International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Classification Algorithms for Predicting Heart Diseases and their Accuracies

Kanduru Sri Ramya[1], Kritika Chopra[2], Madhunika Kulkarni [3], Abdul Rahman Hassan[4], Thatikonda Asritha[5], B. Srinivas[6]

[1] *Student-Computer Science and Engineering & Nalla Malla Reddy Engineering College*
[2] *Student-Computer Science and Engineering & Nalla Malla Reddy Engineering College*
[3] *Student-Computer Science and Engineering & Nalla Malla Reddy Engineering College*
[4] *Student-Computer Science and Engineering & Nalla Malla Reddy Engineering College*
[5] *Student-Computer Science and Engineering & Nalla Malla Reddy Engineering College*
[6] *Assistant Professor-Computer Science and Engineering & Nalla Malla Reddy Engineering College*

Abstract - This project has been undertaken to detect the presence of heart disease in patients by taking some parameters like age, serum cholesterol level, resting blood pressure etc. as input and we have used 4 classification algorithms namely support vector machine, k-nearest neighbor, decision tree and linear regression. In this paper, we calculated the accuracy of machine learning algorithms for predicting heart disease. We have trained the system by using a heart disease dataset from the UCI repository and taken some real-time data for testing purposes. For the implementation, we used Python programming Anaconda (Jupiter) notebook and Django-a high-level Python web framework used in rapid development, for developing our project. We have successfully developed a model to detect the presence of heart disease and calculated the accuracy of all the algorithms used.

Key Words: Heart disease, UCI repository, machine learning, accuracies, algorithms.

1. INTRODUCTION

The heart is one of the most extensive and vital organs of the human body so the care of the heart is essential. Most diseases are related to the heart so the detection of heart diseases is necessary for this purpose comparative study is needed in this field, today most patients die because their diseases are not recognized till their advanced progression and no proper diagnostics due to lack of accuracy of instruments and lack of access to medical professionals. So, there is a need for an efficient model for disease detection by machine learning algorithms.

Machine Learning is found to be one of the efficient technologies for detection, which is based on training and testing. It is the branch of Artificial Intelligence (AI) which is one of the broad areas of learning where machines emulate human abilities. On the other hand, machine learning systems are trained to learn how to process and make use of data hence the combination of both technologies is also called Machine Intelligence.

As the definition of machine learning, it learns from the natural phenomenon, natural things so in this project we use biological parameters such as age, serum cholesterol level, resting blood pressure etc. to do the detection and on the basis of these, a comparison is done in the terms of accuracy of algorithms.

The heart plays a significant role in living organisms. Diagnosis and detection of heart-related diseases require more precision, perfection, and correctness because a little mistake can cause fatigue problems or death of a person. There are numerous death cases related to the heart and their count is increasing exponentially day by day. To deal with the problem there is an essential need for a detection system for awareness about diseases.
In our project, we use the UCI repository dataset for training and testing. For the implementation, we used Python programming Anaconda (jupytor) notebook and Django for development of high-level python web framework for rapid development.

2. REQUIREMENT SPECIFICATION

2.1 Software Requirements:

For developing the application the following are the Software Requirements:

1. Python
2. Django

2.2 Operating Systems supported:

Windows 10-64 bit OS

2.2 Technologies and Languages used to Develop

Python

2.3 Debugger and Emulator

Any Browser (Particularly Chrome)

2.4 Hardware Requirements:

For developing the following are the Hardware Requirements:

- Processor: Intel i3
- RAM: 4 GB
- Space on Hard Disk: minimum 1 TB

3. METHODOLOGY

The methodology section outlines the plan and method that how the study is conducted. This includes the Universe of the study, sample of the study, Data and

Sources of Data, study's variables and analytical framework. The details are as follows;

3.1 Methodology

3.1.1 Data Collection

First step in the system is data collection and deciding about the training and testing dataset. In our project we have used the heart disease dataset from UCI repository as training data and we have given 4 real time patient data as the test data.

Table -1: Attributes of training dataset

s.no	Attribute Name	Description	Type of Value
1	Age	Patient's age (29-77)	Numeric
2	Sex(gender)	Gender Male-0 Female-1	Nominal
3	cp	chest pain type -- Value 1: typical angina -- Value 2: atypical angina -- Value 3: non-anginal pain -- Value 4: asymptomatic	Nominal
4	trestbps	Resting blood pressure (in mm Hg on admission to the hospital 94 to 200)	Numerical
5	Chol	Chol Serum cholesterol in mg/dl, values from 126 to 564)	Numerical
6	fbs	Fasting blood sugar>120 mg/dl (true-1 false-0)	Nominal
7	restecg	Resting electro cardio graphics result (0 to 1) --Value 0: normal -- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) -- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria	Nominal
8	thalach	Maximum heart rate achieved (71 to 202)	Numerical
9	exang	Exercise included angina (1-yes 0-no)	Nominal
10	oldpeak	ST depression introduced by exercise relative to rest (0 to .2)	Numerical
11	slope	The slope of the peak exercise ST segment (1 to 3) -- Value 1: upsloping -- Value 2: flat -- Value 3: downsloping	Nominal
12	ca	number of major vessels (0-3) colored by flourosopy	Numerical
13	thal	3 = normal; 6 = fixed defect; 7 = reversable defect	Nominal
14	Num-the predicted attribute	1 or 0 diagnosis of heart disease (angiographic disease status) -- Value 0: < 50% diameter narrowing -- Value 1: > 50% diameter narrowing (in any major vessel: attributes 59 through 68 are vessels)	Nominal

3.1.2 Attribute Selection

Attributes of dataset are the property of dataset which are used for system and for heart. In our project we are taking 14 attributes for detecting heart diseases. The user can be the WHO, or the public healthcare provider or any healthcare professional. The user can enter the required data of the patient after registering them into the system. The attributes which we are using for testing data are taken from the Heart Disease of UCI repository and the 14 attributes which are being used are age, sex, (Cp) chest pain type,(Trestbps) resting blood pressure, (Chol) serum cholesterol, (Fbs) fasting blood sugar, (Resting) resting electrocardiographics result , (Thali) maximum heart rate achieved, (Exang) exercise included again, (Oldpeak) ST depression introduced by exercise related to rest, (Slope) the slope of the peak exercise ST segment, (Ca) number of major vessels, (Thal) thalium stress tests, and targets. These attributes are given with some ranges under which the values should be filled.

3.1.3 Preprocessing of data

Preprocessing needed for achieving prestigious results from the machine learning algorithms which are linear regression, k-nearest neighbor, support vector machine and decision tree. After preprocessing the data we are going to get the result in the form of graphs of two values that are "success" and "failure". Success represents the people with heart diseases and failure represents the people with no heart diseases.

3.2 System Architecture

The architecture of this project consists of several components that work together to provide a seamless and efficient user experience. Here are the key components of the system:

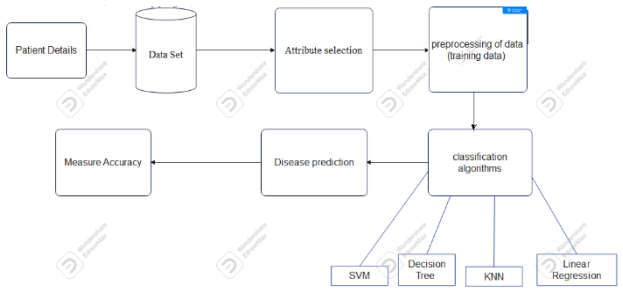

Fig -1: System architecture

i. **Patient details**: This component is responsible for collection of the system. As mentioned in section….. the user can input the details of the patient while registering a new patient and give all the required details asked by the system the details that are mandatory while registered are:
Username: Preferred Login id (user choice) username of the patient (User choice)
Login ID: Preferred Login id (user choice)
Password: Preferred password (user choice)
Mobile: patient's/next of kin's mobile number
Locality: patient's locality
Address: patient's home address
City: patient's city/town
State: state patient resides

ii. **Dataset**: Dataset component is the training data that we took from UCI repository i.e. heart disease data set as mentioned in Section I- introduction.

iii. **Attribute selection:** as mentioned in section 1 Introduction we have used the heart disease dataset from UCI repository as training data for our system. So, the attribute selection of this system is same as the attributes provided by the repository. The list of attributes is mentioned in section 3- 3.1.2 table 1.

iv. **Preprocessing of data:** This component is responsible for the preprocessing of the training data and the test data that we provide. Preprocessing deals with any redundancies, missing values, out of range values in the data and gets the data ready for analysis.

v. **Classification algorithms:** This component shows the 4 algorithms used in prediction. The data that we have is put under this algorithms for analysis. As shown in section 3.2 -fig- 1 there are 4 algorithms that we took for the prediction namely, SVM, KNN, Decision Tree, Linear Regression.

vi. **Disease prediction:** This is the step where the prediction takes place if the patients suffer from any heart diseases or not. The data from the step/component iv, come to this step after analysis and the prediction is made based on the result of step v. the result is displayed to the user whenever asked in pictorial format.

vii. **Measure Accuracy:** This is the last step/component of the system. The accuracy of each algorithm to predict is calculated and displayed.

4. CONCLUSION

The system predicts the presence of heart diseases in the patients given. Heart is one of the essential and vital organs of human body and prediction about heart diseases is also important concern for the human beings so that the accuracy for algorithm is one of parameter for analysis of performance of algorithms. Accuracy of the algorithms in machine learning depends upon the dataset that used for training and testing purpose. When we perform the analysis of algorithms on the basis of dataset whose attributes are shown in section-2.1.1- table 1 and on the basis of confusion matrix, we find KNN is best one with 92% accuracy.

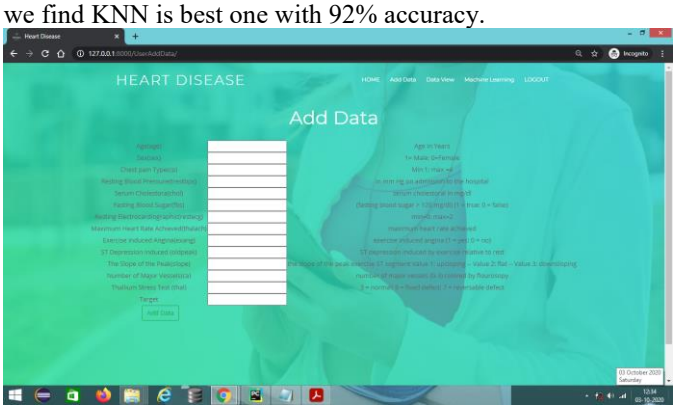

Fig -2: Patient Health Details

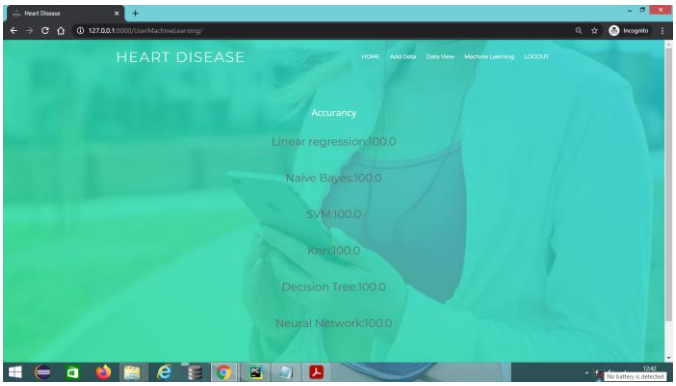

Fig -3: Accuracy of Algorithms for Training data

5. FUTURE SCOPE

For the Future Scope more machine learning approaches can be analyzed for best analysis of the heart diseases and for earlier prediction of diseases with more and better algorithms. So that the dependence on cardiac specialists can be reduced on and dependence on machine

grows in a positive thereby eliminating late diagnosing and scarcity of medical professional as cause of death in this field.

ACKNOWLEDGEMENT

The heading should be treated as a 3rd level heading and should not be assigned a number.

REFERENCES

1. "Heart Disease Prediction Using Machine Learning Algorithms"2020 International Conference on Electrical and Electronics Engineering (ICE3-2020) 978-1-7281-5846-4/20/$31.00 ©2020 IEEE. by Archana Singh Computer Science and Engineering Madan Mohan Malaviya University of Technology Gorakhapur, 273010 archansingh7971@gmail.com Rakesh Kumar Computer Science and Engineering Madan Mohan Malaviya University of Technology Gorakhapur, 273010 rkiitr@gmail.com

2. "An artificial intelligence model for heart disease detection using machine learning algorithms".2772-4425/© 2022 The Author(s). Published by Elsevier Inc., Accepted 2 January 2022, by Victor Chang, Vallabhanent Rupa Bhavani, Ariel Qianwen Xu, MA Hossain.

3. "The Role of Machine Learning Algorithms for Diagnosing Diseases" Vol. 02, No. 01, pp. 10 – 19 (2021) ISSN: 2708-0757 by Ibrahim Mahmood Ibrahim, Adnan Mohsin Abdulazeez.

4. "Analysis and Prediction of Cardiovascular Disease using Machine Learning Classifiers" 2020 6th International Conference on Advanced Computing & Communication Systems (ICACCS) by N. Komal Kumar, G.Sarika Sindhu, D.Krishna Prashanthi, A.Shaeen Sulthana.

BIOGRAPHIES (Optional not mandatory)

1'st Author Photo	**Kanduru Sri Ramya**, a student of Nalla Malla Reddy Engineering college pursuing 4th year B-Tech in the course of Computer Science and Engineering Email: kandurusriramya@gmail.com
	Kritika Chopra, a student of Nalla Malla Reddy Engineering college pursuing 4th year B-Tech in the course of Computer Science and Engineering Email: kritikachopra007@gmail.com
	Madhunika Kulkarni, a student of Nalla Malla Reddy Engineering college pursuing 4th year B-Tech in the course of Computer Science and Engineering Email: kulkarnimadhunika0@gmail.com
	Abdul Rahman Hassan, a student of Nalla Malla Reddy Engineering college pursuing 4th year B-Tech in the course of Computer Science and Engineering Email: mdabdulrahmanh@gmail.com
	Thatikonda Asritha, a Computer Science Engineering Graduate from Nalla Malla Reddy Engineering, working as a recruiter in **Datics Consulting,** Hyderabad. Email: asritha1219@gmail.com
	Mr. B. Srinivas, Assistant professor at Nalla Malla Reddy Engineering College.