INTERNATIONAL JOURNAL OF RESEARCH IN MEDICAL SCIENCES & TECHNOLOGY

e-ISSN:2455-5134; p-ISSN: 2455-9059

EMPLOYMENT OF THE SELECTIVE MACHINE LEARNING ALGORITHM FOR THE EARLY AND EFFECTIVE DETECTION AND DIAGNOSIS OF CARDIOVASCULAR DISEASE

Shreya Bhardwaj
Panjab University, Chandigarh, India

Paper Received: 09th May, 2021; **Paper Accepted:** 09th June, 2021;
Paper Published: 30th June, 2021

DOI: http://doi.org/10.37648/ijrmst.v11i01.024

How to cite the article:

Shreya Bhardwaj, Employment of the Selective Machine Learning Algorithm for the Early and Effective Detection and Diagnosis of Cardiovascular Disease, IJRMST, January-June 2021, Vol 11, 239-246, DOI: http://doi.org/10.37648/ijrmst.v11i01.024

ABSTRACT

Heart and blood artery dysfunction is the root cause of cardiovascular disease, which includes coronary heart disease, cerebrovascular disease, peripheral arterial disease, rheumatic heart disease, deep vein thrombosis, and pulmonary embolism. A model for using a machine-learning algorithm to find cardiovascular diseases is presented in this paper. Utilized the agile methodology in this research, planning, requirements analysis, designing, coding, testing, and documentation are all carried out simultaneously throughout the stages of the production process. Using four distinct machine learning algorithms—a Support Vector Classifier, a K-Nearest Neighbors Classifier, a Random Forest Classifier, and a Decision Tree Classifier—the patient dataset is used to train the model in this paper. An algorithm will make the predictions, resulting in the most accurate results. Flask, a web-based implementation of this model, was used to make a prediction, the user must fill in 13 inputs on the web. Flask and the Python programming language are used to implement the model and the machine learning algorithms. We use a K-Nearest Neighbors Classifier algorithm after considering all four machine learning algorithms. The prediction has a good accuracy of 85.83 per cent, which is good for any model.

INTRODUCTION

Cardiovascular disease (CVD) accounts for 31% of all deaths worldwide, with coronary failure and stroke accounting for 85% of these deaths. More than 75% of CVD deaths occur in low- and middle-income countries. Of the 17 million accidental deaths (those under 70) caused by non-communicable diseases in 2015, 82% occurred in low- and middle-income countries, and 37% were caused.

The most significant risk factor for cardiovascular disease is elevated circulatory stress.

In 2015, the global prevalence of elevated circulatory strain (defined as systolic and diastolic pulses greater than or equal to 140/90 mmHg) in adults over 18 was approximately 24.1 per cent for men and 20.1 per cent for women. In 2015, there were 1.13 billion adults with high blood pressure, up from 594 million in 1975. One in four people dies from cardiovascular disease (CVD), as shown in the following graph. By 2030, cardiovascular diseases (CVDs) will have overtaken infectious diseases as the leading cause of death, according to VillaMedica's analysis.

People with cardiovascular disease or who are at high cardiovascular risk (due to the proximity of at least one risk factor, such as hypertension, diabetes, hyperlipidemia, or effectively cured disease) need early location and the executives utilizing guidance and medications, as appropriate. Most cardiovascular diseases can be prevented by tending to conduct risk factors, such as tobacco use, unhealthy eating habits, obesity, physical inactivity, and harmful alcohol use.

As we talked about, cardiovascular disease (CVD) is a group of diseases that can be caused by the heart and blood vessels. Among these diseases are many, such as coronary heart disease: cerebrovascular disease is a condition that affects the blood artery that supplies blood to the heart muscle. A condition that affects the blood supply to the brain, also known as the peripheral arterial disease: Rheumatic heart disease and disease of the blood artery that supplies blood to the arms and legs: streptococcal bacteria-caused rheumatic fever damage to the heart's tissues and valves.

Using a variety of machine-learning algorithms, this paper presents a model for detecting cardiovascular disease and concludes by determining which Python programming algorithm is superior for creating such a model.

PROPOSED WORK

Machine learning is the process of teaching machines how to perform any task, no matter how simple or difficult it is. The process of making machines think and perceive things like humans is known as machine learning.

Fig 1: Different types of machine learning algorithms

The figure above shows that machine learning is primarily broken down into three main categories:

1) Directed Instruction: Using labels and tags, machines are taught in supervised Learning; It is the process of having direct supervision involved, just like when a teacher directly teaches a class of students based on data that has already been collected. Two steps are involved in supervised Learning:

a) Fallback: Recognizing patterns and making guesses about continuous results is the process. The system must understand the number, values, grouping, etc.

b) Categorization: It manually instructs the algorithm to recognize specific types of objects and classify them appropriately after previous data specimens tag input data.

2) Learning Unsupervised: Unsupervised Learning is a method in which the developer does not direct the process. Unlike supervised Learning, where results are known, unsupervised machine learning algorithms do not know the desired outcomes.

Additionally, unsupervised Learning works with unlabelled data, unlike supervised Learning.

To determine whether a person has a disease, we use four supervised learning algorithms in this paper to evaluate which provides the best results. The model is then constructed using that algorithm.

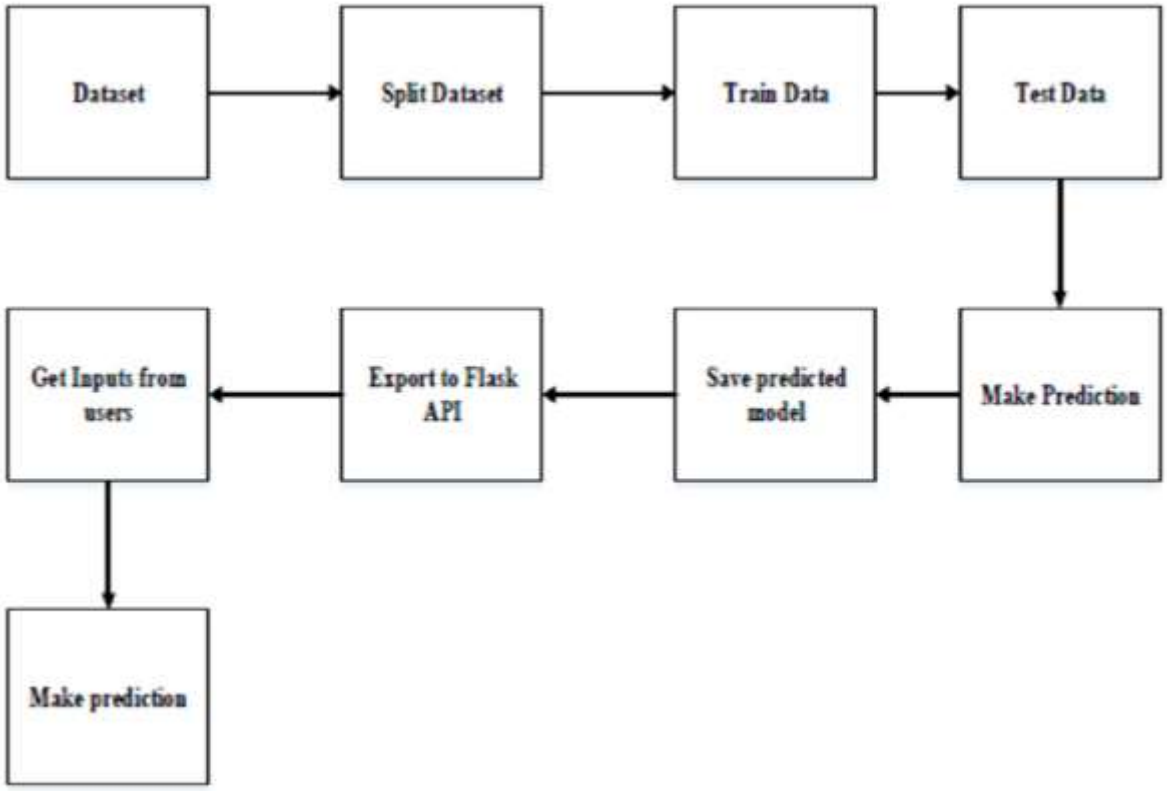

Fig 2: The architecture of the model for prediction

A patient dataset from kaggle.com is used in this system. Fourteen distinct attributes in this dataset represent the test results from 600 individuals. The dataset has a train set (80%) and a test set (20%). Following that, we trained our model using the following four machine learning algorithms: Decision Tree Classifier, Random Forest Classifier, and Support Vector Classifier from K-Nearest Neighbors. Based on the accuracy of the four algorithms used to train the model and the results of training, testing, and checking the predictions, we will now determine which algorithm the model will use to make predictions. Flask, an Application Programming Interface, saves and loads this trained model into the web. Flask loads this trained model and requests user test results. The trained model receives these results and processes them to determine whether a patient has a disease.

DATA ANALYSIS AND OUTPUT

An artificial intelligence (AI) model was developed to determine whether a client has coronary disease in this paper. This model uses a data set of thirteen different results from different people's pathological tests. There are no null values in this dataset, which is extremely clean. We split

our data into x and y variables by importing train_test_split from sklearn.model_selection into python programming. Separated the x and y factors from this dataset. The mixed test results and the 13 qualities are contained in the x variable, while the yield is contained in the y variable.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Fig 3: Description of the dataset with its all 14 attributes

The following outcomes are obtained when the data set is analyzed in light of various attributes, such as the probability that a disease will occur or not:

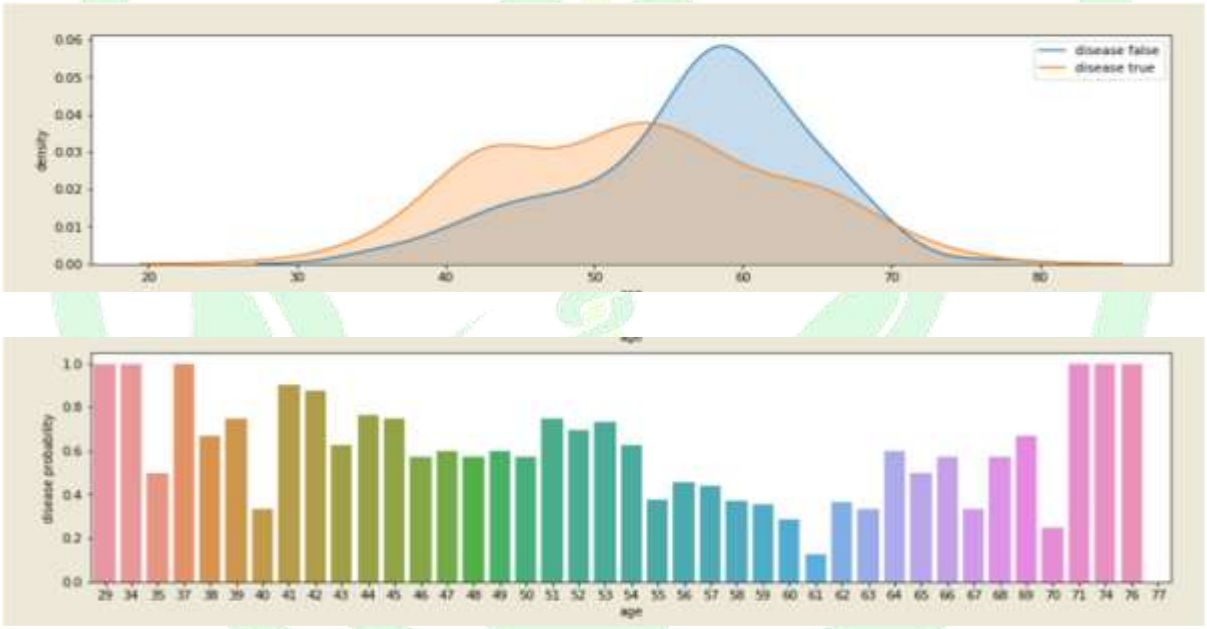

Fig 4: The probability of happening of disease according to the ages of different peoples

Standard Scaler was used to scale the x variable. In addition, the x and y variables were divided into x_train, x_test, y_train, and y_test. Four machine calculations—KNearest Neighbors, Support Vector Machine, Decision Tree, and Random Forest—were fitted or prepared for this x_train and y_train. Used the four calculations to check the level of precise results with different amounts of n esteems. When k = 12, the most important exact result for K neighbours is 84.49 per cent; when the number of estimators is equal to 10, the most notable precise result

for Support Vector Machine is 85.32 per cent. When n is 10, the highest possible outcome for the Decision Tree is 99.83 per cent. When the number of estimators is set to 20, the most notable exact outcome for the Random Forest is 99.83 per cent. After testing precision, we used Random Forest, which has one of the highest exact results for making expectations. Flask, an Application Programming Interface, was being used to save and stack the Random Forest into the web. We created an HTML page with 13 contributions using Flask. The clients will enter their various test results and send the contributions to the model to determine whether they have heart disease.

CONCLUSION

This paper, which proposed "A model to Detect Cardiovascular Disease Using AI Algorithm," utilized an AI strategy. The dataset, which includes the test results of various patients, was prepared and investigated using this AI strategy's four calculations. Additionally, I plotted a diagram using matplotlib to verify the accuracy of these calculations.

On the other hand, K Nearest Neighbors and Support Vector Machine have exact results of 84.49 per cent and 85.32 per cent, respectively. In contrast, Decision Tree and Random Forest have the highest exact results after testing for precision. Can use Deep Learning to bring this investigation into a consistent framework where customers can display their test results. Also integrated the choice tree model into the web via an API called Flask, which predicted excellent results when tested multiple times online without error.

REFERENCES

[1] https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] Nidhi Bhatla, Kiran Jyoti (2012), An Analysis of Heart Disease Prediction using Different Data Mining Techniques, *International Journal of Engineering Research & Technology (IJERT)*, (ISSN:2278-0181) Vol. 1 Issue 8, October – 2012

[3] Rajesh N, T Maneesha, Shaik Hafeez, Hari Krishna, (2018) Prediction of heart disease using machine learning algorithms, *International Journal of Computer Sciences and Engineering & Technology*, 7 (2.32) 363-36

[4] O.E. Taylor, P. S. Ezekiel, F.B. Deedam-Okuchaba, (2019), A Model to Detect Heart Disease using Machine Learning Algorithm, *International Journal of Computer Sciences and Engineering*

(ISSN: 2347-2693) Volume 7 Issue 11 Nov 2019.

[5] Yamla Sandhya, (2020), Prediction of Heart Diseases using Support Vector Machine, *International Journal for Research in Applied Science & Engineering Technology (IJRASET)* (ISSN: 2321-9653) Volume 8 Issue II Feb 2020.

[6] https://villa-medica.com/wp-content/uploads/2017/07/villa-medica-cardiovascular-infographic-2-compressor.jpg