IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 1, January 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001041|

Prediction of Heart Disease in Advance Using Supervised Learning Techniques: A Survey

[1]Arudra Sandeep Kumar, [2]Dr.Hiren Dand, [3]Dr.Venkata Kishore Kumar Rejeti

[1]Research Scholar, Shri Jagdishprasad Jhabarmal Tibrewala University (SJJTU), Jhunjhunu, Rajasthan, India

[2]Professor, Dept of CSE, Shri Jagdishprasad Jhabarmal Tibrewala University (SJJTU), Jhunjhunu, Rajasthan, India

[3]Associate Professor, Dept of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, AP, India

ABSTRACT: Heart related diseases or cardiovascular diseases (CVDs) are the main reason for a huge number of deaths in the world over the last few decades and has emerged as the most life-threatening disease, not only in India but in the whole world. So, there is a need of reliable, accurate and feasible system to diagnose such diseases in time for proper treatment. 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. This paper presents a survey of various models based on such algorithms and techniques andanalyses their performance. Models based on supervised learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbour (KNN), NaiveBayes, Decision Trees (DT), Random Forest (RF) and ensemble models are found very popular among the researchers.

KEYWORDS: Cardiovascular Diseases; Support Vector Machines; K- Nearest Neighbour; Naïve Bayes, Decision Tree; Random Forest; Ensemble Models.

I. INTRODUCTION

According to the World Health Organization, every year 12 million deaths occur worldwide due to heart disease. Heart disease is one of the biggest causes of morbidity and mortality among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of data analysis. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms.

The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduces the complications. Machine learning proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the health care industry. This project aims to predict future heart disease byanalysing data of patients which classifies whether they have heart disease or not using machine-learning algorithm. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analysing to extract the desired data we can say that this technique can be very well adapted to do the prediction of heart disease.

Cardiovascular diseases are a group of ailments that disrupt the proper functioning of the heart. Cardiac failure (HF), coronary artery disease (CAD), vascular disease, heart rhythm abnormalities, and other conditions are among the various types of heart disease. The condition where the blood channels constrict or block, resulting in a heart attack (myocardial infarction) and chest discomfort, is known as heart disease. Chest heaviness, shortness of breath, chest pain (angina), irregular heartbeats, and heart abnormalities are key symptoms of heart disease. Heart failure is a long-term condition that damages the heart chambers. Cardiovascular illness interrupts the heart's natural function, which is to pump enough blood into the whole body without raising intracardiac pressure. When the heart fails to supply enough blood to the entire body, the fluid level of the body will be maintained by the kidney, producing lung obstruction and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001041 |

edema in the legs and arms. Congestive heart failure (CHF) is a major health problem in today's world, impacting 26 million people globally.

Around 17.9 million people are killed each year from cardiovascular disease, accounting for 31% of all deaths globally. Many uncontrolled risk factors for heart failure include gender, family history, rising age, etc., whereas hypertension, high cholesterol, smoking, and obesity are in control label health conditions. We research and review the most common types of heart failure problems in order to better understand heart failure. Figure 1 depicts the four right atria that are responsible for regulating blood pumping. Recently, healthcare manufacturing has collected a vast amount of patient data. However, researchers and clinicians are not effectively utilizing this information for illness diagnosis. The healthcare industry is experiencing significant problems in terms of superiority of service (SS), which assures accurate and fast illness analysis and competent patient treatment. Improper diagnosis leads to negative outcomes that are unacceptable.

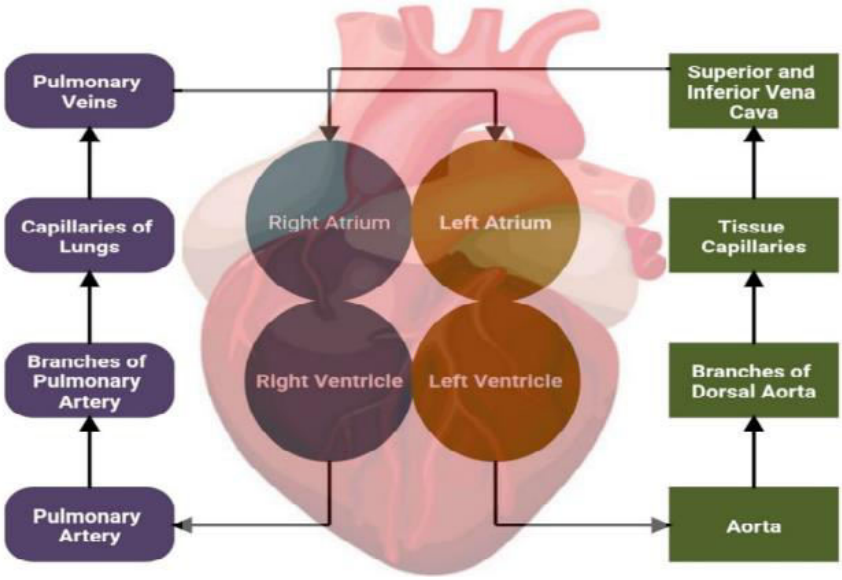

Figure Schematic diagram of heart

Main Kind of Cardiovascular Ailments

A coronary artery disease (CAD) is a kind of cardiovascular disease in which the narrowing or blockage of the coronary arteries is caused by plaque build-up. It is also known as coronary heart disease and ischemic heart disease. It is caused by fatty deposits (plaque) forming inside the arteries. As a result of the artery blockage, the blood supply to the heart muscles is restricted, causing the cardiac function to deteriorate. Coronary arteries are blood vessels that supply oxygen-rich blood to the heart muscle, allowing it to keep hiding. The coronary arteries cross the heart muscle in a straight line. The four principal coronary arteries are the right coronary artery (RCA), left coronary artery (LCA), left forward descending artery (LAD), and left circumflex artery.

The medical name for this ailment is myocardial ischemia. A partial or total blockage of blood vessels causes irreversible damage to the heart, resulting in a heart attack. The four chambers of the human heart are the upper receiving chambers (right and left atria) and the bottom pumping chambers (true and left ventricle (LV)). The right atrium collects deoxygenated blood, while the right ventricle pumps it to the lungs to be oxygenated. The left atrium receives oxygenated blood from the lungs, which is subsequently delivered to all areas of the body by the left ventricle. Because of its size and purpose, the left ventricle chamber is an important and accountable component of the heart.

As a result, left ventricle chamber injury is the most common cause of heart failure. By analysing or monitoring the heart for the course of CAD and the formation of wall motion anomalies, echocardiography assists in the identification of CAD. LV measurement and wall motion score can both be used to determine if a patient has CAD. As a result, LV monitoring is essential for preventing long-term damage to the LV's size, shape, and function. Ultrasound recordings are used in echocardiography to capture different viewpoints, structures, and motions of the heart. Echocardiography is a test that evaluates the functional and anatomical properties of the heart to detect cardiac disease. In addition, echocardiography is utilized to assess the left ventricle discharge portion and cardiac output.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001041|

Congestive Heart Failure

Congestive Heart Failure (CHF) is a kind of heart failure in which the blood arteries become clogged. The cause of congestive heart failure is the insufficient pumping of blood through the heart muscle in the body. Due to this, blood and fluid collection in the lungs causes breathing problems. Because of this, the heart weakens or stiffens over time and causes cardiovascular diseases, such as clogged-up arteries in the heart (coronary artery disease) or extreme blood pressure. As a result, the blood will not be filled efficiently. The long-living life of a human being can be maintained by reducing the signs of cardiac attack and proper treatment. The quality of life can be improved by reducing weight, doing continuous exercise, applying restrictions on salt, and reducing the stresses of life. Congestive heart failure may cause severe symptoms and result in heart surgery, or a myocardial perfusion device may be required. One technique for avoiding heart failure is preventing and treating conditions that might contribute to it, such as atherosclerosis, hypertension, obesity, and being overweight.

Worldwide Heart disease statistics

The World Health Organization has approximated that 12 million deaths happen globally every year owing to the heart diseases. Half of the deaths in the United States and other developed countries occur due to cardio vascular diseases. CVD is also the important reason of deaths in developing countries too to approximate 26% of deaths. On the whole CVD is regardedas the most important cause behind deaths in adults. The term heart disease encompasses the diverse diseases that affect the heart was the major cause of casualties in different countries including India also it is understood from the reports that kill one person every 34 seconds in the United States. The diagnosis of heart disease in most cases depends on a complex combination of clinical and pathological data, which leads to the excessive medical costs affecting the quality of the medical care.

Types of Heart Disease (HD)

Heart disease is a broad term that includes all types of diseases affecting different components of the heart. Heart means cardio. Therefore, all heart diseases belong to the category of cardiovascular diseases. Some of them are as follows:

Angina Pectoris: It is a medical term for chest pain that occurs due to insufficient supply of blood to the heart. Also known as angina, it is a warning signal for heart attack. The chest pain is at intervals ranging for few seconds or minutes.

Congestive Heart Failure: It is a condition where the heart cannot pump enough blood to the rest of the body. It is commonly known as heart failure.

Cardiomyopathy: It is the weakening of the heart muscle or a change in the structure of the muscle due to inadequate heart pumping. Some of the common causes of cardiomyopathy are hypertension, alcohol consumption, viral infections, and genetic defects (Fatima et al. 2017).

Congenital Heart Disease: It also known as congenital heart defect, it refers to the formation of an abnormal heart due to a defect in the structure of the heart or its functioning. It is also a type of congenital disease that children are born with (Kannan et al. 2019).

Arrhythmias: It is associated with a disorder in the rhythmic movement of the heartbeat. The heartbeat can be slow, fast, or irregular. These abnormal heartbeats are caused by a short circuit in the heart's electrical system.

Myocarditis: It is an inflammation of the heart muscle usually caused by viral, fungal, and bacterial infections affecting the heart. It is an uncommon disease with few symptoms like joins pain, leg swelling or fever that cannot be directly related to the heart.

II. LITERATURE REVIEW

[1] Han J, Pei J, Kamber M. (2011)heart disease prediction is more important to prevent the death rate. The death rate increases due to lack of initial detection of heart disease in humans. To predict heart disease in an effective way by using feature selection and classification approach. Thus, an optimized unsupervised technique for feature selection and novel Multi-Layer Perceptron for Enhanced Brownian Motion based on Dragonfly Algorithm (MLP-EBMDA) for classification of heart disease has been proposed in this study. The dataset of the heart disease is obtained as input and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001041|

pre-processing is performed. Features are selected through the optimized unsupervised technique. Based on the selected features, classification of heart disease is performed using the novel hybrid MLP-EBMDA approach. The analysis of the proposed system with various existing systems in terms of accuracy has explored 94.28%. The analysis of the proposed system in terms of precision has showed 96%, in terms of recall the results of the proposed system has been found to be 96%, in terms of F1-score the results of the proposed system has been found to be 96%. Thus, the overall performance analysis of the proposed methodology exhibited efficient results in predicting the heart disease than the various state-of-the-art methods.

[2] Jabbar MA, Deekshatulu BL, Chandra P. (2013)cardiovascular disease (CVD) continues to constitute the leading cause of death globally. CVD risk stratification is an essential tool to sort through heterogeneous populations and identify individuals at risk of developing CVD. However, applications of current risk scores have recently been shown to result in considerable misclassification of high-risk subjects. In addition, despite long standing beneficial effects in secondary prevention, current CVD medications have in a primary prevention setting shown modest benefit in terms of increasing life expectancy. A systems biology approach to CVD risk stratification may be employed for improving risk-estimating algorithms through addition of high-throughput derived omics biomarkers. In addition, modelling of personalized benefit-of-treatment may help in guiding choice of intervention. In the area of medicine, realizing that CVD involves perturbations of large complex biological networks, future directions in drug development may involve moving away from a reductionist approach toward a system level approach. Here, we review current CVD risk scores and explore how novel algorithms could help to improve the identification of risk and maximize personalized treatment benefit. We also discuss possible future directions in the development of effective treatment strategies for CVD through the use of genome-scale metabolic models (GEMs) as well as other biological network-based approaches.

[3] Agrawal R, Srikant R. (2014)in today's world, the advancement of tele-diagnostic equipment plays an essential role to monitor heart disease. The earlier diagnosis of heart disease proliferates the compatibility of treatment of patients and predominantly provides an expeditious diagnostic recommendation from clinical experts. However, the feature extraction is a major challenge for heart disease prediction where the high dimensional data increases the learning time for existing machine learning classifiers. In this article, a novel efficient Internet of Things-based tuned adaptive neuro-fuzzy inference system (TANFIS) classifier has been proposed for accurate prediction of heart disease. Here, the tuning parameters of the proposed TANFIS are optimized through Laplace Gaussian mutation-based moth flame optimization and grasshopper optimization algorithm. The simulation scenario can be carried out using11 different datasets from the UCI repository. The proposed method obtains an accuracy of 99.76% for heart disease prediction and it has been improved up to 5.4% as compared with existing algorithms.

[4] Lakshmi KP, Reddy CRK. (2015)heart disease causes a significant mortality rate around the world, and it has become a health threat for many people. Early prediction of heart disease may save many lives; detecting cardiovascular diseases like heart attacks, coronary artery diseases etc., is a critical challenge by the regular clinical data analysis. Machine learning (ML) can bring an effective solution for decision making and accurate predictions. The medical industry is showing enormous development in using machine learning techniques. In the proposed work, a novel machine learning approach is proposed to predict heart disease. The proposed study used the Cleveland heart disease dataset, and data mining techniques such as regression and classification are used. Machine learning techniques Random Forest and Decision Tree are applied. The novel technique of the machine learning model is designed. In implementation, 3 machine learning algorithms are used, they are 1. Random Forest, 2. Decision Tree and 3. Hybrid model (Hybrid of random forest and decision tree). Experimental results show an accuracy level of 88.7% through the heart disease prediction model with the hybrid model. The interface is designed to get the user's input parameter to predict the heart disease, for which we used a hybrid model of Decision Tree and Random Forest.

[5] Khare S, Gupta D. (2016)among the different causes of human death, heart disease is one of the most common causes of non-communicable and silent death in the world. It is a challenge to early predict heart disease by using clinical data for better treatment. After evolving machine learning, its importance is incessantly being increased in every field of life. From the last couple of years, Machine learning is also the center of attention of researchers in field medical sciences. Researchers use different tools and techniques of machine learning for the early prediction of diseases. Essentially, heart disease prediction with available clinical data is one of the big challenges for researchers. State-of-the art results have been reported using different clinical data using different machine learning algorithms,

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001041|

nevertheless, there is some opportunity for improvement. In this paper, we propose to use a novel method that comprises machine learning algorithms for the early prediction of heart disease. Essentially, the aims of the paper are to find those features by correlation which can help robust prediction results. For this purpose, UCI vascular heart disease dataset is used and compares our result with recently published article. Our proposed model achieved accuracy of 86. 94% which outperforms compare with Hoeffding tree method reported accuracy of 85. 43%

[6] Malarvizhi SP, Sathiyabhama B. (2016)heart disease is one of the most significant causes of mortality in the world today. Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis. Machine learning (ML) has been shown to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. We have also seen ML techniques being used in recent developments in different areas of the Internet of Things (IoT). Various studies give only a glimpse into predicting heart disease with ML techniques. In this paper, we propose a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model is introduced with different combinations of features and several known classification techniques. We produce an enhanced performance level with an accuracy level of 88.7% through the prediction model for heart disease with the hybrid random forest with a linear model (HRFLM).

[7] Altaf W, Shahbaz M, Guergachi A. (2017)coronary artery disease (CAD) is one of the leading causes of death in humans across the world over the last few decades. Coronary artery disease also leads to disability, decreased quality of life and serious illness. CAD can be controlled by identifying the risk factors and on timely diagnosis can also help to reduce the cause of heart failure (death). The conventional method of going through medical history proved that they were less effective in early identification of the disease. So, modern and emerging methods like AI, Machine Learning are more reliable and effective in identifying people with heart disease and can help in containing mortality rate. In the proposed work, three machine learning classification methods (Random Forest, XGBoost and Neural network) are auto-tuned using genetic algorithms to find the most prominent features for maximizing classification performance. These methods are applied to the Z-Alizadeh Sani dataset having demographic examination, ECG, Laboratory and echo data of 303 patients. The computational results of the above application show that three approaches need further ensemble by giving equal importance to all three models to increase the overall performance in assessing the risk and forecasting the presence of disease in the 303 patients.

[8] Methaila A, Kansal P, Arya H, Kumar P. (2017) among the applications enabled by the Internet of Things (IoT), continuous health monitoring system is a particularly important one. Wearable sensor devices used in IoT health monitoring system have been generating an enormous amount of data on a continuous basis. The data generation speed of IoT sensor devices is very high. Hence, the volume of data generated from the IoT-based health monitoring system is also very high. In order to overcome this issue, this paper proposes scalable three-tier architecture to store and process such huge volume of wearable sensor data. Tier-1 focuses on collection of data from IoT wearable sensor devices. Tier-2 uses Apache HBase for storing the large volume of wearable IoT sensor data in cloud computing. In addition, Tier-3 uses Apache Mahout for developing the logistic regression-based prediction model for heart diseases. Finally, ROC analysis is performed to identify the most significant clinical parameters to get heart disease.

[9] Alwidian J, Hammo BH, Obeid N. (2018) Adaptive Neuro Fuzzy Inference System (ANFIS) is among the most efficient classification and prediction modelling techniques used to develop accurate relationship between input and output parameters in different processes. This paper reports the design and evaluation of the classification performances of two discrete Adaptive Neuro Fuzzy Inference System models, ANFIS Matlab's built-in model (ANFIS_ LSGD) and a newly ANFIS model with Levenberg-Marquardt algorithm (ANFIS_LSLM). Major steps were performed, which included classification using grid partitioning method, the ANFIS trained with least square estimates and backpropagation gradient descent method, as well as the ANFIS trained with Levenberg-Marquardt algorithm using finite difference technique for computation of a Jacobian matrix. The proposed ANFIS_LSLM model predicts the degree of patient's heart disease with better, reliable and more accurate results. This is due to its new feature of index membership function that determines the unique membership functions in an ANFIS structure, which indexes them into a row-wise vector. In addition, an attempt was also done to specify the effectiveness of the model's performance measuring accuracy, sensitivity and specificity. A comparison of the two models in terms of training and testing with the Statlog-Cleveland Heart Disease dataset have also been done.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001041|

[10] Amin MS. (2018)cardiovascular disease is the leading cause of death in many countries. Physicians often diagnose cardiovascular disease based on current clinical tests and previous experience of diagnosing patients with similar symptoms. Patients who suffer from heart disease require quick diagnosis, early treatment and constant observations. To address their needs, many data mining approaches have been used in the past in diagnosing and predicting heart diseases. Previous research was also focused on identifying the significant contributing features to heart disease prediction; however, less importance was given to identifying the strength of these features.

[11] Chauhan A, Jain A, Sharma P, Deep V. (2018) the diagnosis of heart disease in most cases depends on a complex combination of clinical and pathological data. Because of this complexity, there exists a significant amount of interest among clinical professionals and researchers regarding the efficient and accurate prediction of heart disease. In this paper, we develop a heart disease predict system that can assist medical professionals in predicting heart disease status based on the clinical data of patients. Our approaches include three steps. Firstly, we select 13 important clinical features, i.e., age, sex, chest pain type, trestbps, cholesterol, fasting blood sugar, resting ecg, max heart rate, exercise induced angina, old peak, slope, number of vessels coloured, and thal. Secondly, we develop an artificial neural network algorithm for classifying heart disease based on these clinical features. The accuracy of prediction is near 80%. Finally, we develop a user-friendly heart disease predict system (HDPS). The HDPS system will be consisted of multiple features, including input clinical data section, ROC curve display section, and prediction performance display section (execute time, accuracy, sensitivity, specificity, and predict result). Our approaches are effective in predicting the heart disease of a patient. The HDPS system developed in this study is a novel approach that can be used in the classification of heart disease.

[12] James SL, et al. (2018)heart disease prediction is a critical task regarding human health. It is based on deriving a Machine Learning model from medical parameters to predict risk levels. In this work, we propose and test novel ensemble methods for heart disease prediction. Randomness analysis of distance sequences is utilized to derive a classifier, which is served as a base estimator of a bagging scheme. Method is successfully tested on medical Spectf dataset. Additionally, a Graph Lasso and Ledoit–Wolf shrinkage-based classifier is developed for Statlog dataset which is a UCI data. These two algorithms yield comparatively good accuracy results: 88.7 and 88.8 for Spectf and Statlog, respectively. These proposed algorithms provide promising results and novel classification methods that can be utilized in various domains to improve performance of ensemble methods.

[13] Kannan AG, Castro TARVC, BalaSubramanian R. (2018)heart disease is the major cause of mortality in the world. The heart disease prediction from the clinical data is deliberate as the most important subject in clinical data analysis. Especially the size of data in health care is vast. Data mining (DM) assists decision and prediction from the raw health care data. DM converts the large collection into useful information. Several existing studies utilize the data mining approaches in heart disease prediction. There is only little research focused on selecting the important features which play a significant role in predicting heart disease is less. The aim of this study is to provide an enhanced approach with novel feature selection and classification technique to predict mortality in congestive heart failure patients. Through this approach the death rate due to heart disease will be decreased gradually. The ant colony optimization (ACO) algorithm is utilized for selecting the best feature for hybrid K-nearest neighbour (KNN) classifier. The proposed approach is compared with the prior classification techniques such as the Support vector machine, Naïve Bayes, KNN, C4.5, and decision tree. UCI Cleveland dataset is utilized for our implementation. Using the NetBeans IDE an experimental was conducted and the result shows that the heart disease prediction model provides a better result with accuracy of 99.2%.The present study shows the efficiency of the HKNN in heart disease prediction system. Initially important features are analysed and then classification is utilized to obtain a better result.

[14] Khan SA, Yadav SK. (2019) Data mining is one of the essential areas of research that is more popular in health organization. Data mining plays an effective role for uncovering new trends in healthcare organization which is helpful for all the parties associated with this field. Heart disease is the leading cause of death in the world over the past 10 years. Heart disease is a term that assigns to a large number of medical conditions related to heart. These medical conditions describe the irregular health condition that directly affects the heart and all its parts. The healthcare industry gathers enormous amount of heart disease data which are not “mined” to discover hidden information for effective decision making. Data mining techniques are useful for analysing the data from many different dimensions and for identifying relationships. This paper explores the utility of various decision tree algorithms in classify and predict the disease

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001041|

[15] Agarwal R, Mittal M. (2019) Heart attack disease is major cause of death anywhere in world. Data mining play an important role in health care industry to enable health systems to properly use the data and analytics to identify impotence that improves care with reduce costs. One of data mining technique as classification is a supervised learning used to accurately predict the target class for each case in the data. Heart disease classification involves identifying healthy and sick individuals. Linear classifier as a Naive Bayes (NB) is relatively stable with respect to small variation or changes in training data. Particle Swarm Optimization (PSO) is an efficient evolutionary computation technique which selects the most optimum features which contribute more to the result which reduces the computation time and increases the accuracy. Experimental result shows that the proposed model with PSO as feature selection increases the predictive accuracy of the Naive Bayes to classify heart disease.

[16] Akbaş KE, Kivrak M, Arslan AK, Çolak C. (2019)heart disease is one of the leading causes for death in men and women across the globe. Several characteristics that can be monitored to predict the heart disease in the earlier stage are blood pressure, cholesterol level, blood sugar level and body weight. The technology is revolutionizing the existing healthcare infrastructure. With the inclusion of Internet of Things (IoT), now we can monitor patients remotely, store their data, and process it for further analysis. However, the need is to propose new and advanced secured algorithms for fast processing and efficient detection of events. In this article, a machine learning based Sine Cosine Weighted K-Nearest Neighbour (SCA_WKNN) algorithm is proposed for the heart disease prediction that learns from the data being stored in blockchain. Since the data stored in the blockchain are tamper resistant, it acts as an authentic source for learning data and also as a secure storage environment for patient information. The performance of proposed SCA_WKNN is assessed in comparison with other algorithms in terms of accuracy, precision, recall, F-score, and root mean square error. Our analysis indicates that SCA_WKNN achieves 4.59% and 15.61% maximum accuracy than W K-NN and K-NN, respectively. Also, blockchain-based storage is compared with peer-to-peer storage in terms of latency and throughput. The blockchain-based decentralized storage achieves 25.03% maximum throughput than peer-to-peer storage.

[17] Amin MS, Chiam YK, Varathan (2019) The objective of the work is to predict heart disease using computing techniques like an oppositional firefly with BAT and rule-based fuzzy logic (RBFL). The system would help the doctors to automate heart disease diagnosis and to enhance the medical care. In this paper, a hybrid OFBAT-RBFL heart disease diagnosis system is designed. Here, at first, the relevant features are selected from the dataset using locality preserving projection (LPP) algorithm which helps the diagnosis system to develop a classification model using the fuzzy logic system. After that, the rules for the fuzzy system are created from the sample data. Among the entire rules, the important and relevant group of rules are selected using OFBAT algorithm. Here, the opposition-based learning (OBL) is hybrid to the firefly with BAT algorithm to improve the performance of the FAT algorithm while optimizing the rules of the fuzzy logic system. Next, the fuzzy system is designed with the help of designed fuzzy rules and membership functions so that classification can be carried out within the fuzzy system designed. At last, the experimentation is performed by means of publicly available UCI datasets, i.e., Cleveland, Hungarian and Switzerland datasets. The experimentation result proves that the RBFL prediction algorithm outperformed the existing approach by attaining the accuracy of 78%.

[18] Bashir, S., Khan, Z. S., Khan, F. H., Anjum, A., & Bashir, K. (2019) Machine learning (ML) is a rapidly developing field in today's world. Use machine learning to extract data from a wide variety of sources. ML can solve various problems based on complex data sets. The prediction of heart disease is the most complex task in the medical field. It cannot be observed with the naked eye, it can appear immediately anywhere, anytime. Many ML algorithms are more capable of handling various algorithms. Due to complexity, the processing of massive data sets is more complicated. By improving these systems, the quality of medical diagnosis decisions can be improved. They can find patterns hidden in large amounts of data that will avoid the use of traditional statistical methods for analysis. In this article, An Enhanced New Dynamic Data Processing (ENDDP) Algorithm is developed to predict the early stages of heart disease. The results prove the performance of the proposed system.

[19] Cengiz AB, Birant KU, Birant D. (2019)heart disease is the primary cause of death nowadays. Treatments of heart disease patients have been advanced, for example with machine-to-machine (M2M) technology to enable remote patient monitoring. To use M2M to take care remote heart disease patient, his/her medical condition should be measured periodically at home. Thus, it is difficult to perform complex tests which need physicians to help.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001041|

Meanwhile, heart disease can be predicted by analysing some of patient's health parameters. With help of data mining techniques, heart disease prediction can be improved. There are some algorithms that have been used for this purpose like Naive Bayes, Decision Tree, and k-Nearest Neighbour (KNN). This study aims to use data mining techniques in heart disease prediction, with simplifying parameters to be used, so they can be used in M2M remote patient monitoring purpose. KNN is used with parameter weighting method to improve accuracy. Only 8 parameters are used (out of 13 parameters recommended), since they are simple and instant parameters that can be measured at home. The result shows that the accuracy of these 8 parameters using KNN algorithm are good enough, comparing to 13 parameters with KNN, or even other algorithms like Naive Bayes and Decision Tree.

[20] Dey L, Mukhopadhyay A. (2019) heart disease is the leading causes of death and hospitalization in the world. With advancement of technology and the contribution of computer engineering, it is easy to detect heart disease and thus treatment is fast and effectively done. Machine learning nowadays is very popular in predicting disease in the medical field. In this paper, authors have tried to predict heart disease using seven machine learning algorithms and attempted to improve the accuracy of weak performing algorithms using ensemble methods like AdaBoost and voting ensemble method. The performance of Linear Discriminate Analysis is good among other algorithms, its mean value is approximately 0.847 and mean absolute error is 0.185, the false acceptance rate is lowest among all i.e.; 0.33 and the false recognition rate is 0.076, accuracy is somehow coming 80% which is less if compared with Logistic Regression.

[21] Dua, D., Graf, C. (2019) In data mining there are several ways, approaches to predict any disease and different researches are still going on. In this survey, we have studied several algorithms (like genetic algorithm, Particle Swarm Optimization, Artificial Neural Network) which play very essential role in determining or predicting heart disease. Here we firstly describe the basic concepts of these three algorithms, and analyse how these algorithms helps in prediction of heart diseases.

[22] Domadiya N, Rao (2018) This model considers the diverse combination of features to make the better classification process. The model shows superior performance with precision via Clinical Decision Support System. The factors that influence the cardiovascular disease need to predict, and better decision is taken during the critical condition. Here, the online available University of California Irvine (UCI) Machine Learning dataset is used for training and testing where 80% data is considered for training and 20% considered for testing purpose. The simulation is done in MATLAB 2020b simulation environment, and the outcomes are compared with various existing approaches. Here, performance metrics like accuracy, precision, F-measure, stability rate, region of curve, and recall is measured to show the model efficiency. The prediction accuracy of the proposed model is 96% which is higher than existing approaches. The overall performance of proposed ensemble model is 96% accuracy, 97% precision, 95% sensitivity, 95% F-measure, 93% Matthew's correlation coefficients, 4.53% False Positive Rate, 3.10% False Negative Rate, and 96% True Positive Rate, respectively.

[23] Domadiya N, Rao (2019) Nowadays, heart disease is the leading cause of death worldwide. Predicting heart disease is a complex task since it requires experience along with advanced knowledge. Internet of Things (IoT) technology has lately been adopted in healthcare systems to collect sensor values for heart disease diagnosis and prediction. Many researchers have focused on the diagnosis of heart disease, yet the accuracy of the diagnosis results is low. To address this issue, an IoT framework is proposed to evaluate heart disease more accurately using a Modified Deep Convolutional Neural Network (MDCNN). The smartwatch and heart monitor device that is attached to the patient monitors the blood pressure and electrocardiogram (ECG). The MDCNN is utilized for classifying the received sensor data into normal and abnormal. The performance of the system is analysed by comparing the proposed MDCNN with existing deep learning neural networks and logistic regression. The results demonstrate that the proposed MDCNN based heart disease prediction system performs better than other methods. The proposed method shows that for the maximum number of records, the MDCNN achieves an accuracy of 98.2 which is better than existing classifiers.

[24] Ibrahim SP, Sivabalakrishnan M. (2019)heart disease is definitely among the many most significant triggers of morbidity and fatality amid the populace among the globe. Prediction of cardiac disease can be considered as one particular among the most crucial topics in the sector of medical info evaluation. The quantity of data through the medical industry is very large. Deep learning becomes the huge range of natural medical care data straight to data which usually may support to identify possibilities and forecasts. This paper reveals the novel algorithm and performance methodology that can forecast the heart disease by ways of CNN modelling. The parameters evaluation

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001041|

will be done for accuracy, sensitivity, specificity, and positive prediction value (PPV). Such parameters can be used in a user-friendly manner by doctors to trace out the possibility of diseases.

[25] Kharya S, Soni S, Swarnkar T. (2019) Data mining for healthcare is an interdisciplinary field of study that originated in database statistics and is useful in examining the effectiveness of medical therapies. Machine learning and data visualization Diabetes-related heart disease is a kind of heart disease that affects diabetics. Diabetes is a chronic condition that occurs when the pancreas fails to produce enough insulin or when the body fails to properly use the insulin that is produced. Heart disease, often known as cardiovascular disease, refers to a set of conditions that affect the heart or blood vessels. Despite the fact that various data mining classification algorithms exist for predicting heart disease, there is inadequate data for predicting heart disease in a diabetic individual. Because the decision tree model consistently beat the naive Bayes and support vector machine models, we fine-tuned it for best performance in forecasting the likelihood of heart disease in diabetes individuals.

[26] Lakshmi KS, Vadivu G. (2019) Conventional clinical decision support systems are based on individual classifiers or simple combination of these classifiers which tend to show moderate performance. This research paper presents a novel classifier ensemble framework based on enhanced bagging approach with multi-objective weighted voting scheme for prediction and analysis of heart disease. The proposed model overcomes the limitations of conventional performance by utilizing an ensemble of five heterogeneous classifiers: Naïve Bayes, linear regression, quadratic discriminant analysis, instance-based learner and support vector machines. Five different datasets are used for experimentation, evaluation and validation. The datasets are obtained from publicly available data repositories. Effectiveness of the proposed ensemble is investigated by comparison of results with several classifiers. Prediction results of the proposed ensemble model are assessed by tenfold cross validation and ANOVA statistics. The experimental evaluation shows that the proposed framework deals with all type of attributes and achieved high diagnosis accuracy of 84.16 %, 93.29 % sensitivity, 96.70 % specificity, and 82.15 % f-measure. The f-ratio higher than f-critical and *p* value less than 0.05 for 95 % confidence interval indicate that the results are extremely statistically significant for most of the datasets.

[27] Mahdi MA, Al-Janabi S. (2019)heart disease diagnosis is found to be a challenging issue which can offer a computerized estimate about the level of heart disease so that supplementary action can be made easy. Thus, heart disease diagnosis has expected massive attention worldwide among the healthcare environment. Optimization algorithms played a significant role in heart disease diagnosis with good efficiency. The objective of this paper is to propose an optimization function on the basis of support vector machine (SVM). This objective function is used in the genetic algorithm (GA) for selecting the more significant features to get heart disease. The experimental results of the GA–SVM are compared with the various existing feature selection algorithms such as Relief, CFS, filtered subset, Info gain, Consistency subset, Chi squared, one attribute based, filtered attribute, Gain ratio, and GA. The receiver operating characteristic analysis is performed to evaluate the good performance of SVM classifier. The proposed framework is demonstrated in the MATLAB environment with a dataset collected from Cleveland heart disease database.

[28] Maji S, Arora S. (2019) Supervised machine learning algorithms are powerful classification techniques commonly used to build prediction models that help diagnose the disease early. However, some challenges like overfitting and underfitting need to be overcome while building the model. This paper introduces hybrid classifiers using the ensembled model with a majority voting technique to improve prediction accuracy. Furthermore, a proposed pre-processing technique and features selection based on a genetic algorithm is suggested to enhance prediction performance and overall time consumption. In addition, the 10-folds cross-validation technique is used to overcome the overfitting problem. Experiments were performed on a dataset for cardiovascular patients from the UCI Machine Learning Repository. Through a comparative analytical approach, the study results indicated that the proposed ensemble classifier model achieved a classification accuracy of 98.18% higher than the rest of the relevant developments in the study.

[29] Ibrahim SS, Sivabalakrishnan M. (2020) Data mining techniques have been widely used in clinical decision support systems for prediction and diagnosis of various diseases with good accuracy. These techniques have been very effective in designing clinical support systems because of their ability to discover hidden patterns and relationships in medical data. One of the most important applications of such systems is in diagnosis of heart diseases because it is one of the leading causes of deaths all over the world. Almost all systems that predict heart diseases use clinical dataset

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001041|

having parameters and inputs from complex tests conducted in labs. None of the system predicts heart diseases based on risk factors such as age, family history, diabetes, hypertension, high cholesterol, tobacco smoking, alcohol intake, obesity or physical inactivity, etc. Heart disease patients have lot of these visible risk factors in common which can be used very effectively for diagnosis. System based on such risk factors would not only help medical professionals but it would give patients a warning about the probable presence of heart disease even before he visits a hospital or goes for costly medical check-ups. Hence this paper presents a technique for prediction of heart disease using major risk factors. This technique involves two most successful data mining tools, neural networks and genetic algorithms. The hybrid system implemented uses the global optimization advantage of genetic algorithm for initialization of neural network weights. The learning is fast, more stable and accurate as compared to back propagation. The system was implemented in Matlab and predicts the risk of heart disease with an accuracy of 89%.

[30] Fitriyani NL, Syafrudin M, Alfan G, Rhee J. (2020) The Healthcare trade usually clinical diagnosis is ended typically by doctor's knowledge and practice. Computer Aided Decision Support System plays a major task in medical field. Data mining provides the methodology and technology to alter these mounds of data into useful information for decision making. By using data mining techniques, it takes less time for the prediction of the disease with more accuracy. Among the increasing research on heart disease predicting system, it has happened to significant to categories the research outcomes and gives readers with an outline of the existing heart disease prediction techniques in each category. Data mining tools can answer trade questions that conventionally in use much time overriding to decide. In this paper we study different papers in which one or more algorithms of data mining used for the prediction of heart disease. As of the study it is observed that Fuzzy Intelligent Techniques increase the accuracy of the heart disease prediction system. The generally used techniques for Heart Disease Prediction and their complexities are summarized in this paper.

III. SCOPE OF THE WORK

Heart disease is a danger to people's health because of its prevalence and high mortality risk. Predicting cardiac disease early using a few simple physical indications collected from a routine physical examination has become difficult. Clinically, it is critical and sensitive for the signs of heart disease for accurate forecasts and concrete steps for future diagnosis. The manual analysis and prediction of a massive volume of data are challenging and time-consuming. In this paper, a unique heart disease prediction model is proposed to predict heart disease correctly and rapidly using a variety of bodily signs. A heart disease prediction algorithm based on the analysis of the predictive models' classification performance on combined datasets and the train-test split technique is presented. Finally, the proposed technique's training results are compared with the previous works. For the Cleveland, Switzerland, Hungarian, and Long Beach VA heart disease datasets, accuracy, precision, recall, F1-score, and ROC-AUC curves are used as the performance indicators. The analytical outcomes for Random Forest Classifiers (RFC) of the combined heart disease datasets are F1-score 100%, accuracy 100%, precision 100%, recall 100%, and the ROC-AUC 100%. The Decision Tree Classifiers for pooled heart disease datasets are F1-score 100%, accuracy 98.80%, precision 98%, recall 99%, ROC-AUC 99%, and for RFC and Gradient Boosting Classifiers (GBC), the ROC-AUC gives 100% performance. The performances of the machine learning algorithms are improved by using five-fold cross validation. Again, the Stacking CV Classifier is also used to improve the performances of the individual machine learning algorithms by combining two and three techniques together. In this paper, several reduction methods are incorporated. It is found that the accuracy of the RFC classification algorithm is high. Moreover, the developed method is efficient and reliable for predicting heart disease.

IV. LIMITATIONS

- It is an unstable classifier i.e.; performance of classifier depends upon the type of dataset. If the dataset is numeric then it generates a complex decision tree
- The processing of ANN network is difficult to interpret and require high processing time if there are large neural networks.
- The main problem is the selection of right kernel function. For every dataset different kernel function shows different results. SVM was designed to solve the problem of binary class. It solves the problem of multi class by breaking it into pair of two classes such as oneagainst-one and one-against all.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001041|

V. CONCLUSION

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. Alternating decision trees when used with PCA, have performed extremely well but decision trees have performed very poorly in some other cases which could be due to over fitting. Random Forest and Ensemble models have performed very well because they solve the problem of over fitting by employing multiple algorithms (multiple Decision Trees in case of Random Forest). Models based on Naïve Bayes classifier were computationally very fast and have also performed well. SVM performed extremely well for most of the cases. Systems based on machine learning algorithms and techniques have been very accurate in predicting the heart related diseases but still there is a lot scope of research to be done on how to handle high dimensional data and over fitting. A lot of research can also be done on the correct ensemble of algorithms to use for a particular type of data.

REFERENCES

[1] Han J, Pei J, Kamber M. Data mining: concepts and techniques, Morgan Kaufmann Publishers is an imprint of Elsevier, 2012 by Elsevier Inc.
[2] Jabbar MA, Deekshatulu BL, Chandra P. Graph based approach for heart disease prediction. In Proceedings of the third international conference on trends in information, telecommunication and computing. New York, NY: Springer. 2013. p. 465–474.
[3] Agrawal R, Srikant R. Fast algorithms for mining association rules. In: Proceedings of 20th international conference very large data bases, VLDB, Vol. 1215, pp. 487–499; 2014.
[4] Lakshmi KP, Reddy CRK. Fast rule-based heart disease prediction using associative classification mining, in 2015 International conference on computer, communication and control (IC4) (pp. 1–5), IEEE, 2015.
[5] Khare S, Gupta D. Association rule analysis in cardiovascular disease. In: Cognitive Computing and Information Processing (CCIP), 2016 Second International Conference on (pp. 1–6), IEEE, 2016.
[6] Malarvizhi SP, Sathiyabhama B. Frequent page sets from web log by enhanced weighted association rule mining. Cluster Computing, 2016, 19(1):269–77.
[7] Altaf W, Shahbaz M, Guergachi A. Applications of association rule mining in health informatics: a survey. ArtifIntell Rev, 2017, 47(3):313–40.
[8] Methaila A, Kansal P, Arya H, Kumar P., American Heart Association, Heart disease and stroke statistics 2017 ata-glance. Geraadpleegd van: https://healthmetrics.heart.org/wp-content/ uploads/2017/06/Heart-Disease-and-Stroke-Statistics-2017-ucm_491265. pdf.
[9] Alwidian J, Hammo BH, Obeid N. WCBA: weighted classification based on association rules algorithm for breast cancer disease. Applications of Soft Computing, 2018, 62:536–49.
[10] Amin MS,identifying significant features and data mining techniques in predicting cardiovascular disease; 2018.
[11] Chauhan A, Jain A, Sharma P, Deep V. Heart disease prediction using evolutionary rule learning, in 2018 4th International conference on computational intelligence & communication technology (CICT) (pp. 1–4), IEEE, 2018.
[12] James SL, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017, Lancet392 (10159), 1789–1858; 2018.
[13] Kannan AG, Castro TARVC, BalaSubramanian R, A comprehensive study on various association rule mining techniques, 2018.
[14] Khan SA, Yadav SK. Class-based associative classification using super subsets to predict the by-diseases in thyroid disorders. in international conference on advances in computational intelligence and informatics (pp. 301–308), Springer, Singapore; 2019.
[15] Agarwal R, Mittal M. Inventory classification using multi-level association rule mining. Int J Dec Supp Syst Technol. (IJDSST), 2019;11(2):1–12
[16] Akbaş KE, Kivrak M, Arslan AK, Çolak C. Assessment of association rules based on certainty factor: an application on heart data set, in 2019 International artificial intelligence and data processing symposium (IDAP) (pp. 1–5), IEEE, 2019.
[17] Amin MS, Chiam YK, Varathan KD Identification of significant features and data mining techniques in predicting heart disease. Telem Inform. 2019;36;82–93.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001041|

[18] Bashir, S., Khan, Z. S., Khan, F. H., Anjum, A., & Bashir, K. (2019). Improving heart disease prediction using feature selection approaches. In 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST) (pp. 619–623). IEEE.
[19] Cengiz AB, Birant KU, Birant D. Analysis of pre-weighted and post-weighted association rule mining, in 2019 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1–5), IEEE.
[20] Dey L, Mukhopadhyay A. Biclustering-based association rule mining approach for predicting cancer-associated protein interactions. IET System Biol. 2019;13(5):234–42.
[21] Dua, D., Graf, C. UCI machine learning repository [http://archive.ics.uci.edu/ ml]. Irvine, CA: University of California, School of Information and Computer Science; 2019.
[22] Domadiya N, Rao UP. Privacy-preserving association rule mining for horizontally partitioned healthcare data: a case study on the heart diseases. Sādhanā. 2018, 43(8):1–9.
[23] Domadiya N, Rao UP. Privacy preserving distributed association rule mining approach on vertically partitioned healthcare data. Procedia Computer Science. 2019; 148:303–12.
[24] Ibrahim SP, Sivabalakrishnan M. An enhanced weighted associative classification algorithm without pre-assigned weight based on ranking hubs, International Journal Adv Computer Science Appl. 10(10); 2019.
[25] Kharya S, Soni S, Swarnkar T. Weighted Bayesian association rule mining algorithm to construct Bayesian Belief network. In: 2019 International conference on applied machine learning (ICAML) (pp. 27–33), IEEE.
[26] Lakshmi KS, Vadivu G. A novel approach for disease comorbidity prediction using weighted association rule mining, Journal of Ambient Intelligence and Humanized Computing, 1–8; 2019.
[27] Mahdi MA, Al-Janabi S. A novel software to improve healthcare base on predictive analytics and mobile services for cloud data centres, in international conference on big data and networks technologies (pp. 320–339), Springer, Cham; 2019.
[28] Maji S, Arora S. Decision tree algorithms for prediction of heart disease. In Information and communication technology for competitive strategies (pp. 447–454), Springer, Singapore, 2019.
[29] Ibrahim SS, Sivabalakrishnan M. An evolutionary memetic weighted associative classification algorithm for heart disease prediction. In Recent Advances on Memetic Algorithms and its Applications in Image Processing (pp. 183–199), Springer, Singapore, 2020.
[30] Fitriyani NL, Syafrudin M, Alfan G, Rhee J. HDPM: an effective heart disease prediction model for a clinical decision support system. IEEE Access, 2020, 8:133034–50.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details