International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-4, April 2020

Cardiovascular Disease Recognition through Machine Learning Algorithms

Rajatdeep Kaur, Kamaljit Kaur

Abstract: ***The heart is more important to the human body than any other circulatory organs. Its function is to provide and pump blood to other organs and brain. So it is very important to have a healthy heart but researches revealed the risk of heart failure increases every day starting from age 30. Many heart specialist can diagnose heart disease with their experience and skills. But some experts lacking the talent or knowledge to predict cardiovascular disease in the early stages, a small mistake can cost a patient's life. Therefore, it is necessary to use specific methods and algorithmic tools to estimate the occurrence of cardiac disorders in the early stages. Different Algorithms for machine learning and data analysis are beneficial in predicting various diseases from patient's data, managed by the Medical Center or hospitals. The data obtained may also help to assess the presence of the disease in the future. Heart Disease or Cardiac related issues can be analyzed by variety of machine learning techniques, Instance Artificial Neural Network, Decision Tree, Random forest, K-nearest neighbor, Naïve Bayes and Support Vector Machine. This study establishes a theoretical understanding of existing algorithms and provides a general understanding of existing work.***

Keywords: Cardiovascular Disease, Data Mining, Machine Learning.

I. INTRODUCTION

Machine Learning is part of the widely used field of Artificial Intelligence that focuses on mimicking humans intellectual abilities by machines [1]. In today's world the volume of data is on the rise, which cannot be ruled out by human involvement as before. Historically, Human Power is used to analyze data, although this method is not systematic but large patterns and data remains hidden [2].Competitive learning theory and model recognition are at the root of Machine Learning. In the field of data analytics these methods are used to predict a business by designing specific models and algorithms. In these ways analysts, scientists, researchers and engineers can produce reliable and valid results. Various hidden patterns or features can also find out past uses and trends in data [1]. Model automation is performed using machine learning methods. In machine learning we have to generate models on the basis of training data and test the model using the test data set. We do not need to teach machines, their functionality allows machines to adapt their methods and produce results depending on new conditions and data [2]. Stages of Machine learning are shown in figure 1.

Revised Manuscript Received on April 25, 2020.
* Correspondence Author
Rajatdeep Kaur*, Department of Computer Engineering and Technology, Guru Nanak Dev University, Amritsar, India. Email: kaur.rajatdeep@gmail.com
Kamaljit Kaur, Department of Computer Engineering and Technology, Guru Nanak Dev University, Amritsar, India. Email: kamal.aujla86@gmail.com

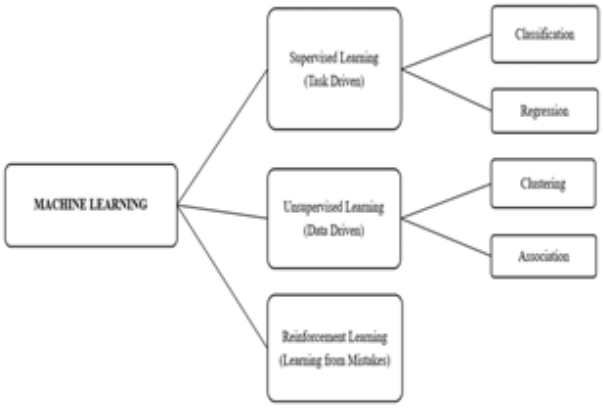

Fig. 1. Stages of Machine Learning

A. Supervised Learning

In supervised learning, Input and Output variables are provided to model for its training and allow it to analysis the relationship between these values. Then the model is expected to predict the unknown variables based on training of model. Classification and Regression are other strategies in supervised learning based on predicted variables. In regression problem, the estimated values are continuous and if the theoretical prediction comes from independent variables, called classes, it is known as the classification problem. Here are some supervised learning algorithms [1]:

- Support Vector Machine
- Decision Tree
- Naïve Bayes
- Random Forest

B. Unsupervised Learning

The purpose in unsupervised learning is to explore unknown styles. Information is organized in groups that do not have related labels. Cluster Analysis and Dimensionality Reduction are some of the least explored learning models. At Cluster we have the same intra cluster similarity and the less inter cluster similarity. In dimension Reduction, replicated and unwanted variables are emitted to generate a smaller subset of the original data [1].

C. Reinforcement Learning

In this kind of machine learning environment, Machine is trained in such a way that it will take appropriate steps to maximize the rewards in a given situation. Various software and machines are used to find the possible method or behavior must be taken in a particular situation.

In this machine you will definitely learn from what happened [3].

Machine Learning and Cardiovascular Problems: Various Medical Centers use machine learning methods to diagnose and predict problems or diseases, e.g. prediction of cancer or heart disease. Clinical parameters or data can be obtained from various health facilities and hospitals.

A major cause of human deterioration is heart disease. It is analyzed that about 17.5 million people are died due to heart illness [4]. The heart is a major part of our body and our health depends on the functioning of the heart. Heart failure can affect cardiac function [5]. Due to the rapid growth of digital technology, medical centers save chunks of data in their repositories, which becomes difficult to analysis. The process of data mining and Machine learning algorithms paly major contribution to dissect the non-identical content in medical institutions. These techniques and algorithms can work on a dataset to generate models and draw conclusions from them. Common features used to detect cardiovascular disease as shown in table I, are Age of the patient, Gender, High blood pressure, Type of Chest Pain, ECG (test used to measure cardiac electrical activity), Number of excellent fluoroscopy, Hypertension, High Cholesterol, Thalach (high-risk coronary heart disease), ST pressure (detected on the electrocardiogram), painloc (chest pain area), Accelerating level of blood sugar, smoking status, eating habits, BMI and obesity as listed below for their symptoms [6].

Table- I: Factors responsible for heart disease [6] and its symptoms [7].

Risk Factors	Description	General Symptoms:
Age	The risk of heart disease increases with age.	Chest pain, Shortness of breath, Abnormal heartbeat, Fatigue, Fainting, Swollen feet.
Gender	Men have a higher risk of having a heart problem than women	
Family History	Sometimes inherited. If blood-related relatives have heart disease then the chances of heart problems in a person increase.	
Smoking	Smokers have a higher rate of heart disease than non-smokers.	
Bad Food	Proper diet is very important for the development of a healthy heart.	
High blood pressure	The arterial blood vessels are affected by Blood Pressure [6], [7].	
Higher cholesterol levels	Plaque formation increases	
Diabetes	Patient suffers Diabetes has a high risk of heart disease	
Obesity	Obesity is a major cause of heart disease.	
Physical inactivity	Physical activities are needed for the heart to function properly.	
Stress	Pressure May damage the blood vessels	
Hygiene	Poor hygiene Increases heart disease	

Table- II: The most common types of heart disease [8]

Type	Effects
Arrhythmia	Improper heartbeat, even if it is not accompanied, very slow or very fast.
Cardiac arrest	The heart stops working, feelings of worthlessness and loss of breath occur frequently and suddenly.
Congestive heart disorder	The heart cannot pump blood and supply essential nutrients and oxygen to other organs according to need.
Congenital heart disease	Cardiac abnormalities.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Coronary artery disorder	The most important blood vein can damage any disease in the blood vessels.
High BP	Pressure is applied on the blood veins while passing through them.
Peripheral artery disease	The arterial blood vessels that slow the flow of blood to the legs.
Stroke	Interruption of blood supply occurs to the brain.

II. LITERATURE REVIEW

Ample work has suggested the use of different Data Mining (DM) techniques and Machine Learning (ML) algorithms in clinical centers used to measure Heart disease. Table III shows the parametric analysis of existing literature.

Marjia Sultana, Afrin Haider, Mohammad Shorif Uddin [9] proposed a model to predict heart disease using K-Star, J-48, SMO, Bayes Net (BN) and MLP programs in the WEKA (Waikato Environment for Knowledge Analysis) software. The UCI Machine Learning Repository is a source of data containing 270 records of which some are non-cardiac data and other are cardiovascular data. SMO and Bayes Network performs much better than the other three separators. The tests performed are based on K-fold cross validation (where k = 10).To provide the best diagnostic decision for diagnostic performance the accuracy needs to be improved.

Kanak Saxena, Richa Sharma [10] develop a functional cardiac prediction system using a decision tree. The University of California, Irvine and Cleveland Clinic Foundation are used for data collection and Dataset testing was performed using WEKA software and Order selection criteria were obtained using KEEL. Specialists and individuals in the field can make an effective decision on the basis of specific symptoms using this technique. The 10 fold method is used for testing and training the system and the accuracy for testing phase is 86.3% and the same for training phase is 87.3%.

Chala Beyene, Pooja Kamat [11] introduced a system for automatic cardiovascular disease monitoring. The program presented helps professionals with less experience and skills. ML algorithms such as the J-48, NB and SVM with K-fold Cross-proofing are used to achieve the appearance of heart failure. Distinguishing medical features such as blood sugar, age, sex and many more are useful to check whether a person has heart disease or not. Data set analysis was performed by WEKA software.

Ashok Dwivedi [12] uses different algorithms like Naïve Bayes (NB), Classification Tree, K-nearest neighbor (KNN), Linear Regression (LR), Support Vector Machine (SVM) and Artificial Neural Network (ANN) to predict heart disease. The Statlog heart disease Dataset found at UCI is used for this study. The study comprises 270 total characteristics of which 150 are free of cardiovascular disorders and 120 have cardiac disorders, while 13 different parameters are considered. Each algorithm have different accuracy, but the Logistic Regression has better value than all the different ML methods with a high accuracy of 85% and the ANN obtained the second highest accuracy.

S. Seema, Deepika kumari [13] focusing on Decision Tree (DT), NB, SVM and ANN to study the heart attack database available at UCI repositories. To calculate the optimal performance at an accurate rate, a comparative analysis was performed. In this study, it is clear that absolute accuracy i.e. 95.2% is obtained from SVM and the second highest accuracy i.e. 94.27% is available from ANN and DT has the lowest accuracy among all.

K. Polaraju, D.Durga Parsad [14] using the Multiple Regression Model for the diagnosis of cardiovascular disease by looking at different patient characteristics. The test is performed on a training data set that has 3000 attributes and the data splits in 7:3 ratio. It is understandable from the results that the Regression Algorithm had a better classification accuracy compared to others and showed that Multiple Linear Regression is better for predicting heart disease.

Jayami Patel, Tejpal Upadhyay, Samir Patel [15] compare the different techniques of DT classification looking for higher interpretations in prognosis of heart disease with the use of WEKA software. It uses J-48, Logistic Model Tree (LMT) and RF for comparisons and uses UCI data that creates cases containing 303 cases and 76 symbols. The purpose of this study, is to use data mining techniques to extract hidden patterns. The highest accuracy is obtained from the J-48 i.e. 56.76% and low LMT i.e. 55.77%. Cardiac prediction is improved by using a combination of algorithms and using more sophisticated models.

S. Prabhavathi, DM Chitra [16] introduced a systematic approach to controlling and extracting the estimated form of successful initiation of a heart attack from data sets of heart attacks. The first aim of this study is to overcome the limitations of the existing system by building a smart and expensive system. Different types of cardiovascular diseases such as CAD, CVD and CHD are predicted using the Decision Tree based Neural Fuzzy System (DNFS) techniques. The proposed system used Data Mining techniques to improve prediction of heart disease, using various data mining techniques. C4.5 performs well for CAD detection and SVM and Neural Network responds well for CHD detection. But the decision tree method is good classifier to diagnose CVD heart disease (after a factor was extracted using GA). As a result of this research study and feature reduction using GA or Principal Component analysis (PCA) can increase the accuracy of the classifier. However, DT and NB classifier are highly recommended for CVD to predict better accuracy.

Sairabi H.Mujawar, PR Devale [17] suggested DM techniques, such as NB and Modified K-methods for analyzing cardiovascular disease. The Cleveland Database for Cardiovascular Disease is used to have 300 records with 13 symptoms. The proposed model to predict the occurrence of heart attack or not and after proper analysis on the diagnosis, doctors may recommend other tests. The classification and clustering techniques are used for prognosis and the modified K-means algorithm is proposed to overcome the limit of the K-means algorithm, which needs to be integrated as input.

R. Sharmila, S. Chllammal [18] proposed a method based on Hadoop and Map Reducing programming. It uses support vector machine (SVM) in the same way to improve predictive accuracy. This study proposed using non-linear classical algorithms that provide better accuracy from which the SVM works best.

M Akhil Jabbar, Bulusu Lakshmana Deekshatulu, Priti Chandra [19], used the Random forest method and Chi Square to estimate cardiovascular disease. The proposed model was tested using different databases such as data collected from various hospitals in Hyderabad, Cleveland and Stalog data. Accurate and alternative methods have been used to make comparisons between the various ML algorithms that were PART C4.5, Naïve Bayes, Decision Table, Neural nets, and Random Forest among which proposed method (Random Forest + Chi Square) has better accuracy.

Yeshvendra K Singh, Nikhil Sinha, Sanjay K Singh [20], Random forest was used to predict cardiovascular disease. The efficacy of the proposed method is evaluated using the Cleveland Heart Disease Dataset and the 10 fold cross validation was performed to verify the authentication of accuracy which is achieved as 85.81%.

Amin UI Haq, Jain Ping Li, Muhammad Hammad Memon, Shah Nazir, Ruinan Sun [21], developed an HD diagnostic system using other ML and Feature Selection Algorithms. Various performance metrics such as category accuracy, precision, sensitivity, Mathew accuracy and performance time are analyzed. The proposed system compared performance on full features, and reduced features and calculated that results on feature reduction data have better accuracy and reduced execution time.

Subhashini Narayan, E Sathiyamoorthy [22], used the Fourier Transform and Machine Learning Machine to accurately predict chronic diseases. The proposed system used a bootstrap model and three different algorithms namely ANN, NB and SVM for evaluation.

Abderrahmane Ed- Daoudy, Khalil Maalmi [23], presented an RT-HD prognosis program based on Apache Spark. The program consists of various stages such as stream broadcasting and DS and visualization. Spark MLlib for spark propagation was used for streaming processing and Apache Cassandra was used by DS and Visualization to accumulate a large amount of data.

Abdar et al. [24], introduced a new N2-Genetic Optimizer system that uses SVM integration with GA or PSO. In this paper various ML Algorithms were used to evaluate the Z-Alizadeh Saini cardiovascular disease, out of which SVM is best selection method and was used for further experiments. Preparation of data reconstruction was performed using Normalization and GA or PSO was used for feature selection and minimization of asset replication. As a result of this process the GA + nuSVM has better accuracy than the PSO + nuSVM in both test and training data.

Miray Akgul, Ozlen Ekal Sonmez, Tunacy Ozcan [25], uses the Artificial Neural Network to monitor the Heart Problem using the default parameters. Subsequently, the combination of GA (Genetic Algorithm) and ANN (Artificial Neural Network), tested a hybrid method using the cardiac database 'Cleveland' and obtained better accuracy than ANN and other ML algorithms.

Table- III: A Parametric Analysis of Existing Literature

Year	Reference no.	Purpose	Technique used	Accuracy
2015	17	Used the Modified K-means and NB to Predict Heart Disease.	K-means algorithm and NB algorithm	Heart Disease detection=93% Heart Disease Not detected = 89%
2016	10	Create an Effective Heart Disorder prognosis system using DT algorithm.	DT Classifier	Testing data set =86.3% Training data set=87.3%
2016	9	Prognosis of Heart Illness using WEKA Tool.	K Star	75%
			J48	86%
			Bayes Net	87%
2016	13	Chronic Disease Prediction by data Mining	Naïve Bayes	93.85%
			ANN	94.27%
			Decision Tree	92.59%
			SVM	95.2%

2016	12	Assess the execution of distinct ML algorithms for Heart Illness Prognosis.	NB	83%
			KNN	80%
			LR	85%
			Classification Tree	77%
2016	15	ML and DM Techniques are used to predict Heart Disease.	J48	56.76%
			Logistic Model Tree	55.75%
2016	16	DNFS Techniques are used to Analysis and Predict different Heart disorders.	DT, c4.5, SVM, NB	85% to 99% for CVD 82% to 92% for CHD
2016	19	Use of Random Forest and Feature Selection for prediction of Heart Disease	Random Forest and Chi Square	83.70%
			Naïve Bayes	78.56%
			Decision Tree	82.43%
			PART C4.5	75.4%
			Neural nets	82.77%
2016	20	Heart Disease prediction using Random Forest	Random Forest	85.81
2018	11	Use of DM Techniques to Predict and Analysis the happening of Heart Disorders.	J48, NB, SVM	The result helps to lessen the cost of diagnosis and provides appropriate results
2018	18	Big Data and DM techniques can intensify the prognosis of Heart Disorders.	SVM in parallel fashion	Parallel SVM =85% and Sequential SVM = 82.35%.
2018	21	Hybrid Framework of Feature selection and Machine Learning for the prediction of heart illness	Logistic Regression	84%
			SVM	86%
			Naïve Bayes	83%
			Artificial Neural Network	73%
			Decision Tree Classifier	74%
			K- Nearest Neighbor	83%
2018	22	FFT based system with Machine Learning for predicting and Identify Heart Disease	Fourier Transformation and Machine Learning	93%
2019	23	Early detection of heart disease using real time Machine Learning and Big data Approach.	Big Data + Machine Learning	87.50%
2019	24	Accurate Diagnoses of CAD using Machine Learning Approach	GA+SVM	93.08%
			PSO + SVM	91.13%
2019	25	Hybrid ANN and GA approach for heart disease diagnosis	GA+ANN	95.82%

III. RESULT ANALYSIS

Various algorithms were used for the prediction of heart disease, it is analyzed from above literature work that each algorithm results into different performance accuracy that effects the outcome of a proposed model. As shown in Figure 2 and Table IV, SVM depicts the highest accuracy among all classifiers, whereas KNN has the lowest performance.

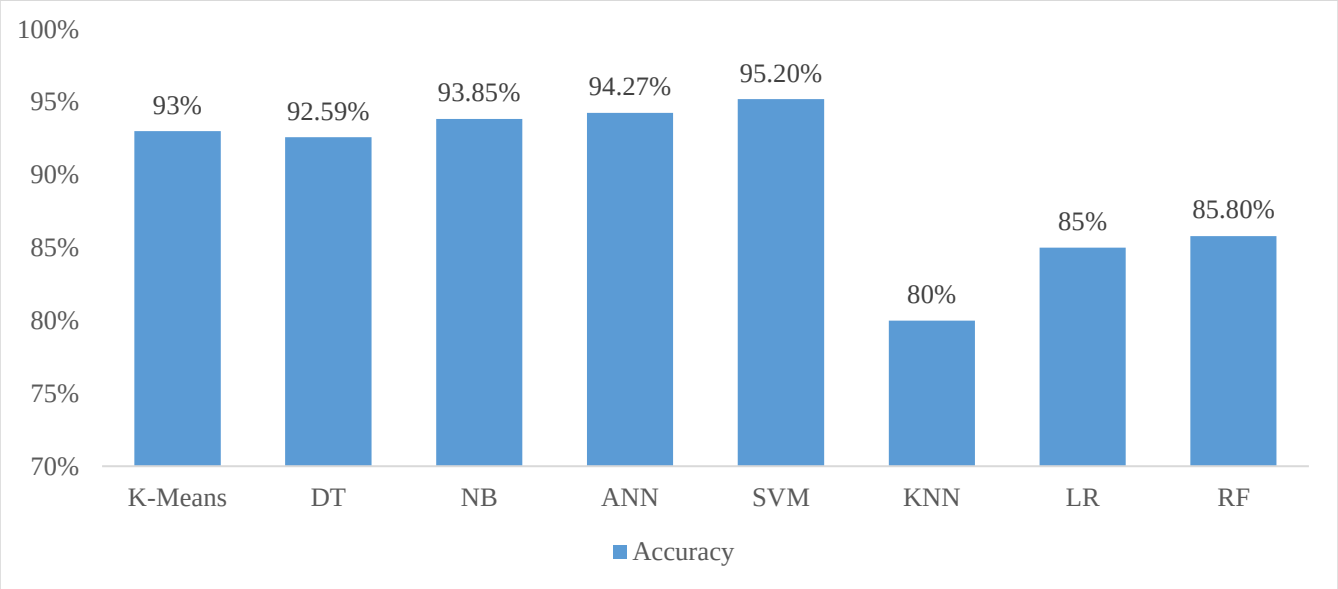

Fig. 2. Shows the performance of Machine Learning Algorithms

Table-IV: Performance of Various Machine Learning Algorithms to detect heart disease.

Algorithm	Accuracy (in Percentage)
K-Means	93%
Decision Tree	92.59%
Naïve Bayes	93.85%
Artificial Neural Network	94.27%
Support Vector Machine	95.20%
K- Nearest Neighbor	80%
Linear Regression	85%
Random Forest	85.80

The accuracy of SVM is observed as 95.20%, however the accuracy of other algorithms that are K-Means, Decision Tree, Naïve Bayes and Artificial Neural Network is fluctuating from 92.59% to 94.27 %. While other algorithms have accuracy less than 90%. Cardiovascular disease can easily predict from the model having best performance.

IV. CONCLUSION

It is recognized that many data mining and machine learning algorithms are used to examine the presence of cardiac disorders using different data sets. It is analyzed that SVM is the most efficient and KNN has the least accuracy. The results can be improved by using proper feature selection methods. Better and more accurate results can help local specialists and others associated with the field be better prepared for high quality care and provide the patient with the right treatment. So early detection helps to provide quality service and saves the health of the individual.

REFERENCES:

1. Rätsch, Gunnar. "A brief introduction into machine learning." *Friedrich Miescher Laboratory of the Max Planck Society* (2004).
2. Alpaydin, Ethem. *Introduction to machine learning*. MIT press, 2020.
3. Sutton, Richard S. "Introduction: The challenge of reinforcement learning." In *Reinforcement Learning*, pp. 1-3. Springer, Boston, MA, 1992.
4. Learning, Machine. "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review." *Advances in Computational Sciences and Technology* 10, no. 7 (2017): 2137-2159
5. Krishnaiah, V., G. Narsimha, and N. Subhash Chandra. "Heart disease prediction system using data mining techniques and intelligent fuzzy approach: A review." International Journal of Computer Applications 136, no. 2 (2016): 43-51
6. Srinivas, K., G. Raghavendra Rao, and A. Govardhan. "Survey on prediction of heart morbidity using data mining techniques." *International Journal of Data Mining & Knowledge Management Process (IJDKP)* 1, no. 3 (2011): 14-34.
7. Heart Disease symptoms caused by heart infections http://www.mayoclinic.org/diseasesconditions/heart-disease/symptoms-causes/dxc-20341558
8. Types of Heart Disease that effect different organs https://www.medicalnewstoday.com/articles/237191
9. Sultana, Marjia, Afrin Haider, and 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)*, pp. 1-5. IEEE, 2016.
10. Saxena, Kanak, and Richa Sharma. "Efficient heart disease prediction system using decision tree. "In International Conference on Computing, Communication & Automation, pp.72-77. IEEE, 2015

11. Beyene, Mr Chala, and Pooja Kamat. "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques." *International Journal of Pure and Applied Mathematics* 118, no. 8 (2018): 165-174.
12. *Dwivedi, Ashok Kumar. "Performance evaluation of different machine learning techniques for prediction of heart disease." Neural Computing and Applications 29 (2016): 685-693.*
13. Deepika, Kumari, and S. Seema. "Predictive analytics to prevent and control chronic diseases." In *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, pp. 381-386. IEEE, 2016.
14. K Polaraju, D.Durga Prasad, "Prediction of Heart Disease using Multiple Linear Regression Model", International Journal of Engineering Development and Research (IJEDR), ISSN:2321-9939, Volume.5, Issue 4, pp.1419-1425, December 2017
15. Patel, Jaymin. "Prof. Tejal Upadhyay, and Dr. Samir Patel, Sep 2015-Mar 2016,"." *Heart Disease Prediction using Machine Learning and Data Mining Technique* 7, no. 1: 129-137.
16. Prabhavathi, S., and D. M. Chitra. "Jan. 2016,"Analysis and Prediction of Various Heart Diseases using DNFS Techniques"." *International Journal of Innovations in Scientific and Engineering Research (IJISER)* 2, no. 1: 1-7.
17. Mujawar, Sairabi H., and P. R. Devale. "Prediction of Heart Disease using Modified K-means and by using Naïve Bayes." *International Journal of Innovative research in Computer and Communication Engineering* 3 (2015): 10265-10273.
18. R. Sharmila, S. Chellammal , "*A conceptual method to enhance the prediction of heart diseases using big data Techniques*", International Journal of Computer Sciences and Engineering, Vol.06, Issue.04, pp.21-25, 2018.
19. Jabbar, M. Akhil, Bulusu Lakshmana Deekshatulu, and Priti Chandra. "Prediction of heart disease using random forest and feature subset selection." In *Innovations in Bio-Inspired Computing and Applications*, pp. 187-196. Springer, Cham, 2016.
20. Singh, Yeshvendra K., Nikhil Sinha, and Sanjay K. Singh. "Heart Disease Prediction System Using Random Forest." In *International Conference on Advances in Computing and Data Sciences*, pp. 613-623. Springer, Singapore, 2016.
21. HAQ, Amin Ul, et al. A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mobile Information Systems*, 2018, 2018.
22. 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.
23. 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. p. 1-5.
24. Abdar, Moloud, Wojciech Książek, U. Rajendra Acharya, Ru-San Tan, Vladimir Makarenkov, and Paweł Pławiak. "A new machine learning technique for an accurate diagnosis of coronary artery disease." *Computer methods and programs in biomedicine* 179 (2019): 104992.
25. Akgül, Miray, Özlen Erkal Sönmez, and Tuncay Özcan. "Diagnosis of Heart Disease Using an Intelligent Method: A Hybrid ANN–GA Approach." In *International Conference on Intelligent and Fuzzy Systems*, pp. 1250-1257. Springer, Cham, 2019.

AUTHORS PROFILE

Rajatdeep Kaur, is a student of Masters of Technology, Computer Engineering and Technology at Guru Nanak Dev University, Amritsar, India. Her main research area is Machine Learning and Data Mining. She is Gold medalist in Bachler's of Technology from Global Institutes of Management and Emerging Technologies, Amritsar, India.

Kamaljit Kaur, was born in Kapurthala, Punjab, India. She completed B.Tech with Distinction from Punjab Technical University, and M.Tech with Gold Medal from Guru Nanak Dev University, Amritsar. She has submitted PhD in the field of Cloud Computing from Guru Nanak Dev University, Amritsar. From 2007 to 2008 she was employed with Dr. B.R Ambedkar NIT, Jalandhar as Lecturer and from 2010 to 2012, she worked as an assistant professor at Lovely Professional University, Jalandhar. She joined Guru Nanak Dev University, Amritsar in July 2012 where she is currently working as Assistant Professor. Her research Interests are in Resource Provisioning in Cloud Computing, Resiliency in Cloud Computing and Distributed Systems, Big Data, IoT. Kamaljit Kaur has published and presented more than 45 papers in scientific journals and international conferences.