How to Cite:
Karthiga, A. S., & Mary, M. S. (2022). A predictive analysis of heart diseases using machine learning algorithms. *International Journal of Health Sciences*, *6*(S3), 3108–3125. https://doi.org/10.53730/ijhs.v6nS3.6309

A predictive analysis of heart diseases using machine learning algorithms

A. Sankari Karthiga
Research Scholar, Department of Computer Science, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli, India

Dr. M. Safish Mary
Assistant Professor, Department of Computer Science, St.Xavier's College, Palayamkottai, Tirunelveli, India

***Abstract*---**So far human beings have passed many centuries apart from many illnesses. The human body is a prodigy created by nature. Among human organs, the very special thing is Human Heart. The development of science in the field of medicine has contributed much to predicting heart diseases. The very recent development technique for predicting heart ailments at the very early stage is data mining. In this audit, a machine learning algorithm is employed to predict the imminent level of heart disease. There are about fifteen drug criterions which are taken into account like age, sex, blood pressure, cholesterol, and obesity for alerting heart ailments. This is applied for prophesying the liability of the stoics who endures heart problems. These tactics are skilled in getting the overall health factors of the nursling especially all about their heart. In this study, K-Nearest Neighbour and Tuned K-NN are applied. The concluding data reveal that the pitfall (risk factor) regarding heart is skillfully guessed with the help of this set diagnostic system.

***Keywords*---**machine learning, cardiovascular disease, KNN, tuned K-NN, disease prediction.

Introduction

Due to the advancement of science, the medical field has attained a splendid development in treating the sick. The recent spreadsheet called data mining is a boon for prophesying the cardiovascular bug at an early stage itself. Using this tool, the eventual risk that the stoic has to face can be diagnosed previously by analyzing the patients' day-to-day plights. This operation is carried out in two ways. i.e. out of a preset class, a miniature is created and it is employed for the

International Journal of Health Sciences ISSN 2550-6978 E-ISSN 2550-696X © 2022.
***Corresponding author**: Karthiga, S. A.; Email: sankarikarthiga.sk@gmail.com*
Manuscript submitted: 18 Dec 2021, Manuscript revised: 27 March 2022, Accepted for publication: 09 April 2022

study of planned or obscure objects. The mechanism of the heart is affected by cardiovascular disease. The conventional class of cardiac arrest is known as the cardio pulmonary arrest. Under this discussion, we have the idea of data mining, the part played by data mining in medicine, the kind of coronary thrombosis, the incitation, and the intention of the developed classifiers for the normal type of angina pectoris known as myocardial infarction are reviewed in this discussion.

Nowadays medical data mining seems to be applied in a greater background. To deal with it, there are some features like incompleteness, incorrectness, and inexactness. Among the latest techniques employed in healthcare applications, data mining in the case of classification analysis provides better quality of medical diagnosis. In human ailment, diagnostic methods, already existing data sets are applied. The classification technique uses a drill sample which is a part of a subset of the so-called present data set. Regarding classification, several types of data mining practices are used. Using predictive analytics, one can encompass a number of polling factors from data mining, predictive modeling, and machine learning which involves current analysis and historical facts so that it can guess the eventual happenings or unknown events. Various approaches take part in predictive analytic statistical techniques like data modeling, machine learning, Artificial Intelligence, Deep learning algorithms, and data mining.

Any disease which results in the malfunctioning of the cardiovascular system is called CVD. The most stoic cardiovascular disorder which is seen all over the world is (Ischemic heart disease) aplastic anemia. Since it is associated with the circulation of blood to the heart muscle this condition is otherwise termed bloodlessness. If the block prevails over a period of time, that condition is referred to as a cardiovascular thrombosis or a cardiovascular stroke. Regarding the 200census, about the total 57, million deaths, 36 million lives were lost for the cause of non-curable diseases like cancer, diabetes, cardiovascular diseases, and respiratory diseases. In the year 2008, about 17 million deaths 48% of deaths are due to non-controllable diseases. The major cause stated by different statistical surveys is what we call aplastic anemia, its threat even happened in India too.

The latest software system is employed in medical diagnosis to help the cardiologists to save the life of the patient (stoics) as far as possible. The most panicking thing is that the stoics who do not show the sign of chest pain seem to be a case of CVD. This is a very crucial thing. The recent studies and research shows how critical it is to diagnose the problem and attempts should be made to speed up the detection of heart disorder through cyber solutions. This is purely the responsibility of the researchers in the medical field to come to a conclusion of applying very new tactics in treating heart diseases. Supposing if a new tool is established in treating heart diseases, the complication should be re-examined over and over to get into a better solution. The stoics (patient) with a high risk of CVD are considered to be odd and others are seen as normal. This research work aims at creating a CAD system that focuses on better neural web criteria evaluated with the help of developmental approaches for better exactness in detection, so that the misclassification rate is reduced.

Related Work

In the suggested paper [1], the authors have proposed that a novel machine learning approach is proposed to predict heart disease. The proposed study used the Cleveland heart disease dataset, and data mining techniques such as regression and classification are used. Machine learning techniques Random Forest and Decision Tree are applied. The novel technique of the machine learning model is designed. In implementation, three machine learning algorithms are used; they are Random Forest, Decision Tree and Hybrid model (Hybrid of random forest and decision tree).

The authors of this study [2], the presence of heart diseases was determined by using machine learning algorithms. In this study, the data obtained from the patients were weighted according to their effects on the success rate. Nowadays, one of the most important illnesses is heart disease which causes of mostly patients dead. Medical diagnosis of heart diseases is very difficult. While heart diseases are diagnosed medically, they can be confused with other diseases that show same symptoms such as chest pain, shortness of breath, palpitations and nausea. This makes it difficult to diagnose heart diseases medically. The authors demonstrated this article [3], we used both of the two machine learning models in the heart disease diagnosis problems. We implemented the algorithms, tuned the parameters, and conducted a series of experiments. We aim to compare the prediction accuracy of the two models under different parameters settings. We used the Cleveland database which is took from UCI learning dataset repository for diagnosis heart disease.

The goal of this study [4], was to use machine learning algorithms to classify patient-reported outcomes (PROs) using activity tracker data in a cohort of patients with stable ischemic heart disease (SIHD). A population of 182 patients with SIHD was monitored over a period of 12 weeks. Each subject received a Fitbit Charge 2 device to record daily activity data, and each subject completed eight Patient-Reported Outcomes Measurement Information Systems short form at the end of each week as a self-assessment of their health status. Two models were built to classify PRO scores using activity tracker data. The first model treated each week independently, whereas the second used a hidden Markov model (HMM) to take advantage of correlations between successive weeks.

In a related study [5], to improve the diagnosis accuracy of CHD based on the analysis of diastolic murmurs, empirical wavelet transform was applied to distinguish the difference of diastolic murmurs between in CHD and in valvular disease. Firstly, the spectrum of diastolic heart sounds was divided into three modes and the segmentation boundaries were set as [150, 200] Hz. Next, the diastolic modal spectrums were obtained. It is found that the essential difference of diastolic murmurs between in CHD and in mitral stenosis is the third modal spectrum. Finally, the characteristic parameter P3 was defined and used to distinguish the diastolic murmurs in CHD and in valvular disease.

The authors of the study [6], proposes an effective heart disease prediction model (HDPM) for a CDSS which consists of Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to detect and eliminate the outliers, a hybrid

Synthetic Minority Over-sampling Technique-Edited Nearest Neighbor (SMOTE-ENN) to balance the training data distribution and XGBoost to predict heart disease. Two publicly available datasets (Statlog and Cleveland) were used to build the model and compare the results with those of other models (naive bayes (NB), logistic regression (LR), multilayer perceptron (MLP), support vector machine (SVM), decision tree (DT), and random forest (RF)) and of previous study results. The purpose of this proposed novel [7], fast conditional mutual information feature selection algorithm to solve feature selection problem. The features selection algorithms are used for features selection to increase the classification accuracy and reduce the execution time of classification system. Furthermore, the leave one subject out cross-validation method has been used for learning the best practices of model assessment and for hyperparameter tuning.

The authors described in this research article [8], to identifying intravenous immunoglobulin-resistant patients is essential for the prompt and optimal treatment of Kawasaki disease, suggesting the need for effective risk assessment tools. Data-driven approaches have the potential to identify the high-risk individuals by capturing the complex patterns of real-world data. To enable clinically applicable prediction of intravenous immunoglobulin resistance addressing the incompleteness of clinical data and the lack of interpretability of machine learning models, a multi-stage method is developed by integrating data missing pattern mining and intelligible models. First, co-clustering is adopted to characterize the block-wise data missing patterns by simultaneously grouping the clinical features and patients to enable (a) group-based feature selection and missing data imputation and (b) patient subgroup-specific predictive models considering the availability of data. Second, feature selection is performed using the group Lasso to uncover group-specific risk factors. Third, the Explainable Boosting Machine, which is an interpretable learning method based on generalized additive models, is applied for the prediction of each patient subgroup.

In the suggested study [9], the factors leading to cardiovascular disease can be diagnosed and recursion enhanced random forest with an improved linear model (RFRF-ILM) to detect heart disease. This paper aims to find the key features of the prediction of cardiovascular diseases through the use of machine learning techniques. The prediction model is adding various combinations of features and various established methods of classification. it produces a better level of performance with precision through the heart disease prediction model. In a related article [10], the author developed a construction based on two distinct classifications: Decision Tree and Naive Bayes. In these two classifications, the Decision Tree gave more reliable results than Naive Bayes. The Decision Tree is 98.27 percent precise, while Naive Bayes is 89.90 percent precise. The authors described in their research paper [11] discusses the most popular classification methods for machine learning, including logistic regression, nearest k-neighborhoods, support vector machines, decision-making tree classification, forestry random classifications and XGBoost classification. SVM algorithms provide smart accurate results and good accuracy in comparison with the above algorithms in this study

Proposed Methodology

After the analysis of the so-called actual work, we come to a conclusion that the below-mentioned algorithms are effectual and apt for classification and prediction regarding exactness. The data given here are examined using the following algorithms.

- K-Nearest Neighbour Algorithm
- Tuned -K-Nearest Neighbour Algorithm

Pre-Processing

Usually, the authentic data are half-done, boisterous, and uncertain. The renewal of data which is otherwise known as codifying is widely employed in calibrating aspect coding to drop inside a specified range. The original or the earliest data undergo a linear alteration by means of min-max normalization within the limits between 0 and 1 the values should be normalized. The process of normalization which is otherwise known as synthesizing is a step involved in data pre-processing and the data attempts are normalized so that it can supply all attributes a balanced weight. Supposing, for predicting the normalization data values for every aspect, a neutral web along with a posterior breeding design valued in the drill tuples will be aided in speeding up the learning phase.

Classification

Classification has performed various functions such as fraud detection, spot marketing, predicting performance, assembling, and medical diagnosis. The term which is called data classification involves two steps, the first thing is a learning step which concentrated on the bricklaying of bracket model and the second thing is the classification step and this model is used for predicting the class labels for the data given. We can say that the percentage of the test set tuples that are duly assorted by the classifier is none other than the delicacy of a classifier on a given test set. The comparison is made between the bound class label of each test tuple and the well-read classifiers class prediction for that tuple. Supposing, if the exactness of the classifier is taken as an acceptable thing, the classifier, in this case, is employed in such a way that is used to assort eventual data tuples for an unknown class label. Classification in this sense is an archetype of data mining technique grounded with machine learning.

K-Nearest neighbour algorithm

An approach that assorts every account in a data set which relies on an amalgam of the classes of the k records uttermost alike to it in a historical dataset where k=1. Rarely termedas K-Nearest Neighbor technique. KNN – K-Nearest Neighbour algorithm functions as follows:

- Decide the value of k
  The initial thing is deciding the value of k. the setting of k value varies widely with respect to the case. In case of using the scikit-learn library, we can get five, which is supposed to be the default value.

- The distance in case of testing data along with training data are casted.
- Within the fresh data the next-door K-Neighbours are detected. Once the distance casted, then search for K- Neighbours which are nearest testing data. In case, if k=3 is applied, searches are made to look for training data which is nearest to the testing or new data.
- Now data class prediction
  In order to refer to a class of new data, there should be a selection of a class of drill data which finds a close relation with the new data and should possess the hignest quantity.
- Evaluation
  This is applied for calculating the exactness of the model, supposing, if the accuracy is flat, then there should be a repetition of this process from step 1.

Tuned K-nearest neighbor

A model parameter that is tuned as a hyperparameter is set before the learning process is initiated. The exhaustic grid search technique can be applicable for hiking. Many hyperparameters are wielded to get an exhaustive grid search for every solitary viable blending of the parameters and so many cross-validations are performed in a way we like. An excellent way of referring an ideal hyperparameter values for usage is nothing but an exhaustive grid search, but it possesses quick time-consuming technique with an addition of every additional parameter value and cross validation. There are three tuning parameters which are in usage like n-neighbours, weight and metric.

- N-neighbours: this is supposed to be the previous computed values which decide an idea one.
- Weights: Here, in this case, there must be a verification of weights as to whether it is advantageous or not no load is assigned with 'uniform', the points regarding 'distance' is weighted by the inverse of their distances which means better or dearer points will possess weight much higher than that of the yon (farther) points.
- Metric: The similarities will be counted by means of the distance metric which is yet to be used.

Tuning params = { 'n_neighbors' : [5,7,9,11,13,15],'weights' : ['uniform','distance'],
'metric' : ['minkowski','euclidean','manhattan']}

As extremely great model exactness is drawn or attained by us. If needed, it is possible to simplify the delicacy by employing additional parameters or we can go for an ideal model. The block diagram of the proposed work is below:

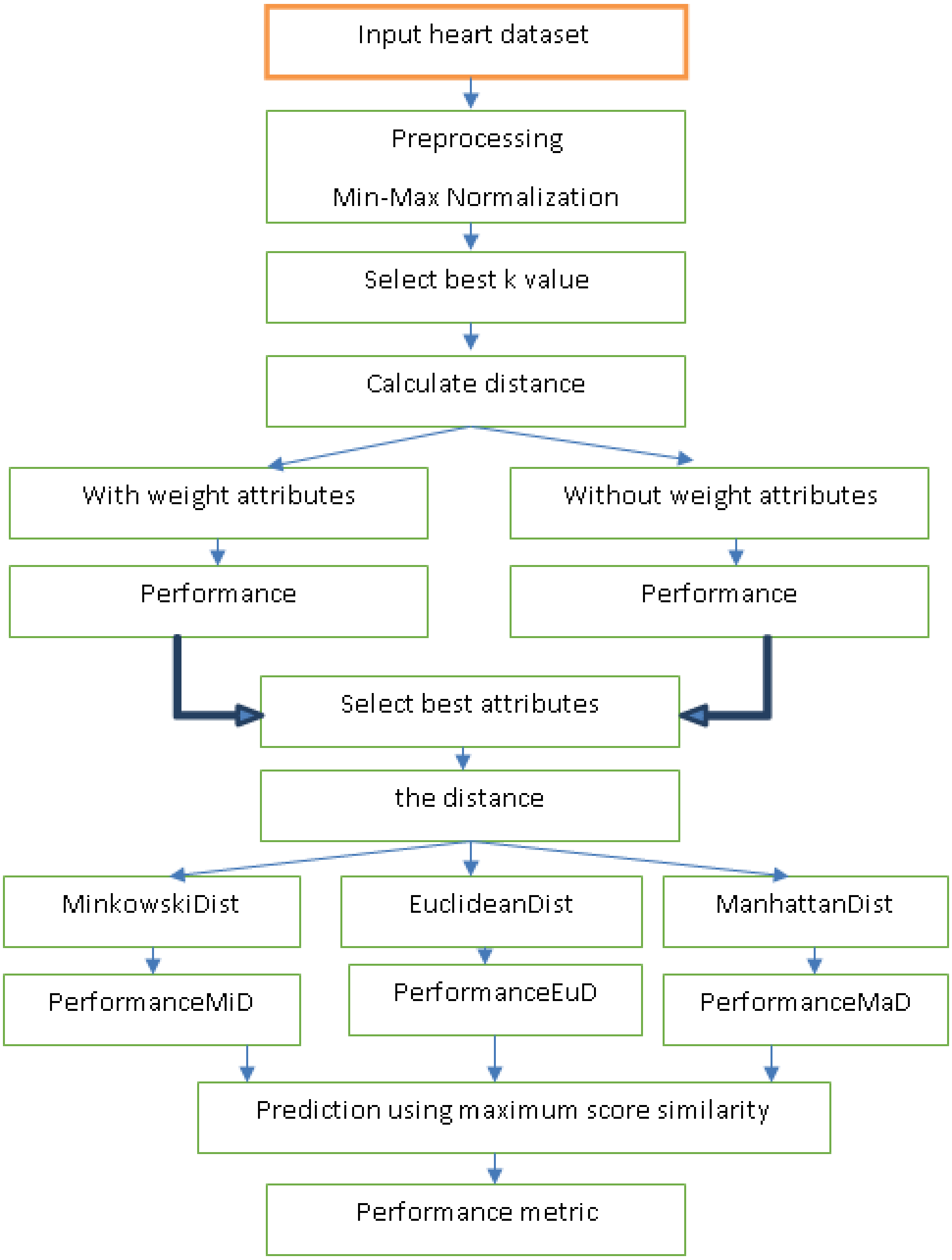

Figure 1. Block Diagram of the Proposed Work

The above diagram shows the proposed work flow. First this work read the input dataset, second it normalizes the dataset values from 0 to 1. After that it calculates the best k value. Then distance is calculated using with and without attributes and its performance. It selects the best accuracy attributes it may be with weight or without weight. After that it calculates the distance using three different Euclidean formula. This work finds out its performance. This work selects the maximum score similarity measure based prediction. Finally it calculates the performance metrics.

Performance analysis

The verifications of the applied techniques is carried out by a common database specially meant for heart disease. The dataset holds a sum of three hundred and ternate records possessing fifteen attributes which in turn break up into two sets of training set as 40% and testing set as 60%. In this operation, the language which is to be used is PYTHON. With a view of presaging the exact condition of the cardiac stoics (patients), an ideal matric was assembled in that case:

A signifies stoics with cardiac complications and **B** symbolises cardiac defects free patients.

Table 1
Confusion Matrix of A and B

	A (patient with heart disease)	B (patient with No heart disease)
A (patient with heart disease)	True Positive	False Negative
B (patient with NO heart disease)	False Positive	True Negative

The recent grid named Table 1 shows Confusion matrix of **A** and **B** encloses data about actual and predict classifications completed by a classification system.
The data enclosed in the matric are estimated so that the functioning of such systems is known.

- True positive
  If the records are actually true, the number of classified records is noted as true.
- False positive
  The number of records is mentioned as true if they were actually false.
- False negative
  The number of records is noted as false if they were really true.
- True negative
  The number of records is denoted as false if they were really false.
  The dataset is downloaded from the UCI repository that contains 303 records with 15attributes.

Table 2 shows the detailed dataset attributes description and distributions are age, sex, chest pain, cholesterol, and so on.

Table 2
The Detailed Dataset Attributes Description and Distribution

ATTRIBUTES				
SL.NO	SYMBOL	DESCRIPTION	TYPE	DATA RANGE
1	Age	Subject Age in years	Numeric	[29, 77]
2	Sex	Subject gender	Binary	1 = male 0 = female
3	Tresbps	Resting blood pressure in mmHg	Numeric	[94, 200]
4	Cp	Chest pain type	Nominal	1= typical angina 2= atypical angina 3=non-anginal pain 4 =asymptomatic
5	Chol	Serum cholesterol in mg/dl	Numeric	[126, 564]
6	Fbs	Fasting blood sugar with value >120mg/dl	Binary	0 – false 1 – true
7	Restecg	Resting electrographic results	Nominal	0 = normal 1=having ST-T wave abnormality 2 = showing probable or definite left ventricular Hypertrophy
8	Thalach	Maximum heart rate	Numeric	[71, 202]
9	Old peak	ST depression induced by exercise relative to rest	Numeric	[0, 6.2]
10	Exang	Exercise angina induced	Binary	0 = no 1 = yes
11	Ca	No of major vessels colored by fluoroscopy	Nominal	0-3 value
12	Slope	Slope of the peak exercise ST segment	Nominal	1=up-sloping 2=flat 3=downsloping
13	Thal	Defect type	Nominal	3-normal 6=fixed defect 7=reversible defect
14	Obes	Obesity	Binary	1=yes 0=no
15	Num	Diagnosis disease of heart	Binary	0-no 1-yes

The collected dataset was a structured dataset. The pre-processing was done by identifying the associated fields and removing all the duplications. After that, all the missing values werefilled and the data were coded according to the domain value.

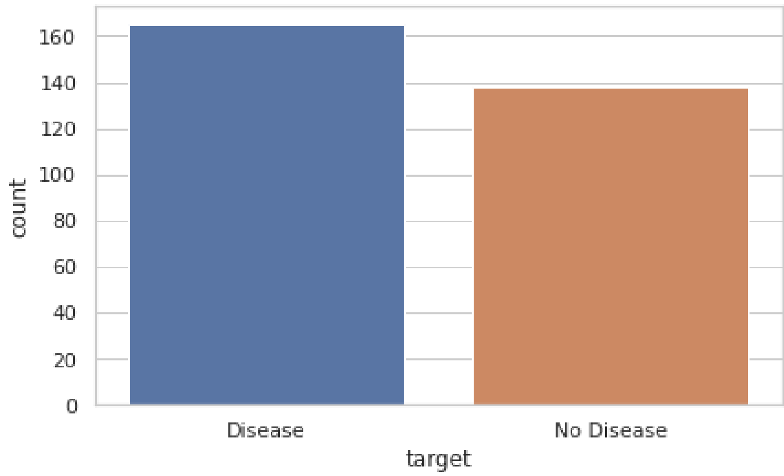

Figure 2. Heart Disease Count

The above figure 2 shows the number of heart disease count. Heart disease is 165 counts. No heart disease is 138 counts.

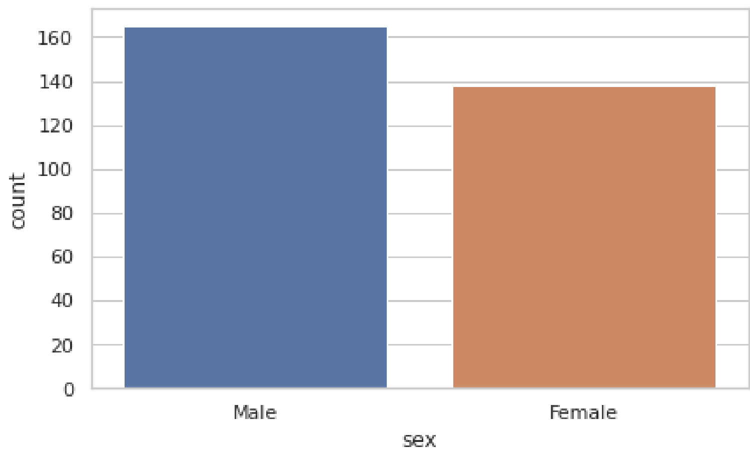

Figure 3. Gender Wise Heart Disease affected in %

The above figure 3 depicts the percentage of heart disease affect based on gender. The male are mostly affected than the female. Male percentage is 68.3% and female percentage is 31.7%.

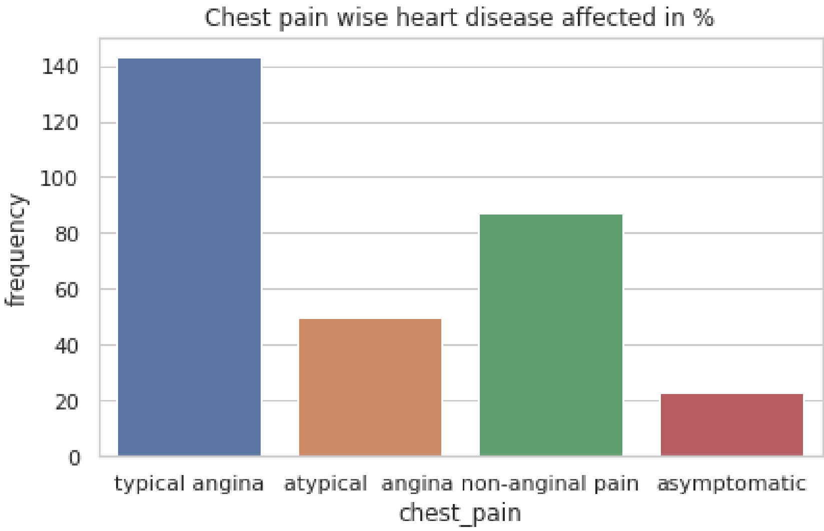

Figure 4. Chest Pain Wise Heart Disease affected in %

The above figure 4 predicts chest pain wise heart disease prediction. There are four types of chest pain. Chest pain (type 1) typical angina is 47.2% affected more than others. Other chest pain percentages are like 16.5% is atypical angina (type 2), 28.7% is non-anginal pain (type 3) and 7.6% is asymptomatic pain (type 4).

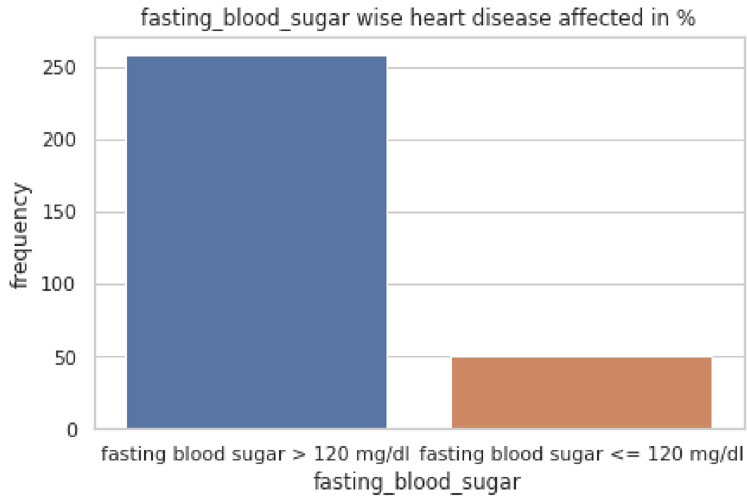

Figure 5. Fasting Blood Sugar Wise Heart Disease affected in %

The above figure 5 shows the heart disease affected percentage based on fasting blood sugar. If fasting blood sugar is greater than 120mg/dl, the heart disease affected percentage will be 83.8%. If fasting blood sugar is less than 120mg/dl, the heart disease affected percentage will be 16.2%.

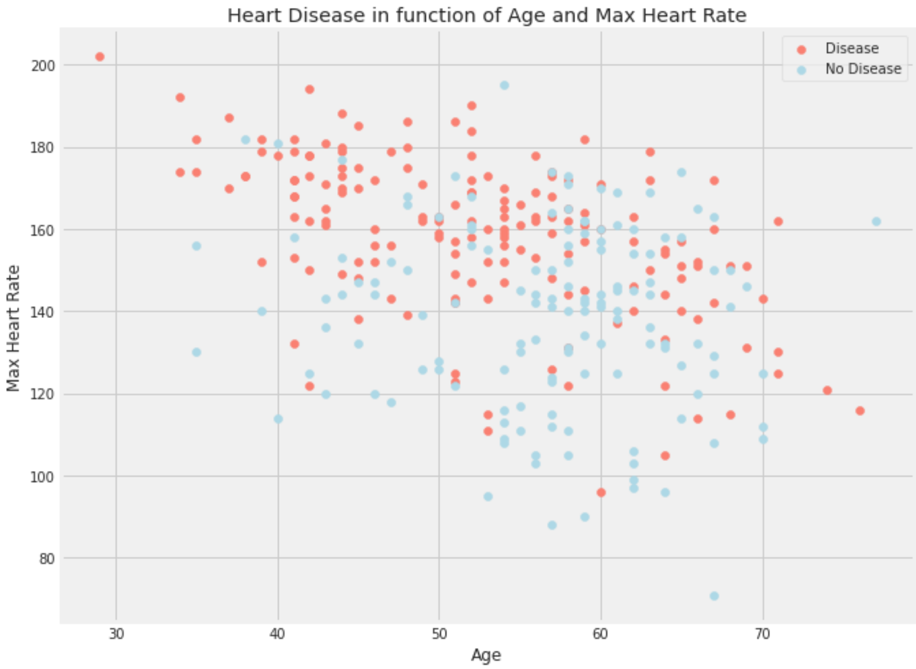

Figure 6. Heart Disease in Function of Age and Max Heart Rate

The above figure 6 shows the age wise maximum heart prediction. The blue dot shows the no disease. The red dot shows the disease. If the heart rate is greater than 140 then the disease possibility is high.

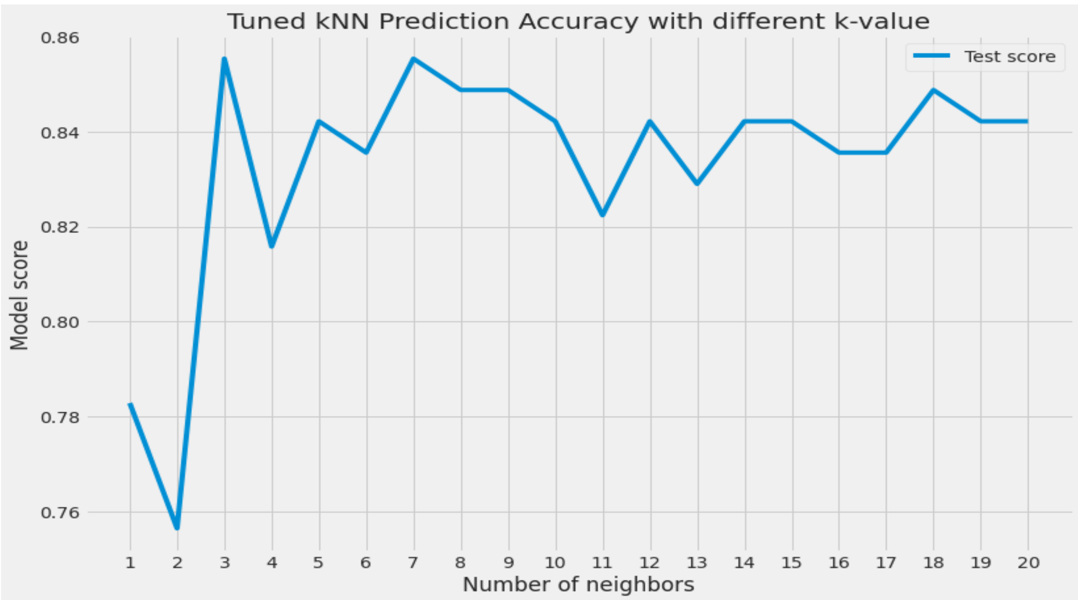

Figure 7. Tuned K-NN Prediction Accuracy

From the above graph figure 7, if we define K as **3-7-9** we will reach maximum score using **Tuned K-NN** model**.**

Table 3 shows the performance analysis of training and testing (70%, 30%) records.

Table 3
Performance Analysis of Training 70% Testing 30%

Performance Analysis of Training 70% Testing 30%					
Sl.No	Model	Testing Accuracy %	Precision	Recall	F1 Score
1.	K-Nearest Neighbours	86.81	87.18	88.70	87.93
2.	Tuned K-NN	89.01	82.54	90.43	86.31

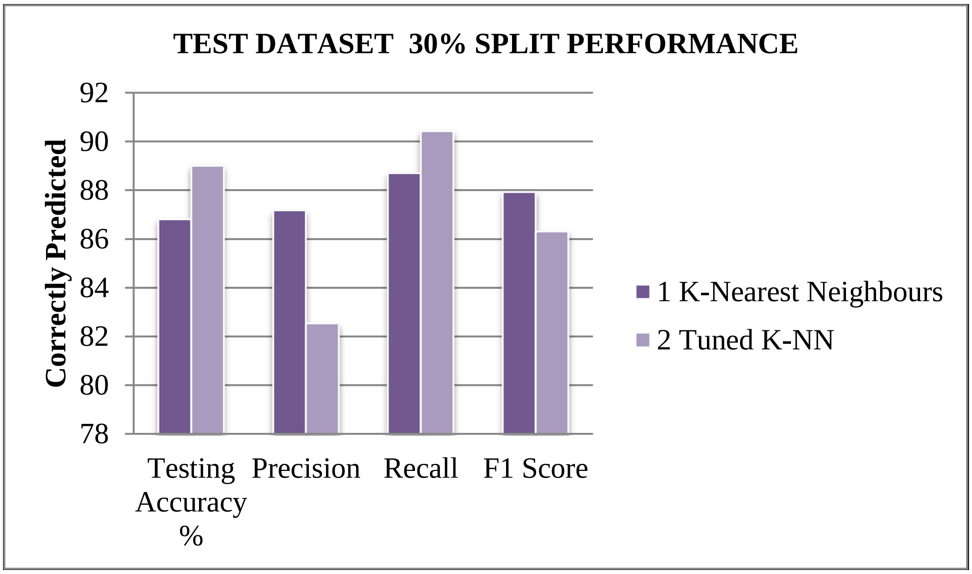

Figure 8. Performance Analysis of Training 70% and Testing 30%

The above figure 8 shows the performance of testing 30% of records. In **Tuned K-NN** performance of testing accuracy obtains **89.01%**, testing precision obtains 82.54%, testing recall obtains 90.43%, and testing f1-score obtains 86.31%.

Table 4 represents the performance analysis of training and testing (60%, 40%) records.

Table 4
Performance Analysis of Training 60% Testing 40%

Performance Analysis of Training 60% Testing 40%					
Sl.No	Model	Testing Accuracy %	Precision	Recall	F1 Score
1.	K-Nearest	86.89	90.91	85.71	88.24

	Neighbours				
2.	Tuned K-NN	86.07	88.41	87.14	87.77

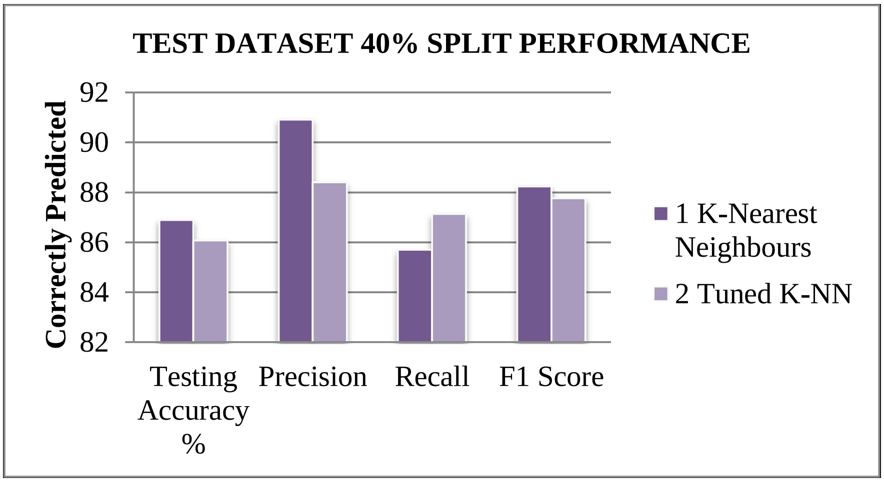

Figure 9. Performance Analysis of Training 60% Testing 40%

The above figure 9 shows the performance of testing 40% of records. In **Tuned K-NN** performance of testing accuracy obtains **86.07%**, testing precision obtains 88.41%, testing recall obtains 87.14%, and testing f1-score obtains 87.77%.

Table 5 shows the performance analysis of training and testing (50%, 50%) records.

Table 5
Performance Analysis of Training 50% Testing 50%

Performance Analysis in Training 50% Testing 50%					
Sl.No	Model	Testing Accuracy %	Precision	Recall	F1 Score
1.	K-Nearest Neighbours	84.21	86.25	84.15	85.19
2.	Tuned K-NN	84.12	86.25	84.15	85.19

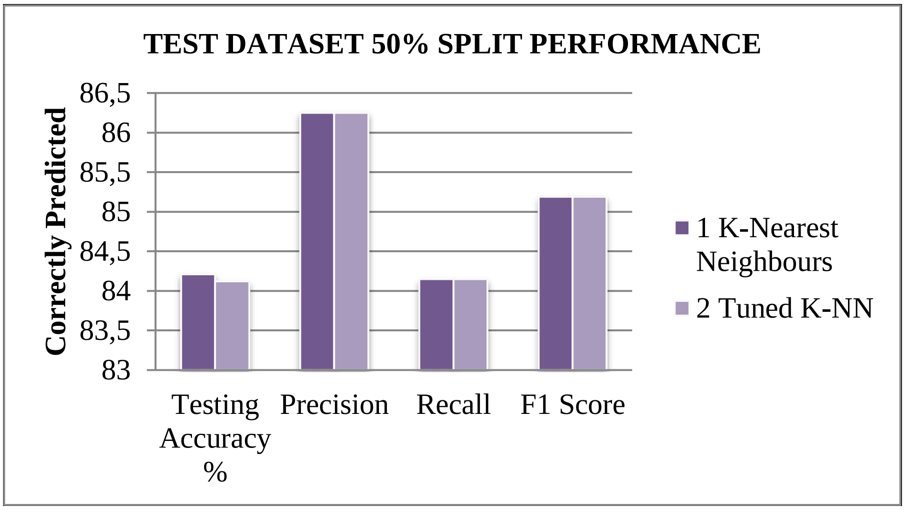

Figure 10. Performance Analysis of Training 50% Testing 50%

The above figure 10 shows the performance of testing 50% of records. In **Tuned K-NN** performance of testing accuracy obtains **84.12%**, testing precision obtains 86.25%, testing recall obtains 87.15%, and testing f1-score obtains 85.19%.

Table 6 represents the performance analysis of training and testing (40%, 60%) records.

Table 6
Performance Analysis of Training 40% Testing 60%

Performance Analysis in Training 40% Testing 60%					
Sl.No	Model	Testing Accuracy %	Precision	Recall	F1 Score
1.	K-Nearest Neighbours	81.87	84.69	82.18	83.42
2.	Tuned K-NN	82.97	85.01	84.16	84.58

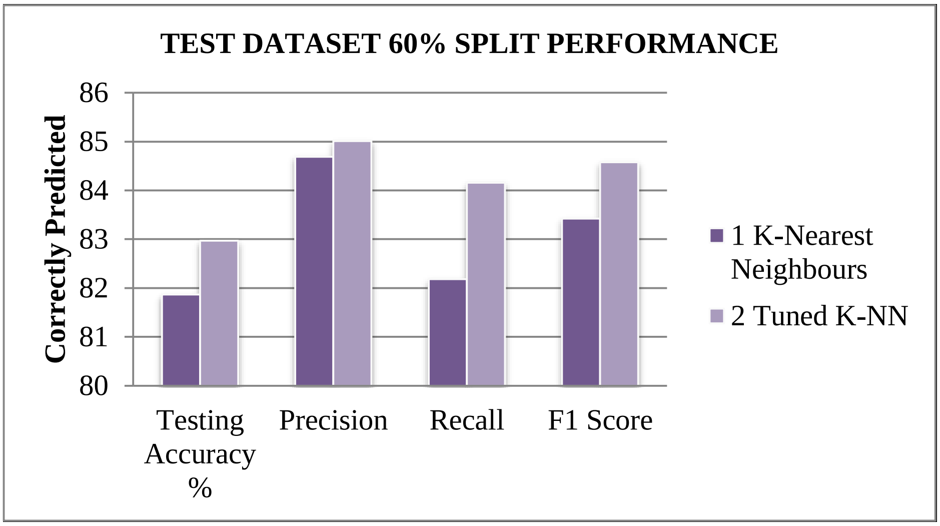

Figure 11. Performance Analysis of Training 40% Testing 60%

The above figure 11 shows the performance of testing 60% of records. In **Tuned K-NN** performance of testing accuracy obtains **82.97%**, testing precision obtains 85.01%, testing recall obtains 84.16%, and testing f1-score obtains 84.58%.

Conclusion

This work mainly focuses on the evolution of cardiac disease disclosure system by mean of the data collected from the patients. This reasoning of data seems to be a challenging task for its range and type of attribute selected. Regarding this prediction system, many apt methodologies are followed and many classifiers are developed so that the marking of normal and abnormal patients will be done by the received data. The records or the data thus collected carrying the very details of the patients showing symptoms of a defective heart. The attributes so far selected are compared with respect to their ability of predicting heart diseases by means of different machine learning algorithm. There are some supervised classifiers namely K- Nearest Neighbour and Tuned K-Nearest Neighbour were estimated with a view of classifying the patients who shows high and low levels of risk of Myocardial Infarction/ the grading of normal and abnormal patients based on quantitative data attributes which shows good classification results by a various trained classifier model. In **Tuned K-NN**, the **K-value** is **3, 7 and 9** predicting maximum accuracy.

References

1. M. Kavitha; G. Gnaneswar; R. Dinesh; Y. Rohith Sai; R. Sai Suraj , “Heart Disease Prediction using Hybrid machine Learning Model”, 6th International Conference on Inventive Computation Technologies (ICICT), ISBN:978-1-7281-8501-9, DOI: 10.1109/ ICICT50816.2021. 9358597, IEEE, Coimbatore, India, 2021.
2. Alperen Erdoğan. Başkent Üniversitesi, Ankara, Türkiye; Selda Güney ,“Heart Disease Prediction by Using Machine Learning Algorithms”,Published in:

2020 28th Signal Processing and Communications Applications Conference (SIU), ISBN:978-1-7281-7206-4, DOI: 10.1109/SIU49456.2020.9302468, IEEE, Gaziantep, Turkey, October 2020.

3. Chu-Hsing Lin; Po-Kai Yang; Yu-Chiao Lin; Pin-Kuei Fu, “On Machine Learning Models for Heart Disease Diagnosis”, Published in: 2020 IEEE 2nd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS), ISBN:978-1-7281-8712-9, DOI: 10.1109/ECBIOS50299.2020.9203614, IEEE, Tainan, Taiwan, May 2020.
4. Yiwen Meng; William Speier; Chrisandra Shufelt; Sandy Joung; Jennifer E Van Eyk; C. Noel Bairey Merz; Mayra, “A Machine Learning Approach to Classifying Self-Reported Health Status in a Cohort of Patients With Heart Disease Using Activity Tracker Data” , IEEE Journal of Biomedical and Health Informatics , ISSN: 2168-2208, Volume: 24, Issue: 3, March 2020.
5. Haixia Li 1, Guofeng Ren1, Xiaopu Yu2, Dongying Wang2,And Shizhe Wu3, “Discrimination of the Diastolic Murmurs in Coronary Heart Disease and in Valvular Disease”, 1Department of Electronics, Xinzhou Teachers University, Xinzhou 034000, China, 2Shanxi Medical University of China, Taiyuan 035100, China, 3Taiyuan Zhida Ivy League Middle School, Taiyuan 030006, Digital Object Identifier 10.1109/ACCESS.2020.3021093, China, September 2020.
6. Norma Latif Fitriyani; Muhammad Syafrudin; Ganjar Alfian; Jongtae Rhee, “HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System”, DOI: 10.1109/ACCESS.2020.3010511, ISSN: 2169-3536, Volume: 8, IEEE, July 2020.
7. Jian Ping Li; Amin Ul Haq; Salah Ud Din; Jalaluddin Khan; Asif Khan; Abdus Saboor, “Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare”,School of Computer Science and Engineering, University of Electronic Science and Technology of China, DOI: 10.1109/ACCESS.2020.3001149, ISSN: 2169-3536, Chengdu, China, IEEE, Volume: 8, June 2020.
8. Haolin Wang; Zhilin Huang; Danfeng Zhang; Johan Arief; Tiewei Lyu; Jie Tian, “Integrating Co-Clustering and Interpretable Machine Learning for the Prediction of Intravenous Immunoglobulin Resistance in Kawasaki Disease”, DOI: 10.1109/ACCESS.2020.2996302, ISSN: 2169-3536, IEEE, Volume: 8, May 2020,
9. Chunyan Guo; Jiabing Zhang; Yang Liu; Yaying Xie; Zhiqiang Han; Jianshe Yu, “Recursion Enhanced Random Forest With an Improved Linear Model (RERF-ILM) for Heart Disease Detection on the Internet of Medical Things Platform”, The Affiliated Hospital of Inner Mongolia Medical University, Hohhot, DOI: 10.1109/ACCESS.2020.2981159, ISSN: 2169-3536, IEEE, Volume: 8, China , March 2020,
10. A. Sankari Karthiga, Dr. M. Safish Mary, M.Phil Scholar, Assistant Professor, Department of Computer Science, Mother Teresa Women’s University, Department of Computer Science, St. Xavier’s College, Palayamkottai, Tirunelveli, “Early Prediction of Heart Disease Using Decision Tree Algorithm” , International Journal of Advanced Research in Basic Engineering Sciences and Technology Vol.3, Issue.3, March 2017.
11. A. Sankari Karthiga, Dr. M. Safish Mary, Research Scholar, Assistant Professor, Department of Computer Science, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli, Department of Computer Science, St.

Xavier's College, Palayamkottai, Tirunelveli, "Cardiovascular Disease Detection Using Machine Learning Techniques", International Journal of Mechanical Engineering, ISSN: 0974-5823, Vol. 7 (Special Issue, Jan.-Feb. 2022).