Evaluating Ensemble Prediction of Coronary Heart Disease using Receiver Operating Characteristics

Tahira Mahboob[1], Rida Irfan[2], Bazelah Ghaffar[3]
Department of Software Engineering
Fatima Jinnah Women University, The Mall, Rawalpindi, Pakistan
tahira.mahboob@yahoo.com

Abstract- Heart diseases may perhaps consequence in debility, severe disorder, and meager quality of lifespan. Furthermore, it could also be lethal. Hence inferring heart disease has turn into foremost distress currently. This paper centers on various machine learning practices which assist ascertaining and perceiving innumerable heart diseases. Multifarious machine learning approaches conversed here are Hidden Markov Models, Support Vector Machine, Feature Selection, Computational intelligent classifier, prediction system, data mining techniques and genetic algorithm. Scrutinizing each approach thoroughly allowed us to select most apposite one. This ultimately permits us to propose an Ensemble Model exploiting pertinent machine learning procedures which perfectly categorizes diverse heart diseases. The evaluation of the proposed technique has been conducted using state of the art technology. The proposed technique has an accuracy of 94.21%, a ROC (Receiver Operating Characteristics) of 0.981, RMSE (Root Mean Square Error) of .2568, Precision of 0.953; showing significant improvement when compared to the performance of K-Nearest Neighbor, Artificial Neural Networks and Support Vector Machines algorithms. Analysis/Evaluation of the implemented algorithms and the proposed Ensemble Model has been done expending the Receiver Operator Characteristics.

Keywords- ANN (Artificial Neural Networks), Ensemble Model KNN (K-Nearest Neighbor), SVM (Support Vector Machines), ROC.

I. INTRODUCTION

In developed and under developing countries, prominent origin of death is heart disease. Person's health is significantly influenced by the heart disease suffered. Cardiovascular disease (CVD) is endured by 80 000 000 inhabitant, alone in united states. Each day approximately 2400 Americans die because of this disease. One very common form of CVD is Anomalous heart rhythm termed as cardiac arrhythmia. The correct functionality of the heart is significantly influenced by Cardiac Autonomic Neuropathy (CAN). Deposits of fatty acids in coronary artery may constrict it down and result in coronary heart disease, which grounds for an occurrence of 1.2 million heart attacks each year.

Providing eminent services is the major concern faced by the health care administrations currently. For instance, it requires early diagnosis of heart disease efficiently and effectively. Hence in order to accomplish this task we are executing various heart disease prediction mechanisms followed by proposing Ensemble models. As the irregular heartbeats are easily perceived by electrocardiogram, therefore ECG seems to be quiet helpful for physicians particularly for the bulky volumes of statistics. This Research paper is systemized in following manner. Section-I is the introduction. Section-II summarizes all the research papers reviewed. Section-III converses the implementation of Ensemble Model along with data set and result analysis, followed by concluding the research paper in fourth section.

II. LITERATURE REVIEW

Expending Artificial Neural Network for prediction of heart disease is major focus of Wijaya et. al[1]. Moreover, Support Vector is also being considered for the prediction process. Predicting heart syndrome is possible within a year by overviewing irregular heart rate. Utilizing various tools such as smart mirror, smart chair, smart mouse and smart phone, data regarding individuals is collected in a server. This is how fatality rate along with number of patients suffering from heart disease decrements significantly. However, in Year 2011 observed accuracy of ANN was 80.06% while SVM observed accuracy was 84.12%. Chen et. al. in 2011[2] present Diagnosis of heart syndrome depends upon the medical data. Heart syndrome prediction system is developed that can help the medical professionals in predicting the heart syndrome by analyzing the medical data of patients. The system takes thirteen medical attributes as input. Then system uses the ANN technique for categorizing the heart syndrome on the basis of these medical attributes. Moreover, ROC is displayed that depicts the performance of

978-1-5090-4815-1/17/$31.00 ©2017 IEEE

system. Outcomes show that the system's prediction accuracy is about 80%, sensitivity is 85%, and specificity about 70%. Sonawane and Patil in 2014[3] detail that in medicine the diagnosis of heart syndrome is extremely problematic job. Hence prediction system using multilayer perceptron neural network is proposed particularly for heart diseases. System takes thirteen medical attributes as data. Back- propagation algorithm is used and the system ultimately predicts whether the heart disease is suffered by patient or not. Moreover, the maximum precision is 98.58% for twenty neurons in unseen layer.Practices of data mining are generally used for prediction of several syndromes as discussed by Amin et. al. [4]. A method is based on neural network and genetic algorithm is proposed for the prediction of heart syndrome by exploiting main detrimental features. The proposed technique predicts the threat of heart syndrome by precision of 89%. Therefore by using this technique a smart system able to predict the syndrome could be constructed easily. In this paper a model for multi-level risk assessment of heart failure is proposed by Aljaaf et. al[5]. By using this model five risk levels of heart failure can be predicted. This model uses the C4.5 decision tree classifier for prediction. In order to improve the prediction accuracy three risk factors are also added to the data set. Efficacy of proposed model is evaluated on the basis of cross validation employing 10 folds. By using this model we can achieve 86.5% sensitivity, 95.5% specificity and 86.53% accuracy. Heart disease prediction and its cure is a job that needs abundant knowledge and experience. Nearly 60 percent of the global inhabitants are targeted by heart syndrome. Goal of this paper is to perceive the reasons of circulatory syndromes by using machine learning methods with the support of medical verdict. C4.5 algorithm, Native Bayesian classifier, Decision tree induction are used in this paper authored by Mohan, K. R et. al.[6]. Palaniappan et. al. [7] discuss various machine learning procedures employed for the evolvement of Intelligent Heart Disease Prediction System (IHDPS) prototype which encompasses Neural Network, Decision Trees, and Naive Bayes. IHDPS is web based application. Traditional decision support system's limitations are significantly outperformed by IHDPS because it responds to "what if" multifaceted queries. It is observed that, Naïve Bayes is the most efficient algorithm compared to other ones, as it produces greatest number of accurate predictions i.e. 86.12% while Neural Network has 85.68% accurate predictions, followed by Decision tree with 80.4%. Cardiac Autonomic Neuropathy (CAN) is a disease which intervenes with regular functionality of heart and hence results in associated disorders as detailed by Cornforth et. al.[8]. For diagnosing early phases of CAN, diverse measures used are portrayed. However, this research paper rely on a case study considering variation of actions to perceive the initial phases of syndrome that disturbs the accurate procedure of the heart and ultimately results in related comorbidities. Machine learning procedures are applied in order to distinguish healthy individuals from the ones with early stage of CAN (Cardiac Autonomic Neuropathy) (using age matching control). For this Genetic Algorithm was used in order to identify the subset of those measures which provided absolute parting amongst two classes. Using this approach 70 percent accuracy was attained. Ordonez et. al[9] state Association rule is a method used for heart diagnosis. But when these methods are practiced using medicinal data set, it yielded immense number of rules which are medically unrelated. Moreover, association rules are mined by exploiting complete data set and do not requisite authentication of distinct sample. In order to overcome problems encountered, a procedure is proposed that practices the four search constraints to decrement the number of rules. Additionally, search for association rules expending trained set and ultimately authenticating them exploiting test set. These search constraints yield a set of rules with great predictive accuracy. Support vector machine (SVM) mechanism is exercised in this paper for interpreting coronary heart disease based on statistical learning technique proposed by Zhang et. al.[10] in 2012. In order to categorize the test data, classifiers are carefully chosen with different kernels, with the help of feature extraction and preprocessing of the original data. Classification results are also compared which depicts greatest accuracy of classifier attained using radial basis function. Grid search technique is implemented for choosing the well suited parameters of penalty factors and kernel functions. This results in highest classification accuracy attained by the classifier. Highest classification accuracy of linear kernel function is 84.1%, polynomial kernel function 81.8%, and RBF kernel function 88.6%.

Using the statistics from ECG, numerous cardiac diseases are categorized efficiently in this research paper by Vanlenzuela [11]. Most related features are picked using these paradigms and are able to attain the classifier capable of discriminating various forms of arrhythmia from ECG signal. It is observed that SVM has the highest precision (i.e. 86.4%) with original matrix (142 features). In this paper by Ho, et al.[12], automatic identification of ECG signal in real time evaluation is executed for constructing clinical support and decision system. Moreover, feature extraction and noise reduction approach for the application of support vector machine (SVM) is implemented with automatic learning algorithms.The results of automatically classifying model are as follows. Sinus classification model observed specificity was 85.9%, for noise

detection model sensitivity observed was 88.4%, while that of disease classification model sensitivity was 89.1%.

In this paper by Bouali et. al. [13] comparative analysis of five classification techniques is presented. The classification techniques are Decision trees, Fuzzy Pattern Tree, Bayesian Network, Artificial Neural Network, and Support Vector Machine. Amongst five machine learning methods being tested, Support Vector Machine gives the better results with greater accuracy. SVM attained greatest percentage of correctly classified instances i.e. 85.76%. In order to evaluate and spot the risk of arrhythmia, the clinical tool which is now extensively used is ECG. Two features are amalgamated in this paper by Oresko et. al.[14], using smartphones for aiding diagnosis. One is Holter monitor's transportability, in which offline evaluation is done by maintaining the track of the signal, other is resting ECG machine with real time processing ability. Explicitly, two platforms developed are wearable CVD recognition based on smartphones. The experiment was performed and the results suggest that prediction accuracy is greater than 90%. For categorizing the electrocardiograms, along with online signal analysis, unique Hidden Markov Model (HMM) is used by Abdreaou et. al. [15]. Multichannel beat assorting and segmentation, demonstrating waveforms into models, and unsupervised adaption to the individual's ECG are the key tasks addressed using this approach. ECG signal scrutiny is performed by the system in two layers. Segmentation of ECG signal is executed at layer 0, in form of the beat waveforms. Then early ventricular contracting beats are detected by the system at layer 1. However, the experiment performed with 59 recordings of test set. The results showed that 99.79% sensitivity achieved for high beat detection and 99.96% of positive prediction.

III. IMPLEMENTATION

Heart diseases are one of the prevailing causes of death in the whole world. Therefore there is dire need for accurate prediction of heart disease. Early symptoms or initial medical data is used for the prediction process. The medical data should be analyzed at regular intervals to figure out chances of developing heart disease and take preventative actions on time. However the prediction system should be as accurate as possible, so that it classifies heart diseases correctly without errors. Hence we intend to propose a model with highest accuracy, precision and with minimum root mean square error.

A. Methodology:

Our Research is focused on three algorithms for heart disease prediction and analysis and the proposed Methodology is depicted in Fig.1. The algorithms used are ANN, KNN and SVM. SVM is usually called black box because the complex transformation and decision boundary that SVM produces are very difficult to interpret. In high dimensional spaces SVM works effectively. If the number of samples is large then KNN will give the good classification but it requires large storage to store the entire dataset. Moreover, more time is consumed to find the k nearest neighbors. To save both time and cost ANN can be used because it takes the samples of data instead of the entire dataset to achieve the desire outcome. But for predictive purposes ANN is rarely used because it mostly over fit the relationship.

B. Proposed Model:

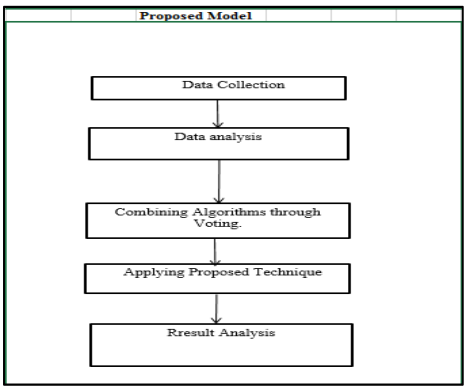


Fig. 1: Proposed Model

We have analyzed each algorithm individually by reviewing various research papers. In year 2011 ANN accuracy observed by Milan Kumari et. al. [7] was about 80.06 % while that of SVM was 84.12%. Heart disease prediction system developed by AH Chen using ANN also had predictive accuracy of 80%. Moreover 85.68% accurate prediction observed by IHDPS developed in (Palaniappan & Awang, 2008) which was based on ANN. However classification accuracy of 88.6% achieved using RBF kernel function of SVM by Zhang et. al. [10] and 86.4% precision acquired using SVM by Valenzuela[15]. In addition to this, disease classification model by Ho Et. al.[11] achieved 89.1% sensitivity. Five classification algorithm tested by [12], out of these SVM attained greatest correctly classified instances i.e. up to 85.76%. However, KNN algorithm used by S. Rajathi [24] for prediction and analysis of Rheumatic Heart Disease had observed accuracy of 68.05% with error rate of 0.675. The KNN algorithm has low accuracy in this scenario but it has good results for large number of samples. Considering these statistics, we have concluded that most efficient amongst all the algorithms studied are SVM, ANN and KNN for heart disease prediction. However we perceive that the results achieved using single algorithm is not as proficient as could be achieved using the

combination of algorithms using voting where two or more sub-models are build, each sub-model, predicting independently. In our scenario voting allows all three algorithms to predict what the output should be and accordingly mean/mode is taken. However by default, the combination Rule parameter is used for amalgamating the predictive results of different sub-models. This parameter calculates the average of the predictive probabilities.

Proposed Ensemble Model: Firstly we assemble the data. 2) Followed by scrutinizing the data set on the basis of number of attributes, instances, type and values of attributes. All attributes in our data set are categorical except age i.e. numeric.3) Subsequently Combine KNN, ANN and SVM using Voting Technique. In voting approach the original data set is used as a basis for training each classifier. And accordingly that class is selected which is supported by greater number of classifiers. 4) Successively apply the proposed technique on Data set.5) lastly Analyze the Results to verify whether the proposed technique gives better results or not.

IV.ANALYSIS AND RESULTS

The ensemble model is presented in the current study. The data set has been acquired from the work by S. U. Amin at. Al [4]. Thirteen attributes are incorporated in total with 50 instances explored. The attributes used to acutely classify multifarious heart diseases are gender, age, cholesterol, heaviness, genetic, smoking, alcohol intake, activity, diabetes, food, obesity, pressure, and Syndrome. Our data set is apportioned in a way that up to 66% is training data and residual 34% is test data. Our data set is presented in Table 1.

Evaluation and Analysis: Receiver Operating Characteristics (ROC) is a popular measure to evaluate and analyze the datasets/algorithms. The analysis is useful for assessing predictive models as presented by (B.Yegnanarayana, 2009)[16] [17]. The text onwards is an in-depth ROC curve analysis for various machine learning algorithms. *K-Nearest Neighbor algorithm*: K- Nearest Neighbor algorithm measures the distance to the nearest K instances, and let them vote. K is typically chosen to be an odd number [19]. The ROC curve obtained employing KNN is depicted in Fig.2. *Artificial Neural Networks:* An artificial neuron network (ANN) is a computational model based on the structure and functions of biological neural networks. Information flows through the network affects the structure of the ANN because a neural network changes - or learns, based on that input and output [20]. ROC curve obtained by exploiting ANN algorithm is depicted in Fig. 3.

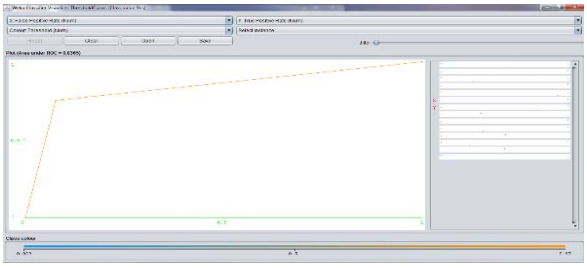


Fig. 2 Receiver Operating Characteristics for K-Nearest Neighbor

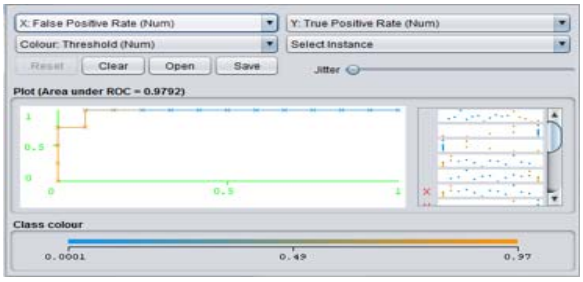
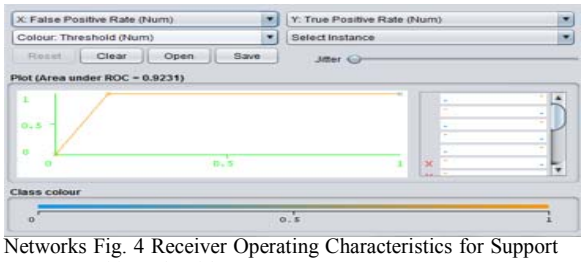


Fig. 3 Receiver Operating Characteristics for Artificial Neural



Networks Fig. 4 Receiver Operating Characteristics for Support Vector Machine Algorithm

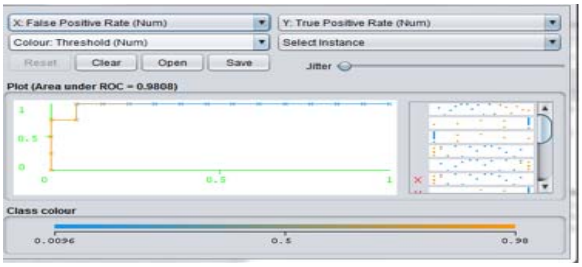


Fig. 5 Receiver Operating Characteristics for Proposed Ensemble Model

Support Vector Machines: A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyper-plane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyper-plane which categorizes new examples [21]. Subsequently applying SVM algorithm following ROC curve presented is presented in Fig.4.

Ensemble Model: Ensemble modeling is the process of running two or more related but different analytical models and then synthesizing the results into a single score or spread in order to improve the accuracy of predictive analytics and data mining applications. [22] Finally the ROC curve acquired by our proposed ensemble model is presented in Fig. 5.The confusion matrix obtained from our proposed Ensemble model technique is represented in Fig.6:

```
=== Confusion Matrix ===

   a   b   <-- classified as
   4   0 |   a = Yes
   1  12 |   b = No
```

Fig. 6: Confusion Matrix for Proposed Ensemble Model

As in confusion matrix all precise categorizations are on the upper left to lowest right diagonal. All off that oblique is improper categorizations of particular kind. So the confusion matrix that is obtained from our technique shows that only one instance out of 17 is incorrectly classified by our model while others are classified correctly. This provides evidence that our model provides highest accuracy. As illustrated by the results achieved, it is apparent that the proposed approach is far more proficient then previously existing approaches. TPR (True Positive Rate), FPR (False Positive Rate), Precision, ROC. In statistics the accuracy of proposed approach is 94.12%, while that of SVM, ANN and KNN is 88.24%, 87.5 and 88.24% respectively. Moreover the root mean square error (RMSE) of our approach is 0.2568, which is also quiet low as compared to SVM, ANN and KNN, with 0.343, 0.270 and 0.3343 respectively. Hence in our proposed ensemble model the true positive rate is immensely high i.e. 0.941, as in comparison to SVM, ANN and KNN with 0.882, 0.875 and 0.882 respectively. And false positive rate of our approach is 0.018 in comparison to SVM, ANN and KNN with 0.036, 0.042 and 0.209 respectively. Furthermore, the greater the ROC curve area the better the classifier is, likewise in our technique the ROC area is 0.981 which is significantly high compared to SVM, ANN and KNN with 0.923, 0.979 and 0.837 respectively. This inculcates that Proposed Ensemble Model is quiet successful to be incorporated in future.

V. CONCLUSION AND FUTURE WORK

An utmost costly and prevalent health problem agonized by today's generation is the multifarious heart diseases. Hence early predicting the disease aids in averting the future sufferings by slight changes in the life style beforehand. In this regard various machine learning procedures conferred and 15 research papers are reviewed. The foremost concern is to identify the most promising approach by examining each technique propagated. Hidden Markov model (HMM) for online beat dissection, SVM for identification of coronary heart disease, computational intelligent classifier to sense cardiac pathologies, CDSS for programmed acknowledgment of the ECG, prediction system employing data mining methods and perceiving premature phases of Cardiac Autonomic Neuropathy (CAN) are some of the machine learning practices disseminated in different research papers [23]. Exploring the above mentioned techniques we are able to design ensemble model assimilating SVM, KNN, and ANN. Preeminent competence is attained consuming these classification procedures, and relative sorting between them is effortlessly performed. Intelligent ensemble model is fabricated because of the very auspicious acquired outcomes. The proposed ensemble model is further aimed to magnify and augment in future by assimilating more machine learning methods and additional medical attributes.

References

[1]. Wijaya, R., & Prihatmanto, A. S. (2013, November). Preliminary design of estimation heart disease by using machine learning ANN within one year. InRural Information & Communication Technology and Electric-Vehicle Technology (rICT & ICeV-T), 2013 Joint International Conference on (pp. 1-4). IEEE.

[2]. Chen, A. H., Huang, S. Y., Hong, P. S., Cheng, C. H., & Lin, E. J. (2011, September). HDPS: Heart disease prediction system. In Computing in Cardiology, 2011 (pp. 557-560). IEEE.

[3]. Sonawane, J. S., & Patil, D. R. (2014, February). Prediction of heart disease using multilayer perceptron neural network. In Information Communication and Embedded Systems (ICICES), 2014 International Conference on (pp. 1-6). IEEE.

[4]. Amin, S. U., Agarwal, K., & Beg, R. (2013, April). Genetic neural network based data mining in prediction of heart disease using risk factors. InInformation & Communication Technologies (ICT), 2013 IEEE Conference on(pp. 1227-1231). IEEE.

[5]. Aljaaf, A. J., Al-Jumeily, D., Hussain, A. J., Dawson, T., Fergus, P., & Al-Jumaily, M. (2015, April). Predicting the likelihood of heart failure with a multi level risk assessment using decision tree. In Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE), 2015 Third International Conference on (pp. 101-106). IEEE.

[6]. Mohan, K. R., Paramasivam, I., & Narayan, S. S. (2014, February). Prediction and Diagnosis of Cardio Vascular Disease--A Critical Survey. InComputing and Communication Technologies (WCCCT), 2014 World Congress on (pp. 246-251). IEEE.

[7]. Palaniappan, S., & Awang, R. (2008, March). Intelligent heart disease prediction system using data mining techniques. In Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on(pp. 108-115). IEEE

[8]. Cornforth, D., Tarvainen, M. P., & Jelinek, H. F. (2014, September). Automated selection of measures of heart rate variability for detection of early cardiac autonomic neuropathy. In Computing in Cardiology Conference (CinC), 2014 (pp. 93-96). IEEE.

[9]. Ordonez, C. (2006). Association rule discovery with the train and test approach for heart disease prediction. IEEE Transactions on Information Technology in Biomedicine, 10(2), 334-343.

[10]. Zhang, Y., Liu, F., Zhao, Z., Li, D., Zhou, X., & Wang, J. (2012, June). Studies on application of Support Vector Machine in diagnose of coronary heart disease. In Electromagnetic Field Problems and Applications

(ICEF), 2012 Sixth International Conference on (pp. 1-4). IEEE.

[11]. Ho, T. W., Lai, H. Y., Wang, Y. J., Chen, W. H., Lai, F., Ho, Y. L., & Hung, C. S. (2014, October). A clinical decision and support system with automatically ECG classification in telehealthcare. In e-Health Networking, Applications and Services (Healthcom), 2014 IEEE 16th International Conference on (pp. 293-297). IEEE.

[12]. Bouali, H., & Akaichi, J. (2014, December). Comparative study of different classification techniques: Heart disease use case. In Machine Learning and Applications (ICMLA), 2014 13th International Conference on (pp. 482-486). IEEE.

[13]. Oresko, J. J., Jin, Z., Cheng, J., Huang, S., Sun, Y., Duschl, H., & Cheng, A. C. (2010). A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing. IEEE Transactions on Information Technology in Biomedicine, 14(3), 734-740

[14]. Andreão, R. V., Dorizzi, B., & Boudy, J. (2006). ECG signal analysis through hidden Markov models. IEEE Transactions on Biomedical engineering, 53(8), 1541-1549.

[15]. Valenzuela, O., Rojas, F., Herrera, L. J., Ortuño, F., Pomares, H., & Rojas, I. (2013, December). Comparison of different computational intelligent classifier to autonomously detect cardiac pathologies diagnosed by ECG. In Intelligent Systems Design and Applications (ISDA), 2013 13th International Conference on (pp. 301-306). IEEE.

[16]. Zou, K. H., O'Malley, A. J., & Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation, 115(5), 654-657.

[17]. Mirmiran, P. A. R. V. I. N., Esmaillzadeh, A., & Azizi, F. (2004). Detection of cardiovascular risk factors by anthropometric measures in Tehranian adults: receiver operating characteristic (ROC) curve analysis. European Journal of Clinical Nutrition, 58(8), 1110-1118.

[18]. Zhang, M. L., & Zhou, Z. H. (2005, July). A k-nearest neighbor based algorithm for multi-label classification. In Granular Computing, 2005 IEEE International Conference on (Vol. 2, pp. 718-721). IEEE.

[19]. Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... & Zhou, Z. H. (2008). Top 10 algorithms in data mining. Knowledge and information systems, 14(1), 1-37.Yegnanarayana, B. (2009). Artificial neural networks. PHI Learning Pvt. Ltd..

[20]. Steinwart, I., & Christmann, A. (2008). Support vector machines. Springer Science & Business Media.

[21]. Caruana, R., Niculescu-Mizil, A., Crew, G., & Ksikes, A. (2004, July). Ensemble selection from libraries of models. In Proceedings of the twenty-first international conference on Machine learning (p. 18). ACM.

[22]. Austin, P. C., Tu, J. V., Ho, J. E., Levy, D., & Lee, D. S. (2013). Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. Journal of clinical epidemiology, 66(4), 398-407.

[23]. Rajathi, S., & Radhamani, G. (2016, March). Prediction and analysis of Rheumatic heart disease using kNN classification with ACO. In Data Mining and Advanced Computing (SAPIENCE), International Conference on (pp. 68-73). IEEE.

Table 1 Dataset for Heart Diseases

Data Set for Heart Disease													
No	Sex	Age	Cholesterc	Pressure	Hereditary	Smoking	Intake	Activity	Diabetes	Diet	Obesity	Stress	Disease
1	Female	35	High	Normal	No	No	Yes	Low	Yes	Poor	Yes	Yes	Yes
2	Male	70	Low	Low	No	No	Yes	High	Yes	Normal	No	No	No
3	Female	60	High	High	No	No	No	Normal	Yes	Poor	Yes	Yes	Yes
4	Female	36	Low	Normal	No	No	No	Normal	No	Good	No	No	No
5	Male	30	Low	Normal	No	No	Yes	High	No	Normal	No	No	No
6	Female	39	Low	Normal	Yes	No	Yes	High	Yes	Normal	No	Yes	No
7	Female	41	High	Normal	No	No	No	Low	No	Poor	Yes	No	No
8	Male	70	High	Normal	No	No	Yes	Low	No	Poor	Yes	No	Yes
9	Male	65	Normal	High	Yes	Yes	Yes	Normal	Yes	Poor	Yes	No	Yes
10	Male	30	Normal	High	No	Yes	No	Normal	No	Good	No	Yes	No
11	Female	31	Low	Normal	No	No	No	High	No	Normal	No	No	No
12	Female	29	Low	Normal	No	No	Yes	High	No	Good	No	No	No
13	Male	30	Low	Normal	No	No	Yes	Normal	No	Normal	No	No	No
14	Female	45	Normal	High	Yes	Yes	No	Normal	Yes	Normal	Yes	Yes	No
15	Male	25	High	Normal	Yes	Yes	Yes	Low	Yes	Normal	No	No	Yes

Table 2 Comparative Performance Analysis of Algorithms with Proposed Technique

	Results											
Classifier	Mean Absolute Error	Root mean Squared Error	Relative Absolute Error	Root Relative Squared Error	True positive rate	False Positive Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Accuracy
KNN	0.1395	0.3343	30.18%	71.71%	0.882	0.209	0.882	0.882	0.882	0.673	0.837	0.843	88.24%
ANN	0.133	0.2701	29.02%	58.20%	0.875	0.042	0.917	0.875	0.882	0.745	0.979	0.983	87.50%
SVM	0.1176	0.343	25.45%	73.58%	0.882	0.036	0.922	0.882	0.889	0.751	0.923	0.894	88.24%
Proposed Technique	0.1277	0.2568	27.62%	55.09%	0.941	0.018	0.953	0.941	0.943	0.859	0.981	0.984	94.12%