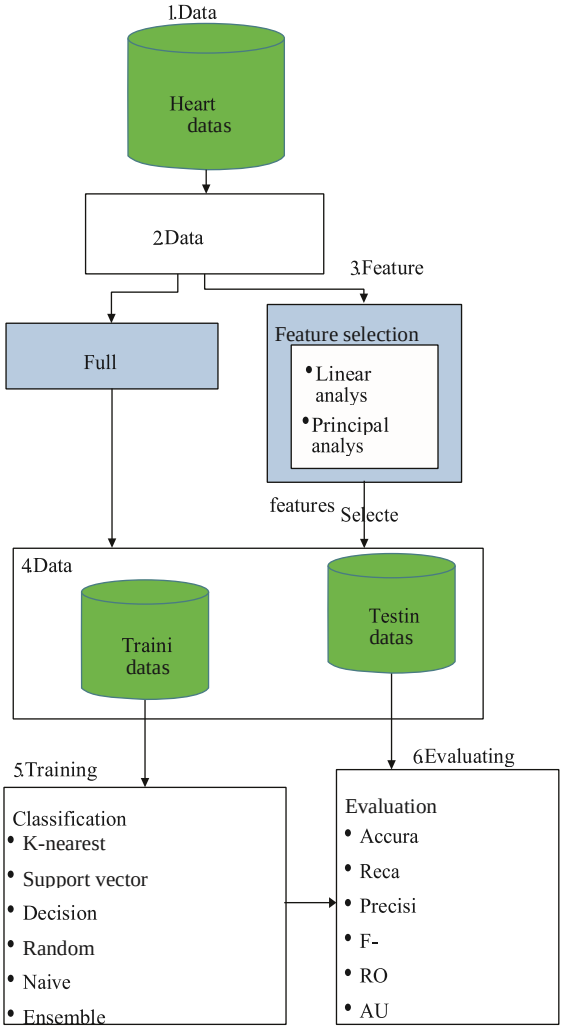
2. Algorithm

The following provides a detailed explanation of each phase of the suggested approach.
2.1. Data Collection: To train and test models, the heart disease dataset is used. 13 characteristics, 1025 records, and one target column make up this dataset. The goal column has two classes: 1 for cardiac ailments for illnesses other than those of the heart. Specifics of the features are described in Table 1.
2.2 Data Preprocessing*:* The features have been resized to fall between . It is important to note that the dataset is purge missing values.
2.3 Feature Extraction *(FE)*. An important step since the machine learning classifier's classification effectiveness is frequently impacted by irrelevant features.
2.4 Data Splitting: testing set is 25% and training set is 75%.
2.5 Training Models Heart illness is categorised using machine learning algorithms, among other types.

(1)That enables one to predict how new samples will be categorised. It is used by several groups. It can be applied to regression and classification forecasting issues.
The two different ensemble procedure are bagging and boosting.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Special Issue: 1 Volume No.7 April – 2023
DOI:10.46647/ijetms.2023.v07si01.092 ISSN: 2581-4621

(a)Boosting is the process of creating a model sequence with the intention of fixing mistakes that have occurred in the models. The dataset is meticulously the classification method is then trained on a sample to build a series of average efficiency models. As a result, new samples are generated using the elements of the prior model that were incorrectly classified. The ensemble technique then improves upon itself by mixing the weak models. Algorithm 1 provides the boosting pseudocode.
(b)Bagging for each subset in a replacement training set that has several subsets. The final performance forecast is consistent with the sub models taken together. Following that, a vote process is carried out for each categorization model, as can be seen in the bagging algorithm's pseudocode. As a result, the classification result is established using the vast majority of the average values. Algorithm 2 provides the bagging pseudocode.



3. Result:

After performing the machine learning approach for training and testing we find that accuracy of the xgboost is better compared to other algorithms. Accuracy is calculated with the support of the confusion matrix of each algorithm, here the number count of TP, TN, FP, FN is given and using the equation of accuracy, value has been calculated and it is concluded that extreme gradient boosting is best with 81% accuracy and the comparison is shown below.

TABLE: Accuracy comparison of algorithms

Algorithm	Accuracy
XG Boost	**81.3%**

DOI:10.46647/ijetms.2023.v07si01.092 ISSN: 2581-4621

SVM	**80.2%**
Logistic Regression	**79.1%**
Random Forest	**79.1%**
Navie Bayes	**76.9%**
Decision Tree	**75.8%**
Adaboost	**73.6%**

4. Conclusion:

Heart diseases are a major killer in India and throughout the world, application of promising technology like machine learning to the initial prediction of heart diseases will have a profound impact on society. The early prognosis of heart disease can aid in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine. The number of people facing heart diseases is on a raise each year. This prompts for its early diagnosis and treatment. The utilization of suitable technology support in this regard can prove to be highly beneficial to the medical fraternity and patients. In this paper, the seven different machine learning algorithms used to measure the performance are SVM, Decision Tree, Random Forest, Naïve Bayes, Logistic Regression, Adaptive Boosting, and Extreme Gradient Boosting applied on the dataset. The expected attributes leading to heart disease in patients are available in the dataset which contains 76 features and 14 important features that are useful to evaluate the system are selected among them. If all the features taken into the consideration then the efficiency of the system the author gets is less. To increase efficiency, attribute selection is done. In this n features have to be selected for evaluating the model which gives more accuracy. The correlation of some features in the dataset is almost equal and so they are removed. If all the attributes present in the dataset are taken into account then the efficiency decreases considerably. All the seven machine learning methods accuracies are compared based on which one prediction model is generated. Hence, the aim is to use various evaluation metrics like confusion matrix, accuracy, precision, recall, and f1-score which predicts the disease efficiently. Comparing all seven the extreme gradient boosting classifier gives the highest accuracy of 81%.

5. References:

[1] M. A. Abushariah, A. A. Alqudah, O. Y. Adwan et al"Automatic heart disease diagnosis system based onartificial neural network (ann) and adaptive neuro-fuzzy inference systems (anfis) approaches," Journal of Software Engineering and Applications, vol. 7, p. 1055, 2014.
[2] R. Bhat, S. Chawande, and S. Chadda, "Prediction of test for heart disease diagnosis using artificial neural network," Indian Journal of Applied Research, vol. 9, 2019.
[3] S.Nalluri, R.V.saraswathi, S.Rama subba reddy,K.Govinda, and E.Swetha, "chronic heart disease prediction using datamining tecniques," In data engineering and communication technology,pp.903_912,2020.
[4] A. N. Oo and K. T. Win: Feature Selection Based Sequential Minimal Optimization (Smo) Classifier for Heart.
[5] K. Raza, "Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule," InUHealthcare Monitoring Systems, pp. 179–196, 2019.
[6] R. Kannan and V. Vasanthi, "Machine learning algorithms with roc curve for predicting and diagnosing the heart disease," InSoft Computing and Medical Bioinformatics, pp. 63– 72, 2019.
[7] N. S. R. Pillai, K. K. Bee, and J. Kiruthika, "Prediction of heart disease using rnn algorithm," International Research Journal of Engineering and Technology, vol. 5, 2019.

DOI:10.46647/ijetms.2023.v07si01.092 ISSN: 2581-4621

[8] R. Atallah and A. Al-Mousa, "Heart disease detection using machine learning majority voting ensemble method," in Proceedings of the 2019 2nd International Conference on New Trends in Computing Sciences (ICTCS), pp. 1–6, Amman, Jordan, October 2019.
[9] N. S. C. Reddy, S. S. Nee, L. Z. Min, and C. X. Ying,"Classification and feature selection approaches by machine learningtechniques: heart disease prediction," *International Journal of Innovative Computing*, vol. 9, 2019.
[10] V. Pham, Q. De Hemptinne, J.-M. Grinda et al., "Giant coronary aneurysms, from diagnosis to treatment: a literature review," *Archives of Cardiovascular Diseases*, vol. 113, pp. 59–69, 2020.
[11] A. A. Ali, H. S. Hassan, and E. M. Anwar, "Heart diseases diagnosis based on a novel convolution neural network and gaterecurrent unit technique," in Proceedings of the 2020 12th International Conference on Electrical *Engineering* (ICEENG), vol. 145–150, Cairo, Egypt, July 2020.
[12] T. Obasi and M. O. Shafiq, "Towards comparing and using machine learning techniques for detecting and predicting heartattack and diseases," in Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), pp. 2393–2402, Los Angeles, CA, USA, April 2019.
[13] M. Aljanabi, H. Qutqut, and M. Hijjawi, "Machine learning classification techniques for heart disease prediction: a review," *International Journal of Engineering and Technology*, vol. 7, pp. 5373–5379, 2018.
[14] A. Sagheer, M. Zidan, and M. M. Abdelsamea, "A novel autonomous perceptron model for pattern classification applications," *Entropy*, vol. 21, no. 8, p. 763, 2019.
[15] A.-H. Abdel-Aty, H. Kadry, M. Zidan et al., "A quantum classification algorithm for classification incomplete patterns based on entanglement measure," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 9, pp. 1–8, 2020.
[16] K. Vanisree and J. Singaraju, "Decision support system for congenital heart disease diagnosis based on signs and symptomsusing neural networks," *International Journal of Computer Application*, vol. 19, pp. 6–12, 2011.
[17] S. N. Blair, "Commentary on Wang Y et al. "An Overview of Non-exercise Estimated Cardiorespiratory Fitness: estimation Equations, Cross-Validation and Application"" *Journal of Science in Sport and Exercise*, vol. 1, no. 1, pp. 94-95, 2019.
[18] J. Nahar, T. Imam, K. S. Tickle, and Y.-P. P. Chen, "Association rule mining to detect factors which contribute to heart disease in males and females," *Expert* Systems with Applications, vol. 40, no. 4, pp. 1086–1093, 2013.
[19] World Health Organization. cardiovascular diseases/en. 2019..
[20] . M. Sanz, A. Marco del Castillo, S. Jepsen et al., "Periodontitis and cardiovascular diseases: consensus report," Journal of Clinical Periodontology, vol. 47, no. 3, pp. 268–288, 2020.