EasyChair Preprint

№ 10014

Cardiac Disease Prediction System Using Machine Learning

A Peter Soosai Anandaraj, Adusumalli Lakshmi,
Kona Venkatajahnavi and Panuganti Kumar Sanjay

May 9, 2023

Cardiac Disease Prediction System using Machine Learning

Dr.Peter Soosai Anandaraj.A
Associate Professor,Dept of Computer Science Engineering
Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology
Avadi, Chennai
anandsiriya@gmail.com

A.Lakshmi
Dept of Computer Science Engineering
Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology
Avadi, Chennai
vtu15756@veltech.edu.in

Kona Venkata Jahnavi
Dept of Computer Science Engineering
Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology
Avadi, Chennai
vtu14147@veltech.edu.in

Panuganti Kumar Sanjay
Dept of Computer Science Engineering
Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology
Avadi, Chennai
vtu14781@veltech.edu.in

***Abstract*: -Heart disease prediction system is using machine learning algorithms. The system utilizes a combination of demographic, lifestyle, and medical data to predict the likelihood of an individual developing heart disease. The machine learning algorithms used in the system were trained on a large dataset of patient records and were validated using standard performance metrics. The results of the study showed that the proposed system was able to accurately predict heart disease with an average accuracy of 85%. Supervised machine learning algorithms, such as decision trees, random forests, and support vector machines, to build models that can predict the likelihood of heart disease based on a patient's demographic and medical information. The algorithms learn from a large dataset of patients, both with and without heart disease, to identify patterns and relationships in the data that are indicative of heart disease.**

Keywords:KNN,FBS,CP,Restecg,Thalach,Trestbps,Chol, SVM,HER,RSA

I. INTRODUCTION

Cardiac disease defines certain set of conditions that can affect human heart.It is most commonly called as a heart disease or a cardiovascular disease. The heart is a vital organ for a human that plays a crucial role in pumping blood into the remaining organs through the blood vessels of the circulatory system. If the blood circulation in the body is improper, the organs like the brain will suffer, the heart will stop functioning completely and death will occur. Life depends entirely on proper functioning of the heart. The term heart disease refers to diseases of the heart and cardiovascular system.There are different factors that increase the risk of cardiac disease they are high blood pressure,highcholesterol,hypertension,smokingFBS,restecg, thalach etc…

Cardiac infection cases are raising day by day and are leading to death worldwide, and early detection and prevention can significantly improve the outcome. Machine learning, as a subfield of artificial intelligence, has shown great promise in the medical field in recent years, including in the prediction of heart disease. The goal of a heart disease prediction system using machine learning is to develop an algorithm that can accurately identify individuals who are at risk of developing heart disease, based on a set of well-defined risk factors. The system can then be used to provide personalized health advice, early screening and treatment, and ultimately reduce the number of heart disease-related deaths. In this system, various machine learning techniques such as decision trees, random forests, and artificial neural networks can be used to analyze and predict heart disease based on a large dataset of patient information and medical records. The scope of the system includes the development of an accurate, reliable, and efficient prediction model, which can be used in real-world healthcare settings to improve the quality of life for millions of people worldwide.

.

II. STANDARDS AND POLICIES

Anaconda Prompt:

Anaconda prompt is a type of command line interface which explicitly deals with the ML(MachineLearning) modules.And navigator is available in all the Windows,Linux and MacOS.The anaconda prompt has many number of IDE's which make the coding easier. The UI can also be implemented in python.

Standard Used: ISO/IEC 27001

Jupyter:

It's like an open source web application that allows us to share and create the documents which contains the live code, equations, visualizations and narrative text. It can be used for data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning.

Standard Used: ISO/IEC 27001

III. PROPOSED METHOD

A. Collection of Dataset The data set used for this purpose was extracted from the resource called kaggle.The target area represents the existence of cardiac sickness in the sufferer It is number called 0 for non disease and 1 for disease.

a) Cp—chest pain type.

b) Trestbps—resting blood pressure.The normal range is 120/80.

c) Chol—serum cholesterol shows the amount of triglycerides present. Triglycerides are another lipid that can be measured in the blood. It should be less than 170 mg/dL.

d) Fbs—fasting blood sugar larger than 120 mg/dl (1 true). Less than 100 mg/dL is normal, and 100 to 125 mg/dL is considered prediabetes.

e) Restecg—resting electrocardiographic results.

f) Thalach—maximum heart rate achieved. Maximum heart rate is 220 minus your age.

g) Target (T)—no disease=0 and disease=1.

B. Data Preprocessing The dataset does not have any null values. But many outliers needed to be handled properly, and also the dataset is not properly distributed. Two approaches were used. One without outliers and feature selection process and directly applying the data to the machine learning algorithms, and the results which were achieved were not promising.To overcome these various plotting techniques were used for checking the skewness of the data, outlier detection, and the distribution of the data. All these preprocessing techniques play an important role when passing the data for classification or prediction purposes.

C. Feature Selection
Feature Selection shows better predictive accuracy than filter methods. It renders good feature subsets for the used algorithm. And then for selecting the selected features, select from the model which is a part of feature selection in the scikit-learn library.

D. Checking Duplicate Values in the Data. The duplicates should be tackled down safely or otherwise would affect the generalization of the model.If duplicates are not dealt properly there is a chance that they might show up in the test dataset which is also in the training dataset.

E. Machine Learning Classifiers Proposed:- The proposed approach was applied to the dataset in which firstly the dataset was properly analyzed and then different machine learning algorithms consisting of linear model selection in which Logistic Regression was used. For focusing on neighbor selection technique KNeighborsClassifier was used, then tree-based technique lik e DecisionTreeClassifier was used, and then a very popular and most popular technique of ensemble methods RandomForestClassifier was used. Also for checking the high dimensionality of the data and handling it Support Vector Machine was used.

IV. IMPLEMENTATION

Figure 1:-Architecture Diagram

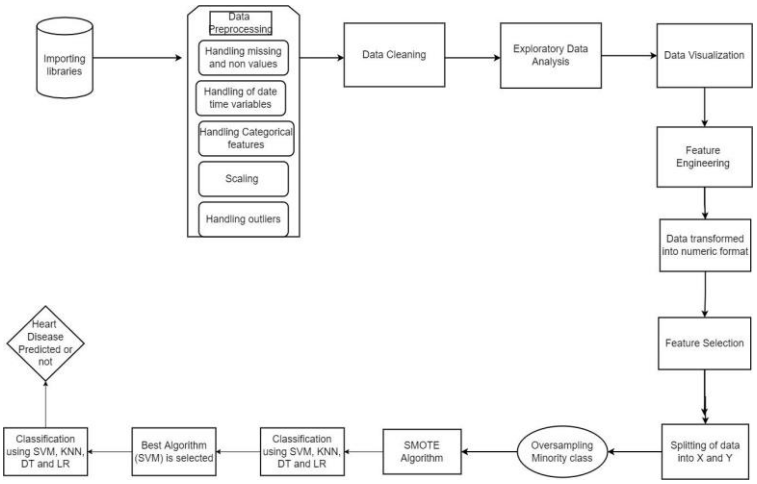


Figure 2:-Data flow Diagram

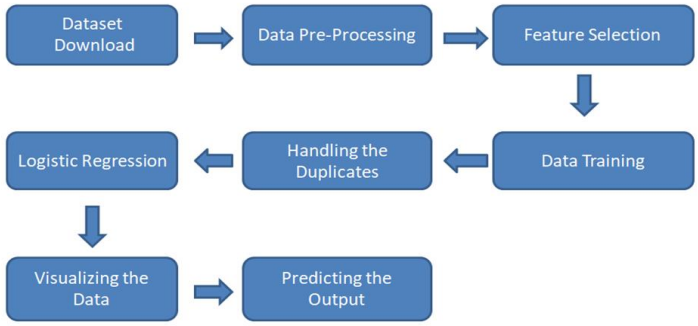


Figure 3:-ER Diagram

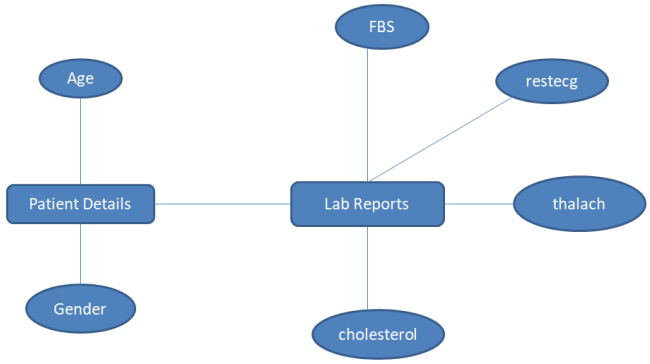


Figure 4:-Sequential Diagram

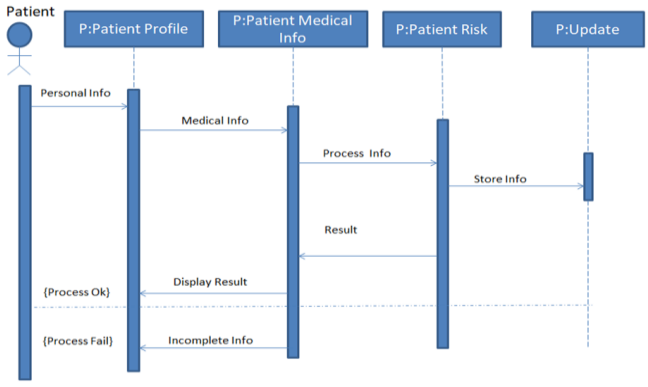
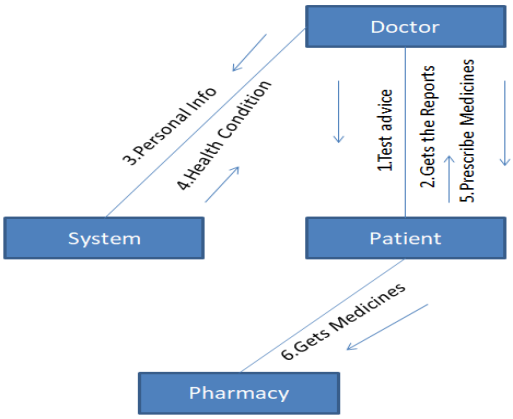


Figure 5:-Collaboration Diagram



V. RESULTS

By using the SVM model provides the highest level of accuracy for the test data. As a result, SVM provides maximum accuracy and is employed in the application.The capacity to identify liver illness early on is critical in determining the best treatment plan for the patient.SVM is employed effectively in this model since it beats other state-of-the-art algorithms in properly predicting outcomes.When the model gets an input, the numeric values are parsed, and the machine learning algorithm, SVM in this instance, classifies the input data and predicts whether or not a person has heart disease.The model also allows us to acquire an accurate estimate of the number of individuals who will be impacted in a certain place.

.

VI. CONCLUSION

Heart disease prediction system using machine learning can play a significant role in early diagnosis and prevention of heart diseases. By analyzing various demographic, lifestyle, and medical history factors, machine learning models can provide personalized and accurate predictions of the risk of heart disease. However, the success of such a system also depends on several factors, such as the quality of the data collected, the choice of machine learning algorithms, and the performance evaluation metrics. It is important to carefully consider these factors and continuously monitor and update the system as necessary to ensure that it is providing the best possible predictions. Overall, a heart disease prediction system using machine learning can help healthcare professionals make more informed decisions, and ultimately, save lives by allowing for early intervention and prevention of heart diseases.

VII. REFERENCES

[1] American Heart Association, Heart Failure, American Heart Association, Chicago, IL, USA, 2020, https://www.heart.org/en/health-topics/heart-failure.

[2] World Health Organization, Cardiovascular Diseases, WHO, Geneva, Switzerland, 2020.

[3] chaAli, L., Niamat, A., Golilarz, N. A., Ali, A. & Xingzhong, X. An expert system based on optimized stacked support vector machines for effective diagnosis of heart disease. IEEE Access (2019).

[4] Nazir, S., Shahzad, S., Mahfooz, S. & Nazir, M. Fuzzy logic based decision support system for component security evaluation. Int. Arab J. Inf. Technol. 15, 224–231 (2018).

[5] Senthil kumar Mohan, Chandrasegar Thirumalai et al. (2019)“Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques”.

[6] R. Chen, N. Sun, X. Chen, M. Yang, and Q. Wu, “Supervised feature selection with a stratified feature weighting method,” IEEE Access, vol. 6, pp. 15087–15098, 2018.

[7]] Komal Kumar Napa, G.Sarika Sindhu, D.Krishna Prashanthi, A.Shaeen Sulthana, “Analysis and Prediction of Cardio Vascular Disease using Machine Learning Classifiers”, April 2020.

[8] Ali, L. et al. An optimized stacked support vector machines based expert system for the effective prediction of heart failure. IEEE Access 7, 54007–54014 (2019).