International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 07 | July 2022 www.irjet.net p-ISSN: 2395-0072

E-Healthcare monitoring System for diagnosis of Heart Disease using Machine Learning

Rakshitha G R[1], Shejal Shankar[2], Vijay Kumar S[3]

[1]*Student, Dept. of Information Science and Engineering, BNM Institute of Technology, Karnataka, India*
[2] *Student, Dept. of Information Science and Engineering, BNM Institute of Technology, Karnataka, India*
[3] *Assistant Professor, Dept. of Information Science and Engineering, BNM Institute of Technology, Karnataka, India*

Abstract – Modern healthcare technology advancements have a tremendous impact on how well medical services are provided and how many lives are saved. Cardiovascular illness, often known as heart disease, is the most deadly and complex disease and is difficult to diagnose with the unaided eye. According to the WHO cardiovascular disease (CVD) claims millions lannually. According to Global Burden of Disease, CVDs account for about 24.8% of all deaths in India. The number of CVD-related deaths in India has increased annually, from 2.26million in 2019 to 4.77 million in 2020. Heart disease is typically diagnosed by a doctor after reviewing the patient's medical history, the results of their physical exam, and any troubling symptoms. However, the results of this method of diagnosis do not reliably identify patients who have cardiac disease. Additionally, it is costly and computationally challenging to assess. Most clinical diagnoses are made by doctors with training and experience. However, incidences of incorrect diagnosis and treatment continue to be reported. Numerous diagnostic tests are required of patients. Many times, not every test helps with a disease's accurate diagnosis. Machine learning algorithms can foresee this kind of illness. Utilizing various machine learning approaches to quickly analyze and diagnose HD is one of the project's primary research goals. Additionally, it appears that the machine learning prediction model is a crucial feature in this field of study. With the use of certain methods like Feature Selection, Record, Attribute Minimization, and Classification, this work intends to offer a new heart disease prediction model in this situation.

Key Words: Feature Selection, Classification, Machine Learning.

1. INTRODUCTION

Systems can learn and develop automatically thanks to machine learning, a subset of AI. Its the study of computer algorithms that can get better on their own over time by exploiting data and experience. Along with being capable ofsupervised learning, unsupervised learning, and reinforcement learning, ML and DL models can also learn in other ways. To categorize or generate predictions, supervised learning uses labelled datasets; this process necessitates human interaction to accurately identify the incoming data. Unsupervised learning, in contrast, does not require labelled datasets; instead, it finds patterns in the dataand groups them according to any differentiating traits. A model learns to improve its accuracy for completing an activity in an environment based on feedback through the process of reinforcement learning. Machine learning is becoming increasingly popular as a result of its superior accuracy when taught on massive amounts of data. When it comes to challenging issues like speech recognition, natural language processing, and image classification, it really excels.

2. RELATED WORK

In [3], introduces a method that uses fewer criteria to predict the presence of heart disease more effectively. Thirteen factors were first taken into consideration for predicting heart disease. In our research, a genetic algorithm is utilized to identify the characteristics that are most helpful in the detection of cardiac conditions. Using genetic search, thirteen traits are reduced to six. The diagnosis of patients is then predicted using three classifiers—Naive Bayes, Classification by Clustering, and Decision Tree—with the same accuracy as before the reduction in the number of characteristics. The observations also show that after including feature subset selection the Decision Tree data mining technique beats the other two data mining strategies.

In [4], offers to use data mining techniques to uncover hidden patterns that are significant to cardiac illnesses and to forecast whether the patient has a bad heart. The Weka tool is used to implement the selected strategy. It has tools for pre- processing information, categorization, regression, clustering, association rules, and visualization. For this study, decision tree classifiers with 10-fold Cross-Validation in Test mode aretaken into consideration.

In [2], They suggested that the prediction model include procedures like Classification, Record and Attribute reduction and Feature Extraction. This paper's main contribution is a novel hybrid optimization technique that combines the PSO and the well-known Lion algorithm. The suggested algorithm teaches the neural network (NN) how to learn the dimension-reduced data from PCA. The framework for predicting heart disease involves the extraction of features, minimizing of records and characteristics when Compared to LA-NN, an improvement in accuracy of 7.41%.

3. KEY TECHNOLOGIES

3.1 NumPy

The library known as NumPy, or "Numerical Python," contains multidimensional array objects and a selection of procedures for handling those arrays. Arrays can be subjected to logical and mathematical operations using NumPy.

3.2 Pandas

Pandas is a data analysis and manipulation software package created for the Python programming language. It provides data structures and methods for working with time series and numerical tables.

3.3 Flask

It is a Python-based micro web framework. Because it doesn't need any tools or libraries, it is referred to as a microframework. It lacks any components where pre-existing third-party libraries already provide common functions, such as a database abstraction layer, form validation, or other components.

3.4 Scikit Learn

The most effective and reliable Python machine learning library is called Sklearn (Skit-Learn). Through a Python consistency interface, it offers a variety of effective tools for statistical modelling and ML, including classification, regression, clustering, and dimensionality reduction. This library is based on NumPy, SciPy, and Matplotlib and was written primarily in Python.

4. PROPOSED METHOD

Modified SVM is based on the modified particle swarm optimization. It does a simultaneous search for the SVM classifier's regularization parameter value, kernel function type, and kernel function parameter values.
Such a classifier provides high quality of data classification. The SVM to calculate the which class does the testing data belongs and which also reduce time consumption.

The particle swarm algorithm (Particle Swarm Optimization, PSO algorithm), which is based on the notion that optimization issues may be solved by modelling the behavior of animal groupings. A new strategy that implements the simultaneous search for the best type of the kernel function, values of the kernel function parameters, and value of the regularization parameter is presented in addition to the conventional approach of applying the PSO algorithm. Modified SVM reduces misclassification and optimizes time complexity by transforming data using a method known as kernel function and then determining the best boundary between the potential outputs based on these transformations. SVM selects the extreme vectors and points that aid in the creation of the hyperplane. Support vectors, which are used to represent these extreme instances, form the basis for the SVM method.

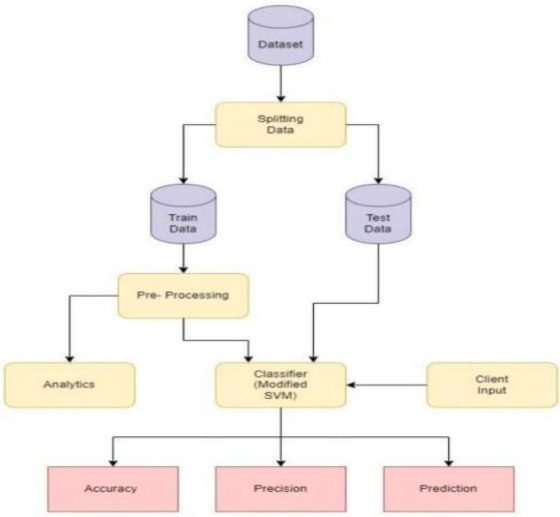

Fig -1: Block Diagram of Proposed Method

4.1 Dataset Preparation

Data preparation, which includes data visualization, analysis, and information mining using machine learning applications, is the most popular method for obtaining, collecting, categorizing, and organizing information. The Cleveland heart disease dataset from the UCI Repository was the one used in this experiment. Here, 13 characteristics are considered.

4.2 SVM

Support Vector Machine (SVM), one of the most popular supervised learning methods, is applied to classification and regression problems. In order to swiftly categorize new data points in the future, the SVM algorithm aims to determine the optimum line or decision boundary that can divide n-dimensional space into classes. The name for this ideal decision boundary is a hyperplane. SVM designates the extreme vectors and points that aid in the creation of the hyperplane. SVM selects the extreme vectors and points that aid in the creation of the hyperplane. Support vectors are used in these extreme situations.

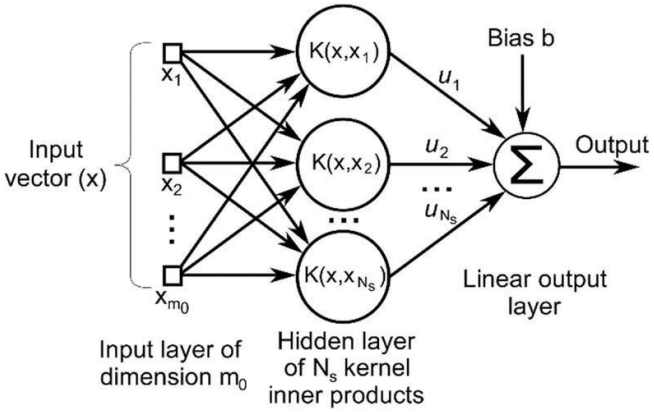

Fig -2: Architecture Diagram of Modified SVM

5. RESULTS AND ANALYSIS

Initially a correlation matrix was constructed as part of the pre-processing to check the dependency of the target variable on each of the independent carriable.

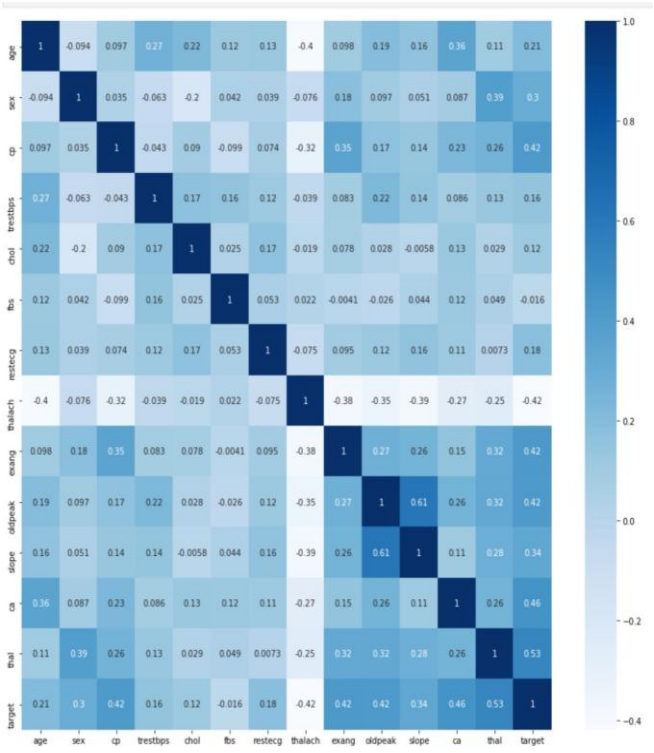

Fig-3: Correlation Matrix

In this work we have also conducted an analysis of various supervised ML algorithms to understand and show that our algorithm- Modified SVM gives a higher accuracy when compared to the other algorithms.

The suggested machine learning models, Random Forest, Gradient Boosting, Ada Boost, SVM and Modified SVM are trained. After the models have been trained, the results are examined to identify which model has the highest accuracy.

```
RandomForestClassifier

classification_report :
              precision    recall  f1-score   support

           1       0.76      0.73      0.75        30
           2       0.68      0.71      0.69        24

    accuracy                           0.72        54
   macro avg       0.72      0.72      0.72        54
weighted avg       0.72      0.72      0.72        54


confusion_matrix :
[[22  8]
 [ 7 17]]

-----
```

Fig -4: Accuracy of Random Forest Classifier

```
AdaBoostClassifier

classification_report :
              precision    recall  f1-score   support

           1       0.75      0.80      0.77        30
           2       0.73      0.67      0.70        24

    accuracy                           0.74        54
   macro avg       0.74      0.73      0.73        54
weighted avg       0.74      0.74      0.74        54


confusion_matrix :
[[24  6]
 [ 8 16]]

-----

GradientBoostingClassifier

classification_report :
              precision    recall  f1-score   support

           1       0.79      0.73      0.76        30
           2       0.69      0.75      0.72        24

    accuracy                           0.74        54
   macro avg       0.74      0.74      0.74        54
weighted avg       0.74      0.74      0.74        54


confusion_matrix :
[[22  8]
 [ 6 18]]

-----
```

Fig -5: Accuracy of Gradient Boosting

```
SVC

classification_report :
              precision    recall  f1-score   support

           1       0.71      0.90      0.79        30
           2       0.81      0.54      0.65        24

    accuracy                           0.74        54
   macro avg       0.76      0.72      0.72        54
weighted avg       0.76      0.74      0.73        54


confusion_matrix :
[[27  3]
 [11 13]]

-----
```

Fig -6: Accuracy of SVM

```
-----REPORT-------
Modified SVM
Accuracy on Custom Kernel: 0.8481481481481481
The Prediction is:
[2 2 1 2 1 1 2 2 2 2]
```

Fig -7: Accuracy of modified SVM

6. CONCLUSION AND FUTURE SCOPE

The fundamental measures in preventing the disease are the early identification and diagnosis of heart disease using ML approaches with the least amount of expense and problems. The introduction of DL algorithms will soon make it possible to diagnose this condition more quickly, affordably, and safely. On conduction of the experiment, it was realized that modified kernel of SVM showed better results than the traditional models. It showed better accuracy by 8.8%. The developed model could be used for initial diagnosis that could indicate to a user the necessity to undergo more reliable tests. We intend to train the models on more features in the future in order to enhance their performance. If a person is diagnosed, we would also discover, that characteristic which contributes to the heart condition. Finally, it can be said that ML andDL methods were crucial for the early detection, classification, screening, and prediction of CVD.

REFERENCES

[1] Amin MS. Identifying significant features and data mining techniques in predicting cardiovascular disease; 2018.

[2] Javier RodriÃÅguez, Signed Prieto, Leonardo Juan RamiÃÅrez Lo ÃÅpez, A novel heart rate attractor for the prediction of cardiovascular disease, Informatics in Medicine Unlocked, vol. 15, 2019.

[3] T. Vivekanandan, N. Ch Sriman Narayana Iyengar, Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease, Comput. Biol. Med. 90 (2017) 125–136.

[4] Fitriyani NL, Syafrudin M, Alfian G, Rhee J. HDPM: an effective heart disease prediction model for a clinical decision support system. IEEE Access. 2020; 8:133034–50.

[5] Ibrahim SS, Sivabalakrishnan M. An evolutionary memetic weighted associative classification algorithm for heart disease prediction. In Recent Advances on Memetic Algorithms and its Applications in Image Processing (pp. 183–199). Springer, Singapore; 2020.

[6] Oh, Yujin; Park, Sangjoon; Ye, Jong Chul (2020). "Deep Learning based Heart Diseaseprediction Features on CXR using Limited Training Data Sets". IEEE Transactions on Medical Imaging, (), 1–1. doi:10.1109/TMI.2020.299329.

[7] Chauhan A, Jain A, Sharma P, Deep V. Heart disease prediction using evolutionary rule learning, in 2018 4th International conference on computational intelligence & communication technology (CICT) (pp. 1–4). IEEE; 2018.

[8] S. Maheswari and R. Pitchai, "Heart disease prediction system using decision tree and naive bayes algorithm," Current Medical Imaging 15, 712–717 (2019).

[9] M. Jabbar, "Prediction of heart disease using k-nearest neighbor and particle swarm optimization," (2017).

[10] A. T. Azar, H. I. Elshazly, A. E. Hassanien, and A. M. Elkorany, "A random forest classifier for lymph diseases," Computer methods and programs in biomedicine 113, 465– 473 (2014).