International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 05, 2020
ISSN: 1475-7192

A New Approach for Cardio Vascular Disease Prediction Using Decision Tree

Dr.C. Saravanabhavan, C. Kalpana, E. Saranya, M. Suganthi and P. Preethi

Abstract--- *Heart diseases are the central reason for death everywhere throughout the world in the course of the most recent couple of decades. To stay away from heart malady or coronary sickness and find signs early, people more than 55 years must have an all out cardiovascular exam. Scientists and experts created different shrewd procedures to improve limit of the human services experts in acknowledgment of cardiovascular infection. In cardiovascular infection finding and treatment, single information mining procedures are giving the sensible exactness and accuracy. All things considered the use information mining method is equipped for diminishing the quantity of test that is required to be done. So as to diminish the figure of passing's from heart diseases there must be a snappy and proficient recognition method giving better accuracy and exactness. The point of this paper is to introduce an effective method of anticipating heart diseases utilizing machine learning draws near. In this paper, hybridization method is proposed in which decision tree and counterfeit neural system classifiers are hybridized for better execution of forecast of heart illness. This is finished utilizing WEKA. To approve the presentation of the proposed algorithm, ten times approval test is performed on the dataset of heart sickness patients which is taken from UCI storehouse. The accuracy, affectability, and explicitness of the individual classifier and half and half procedure are analyzed.*

Keywords--- *Cardio, Disease, Prediction, Machine Learning, Decision Tree.*

I. INTRODUCTION

Heart disappointment is an ordinary degeneration of the heart work highlighted by the diminished capacity for the heart to siphon blood productively [1], and is a typical and exorbitant clinical condition, related with critical bleakness and mortality [2]. Auspicious determination is imperative to advance proof based treatment openings, which can postpone mortality and improve side effects. In this manner, it mixes with deoxygenated blood and is given to the lungs, regardless of the way that it's starting at now empowered with oxygen. If the atrial septal distortion is significant, this extra blood volume can pack the lungs and fumes the right half of the heart. If not treated, the right half of the heart over the long haul expands and weakens. If this methodology continues with, the circulatory strain in your lungs may augment additionally, inciting pneumonic hypertension [3].

Concealed example from information and existing relationship can be extricated from huge information sources utilizing information mining methods that consolidate factual examination, machine learning and database innovation [9]. In clinical focuses (medical clinics or in facilities) information mining method helps in

Dr.C. Saravanabhavan, Professor and Head, Department of Computer Science and Engineering, Kongunadu College of Engineering and Technology, Trichy.
C. Kalpana, Assistant Professor, Computer Science and Engineering, Karpagam Institute of Technology, Coimbatore.
E. Saranya, Assistant Professor, Information Technology, Karpagam Institute of Technology, Coimbatore.
M. Suganthi, Assistant Professor, Computer Science and Engineering, Roever College of Engineering and Technology, Perambalur.
P. Preethi, Assistant Professor, Computer Science and Engineering, Kongunadu College of Engineering and Technology, Trichy.

DOI: 10.37200/IJPR/V24I5/PR2020796
Received: 24 Apr 2020 | Revised: 30 Apr 2020 | Accepted: 25 May 2020

distinguishing that if the individual has any sort of diseases or not. It is likewise utilized for early programmed finding of patients from their diseases for example in brief timeframe. It is utilized for programmed conclusion of the infection and gives good administrations in social insurance habitats for sparing the life of people. Forecast strategy encourages the partners to take sensible decision, especially for pros to give method of reasoning decision to treat patients. In this paper, a crossover strategy utilizing Random timberland classifier and straightforward k-implies algorithm for anticipating heart illness is proposed [6]. Machine learning is a specialty of acing framework without being unequivocally figured. They are utilized to break down the investigative course of action in high dimensional, differing informational collections like heart ailment [7].

As indicated by the cardiovascular sickness is the main source of death in the UK, US, Canada, and Australia and will happen because of cardiovascular malady. Coronary heart malady, arrhythmia, and myocardial localized necrosis are a few instances of heart ailment. Some significant reasons of heart ailment are age, smoking, diabetics, bloatedness, genetic, sorrow, hyper strain, pulse, cholesterol and so on [4-7].Usually cardio vascular sickness can be use with medical procedure or drug. Yet, its powerful counteraction isn't yet being finished. The viable avoidance heart infection is additionally an objective of the exploration [8-2].

The remainder of this paper the review is quickly examined in Section II. Issues are recognized in Section III. Technique examined in Section IV. Result is given in Section V. at long last gives Conclusion in Section VI.

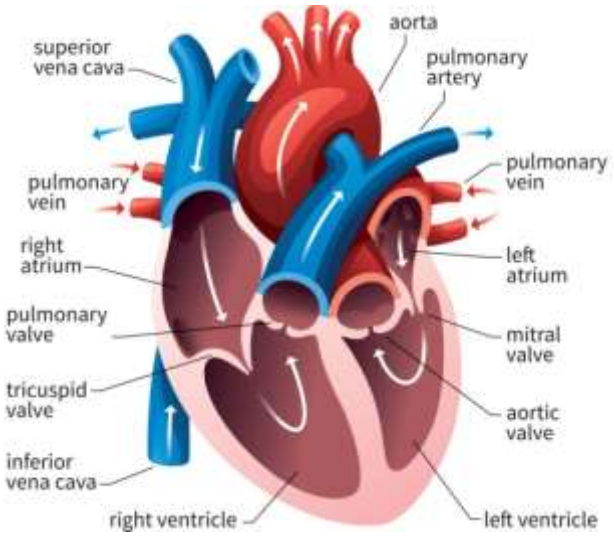

Figure 1: Human Heart Structure

II. BACKGROUND WORK

Borkar, Sneha, et al. [3] the customized approach for distinguishing proof of ASD (Atrial Septal Defect) And DCM (Dilated Cardiomyopathy) is shown. The ultrasound chronicles are taken as data, the housings are preprocessed by center channel to oust salt and pepper commotion. The features are expelled by authentic parameters. The isolated features were requested using oversaw support vector machine computation. The proposed approach achieved the ID pace of 98.30%. In future, the significant learning figurings can be completed by social event immense database.

DOI: 10.37200/IJPR/V24I5/PR2020796
Received: 24 Apr 2020 | Revised: 30 Apr 2020 | Accepted: 25 May 2020

Omar Boursalie, et al. [4] the authors are have exhibited another framework for the observing of cardiovascular ailment. The proposed framework depends on the SVM which breaks down highlights extricated from wearable sensors and clinical databases. Such a framework will help medicinal services experts by introducing applicable and summed up data on a patient's condition from the entirety of the gathered information. The verification of-idea framework effectively grouped a patient's hazard level utilizing engineered information. Taking into account these outcomes, the authors are actualizing this framework on a cell phone utilizing genuine information from clinical databases and wearable sensors. The checking gadget will be assessed by its accuracy in ordering live observed information. The authors are researching strategies to test the framework's affectability to changing patient conditions.

Chen, Rui, et al. [5] ML gave a powerful method to anticipate the danger of cardiovascular occasions in patients with serious DCM during 1-year development.

Dhar, Sanchayita, et al. [6] the goal was to devise a prescient model for recognition of cardiovascular heart ailment utilizing machine learning strategies using fluctuated parameters identified with heart. The dataset was pre-handled and three distinctive managed machine learning order algorithms for example J48 Classifier, Naive Bayes and Random Forest utilizing Weka device of machine learning programming. The presentation of the method was assessed using benchmark measurements of accuracy, exactness, review and F-measure and so on. The best model to foresee patients with heart infection gives off an impression of being a Random Forest classifier actualized on chosen properties with an order accuracy of 100%. Essentially heart malady is a fatal illness and postponement in analysis or misdiagnosis of this infirmity can ground to serious or life startling issues like heart failure and even demise. The best model chose for foreseeing heart sickness couldn't surpass accuracy of arrangement 100%. Henceforth it demonstrates that machine learning or information mining systems can be effectively used to anticipate cardiovascular heart diseases. The finish of this paper can be used as a subordinate apparatus via cardiologists to define solid analysis of heart diseases.

Dinesh, Kumar G., et al. [7] contributes the correlative application and examination of particular machine learning algorithms in the R programming which gives a prompt system for the client to utilize the machine learning algorithms in R programming for anticipating the cardiovascular diseases. This is non-moral investigation intends to utilize accessible machine learning strategies in R programming.

Terrada, Oumaima, et al. [8] utilized four machine learning algorithms, for example, the ANN, KNN, k-means, and K-medoids algorithms prepared on Cleveland dataset for heart ailment. The consequence of this work was assessed with numerous exhibition assessment strategies, for example, sensitivities, specificities, accuracy, and MCC of the testing set contains 83 patients. The proposed framework results arrived at a general accuracy of 96% and 93% of MCC.

Fathalla, Karma M., et al. [9] proposed RVA-based cardiovascular hazard forecast models of tantamount arrangement execution to customary QRisk and Framingham models.

DOI: 10.37200/IJPR/V24I5/PR2020796
Received: 24 Apr 2020 | Revised: 30 Apr 2020 | Accepted: 25 May 2020

III. PROBLEMS IDENTIFIED

The inauspicious location of heart diseases can stop the passing. However, in each circumstance it is seeing at the last phases of infection. Or then again after death. So the Health cares are point to recognize the sickness at beginning times. For this situation Data mining strategy is the acceptable procedure can identify ailment. Totally fix the illness by appropriate determination. In any case, the fundamental issue of Data mining is utilizing various algorithms for location of heart illness. A few algorithms are analyze is less exact and tedious.

IV. SYSTEM MODEL

To analyze the sickness at beginning time at moderate expense is the significant point of this paper. Utilizing information mining Technique we can distinguish infection from the start stage. We can totally fix the infection by legitimate finding. Medicinal services industry gathers immense measure of data.

A) Datasets

Heart disease data was collected from the UCI machine learning repository. Cleveland Clinic Foundation (cleveland. data). The Cleveland database contains total 76 raw attributes, but in experiments only 14 of them is actually used. The dataset used in this experiment contains different important parameters like ECR, cholesterol, chest pain, fasting sugar, MHR (maximum heart rate) and many more. The detailed information about these attributes and their domain range are as follows:

@relation Cleveland
@attribute age real [29.0, 77.0]
@attribute sex real [0.0, 1.0]
@attribute cp real [1.0, 4.0]
@attribute trestbps real [94.0, 200.0]
@attribute chol real [126.0, 564.0]
@attribute fbs real [0.0, 1.0]
@attribute restecg real [0.0, 2.0]
@attribute thalach real [71.0, 202.0]
@attribute exang real [0.0, 1.0]
@attribute oldpeak real [0.0, 6.2]
@attribute slope real [1.0, 3.0]
@attribute ca real [0.0, 3.0]
@attribute thal real [3.0, 7.0]
@attribute num {0, 1, 2, 3, 4}
@inputs age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal
@outputs num

We have used the Classification model by covering rules (based on DT) on the above modified dataset and find out the generated rule sets with different priority.

DOI: 10.37200/IJPR/V24I5/PR2020796
Received: 24 Apr 2020 | Revised: 30 Apr 2020 | Accepted: 25 May 2020

B) Data Pre-processing

Heart issues information is really pre-handled just after assortment of various records. The genuine dataset incorporates an all out related with 303 individual records, any place 6 information are which incorporates missing convictions. Those six records happen to be expelled from the real dataset and furthermore the staying 297 patient data are utilized inside pre-preparing.

C) Feature Selection and Reduction

From among the 13 characteristics of the informational index, two credits relating to age and sex are utilized to recognize the individual data of the patient. The staying 11 characteristics are viewed as significant as they contain indispensable clinical records. Clinical records are fundamental to determination and learning the seriousness of heart infection.

D) Classification Modelling

The grouping of datasets is done based on the factors and standards of Decision Tree (DT) highlights. At that point, the classifiers are applied to each grouped dataset so as to assess its presentation. The best performing models are distinguished from the above outcomes dependent on their low pace of blunder. The exhibition is additionally streamlined by picking the DT group with a high pace of blunder and extraction of its comparing classifier highlights. The presentation of the classifier is assessed for blunder improvement on this informational index.

E) Decision Trees

For preparing tests of information D, the trees are developed dependent on high entropy inputs. These trees are straightforward and quick developed in a top down recursive separation and overcomes (DAC) approach. Tree pruning is performed to expel the unessential examples on D.

$$Entropy = -\sum_{j=1}^{m} pijlog2pij - - - - - - - -(1)$$

Algorithm 1 Decision Tree-Based Partition

Require: Input: D dataset - features with a target class

for features **do**

for Each sample **do**

Execute the Decision Tree algorithm

end for

Identify the feature space f_1, f_2,..., f_x of dataset UCI.

end for

Obtain the total number of leaf nodes l_1, l_2, l_3,..., l_n with its constraints

Split the dataset D into d1, d2, d3,..., d*n* based on the leaf nodes constraints.

Output: Partition datasets d_1, d_2, d_3,..., d_n

Support Vector Machine: SVM is a managed machine learning algorithm. It is generally utilized in order issues. This has a preparation stage in the underlying. We have to give the information to the SVM algorithm appears in Fig. 3 the information which is arranged. At that point information are given to the algorithm which will be

DOI: 10.37200/IJPR/V24I5/PR2020796
Received: 24 Apr 2020 | Revised: 30 Apr 2020 | Accepted: 25 May 2020

characterized with the assistance of the less incitement of the human. It has a stage called highlight determination. That it recognizes the know classes. At the point when the obscure example. Its expectation is hub required a lot of then we will utilize SVM. These both are likewise utilized for recognizing the keysets. That is engaged with various procedures that help to separate the classes. The accompanying chart is the procedure of SVM.

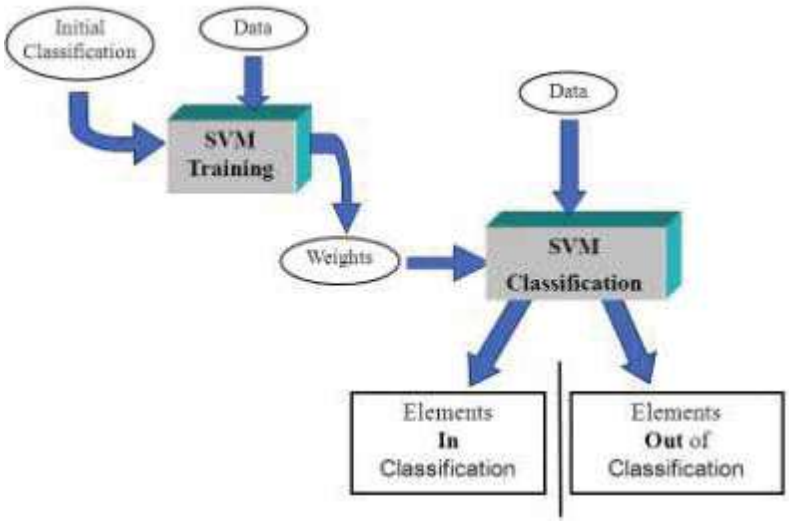

Figure 2: SVM Model

Random Forest: Random forest is a decision tree is a decision bolster instrument. That utilizes a tree like model of decisions it is overwhelmingly a band of unprimed grouping trees. It gives amazing execution on various down to earth issues.Such as medicinal services expectation issues. It isn't touchy to clamor in the informational index. It isn't exposed to over fitting. It is worked by joining the forecasts of a few trees. Every one of them is prepared independently. It works quickly. Generally displays a noteworthy exhibition improvement over numerous other tree-based algorithms.

KNN: KNN is a non parametric method in Fig. 3 Utilized for grouping. It is a sort of occurrence based learning. All calculation is conceded shift till order. This algorithm is among the most straightforward of all machine learning algorithms. Valuable procedure can be utilized to dole out weight to the commitments of the neighbors. A typical weighting plan comprises in giving each neighbor a load of 1/d. Where d is the separation to the neighbor. It is delicate to the nearby structure of the information.

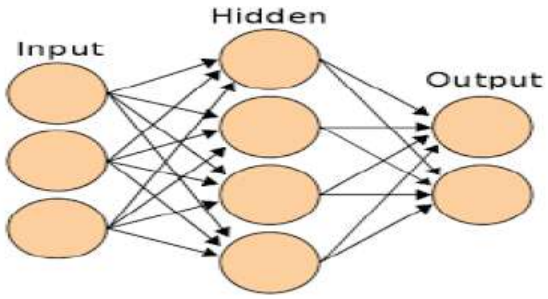

Figure 3: KNN Model

DOI: 10.37200/IJPR/V24I5/PR2020796
Received: 24 Apr 2020 | Revised: 30 Apr 2020 | Accepted: 25 May 2020

Comparison between Different Algorithms

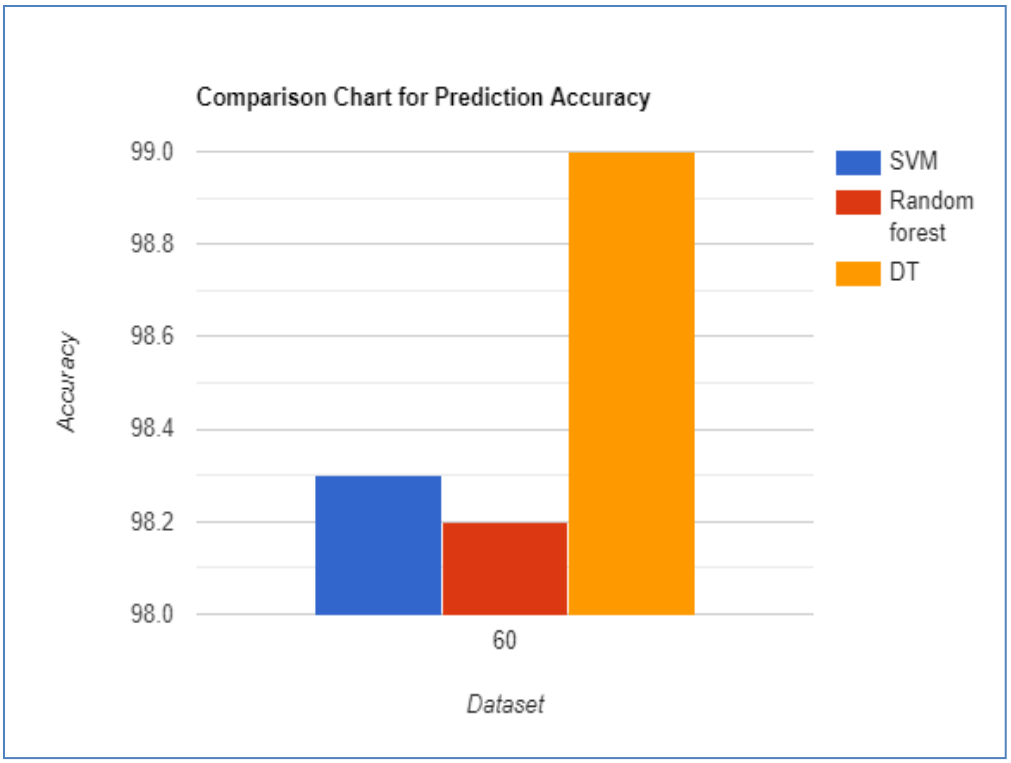

Figure 4: Comparison Chart for Different Algorithms

V. RESULTS AND DISCUSSION

This examination consolidated the data of cardiovascular time arrangement fluctuation and stage space remaking to group typical subjects and heart disappointment patients. Recognizing the handling of crude social insurance information of heart data will help in the drawn out sparing of human lives and early location of variations from the norm in heart conditions. Machine learning methods were utilized in this work to process crude information and give another and novel insight towards heart malady. Heart illness expectation is testing and significant in the clinical field. Be that as it may, the death rate can be definitely controlled if the malady is distinguished at the beginning periods and protection measures are embraced at the earliest opportunity.

The term cardiovascular framework affliction alludes to disorder of cardiovascular framework and pontoon structure inside it. There are measures of components which as a rule construct the danger of Heart and soul contamination:

- Family history of coronary illness
- Smoking
- Poor eating methodology

High pulse

- Cholesterol

DOI: 10.37200/IJPR/V24I5/PR2020796
Received: 24 Apr 2020 | Revised: 30 Apr 2020 | Accepted: 25 May 2020

- High blood cholesterol
- Obesity
- Physical inertia

The subset of 13 attributes (Age, sex, cp, treetops, chol, FBS, restecg, thalach, exang, oldpeak, slope, ca, that, target) is selected from the pre-processed data set of heart disease. The three existing models for heart disease prediction (DT, LR, SVM) are used to develop the classification. The evaluation of the model is performed with the confusion matrix. Totally, four outcomes are generated by confusion matrix, namely TP (**True Positive**), TN (**True Negative**), FP (**False Positive**) and FN (**False Negative**). The following measures are used for the calculation of the accuracy, sensitivity, specificity.

Accuracy = (TN+TP) / (TN+TP+FN+FP)

Sensitivity =(TP/TP+FN)

Specificity = (TN/TN+FP)

Precision = TP /TP+FP

F-Measure = 2TP/ 2TP+FP+FN

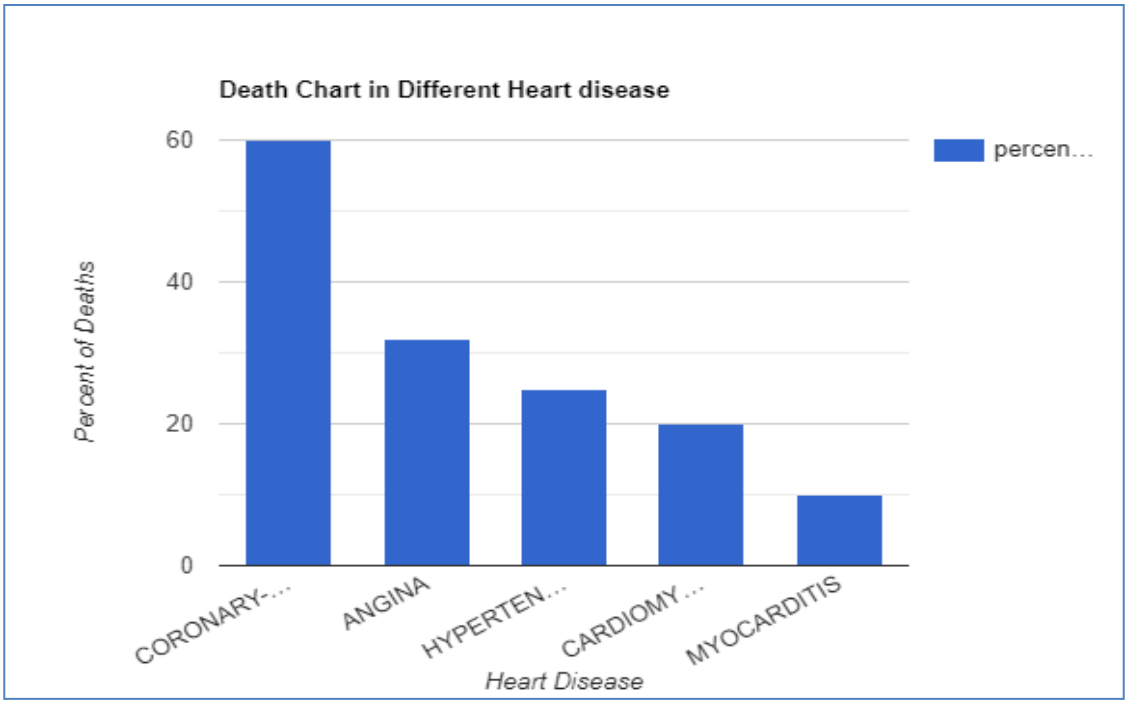

Fig. 5: Bar Chart Displaying the Percent of Deaths Caused due to different Type of Heart Diseases

Recognizing the preparing of crude social insurance information of heart data will help in the drawn out sparing of human lives and early identification of variations from the norm in heart conditions. Machine learning strategies were utilized in this work to process crude information and give another and novel acumen towards heart ailment. Heart malady forecast is testing and significant in the clinical field. Be that as it may, the death rate can be radically controlled if the malady is distinguished at the beginning periods and deterrent measures are embraced at the earliest opportunity.

DOI: 10.37200/IJPR/V24I5/PR2020796
Received: 24 Apr 2020 | Revised: 30 Apr 2020 | Accepted: 25 May 2020

VI. CONCLUSIONS

Measurable models are not helpful for the forecast of diseases as these models are unequipped for putting away straight out qualities and enormous number of missing qualities. Machine learning procedure (MLT) is utilized. It is utilized in mix with information mining procedures to group various types of diseases dependent on patient's records. Accuracy is significant for information mining in clinical industry. Various algorithms can be applied for identifying various kinds of diseases. This thus makes the framework canny. Progressively combinational models are developed to foresee the heart ailment which can help specialists in forecast of various kinds of heart diseases at a beginning time. It is analyzed that half breed DT works out the best for the given dataset as for the individual algorithms. Along these lines, it advances better outcomes for expectation of heart ailment. For Further we improve the forecast accuracy.

REFERENCES

[1] Bhatnagar, Prachi, et al. "Trends in the epidemiology of cardiovascular disease in the UK." *Heart* 102.24 (2016): 1945-1952.

[2] Zhao, Lina, et al. "Enhancing Detection Accuracy for Clinical Heart Failure Utilizing Pulse Transit Time Variability and Machine Learning." *IEEE Access* 7 (2019): 17716-17724.

[3] Borkar, Sneha, and M. N. Annadate. "Supervised Machine Learning Algorithm for Detection of Cardiac Disorders." *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). IEEE*, 2018.

[4] Omar Boursalie, Reza Samavi, Thomas E. Doyle. "M4CVD: Mobile Machine Learning Model for Monitoring Cardiovascular Disease." *Procedia Computer Science*, 63 (2015): 384-391.

[5] Chen, Rui, et al. "Using Machine Learning to Predict One-year Cardiovascular Events in Patients with Severe Dilated Cardiomyopathy." *European Journal of Radiology* (2019).

[6] Dhar, Sanchayita, et al. "A Hybrid Machine Learning Approach for Prediction of Heart Diseases." *2018 4th International Conference on Computing Communication and Automation (ICCCA). IEEE*, 2018

[7] Dinesh, Kumar G., et al. "Prediction of Cardiovascular Disease Using Machine Learning Algorithms." *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT). IEEE, 2018.*

[8] Terrada, Oumaima, et al. "Classification and Prediction of atherosclerosis diseases using machine learning algorithms." *2019 5th International Conference on Optimization and Applications (ICOA). IEEE*, 2019.

[9] Fathalla, Karma M., et al. "Cardiovascular risk prediction based on Retinal Vessel Analysis using machine learning." *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE*, 2016.

DOI: 10.37200/IJPR/V24I5/PR2020796
Received: 24 Apr 2020 | Revised: 30 Apr 2020 | Accepted: 25 May 2020