International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2020 IJSRCSEIT | Volume 6 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT2062136

Study of Machine Learning Algorithms for Prediction and Diagnosis of Cardiovascular Diseases : A Review

Manoj Patil[1], Dr. Harsh Mathur[2]
[1]PhD Scholar, Department of Computer Science and Engineering, MPU Bhopal, Madhya Pradesh, India
[2]Associate Professor Department of Computer Science and Engineering, MPU Bhopal, Madhya Pradesh, India

ABSTRACT

We are living in a post modern era and there are tremendous changes happening to our daily life which make an impact on our health positively and negatively. As a result of these changes various kind of diseases are enormously increased. In the medical field, the diagnosis of cardiovascular disease is the most difficult task. The diagnosis of cardiovascular disease is difficult as a decision relied on grouping of large clinical and pathological data. Due to this complication, the interest increased in a significant amount between the researchers and clinical professionals about the efficient and accurate heart disease prediction. In case of heart disease, the correct diagnosis in early stage is important as time is the very important factor. Heart disease is the principal source of deaths widespread, and the prediction of Heart Disease is significant at an untimely phase. Machine learning in recent years has been the evolving, reliable and supporting tools in medical domain and has provided the greatest support for predicting disease with correct case of training and testing. This research paper intends to provide a survey of techniques of knowledge discovery in databases using data mining techniques that are in use in today's medical research particularly in Cardiovascular Disease Prediction.

Keywords : Cardiovascular disease, Heart, Machine learning, Prediction, Training, Data Mining.

I. INTRODUCTION

As per statistics given by WHO (World Health organization) globally cardiovascular diseases (CVDs) are the number one cause of death as more people die annually from CVDs than from any other cause. An estimated 17.9 million people died from CVDs in 2016, representing 31% of all global deaths. Of these deaths, 85% are due to heart attack and stroke. Over three quarters of CVD deaths take place in low- and middle-income countries. Out of the 17 million premature deaths (under the age of 70) due to non communicable diseases in 2015, 82% are in low- and middle-income countries, and 37% are caused by CVDs. Most cardiovascular diseases can be prevented by addressing behavioral risk factors such as tobacco use, unhealthy diet and obesity, physical inactivity and harmful use of alcohol using population-wide strategies [1].

Heart is an important part of human body. Life depends on an efficient working of heart. If working of heart is not good then it will affects the other parts of our human body like kidney and brain. Heart disease is predicted based on the performance of heart. Some of the factors that are used to predict heart diseases are:

- High blood pressure
- Cholesterol
- Lack of physical exercise
- Smoking
- Obesity
- Family history of heart disease

As heart disease is the major cause of human deaths, predictions should be taken to reduce the risk of heart disease. Generally, doctors will diagnosis heart disease based on the symptoms and physical examination of the patient body. Heart disease prediction is a difficult task in healthcare industry. Nowadays, healthcare industry contains huge amount of data of patients, disease diagnosis, electronic patient records and medical devices. It is a key resource that needs to be processed while knowledge extraction and it will support decision making process [20].

Heart disease is an umbrella term that includes many types, including congenital, coronary, and rheumatic heart diseases. Among those conditions, coronary heart disease is the most common, causing over 360,000 American deaths due to heart attacks in 2015. According to the Centers for Disease Control and Prevention, it is estimated that approximately every 40 seconds, an American experiences a heart attack. Consequently, heart disease expenditures have risen to over $200 billion annually in the United States alone [18]. Furthermore, by 2030, health care costs due to heart disease are expected to double according to the American Heart Association [19].

Some of the most common heart diseases are listed in the table below with their description [21] [22].

TABLE – 1 : TYPES OF HEART DISEASE

Types of heart diseases	Description
Angina	Chest pain due to a deficiency of blood to the heart muscle
Acute coronary syndrome	Blood supply to the heart muscle is swiftly obstructed
Arrhythmia	Atypical heart rhythm
Cardiomyopathy	Heart muscle disease
Congenital heart disease	Heart disfigurements that are present at birth
Coronary artery disease	Occurred when Arteries supplying blood becomes obstructed
Rheumatic heart diseases	Rheumatic fever

There are also different heart disease factors, from that most common are listed in the table below with their symptoms [21] [22].

TABLE - 2: FACTORS OF HEART DISEASE

Risk factors	Description	General Symptom
Age	Old people are more suffers from heart disease	Chest pain Shortness of breath Irregular heartbeat Fatigue Fainting Swollen feet
Sex	Males are at greater risk than females	
Family History	If relatives have heart disease the probability of a person to have cardiovascular disease is high	
Blood Pressure	Blood pressure can effect in narrowing hardening arteries, as well as thickening blood vessels[1], [2].	
Smoking	Heart disease higher in smokers than nonsmokers people	

Poor Diet	Diet food is essential for development of heart	
High Blood Cholesterol Level	It increases formation of plaques	
Diabetes	It is the disease as a result of sugar in our body	
Obesity	Overweight body is one of the cause for heart diseases	
Physical Inactivity	Physical activity helps heart to function properly	
Stress	Damage arteries	
Poor Hygiene	It increases heart disease	

Prediction is a good methodology in healthcare centers where clinicians do not have more knowledge and skill as well as where there are no specialists, for instance, such clinicians may give their own decision that may give poor result and lead the patients to death. Prediction of heart disease is used for automatic diagnosis of the disease and give sufficient qualities of services in healthcare centers to save the life of individuals. Prediction technique helps to make an accurate decision for the stakeholders, particularly for specialists to give reasonable decision to treat patients [4]. Prediction of cardiovascular disease is challenging and more complicated task to achieve an automatic diagnosis of sickness. Because an enormous amount of data are stored in healthcare centers that are very complex and challenging to analyses. Even if it is challenging task using prediction of heart diseases in medical centers is plays significant roles to save the lifestyle of individuals and to make active and accurate decision-making for stakeholders [4].

Medical data mining has played important role for exploring the hidden patterns in the data sets of the medical domain. These patterns can be utilized for clinical diagnosis. However, the available raw medical data are widely distributed, heterogeneous in nature, and voluminous. These data need to be collected in an organized form and can be then integrated to form a hospital information system. Data mining technology provides a user oriented approach to novel and hidden patterns in the data.

II. LITERATURE REVIEW

The objective of this review portion is to explore the work happening on the classification and feature selection process for the prediction of CVD and to establish the extent and depth of existing knowledge on CVD prediction process.

Introduction

Medical organizations, all around the world, collect data on various health related issues. These data can be exploited using various machine learning techniques to gain useful insights. But the data collected is very massive and, many a times, this data can be very noisy. These datasets, which are too overwhelming for human minds to comprehend, can be easily explored using various machine learning techniques. Thus, machine Learning algorithms have become very useful, in recent times, to predict the presence or absence of heart related diseases accurately. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart related diseases [3].

A. Literature review on prediction of CVD

N. Satish Chandra Reddy et al. [5] This paper is focused on the classification and feature selection required for

the prediction of Heart disease using various machine learning algorithms. The paper uses KNN, SVM, Random Forest, Naïve Bayes and Neural Network. The algorithms used in the paper gives the better result with average accuracy of 85.92-89.41% over the combined dataset.

The Marjia et al.[7] Propose heart disease prediction using K Star, J48, SMO, and Bayes Net and Multilayer perceptron using WEKA software. Based on performance from different factor SMO (89% of accuracy) and Bayes Net (87% of accuracy) achieve optimum performance than KStar, Multilayer perceptron and J48 techniques using k-fold crossvalidation. The accuracy performance achieved by those algorithms are still not satisfactory. So that if the performance of accuracy is improved more to give better decision to diagnosis disease.

The Azam et al.[8] The paper describe automatic diagnosis of coronary artery disease (CAD) patients using optimized SVM, in this parameters of SVM are optimized to improve the accuracy of prediction, which gives an accuracy of 99.2% using k-fold cross-validation. The paper helps to diagnosis disease at an early stage and to reduce the cost. The accuracy obtained is good to predict if the individual has heart disease or not.

The Cemil et al. [9] Propose application of knowledge discovering process on prediction of stroke patients based on Artificial Neural Network (ANN) and Support Vector Machine (SVM), which give accuracy of 81.82% and 80.38% for ANN and SVM respectively for training data set and 85.9% and 84.26% for Artificial Neural Network (ANN) and Support Vector Machine (SVM) in test dataset respectively. ANN show more accurate result than Support Vector Machine (SVM) for proposed work. The accuracy obtains by the paper is not enough to prediction the stroke patients.

The Shailendra Narayan Singh et al. [10] Prediction of heart disease using techniques of data mining, the paper describes different classifier merits and demerits for data classification and knowledge extraction to implement algorithms that are most useful in health organizations. The paper is a help to understand how and when to use the algorithms based on the merits and demerits given.

The Sanavar et al. [11] The description of survey paper on heart disease prediction. It describes the different methodology and the way in which proposed methods are implemented. It also provides some overview of heart disease, as well as the role of data mining in healthcare centers and how to apply oruse data mining in a healthcare organization, is explained

The Vidya K. Sudarshan et al. [12] The paper is focused on the Application of higher-order spectra for the characterization of coronary artery disease using electrocardiogram signals. The paper use K-Nearest Neighbor and Decision Tree. The performance of accuracy calculated by this algorithms is 98.17% and 98.99% are obtained respectively. The algorithms used in the paper give the better result of the performance of accuracy to characterize the coronary artery disease.

The Emrana KabirHashi,[13] An expert clinical decision support system to predict disease using classification techniques. The paper is based on WEKA software and percentage ratio method for train and test dataset using C4.5 and KNN, They give 90.43% and 76.96% accuracy respectively. C4.5 Decision Tree gives better accuracy compared to KNN and helps for clinical decision support system.

The Megha Shahi et al.[14] The objective of the paper is Heart disease prediction system using data mining techniques and WEKA software for automatic diagnosis of disease and to give qualities of services in healthcare centers. The paper used various algorithms like SVM, Naïve Bayes, Association rule, KNN, ANN

and Decision Tree. The paper explains that in some paper SVM effective and efficient accuracy about 85% as compared to other data mining algorithms.

The Priti Chandra et al.[15] The paper describes Computational Intelligence Technique for Early Diagnosis of Heart Disease using WEKA and 10-Fold cross-validation. The algorithm used in this research paper is Naïve Bayes which give 86.29% accuracy.The accuracy obtained is good, but there is not satisfactory to the automatic diagnosis of heart disease.

The Syed Muhammad Saqlain Shah et al.[16] The paper is proposed for the analysis of Heart Disease Diagnosis based on feature extraction using K-Fold cross-validation. The algorithm used by these paper is SVM which gives 91.30%. The algorithm accuracy is better for prediction of heart disease and for automatic diagnosis of the disease.

The Muhammad Saqlain et al.[17] The paper is focused on identification of Heart Failure by using unstructured Data of Cardiac Patients. The paper uses Logistic Regression, Neural Network, SVM, Random Forest Decision Tree and Naïve Bayes. The algorithms achieve accuracy of 80%, 84.8%, 83.8%, 86.6%, 86.6% and 87.7% respectively for each individual's algorithm. Naïve Bayes provide the highes taccuracy compared to others algorithms.

B. Literature review on algorithms

K-Nearest Neighbor's Algorithm

It is a simple classifier that cannot handle noises, easy to implement and understand, requires short training time and whole training set is used for prediction. The K-Nearest Neighbors (K-NN) with weighting parameter has been used for the prediction of heart disease. The similar techniques (KNN) have been used along with the feature selection technique such as particle swarm optimization (PSO) for the heart disease prediction where it shows much better accuracy

Support Vector Machine (SVM)

Based on the kernel functions the support vector classifiers are divided into different types such as linear, nonlinear, radial basis function (RBF), sigmoid and polynomial. The hyperplane or support vector machine separates the support vector or data points. Binary SVMs are classifiers which discriminate data points of two categories. Each data object (or data point) is represented by an n dimensional vector. Each of these data points belongs to only one of two classes. A linear classifier separates them with a hyper plane. There exist many hyper planes which can separate the data samples. In order to achieve maximum separation between the two classes, SVM picks the hyper plane which has the largest margin. The margin is the summation of the shortest distance from the separating hyper plane to the nearest data point of both categories. Such a hyper plane is likely to generalize better, meaning that the hyper plane correctly classify unseen or testing data points. SVM does the mapping from input space to feature space to support nonlinear classification problems.

Random Forest

Random forest constructs a multitude of decision trees at training time and output the mode prediction of the classes for classification and the mean prediction for regression. The different patterns in the data are evaluated by the decision tress. The class prediction is based on the majority vote for classification.

Naive Bayes Classifier

Naive Bayes is a classification algorithm based on bayes theorem. Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even

highly sophisticated classification methods. It assumes that the presence or absence of particular feature of a class is unrelated to the presence or absence of any other feature. It is based on conditional probabilities. An advantage of the Naive Bayes classifier is that it requires only a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Since independent variables are assumed, only the variances of the variables for each class need to be determined. It can be used for both binary and multiclass classification problems.

Decision Tree

Decision tree learning is one of the most widely used and practical machine learning methods for inductive inference. It is method for approximating discrete valued functions that is robust to noisy data and capable of learning disjunctive expressions. This learned function is represented by decision tree. Tree models where the target variable can take a finite set of values are called classification trees. In these tree structures, leaves represent classification results and branches represent conjunctions of features that point to those classification results.

III. OBSERVATION

Statistical models for estimation that are not capable to produce good performance results have flooded the assessment area. Statistical models are unsuccessful to hold categorical data, deal with missing values and large data points. All these reasons arise the importance of MLT. ML plays a vital role in many applications, e.g. image detection, data mining, natural language processing, and disease diagnostics. In all these domains, ML offers possible solutions. This paper provides the survey of different machine learning techniques for diagnosis of different diseases such as heart disease, diabetes disease, liver disease, dengue and hepatitis disease. It is observed that for the detection of heart disease, SVM provides improved accuracy. Survey highlights the advantages and disadvantages of these algorithms. Improvement graphs of machine learning algorithms for prediction of Diseases. From analysis, it can be clearly observed that these algorithms provide enhanced accuracy on different diseases so provide opportunity for the improved decision making process [2].

Based on the above review, it can be concluded that there is a huge scope for machine learning algorithms in predicting cardiovascular diseases or heart related diseases. Each of the above-mentioned algorithms have performed extremely well in some cases but poorly in some other cases which could be due to overfitting. Random Forest and Ensemble models have performed very well because they solve the problem of overfitting by employing multiple algorithms i.e. multiple Decision Trees in case of Random Forest. Systems based on machine learning algorithms and techniques have been very accurate in predicting the heart related diseases [3].

The heart disease prediction has been evaluated with the classification and feature selection algorithms implemented in CARET package of R tool using combined dataset. The accuracy of the model depends on the database, preprocessing, analytical tools and techniques. It is important to select minimum and prominent attributes to improve the performance when compared to the use of whole features from the dataset. The highest accuracy is shown by random forest in three percentage split (without and with feature selection). This study shows that the random forest can be used as a good classification algorithm for the accurate prediction of heart disease with an accuracy of 90–95 %. The less variation of accuracy differences between dataset and selected features (8 and 6) indicates these features can be useful for the prediction of heart diseases [5].

The diagnosis of heart disease is difficult as a decision relied on grouping of large clinical and pathological

data. This paper express idea to study diverse prediction models for the heart disease and selecting important heart disease feature using genetic algorithm. In this work, different prediction models were studied and the experiments are conducted to find the best classifier for predicting the heart disease. Four classifiers Random Forest, Naïve Bayes, Decision Tree, Support Vector Machine were used for prediction of patients with heart diseases. Observation shows that in most of the cases Naive bayes classifier performance is having more accuracy. The other observation from this study is that, the genetic algorithm feature selection technique suggest the most important attributes for heart diseases The result also shows that integration genetic algorithm with prediction models improves the performances of the models[6].

IV. PROPOSED STUDY

Representation of proposed system:

Machine learning is the subfield of AI study. There are many types of Machine Learning Techniques i.e. Supervised, Unsupervised, Semi Supervised, Reinforcement, Evolutionary Learning and Deep Learning.

1) Supervised learning: Offered a training set of examples with suitable targets and on the basis of this training set, algorithms respond correctly to all feasible inputs Learning from exemples is another name of Supervised Learning. Classification and regression are the types of Supervised Learning.
Classification: It gives the prediction of Yes or No, for example, “Is this tumor
cancerous?”, “Does this cookie meet our quality standards?”
Regression: It gives the answer of “How much” and “How many”.

2) Unsupervised learning: Correct responses or targets are not provided. Unsupervised learning technique tries to find out the similarities between the input data and based on these similarities, un-supervised learning technique classify the data. This is also known as density estimation. Unsupervised learning contains clustering.
Clustering: it makes clusters on the basis of similarity.
3) Semi supervised learning: Semi supervised learning technique is a class of supervised learning techniques. This learning also used unlabeled data for training purpose (generally a minimum amount of labeled-data with a huge amount of unlabeled-data). Semi-supervised learning lies between unsupervised-learning (unlabeled-data) and supervised learning (labeled-data). .
4) Reinforcement learning: This learning is encouraged by behaviorist psychology. Algorithm is informed when the answer is wrong, but does not inform that how to correct it. It has to explore and test various possibilities until it finds the right answer. It is also known as learning with a critic. It does not recommend improvements. Reinforcement learning is different from supervised learning in the sense that accurate input and output sets are not offered, nor suboptimal actions clearly précised. Moreover, it focuses on on-line performance.

5) Evolutionary Learning: This biological evolution learning can be considered as a learning process: biological organisms are adapted to make progress in their survival rates and chance of having off springs. By using the idea of fitness, to check how accurate the solution is, we can use this model in a computer

6) Deep learning: This branch of machine learning is based on set of algorithms. In data, these learning algorithms model high-level abstraction. It uses deep graph with various processing layer, made up of many linear and onlinear transformation.

The huge amounts of data produced by different sources have become a fundamental importance for capturing, storing, searching, sharing, and are hard to interpret and analyze. The huge volume of data and the increase in diagnosis cost made to look for feature selection which in turn increases the accuracy of the model, and give a better result for the prediction of disease.

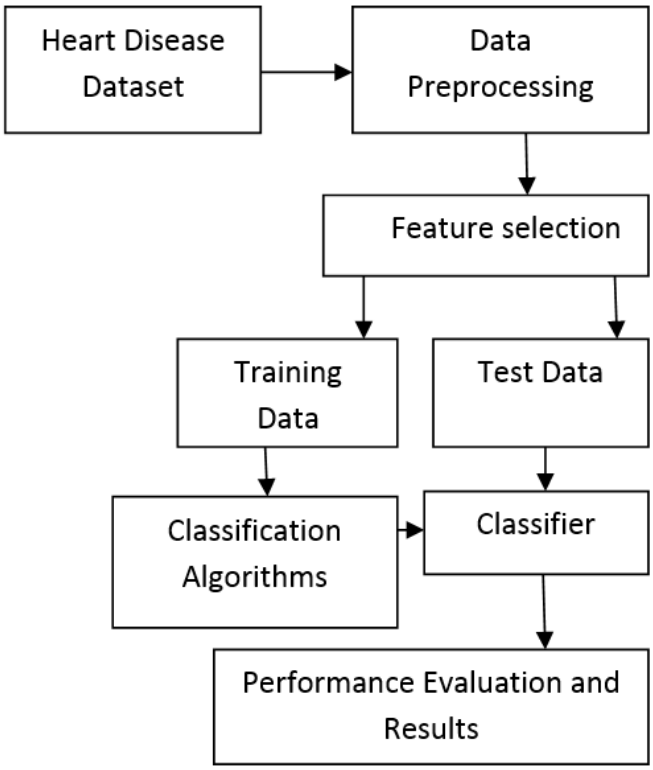

Figure 1: Proposed framework for prediction of CVD

Data Preprocessing:

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. Data preprocessing prepares raw data for further processing.

In the data preprocessing stage, missing values are replaced with mode value based on the particular datasets source. Second, taking into consideration that heart disease patients might have high values of respective attributes (i.e., referred as outliers in the dataset) are not removed. The normalization (normalize <- function(x) {return ((x - min (x)) / (max(x) - min(x)))} has been carried out since dataset consists of different measuring units. The initial dataset consists of number of attributes, some of them may not be useful thus it is necessary to remove them during data preprocessing.

Feature Selection:

One of the important steps in data preprocessing is feature selection, by this unnecessary features can be removed and improve the performance to build a better classification model. The feature selection is performed on the dataset to select a subset of relevant features for model building that aims to improve model accuracy. Feature Selection is an efficient data preprocessing technique in data mining for reducing dimensionality of data. In medical diagnosis, it is very important to identify most significant risk factors related to disease. Relevant feature identification helps in the removal of unnecessary, redundant attributes from the disease dataset which, in turn, gives quick and better results

Classification:

After data normalization to build a classification model, the dataset with their attributes is divided into training and testing data. Classification and prediction is a data mining technique which first uses training data to develop a model and then the resulted model is applied on testing data to get results of prediction. Various classification algorithms like K-Nearest Neighbors (K-NN), *Support Vector Machine (SVM), Random Forest,* Naïve bayes have been applied on disease datasets for the diagnosis of disease. There is an utmost need to develop a novel classification technique which can expedite and simplify the process

of diagnosis of disease. The performance of a model on test data is calculated by accuracy, sensitivity/recall, and specificity. Sensitivities and specificity measures the true positives (risk class) and the true negatives (normal class) respectively. Thus the predictive capabilities of the classifiers are measured by sensitivity and specificity values.

V. CONCLUSION

From the knowledge of reviewed literature it has been found that the heart disease is said to be major causes of death globally. The group of diseases related to both the heart and blood vessels are referred as cardiovascular disease (CVD). Also the diagnosis of heart disease is difficult as a decision relied on grouping of large clinical and pathological data. The main idea behind this study is to find out diverse prediction models for the cardiovascular disease (CVD) and selecting important disease feature using machine learning algorithm.

The proposed work need to consider risk factors like hypertension and family history as a predictor and use selected attributes for efficient heart disease prediction. The presence and absence of heart disease are predicted using various classification models. The performance of the prediction models are measured using various measures such as accuracy, sensitivity and specificity

VI. REFERENCES

[1]. M. Fatima, M. Pasha" Survey of Machine Learning Algorithms for Disease Diagnostic" Journal of Intelligent Learning Systems and Applications, 2017, 9, 1-16

[2]. V.V. Ramalingam, Ayantan Dandapath, M Karthik Raja "Heart disease prediction using machine learning techniques: A Survey" International Journal of Engineering & Technology, 7 (2.8), 2018, 684-687

[3]. Mr. Chala Beyene, Pooja Kamat "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques" International Journal of Pure and Applied Mathematics , ijpam Volume 118 No. 8 ,2018, 165-174

[4]. N. Satish Chandra Reddy, Song Shue Nee, Lim Zhi Min & Chew Xin Ying" Classification and Feature Selection Approaches by Machine Learning Techniques: Heart Disease Prediction."International Journal of Innovative Computing , IJIC Vol. 9:1, 2019,39-46

[5]. P. Suresh and M.D. Ananda Raj "Study and Analysis of Prediction Model for Heart Disease: An Optimization Approach using Genetic Algorithm" International Journal of Pure and Applied Mathematics , ijpam, Volume 119, No. 16, 2018, 5323-5336

[6]. M. Sultana, A. Haider, and M. S. Uddin, "Analysis of data mining techniques for heart disease prediction," 2016 3rd Int. Conf. Electr. Eng. Inf. Commun. Technol. iCEEiCT 2016, 2017.

[7]. A. Davari Dolatabadi, S. E. Z. Khadem, and B. M. Asl, "Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM," Comput. Methods Programs Biomed., vol. 138, 2017,pp. 117–126.

[8]. C. Colak, E. Karaman, and M. G. Turtay, "Application of knowledge discovery process on the prediction of stroke," Comput. Methods Programs Biomed., vol. 119, no. 3,2015, pp. 181–185.

[9]. M. Gandhi, "Predictions in Heart Disease Using Techniques of Data Mining," Int. Conf. Futur. Trend Comput. Anal. Knowl. Manag., 2015.

[10]. S. Kiruthika Devi, S. Krishnapriya, and D. Kalita, "Prediction of heart disease using data mining techniques," Indian J. Sci. Technol., vol. 9, no. 39,2016 pp. 21–24.

[11]. U. R. Acharya et al., "Application of higher-order spectra for the characterization of

Coronary artery disease using electrocardiogram signals," Biomed. Signal Process. Control, vol. 31,2017, pp. 31–43

[12]. E. K. Hashi, M. S. U. Zaman, and M. R. Hasan, "An expert clinical decision support system to predict disease using classification techniques," 2017 Int. Conf. Electr. Comput. Commun. Eng.,2017 pp. 396–400.

[13]. M. Shahi and R. Kaur Gurm, "Heart disease prediction system using data mining techniques," Orient. J Comput. Sci. Technol., vol. 6, no. 4, 2013,pp. 457–466.

[14]. Priti Chandra; M. A. Jabbar ; B. L. Deekshatulu "Computational intelligence technique for early diagnosis of heart disease," IEEE International Conference on Engineering and Technology (ICETECH), 10.1109/ICETECH.2015.7275001.

[15]. S. M. S. Shah, S. Batool, I. Khan, M. U. Ashraf, S. H. Abbas, and S. A. Hussain, "Feature extraction through parallel Probabilistic Principal Component Analysis for heart disease diagnosis," Phys. A Stat. Mech. Its Appl., vol. 482, 2017,pp. 796–807.

[16]. M. Saqlain, W. Hussain, N. A. Saqib, and M. A. Khan, "Identification of Heart Failure by Using Unstructured Data of Cardiac Patients," 2016 45th Int. Conf. Parallel Process. Work.,2016, pp. 426–431.

[17]. Centers for Disease Control and Prevention (CDC), "Heart disease fact sheet" https://www.cdc.gov/dhdsp/data_statistics/fact_sheets/fs_heart_disease.htm, 2015.

[18]. E. J. Benjamin, S. S. Virani, C. W. Callaway, A. R. Chang, S. Cheng, S. E. Chiuve, M. Cushman, and F. N. Delling, et. al., "Heart disease and stroke statistics-2018 update: a report from the American Heart Association," Circulation, Vol. 137, No. 12, pp. 67- 492, 2018.

[19]. T. Revathi, S. Jeevitha " Comparative Study on Heart Disease Prediction System Using Data Mining Techniques" International Journal of Science and Research (IJSR) ISSN (Online) Volume 4 Issue 7, July 2015.

[20]. R. Rao, "Survey on Prediction of Heart Morbidity Using Data Mining Techniques," Knowl. Manag., vol.1, no. 3, pp. 14–34, 2011.

[21]. M. C. Staff, "Heart disease," Mayo Clinic. Online]. Available: http://www.mayoclinic.org/diseasesconditions/heart-disease/symptoms-causes/dxc-20341558

Cite this article as :

Manoj Patil, Dr. Harsh Mathur, "Study of Machine Learning Algorithms for Prediction and Diagnosis of Cardiovascular Diseases : A Review", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 2, pp. 480-489, March-April 2020. Available at doi : https://doi.org/10.32628/CSEIT2062136
Journal URL : http://ijsrcseit.com/CSEIT2062136