Efficient Classification of Heart Disease using K-Means Clustering Algorithm

M. Thangamani[1], R. Vijayalakshmi[2], M. Ganthimathi[3], M. Ranjitha[4], P. Malarkodi[5],
S. Nallusamy[6]

[1]Associate Professor, Department of Computer Science and Engineering, Kongu Engineering College, Tamil Nadu, India
[2]Associate Professor, Department of Computer Science and Engineering, Velammal College of Engineering and Technology, Tamil Nadu, India
[3] Associate Professor, Department of Computer Science and Engineering, Muthayammal Engineering College, Tamil Nadu, India
[4]Assistant Professor, Department of Electrical and Electronics Engineering, Muthayammal Engineering College, Tamil Nadu, India
[5] Assistant Professor, Department of Physics, Kongu Engineering College, Tamil Nadu, India
[6] Professor, Department of Mechanical Engineering, Dr. M.G.R. Educational and Research Institute, Chennai, Tamilnadu, India

[1]manithangamani2@gmail.com, [2]viji21076@gmail.com, [3]gandhimanimay02@gmail.com

Abstract - *Arteria Coronaria Heart Disease (CAD) is brought about by atherosclerosis in coronary supply routes and consequences in heart failure and besides respiratory failure. For the conclusion of CAD, angiography is utilized as an expensive tedious, and profoundly specialized obtrusive strategy. Scientists are consequently provoked for elective techniques, for example, Artificial Intelligence (AI) calculations that could utilize non-obtrusive clinical information for the coronary illness analysis and evaluating its seriousness. This research illustrates a technique crossbreed strategy intended for CAD determination, containing hazard factor recognizable proof utilizing particle swam optimization with component subset and K-meanss scheme. This implementation compares Multi-Layer Perceptron (MLP), Multinomial Strategic Relapse (MLR), Fluffy Unordered Standard Acceptance Calculation (FURIA), and C4.5 for CAD disease detection. MLR beats different procedures. The proposed hybridized model improves the precision of characterization calculations is 11% for the Cleavelanddata. The anticipated strategy is, along these lines, a capable apparatus for recognizable proof of CAD affected role with progress forecast exactness.*

Keywords: *Machine Learning, Heart Disease, Risk Prediction, Feature Selection, Prediction Model, Classification Algorithms, Cardiovascular Disease*

I. INTRODUCTION

A. Machine Learning

Machine Learning (ML) that progress through experience. It is observed as a subgroup of manmade intelligence. AI techniques are used widely now a day to handle large datasets. ML and AI do soft work with intelligence. Perform For further developed assignments, and it tends to be trying for a human to physically make the required calculations. By and by, it can end up being more powerful to enable the machine to build up its own calculation, as opposed to having human developers indicate each required step. The order of AI utilizes different ways to deal with instruct PCs to achieve errands where no completely palatable calculation is accessible. In situations where tremendous quantities of potential answers exist, one practice is to name a helping of the exact answers as significant. This would then be able to be utilized as preparing information for the PC to improve the algorithm

B. Heart Disease

Illnesses include vein diseases, coronary conduit infection, arrhythmias, and heart abandons. The representation of coronary illness means cardiovascular infection. Heart sickness indicates conditions that contain obstructed veins.

Coronary artery disease is an issue with the veins that convey blood to the heart muscle. In the event that these veins get minuscule, or in the event that they become impeded, blood can't move through them regularly. Since less blood is provided to the heart muscle, the muscle can't work at the typical limit. The heart muscle can get debilitated and frail. Heart muscle can even bite the dust if the bloodstream stops. Obstructed courses in the heart are frequently brought about by smoking, elevated cholesterol, hypertension, diabetes, and acquired qualities from guardians. These issues harm the covering of the heart's veins and cause them to become limited or hindered completely. Congestive Heart Failure is an illness that implies that the heart isn't siphoning at ordinary levels. Two basic causes are powerless or wiped-out heart muscle and

strange heart valves. The valves may not let enough blood through in light of the fact that they are excessively restricted. Or then again, the valve may "hole" and let the bloodstream in reverse (a misguided course) inside the heart. At the point when the heart valves don't work ordinarily, the heart muscle needs to accomplish additional work, and it can get drained.

C. Risk Prediction

Risk forecast instruments are created to recognize patients in danger and to encourage doctor dynamic. The consequence of the expectation models can be utilized to choose the most fitting/suggested strategy. In the ideal world, the choice has been settled on as a shared dynamic to improve the patient's inspiration and adherence. An investigation exhibited that the usage of a danger forecast device encouraged doctors in getting more mindful of the results, in getting more educated on danger factors, and to have a more inspirational demeanor toward preemptive administration.

D. Feature Selection

There are two primary kinds of highlight choice procedures: administered and unaided, and managed techniques might be partitioned into covering, channel, and natural. Channel-based element determination techniques utilize factual measures to score the connection or reliance between input factors that can be sifted to pick the most pertinent highlights. Factual measures for include choice must be painstakingly picked dependent on the information sort of the info variable and the yield or reaction variable. Feature selection is an essential data mining functionality. It is used to identify relevant features to increase the performance of the clusters.

E. Predictive Model

Prescient demonstrating utilizes measurements to foresee outcomes. Most frequently, the occasion one needs to anticipate is later on, yet prescient displaying can be applied to an obscure occasion, paying little mind to when it happened. For instance, prescient models are regularly used to distinguish violations and recognize suspects after the wrongdoing has occurred.

Depending on definitional limits, prescient demonstrating is inseparable from or generally covered with the field of AI, as it is more ordinarily alluded to in scholastic or innovative work settings. When conveyed monetarily, prescient demonstrating is frequently alluded to as prescient examination. Prescient displaying is regularly stood out from causal demonstrating/investigation. In the previous, one might be totally fulfilled to utilize markers of, or intermediaries for, the result of interest. In the last mentioned, one looks to decide genuine circumstances and logical results connections. This differentiation has offered to ascend to expanding writing in the fields of examination techniques and insights of the model.

F. Classification Algorithms

Data set contain label variables as considered for classification, and unknown labels represent clustering techniques. For example, the idea of arrangement calculations with sexual orientation grouping utilizing hair length (in no way, shape, or form am I attempting to generalize by sex, this is just a model). To order sex (target class) utilizing hair length as highlight boundary, we could prepare a model utilizing any grouping calculations to concoct some arrangement of limit conditions that can be utilized to separate the male and female sexes utilizing hair length as the preparation include. In the sex characterization case, the limit condition could affect the best possible hair length esteem. In bunching, the thought isn't to foresee the objective class as in order; it's additionally attempting to assemble the comparative sort of things by thinking about the most fulfilled condition, all the things in a similar gathering should be comparable and ought no two distinctive gathering things to not be comparative.

G. Cardio Vascular Disease

Cardio Vascular Disease (CVD) is a period of disease that affects the heart or veins due to respiratory failure, thromboembolic infection, hypertensive and rheumatic coronary illness, stroke, venous apoplexy, and irregular heart rhythms. Diabetes, blood lipids, hypertension, and throat infection also key factors to occurred cardiovascular diseases. The hidden components shift contingent upon Treating individuals who have strep throat with anti-infection agents can diminish the danger of rheumatic coronary illness. The utilization of anti-inflammatory medicine in individuals who are generally sound is of indistinct advantage.

II. PROBLEM IDENTIFICATION

A computer-aided design approach can be used to find CVD. The existence of atherosclerotic plaques in coronary corridors prompts dead myocardial tissue or abrupt heart passing. To analyze the positive indication of coronary illness to evaluate the degree of harm of heart muscles by taking the atomic sweep, ECG, etc. This constraint prompts angiography; it's an intrusive determination to affirm CAD cases and is considered as the highest quality level for coronary illness recognition and seriousness investigation. In any case, it is expensive and requires a significant level of specialized ability. Scientists are, accordingly, looking for more affordable and compelling other options, state, utilizing information digging for foreseeing CAD cases.

Throughout the day, picture handling, signal processing methods, measurable, and AI methods have been progressively applied to help clinical conclusions utilizing ECG and echocardiogram. ECG and echocardiogram are particular cycles led via prepared practitioners. Here and there, ECG can't affirm CAD cases. This cycle is intricate, exorbitant, includes part of time and exertion. To beat these constraints, numerous scientists utilized other danger factors barring angiography from anticipating CAD cases.

III. RELATED WORK

S Mohan [1] centered another procedure on discovering the significance by giving Artificial Intelligence methods following the improvement of exactness in the estimation of CVD sickness, and it brings about 88.7% all through the forecast model for heart illness with the half breed irregular woods with a direct model. Archana Singh [2] specified the exactness of AI calculations for discovering the coronary illness, k-closest neighbor, choice tree, straight relapse, and backing vector machine calculations were utilized, and UCI vault dataset for preparing and testing. The exhibition of prescient investigation utilizing these information mining, AI calculations on heart infections and break down the different mining, Machine Learning calculations utilized and close which procedures are viable and productive was certified [3].

Remote sensor organizations were utilized to go about as an assortment of center correspondences that gather realities from the city to execute shrewd administrations [4]. We utilize a powerful, versatile possibility revelation component to recognize obscure attacks. A half breed choice help technique of ANN and Fuzzy-AHP were talked for cardiovascular breakdown forecast [5]. In this examination, the normally utilized 13 HF ascribes were thought of, and their commitments were dictated by an accomplished cardiovascular clinician, and the Fuzzy scientific order measure strategy was utilized to register the worldwide loads for the characteristics dependent on their individual commitment. Wagh and Paygude [6] listed the Neuro-Fuzzy model (NFS) as accessible information for the forecast of coronary illness. The half and half framework executed utilizations the worldwide improvement bit of leeway of hereditary calculation for reinstatement of neural organization loads.

The neuro-fluffy model consolidates the neural organization versatile abilities and the fluffy rationale thinking approach for the expectation of coronary illness. The proposed framework will deliver a canny framework with great accuracy. The displaying results uncovered that the proposed cross breed conspires successfully order coronary illness and beat the run of the mill, single-stage ANN method [7]. In this examination paper, a heart illness estimation framework is created using neural organization. The proposed system used 13 clinical requirements for heart illness estimations. When comparable methodologies of the cutting edge, the examinations directed in this work have indicated the great implantation ion of the proposed calculation is contrasted [8]. Resul Das Ibrahim Turkoglu, Abdulkadir Sengur were obtained 89.01% order precision from the investigations made on the data taken from the Cleveland heart illness data set and furthermore acquired 80.95% and 95.91% affectable and particularity esteems, separately, in heart illness finding [9].

Mahboob [10] told the outfit student creates more precision of heart recognition forecast. Such uncovered AI calculation in coronary illness prediction [11]. Li Yang, Haibin Wu, Xiaoqing Jin [12] told that the Random Forest was better than different strategies with an area under the ROC Curve (AUC) of 0.787 and accomplished the critical development over the standard. It gave a Cardiovascular disease expectation model to change the appraisal of CVD. The precise arrangement utilizing K-Mean calculation was explained [13]. Experiments on the gathering of chi-square with Principal Component Analysis (PCA) accomplish better introduction in the greater part of the classifier. The act of PCA sincerely from crude information processed under results and would need more prominent dimensionality to increase the outcomes [14].

The work classification trees accomplish an affectability and a particularity pace of 93.3% and 63.6%, separately, in recognizing higher danger patients [15]. Nancy Masih [16] here assess the introduction of a Neural Network (NN) organization, a help vector machine, a technique with a fluffy system genetically shaped, and an arrangement and relapse tree and its immediate development, which is the arbitrary backwoods, in dissecting our information base. We need quick and efficient detection techniques to reduce the number of deaths [17]. One of the effective data mining techniques used is the decision tree. In heart disease, diagnosis researchers compare different algorithms of decision trees to attain better performance using the WEKA tool. The different algorithms of decision trees are the Random forest algorithm, J48 algorithm, and Logistic model tree algorithm. The main idea of this paper is to identify the unseen pattern by utilizing DM techniques.

The most lethal one in today's modern world is cardiovascular disease. Nowadays, the most provocating task is to detect the disease at the correct time. In India, the treatment for heart disease patients is very high and not reasonable for the patients. The main idea is to give a least-cost treatment using the DM technique by the decision support system. The huge amount of data in varied forms requires lots of effort to make intelligent decisions. Using different features or symptoms to diagnose the disease is a complex activity. The main idea to assist the diagnosis of heart disease varied DM techniques are utilized [18].

A maximum number of deaths are caused due to heart disease in the present scenario. To predict and prevent heart disease, Scientists use various computational techniques. The objective of this paper is to give the better performance of the coronary disease prediction hybridization techniques are used in which artificial neural networks and decision trees are used. This is all done by the Waikato Environment for Knowledge Analysis (WEKA) tool. The dataset is taken from the UCI repository [19, 20]. The individual classifier and hybrid technique's accuracy, sensitivity, and specificity are analyzed. A number of medical terms associated with the heart are called heart disease. The abnormal condition of the heart affects not only the heart but also other parts of the body. In today's time, the major health problem is coronary disease. The main idea of this system is to analyze the various DM technique, which is utilized in recent years for the identification of coronary disease. With the support of

feature subset selection and genetic algorithm, the decision tree is showing good accuracy results [21].

The main reason for death throughout the world is considered a coronary disease. The aim of the paper is to address the issues of coronary disease prediction according to input attributes. Coronary disease identification is investigated using SMO, Multilayer Perception, Bayes Net, J48, and KStar by the WEKA tool. The execution is calculated by the results of the ROC curve, accuracy, AUC value. Bayes Net and SMO techniques show better results than KStar, Multilayer Perception, and J48 techniques [22].

Some DM techniques are utilized for giving correct results and creating efficient decisions on the information. In this study, for identifying the hazard level of coronary disease. Effective Heart Disease Prediction System (EHDPS) is expanded by utilizing the neural network. For identification, the system uses 15 medical parameters [23].

Clustering, classification, regression, association rules, and etc., are the several DM techniques available for identifying coronary diseases. Sequential Minimal Optimization (SMO) algorithms, multi-layer perception, and logistics are algorithms utilized for classification function and tested in this work [24]. The WEKA is used for Comparative analysis. Clustering accuracy and error rate are factors utilized for identifying efficiency and performance. The result reveals that logistics classification function efficiency and minimal sequential optimization is better than multi-layer perception. Using data mining techniques, this paper aims to give a survey of present methods of knowledge discovery in DB techniques, which will be helpful for doctors to make correct decisions [25]. To identify theoccurances of coronary disease with less number of attributes and is the goal of this work. The objective of this study is to predict heart disease with supervised machine learning algorithms [26]. Techniques like categorization, preprocessing are important techniques in Data mining.

Data mining tests could be reduced. Efficient decisions are created with unseen data by implementing it to the patient [27]. This paper analyzes based total on the data mining approaches some parameters and identifies coronary diseases, thereby proposes a Heart Diseases Prediction System (HDPS). The aim of this paper is the identification of coronary disease in a patient by the withdrawal of interesting patterns from the coronary dataset using important parameters by utilizing and analyzing various DM methods [28].

Machine learning can be implemented where the identification of coronary disease is one of the sectors, with good flexibility and adaptability [29]. For redundant filter features, they use Fast Correlation-Based Feature Selection (FCBFS), and it also increases the standard of coronary diseases classification. Therefore, this study examines that outcome utilities various measures, i.e.f1-score, precision, accuracy, recall, etc. [30].

To predict various diseases, DM is helpful for the identification of various data, and it useful for the medical field. To identify heart problems, some of the DM and ML techniques such as Support Vector Machine (SVM), Decision tree, K-Nearest Neighbour (KNN), Fuzzy Logic, Naïve Bayes, and Artificial Neural Network (ANN). This result gives an overall summary of the work. As health care having different disease characteristics, domains, and various methods have their own identification efficiency and are enhanced. Using performance measures like kappa statistics, MAE RMSE, and ROC using the WEKA tool, the Authors have compared Random Forest, J48, SVM, k-NN classifiers. And Naive Bayes [31]. In heart disease prediction, the study aims to use data mining techniques with a simple framework to be utilized. To improve accuracy, K Nearest Neighbors (KNN) is utilized for framework weighting techniques. Out of 13 parameters, only 8 parameters are used [32].

IV. PROPOSED SYSTEM

In the proposed work client will look for the coronary illness analysis (coronary illness and treatment-related data) by giving manifestations as a question in the internet searcher. These manifestations are preprocessed to make the further cycle simpler to discover the indications watchword, which assists with distinguishing the coronary illness rapidly. The side effects which watchword is coordinated with the put away clinical information base to recognize the numerous heart diseases identified with that catchphrase. Different heart diseases are distinguished, and it will make the example coordinating about the numerous heart diseases and furthermore discover the likelihood of heart diseases. At that point, coronary illness will make a differential analysis to discover coronary illness exactness.

The watchword, which is a preprocessed manifestation, is coordinated with the heart diseases put away in the neighborhood information base to distinguish the comparing coronary illness identified with those side effects given by the client. This needs to look through a record information base of in excess of 20000 heart diseases and much more manifestations, which is very tedious, so CFS+PSO grouping was applied to characterize heart Disease highlights into subgroups. On the off chance that a gathering of side effects coordinate higher inclination is given to that subgroup, and looking in that new, more modest subgroup subsequently decreases information base access.

An example acknowledgment, CFS with PSO Feature Selection calculation is a technique for grouping objects dependent on nearest preparing models in the element space. CFS+PSO are a kind of occurrence based learning, or sluggish realizing where the capacity is just approximated locally, and all calculation is conceded until characterization. This element has been distinguished as the most reasonable for the current framework.

A. Clevel and Heart Disease Training Dataset

The Cleveland heart disease data found in the UCI AI store contains 14 factors predictable on 303 people who have the coronary illness. The people had been gathered into five

degrees of coronary illness.

In the dataset, attributes are age, sex, chest pain type, chol, etc. Age can be mentioned in years. Chest pain type one means it represents typical angina, 2 indicates atypical angina, 3 means nonanginalpain, and four indicates asymptomatic. Trestbps is a continuous value attribute, and it represents the level of resting blood pressure. Serum cholesterol can be measured in the form of mg.

Fbs is a discrete value, indicates fasting blood sugar. Maximum heart rate is indicated by the Thalach attribute. Thal three means normal, six indicate a fixed defect, and seven denotes reversible defect. The attribute diagnosis shows the healthy patient by zero, and the patient who is subject to conceivable heart diseases is shown by one.

V. RESULT AND DISCUSSION

Heart disease prediction of the classification model is claimed through accuracy and misclassification rate. Table 1 and Figure 1 show the comparison result of the classification models. The precision of order models is expanded by 11% if there should arise an occurrence of MLP, 9.3 % in the event of MLG, 9.2% if there should be an occurrence of FURIA, and 9.4 % if there should be an occurrence of C4.5. We analyzed exactness accomplished by before utilized models for the Cleveland dataset.

Table 1: Overall Result of MLP, MLG, FURIA, and C4.5 in CAD

Measures	MLP	MLR	FURIA	C4.5
Kappa statistic	63.2	10.91	6.9	0.261
Mean absolute error	50.59	47.66	45.4	44.77
Root mean squared error	53.08	48.83	624.9	73.1
Relative absolute error	101.18	95.32	90.7	89.5
Root relative squared error	106.39	97.2	124.98	94.63
Coverage of cases	98.98	100	67.33	100
Mean rel. region size	98.99	100	62.71	98.35
Accuracy in %	93.67	92.7	94.7	94.9
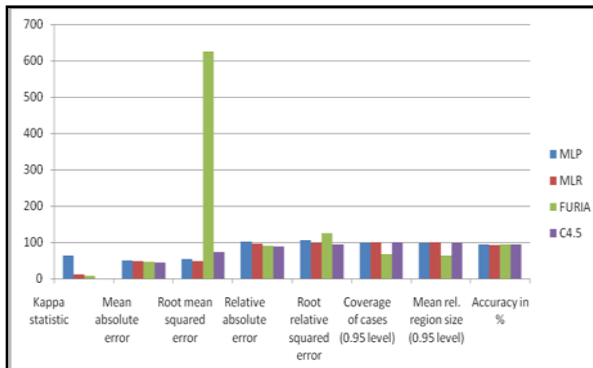
Fig. 1: Results of MLP, MLG, FURIA, and C4.5 in CAD

VI. CONCLUSION AND FUTURE WORK

The clinical finding is a significant region of exploration, which assists in recognizing the event of a coronary illness. The framework, utilizing different methods referenced, will thus uncover the coronary root illness alongside the arrangement of most plausible heart diseases that have comparative side effects. The information base utilized is a portrayal data set, so to decrease the dataset tokenization, separating and stemming is finished. The venture shows a mixture model to recognize and affirm CAD cases requiring little to no effort by utilizing clinical information that can be effectively gathered at clinics. In the feature, this model can be applied to the CAD big data sets.

REFERENCES

[1] Senthilkumar Mohan, Chandrasegar Thirumalai, and Gautam Srivastava, Effective heart disease prediction using hybrid machine learning techniques, IEEE Access, 7 (2019) 81542-81554.

[2] Archana Singh and Rakesh Kumar, Heart Disease Prediction Using Machine Learning Algorithms, International Conference on Electrical and Electronics Engineering, IEEE, (2020) 452-457.

[3] M. Nikhil Kumar, K.V.S. Koushik and K. Deepak, Prediction of heart diseases using data mining and machine learning algorithms and tools, International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 3(3) (2018) 887-898.

[4] J. Wu, K. Ota, M. Dong, and C. Li, A progressive security structure for safeguarding against complex assaults on remote sensor networks in brilliant urban areas, IEEE Access. 4 (2016) 416-424.

[5] O.W. Samuel, G.M. Asogbon, A.K. Sangaiah, P. Tooth, and G. Li, A coordinated choice emotionally supportive network dependent on ANN and Fuzzy_AHP for cardiovascular breakdown hazard prediction, Expert Systems with Applications: An International Journal. 68 (2017) 163-172.

[6] R. Wagh and S.S. Paygude, CDSS for coronary illness forecast utilizing hazard factors, International Journal of Innovative Research in Computer and Communication Engineering, 4(6) (2016) 12082-12089.

[7] Y.E. Shao, C.D. Hou, and C.C. Chiu, Hybrid clever demonstrating plans for coronary illness characterization, Appl. Delicate Comput. Journal 14 (2014) 47-52.

[8] N. Al-milli, 'Backpropagation neural organization for the forecast of coronary illness, J. Theor. Appl.Inf. Technol. 56(1) (2013) 131-135.

[9] Resul Das Ibrahim Turkoglu, Abdulkadir Sengur, Effective determination of coronary illness through neural organizations outfits, Elsevier, Expert Syst. Appl. 36(4) (2009) 7675-7680.

[10] T. Mahboob, R. Irfan, and B. Ghaffar Evaluating troupe expectation of coronary illness utilizing beneficiary working attributes, in Proc. Web Technol. Appl. (2017) 110-115.

[11] Sudha. K, An upgrading coronary illness expectation is utilizing AI calculations, International Journal of Innovative Research in Science, Engineering, and Tech. 9(6) (2020) 2319-8753.

[12] Li Yang, Haibin Wu, Xiaoqing Jin, Pinpin Zheng, Shiyun Hu, Xiaoling Xu, Wei Yu, and JingYan, Study of cardiovascular infection expectation model dependent on irregular timberland in eastern China Scientific Reports, (2020) 86-92.

[13] Md. Zakir Hossain, Md. Nasim Akhtar, R.B. Ahmad, and Mostafijur Rahman, A unique K-implies grouping for information mining,

Indonesian Journal of Electrical Engineering and Computer Science. 13(2) (2019) 521-526.

[14] Panelanna Karen, Garate Escamila, AmirHajjam El, Hassania Emmanuel, Andres, Classification models for coronary illness forecast utilizing highlight choice and PCA, Informatics in Medicine Unlocked.19 (2020) 1-11.

[15] S. Nallusamy, Christina Mary P. Paul and K. Sujatha, Bio-medical waste management system in the multi-specialty hospital using Birnbaum's measures - A case study, Indian Journal of Public Health Research and Devpt. 9(2) (2018) 276-289.

[16] Guidi, G., Pettenati, M.C., Melillo, P. and Iadanza, E., An AI framework to improve cardiovascular breakdown understanding Assistance, IEEE J Biomed Health Inf. 18(6) (2014) 1750-1756.

[17] Bhatla, N. and K. Jyoti, An analysis of heart disease prediction using different data mining techniques, International Journal of Engineering Research & Technology. 1(8) (2012) 01-04.

[18] Sultana, M. et al., Analysis of data mining techniques for heart disease prediction, 3rd International Conference on Electrical Engineering and Information Communication Technology, IEEE. (2016) 116-124.

[19] Patel, J. et al. Heart disease prediction using machine learning and data mining technique, Heart Disease. 7(1) (2015) 129-137.

[20] P. Seetha Subha Priya, S. Nandhinidevi, M. Thangamani, S. Nallusamy, A review on exploring the deep learning concepts and applications for medical diagnosis, Int. Journal of Engineering Trends and Technology. 68(10) (2020) 63-66.

[21] Taneja, A. Heart disease prediction system using data mining techniques, Oriental Journal of Computer Science and Tech. 6(4) (2013) 457-466.

[22] D. Sobya, S.K. Muruganandham, S. Nallusamy and P.S. Chakraborty, Wireless ECG monitoring system using IoT based signal conditioning module for real-time signal acquisition, Indian J. of Public Health Research and Devpt. 9(2) (2018) 294-299.

[23] S.K. Muruganandham, D. Sobya, S. Nallusamy, Dulal Krishna Mandal and P.S. Chakraborty, Study on leaf segmentation using k-means and k-medoid clustering algorithm for identification of disease, Indian Journal of Public Health Research and Development. 9(2) (2018) 289-293.

[24] Vijayarani, S. and Sudha, S., Comparative analysis of classification function techniques for heart disease prediction, International Journal of Innovative Research in Computer and Communication Engineering. 1(3) (2013) 735-741.

[25] Chaurasia, V. and Pal, S., Data mining approach to detect heart diseases, International Journal of Advanced Computer Science and Information Technology. 2 (2014) 56-66.

[26] Kanchan, B.D. and Kishor, Study of machine learning algorithms for special disease prediction using the principle of component analysis, International Conference on Global Trends in Signal Processing, Information Computing, and Communication, IEEE. (2016) 220-229.

[27] Sarangam Kodati, D.R.V., Analysis of heart disease using in data mining tools Orange and Weka, Global Journal of Computer Science and Technology C: Software & Data Engineering. 18(1) (2018) 17-21.

[28] Chadha, R. and S. Mayank, Prediction of heart disease using data mining techniques, CSI Transactions on ICT. 4(2-4) (2016) 193-198.

[29] Khourdifi, Y. and M. Bahaj, Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization, Int. J. Intell. Eng. Syst.12(1) (2019) 242-252.

[30] Marimuthu, M. et al. A review on heart disease prediction using machine learning and data analytics approach, International Journal of Computer Applications, 975 (2018) 8887-8892.

[31] Kumar, N. and S. Khatri Implementing WEKA for medical data classification and early disease prediction, 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT), IEEE (2017) 56-65.

[32] Enriko, I.K.A. et al. Heart disease prediction system using the k-nearest neighbor algorithm with simplified patient's health parameters, Journal of Telecommunication, Electronic, and Computer Engineering, 8(12) (2016) 59-65.