Design Engineering

ISSN: 0011-9342 | Year 2021
Issue: 6 | Pages: 4119- 4133

ENHANCED WEIGHTED QUADRATIC RANDOM FOREST ALGORITHM FOR HEART DISEASE PREDICTION

[1]Ms. C. Keerthana, [2]Dr.B.Azhagusundari

Assistant professor, Department of computer science,
Nallamuthu Gounder Mahalingam College,
Pollachi, Tamilnadu.

Abstract – Heart disease is the leading reason for death in the U.S. Sooner or later in your life, possibly you or one of your friends and family will be forced to settle on decisions about some part of heart disease. Heart disease can strike out of nowhere and expect you to settle on decisions rapidly. In this phase, the proposed Enhanced Weighted Quadratic Random Forest Algorithm is applied to patient heart disease data with high-dimensional lopsided attributes. First, the random forest algorithm is utilized to arrange feature significance and diminish dimensions. Second, the chose features are utilized with the random forest algorithm and the F-measure esteems are determined for every decision tree as weights to construct the prediction model for patient heart disease data. Weighted F-measure into the RF algorithm, which creates a superior performance for patient heart disease prediction by assigning different weights to different decision trees with the proper version of proposed classifiers, this technique could thus build up a most ideal measure of covered up units for guaranteed.

Keywords - Random Forest Algorithm, Weighted Quadratic, Classification, F-Measure, Data Mining.

1. INTRODUCTION

The health of a human heart depends on the encounters in an individual's life and is totally subject to professional and individual practices of an individual. There may likewise be a few hereditary factors through which a kind of heart disease is passed down from ages. According to the World Health Organization, consistently in excess of 12 million passing are occurring worldwide because of the different sorts of heart diseases which are additionally known by the term cardiovascular disease. the heart disease can be anticipated for certain essential ascribes taken from the patient and in their work have introduced a framework that includes the qualities of an individual person dependent on absolutely 13 fundamental credits like sex, circulatory strain, cholesterol and others to foresee the probability of a patient getting affected by heart disease. They have added two additional credits for example fat and smoking conduct and expanded the exploration dataset. The data mining classification algorithms, main inspiration of doing this examination is to introduce a heart disease prediction model for the prediction of event of heart disease. Further, this exploration work is pointed towards identifying the best classification algorithm for identifying the chance of heart disease in a patient. This work is justified by performing a relative report. A key test confronting healthcare organization (hospitals, medical focuses) is the facility of quality services at sensible costs. Quality conveniences propose diagnosing patients precisely and regulating drugs that are effective. Poor clinical decisions can incite wretched outcomes, which are thusly unsatisfactory. Hospitals should restrict the expense of clinical tests. Data Mining is an undertaking of extracting the essential decision-making information from a group of past records for future investigation or prediction. The information might be covered up and isn't identifiable without the utilization of data mining. The classification is one data mining technique through which the future result or predictions can be made dependent on the chronicled data that is accessible. The medical data mining made a potential answer for integrate the classification techniques and give electronic training on the dataset that further prompts exploring the secret examples in the medical data sets which is utilized for the prediction of the patient's future state. In this way, by using medical data mining it is feasible to give insights on a patient's set of experiences and can offer clinical help through the examination. For clinical investigation of the patients, these examples are a lot of fundamental. In Basic English, the medical data mining utilizes classification algorithms that are a fundamental part for identifying the chance of heart assault before the event.

Heart disease

Heart disease is the leading reason for death in the U.S. Eventually in your life, possibly you or one of your friends and family will be forced to settle on decisions about some part of heart disease. Knowing something about the life structures and functioning of the heart, specifically how angina and heart attacks work will empower you to settle on informed decisions about your health. Heart disease can strike out of nowhere and expect you to settle on decisions rapidly.

Data mining is interaction of extracting useful information from enormous measure of databases. Data mining is generally useful in an exploratory investigation on account of nontrivial information in huge volumes of data. Data mining is the way toward extracting data for finding covered examples which can be transformed into significant. Data mining information afford a user-situated way to deal with new and disguised examples in the data. The information which is uncovered can be utilized by the healthcare professionals to improve quality of service and to lessen the degree of unfavourable medicine effect. Hospitals need to diminish the charge of medical tests. Medical consideration organizations should have ability to investigate data. Treatment records of millions of patients can be accumulated and data mining techniques will help in answering various fundamental and unequivocal inquiries interrelated to health care. Data mining techniques has been performed in healthcare domain. This acknowledgment is in the stimulate of blast of difficult medical data.

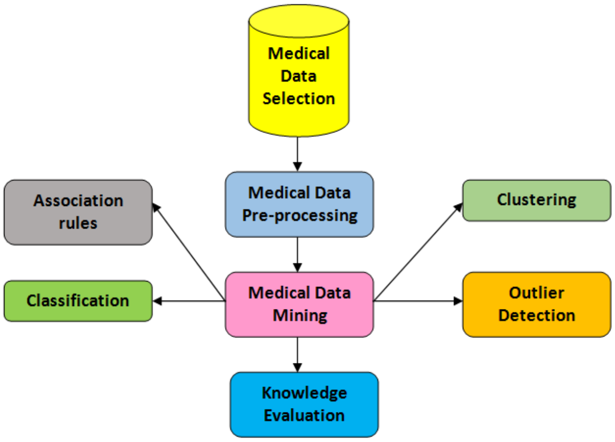

Figure 1.Heart disease Detection using Data Mining

Medicinal data mining can use the hidden examples present in enormous medical data which in any case is left unseen. Data mining techniques which are useful to medical data include affiliation rule mining for finding frequent examples, prediction, classification and clustering. Data mining techniques are more useful in predicting heart diseases, breast cancer lung cancer, diabetes and so forth.

2. EXISTING APPROACHES

2.1 Coactive Neuro-Fuzzy Inference System (CANFIS)

Parthiban, et al. have proposed a new work in which the heart disease is identified and

Predicted using the proposed Coactive Neuro-Fuzzy Inference System (CANFIS). Their model works dependent on the aggregate idea of neural network versatile capacities and dependent on the hereditary algorithm alongside fuzzy logic in request to diagnose the event of the disease. The performance of the proposed CANFIS model was assessed as far as training Performances and classification correctness's. Finally, their outcomes show that the proposed CANFIS model has incredible planned in predicting the heart disease.

2.2 clustering algorithm (K-Means) and one hierarchical clustering algorithm (agglomerative)

Singh, et al. has done a work using, one partition clustering algorithm (K-Means) and one hierarchical clustering algorithm (agglomerative). K-means algorithm has higher effectiveness and scalability and merges fast when production with huge data sets. Hierarchical clustering builds a progressive system of clusters by either frequently merging two more modest clusters into a bigger one or splitting a bigger bunch into more modest ones. Using WEKA Data mining tool, they have determined the performance of k-means and hierarchical clustering algorithm based on accuracy and running time.

2.3 k-means clustering with Naive Bayes

Shouman et.al presented work by integrating k-means clustering with Naive Bayes using different initial centroid selection to improve the Naive Bayes accuracy for diagnosing heart disease patients and accuracy was 84.5%. Rupali et al. decision support in Heart Disease Prediction.

2.4 Intelligent Heart Disease Prediction System (IHDPS)

Palaniappan, et al. has carried out a research work and have built a model known as Intelligent Heart Disease Prediction System (IHDPS) by using several data mining techniques

such as Decision Trees, Naïve Bayes and Neural Network.

2.5 Multi-Layer Perceptron with Back-Propagation

Shantakumar, et al. have done a research work in which the intelligent and effective heart attack prediction system is created using Multi-Layer Perceptron with Back-Propagation. Accordingly, the frequency examples of the heart disease are mined with the MAFIA algorithm dependent on the data separated.

2.6 classification method based on the origin of multi parametric features

Yanwei, et.al have built a classification method dependent on the origin of multi parametric features by assessing HRV (Heart Rate Variability) from ECG and the data is pre-prepared and heart disease prediction model is constructed that classifies the heart disease of a patient.

3. PROPOSED MODEL

This chapter proposed an improved RF algorithm, the WQRF dependent on the weighted F-measure. The main idea is to follow two stages. In the first place, the random backwoods algorithm is utilized to order feature significance and reduce dimensions. Second, the selected features are utilized with the random woodland algorithm and the F-measure esteems are determined for every decision tree as loads to assemble the prediction model.

Random Forest Algorithm

Random Forest is a supervised learning algorithm. Random forest can be utilized for both classification and regression problems, by utilizing random forest regressor we can utilize random forest on regression problems. Yet, we have utilized random forest on classification in this phase. So this phase will just consider the classification part.

Random Forest pseudo code

Step 1. Randomly select “k” features from total “m” features. Where k << m.

Step 2. Among the "k" features, calculate the node "d" using the best split point.

Step 3. Split the node into daughter nodes utilizing the best split.

Step 4. Repeat 1 to 3 steps until "l" number of nodes has been reached.

Step 5. Construct forest by repeating steps 1 to 4 for "n" number times to create "n" number of trees.

Random forest prediction pseudo code

Step 1. Takes the test features and utilize the guidelines of each randomly created decision tree to predict the outcome and stores the predicted outcome (target).

Step 2. Calculate the votes for each predicted target.

Step 3. Consider the high voted predicted target as the final prediction from the random

Forest algorithm.

Accuracy score of Random Forest is 86.9%

Classifier Evaluation Index

The basic assessment records for the prediction model's performance are accuracy (ACC), recall, precision (PPV), and the region under the bend (AUC). To ascertain these records, the disarray network is utilized. In the grid, the segments represent the prediction categories and the amount of the worth in the section is the information observations in the class. Moreover, the lines in the framework represent the actual categories and the amount of the qualities in the lines represents the information observations in that class. In this section centre is around whether there is coronary illness, which is a binary grouping.

Heart disease is set as the positive category and no heart disease set as the negative category. As shown in below table, TP denotes that the actual heart disease is predicted as heart disease; FN denotes that the actual heart disease is predicted as no heart disease; TN denotes that actual no heart disease is predicted as no heart disease and FP denotes that actual no heart disease is predicted as heart disease.

Prediction			
Actual		Heart Disease	No Heart Disease
	Heart Disease	TP	FN

	No Heart Disease	**FP**	**TN**

Recall means the true positive rate (TPR) and the equation is

$$\text{Recall} = \text{TPR} = \frac{TP}{(TP+FN)} \qquad (1)$$

FPR means the false positive rate the equation is

$$\text{FPR} = \frac{FP}{(FP+TN)} \qquad (2)$$

Precision means the positive predictive value (PPV) and the equation is

$$\text{PPV} = \frac{TP}{TP+FP} \qquad (3)$$

ACC means accuracy and the equation is

$$\text{ACC} = \frac{(TP+TN)}{(TP+FP+FN+TN)} \qquad (4)$$

The AUC signifies the area under the receiver operating characteristics curve (ROC). It is a significant index for passing judgment on the benefit and detriment of a binary prediction model. In the event that its worth is greater, the exhibition of the model is better. The distinctions in the marks of the ROC mirror the various reactions to a similar signal stimulation. Likewise, the x-coordinate of the ROC curve is FPR and the y-coordinate is recall.

Proposed Improved Weighted Random Forest Algorithm

Ordinarily, all decision trees of the RF have a similar weight value while voting for the classification. Nonetheless, it has a fatal defect when utilized with the unbalanced data classification prediction. To tackle this, we bring the weighted F-measure into the RF algorithm, which produces a superior presentation for worker heart disease prediction by appointing various weights to various decision trees. From the perspective of data mining, the issue in representative heart disease predictions is the twofold classification of the unbalanced data.

This stage set "heart disease" as the positive class, while "not leaving" as the negative classification. Clearly the positive class is minor and the negative classification is the major class. It isn't adequate to quantify the exhibition of the model for exactness with unbalanced data. For instance, if an organization's patients heart disease rate is 2% and nobody is required to leave, the precision rate could be just about as high as 98%, be that as it may, this doesn't bode well here.

Since this research focus on minor categories, regardless of whether every one of the minor categories are falsely classified as major categories, the precision is still exceptionally high however the model has no incentive for patient-heart disease research. In this examination, it is smarter to misconceive a patient who has no goal of leaving as a potential departure than to neglect an individual with a certifiable aim of leaving. This research joins the two evaluation indices of precision and recall and uses the harmonic mean of the F-measure to assess the performance of every decision tree and compute the heaviness of the vote. Contrasted and the regular RF algorithm, this refreshed algorithm improves the performance of the unbalanced data classification.

The F-measure here refers to F1 (in particular, α is set as one) and its formula is in condition 1. The F-measure joins the after effects of precision and recall and the higher the F-measure, the better the classification performance.

$$\textbf{F-measure} = \frac{2\times recall\times precision}{recall+precision} = \frac{2TP}{2TP+FP+FN}$$

The algorithm follows the steps as follows.

Step 1. Affirm the training set, validation set, and test set. Arbitrarily extricate K+1 datasets from the first dataset, with known classifications, by the bootstrap method. The limit of each dataset is n, specifically, equivalent to the first dataset. Among the K+1 datasets, K sets are utilized as training sets and the excess set is utilized as the validation set. The example that has not been drawn addresses about 33% of the absolute example of the first training set and establishes the test set.

Step 2. Develop the RF classifier. Info K training sets and utilize the RF algorithm to fabricate the model; apply the combination classifier made out of K characterization decision trees.

Step 3. Get the weight value by computing the F-measure of the sub classifier. Info the validation set and classifies each example in the validation set by with respect to each decision tree in the forest as an independent classifier. At that point get the TP, FP, FN, TN, precision rate, and recall rate values of each classifier. Then, compute the F-measure, comparing to the weight of each sub classifier. The weight of the sub classifier for j is as appeared in (7).

Step 4. Info the test set to assess the performance of the model.

Step 5. Information the unclassified samples. Classify the samples by the F-measure weighted random forest. The outcome H relies upon the weighted vote of the classification after-effects of each sub classifier. The classification consequence of sub classifier j is and the classification weight after-effect of sub classifier j is,

The final classification decision can be communicated as

$$W_j = F_j = \frac{2Tp_j}{2TP_{J+FP_{j+FN_j}}} \qquad \textbf{(7)}$$

$$\mathbf{H}\ (x) = \underset{y}{arg\,max} \sum_{j=1}^{K} W_j\ \mathbf{I}\ (h_{j(x)}=\mathbf{Y}) \qquad (8)$$

In (8), H(x) addresses the combined classification model got by the weighted RF algorithm is the sub classifier (i.e., single choice tree), Y addresses the yield variables (i.e., the classification type), and function I is the indicator function.

In the patient heart disease prediction question, parameter Y has two alternatives: disease and not disease. Consequently, in (8), when Y signifies disease, every one of the weighted values of the sub classifier, named heart disease, will be added together as the score of H(x). Then again, when Y signifies not disease, every one of the weighted values of the sub classifiers, delegated not heart disease, will be added together as the score of H(x). The correlation of the two scores and the relating estimation of Y of the maximum values of the two scores address the prediction classification values of the combined classification model. Figure 2 presents the construction of the weighted random forest.

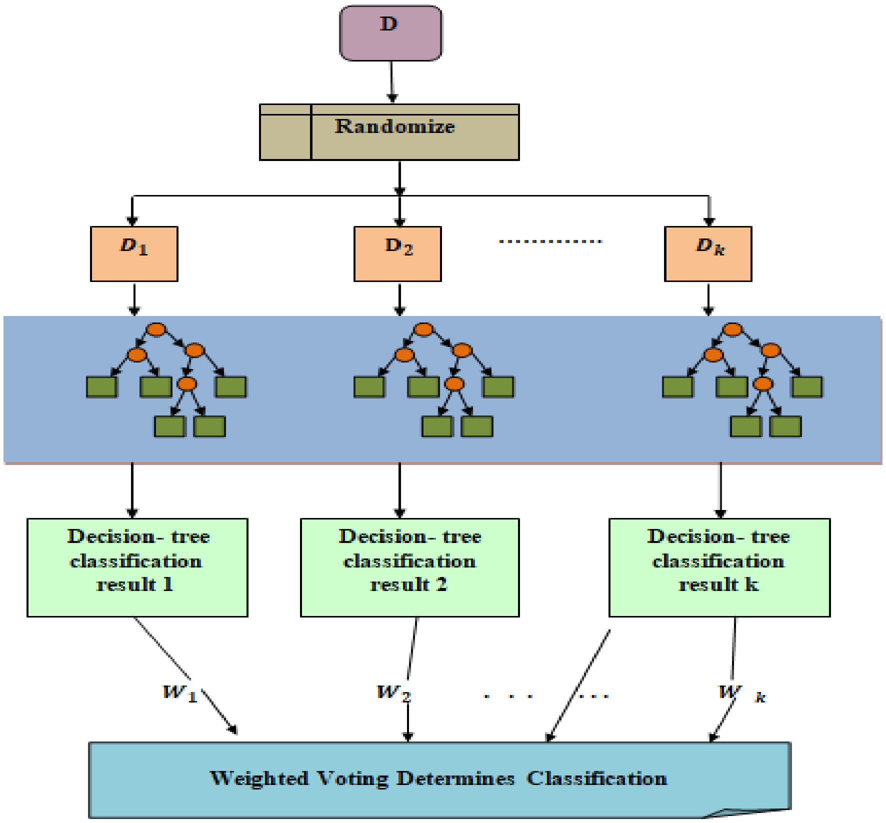
D
Randomize
D_1
D_2
....
D_k
Decision- tree classification result 1
Decision- tree classification result 2
Decision- tree classification result k
W_1
W_2
. . .
. . .
W_k
Weighted Voting Determines Classification

Figure 2.Structure of weighted random forest.

The enhanced weighted quadratic random forest algorithm (EWQRF) Method

Here, by introducing the combination forecasting hypothesis into this field of data mining, the current patient data (leaving and not leaving patients and the leaving status names) are trained and displayed to foresee whether a patient will leave the work in the future.

The EWQRF algorithm is proposed to fabricate the prediction model. The original data (recorded) are isolated into the training set and the validation set randomly. Likewise, the not chose data (OOB) are utilized as the test set. In the first calculation, the training set is utilized to rank the features by significance applying the RF algorithm. The m of the main features for patient heart disease is chosen. In the subsequent calculation, these m features are included in the weighted RF algorithm with the training set to assemble the prediction model of patient turnover. The weights in the weighted RF algorithm are obtained by using the validation set to figure the F-measure of each tree. For any patient, the m features are obtained and input into the prediction model, so the intention of the patient to leave in the future can be predicted. The prediction model can be assessed by a 10-fold cross validation technique to obtain the accuracy, sensitivity, exactness, and AUC. In the following area, the trial will approve the performance of the algorithm proposed here; it will show that it is in reality better compared to basic predictive algorithms like the RF, the decision tree (e.g., the C4.5 algorithm), the Logistic relapse (Logistic), and back propagation (BP) neural network. Figure 3 is a schematic diagram of the EWQRF algorithm.

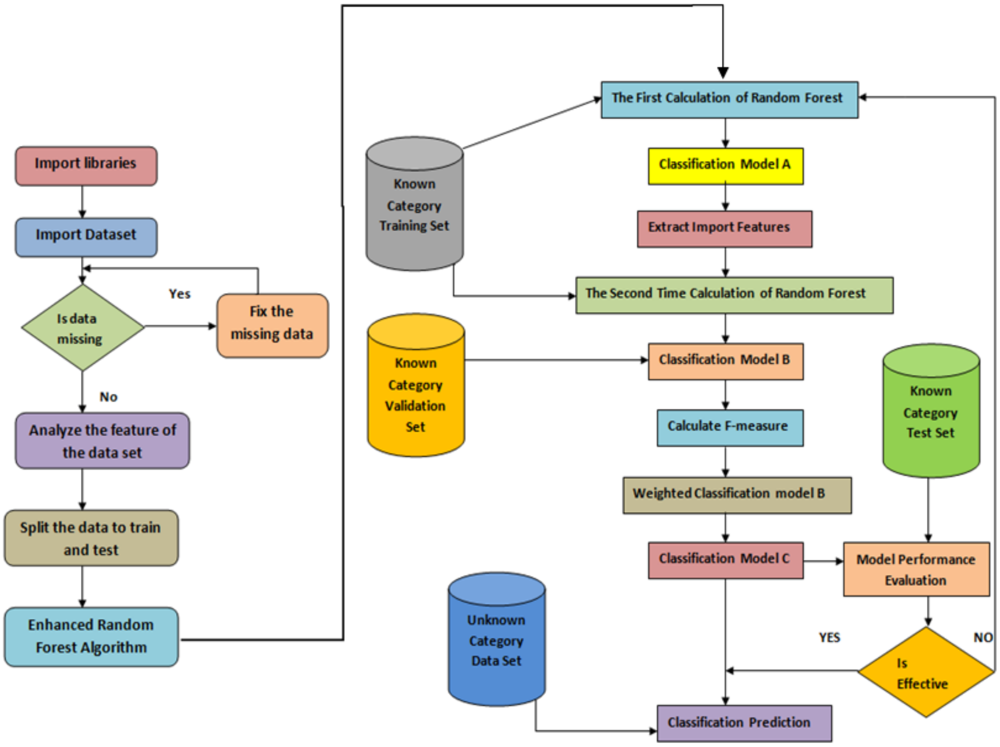
Import libraries
Import Dataset
Is data missing
Yes
Fix the missing data
No
Analyze the feature of the data set
Split the data to train and test
Enhanced Random Forest Algorithm
Known Category Training Set
The First Calculation of Random Forest
Classification Model A
Extract Import Features
The Second Time Calculation of Random Forest
Known Category Validation Set
Classification Model B
Calculate F-measure
Weighted Classification model B
Known Category Test Set
Classification Model C
Model Performance Evaluation
YES
NO
Is Effective
Unknown Category Data Set
Classification Prediction

Figure 3.Flowchart of weighted quadratic random forest algorithm

4. EXPERIMENTAL RESULTS

1. Sensitivity

The sensitivity measures the chance of positive classification of the patient as being sick to the measure of positive theory. It presents how often one finds what is one looking for. The sensitivity is a rate extent of True Positive cases to the amount of every positive speculation (True Positive and False Negative). The sensitivity is introduced by the formula.

Sensitivity = TP / (TP + FN)

Values	C4.5 (%)	Logistic Regression (%)	Proposed EWQRF (%)
TP	77.4	80.4	82.3
TN	85.1	87.3	89.1
FP	90.2	92.6	94.7
FN	95.4	97.5	98.5

Table 1.Comparision table of Sensitivity

The Comparison table 1 of Sensitivity of Value explains the different values of existing algorithms C4.5, Logistic Regression and proposed improved EWQRF Algorithm. While comparing the Existing algorithm, the proposed improved EWQRF provides the better results. The existing algorithm (C4.5, Logistic Regression) values start from 77.4 to 95.4, 80.4 to 97.5 and proposed EWQRF Algorithm starts from 82.3 to 98.5, provides the great results

.

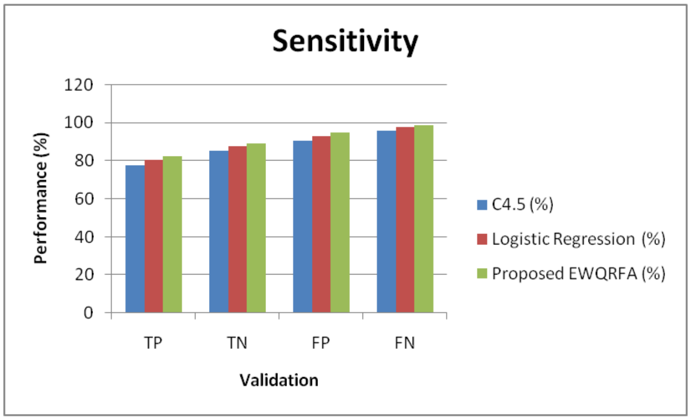

Figure 4.Comparision chart of Sensitivity

Figure 4 shows the comparison chart of Sensitivity values demonstrates the existing systems C4.5, Logistic Regression and proposed EWQRF algorithm. The proposed algorithm is better than the existing algorithm. X axis denotes the validation values and Y axis denotes the performance in %. The existing algorithm (C4.5, Logistic Regression) values start from 77.4 to 95.4, 80.4 to 97.5 and proposed EWQRF Algorithm starts from 82.3 to 98.5, provides the great results.

2. Specificity

The specificity, the following measure, shows the chance of right, negative classification of the patient as being healthy to all negative speculations. The specificity is a rate extent of True Negative cases to the amount of every single negative theory (False Positive and True Negative). The specificity is introduced by the formula,

$$\textbf{Specificity} = \textbf{TN} / (\textbf{TN} + \textbf{FP})$$

Values	C4.5 (%)	Logistic Regression (%)	Proposed EWQRF (%)
TP	53.2	57.4	67.8
TN	72.6	81.3	83.2
FP	55.2	65.9	86.4
FN	90.4	78.5	95.7

Table 2.Comparision table of Specificity

The Comparison table 2 of Specificity of Value explains the different values of existing algorithms C4.5, Logistic Regression and proposed improved EWQRF Algorithm. While comparing the Existing algorithm, the proposed improved EWQRF provides the better results. The existing algorithm (C4.5, Logistic Regression) values start from 53.2 to 90.4, 57.4 to 78.5 and proposed EWQRF Algorithm starts from 67.8 to 95.7, provides the great results.

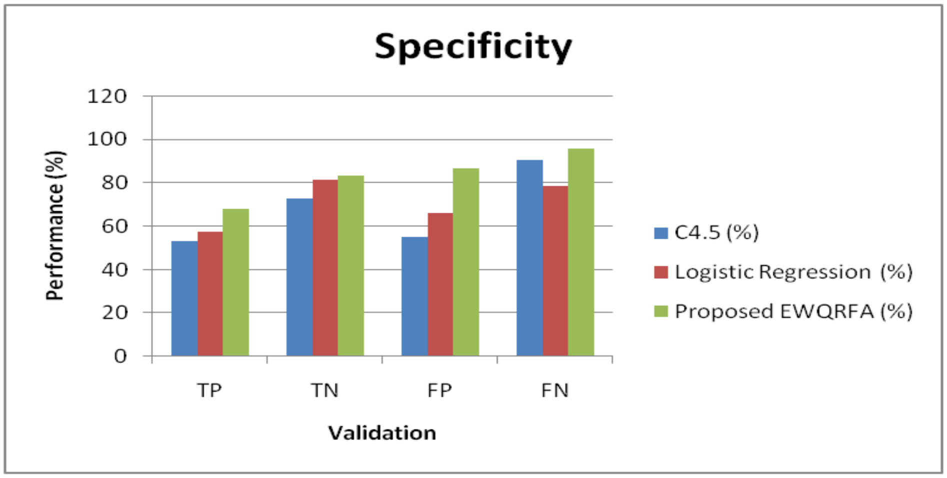

Figure 5.Comparision chart of Specificity

Figure 5 shows the comparison chart of Specificity values demonstrates the existing systems C4.5, Logistic Regression and proposed EWQRF algorithm. The proposed algorithm is better than the existing algorithm. X axis denotes the validation values and Y axis denotes the performance in %. . The existing algorithm (C4.5, Logistic Regression) values start from 53.2 to 90.4, 57.4 to 78.5 and proposed EWQRF Algorithm starts from 67.8 to 95.7, provides the great results.

3. Accuracy

The third measure of accuracy is called predictive accuracy. This measure shows the proportion of accurately classified cases to all cases in the set. The bigger the predictive accuracy, the better the performance is. The predictive accuracy is introduced by the formula.

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

Values	C4.5 (%)	Logistic Regression (%)	Proposed EWQRF (%)
TP	77.4	81.4	83.3
TN	85.1	88.3	90.9
FP	92.7	93.6	95.1
FN	96.2	98.4	99.6

Table 3.Comparision table of Accuracy

The Comparison table 3 of Accuracy of Value explains the different values of existing algorithms C4.5, Logistic Regression and proposed improved EWQRF Algorithm. While comparing the Existing algorithm, the proposed improved EWQRF provides the better results. The existing algorithm (C4.5, Logistic Regression) values start from 77.4 to 96.2, 81.4 to 98.4 and proposed EWQRF Algorithm starts from 83.3 to 99.6, provides the great results.

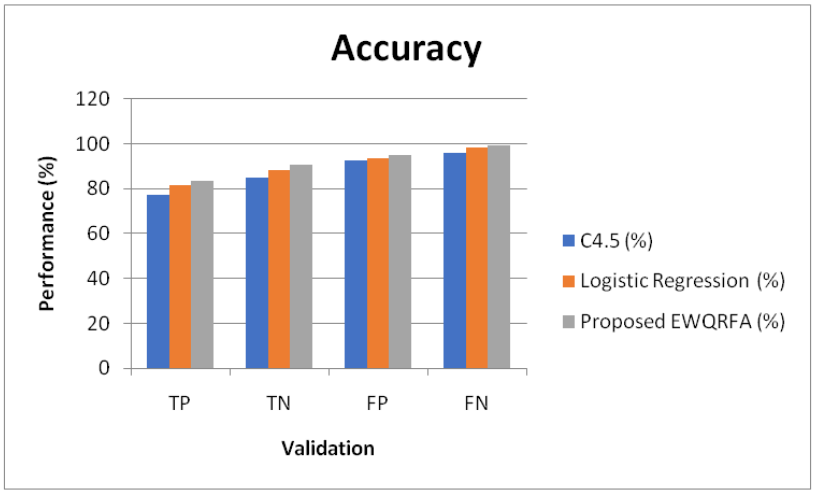

Figure 6.Comparision chart of Accuracy

Figure 6 shows the comparison chart of Specificity values demonstrates the existing systems C4.5, Logistic Regression and proposed EWQRF algorithm. The proposed algorithm is better than the existing algorithm. X axis denotes the validation values and Y axis denotes the performance in %. The existing algorithm (C4.5, Logistic Regression) values start from 77.4 to 96.2, 81.4 to 98.4 and proposed EWQRF Algorithm starts from 83.3 to 99.6, provides the great results.

The classification errors False Positive Rate (FPR) and False Negative Rate (FNR) were utilized to measure the errors brought about by the classification technique. FPR was the quantity of things incorrectly marked as belonging to the positive class and FNR was the things which were not named as belonging to the positive class yet ought to have been by the classification strategy which is given by

$$\textbf{FPR} = \frac{|\boldsymbol{FP}|}{|\boldsymbol{TN}|+|\boldsymbol{FP}|}$$

$$\textbf{FNR}=\frac{|\boldsymbol{FN}|}{|\boldsymbol{TP}|+|\boldsymbol{FN}|}$$

5. CONCLUSION

Treatment records of millions of patients can be accumulated and data mining techniques will help in answering various fundamental and unequivocal inquiries interrelated to health care. Data mining techniques has been performed in healthcare domain. The main goal of this phase was to build up an Enhanced Weighted Quadratic Random Forest Algorithm for predicting the danger of heart disease to a patient with the medical records got from the patients. By exploiting this distinct feature of the Proposed, an algorithm for prediction was constructed which would accurately anticipate heart disease and was likewise efficient. If hospitals can anticipate ahead of time which patients may heart disease, it can build up an arrangement and take measures to lessen this chance, address the need to medicine substitutions quickly, and make different changes in accordance with fix patients in key positions. Using the EEWQRF approach, specialists can anticipate better patient and take ideal activity.

REFERENCES

[1]. Ankita Dewan and Meghna Sharma, Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classi_cation, vol. 15, pp. 13-24, 2014.

[2]. K. Thenmozhi and P. Deepika, Heart Disease Prediction Using Classi_cation with Di_erent Decision Tree Techniques, pp. 2227-2235, 2011.

[3]. G.V. Nadiammai and M. Hemalatha, E_ective approach toward Intrusion Detection System using data mining techniques, vol. 15, pp. 371750, 2014.

[4]. M. Tavallaee, E. Bagheri, W. Lu and A Ghorani, A detailed anaylysis of the KDD CUP 99 data set, 2009.

[5]. Siva S. Sivanath, S. Geetha and A. Kannan, Decision tree based light weight intrusion detection using a wrapper approch, 2012.

[6]. G.V. Nadiammai and M. Hemalatha, E_ective approach toward Intrusion Detection System using data mining techniques, vol. 15, pp. 371750, 2014.

[7]. Robert Mitchell and Ing-Ray Chen, A survey of intrusion detection in wireless network applications, vol. 42, pp. 11723, 2014.

[8]. Panos Louvieris, N, Natalie Clewley and Xiaohui Liu, E_ects-based feature identi_cation for network intrusion detection, pp. 26517273, 2013.

[9]. G.M. Nasira and N. Hemageetha, "Vegetable Price Prediction Using Data Mining Classification Technique", Proceedings of the International Conference on Pattern Recognition Informatics and Medical Engineering, March 21–23, 2012.

[10]. X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu, Q. Wang, et al., "A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method", Computational and Mathematical Methods in Medicine, vol. 2017, pp. 11, [online] Available: https://doi.org/10.1155/2017/8272091.

[11]. C. B. C. Latha and S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques", Informatics in Medicine Unlocked, vol. 16, 2019.

[12]. C. Gazeloglu, "Prediction of heart disease by classifying with feature selection and machine learning methods", Progress in Nutrition, vol. 22, no. 2, pp. 660-670, 2020.

[13]. Heart Disease UCI, June. 2020, [online]Available:https://www.kaggle.com/ronitf/heartdisease-uci.

[14]. L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[15]. S. Malek, R. Gunalan, S. Kedija et al., “Random forest and Self Organizing Maps application for analysis of pediatric fracture healing time of the lower limb,” Neurocomputing, vol. 272, pp. 55–62, 2018.

[16]. S. Wang, X. Liu, T. Yang, and X. Wu, “Panoramic crack detection for steel beam based on structured random forests,” IEEE Access, vol. 6, pp. 16432–16444, 2018.

[17]. N. Guru, A. Dahiya and N. Rajpal, "Decision Support System for Heart Disease Diagnosis using Neural Network", Delhi Business Review, vol. 8, no. 1, pp. 1-6, 2007.

[18]. S. Palaniappan and R. Awang, "Intelligent Heart Disease Prediction System using Data Mining Techniques", International Journal of Computer Science and Network Security, vol. 8, no. 8, pp. 1-6, 2008.

[19]. S. B. Patil and Y.S. Kumaraswamy, "Intelligent and Effective Heart Attack Prediction System using Data Mining and Artificial Neural Network", European Journal of Scientific Research, vol. 31, no. 4, pp. 642-656, 2009.

[20]. Sri HartatiKusrini, Implementation of c4.5 algorithm to evaluate the cancellation possibility of new student applicants at stmikamikomyogyakarta, pp. 17-19, 2007.

[21]. Sellappan Palaniappan and Rafiah Awang, "Intelligent Heart Disease Prediction System Using a Data Mining Techniques", IJCSNS International Journal of Computer Science and Network Security, vol. 8, no. 8, August 2008.

[22]. S.H Ishtake and S.A. Sanap, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", International J. of Healthcare & Biomedical Research, vol. 1, no. 3, pp. 94-101, April 2013.

[23]. R. Chitra and V. Seenivasagam, "review of heart disease prediction system using data mining and hybrid intelligent techniques", ICTACT journal on soft computing, vol. 03, no. 04, july 2013.

[24]. N. Singh and D. Singh, Performance Evaluation of K-Means and Hierarchal Clustering in Terms of Accuracy and Running Time, 2012.

[25]. J. R. Quinlan, "Induction of Decision Trees", Machine Learning, vol. 1, no. 1, pp. 81-106, 1986.

[26]. J. U. Ansari and S. Dipesh, "Predictive data mining for medical diagnosis:an overview of heart disease prediction‖", Int. J. Comput. Appl, vol. 17, no. 8, pp. 0975-8887, March 2011.

[27]. H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases", Expert Systems with Applications, vol. 35, no. 1-2, pp. 82-89, 2000.

[28]. H. T. Malazi and M. Davari, “Combining emerging patterns with random forest for complex activity recognition in smart homes,” Applied Intelligence, vol. 48, no. 2, pp. 315–330, 2018.

[29]. F. B. de Santana, A. M. de Souza, and R. J. Poppi, “Visible and near infrared spectroscopy coupled to random forest to quantify some soil quality parameters,” Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 191, pp. 454–462, 2018.

[30]. R. Anitha and S. R. D. Siva, “Development of computer-aided approach for brain tumor detection using random forest classifier,” International Journal of Imaging Systems and Technology, vol. 28, no. 1, pp. 48–53, 2018.