Semiconductor Optoelectronics
ISSN: 1001-5868

MACHINE LEARNING CLASSIFICATION METHODS FOR IDENTIFYING HEART DISEASE

D Vijendra Kumar[1],Sathi Rushmitha siva nageswari[2], Kakunuri Kalyani[3],Nimmalapudi Sriram[4],Marukurthi Ashok Kumar[5]
[1]AssistantProfessor, Dept of ECE, Godavari Institute of Engineering and Technology(A),Rajahmundry,AP
[2,3,4,5]Students, Dept of ECE, Godavari Institute of Engineering and Technology (A),Rajahmundry,AP

Abstract- Despite the serious nature of the illness, heart problems affect a sizable population worldwide. Just as in any other medical discipline, cardiology depends greatly on quick and precise diagnosis of heart ailment. We describe an approach that uses machine learning methods that is both effective and practical for diagnosing cardiac disorders. Classification algorithms such as the Support Vector Machine, the K-nearest neighbour, the Naive Bayes classifier, the Decision Tree, and the Hierarchical Random Forest Logistic Model (HRFLM) are the backbone of the system, and standard feature selection algorithms such as Recursive Feature Elimination are used to remove unnecessary data from the dataset. As an alternate approach, Our quick conditional mutual information feature selection method is proposed as a solution to this issue. Classification algorithms may run more quickly with the help of feature selection techniques, without having to compromise on accuracy. Moreover, the leave-one-subject-out cross-validation method has been employed to identify the most productive methods of assessing models and enhancing its hyper parameters. The performance of a classifier may be measured using these indicators. In this study, we examined the performance of several classifiers by analyzing the characteristics selected by each method. The experimental findings show that the suggested feature selection algorithm (RFE) may be successfully combined with the classifier Decision Tree to provide an innovative intelligent system for detecting heart disease. Our proposed diagnostic system, RFE-DT, has been shown to be more accurate than previous approaches. In addition, the suggested strategy for detecting cardiovascular disease is easy to adopt in clinical settings.

Keywords: prediction model, heart disease prediction, feature selection, classification algorithms, cardiovascular disease (CVD)

1. INTRODUCTION

Identifying cardiac disease has always been a difficult task because of the wide variety of variables that may play a role in causing it. An array of machine learning techniques has been used to the problem of human heart disease incidence estimation. Sickness severity is classified using a variety of approaches, such as K-Nearest Neighbor (KNN), Decision Trees (DT), and Naive Bayes (NB). Due to the intricacy of heart disease, managing it is a delicate procedure. The heart is particularly vulnerable to the effects of not doing so. Machine learning and medical research methodologies may help identify a wide variety of metabolic disorders. Machine learning with a focus on classification may help with both the detection of cardiac disease and the investigation of previously undiscovered patterns in data.

In addition, decision tree have been demonstrated to be effective in predicting cardiac events. Many alternative knowledge-abstraction techniques based on well-established machine learning practises have been tried for predicting cardiovascular disease. In order to build a prediction model that incorporates a wide range of analyses and their reciprocal linkages, a large amount of reading has gone into this project. Hybrid techniques are used to define the recently created strategies that mix preexisting ones. For this classification task, we use a radial basis function network (RBFN), allocating 20% of the data to the actual classification and 80% to network training. A brief introduction to the Computer Assisted Decision Support System (CADSS) and its applications in the hard sciences and the medical field is also included. It has been shown that Machine Learning techniques may be utilised to provide more precise and timely illness predictions in the healthcare industry. The well-known Cleveland dataset, built from a UCI machine learning repository, is used to check our work. When time permits, we'll examine how our results stack up against competing supervised learning strategies. Rules for cardiovascular disease are developed using the strongest evolutionary approach, the confusion matrix. The random application of rules utilising encoding approaches has increased accuracy overall. Heart illness may be diagnosed using a wide range of parameters including heart rate, sex, age, and many more. Classification algorithms are used into the ML methodology to get more reliable findings. Classification approaches are currently the norm for predicting complex diseases including cardiovascular disease and neurological problems. Our proposed method for predicting cardiovascular disease includes thirteen factors. When compared to state-of-the-art methods used in previous research, the results show substantial improvement. Stent placement in the carotid artery has been a regular treatment in recent years. A CAS increases the risk of a major adverse cardiac event in the elderly with heart illness. This evaluation becomes essential. To produce results, we use a voting classifier, whose stellar track record for foreseeing cardiac abnormalities motivated us to continue this avenue of inquiry. Methods for categorising data are presented, such as the use of posterior probability and the projection of values based on different prior processes. With an outstanding maximum accuracy of 98%, the DT model stands head and shoulders above competing approaches. We use the Cleveland heart dataset to better forecast cardiac abnormalities in all of our tests, and we do so by using classification algorithms. Due in large part to the IoT, machine learning (ML) has also made tremendous strides in recent years (IoT). It has been shown that device identification among IoT devices may be achieved utilising data from network traffic monitoring with ML algorithms. Nine different kinds of Internet of Things (IoT) devices, personal computers, and mobile phones had their network traffic gathered and categorised by Median et al. The multi-stage Meta classifier was developed through supervised learning. Classifier functionality begins with determining whether the incoming data was generated by an IoT device. Assigning a class to each piece of IoT hardware is the second step. There is enormous potential in applying deep learning to the task of sifting through the vast volumes of data produced by IoT sensors in complex environments. Since machine learning is implemented on several levels, it may also thrive in a distributed computing environment, such as that seen in edge nodes.

Objective:

The main goal of this research is to forecast cardiac issues utilising standard feature selection techniques like the Recursive Feature Elimination (RFE) method and other classification algorithms like the Support vector machine and K-nearest neighbour. In addition, it developed a novel, speedy conditional mutual information feature selection method for solving the feature-selection issue. The method of feature selection is used to improve classification accuracy and computational performance.

Problem Definition:

Despite the intricacy, many people all around the world suffer from cardiovascular disease. Diagnosing heart sickness quickly and accurately is of paramount importance in cardiology, as it is in other fields of medicine. Without access to cutting-edge diagnostic technologies and trained medical professionals, treating heart illness is an uphill battle. An accurate diagnosis and effective treatment might save many lives. Over 26 million individuals are living with HD, and 3.6 million new instances are diagnosed annually, as reported by the European Society of Cardiology. In sum, the goal of this study is to apply a number of machine learning approaches to improve the possibility for early and accurate identification of cardiac ailment.

2. PROPOSED SYSTEM

Our proposed heart illness diagnosis approach is rapid and very accurate. The system was built using a variety of classification methods, including the Support vector machine, K-nearest neighbour, Naive bayes, Decision tree, and Voting Classifier (HRFLM), with common feature selection procedure, such as Recursive Feature Elimination, used to prune irrelevant data. Using a features selection strategy may improve the classification accuracy and speed of a classification system.

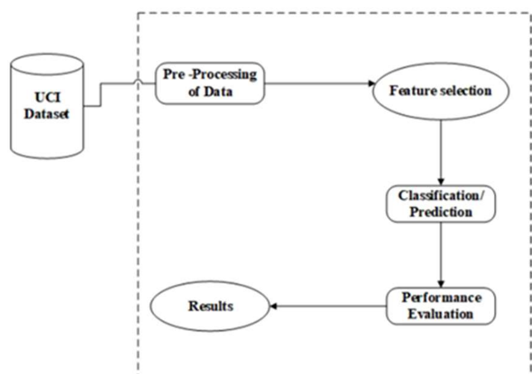
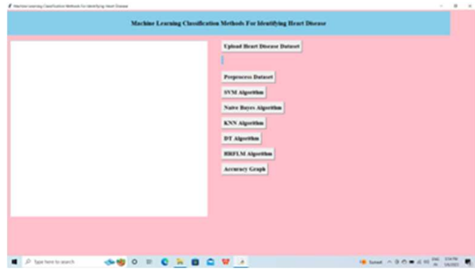


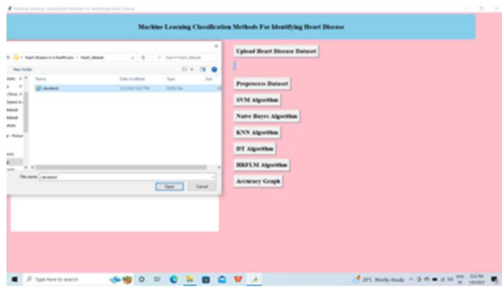
Fig1: System Architecture

Numerous classification and prediction methods, including Support Vector Machines, Naive Bayes, and others, are used to make predictions about heart diseases. Inaccuracy plagues every single method of prediction now in use. To improve the accuracy of her predictions for the heart dataset, the scientists have devised a novel approach they call Hybrid Machine Learning by merging two existing classification algorithms, Linear Model and Random Forest. In order to help create a hybrid algorithm using a Linear Model and a Random Forest, the classification voting algorithm will compare the two algorithms' prediction accuracy and choose the one that produces the most reliable results. That's why it's feasible that constantly using a hybrid model will allow for the development of a more precise algorithm for predicting cardiac disease.

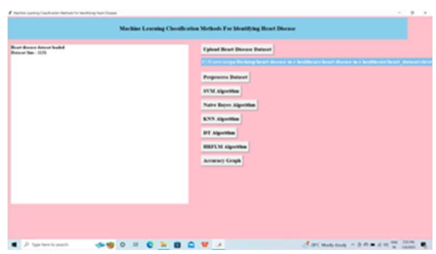
3. RESULTS



Follow the aforementioned steps to upload a dataset related to cardiovascular illness.



Previous window is now uploading the 'cleveland.data' dataset. After your data set has been uploaded, it will be shown in the section below the screen.



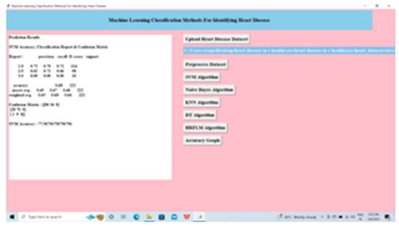
There are 1131 items in the dataset, as seen in the preceding window; utilise the 'Pre-process Dataset' option to remove any characters or other non-numerical information.

Preprocessing:

The above window displays the results of the program's first processing, which was to randomly split the whole dataset into train and test sets. The datasets are comprised of 887 training records and 222 test records.

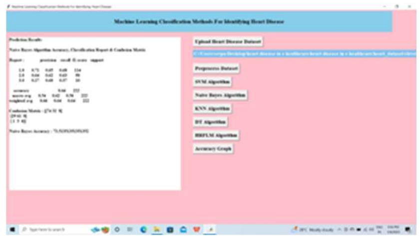
SVM Algorithm

To develop a support vector machine (SVM) model using the train dataset, click the 'Run SVM Algorithm' button. This model may then be used to classify data from the test dataset.



Click the "Run Naive Bayes Algorithm" button to see how accurate the Naive Bayes algorithm is compared to the SVM's 77.5% accuracy shown above.

Naïve Bayes:



Naive Bayes showed an accuracy of 73.5 percent; to check how well the KNN algorithm does, choose the option that reads "Run KNN Algorithm".

KNN Algorithm:

You can see that KNN got an accuracy of 80% in the window just above this one; to learn about the success rate of the DT Algorithm, choose the button labelled "Run DT Algorithm".

DT Algorithm:

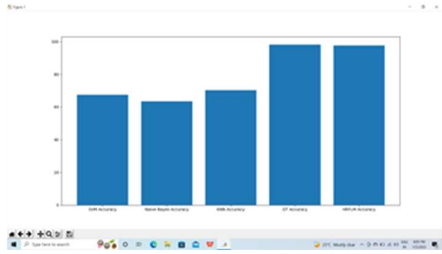
We can now test how well the suggested solution works by running the HRFLM algorithm, knowing that DT has the greatest accuracy of any method (98%).

HRFLM:

Based on the results shown in the previous picture, HRFLM was able to reach an accuracy of

97.7%. Just click the "Accuracy Graph" link below to see the infographic.

Accuracy Graph:



The accompanying graph plots the precision of several algorithms versus their execution times. Among the suggested algorithms, the Decision Tree and HRFLM have the best accuracy.

4. CONCLUSION

In this study, we introduce a robust machine learning-based cardiac diagnostic system's design and implementation. KNN, HRFLM, SVM, NB, and DT are only few of the machine learning classifiers used into the system's design. This is tested using the Cleveland heart disease dataset. Performance evaluation method is also used to evaluate the effectiveness of the identifying system. In order to improve classification accuracy while decreasing the amount of time required for processing by diagnostic systems, we use feature selection algorithm, a unique part of our study.

Future Scope

A variety of methods, including the Support Vector Machine, Naive Bayes, Random Forest Classifier, K-Nearest Neighbor, and Decision Tree, may be used to train the heart disease dataset and enhance its classification accuracy for early detection.

REFERENCES

[1] R. Tao et al., “Magnetocardiography-Based Ischemic Heart Disease Detection and Localization Using Machine Learning Methods”, vol. 66, no. 6, pp. 1658–1667, 2019.

[2] A. Rahim, Y. Rasheed, F. Azam, M. W. Anwar, M. A. Rahim, and A. W. Muzaffar, “An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases,” IEEE Access, vol. 9, pp. 106575–106588, 2021, doi: 10.1109/ACCESS.2021.3098688.

[3] E. Puyol-Anton et al., “Regional Multi-View Learning for Cardiac Motion Analysis: Application to Identification of Dilated Cardiomyopathy Patients,” IEEE Trans. Biomed. Eng., vol. 66, no. 4, pp. 956–966, 2019, doi: 10.1109/TBME.2018.2865669.

[4] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, “HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System,” IEEE Access, vol. 8, pp. 133034–133050, 2020, doi: 10.1109/ACCESS.2020.3010511.

[5] S.Mohan, C.Thirumalai, &G. Srivastava, “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques,” IEEE Access, 81542–81554,2019, Https://Doi.Org/10.1109/Access.2019.2923707

[6] A. S. Abdullah and R. R. Rajalaxmi, ‘‘A data mining model for predicting the coronary heart disease using random forest classifier,’’ in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 22–25.

[7] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, ‘‘Using PSO algorithm for producing best rules in diagnosis of heart disease,’’ in Proc. Int. Conf. Comput.

Appl. (ICCA), Sep. 2017, pp. 306– 311.
[8] N. Al-milli, ‘‘Backpropogation neural network for prediction of heart disease,’’ J. Theor. Appl.Inf. Technol., vol. 56, no. 1, pp. 131–135, 2013.
[9] C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, ‘‘Analysis of neural networks based heart disease prediction system,’’ in Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233–239.
[10] P. K. Anooj, ‘‘Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules,’’ J. King Saud Univ.-Comput. Inf. Sci., vol. 24, no. 1, pp. 27–40, Jan. 2012. doi: 10.1016/j.jksuci.2011.09.002.
[11] L. Baccour, ‘‘Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets,’’ Expert Syst. Appl., vol. 99, pp. 115–125, Jun. 2018. doi: 10.1016/j.eswa.2018.01.025.
[12] C.-A. Cheng and H.-W. Chiu, ‘‘An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database,’’ in Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2017, pp. 2566–2569.
[13] H. A. Esfahani and M. Ghazanfari, ‘‘Cardiovascular disease detection using a new ensemble classifier,’’ in Proc. IEEE 4th Int. Conf. Knowl.- Based Eng. Innov. (KBEI), Dec. 2017, pp. 1011–1014.