© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 03 | NCASIT-2022 | April 18th 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

A Survey Paper on Heart Disease Prediction Using Multiple Models

Ms. Neha V. Mogre
Guide, Asst. Prof. [CSE]
Tulsiramji Gaikwad-Patil College of Engineering and Technology
Nagpur, India
neha.cse@tgpcet.com

Ms. Payal K. Rahangdale
B.E. [CSE] Student
Tulsiramji Gaikwad-Patil College of Engineering and Technology
Nagpur, India
prahangdale468@gmail.com

Ms. Ratneshwari S. Rahangdale
B.E. [CSE] Student
Tulsiramji Gaikwad-Patil College of Engineering and Technology
Nagpur, India
ratnarahangdale1999@gmail.com

Mr. Saurabh Y. Gondane
B.E. [CSE] Student
Tulsiramji Gaikwad-Patil College of Engineering and Technology
Nagpur, India
saurabhgondane358@gmail.com

Ms. Dolly D. Rahangdale
B.E. [CSE] Student
Tulsiramji Gaikwad-Patil College of Engineering and Technology
Nagpur, India
dollyrahangdale05@gmail.com

Abstract— **This**
Heart Disease prediction is one of the most complicated tasks in medical field. For medical purposes, the diagnosis of heart sickness is the difficult ventures. Heart diseases or cardio vessel Diseases (CVDs) unit for most reason for an enormous style of death among the global. The latest statistics of World Health Organization anticipated that cardiovascular diseases including vascular disease, Heart attack, Coronary Heart Disease, In the world as the biggest pandemic. On monthly basis huge amount of patient related data is maintained. The occurrence of future disease the stored data can be useful for source of predicting. This paper is presenting a comprehensive survey on heart disease prediction models.
Keywords-component; formatting; style; styling; insert (key words)

I. INTRODUCTION

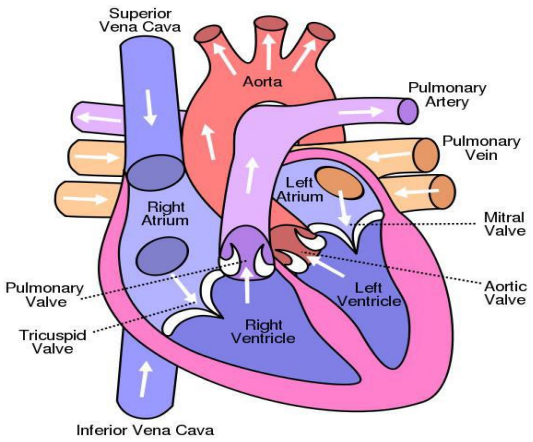
Heart is a vital organ of the organic structure it pumps blood to a part of our body. Heart disease can lead to reduce the lifespan of human beings nowadays. Every year 17.5 million of people are dying due to heart disease. Digital technologies are rapidly growing. In healthcare centers store huge amount of data in their database that is very challenging and complex to analysis. Data mining technique and machine learning algorithms plays vital role in analysis of different data in medical centers. The techniques and algorithms can be directly used in dataset for creating some models or to draw vital conclusions, and inferences from the dataset. Common attributes used for heart disease which is Age, Sex, Fasting Blood Pressure, Chest Pain type, Resting ECG (test that measures the electrical activity of the heart), Number of major vessels colored by fluoroscopy, Threst Blood Pressure (high blood pressure), Serum Cholesterol (determine the risk for developing heart disease), Thala Ch (maximum heart rate achieved), ST depression (finding on electrocardiogram, trace in the ST segment is abnormally low and below the baseline). painloc (chest pain location (substernal=1, otherwise=0)). Fasting blood sugar, Exang (exercise included angina), Food habits, smoke, Hypertension, height, weight and obesity [4]. Table 1 In the heart disease summarizes the most common types as follows.

TABLE I. DIFFERENT TYPE OF HEART DISEASE [5]

Heart Disease	Affected on the Function of Heart
Arrhythmia	The heart beat is improper whether it may irregular, too fast or too slow.
Cardiac arrest	An unexpected loss of heart function breathing and consciousness occur suddenly.
Congestive heart failure	Heart does not pump blood as well as it should., it is condition of the chronic.
Congenital heart disease	The heart's abnormality which develops before birth.
High Blood Pressure	It has a condition that be force of the blood against the artery walls is too high.
Peripheral artery disease	The narrowed blood vessels which reduce flow of blood in the limbs is the circulatory condition.
Stroke	Interruption of blood supply occur damage to the brain.

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 03 | NCASIT-2022 | April 18th 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

FIGURE I. HUMAN HEART[5]



Sr no.	Person	Systolic, mm Hg	Diastolic, mm Hg
1	Normal	130	85
2	High normal	130 -139	85- 89
3	Hypertension stage I	140 – 159	90 – 99
4	Hypertension stage II –IV	≥160	≥100

TABLE II. HUMAN HEART[4]

II. LITERATURE REVIEW

A. Data Collection

Data Description There are three important features of input classified as Objective, Examination, and Subjective.

1. Objective features for factual information from patient;
2. Examination features for results collected after medical examination;
3. Subjective features for information collected from the patient

B. Attribute Selection

Age, Glucose, Height, Cardiovascular disease (Presence), Weight, Physical activity, Gender, Alcohol intake, Systolic blood pressure, Smoking, Diastolic blood pressure, Cholesterol are the features of dataset used in model for description refer (table no.1) and BMI is added using calculation.

C. Preprocessing of data

Preprocessing needed for achieving prestigious results from the machine learning algorithms. For Example, Random Forest algorithm doesn't support null values dataset and for this we have to manage null values from original raw data. For our project we have to convert some categorized value by the dummy value means in the form of "0"and "1".

D. Data Balancing

Data balancing is essential for accurate result because by data balancing graph we can see both the target classes are equal. represent the target classes where "0" represent with heart diseases patient and "1" represent no heart diseases patients.

III. CLASSIFICATION ALGORITHMS

Classification means classifying the data in different groups based on the similarities present in different data points. Here classification is used in the prediction of heart disease. Various machine learning models are available, but in the proposed method, any one of the following algorithms or models can be used in given problem is a classification and regression problem. We are using various algorithms to find (or predict) the relation between the target variable (i.e., survived or not) with other variables (BMI, Gender, Age, and more). The following algorithms are used

A. Logistic Regression

It is a predictive analysis technique which is used when the target variable is dichotomous (binary). Logistic Regression model explains the relationship between one dependent binary variable and one or more independent variables. It predicts the probability of the target value.

B. k-Nearest Neighbors algorithm

k-Nearest Neighbors algorithm is a classification algorithm. The class of a particular data point is determined based on the class which is most common among its k-Nearest neighbors where k is a small positive integer.

C. Naive Bayes

It is a set of supervised learning algorithms based on applying Bayes' theorem. The naïve assumption being the conditional independence between every pair of features.

D. Neural Networks

A neural network is a series of algorithms which recognizes underlying relations in a training data set through a process that vaguely mimics the way of working of human brain. Basically,

neural networks are a system of neurons which are either artificial or organic in nature. Neural network adapts to the changing input which allows it to generate best output.

E. Decision Tree Classifier

This classifier model applies a Decision Tree as a predictive model. It organizes the characteristics (tree branches) to inferences about the target value (tree leaves). The classification trees are the tree models in which the target parameter can acquire a finite set of values. In these tree frameworks, the class labels are signified by the leaves, and the branches describe the concurrences of features that guide to those class labels. The regression trees are the decision trees in which the target parameter can take the continuous values (generally real numbers).

IV. CONCLUSION

In this paper, we have survey on two things:
The first part of the study is finding important factors affecting the heart disease, the important attributes and their minimum support values for no heart disease. Then the use of machine learning algorithm for prediction of heart disease. From risk factors we have selected number of attributes and their minimum value for normal and disease person. If the values of the attributes more than the minimum value mean you have a risk of heart disease. That means a person who have diabetes or smoking habit last 10 years, hypertension, chest pain, ST-T depression, older age more than 60 years, women with early menopause, over dirking, higher values of cholesterol more than 200mg/dl and person with blockage in heart vessels are more likely to be the heart disease person.

The Second part of the studies of supervised machine learning algorithms for prediction of heart disease. In these various algorithms studied are Support vector machine, Decision tree, Random Forest, Linear regression and Naive Bayes classifier. The dataset is quite old and it has no new attributes added in it. There is no pruning and cleaning of data. Uncleaned and missing values in the dataset has no use for classification and prediction. Moreover, no one has worked on the size of dataset. The small size of the data set is a problem for machine learning algorithms. Large size of the dataset is needed for the better prediction.

REFERENCES

[1] Olga Krzywińska, Marietta Bracha, Caroline Jeanniere, Emeline Recchia, Kornelia Kędziora Kornatowska, and Mariusz Kozakiewicz, “Meta-Analysis of the Potential Role of miRNA-21 in Cardiovascular System Function Monitoring”, Hindawi, BioMed Research International, Volume 2020, Article ID 4525410, 6 pages.

[2] Camila Oliveira, Erika Aparecida Silveira, Lorena Rosa, Annelisa Santos, Ana Paula Rodrigues, Carolina Mendonça, Lucas Silva, Paulo Gentil, and Ana Cristina Rebelo, “Risk Factors Associated with Cardiac Autonomic Modulation in Obese Individuals”, Hindawi, Journal of Obesity, Volume 2020, Article ID 7185249, 8 pages.

[3] Kumar V.V.

Healthcare Analytics Made Simple: Techniques in Healthcare Computing using Machine Learning and Python Packt Publishing Ltd. (2018)

Available at: https://books.google.com/books?hl=en&lr=&id=nwZnDwAAQBAJ&oi=fnd&pg=PP1&dq=application+of+python+programming+language+in+health+care+sectors&ots=BJz0Qe_09q&sig=Bhv_iaxOuZoKJu-WVEKPDr6B9y0Google Scholar

[4] P. Santhi a, R. Ajayb, D.Harshini c and S.S.Jamuna Sri d, “A Survey on Heart Attack Prediction Using Machine Learning”, Turkish Journal of Computer and Mathematics Education, Vol.12 No.2 (2021), 2303 – 2308.

[5] M.Marimuthu, M. Abhinaya, K.S. Hariesh, K. Madhankumar, V. Pavithra, “A Review on Heart Disease Prediction Using Machin Learning and Data Analytics Approach”, International Journal of Computer Applications (0975-8887) Volume 181-No. 18,September 2018.

[6] Ali, Liaqat, et al, "An optimized stacked support vector machines based expert system for the effective prediction of heart failure." IEEE Access 7 (2019): 54007-54014.

[7] Bigsby, K.G., Ohlmann, J.W. and Zhao, K., 2019. The turf is always greener: Predicting decommitments in college football recruiting using Twitter data. Decision Support Systems, 116, pp.1-12.

[8] Cardiovascular disease: Types, symptoms, prevention, and causes (medicalnewstoday.com).

[9] Brownlee, J. (2016). Naive Bayes for Machine Learning. Retrieved March 4, 2019, from https://machinelearningmastery.com/naive-bayes-for-machine-learning

UCI Machine Learning Repository: Heart Disease Data Set