ISSN 2278-3091
Volume 9, No.4, July – August 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse375942020.pdf
https://doi.org/10.30534/ijatcse/2020/375942020

Heart Disease Prediction Using Machine Learning Model Ensemble-Random Forest with Simple Regression

Dr.K.Venkata Subba Reddy[1], Anusha Surakarapu[2], Dr.Mohammed Umar Khan[3] , Sunitha Avula[4]
[1]Kallam Haranadhareddy Institute of Technology, India,:kvsreddy2012@gmail.com
[2]Kallam Haranadhareddy Institute of Technology, India, anushasurakarapu@gmail.com
[3]Kallam Haranadhareddy Institute of Technology, India, v.umar2009@gmail.com
[4]Kallam Haranadhareddy Institute of Technology, India, sunitha.avula@gmail.com

ABSTRACT

Heterogeneous Criterion based Decision-making is the need of the hour when dealing with clinical data analysis such as Heart Disease data. A Machine Learning (ML) based model can effectively perform complex analysis and then predict the presence or absence of heart disease in an individual accurately. This analysis will aid the doctors in their diagnosis and save countless human lives. This paper proposes a hybrid model Random Forest with Simple Linear Regression (RFSLR) for classifying individuals as healthy individual or Heart Disease individual. The experimental analysis done by this work shows that RFSLR outperforms the existing classifiers Linear Regression(LR), Support Vector Machine (SVM), Random Forest (RF), Naïve Bayes (NB) and DTs(DT) in terms of performance metrics. This paper also recognizes the best features from the Cleveland Heart Disease Dataset (CHDD) using three Feature Selection Algorithm (FSA) RELIEF-F, LASSO and FOCUS.

Key words: Machine Learning, Heart Disease Prediction, Feature Selection, Ensemble

1. INTRODUCTION

Heart disease is a major cause for suffering among human beings globally. World Health Organization (WHO) reports estimate that there are around 17.5 million deaths annually due to heart disease. Heart disease is the major reason of death for both male and female patients [1]. Annually, about 735,000 Americans have a heart attack. Of these, 525,000 are first timers and remaining 210,000 happen in people who have already had a heart attack. Heart stroke, cardiomyopathy, cardiovascular disease, murmurs, Artery blocks, congenital heart ailments, pulmonary stenosis, hypertension and rheumatic are some of the heart ailment classes. Prediction of heart disease is considered the most significant areas in the class of clinical data analysis [2]. With the advent of health clouds, hospitals can store huge amounts of patient data. ML algorithms are employed to analyze & assess this data for predicting & critical decision making. Heterogeneous criterion based decision-making (HCDM) should be performed for risk assessment & prediction in cases pertaining to heart diseases. Prediction of heart disease helps in effective analysis of the disease and there by save the lives of heart patients. ML algorithms enable machines to train from data, improve without being explicitly programmed [3]. Figure 1 displays the basic process of ML.

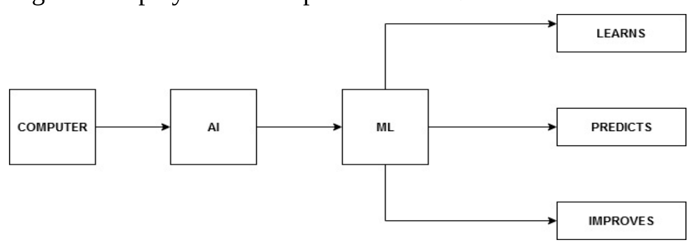

Figure 1: Machine Learning Process (MLP)

ML Models can be classified as Supervised, Unsupervised & Reinforcement Learning. This work proposes a novel supervised ML model, "RF with a Generalized Linear Model" for classification and prediction. The proposed Model is applied on Cleveland dataset. This data set is quite popular and extracted from a UCI ML repository[4].The results obtained are compared with that of other Supervised Learning algorithms like LR, SVM, RF, NB, DT and certain conclusions are drawn out.
The remainder of the paper is composed as follows; Section II throws light on the related work. Section III explains about the proposed model in detail, while Section IV discusses about the process and Section V performs result analysis and lastly Section VI draws out conclusions about the work and future enhancements that can be done on current work.

2. RELATED WORK

This section of the paper essays the already existing work on the topic. In the Literature many works with the objective of building an automotive heart disease Predictive System exist [5]. H.Alkeshuosh Et al. derived some rules based on particle Swarm Optimization and combined them with encoding techniques for detecting Heart Disease. P. S. Kumar et al.

applied Artificial Neural Networks for predicting heart diseases in people [6]. N.Al-milli used back-propagation neural network for prediction of heart disease with a good success rate [7]. Marjia Et al. through their work proposed a Heart Disease Prediction System which was based on K-Star, J48, SMO, and Bayes Net and Multilayer perception using WEKA Software [8]. B.S.S.Rathnayakc and G. U.Ganegoda used hybrid techniques for heart disease predicting. They amalgamated multiple techniques in to a one technique for carrying out their research [9]. P. K. Anooj used fuzzy logic techniques for carrying out their research on heart disease detection [10]. Data mining algorithms & intelligent fuzzy methods were employed by V. Krishnaiah Et al. for solving the issues in automated heart disease detection [11]. L. Baccour introduced Artificial Neural Networks to produce the highest accuracy prediction in the medical field [12]. R. Das Et al. used MLP of ANN to predict HD [13]. C. Cheng and H. Chiu developed an Artificial Neural Network (ANN) based system for heart disease prognosis on a nationwide database [14]. D. K. Ravish Et al. designed an ANN for monitoring heart, predicting and preventing heart disease [15]. W. Zhang and J. Han, in their research on heart disease detection used Convolution Neural Network with a satisfactory rate of accuracy [16]. M.J.Liberatore and R.L.Nydick, for their experiments on automated divination of HD used the CHD with a NN [17]. M. Durairaj and V. Revathi conducted experiments on heart disease forecasting using back propagation algorithm [18].

A.Gavhane Et al. presented their work on Heart disease Detection using ML Algorithms like DTs (DT), K-Nearest Neighbor Algorithm (KNN), NB (NB) and Genetic algorithm (GA) [19]. M.Ghazanfari and H.A.Esfahani implemented an ensembler classifier for classifying people with heart disease & without heart disease [20].The classifier analyzed symptoms like sex, pulse rate, age and other factors for its prediction. Similar methodology was applied by M.Gandhi and S.N.Singh in their work [21]. V.Krishnaiah Et al. presented a systematic review of existing research on predicting Heart disease by mixing data mining techniques and intelligent fuzzy approaches [22]. P.K.Anooj designed a Clinical Decision Support System to assist doctors for diagnosing heart related diseases. He used Weighted Fuzzy rules for building the System [23]. S.S.Salankar and J.P.Kelwade, applied Neural Network based Radial basis function on heart rate time series to detect cardiac arrhythmias [24]. A.S.Abdullah and R.R.Rajalaxmi employed Data mining techniques for predicting the events occurring in heart disease [25]. A.Davari Dolatabadi et al. automatically diagnosed heart diseases using Support Vector Machine for designing the model and K-fold Cross-Validation of dataset [26]. A.K.Dwivedi performed an evaluation of the ML algorithms for predicting the heart disease detection [27]. R.Detrano Et al. developed an algorithm based on probability theory [28]. K. Vanisree and J. Singaraju worked in this area by developing a computer based prognosis model which employed Neural Networks [29]. Kevin Buchanan Et al. proposed a system to predict heart disease from clinical narratives. The system extracted features using ontology-guided techniques and compared them with two classic feature selection techniques [30].

The current work proposes a hybrid algorithm, “RF with a General Linear Model (RFGLM)”, for designing an ML based Intelligent Decision System for early prognosis of Heart Disease. This work classifies people as healthy and sick using a hybrid RF predictive algorithm. FSAss like RELIEF-F, FOCUS, LASSO will be employed to reduce the dimensionality of the features. Moreover, data cleaning techniques were employed on the heart disease data. CHDD, 2016 is used for testing and training purpose [31]. Train Split Validation method is used to validate the proposed Classifier. Classification accuracy, classification error, specificity and sensitivity are the performance evaluation metrics used by the proposed Classifier [32].

3. PROPOSED METHODOLOGY

The main contributions of this work are-

1) Implementing feature extraction algorithms and optimization techniques to boost up the accuracy of the classification techniques for HD prognosis.
2) Performing the feasibility study of feature extraction algorithms with Classifier for accurate prediction of heart disease in human beings.
3) Designing a decision support system for intelligently classifying & predicting heart diseases.
4) Evaluating the classification accuracy of the proposed framework and execution time with industry standard methodologies.

3.1 SYSTEM FRAMEWORK

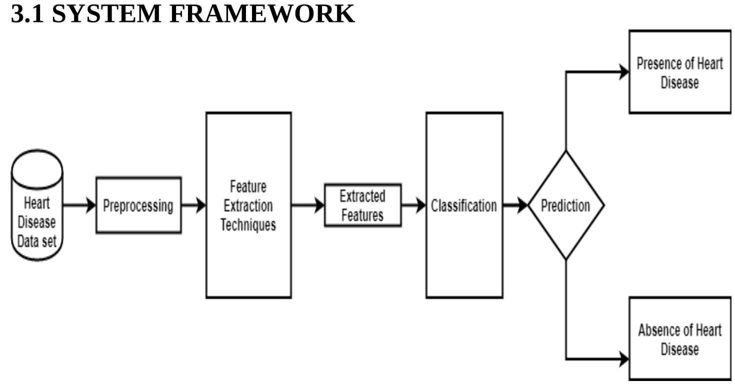

Figure 2: Framework of the proposed system

The proposed model is built with the objective to classify individuals with heart disease and without heart disease. Figure 2 shows the framework of the proposed system. Feature extraction techniques like RELIEF-F, FOCUS, and LASSO can be used to extract relevant features, and on these descriptors, the accuracy of the classification algorithms can be tested. This paper uses the CHDD, for making its investigations. This work proposes a novel model which implements Hybrid RF Algorithm with a Generalized Linear Model for classification and prediction. The results were compared with other popular Classifiers like LR, RF, ANN, SVM, DT, and NB. The technique of the proposed system can be divided into 5 stages viz. (1) Preprocessing, (2) Feature Extraction, (3) ML classifiers, (4) Prediction, (5) Result generation.

4. EXPERIMENTAL SETUP

4.1 Data Set

This work referred to popular databases like Cleveland Heart Disease Data, Z-Alizadeh Sani Data Set [32], AJA University of Medical Sciences [33] but finally selected the Cleveland Heart Disease Data because the other two data sets have data related to patients with a specific heart disease like coronary artery disease. The Cleveland Heart Disease Data is available in the UCI ML collection. It consists of 76 attributes extracted from 303 patients who have a medical record of heart disease. The attributes of the dataset and related information is listed as Table 1.

Table 1: Major Attribute list of CHDD

S.No	Attribute Name	Attribute Symbol
1.	Thallium scan	THA
2.	Exercise-induced angina	EIA
3.	Chest Pain class	CPT
4.	Peak Exercise Slope	PES
5.	Number of Prominent Vessels Colored by Fluoroscopy	VCA
6.	Maximum Heart Rate	MHR
7.	Serum Cholesterol	SCH
8.	Sex	SEX
9.	Fasting Blood Sugar	FBS
10.	Old Peak	OPK
11.	Resting Blood Pressure	RBP
12.	Age	AGE
13.	Resting ECG	RES

4.2. Data Processing

Data Preprocessing makes data relatively suitable for further analysis. It involves selection of attributes, modifying the attributes. Popular techniques are Feature Subset Selection, Aggregation, Feature Creation, Sampling, Dimensionality Reduction, Discretization, Binarization and Variable Transformation. Out of the 303 records, only 297 were considered for this work as there is some data missing in remaining 6 records. Out of the 76 attributes only 13 were deemed to be the most appropriate features for this research.

4.3. Feature Selection

The objective of any FSA is to recognize and select relevant descriptors according to a definition of relevance. Although, the meaning of relevance in ML is not established in a common manner [34].The basic objective of a FSAs is to improve the learning speed and representation so that the accuracy of the classifier increases [35]. Relevance [36] is the notion of being "relevant with respect to an objective", in our case classification objective. This work used RELIEF-F [37], Least Absolute Shrinkage and Selection Operator (LASSO) [38] and FOCUS [39] algorithms for the task of feature selection. The idea behind Relief algorithms is to predict the quality of attributes on the fact that how well the attribute can differentiate among instances that are close to each other. RELIEF-F is an extension of Relief. It can work with multiclass problems. LASSO extracts attributes that are generated by recording the absolute value of attributes coefficient.

4.4. Machine Learning Classifier

This work employed popular ML classification algorithms for classifying the individuals into healthy and sick categories. LR (LR), SVM, K-NN, NB (NB), DT (DT) [40] and a Hybrid algorithm RF with Simple LR (RFSLR) (proposed). The proposed hybrid classifier is a combination of RF and Simple LR models. A RF(RF) refers to a collection of one or more DTs like C4.5, CART and CHAID that use bagging technique for learning. The idea of a RF is to ensemble numerous models of DTs into a RF, and make use of the diversity of those trees for better accuracy [41].The RF Model is then mixed with Simple LR which is a General Linear Model, for enhanced accuracy in prediction. "The term General Linear Model (GLM) usually refers to conventional LR models for a continuous response variable given continuous and/or categorical predictors [42]".This work considered simple LR as the GLM [43]. The Simple LR model assumes that multi-collinearity, Auto-correlation in the data. It also assumes that there is a linear relationship between response and failure variables.

4.4.1 Simple Linear Regression (SLR)

"LR assumes a linear or straight line relationship between the input variables (X) and the single output variable (Y)".

$$y = p_0 + p_1 * x \qquad \textit{Equation 1}$$

Where p_0, p_1 are the coefficients that are calculated from the training data.

4.4.2. Random Forest Algorithm Working

"For every tree in the forest, a bootstrap sample is chosen from S where S(i) represents the i^{th} bootstrap". A modified decision-tree learning algorithm is trained. The basic DT algorithm is changed as follows: "*At each node of the tree, the algorithm randomly chooses an extract of the features f ⊆ F. where F is the set of features. The node then splits on the best feature in f rather than F. f is relatively minute than F*". Figure 3 depicts below depicts the RF Algorithm.

4.4.3. RandomForest Algorithm with Simple Linear Regression

This work creates an ensemble of the above two discussed Classifiers. Figure 4 algorithm describes the process of creating an ensemble of Classifiers. For this work, data set ***U*** is Cleveland Heart disease data, ***x*** the no. of classifiers is two, and ***P*** = {***RF,SLR***}. The output of the algorithm is an ensemble ***C*** that is a Hybrid Classifier.

Algorithm: RandomForest
Precondition: "*A training set* $\coloneqq (p_1,q_1),\ldots\ldots(p_n,q_n)$, features F, and number of trees in forest B".
function RandomForest(S,F)

1. $M \leftarrow \emptyset$
2. **for** *k* ? 1,2….,B do
3. $S(k) \leftarrow A$ **bootstrap sample** from S
4. $m_k \leftarrow RandomizedTreeLearn(S(k),F)$
5. $M \leftarrow M$? $\{m_k\}$
6. end **for**
7. **return** *M*
8. end **function**

function RandomizedTreeLearn(S,F)

1. At each node:
2. $e \leftarrow$ very small subset of *E*
3. **Split** on best feature in *e*
4. **return** ***thelearnedtree***
5. end **function**

Figure 3: RandomForest Algorithm

Algorithm: Ensemble of GLM & RF
Input: A data set ***U***, a set of classification algorithms ***P***, the number of classifiers ***x***
Output: An ensemble ***C***
function RFSLR(*P*,*x*,*U*)

1. **for** *i* **=** 1 to *x***.**
2. Apply **bootstrap** sampling to sample ***U*** and generate ***T****i*, which is the same size of ***U.***
3. **Select** the (⌊*i* modulo|*P*|⌋ + 1)th element in *U* as A_i
4. **Train** C_i by applying A_i on T_i
5. end **for**
6. **return** *C*= **Union** of C_i *where i = 1 to n.*
7. end **function**

Figure 4: An ensemble of GLM & RF

4.5. Validation Method of Classifier

The Hybrid Classifier ***C***, is validated to test its effectiveness using the Train Split approach. In this approach the available dataset is randomly divided into two classes namely Training data and Test data in a 70:30 ratio. The Hybrid Classifier is trained on this 70% subset and tested with the 30% subset. Table 2 represents the Confusion Matrix where, a True Positive implies the individual is rightly classified with Heart Disease, a True Negative means the individual is rightly analyzed as healthy, a False Positive means a fit individual is wrongly analyzed as an individual with Heart Disease and finally False Negative refers to an individual with Heart Disease classified as healthy person. Performance metrics like Classification accuracy, classification error, sensitivity, specificity and precision are extracted for each classifier by using the confusion matrix.

Table 2: Confusion Matrix

	Predicted HD individual(1)	Predicted healthy individual(0)
Correct HD individual(1)	True Positive	False Negative
Incorrect HD individual(0)	False Positive	True Negative

4.6. Performance Metrics

Classification accuracy, Classification error, Sensitivity, Specificity and precision are the performance indictors considered by this work. Accuracy refers to the correctness of a model. Classification error is the measure of wrong classification performed by the model. In this paper, Sensitivity is calculated by dividing the newly analyzed heart patients to the total number of heart patients. Specificity here refers to a case where the test result is negative and the person is healthy. Finally precision refers to the percentage of the obtained results that are relevant.

4.7. Hardware and Software Tools

A machine with basic hardware is used for obtaining the results. The algorithms and techniques discussed in this paper were implemented using packages like sci-kit learn, pandas, numpy, matplotlib of Python Programming Language. For visualization of the results this paper Microsoft Charts were used.

5. EXPERIMENTAL RESULTS AND DISCUSSION

In this section of the paper, a detailed analysis of the results obtained is presented. To start off, the existing models LR, SVM, RF, NB, DT are applied on the CHDD and the results are recorded for analysis. The second step is to systematically apply the steps highlighted in section III of the paper. The 13 features considered where given as input for the FSAs. Relief-F selected CPT, SCH, PES, VCA, SEX and THA as the top 6 features from available 13 features. Table 3 represents the ranking given by the three FSAs. The score obtained by each feature from the feature selector is also listed. LASSO algorithm selected SEX, VCA, EIA, CPT, PES and THA. Similarly, FOCUS selected CPT, PES, THA, SCH, SEX and VCA as its top 6 features out of the 13 features.

The proposed RF with Simple LR Classifier is applied on the Subsets of the Cleveland Heart Disease Data set generated by FOCUS FSAs. The next step is to compare and analyze the existing classifiers with the proposed RFSLR classifier based on the performance metrics.

Table 4 represents the performance of the classifiers namely, LR, SVM, RF, NB, DTs and proposed RFSLR. In terms of accuracy RFSLR (proposed) achieves 91% followed by SVM, RF, DTs, LR and NB with a rate of 86.1%, 85.1%,85%,82.9% and 75.8% respectively.

Table 3: Top features selected by the FSAs Relief-f, LASSO and FOCUS

	RANK											
FSA	**1**		**2**		**3**		**4**		**5**		**6**	
	F	SCORE	F	SCORE	F	SCORE	F	SCORE	F	SCORE	F	SCORE
RELIEF F	CPT	0.59	SCH	0.57	PES	0.574	VCA	0.542	SEX	0.523	THA	0.486
LASSO	SEX	0.15	VCA	0.14	EIA	0.13	CPT	0.1	PES	0.08	THA	0.08
FOCUS	CPT	0.83	PES	0.82	THA	0.79	SCH	0.75	SEX	0.75	VCA	0.56

Table 4: Results of the Performance Metrics for the classifiers implemented

	LR	**SVM**	**RF**	**NB**	**DTs**	**RFSLR proposed**
Classification Accuracy	82.9	86.1	85.1	75.8	85	91
Classification Error	24.2	13.9	13.9	24.2	15.0	8.025
Sensitivity	79.8	100	98.8	79.8	98.8	90
Specificity	60.0	0	10.0	60.0	0.0	87
Precision	90.5	86.1	87.1	90.5	86	93.75

Figure 5 visualizes Classification Accuracy using Scattered Graph. RFSLR achieves the least error rate of 8.025% followed by SVM(13.9%), RF(13.9%), DTs(15.0%), LR(24.2%) and NB (24.2%).

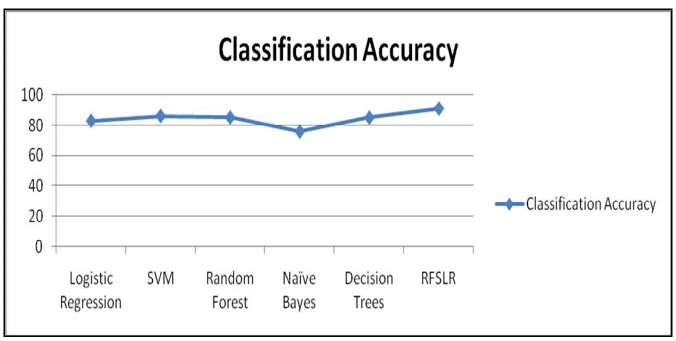

Figure 5: Visualizing Classification Accuracy using Scattered Graph

Figure 6 visualizes Classification error using Scattered Graph. In terms of Sensitivity the order of algorithms will be SVM, RF, DTs, RFLSR, LR and NB.

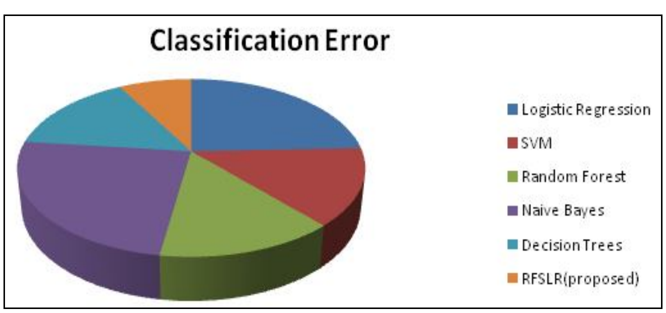

Figure 6: Visualizing Classification Error using Pie Graph

Figure 7 visualizes Sensitivity using Bar Graph. RFSLR achieves a Specificity of 87%.

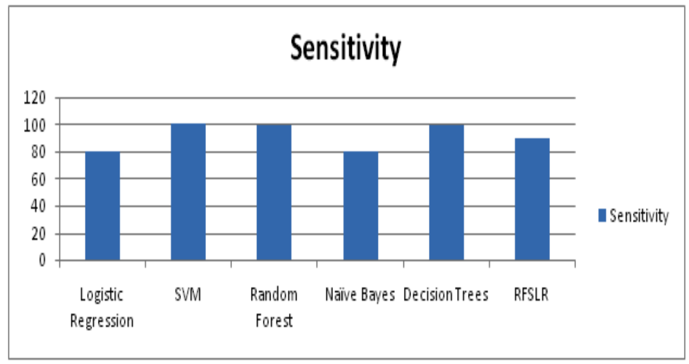

Figure 7: Visualizing Sensitivity using Bar Graph

Figure 8 visualizes Specificity achieved by the classifiers using cone bar graphs. In terms of precision RFSLR achieves a rate of 93.75% followed by NB, Logistic Regression, RF, SVM and DTs.

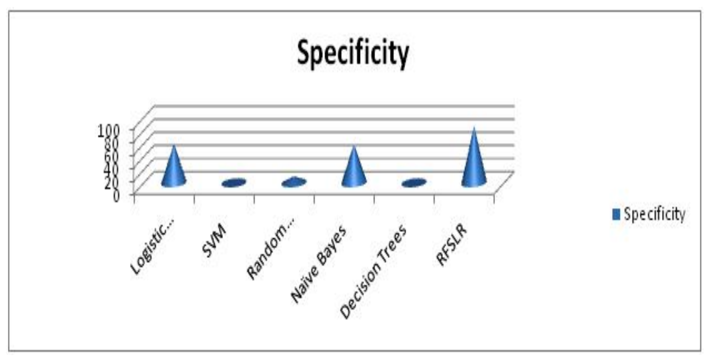

Figure 8: Visualizing Specificity using Cone-bar Graph.

Figure 9 visualizes precision using a 3D graph. Finally all these performance metrics for the classifiers implemented by this work is visualized using a Radar graph in Figure 10. Figure 11. represents the frequency of heart disease in male & female patients. A value 0 represents no disease whereas 1 represents diagnosed with heart disease.

Figure 12. represents the Frequency of Heart Disease in patients with attribute “Age”. It suggests that around 11 patients with age 54 suffered with heart disease, next around 9 patients with ages 9, 51 and 52 suffer with heart related anomalies.

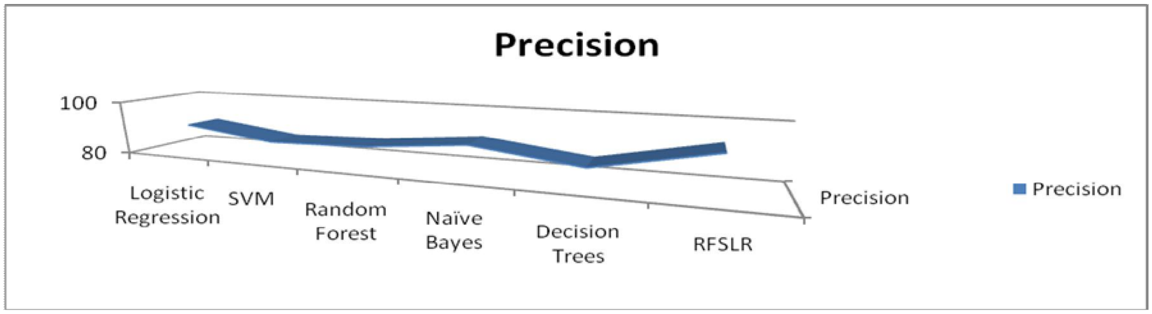

Figure 9: Visualization of Precision using 3D Graph.

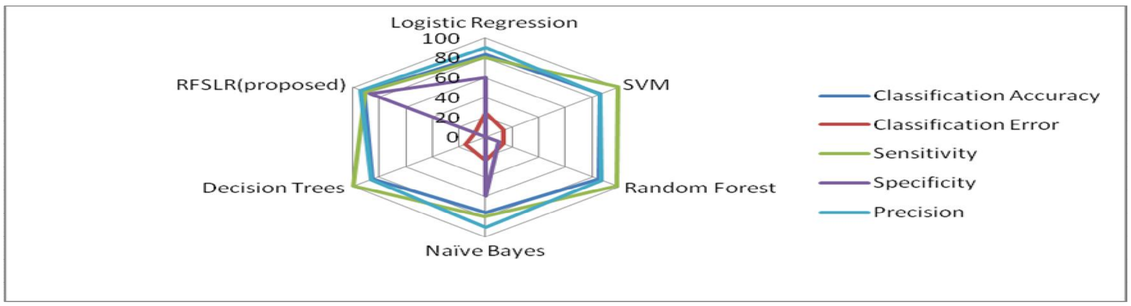

Figure 10: Visualizing Performance metrics for the classifiers using Radar Graph

.

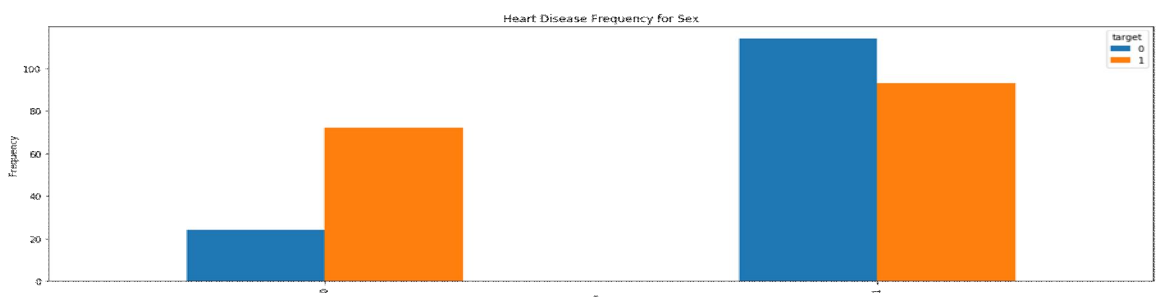

Figure 11: Frequency of Heart Disease in patients for attribute "Sex(Male/Female)"

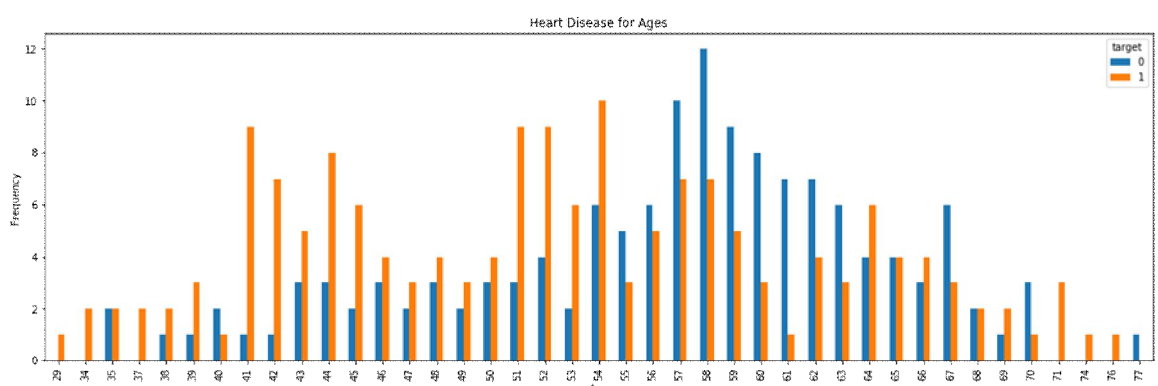

Figure 12: Frequency of Heart Disease in patients with attributes Age

6. CONCLUSION AND FUTURE WORK

Early detection of abnormalities related to heart can save human lives in long term. ML Models analyze the patient's heart health data and predict potential heart diseases. This paper investigated the CHDD using FSAs Relief-f, LASSO and FOCUS algorithms. The proposed ensemble classifier RF Simple LR has performed satisfactorily along with the other classifiers SVM, RF, DT, LR and NB. The proposed model used Data preprocessing techniques, FOCUS FSAs, RF Simple LR ensemble classifier and Train Split validation approach. Finally the classifiers were compared using standard performance metrics. In future this research can be performed with ensembles of diverse ML techniques to enhance prediction. Moreover, novel feature extraction techniques can be investigated to extract robust features.

REFERENCES

1. Mr. Chala Beyene, Prof. Pooja Kamat **Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques**, *International Journal of Pure and Applied Mathematics*, Volume 118, pp. 165-174, Nov 2018.
2. Amin Ul Haq, Jian Ping Li ,Muhammad Hammad Memon ,Shah Nazir , and Ruinan Sun **A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using ML Algorithms**, *Mobile Information Systems*- Hindawi, Volume 2018.
3. Parth Bhavsar,Ilya Safro,Nidhal Bouaynaya,Robi Polikar and DimahDera, **"Data Analytics for Intelligent Transportation Systems"** Chapter 12 in *ML in Transportation Data Analytics* 2017, Pages 283-307.
4. M. J. Liberatore and R. L. Nydick, **``The analytic hierarchy process in medical and health care decision making:Aliterature review**, *Eur. J. Oper. Res.*, vol. 189, no. 1, pp. 194_207, 2008.
5. Chala Beyene and Pooja Kamat, **"Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques"**, *International Journal of Pure and Applied Mathematics* Volume 118 No. 8 2018, 165-174
6. P. S. Kumar, D. Anand, V. U. Kumar, D. Bhattacharyya, and T.-H. Kim**, A computational intelligence method for effective diagnosis of heart disease using genetic algorithm,** *Int. J. Bio-Sci. Bio-Technol.*, vol. 8, no. 2, pp. 363-372, 2016.
7. N. Al-milli, **Backpropogation neural network for prediction of heart disease,** *J. Theor. Appl.Inf. Technol.*, vol. 56, no. 1, pp. 131_135, 2013.
8. Marjia Sultana, Afrin Haider, **"Heart Disease Prediction using WEKA tool and 10-Fold cross-validation"**, *The Institute of Electrical and Electronics Engineers*, March 2017
9. B. S. S. Rathnayakc and G. U. Ganegoda, **``Heart diseases prediction with data mining and neural network techniques,''** in *Proc. 3rd Int. Conf. Converg. Technol. (I2CT)*, Apr. 2018, pp. 1_6.
10. P. K. Anooj, **``Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules,''** *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 24, no. 1, pp. 27_40, Jan. 2012
11. V. Krishnaiah, G. Narsimha, and N. Subhash, **``Heart disease prediction system using data mining techniques and intelligent fuzzy approach: A review,''** *Int. J. Comput. Appl.*, vol. 136, no. 2, pp. 43_51, 2016.
12. L. Baccour, **``Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets,''** *Expert Syst. Appl.*, vol. 99, pp. 115_125, Jun. 2018.
13. R. Das, I. Turkoglu, and A. Sengur**, ``Effective diagnosis of heart disease through neural networks ensembles,''** *Expert Syst. Appl.*, vol. 36, no. 4,pp. 7675_7680, May 2009.
14. C. Cheng and H. Chiu, **"An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database,"** 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, 2017, pp. 2566-2569.
15. D. K. Ravish, K. J. Shanthi, N. R. Shenoy, and S. Nisargh, **``Heart function monitoring, prediction and prevention of heart attacks: Using artificial neural networks,''** in *Proc. Int. Conf. Contemp. Comput. Inform. (IC3I)*, Nov. 2014, pp. 1_6.
16. W. Zhang and J. Han, **``Towards heart sound classi_cation without segmentation using convolutional neural network,''** in *Proc. Comput. Cardiol. (CinC)*, vol. 44, Sep. 2017, pp. 1_4.
17. M. J. Liberatore and R. L. Nydick, **``The analytic hierarchy process in medical and health care decision making:Aliterature review,''** *Eur. J. Oper. Res.*, vol. 189, no. 1, pp. 194_207, 2008.
18. M. Durairaj and V. Revathi, **``Prediction of heart disease using back propagationMLPalgorithm,''** *Int. J. Sci. Technol. Res.*, vol. 4, no. 8, pp. 235_239,2015.
19. A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, **``Prediction of heart disease using ML,''** in *Proc. 2nd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, Mar. 2018, pp. 1275_1278.
20. H. A. Esfahani and M. Ghazanfari, **``Cardiovascular disease detection using a new ensemble classifier,''** in *Proc. IEEE 4th Int. Conf. Knowl.-Based Eng. Innov. (KBEI)*, Dec. 2017, pp. 1011_1014.
21. M. Gandhi and S. N. Singh, **``Predictions in heart disease using techniques of data mining,''** in *Proc. Int. Conf. Futuristic Trends Comput. Anal. Knowl. Manage. (ABLAZE)*, Feb. 2015, pp. 520_525.

22. Krishnaiah, G. Narsimha, and N. Subhash, **``Heart disease prediction system using data mining techniques and intelligent fuzzy approach: A review,''** *International Journal of Computer applications*., vol. 136, no. 2, pp. 43_51, 2016.
23. P. K. Anooj, **``Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules,''** *J. King Saud Univ.-Computer. Inf. Science.*, vol. 24, no. 1, pp. 27_40, Jan. 2012.
24. J. P. Kelwade and S. S. Salankar, **``Radial basis function neural network for prediction of cardiac arrhythmias based on heart rate time series,''** in *Proc. IEEE 1st Int. Conf. Control, Meas. Instrum. (CMI),* Jan. 2016, pp. 454_458.
25. A. S. Abdullah and R. R. Rajalaxmi**, ``A data mining model for predicting the coronary heart disease using RF classifer,''** in *Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls*, Apr. 2012, pp. 22_25.
26. A. Davari Dolatabadi, S. E. Z. Khadem, and B. M. Asl, **"Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM,"** *Comput.Methods Programs Biomed.*, vol. 138, pp. 117–126,2017.
27. A. K. Dwivedi, **"Performance evaluation of different ML techniques for prediction of heart disease,"** *Neural Comput. Appl.*, pp. 1–9, 2016.
28. R. Detrano, A. Janosi, and W. Steinbrunn, **"International application of a new probability algorithm for the diagnosis of coronary artery disease,"** *American Journal of Cardiology,* vol. 64, no. 5, pp. 304–310, 1989.
29. K. Vanisree and J. Singaraju, **"Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks,"** *International Journal of Computer Applications*, vol. 19, no. 6, pp. 6–12, 2011.
30. KevinBuchan, MicheleFilannino and ÖzlemUzunerb, **"Automatic prediction of coronary artery disease from clinical narratives",** *Journal of Biomedical Informatics* Volume 72, August 2017, Pages 23-32.
31. Akhilesh, Kumar & Tomar, Divya & Agarwal, Sonali. (2013). **Clustering of Lung Cancer data using Foggy K-means.** *10.1109/ICRTIT.2013.6844173*.
32. B.Fida, M. Nazir, N.Naveed, and S.Akram, **"Heart disease classification ensemble optimization using genetic algorithm,**"*IEEE 14th International Multitopic Conference*, pp. 19–24.2011
33. Nematzadeh, Zahra. (2015). Comparative Studies on Breast Cancer Classifications with K-Fold Cross Validations Using ML Techniques.
34. M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, **"Classification of heart disease using artificial neural network and feature subset selection,"** *Global Journal of Computer Science and Technology Neural & Artificial Intelligence,* vol. 13, no. 11,2013.
35. O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, **"An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction,"** *Expert Systems with Applications,* vol. 68, pp. 163–172, 2017.
36. K. Vanisree and J. Singaraju, **"Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks,"** *International Journal of Computer Applications*, vol. 19, no. 6, pp. 6–12, 2011.
37. S. Nazir, S. Shahzad, S. Mahfooz, and M. Nazir, **"Fuzzy logic based decision support system for component security evaluation,"** *International Arab Journal of Information Technology,* vol. 15, pp. 1–9, 2015.
38. S. Nazir, S. Shahzad, and L. Septem Riza, **"Birthmark-based software classification using rough sets,"** *Arabian Journal for Science and Engineering,* vol. 42, no. 2, pp. 859–871, 2017.
39. A. Methaila, P. Kansal, H. Arya, and P. Kumar, **"Early heart disease prediction using data mining techniques,"** in *Proceedings of Computer Science & Information Technology (CCSIT-2014),*vol. 24, pp. 53–59, Sydney, NSW, Australia, 2014.
40. Dua, D. and Graff, C. (2019). UCI ML Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.
41. R. Alizadehsani et al., **A data mining approach for diagnosis of coronary artery disease,** *Computer Methods and Programs in Biomedicine,* vol.111, no.1, pp.52-61, Jul. 2013.
42. Ayatollahi H, Gholamhosseini L, Salehi M. Predicting coronary artery disease: a comparison between two data mining algorithms. BMC Public Health. 2019;19(1):448. Published 2019 Apr 29. doi:10.1186/s12889-019-6721-5
43. D. Bell and H. Wang. A Formalism for Relevance and its Application in Feature Subset Selection. ML, 41(2):175–195, 2000.