International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-09, Issue-01, Apr 2023

Comparative Study to Identify the Heart Disease Using Machine Learning Algorithm

[1]Prof. Vishal Shinde, [2]Mr. Kailash Poonaram Choudhary, [3]Mr. Mayur Sanjay Adhalage, [4]Mr. Bhavin Harish Patel.

[1]Asst.Professor,[2,3,4]UG Student,[1,2,3,4]Computer Engg. Dept. Shivajirao S. Jondhle College of Engineering & Technology, Asangaon, Maharashtra, India.

[1]mailme.vishalshinde@gmail.com, [2]kailashchaudhary523@gmail.com, [3]mayuradhalage@gmail.com, [4]patelbhavin16523@gmail.com

Abstract - Every year, this sickness affects many people, first in the USA and then in India. Heart disease, according to medical experts and academic studies is not an illness that appears out of nowhere rather, it develops over time as a result of an irregular lifestyle and a variety of physiological functions. when those symptoms start to show, people seek medical attention and undergo a variety of pricey tests and treatments. Therefore, prior to developing this sickness, people can gain knowledge about the patient condition from the findings of this research. This study gathered data from various sources and divided it into two groups, with 80% going to the training dataset and the remaining 20% going to the test dataset. It was attempted to improve accuracy using various classifier algorithms and then to summarize that accuracy. These methods include Nave Bayes, Logistic Regression, K-nearest Neighbor, Support Vector Machine, Random Forest Classifier, Decision Tree Classifier, and Support Vector Machine. SVM, Logistic Regression, and KNN provided accuracy comparable to or greater than other techniques. This essay makes a development about which factor, given a fundamental prefix such as sex, glucose, blood pressure, heart rate, etc., is vulnerable to heart disease.

Keywords- Machine Learning, Decision Tree Classifier, Random Forest Classifier, SVM, Logistic Regression, K-Nearest Neighbors, Big Data.

I. INTRODUCTION

Our body's main organ after the heart is the brain. It is therefore a highly important component of our body and its major duty is to supply blood to every area of the body. However, there are a few potential causes for failure. Various forms of cardiac disease exist in the world, although cardiac failure cad are the most frequent causes. The main cause of coronary artery disease is coronary artery blockage or narrowing [4]. More than 26 million individuals worldwide have heart disease which is growing by around 2 each year due to cad, making it the major cause of heart disease-related deaths. In 2005 there were 17.5 million fatalities worldwide2.

Medical science divides health risks into two major categories: those that are unchangeable and those that are changeable. Age, family history, sex, etc. are unchangeable factors.

Smoking, eating habits, physical activity, blood pressure (BP), cholesterol, etc. are all factors. Heart disease is one of the most serious problems in the world today. It requires diagnosis and prognosis, and there are several ways to do it. The common but somewhat pricey method is angiography. This paper can therefore employ a machine learning model to anticipate the patient's status rather than the conventional approach. Over the past year, machine learning gives us this advancement to predict the heart disease condition for a medical data set.

II. AIMS AND OBJECTIVE

a) Aim

This study intends to compare various algorithms that are used to predict heart problems. The findings of this study will aid in the development of knowledge about current approaches employed for heart disease prediction models. In this research, key data mining approaches that can be utilized to create extremely precise and effective prediction models are analyzed. These models will aid physicians in lowering the number of heart disease-related deaths.

b) Objective

- Information collection from Kaggle can be done without pre-handling systems allowing for prediction and decision making at different levels.

 DOI : 10.5281/zenodo.7863649

- Applying categorization, correctness, sensitivities, and particular analyses, this research proposed a predicting model to detect or diagnose heart disease.
- The effectiveness of various methods is compared in order to find a model that accurately predicts cardiac disease.

III. LITERATURE SURVEY

Paper 1: Machine Learning and End-to-End Deep Learning for the Detection of Chronic Heart Failure from Heart Sounds.

In this paper provide a technique for detecting CHF based on heart sounds. The technique integrates traditional machine learning (ml) and full-scale deep learning (dl). 947 patient's recordings from six publicly available datasets and one CHF dataset used in this study's evaluation of the approach. The proposed technique received a score of 89.3 using the same assessment criteria as a recent physonet challenge, which is 9.1 points better than the challenges baseline method. Finally, with an accuracy of 93.2%, this paper discovered 15 expert features that are helpful for creating ml models to distinguish between CHF phases (i.e., during the decompensated phase during hospitalization and in the recompensated phase). The suggested method yields encouraging results for differentiating recordings between healthy participants and patients as well as for the identification of various CHF phases [2].

Paper 2: An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection

In the world, heart failure is one of the main causes of death. Heart failure diagnosis is a difficult process, particularly in undeveloped and underdeveloped nations where there are few human expertise and resources. Therefore, a variety of sophisticated systems have been created by researchers for the automatic identification of heart failure. While recently published methods improved heart failure detection accuracy on testing data while sacrificing heart failure detection accuracy on training data, most of these algorithms are still plagued by the overfitting problem. As a result, the built models overfit the testing data. In this research, this paper constructs a novel diagnostic method in order to create an intelligent system that would perform well on both training and testing data [5].

Paper 3: An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure:

For just very early recognition of HF and to aid cardiologists in improving the diagnosis process, researchers has produced a variety of machine training models over the years. Inside this research, this paper present an expert system for such accurate prediction of HF that combines two support vector machine (SVM) models. Linear and L-1 regularized describe the first SVM model. A L2 regularized second SVM model is used. The Matthews correlation coefficient (MCC), ROC charts, accuracy, sensitivity, specificity, and area under the curve are six evaluation metrics that are used to assess the success of the suggested approach (AUC). The experimental results show that the suggested strategy enhances a traditional SVM model's performance by 3.3%. Additionally, the suggested strategy outperforms the ten previously suggested methods, which had accuracy ranges between 57.85% and 91.83%. Additionally, the suggested approach outperforms existing cutting-edge machine learning ensemble models in terms of performance [6].

IV. EXISTING SYSTEM

Diabetic, elevated blood pressure, more stress, triglycerides, obesity, ageing, and genetic predisposition are the main causes of heart disease. Only a very small percentage of systems employ the available clinical data for prediction, and even then, they are constrained by the numerous association rules that are in place. The doctor's clinical judgement and the patient's medical history are the only factors used in the diagnosis. Researchers in the system in place discovered that people are unable to organize vast amounts of data in an orderly or systematic fashion [1].

V. COMPARATIVE STUDY

Sr. No.	Author	Project Title	Publication	Technology	Purpose
1.	Awaise Niamit, Javed Ali Khan	An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure.	IEEE,2019	Support Vector Machine	This model eliminates noisy and irrelevant features while also used as a predictive model.
2.	Ashir Javed, shijie Zhou, Liao Youngjian	An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection	IEEE,2019	Random Search Algorithm	The proposed diagnostic system uses random search algorithm (RSA) for features selection and random forest model for heart failure prediction.
3.	Martin Gjoreski, Anton Gradisek, Borut Budna	Machine Learning and End-to-End Deep Learning for the Detection of Chronic Heart Failure from Heart Sounds.	IEEE,2020	Machine Learning & End-to-End Deep Learning	The method combines classic Machine-Learning and end-to-end Deep Learning and is used for detection of CHF phases for CHF Detection based on heart sounds.

Table no.1: Comparative Analysis

 DOI : 10.5281/zenodo.7863649

VI. PROBLEM STATEMENT

Finding a cure for heart disease is a huge task there are tools that can predict heart problem but they are either expensive or ineffective at estimating the likelihood of heart disease in humans the death rate and total consequences can be reduced by heart illnesses that are detected early a doctor cannot confer with a patient for a whole 24 hours because it would take too much time patience and knowledge however it is not always practicable to adequately monitor patients throughout the day in the modern world this paper have a lot of data so we can use a variety of machine learning algorithms to examine the data and look for hidden patterns in medical data hidden patterns can be exploited to diagnose health issues.

VII. PROPOSED SYSTEM

This paper used a heart illness dataset that downloaded from the university of California Irvine UCI repository in this study the data were gathered from various sources and divided into two groups 80 for the training dataset and the remaining 20 for the test dataset a variety of classifier methods were used in an effort to improve accuracy and the accuracy was then summarized the remaining feature[7] in this dataset is the lone target value each of those traits has an age and a gender which are denoted by the numbers 0 and 1 for male and female respectively their history of smoking as well as their blood pressure glucose heart rate BMI and previous stock history some records are given a classification of 0 or 1 where 0 and 1 denote absence and presence respectively.

VIII. ALGORITHM

The general idea of working of proposed system algorithm as follow:

Step 1: Start
Step 2: User Input
Step 3:

```
path = settings.MEDIA_ROOT + "\\" + "heart1.csv"
    data = pd.read_csv(path)
    x = data.iloc[:, :-1].values
    # print(x)
    y = data.iloc[:, -1].values
    from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y,
test_size=0.20, random_state=0)
    from sklearn.ensemble import RandomForestClassifier
    model = RandomForestClassifier()
    # print('x-train:', x_train)
    test_set = [age, gender, cp, trestbps, chol, fbs, restecg,
thalach, exang, oldpeak, slope, ca, thal]
    print(f" natu kodi {test_set}")
    # print('y_train:', y_train)
    model.fit(x_train, y_train)
    y_pred = model.predict([test_set])
    print("=====",y_pred)
```

Step 4:

```
if y_pred[0] == 0:
    msg = "you'r health is Good"
else:
    msg = "you'r health is Bad"
return render(request, 'users/ML.html', {'msg': msg})
path = settings.MEDIA_ROOT + "\\" + "heart1.csv"
data = pd.read_csv(path)
x = data.iloc[:, :-1].values
# print(x)
y = data.iloc[:, -1].values
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y,
test_size=0.20, random_state=0)
```

Step 5.

```
from sklearn.ensemble import RandomForestClassifier
    model = RandomForestClassifier()
    # print('x-train:', x_train)
    test_set = [age, gender, cp, trestbps, chol, fbs, restecg,
thalach, exang, oldpeak, slope, ca, thal]
    print(f" natu kodi {test_set}")
    # print('y_train:', y_train)
    model.fit(x_train, y_train)
    y_pred = model.predict([test_set])
    print("=====",y_pred)
```

Step 6:

```
if y_pred[0] == 0:
    msg = "you'r health is Good"
else:
    msg = "you'r health is Bad"
return render(request, 'users/ML.html', {'msg': msg})
```

Step 7: Stop

IX. MATHEMATICAL MODEL

1. SVM

Support Vector Machines (SVMs) classify data by evaluating a hyperplane that raises the dividing line between classes in the training data. Hyperplane can be formulated as

$$f(x) = a^T x + c$$

Where, a = dimensional coefficient, c = offset

2. Logistic Regression

Logistic regression (LR) is typically used to assess the likelihood that an instance will fit a given class. Logistic regression describes the relationship between a categorical outcome and one or more categorical predictors using a logit transformation of the dependent variable, where the logit model predicts the logit of the dependent variable from the independent variable [3].

$$\hat{p} = h_\theta(x) = \sigma(x^T\theta)$$

 DOI : 10.5281/zenodo.7863649

Where,

$$\sigma(t)=\frac{1}{1+e^{-t}}$$

3. K Nearest Neighbor

K-Nearest Neighbor (KNN) was a method which looks at K dataset instances close to the observation. The method itself then evaluates the inspection variable y to be anticipated using the output. For calculating the distance of two observations, Euclidean distance is used, and the equation is as follows:

$$d(x_i,y_i)=\sqrt{(x_{i,1}-y_{i,1})^2+ \ldots. +(x_{i,m} - y_{i,m})^2}$$

4. Decision Tree

The decision-tree method recognizes various outcomes, especially event consequences, using a graph-based framework. model trees target variables may have a discrete range of values, but in trees, the leaves stand in for class designations, and the branches for feature links that imply class labels. The equation of entropy is provided underneath,

$$E=-\sum_{b=1}^{n} p_{ab}\log_2 p_{ab}$$

X. SYSTEM ARCHITECTURE

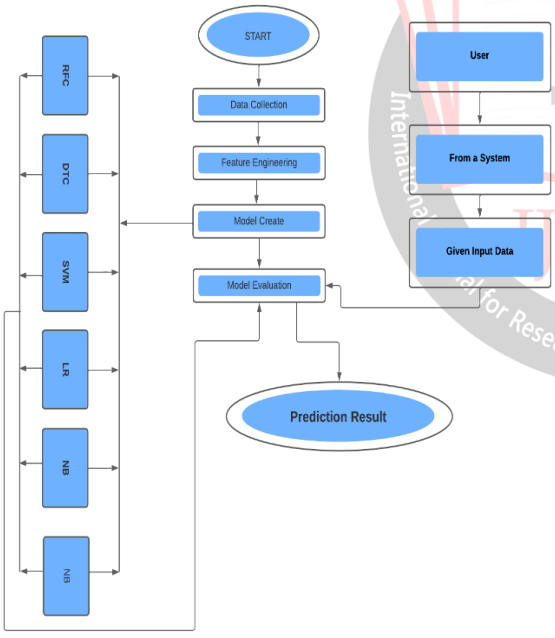

Fig.1: System Architecture

Description:

1)The user will enter the system by logging in.

2)User authentication will be done by administrator.

3) User will enter the details

4)The system uses trained model and three ml algorithms which are KNN, Logistic Regression, and SVM.

5) Heart Disease will be predicted as a final output.

XI. ADVANTAGES

It improves the exactness of cardiac illness diagnosis and handles massive amounts of information with the method of random forest and extraction of characteristics.

It automatically identifies major characteristics without a requirement for human intervention. Less computational power or resources, such as RAM, CPU, GPU, or TPU, for example. This paper can achieve greater accuracy with less data [8].

This aids in solving complicated real-life situations under a variety of constraints while reducing the time complexity for doctors and being cost-effective for patients.

Understanding move between one system onto another depending on tasks and domains. offers a route to achieve artificial general intelligence someday.

XII. DESIGN DETAILS

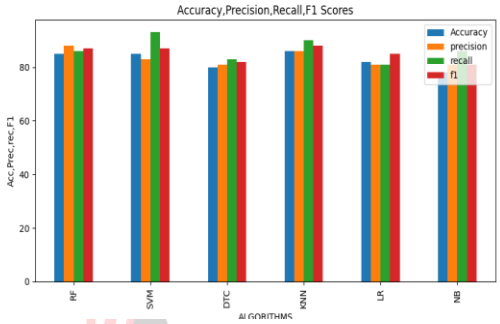

Fig 2: Result

The above graph shows the comparison between different algorithms like RF, SVM, DTC, KNN, LR, and NB to predict heart disease.

XIII. CONCLUSION

Thus we have tried to implement the paper "Comparative Study to Identify the Heart Disease Using Machine Learning Algorithms", Rezaul Karim; Wang Chengliang; Mohaiminul Islam, IEEE, 2021 and the conclusion as follow:

KNN, Logistic Regression, and SVM have provided superior accuracy than those of other algorithms, according to comparisons made between them and all other algorithms used in this research. The other four, of which two are close to having higher accuracy and Decision Tree Classifier having the lowest, all played minor roles in this research, but these three were crucial. Additionally, this research model provides improved accuracy, and by using it, users can conduct tasks from a system and can be informed about their present heart disease status. This paper may also contrast various hybrid methods and data mining techniques to improve the accuracy of the current results.

REFERENCE

[1] Rezaul Karim, Wang Chengliang, Mohaiminul Islam " Comparative Study to Identify the Heart Disease Using

 DOI : 10.5281/zenodo.7863649

Machine Learning Algorithms" Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS-2021) IEEE Xplore Part Number: CFP21OAB-ART; ISBN: 978-1-7281-9537-7AccessionNumber: 20513286DOI: 10.1109/ICBAIE52039.2021.9390032

[2] Martin Gjoreski; Anton Gradišek; Borut Budna; Matjaž Gams; Gregor Poglajen, "Machine Learning and End-to-End Deep Learning for the Detection of Chronic Heart Failure From Heart Sounds." IEEE Access, vol. 8, pp. 20313 - 20324, 2020, doi: 10.1109/ACCESS.2020.2968900

[3] Molnar, Christoph, "Interpretable machine learning. A Guide for Making Black Box Models Explainable", 2020

[4] L. Ali et al., "An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure," IEEE Access, vol. 7, pp. 54007-54014, 2019, doi: 10.1109/ACCESS. 2019.2909969.

[5] Ashir Javeed; Shijie Zhou; Liao Yongjian; Iqbal Qasim; Adeeb Noor; Redhwan Nour "An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection" IEEE Access, vol. 8, pp. 180235 - 180243, 2019, doi : 10.1109/ACCESS.2019.2952107

[6] Liaqat Ali; Awais Niamat; Javed Ali Khan; Noorbakhsh Amiri Golilarz; Xiong Xingzhong; Adeeb Noor; Redhwan Nour; Syed Ahmad Chan Bukhari, "An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure" IEEE Access, vol. 7, pp. 54007-54014, 2019, doi: 10.1109/ACCESS.2019.2909969

[7] Prof. Vishal R. Shinde, "Machine learning algorithm for stroke disease classification" in IJREAM, ISSN: 2454-9150, Volume 08, Issue 01, APR 2022 Special Issue

[8] Prof. Vishal R. Shinde, "Analysis and prediction of cardiovascular disease using machine learning classifiers" in IJREAM, ISSN: 2454-9150, Volume 08, Issue 01, APR 2022 Special Issue (indexed in scope database https://sdbindex.com/documents/00000217/00001-75837. pdf)

 DOI : 10.5281/zenodo.7863649