ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

Internet of Things-based Hybrid Machine Learning Technique for Heart Disease Prediction

Dr. D.V. Chandra Sekhar,

Professor, PG Dept of Computer Science, TJPS College, Guntur,

dvchand.info@gmail.com

N. SreeRam

Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, India, sriramnimmagadda@gmail.com

Abstract.

It is very vital and alarming to be able to predict heart diseases before it manifest, as the number of people diagnosed with cardiac diseases is rising at an exponential rate day by day. Since determining this diagnosis is a challenging process, it is essential that it be carried out in an accurate and timely manner. The proposed study effort centres mostly on the issue of which patients are more prone to suffer from heart disease depending on a variety of different medical parameters. By using patient's past medical history, a method called the Hybrid Machine Learning Technique (HMLT) that is based on the Internet of Things (IOT) has been presented as an approach for determining whether or not a patient is likely to be identified with a cardiovascular disorders. The performance of the proposed HMLT is compared to that of a number of traditional methods, including Decision Tree (DT), Support Vector Machine (SVM), Naive Bayes (NB), X-GBOOST, Artificial Neural Network (ANN) and Random Forest (RF). The results of the experiments demonstrate that the effectiveness of the HMLT-based cardiac disease prediction system that has been developed is superior to that of other techniques. This conclusion was reached as a result of the findings of the studies. The newly developed methodology indicates that HMLT has a performance accuracy of 96%, which is superior to that of the traditional classification algorithms that are presently in use.

Keywords: heart disease, Hybrid Machine learning technique, Internet of Things, SVM, and Random Forest

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

1. Introduction

Ischemic heart disease was found to be the leading cause of death worldwide in 2016, with stroke following close behind as the second leading cause of death, according to research conducted by the World Health Organization. It is generally accepted that people should have access to sufficient health care services so they can have annual checkups. Nearly 31% of all deaths occur from cardiovascular disease [14], making it the top cause of death worldwide. Early identification [15, 16, 17] and therapy of a variety of various cardiac ailments can be particularly challenging due to a lack of diagnostic facilities, experienced medical experts, and other factors that influence the precise prognosis of heart disease. Developing countries are particularly affected by this. In response to this problem, a medical aid programmed has been developed to serve as a diagnostic support system for heart disease using cutting-edge computer technology and machine learning approaches.

The likelihood of dying from a heart-related condition is drastically decreased if treatment begins at an early stage. In order to understand the pattern of the data and make a prediction from it, multiple machine learning algorithms are used to medical data. When it comes to healthcare, the data required are often massive and complex. The capacity of ML methods to handle and analyse massive amounts of data for useful information is a major benefit [18, 19]. Machine learning algorithms learn from the past and make predictions based on the present. By using machine learning to better forecast heart problems, cardiologists may be encouraged to intervene more quickly, which could save many lives. This is due to the increased accessibility of drugs, which allows more people to become high in less time. Machine learning, a branch of AI research [2] that has exploded in popularity in recent years, is currently one of the most talked-about facets of data science. Machine learning is based on the idea that one set of algorithms may be utilised for a wide range of applications, from prediction and classification to decision making and beyond. ML models were put through their paces by being evaluated and validated on a series of new, real-time test datasets. The final efficiency of the model is then compared to the actual number to verify the overall accuracy of the predicted outcome.

Therefore, it is of paramount importance to establish a fast and accurate method of diagnosing cardiovascular illness. Numerous efforts have been undertaken to predict heart illness using machine learning algorithms [20, 21, 22, 23], however the research reported

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

here compares IoT healthcare datasets to widely used ML algorithms to identify the most precise ML approach.

The planned study is being conducted to determine, with the help of a patient's medical records, which detects persons who are most likely to be diagnosed with a heart disease [6], since cardiovascular disease is the primary cause of mortality in adults. The proposed HMLT can help in detecting the illness with fewer diagnostic exams and effective care, so that patients can be cured effectively. This is because it determines people who are experiencing any heart disease-related symptoms, such as chest tightness or high blood pressure. Symptoms like chest discomfort and elevated blood pressure are examples.

The goal of this research is to see if there is a correlation between the patient's medical history (gender, age, heart palpitations, fasting insulin levels, etc.) and the presence or absence of cardiac heart disease. The results of this investigation will be applied to follow-up studies. The sensor data is collected and added to a healthcare dataset based on the Internet of Things. The severity of the patient's heart ailment can be assessed with the use of this dataset. Patients were categorised based on 13 medical factors to evaluate their risk for developing heart disease. The attributes of the dataset are trained using a variety of machine learning techniques, including Decision Tree, Support Vector Machine, Naive Bayes, X-GBOOST, Artificial Neural Network, and Random Forest.

When compared to other classifiers, the HMLT approach's 96% accuracy rate stands out as the highest. Finally, this proposed treatment is fully cost effective, and it classifies people according to whether or not they are at risk for getting heart disease. There are five main parts to this paper's structure.

2. Proposed work

Men and women have the same chance of dying from heart disease, which is the main cause of death. Based on the information we have access to; heart disease is the cause of about one third of all deaths in the world. There has been a lot of study done to try and make it easier to tell if someone has a heart disorder. Professionals in the healthcare field have a hard time identifying diseases when there is a lot of data. Both Data Mining and Machine Learning take huge amounts of data and turn them into data that can be used to make good guesses and decide what to do next. Most of the time, detailed information is needed to make

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

an estimate about heart disease because it is so complicated. Because of this, experts are feeling better about the outlook for diseases, especially heart diseases. There was progress made on a strategy for predicting heart disease in this study. The framework, which uses 13 features of the standard UCI heart disease dataset, is meant to help doctors find heart disease early. The suggested method's steps are broken down into several sections, which are each talked about below. Look at figure 1, which shows a flowchart of a suggested model for predicting heart disease

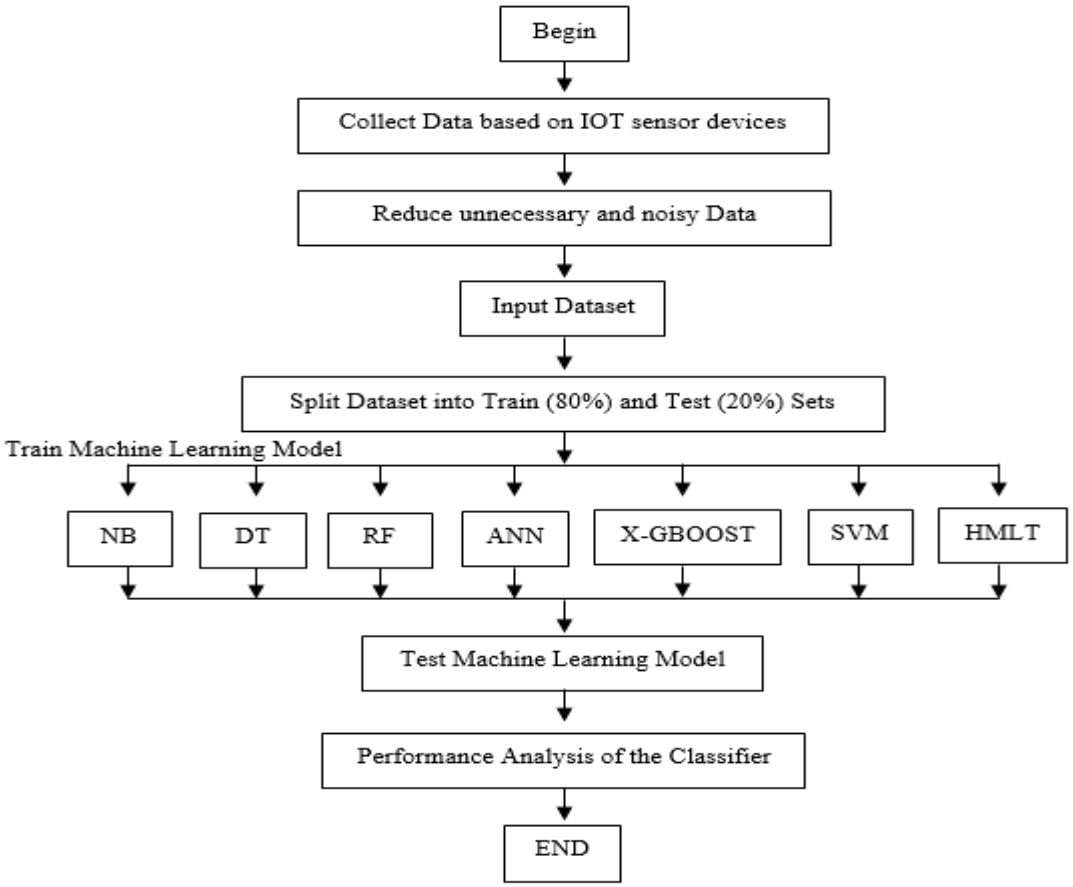

Figure 1: Flow Chart of Proposed framework for predicting heart disease

Collection of Information:

Using Internet of Things gadgets that can be worn or implanted, it is possible to get patient data from people who are in different places. IoT devices that are implanted into people, data

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

from benchmark datasets, and health records are all used to collect data in this first step. In order to make predictions about heart disease, the information in the 304 records that make up the normal UCI heart disease dataset is used. The following table 1 lists the dataset's properties.

Table 1: Data set Description with attributes

1	Age (in years)	Continuous
2	Gender (male or female)	1- Male,0- Female
3	Trestbps (resting blood pressure)	Continuous value
4	Cp (type of chest pain)	Chest pain type, values are taken between 1 to 4 1- typical angina 2- atypical angina 3- non-anginal pain 4- asymptomatic
5	fbs (fasting blood sugar)	Measured in mg/dL <=100 - normal 110 to 125 pre- diabetes >=126 - diabetes
6	restecg (ECG result)	0-normal 1-having ST- T 2-hypertrophy
7	thal (Heart rate of patient)	It takes following values 3-normal 6- fixed defect 7-reversible defect
8	chol (serum cholesterol)	Continuous value
9	exang (exercised induced angina)	0 - no 1 - yes
10	thalach (max heartbeat rate)	Continuous value
11	Oldpeak (ST depression)	Continuous value
12	Ca (major vessels number colored by fluoroscopy)	Number of major vessels from 0-3
13	Slope (peak slope)	Takes values from 1 to 3

Pre-Processing Data:

After getting a lot of different records, the heart disease statistics are first pre-processed. There were 300 records for patients in the dataset, but 5 of them are missing numbers. These 5 records were taken out of the collection, and the pre-processing was done on the 257 patient records that were left. As part of the collection, a multi-class variable and a binary classifier were both used. It is necessary to use the multi-class variable to check for the presence or lack of heart disease. This number is set to 1 if the patient is found to have heart disease and to 0 if the patient is found not to have heart disease. This means that the patient does not have heart disease. As a very important part of pre-processing data, medical information must be turned into diagnostic values. After the first round of handling the data for 295 patient records, 130 of them showed a value of 1, which means the patient had heart disease. The other 165 records showed a value of 0, which meant the patient did not have heart disease.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

Twelve of the thirteen traits that make up the data set are used to figure out personal information about each patient. These are the patient's age and gender. The last 12 traits are important because they hold important medical records. If you want to diagnose heart disease and figure out how bad it is, you have to have clinical records. DT, NB, SVM, ANN, RF, KNN, Xg-Boost, and HMLT are just some of the machine learning (ML) methods used in this study. All 14 attributes and all four ML methods were used in the experiment.

A lot of different classification models that come from machine learning are used in this study. The given information is used to figure out how accurate each model is, and then that result is compared to the HMLT

3. Results and Discussion

In order to find out if the patient has heart disease, experiments are done. A patient's age, gender, chol, restecg, thal, trestbps, fps, cp, exang, slope, ca, oldpeak, and thalach readings are some of the things that these studies look at. Therefore, a dataset with 300 records each characterized by 14 different characteristics is trained so that this experiment can be done. Figure 2 is a picture of a graph that shows how heart problems affect people with different types of traits. If someone is under a lot of stress at work and doesn't eat well, they are likely to develop heart syndrome quickly and could have a major coronary event.

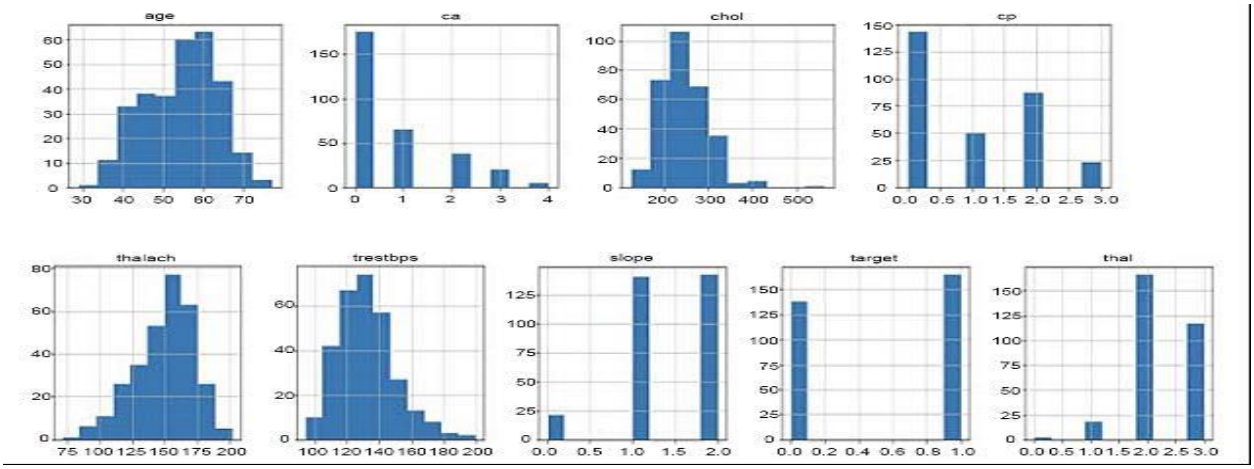

Figure. 2. Histogram of the distribution of target on features of dataset

On the same data set, seven different machine learning classification algorithms were tested, and the results of those tests were compared with those produced by the proposed HMLT as shown in table 2.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

Table 2: Performance measurements of seven ML classifiers and proposed HMLT

Classifier	Accuracy	Precision	Recall	F-Measure
RF	0.93	0.95	0.97	0.98
SVM	0.97	0.98	0.99	0.99
X-GBOOST	0.96	0.97	0.98	0.99
NB	0.85	0.90	0.97	0.93
ANN	0.95	0.97	0.98	0.98
HMLT	0.97	0.99	0.98	0.98
DT	0.93	0.96	0.97	0.97

Look at Figure 3, which shows how well different traditional classifiers and the suggested HMLT did in tests

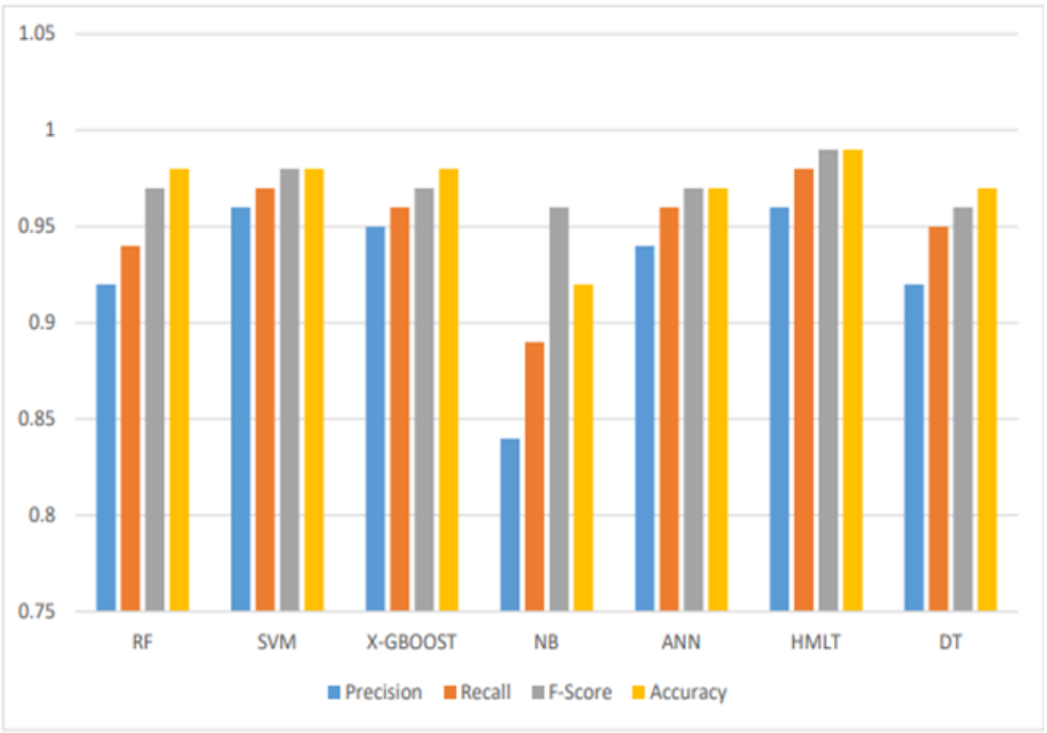

Figure. 3: Performance measurements of seven ML classifiers and proposed HMLT

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

As you can see from Figure 3, the proposed hybrid machine learning method works better than the other standard techniques.The graph in Figure 4 shows the HMLT's receiver operating characteristic (ROC), which is used to guess if someone will get heart disease

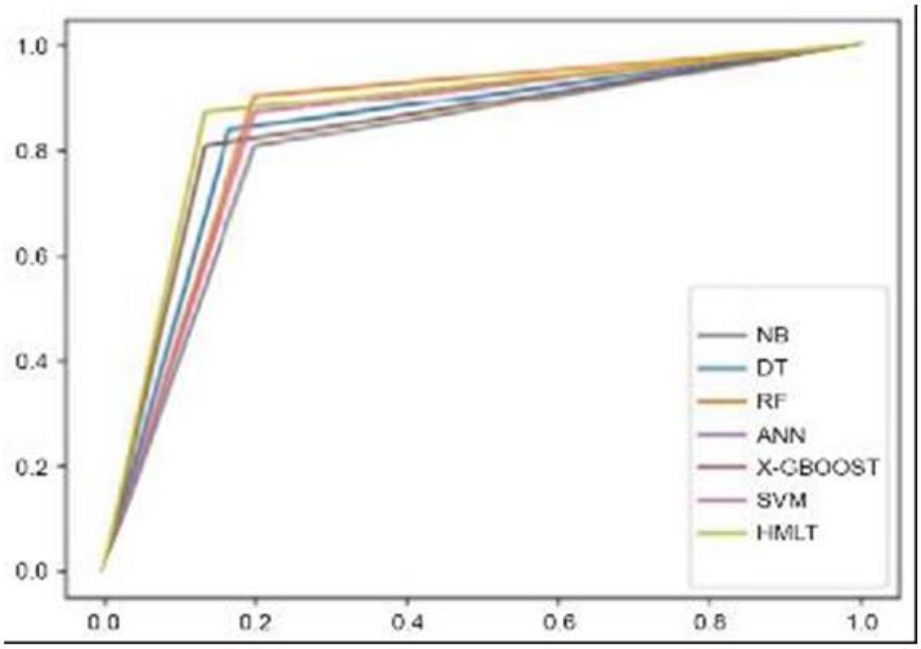

Figure. 4. ROC curve of seven algorithms in predicting heart disease

The ROC curve shows how precision and sensitivity are related to each other. The X-axis's abscissa is used to measure specificity. The closer it is to zero, the more accurate it is. Along the Y-axis of the ordinate, sensitivity is recorded. The accuracy is better when it is closer to one. Figure 4 shows that, out of the seven methods, HMLT makes the most accurate predictions.As a result, the data were looked at using seven different traditional classifications along with the suggested HML method. In this study, different evaluation metrics are used to compare the suggested HMLT's performance to that of 7 traditional classifiers. The suggested HMLT has a higher classification accuracy (96% vs. 89%) than the other models.

4. Conclusions

The suggested mixed machine learning method is more accurate at classifying things (96% of the time) than Decision Tree, Support Vector Machine, Naive Bayes, X-GBOOST,

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

Artificial Neural Network, and Random Forest. In the future, when this study is picked up again, it might use more combinations of different machine learning methods to get better predictions. This shows that there is a lot of room for improvement in how ensemble learning is used in medicine. On the other hand, picking out people who might get heart disease can be made more accurate by coming up with new ways to pick out features that help us learn more about important traits.

References

[1] S. Ouyang, "Research of Heart Disease Prediction Based on Machine Learning," 2019 5th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), Wuhan, China, 2022, pp. 315-319, doi: 0.1109/AEMCSE55572.2022.00071.

[2] K. S. K. Reddy and K. V. Kanimozhi, "Novel Intelligent Model for Heart Disease Prediction using Dynamic KNN (DKNN) with improved accuracy over SVM," 2022 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 2018, pp. 1-5, doi: 10.1109/ICBATS54253.2022.9758996.

[3] V. Kannagi, M. Rajkumar, I. Chandra, K. Sangeethalakshmi and V. Mohanavel, "Logical Mining Assisted Heart Disease Prediction Scheme in Association with Deep Learning Principles," 2016 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2022, pp. 1409-1415, doi: 10.1109/ICEARS53579.2022.9751820.

[4] A. Basak, M. S. Rahman and M. Rahman, "Prediction of Heart Disease Using an Approach Based on Machine Learning," 2015 13th International Conference on Computing Communication and Networking Technologies International Journal of Intelligent Systems and Applications in Engineering IJISAE, 2023, 11(5s), 539–549 | 549 (ICCCNT), Kharagpur, India, 2022, pp. 1-5, doi: 10.1109/ICCCNT54827.2022.9984555.

[5] G. Kumar Sahoo, K. Kanike, S. K. Das and P. Singh, "Machine Learning-Based Heart Disease Prediction: A Study for Home Personalized Care," 2014 IEEE 32nd

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

International Workshop on Machine Learning for Signal Processing (MLSP), Xi'an, China, 2014, pp. 01-06, doi: 10.1109/MLSP55214.2022.9943373.

[6] Z. Lin, Y. Wang, Z. Xiao and L. Xie, "A Survey of Heart Disease Risk Computer Intelligent Prediction Model," 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 2015, pp. 1433-1436, doi: 10.1109/IPEC54454.2022.9777302.

[7] T. Kavitha, S. Hemalatha, S. Mownika, G. M. Priya and J. K. Vigashini, "Review on Cardio Vascular Disease Prediction Using Machine Learning," 2012 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2012, pp. 1-4, doi: 10.1109/ICCCI54379.2022.9740953.