



Received: 09-04-2022

Accepted: 19-05-2022

International Journal of Advanced Multidisciplinary Research and Studies

ISSN: 2583-049X

Survey report of heart disease prediction using machine learning

¹Dr. Rachhpal Singh, ²Dr. Rupinder Singh, ³Prabhjot Kaur

^{1,2,3}PG, Department of Computer Science and Applications, Khalsa College, Amritsar, India

Corresponding Author: Dr. Rachhpal Singh

Abstract

Heart anxiety is a situation when there is an irregularity of heart beat means either too fast or too slow. Due to occurrences of electrical impulses improperly heart beat moves up and down either fast or very slow and create anxiety that becomes the cause of sudden death. Abnormality of heart vessels identified by a test called heart angiography. It is a very important organ of human body as compared to other organs that pumps the blood and provides blood to brain and other parts of the body for better functioning. So healthy heart is very significant for human body but due to increase in heart failure risk from young age onwards there is need of more precautions and knowledge of

heart and its related diseases. Numbers of expertise in the field of heart diseases diagnose the heart related problems but failed early predictions. Therefore, some of the specific algorithms and mechanisms are used for estimation and prediction of the occurrence of heart problems at early stages. Data Analysis (collected from different hospitals and medical centers) and applying some machine learning algorithms becomes more helpful for prediction of heart failure on the basis of some parameters of patient's data. Here we study, analyzed and discussed some of the research papers regarding the prediction of heart blockage problems with machine learning technique.

Keywords: Heart Disease, Data Mining, Data Analysis, Feature Selection, Data Classification, Random Forest and Machine Learning

1. Introduction

Health care system has huge amount of data and to organize and use such data some mechanisms used. Data mining is a better mechanism for handling such data. As death cases are increasing day by day in worldwide due to Heart failure or disease ^[1]. In cardio problems blockage or Abnormality of coronary artery only can be detected when it is at high risk. In previous years, many developments were done for the improvement in the diagnosis of heart related diseases that increases the survival rate. Angiography is best technique for diagnosing the heart related problems. This will give high predictions of heart disease. The outcomes occur here for prediction during diagnosis by using the data is considered in percentage. Medical parameters are the best way for classification of given or obtained data sets by using the data mining technique. Some of the important algorithms of Machine Learning (ML) for processing and handling the datasets using python are Naive Bayes Algorithm and Decision Tree Algorithm for finding the level of accuracy for occurrences of heart disease. ML is the widely used part of Artificial Intelligence (AI) that emphasizes on imitating of intellectual abilities of humans by computers or machines ^[2]. As today's world has huge amount of data cannot handled by human beings directly or indirectly and need some computerized mechanisms for handling, organizing and manipulating it ^[3].

Machine learning has two major roots like model recognition and Competitive learning theory for better prediction in data analytics field by developing some algorithmic approaches and designing some specific models. By doing such approaches scientists, researchers, analysts and engineers get the valid and reliable outcomes. Number of features or hidden patterns becomes helpful for best approximation of prediction by using model automation in ML. Here we can develop a model by considering the training data. Further data is tested for this model by using the test data set. There is no need to teach or tell the machines regarding what to do, when to do and where to do, only the function mechanisms become helpful for machines to use the models or methods for results production depending upon some of the new constraints and obtained data ^[4]. Before processing such model, data passes through any one of the machine learning stages like either from Supervised Learning or from Unsupervised Learning or from Reinforcement Learning. Supervised learning used for building a model that work in unknown circumstances. It provides us the way about how to use all the given inputs and obtained outputs and further allow such set of values for creating a relationship for optimized accurate prediction. Supervised learning has naming and classification two major strategies on the basis of predicted variables. If estimated values are in the continuous form, then it is a

regression problem, but if estimation occurs from independent variables, then it is classification problem. RF, SVM, NB, DT, ANN etc. are some supervised learning techniques. By exploring the unknown values or data sets, unsupervised learning is applied. When information obtained in unorganized and unrelated groups then some least explored Dimensionality Reduction or Cluster Analysis learning methods are used. In Cluster approach, we get same value sets of intra cluster similarity and very few value sets inter cluster similarity. Unwanted and replicated variables or value sets occurs in case of dimension Reduction that generate a very smaller data values or data subset from the original data [2]. In ML, machines are trained in such a manner that there is less time occurs during processing for maximization of outcomes under given constraints. Number of machines and software are used for finding the possible behaviors and approaches under particular situation. Note that something will learn definitely from process model by using the machine learning concept [5]. Number of Medical Centers uses the ML techniques for diagnosing and prediction of heart failure related problems. Data sets or clinical parameters can be easily available from different hospitals or health centers. Heart diseases becomes human deterioration that affects the cardiac function and so it is stimulated that 18 million people died by heart disease [6, 7]. Rapid growth in digital technology, now mostly all the medical organizations are storing the data from various data sources for further processing and use for prediction. The Learning approaches and data mining techniques are paly a vital role for processing of uneven or unorganized data and so applied a classification approach. Such type of models or mechanisms becomes helpful for generating the dataset or models so that some better conclusions can be drawn from them. Some of the common parameters or features considered for heart related disease are sex, Age, Chest pain, blood pressure, weight, ECG, Hypertension, excellent fluoroscopy set values, High Cholesterol level, ST pressure, Thalach, painloc value, smoke, accelerating blood sugar, eighting habits, obesity and height as shown in the below table with their symptoms [8, 9].

Table 1: Risk Factors for heart problems with description

Risk Factors	Description
Type of Gender	High risk in men
Age	High in increases in age
High BP (blood pressure)	Higher the BP, high the risk rate [8, 9].
Bad habit of Smoking	higher rate in smoker
Family Background	Inherited property of occurrence of attack effects the high risk
Cholesterol	High the Cholesterols points, more be the risk
Poor Food	Health the food, less be the risk
Obesity	It is a major factor
Heavy the Stress	More be the risk
Diabetes	High risk in diabetes patients
Hygienety	High risk in bad hygienic food and habits
Physical activity	Physical activities avoid the occurrence

Some of the more general symptoms for occurrences of heart problems are as below:

Chest pain, Fatigue, Shortness of breath, Swollen feet, Failing and Abnormal heartbeat

A few common types of heart related diseases are discussed in the table below [10]:

Table 2: Common Heart Diseases with some symptoms

Disease Name	Symptoms
Cardiac arrest	Feelings of worthlessness, loss in breath and heart stops suddenly.
Arrhythmia	Occurrence of improper heartbeat
Stroke	Blood supply interruption to brain.
Congestive heart disorder	Pumping of blood in heart stops and so supply of oxygen stopped.
High BP	Blood passes with high force in blood veins.
Coronary artery disorder	The vital blood vein damages the blood vessels.
Peripheral artery disease	Flow of blood becomes slow to the lower part of the body.

2. Literature survey

Plenty work has been done in various health centers for measuring the diseases using different ML and other techniques. Some of the research articles studied and discussed as below:

Langley *et al.* [11] created myocardial ischemia classification algorithm that increase the sensitivity and specificity by using the Support Vector Machine (SVM) for classification. Marjia Sultana *et al.* [12] predicted the heart related problems using some of the basic programs like SMO, BN, K-Star etc. in the WEKA software (Waikato Environment for Knowledge Analysis) software. In this research UCI Machine Learning Repository cardiac data source is used having 270 records so that accuracy was improved by using the best diagnostic decision system by comparing all the given systems.

Purushottam *et al.* [13] developed a cardiac functional prediction mechanism using the decision tree approach after collecting the data sets from University of California and was tested on WEKA software for effective decision that improves the accuracy of prediction of heart problems.

Chala Beyene *et al.* [14] monitored an automatic cardiovascular disease system using ML technique that can be used easily by less experience persons. WEKA software is used for data analysis and finding the accuracy of prediction.

Ashok Dwivedi [15] predicted the heart diseases using some popular KNN, SVM and ANN etc. like algorithms that use the UCI data set naming Stat log heart disease Dataset by discussing 270 various features for obtaining the high accuracy of prediction.

S. Seema *et al.* [16] studied the heart diseases related UCI database on the basis of SVM, ANN like algorithms to compute the optimal accurate rate of prediction by comparing all the techniques.

K. Polaraju *et al.* [17] discussed Multiple Regression Model by diagnosing the cardiovascular disease on the basis of various characteristics of patient. Testing was done by taking the 3000 attributes of training data set values and comparing all the given approaches, found the high accuracy in prediction.

Jayami Patel *et al.* [18] compared the various DT classification techniques for predicting heart diseases using WEKA software and UCI data set having 76 symbols and 303 cases. It extracts the hidden pattern by suing data mining. After comparisons of all the techniques, highest accuracy prediction level obtained.

S. Prabhavathi *et al.* [19] discussed the systematic approach for extracting and controlling the estimated heart attack initiation having high success rate from the heart attack data sets. Here data mining method is used for improving the forecasting of heart disease problems and that will obtain a better CAD detection performance. After reducing the feature and classification approach with GA will increase the accuracy level of prediction.

Sairabi H. Mujawar *et al.* [20] analyzed some Modified K-methods and NB based DM techniques for cardiovascular disease by analyzing the 300 records having 13 symptoms on Cleveland Database. Assembly and partitioning procedure used for modification and forecasting by using modified K-means proposed model.

R. Sharmila *et al.* [21] discussed a Map Reducing programming with Hadoop using support vector machine (SVM) for improving the accuracy of prediction of occurrence of heart diseases.

Jabbar *et al.* [22] developed the Chi Square and Random Forest (RF) methods for the estimation of heart disease. After collecting the data base from different health centers in Hyderabad, testing was done on various databases that become very much helpful for prediction accuracy in proposed model.

Yeshvendra *et al.* [23] evaluated the Random Forest (RF) method using Cleveland Dataset for the prediction of cardiovascular disease and find the efficiency and accuracy of prediction.

Amin UI *et al.* [24] analyzed different performance metrics

like sensitivity, accuracy, performance time etc. and developed a diagnostic module using Feature selection techniques and ML approach. After comparing all the given or obtained performance, it was found that proposed method gave better accuracy.

Narayan *et al.* [25] discussed the ML and Fourier Transform for finding the accuracy of chronic diseases prediction.

Abderrahmane *et al.* [26] elaborated a RT-HD naming prognosis software using Apache Spark for spark propagation that visualize and accumulate the generated data and show the best streaming of prediction.

Abdar *et al.* [27] developed a novel N2-Genetic Optimizer using SVM and PSO or GA for the evaluation of cardiovascular disease with different ML techniques and find the best SVM selection algorithm. This will give better accuracy of prediction.

Akgul *et al.* [28] developed a hybrid approach having combination of ANN with GA for the occurrence of heart related problems and monitor it with some of the default parameters. This will give better accuracy than any other ML approaches.

Kumar *et al.* [29] analyzed and predicted all the possibilities of cardio vascular problem with the help of machine learning related mechanisms and its classifiers that gives better results as compared to other techniques.

Nikhar *et al.* [30] studied all the heart related problems and developed a machine learning process for prediction and apply all the medical aids for fast recovery.

Table 3: A complete view of existing literature having Parametric Analysis

Author Name	Research Purpose	Technique used	Accuracy in Outcomes
Mujawar <i>et al.</i> (2015)	Modified K-means used and prediction takes place for heart disease by using NB.	NB and Modified K-means algorithm	Detection of Heart Disease is 93%
Sultana, Marjia <i>et al.</i> (2016)	Using WEKA tools, Prognosis of Heart Illness detected.	K Star	75%
Deepika, Kumari, and S. Seema (2016)	Data Mining used for Chronic Disease Prediction	NB = 93.85%, ANN= 94.27% DT= 92.59%, SVM = 95.2%	
Dwivedi, Ashok Kumar (2016)	For finding Heart problem, execution of distinct ML algorithms	NB	83%
Patel, Jaymin (2016)	DM and ML techniques used for prediction	LMT	56.76%
Prabhavathi, S., and D. M. Chitra (2016)	DNFS Techniques used.	SVM, NB	Diagnosis in range 82% to 92%
Jabbar <i>et al.</i> (2016)	Use of Feature Selection and RF for heart problem prediction	Random Forest	83.70%
Singh <i>et al.</i> (2016)	RF predict heart disease	RF	86%
Beyene, MrChala, and Pooja Kamat (2018)	DM Techniques used for analysis and prediction of Heart Disorders.	SVM	Less in range
R. Sharmila, S. Chellammal (2018)	DM and Big Data approaches used	SVM	SVM =85%
Haq <i>et al.</i> (2018)	Machine Learning for the prediction and Hybrid Framework having Feature selection used	SVM is 86% and Artificial Neural Network is 75%	
Naryan <i>et al.</i> (2018)	FFT based system with Machine Learning for predicting and Identify Heart Disease	Fourier Transformation and Machine Learning	93%
ED-DAOUDY <i>et al.</i> (2019)	ML and Big Learning will detect of heart disease	Big Data and ML	89%
Abdar <i>et al.</i> (2019)	ML used for Accurate Diagnoses	GA+SVM and PSO + SVM	94%
Akgül <i>et al.</i> (2019)	Hybrid GA and ANN applied	GA+ANN	96%
Kumar <i>et al.</i> (2020)	Predicted the heart diseases	Machine learning with RF	92% accuracy
Nikhar <i>et al.</i> (2016)	Developed a machine learning system	NB with decision tree classification	88% accuracy of results

3. Conclusion

Number of data mining techniques used for working and operational environment for the occurrence's cardiac disorders with various data sets. The performance improved by using some of the best optimized feature selection method. Note that SVM has highly efficient whereas KNN has low accuracy. More accurate and better outcomes can help the researchers associated with the concerned field that provide better treatment to the patients. So, detection of heart problem at early stage using the prediction will provide good quality services and make the health of every person in safe mode.

4. References

- Prinyakupt J, Pluempitiwiriyaewej C. Segmentation of white blood cells and comparison of cell morphology by linear and naïve Bayes classifiers," *BioMedical Engineering Online*. 2015; 14(1):63.
- Rätsch, Gunnar. A brief introduction into machine learning. Friedrich Miescher Laboratory of the Max Planck Society, 2004.
- Alpaydin, Ethem. Introduction to machine learning. MIT press, 2020.
- Ukil A, Bandyopadhyay S. Automated cardiac health screening using smartphone and wearable sensors through anomaly analytics. in *Mobile Solutions and Their Usefulness in Everyday Life*, Springer, Berlin, Germany, 2019, 145-172.
- Sutton, Richard S. Introduction: The challenge of reinforcement learning. In *Reinforcement Learning*, pp. 1-3. Springer, Boston, MA, 1992.
- Learning, Machine. Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review. *Advances in Computational Sciences and Technology*. 2017; 10(7):2137-2159.
- Krishnaiah V, Narsimha G, Subhash Chandra N. Heart disease prediction system using data mining techniques and intelligent fuzzy approach: A review. *International Journal of Computer Applications*. 2016; 136(2):43-51.
- Srinivas K, Raghavendra Rao G, Govardhan A. Survey on prediction of heart morbidity using data mining techniques. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*. 2011; 1(3):14-34.
- Web site:
<http://www.mayoclinic.org/diseasesconditions/heart-disease/symptoms-causes/dxc-20341558>.
- Web site:
<https://www.medicalnewstoday.com/articles/257484.php>.
- Asher E, Abu-Much A, Bragazzi NL. CHADS2 and CHA2DS2-VASc scores as predictors of platelet reactivity in acute coronary syndrome. *Journal of Cardiology*. 2020; 4.
- Sultana, Marjia, Afrin Haider, Mohammad Shorif Uddin. Analysis of data mining techniques for heart disease prediction. In 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), IEEE, 2016, 1-5.
- Purushottam, Kanak Saxena, Richa Sharma. Efficient Heart Disease Prediction System using Decision Tree, 2016, 15.
- Beyene, MrChala, Pooja Kamat. Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques. *International Journal of Pure and Applied Mathematics*. 2018; 118(8):165-174.
- Dwivedi, Ashok Kumar. Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Computing and Applications*. 2016; 29:685-693.
- Deepika, Kumari, Seema S. Predictive analytics to prevent and control chronic diseases. In 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), IEEE, 2016, 381-386.
- Polaraju K, Durga Prasad D. Prediction of Heart Disease using Multiple Linear Regression Model. *International Journal of Engineering Development and Research (IJEDR)*. 2017; 5(4):1419-1425. ISSN:2321-9939
- Patel, Jaymin, Tejal Upadhyay, Dr. Samir Patel. Heart Disease Prediction using Machine Learning and Data Mining Technique. 2016; 7(1):129-137.
- Prabhavathi S, Chitra DM. Analysis and Prediction of Various Heart Diseases using DNFS Techniques. *International Journal of Innovations in Scientific and Engineering Research (IJISER)*. 2016; 2(1):1-7.
- Mujawar, Sairabi H, Devale PR. Prediction of Heart Disease using Modified K-means and by using Naïve Bayes. *International Journal of Innovative research in Computer and Communication Engineering*. 2015; 3:10265-10273.
- Sharmila R, Chellammal S. A conceptual method to enhance the prediction of heart diseases using big data Techniques. *International Journal of Computer Sciences and Engineering*. 2018; 6(4):21-25.
- Jabbar Akhil M, Bulusu Lakshmana Deekshatulu, Priti Chandra. Prediction of heart disease using random forest and feature subset selection. In *Innovations in Bio-Inspired Computing and Applications*. Springer, Cham, 2016, 187-196.
- Singh, Yeshvendra K, Nikhil Sinha, Sanjay Singh K. Heart Disease Prediction System Using Random Forest. In *International Conference on Advances in Computing and Data Sciences*, Springer, Singapore, 2016, 613-623.
- HAQ, Amin UL, *et al.* A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mobile Information Systems*, 2018.
- Narayan, Subhashini, Sathiyamoorthy E. A novel recommender system based on FFT with machine learning for predicting and identifying heart diseases. *Neural Computing and Applications*. 2019; 31(1):93-102.
- Ed-Daoudy Abderrahmane, Maalmi Khalil. Real-time machine learning for early detection of heart disease using big data approach. In: 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS). IEEE, 2019, p1-5.
- Abdar, Moloud, WojciechKsiążek U, Rajendra Acharya, Ru-San Tan, Vladimir Makarenkov, *et al.* A new machine learning technique for an accurate diagnosis of coronary artery disease. *Computer methods and programs in biomedicine*. 2019; 179:104992.
- Akgül, Miray, Özlen Erkal Sönmez, Tuncay Özcan. Diagnosis of Heart Disease Using an Intelligent Method: A Hybrid ANN-GA Approach. In

- International Conference on Intelligent and Fuzzy Systems, Springer, Cham, 2019, 1250-1257.
29. Kumar Komal N, Sarika Sindhu G, Krishna Prashanthi D, Shaeen Sulthana A. Analysis and prediction of cardio vascular disease using machine learning classifiers. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), IEEE, 2020, 15-21.
 30. Nikhar, Sonam, Karandikar AM. Prediction of heart disease using machine learning algorithms. International Journal of Advanced Engineering, Management and Science. 2016; 2(6):239484.