Enhanced Accuracy of Heart Disease Prediction using Machine Learning and Recurrent Neural Networks Ensemble Majority Voting Method

Irfan Javid[1], Ahmed Khalaf Zager Alsaedi[2], Rozaida Ghazali[3]

Faculty of Science Computer & Information Technology, Universiti Tun Hussein Onn, Malaysia[1, 3]
Department of Physics, College of Science, University of Misan, Maysan, Iraq[2]
Department of Computer Science & IT, University of Poonch,Rawalakot, AJK, Pakistan[1]

Abstract—**To solve many problems in data science, Machine Learning (ML) techniques implicates artificial intelligence which are commonly used. The major utilization of ML is to predict the conclusion established on the extant data. Using an established dataset machine determine emulate and spread them to an unfamiliar data sets to anticipate the conclusion. A few classification algorithm's accuracy prediction is satisfactory, although other perform limited accuracy. Different ML and Deep Learning (DL) networks established on ANN have been extensively recommended for the disclosure of heart disease in antecedent researches. In this paper, we used *UCI* Heart Disease dataset to test ML techniques along with conventional methods (i.e. random forest, support vector machine, K-nearest neighbor), as well as deep learning models (i.e. long short-term-memory and gated-recurrent unit neural networks). To improve the accuracy of weak algorithms we explore voting based model by combining multiple classifiers. A provisional cogent approach was used to regulate how the ensemble technique can be enforced to improve an accuracy in the heart disease prediction. The strength of the proposed ensemble approach such as voting based model is compelling in improving the prognosis accuracy of anemic classifiers and established adequate achievement in analyze risk of heart disease. A superlative increase of 2.1% accuracy for anemic classifiers was attained with the help of an ensemble voting based model.**

Keywords—Deep learning; machine learning; heart disease; majority voting ensemble; University of California; Irvine (UCI) dataset

I. INTRODUCTION

Heart disease is particular reason of millions of worldwide death per year confer to the World heart federation Report of 2018. Stroke or CVDs are medically familiar as Heart disease (HD) along with blood pressure (BP), artery disease (AD) and debilitated heart disease by cause of diminish, blockade or reinforce capillaries that hamper the required amount of blood circulation to brain, heart, lungs and other body parts. Congestive heart failure is the most trivial kind of heart disease in all other categories of cardiovascular disease. In human body, work of blood vessels is to provide blood to the heart. Alternate, there are some other reasons of heart disease as well alike valves in the heart not supply properly and may be the reason of heart failure. Chest pain, anesthesia, jaw pain, neck ache, throat burn and back agony, cramp in upper abdomen are the most prevailing syndromes of heart disease.

Withal to curtail imperil of heart disease, there are a few predominant aspects such as inhibited blood pressure, under control cholesterol and legitimate exercise. Particularly, heart disease is diagnose after angina, dilated cardiomyopathy, stroke or congestive heart failure. Thus, it is significant to pay attention to CVDs parameter and turn to doctors.

Moreover, confer to the WHO, people expire around 17.9 million every year due to CVDs which coincide to 31% of all deaths globally [1]. This provoke a demand of acquiring an economical arrangement especially capable to provide preamble appraisal of patient established on comparatively elementary medical tests that are economical to everybody. Machine learning (ML) [2] methods have drawn maximum amount of understanding in research society. As illustrate in diverse ongoing studies ML techniques have eventual offering maximum accuracy in classification as associated to alternative procedures for testimony classification. Carry out spectacular accuracy in prediction is crucial as it can edge to pertinent stability. Different machine learning techniques may varies in prediction accurateness. Therefore, it is demanding to perceive gimmick efficient of generating maximum accuracy in heart disease (HD) prediction. Prediction accuracy adept in the take up work is coordinated with earlier research studies. The uttermost practical appraisal formation approach is ML classification for the here and now along with experimental position. Three machine learning (ML) techniques have been practiced consist of random forest (RF), Support Vector Machine (SVM), k-nearest neighbor (KNN). In biomedical field like in diabetes prediction [3] [4], accomplice of diabetes and CVDs [5], reasoning of diabetes proteins [6], machine learning (ML) has already been practiced. There are the divergent conventional approach to use these fettle data to grab the latent material, but the accuracy of the conventional approach is very low, along with prolonged. So, we require contemporary technology which can backing this complex data to be appraised and grab conducive information. Deep learning (DL) algorithms have the ability to learn features from the provided training data, which outrun extracted features used in traditional machine learning algorithms. There are modernity architectures like recurrent neural network (RNN), convolutional neural network (CNN), Long-short-Term memory (LSTM) and gated recurrent unit (GRU). The extant networks confide on disease definitive approach. For classification of cardiac disease in patient modernity

architectures like LSTM and GRU is applied on the extant dataset to evaluate the performance.

The Cleveland dataset from familiar *UCI* database was used to train and testing ML and DL models. It is substantiate dataset and it is extensively used for testing and training in deep learning (DL) and machine learning (ML) models [7]. The dataset consist of 303 patient records and 14 attribute features that are placed on acclaimed aspects and these features are consider to tie with risk of CVDs. We proposed a new hard voting ensemble method in this paper in which various deep learning and machine learning models are mixed and majority vote method is used to predict the result. By using this technique we can improved the overall accuracy in prediction result while aggregation of models produces collective comprehensive model.

The rest of the paper is formulated as follows. Section II, we have reviewed the earlier relevant work to the heart disease prediction and then in Section III we proposed the convoluted particulars of dataset, DL and ML techniques used and data preprocessing. Section IV shows the results produced by each model as well as the accuracy of the prediction proposed by hard voting model. Conclusion and future enhancement is outlined in Section V.

II. REVIEW OF RELEVANT WORKS

Deep learning and machine learning is advantageous for a divergent set of complications. One of the major application of these techniques is to predict the vulnerable variable from the values of autonomous variables. Even in the advanced countries one of the major reason of deaths is CVD [8]. In medical field artificial neural network (ANN) has been popularized to produce maximum accuracy [9]. The research conferred in [10] used the similar heart disease data as this study but divergent ML algorithms were enforced. Four discrete classification techniques were used which comprised Decision Tree, Naïve Bayes, Multi-layer perception and C4.5. Each of these models predict heart disease with maximum accuracy of 85.12% in the MLP classifier. Tree algorithms like J48 and Logistic model were implemented to predict CVDs also used the Cleveland HD dataset [11]. An observation of these approaches was conducted and maximum accuracy 84% was achieved with J48 algorithm.

With web base interface an application named "Intelligent Heart Disease Prediction System" was developed based on three classifier: DT, NB and ANN [12]. Several surveys conducted related to the ML utilization in Healthcare applications, especially in heart disease prognosis. The survey [13] conclude that Bayesian classification and DT surpass the others techniques like k-nearest neighbor, artificial neural network and clustering-based classification. Confer to the new study [14] by Kadi et al. has completed a pragmatic research after hands-on 149 papers proclaimed during the period from 2000-2015 for the prognosis of CVDs, DT, SVM and ANN were established to be the most periodically used ML techniques. An extreme machine learning (EML) were also implemented to predict heart disease (HD) by using UCI datasets repository and achieved highest accuracy of 80% [15]. GA and fuzzy logic (Hybrid genetic Fuzzy) approach

trained and certified over similar UCI repository dataset with maximum accuracy of 86% [16].

According to [7] Raihan et al. developed an android based application to recommend a mock-up for data compilation for IHD. By practicing the P-value strategy and mobile interface they possessed 787 attributes and establish interrelationship amidst symptoms and Ischemic Heart Disease. They established a compelling correlation amidst features with P-value=0.0001 and Ischemic Heart Disease. Likewise, for scoring the symptoms statistical test chi-square, Fisher's exact test and risk score tree are used. BP algorithm is used to extract attributes and syndromes in recent past 2018 [17]-[21].

In RNN section, LSTM consider as the determination with four important factors (forget gate (f_g), input gate (I_g), output gate (O_g) and cell state) have an ample usage for the image analysis along with text and audio signal analysis but is extensively usage in time series analysis, transcribed analysis, voice recognition and health testimony [22]. The major detriment of the RNN model was vanishing gradient problem, LSTM increased the input and output capability of RNN to solve these issues and it uses logical memory to learn sequence vector. To deal with CVDs data temporal features could be learn by Intelligent Healthcare Platform (IHP) established on attention module based LSTM framework [23]. Moreover, to predict CVDs 4. distinct repositories in conjunction with Cleveland dataset is used [24]. Decision Tree (DT) algorithm is the only algorithm comprises of C4.5 and Fast Tree Decision. Formerly, trained technique is established on every attributes of dataset. Later the best sample from datasets are preferred and used to train the model. This approach enhanced the prediction accuracy of the technique from 76.3% to 77.5% adopting C4.5 (average accuracy from datasets) along with enhancement in average accuracy of Fast Tree Decision from 75.48% to 78.06%.

Furthermore, to achieve highest accuracy in the prediction of CVDs distinct methods were used in contemporary research, a few classification algorithms determine CVDs with low accuracy. In contrast with traditional algorithm, hybrid method (include classification algorithms) have produce high accuracy. Our research work proposed a technique to enhance the accuracy of weak classification algorithms by linking them with rest of the classification algorithms. Thus, this technique enhanced the competence of such algorithms along with prediction accuracy for CVDs. The proposed study using ensemble majority voting techniques is done and the results are figure out. The results are compute to illustrate that aforementioned models can have adequate significant usage in medical field.

III. EXPERIMENTAL RESULT ANALYSIS

In this paper, the main objective is to demonstrate CVDs prediction system using prior dataset. The purpose of this research is to use dataset which reflect real life data and grant the prediction system to conclude to any advanced data.

A. Dataset Features Information

For the experiment *UCI* Cleveland heart dataset repository has been used. The most effective 14 attributes were found amongst the 76 based on the comprehensive experiment. The

Cleveland dataset consist of most dominant 14 attributes and 303 samples. Along with 8 absolute features and 6 numeric features. Table I depicts the description of dataset.

In this dataset selected patients had age from 29 to 77. The value 0 is used to depict the female patients and value 1 is used to depict the male patients. There are 3-types of chest pain might be an indicators of heart disease. Typical angina type-1 is because of the blocked heart arteries due to decreased blood discharge to the heart muscles. The basic reason of type 1 angina is mental or emotional stress. And, second type occurs due to numerous reasons but sometime it may not be the reason of actual HD are known as Non-angina chest pain. The next feature is trestbps depicts the readings of resting blood pressure. Cholesterol level is depicted by Chol. Fasting blood sugar level is represented by Fbs. If Fbs is above 120 mg/dl then the value 0 is assigned and value 1 depicts if the Fbs is below the 120 mg/dl. Resting electrocardiography result is represented as Restecg. Maximum heart rate is represented by thalach, exercise cajoled by angina reported as 0 depicts no pain and 1 depicts pain is represented by exang, ST depression is cajoled by exercise is represented as oldpeak, Peak exercise slope ST segment is depicts by slope, number of major vessels colored by fluoroscopy is represented by ca, exercise test duration is represented by thal and the last one target is as class attribute. Class attribute value is used to distinguish the patient with heart disease and patient with no heart disease. Value 1 depicts patients with heart disease and value 0 depicts normal.

A correlation value was determined among every attributes of dataset and the target diagnosis in order to evaluate the data. Oldpeak, Exang, cp and thalach features have the highest correlated value with target feature. Table II depicts the correlated value with target attribute. This is very helpful in making an analysis against the data that is being handle with.

Furthermore, a heat map is also used to show the clear analysis of the correlation among all the attributes in Fig. 1.

Along with, a bar chart depicts in Fig. 2 gender dissemination of samples in *UCI* Cleveland dataset. The male percentage is almost 68.3% and percentage of females is 31.7% in dataset.

Moreover, histograms are devise for discrete features data visualization to depict the marginal features distribution compared for disease and not disease as represented in Fig. 3 to Fig. 8. It is observed that all the discrete features acquire normal distribution. Age vs. Thalach is shown in Fig. 9.

For the age distribution attribute, Fig. 3 represents the people with CVDs and people with no CVDs commonly. It can be viewed that maximum measurements exist between 40-52 years old. It is also realized that if age has a relation to having CVDs, then people in age range from 50-52 and 40-41 had a dominant consolidation with heart diseases.

Furthermore, to depict the possibility of any relation, Fig. 4 represents the maximal correlated discrete feature (thalach) is devise adjacent to age. It is observed that heart rate is commonly higher for the people with heart disease as compared to the people with no heart disease. Moreover, maximal heart rate decreased noted to a -ve correlated value of -0.3 as age increased. It is represented previously in Fig. 1.

TABLE. I. FEATURES DESCRIPTION OF THE CLEVELAND HEART DISEASE DATASET

S#	Features	Features Description	Data Type
1	Age	Patient age (completed in years)	Numeric
2	Sex	Gender of the patient [0= Female, 1= Male]	Binary
3	Cp	Type of chest pain classified in to four values [1- Typical angina Type, 2- Atypical angina Type, 3- Non-angina pain]	Nominal
4	Trestbps	Level of the patient blood pressure at resting mode in mm/Hg	Continuous
5	Chol	Cholesterol serum, mg/dl	Numeric
6	Fbs	Level of blood sugar on fasting (>120 mg/dl): 1 depict in case of true & 0 depict in case of false.	Binary
7	Restecg	At resting result of ECG is depict in three different values: 0 represented Normal state, 1 represented abnormality in ST-T wave, 2 having LV hypertrophy defined	Nominal
8	Thalach	Maximum rate of Heart recorded	Continuous
9	Exang	Angina-induced by exercise (1 represent 'yes' and 0 represent 'no'	Binary
10	Old peak ST	Exercise tempted ST depression comparative to rest state	Continuous
11	Slope	During peak exercise measured the ST segment in terms of slope represent in 3 values: [1. Up-sloping, 2. Flat, 3. Down-sloping]	Continuous
12	Ca	Ranges from 0-3 represent the number of vessels colored by fluoroscopy	Nominal
13	Thal	Status of the heart: [3. Normal, 6. Fixed defect, 7. Reversible defect]	Discrete
14	Target	Diagnosis represent in two categories: [0= Well, 1= possibility HD]	Binary

TABLE. II.　Correlated Values with Target Attribute Analysis

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
age	1.000000	-0.098447	-0.068653	0.279351	0.213678	0.121308	-0.116211	-0.398522	0.096801	0.210013	-0.168814	0.276326	0.068001	-0.225439
sex	-0.098447	1.000000	-0.049353	-0.056769	-0.197912	0.045032	-0.058196	-0.044020	0.141664	0.096093	-0.030711	0.118261	0.210041	-0.280937
Cp	-0.068653	-0.049353	1.000000	0.047608	-0.076904	0.094444	0.044421	0.295762	-0.394280	-0.149230	0.119717	-0.181053	-0.161736	0.433798
Trestbps	0.279351	-0.056769	0.047608	1.000000	0.123174	0.177531	-0.114103	-0.046698	0.067616	0.193216	-0.121475	0.101389	0.062210	-0.144931
chol	0.213678	-0.197912	-0.076904	0.123174	1.000000	0.013294	-0.151040	-0.009940	0.067023	0.053952	-0.004038	0.070511	0.098803	-0.085239
fbs	0.121308	0.045032	0.094444	0.177531	0.013294	1.000000	-0.084189	-0.008567	0.025665	0.005747	-0.059894	0.137979	-0.032019	-0.028046
Restecg	-0.116211	-0.058196	0.044421	-0.114103	-0.151040	-0.084189	1.000000	0.044123	-0.070733	-0.058770	0.093045	-0.072042	-0.011981	0.137230
thalach	-0.398522	-0.044020	0.295762	-0.046698	-0.009940	-0.008567	0.044123	1.000000	-0.378812	-0.344187	0.386784	-0.213177	-0.096439	0.421741
Exang	0.096801	0.141664	-0.394280	0.067616	0.067023	0.025665	-0.070733	-0.378812	1.000000	0.288223	-0.257748	0.115739	0.206754	-0.436757
Oldpeak	0.210013	0.096093	-0.149230	0.193216	0.053952	0.005747	-0.058770	-0.344187	0.288223	1.000000	-0.577537	0.222682	0.210244	-0.430696
Slope	-0.168814	-0.030711	0.119717	-0.121475	-0.004038	-0.059894	0.093045	0.386784	-0.257748	-0.577537	1.000000	-0.080155	-0.104764	0.345877
Ca	0.276326	0.118261	-0.181053	0.101389	0.070511	0.137979	-0.072042	-0.213177	0.115739	0.222682	-0.080155	1.000000	0.151832	-0.391724
thal	0.068001	0.210041	-0.161736	0.062210	0.098803	-0.032019	-0.011981	-0.096439	0.206754	0.210244	-0.104764	0.151832	1.000000	-0.344029
Target	-0.225439	-0.280937	0.433798	-0.144931	-0.085239	-0.028046	0.137230	0.421741	-0.436757	-0.430696	0.345877	-0.391724	-0.344029	1.000000
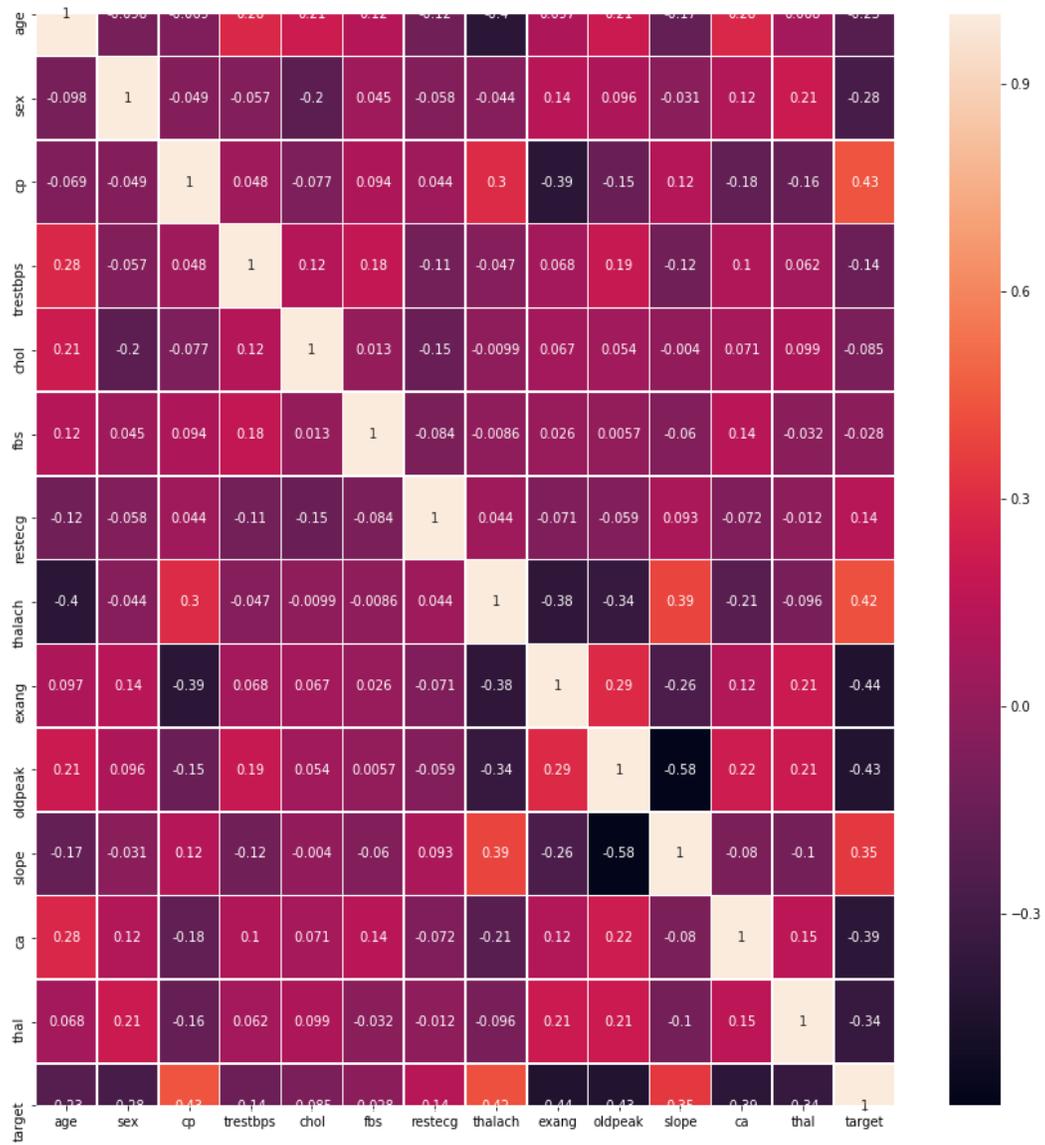
Fig. 1.   Correlation-Cross Values Heat Map.
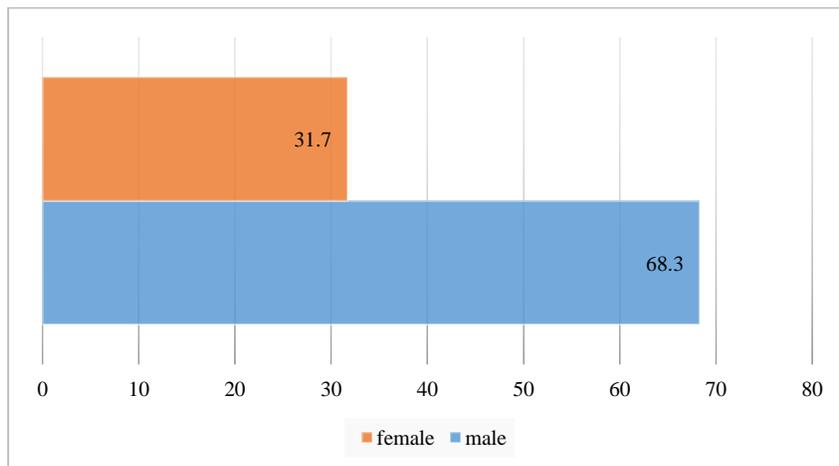
Fig. 2.   Gender Dissemination
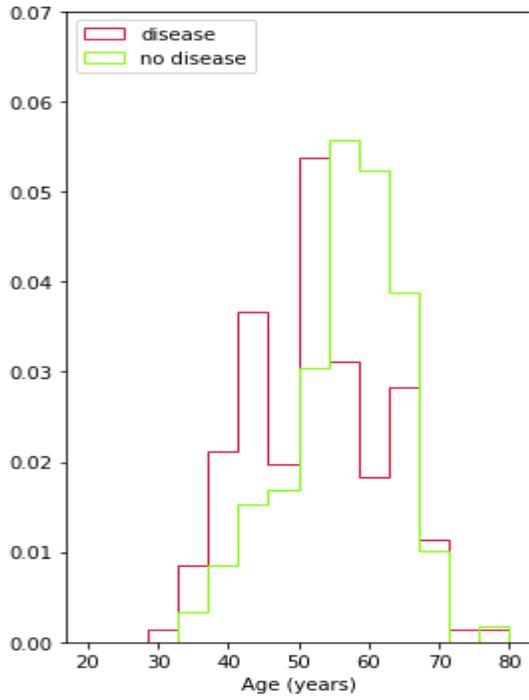
Fig. 3.   Age Distribution.
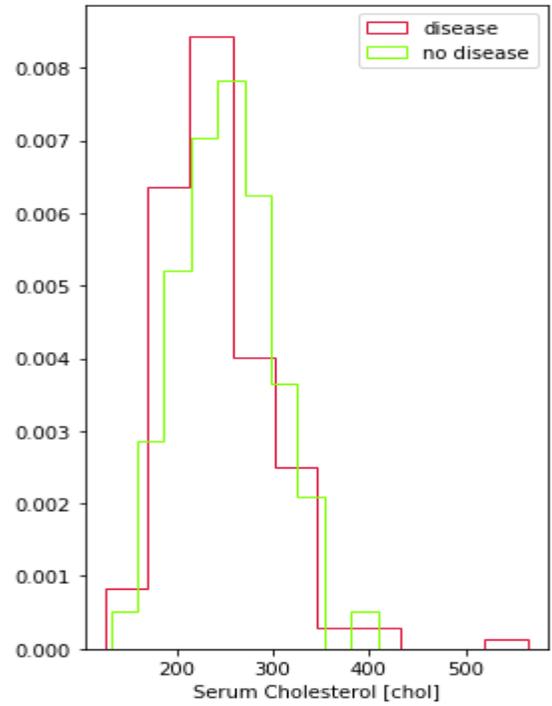
Fig. 5.   Serum Cholesterol Distribution.
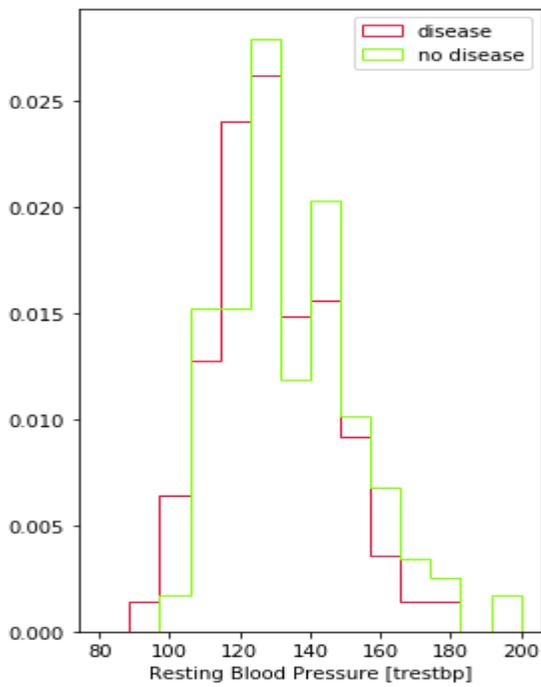
Fig. 4.   Blood Pressure Distribution (at Rest).
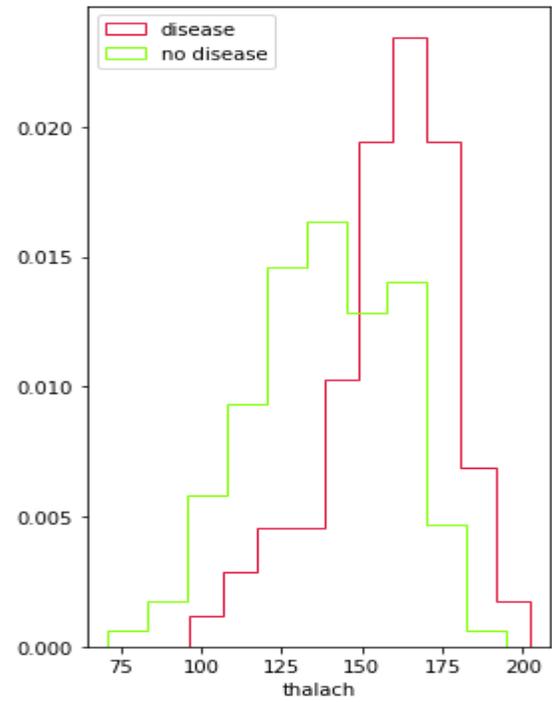
Fig. 6.   Maximal Heart Rate Acquire.
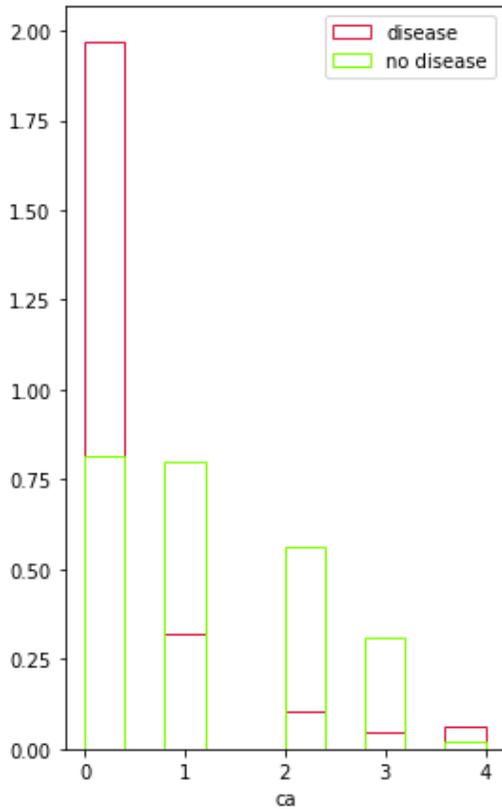
Fig. 7. Calcium Distribution.
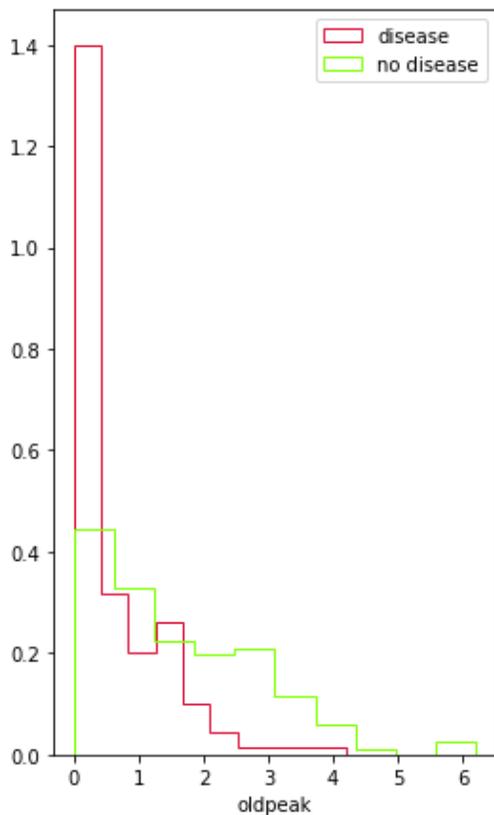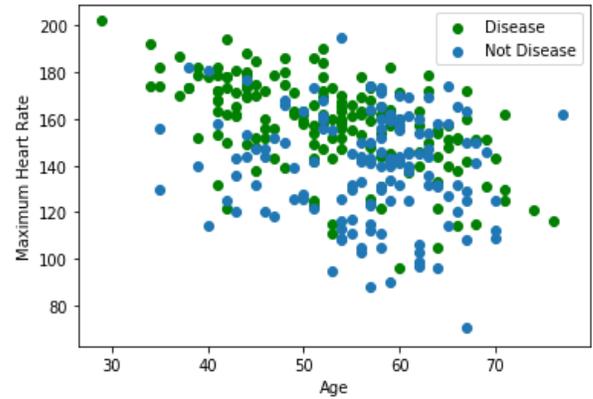
Fig. 9. Age vs. Thalach.

B. Attribute Preprocessing

In order to scale the maximum discrete values by using the Minimum and Maximal normalization approach, the attributes in Cleveland dataset acquire distinct proportions. As shown in eq (1), by using mentioned strategy data is transformed linearly by deducting the smallest and divide over the data range. So, the sample is categorized between zero and one which stimulate learning models to normalize the impact of distinct parameters and form a fair direction between data.

$$Z = \frac{N - \text{minimum}}{\text{maximum} - \text{minimum}} \tag{1}$$

IV. MACHINE LEARNING VS. DEEP LEARNING MODELS

The Cleveland heart disease dataset has been split into a testing set and train set in the scale of 80% of training set and 20% of testing data and training data set is used to train particular models. Test data is used to check the ability of a models. The working of the particular models are described in the later part.

A. Random Forest Classifier

It is also known as tree based classifier algorithm. Basically, name of the classifier is the indication that the algorithm build a woodland surrounded by huge number of trees. In order to get a maximum accuracy and substantial prediction, RF is an ensemble algorithm comprises on constructing numerous trees and integrate them together. This model used random samples from the training set to build set of decision trees. RFC rerun with numerous samples and compose an eventual decision established on majority voting. To handle missing information RFC is very effective but it is prone to over fitting.

B. Support Vector Machine

SVM was first suggested by Vladimir N. V and Alexey Ya in his study related to theory of statistical learning [25, 26]. For classification and regression purposes a supervised learning machine approach known as support vector machine (SVM) is used. In SVM a technique named trick kernel is used to revamp the information and then it identify most appropriate solution based on these alteration. At present, patient with heart disease and patient with no heart disease are classified by SVM on the basis of binary classification for $k_i = +1, -1$ additionally. This approach can be protected for

Fig. 8. Old Peak Distribution.

classification in multiple classes by formulating two-multiclass classifiers [25]. A support vector machine classifier is a best approach to get reprieve hyper-plane which lie between two classes [27]. This reprieve clear hyperactive plane has numerous adequate statistical aspects. Finally, slack fickle is very informative to provide adversities of noisy data.

C. K-NN Classifier

The third classifier that was presented is the K-NN algorithm. The main purpose of this algorithm is to find the distance between the current sample along with all the trained samples, K depicts the predefined figures of adjacent points which are used for voting to the current test data's class. Certainly, classification follow established on the more classes of the K data points elected. On the bases of Grid-Search-CV more accurate results are produced and the predefined number for K in this study was selected to be 7.

D. Long-Short term Memory (LSTM)

LSTM was first proposed by Hochreiter al. is a special kind of Recurrent Neural Network (RNN) [28]. LSTM have two distinct states passed between the neurons – the cell state and the hidden layer. Cell state act as short term memory while hidden layer carry the long-term memory, commonly. There was a vanishing gradient problem with original RNN model. Therefore, RNNs are not suitable for long-term dependency data calculations. The vectors in the LSTM are added to the current node on the support of standard RNN model, which helps to solve the problems of RNN with long-term data calculations. Furthermore, LSTM model has been extensively used. LSTM layers consist of three vectors i.e., a forget vector, an input vector, and an output vector. With the passage of time many researchers proposed trivial changes to the standard LSTM model. One of the most attractive LSTM variant "peephole-connections" was introduced by Gers et al. [29]. There are numerous adaptations with small changes regarding the gated structure in the LSTM units. Here we will consider the one proposed by Graves et al. [32].

$$i_t = \sigma\,(W_{xi}\,x_t + W_{hi}\,h_{t-1} + W_{ci}\,c_{t-1} + b_i\,) \tag{2}$$

$$r_t = \sigma\,(W_{xr}\,x_t + W_{hr}\,h_{t-1} + W_{cr}\,c_{t-1} + b_r\,) \tag{3}$$

$$c_t = r_t\,\iota\,c_{t-1} + i_t\,\iota\,\tanh(W_{xr}\,x_t + W_{xc}\,x_t + W_{hc}\,h_{t-1} + b_c\,) \tag{4}$$

$$o_t = \sigma\,(W_{xo}\,x_t + W_{ho}\,h_{t-1} + W_{co}\,c_{t-1} + b_o\,) \tag{5}$$

$$h_t = o_t\,\iota\,\tanh(c_t) \tag{6}$$

where ι represent element wise product and r, o, i are the forget vector, output vector and input vector respectively. It is observed that the gating structure regulates how the new input and previous hidden state value must be unite to produce the new hidden state value.

The most attractive variant of LSTM is gated-recurrent unit (GRU) was introduced by Chung et al. [30]. The idea was to combine forget vector and input vector as single update vector. In GRU, cell state and hidden state are also merges and make some numerous changes as well. The GRU support the long term sequences and also carry the long-term memories. Therefore, proposed GRU architecture is simpler and most attractive than the original LSTM model.
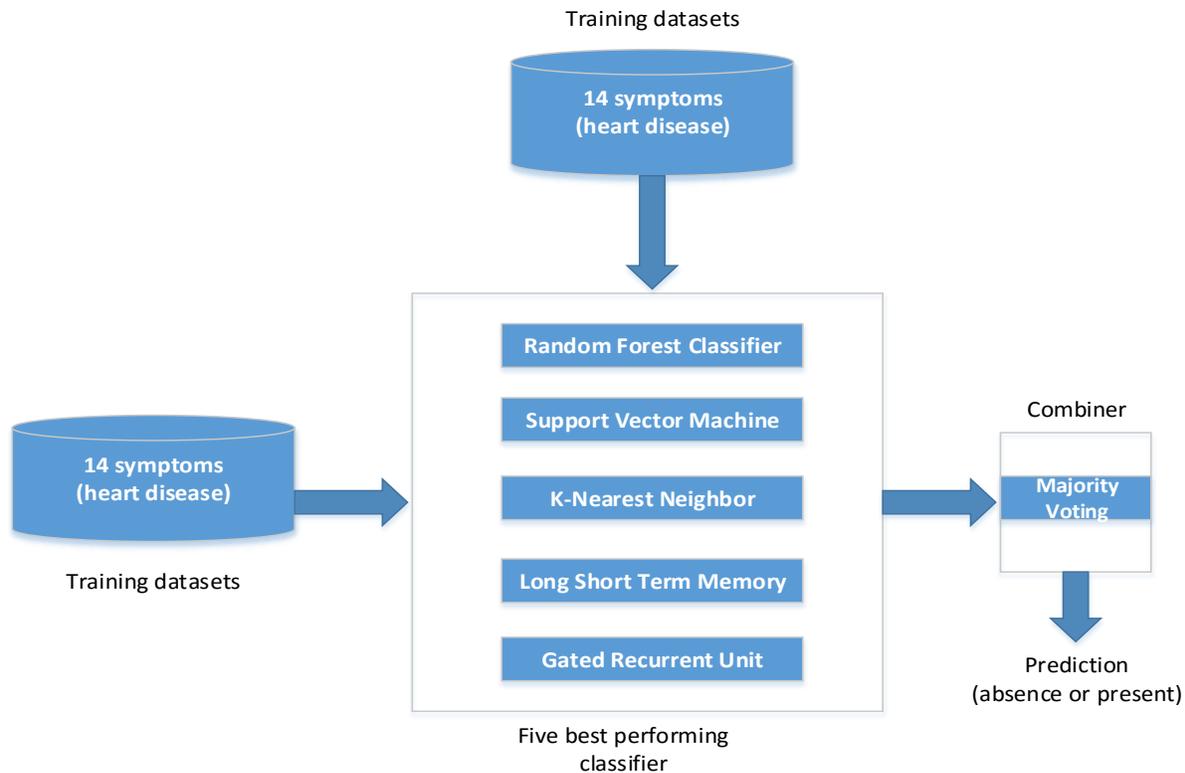
Fig. 10. Workflow of the Proposed Ensemble Vote based Model.

E. Gated Recurrent Unit (GRU)

Cho et al. [30] proposed another gating structure known as GRU (gated recurrent unit) with the purpose to carry long-term dependencies from the calculations within the GRU neuron to produce the hidden state. GRU have only one hidden state conveyed between time steps. Following are the equations determined by Chung et al. [31].

$$r_t = \sigma (W_r x_t + U_r h_{t-1}) \tag{7}$$

$$\tilde{h}_t = \tanh (W x_t + U(r_t \square h_{t-1})) \tag{8}$$

$$z_t = \sigma (W_z x_t + U_z h_{t-1}) \tag{9}$$

$$h_t = (1 - z_t) h_{t-1} + z_t \tilde{h}_t \tag{10}$$

Where r and z are commonly the reset and update gates. It can be observed that, GRU is most simple than LSTM, and performance is far better in different experiments. In [31], Chung et al. provide a comparison related to the performance of original RNN, LSTM and GRU, using numerous datasets. It was observed that gated recurrent unit surpass the other techniques in different situations.

F. Ensemble Classifier

At the end, five models aforementioned are unite in an ensemble method where hard voting (majority vote of the models) technique is used for classification. The voting is based on the prediction of each model about each sample and final prediction is based on the majority votes, one that obtains more than 50% of the votes.

The independent classifiers output is united and plays an important role in the final output prediction of an ensemble system. As shown in the Fig.10. Therefore, one of the interesting research study is combination of classifiers in ensemble system. Majority voting approach is extensively used method for labeling the output [33]. In case of discrete outputs, like linear combination, a maximum, minimum, average or any other alternate like derriere possibilities may be used. Many times a classifier may be used as a meta-classifier for uniting outputs of ensemble-members. Due to better performance of majority voting approach over other linear and meta-classifiers has been applied in this work. Therefore, majority voting rule lies in 3 categories: (1) Unanimous-Voting method, here every models must acknowledge the prediction, (2) simple majority method, here prediction required to be partially higher than 50% of classifiers, and (3) majority voting method, here maximum figures of votes is required for the ensemble-decision. If the output of the individual classifier is independent than the majority voting rule combiner constantly enhance the prediction accuracy [34]. Suppose that a class define outputs of classifier O_i are shown as d-dimensional binary vectors:

$$[O_{i, 1} \ldots O_{i, d}] \in \{0,1\}^d, i=1,\ldots\ldots,N \tag{11}$$

Where $O_{i,1}=1$, if classifier O_i label y in w_j, and 0 differently. The majority voting method would provide an ensemble decision for class w_k, if the below equation is satisfied:

$$\sum_{i=1}^{N} O_{i, k} = \max{}^c j=1 \sum_{i=1}^{N} O_{i, j} \tag{12}$$

If we have 2 classes (c=2), the majority voting method correspond with simple majority approach (50% of vote +1). According to the equation (4) majority voting approach would predict an accurate class define at least [N/2+1] classifiers correctly predict the define class [35]. In our proposed research work, $N = 5$, it observes that our proposed approach would be able to predict correctly if more than half (at least 3 classifiers) predict the define class correctly.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The first classifier Random forest, to study unseen data prediction was run on the test dataset so that the approach has never overcome. Default parameters of the approach are used to run the early test and composed an accuracy of 83.6%. Along with, attributes importance was calculated in this approach and most important attributes were (ca, thalach, oldpeak). Confusion matrix obtained from this approach is shown in Fig. 11.

Also, the second classifier was Support Vector Machine (SVM) algorithm. To run the unseen test dataset the approach was developed with the default parameters. The prediction accuracy of this model was 81.31%. In Fig. 12, confusion matrix obtained from this classifier is depicted.

The third approach, known as K-Nearest Neighbor model. To run the unseen test dataset using default parameters we developed the model. The prediction accuracy get out to be 82.8%. Fig. 13 depicts confusion matrix obtained from this algorithm.

Fig. 11. Random Forest Model Confusion-Matrix.

Fig. 12. SVM Model Confusion-Matrix.

KNN
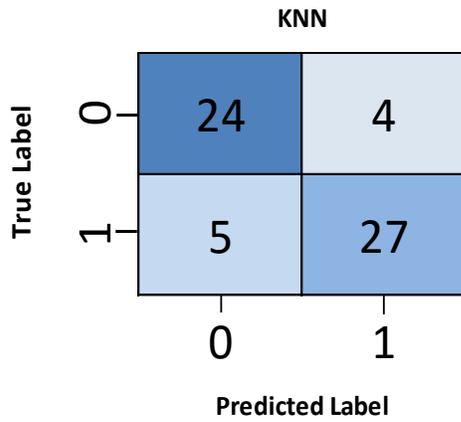
Fig. 13. K-Nearest Neighbor Model Confusion-Matrix.

Additionally, fourth approach that was developed known as LSTM model. Using the default parameters this approach was developed and classification established based on the hidden data test set. The prediction accuracy get out to be 81.31%. In Fig. 14, Confusion matrix obtained from this model is depicted.

Finally, the fifth approach that was developed was the GRU model. Using the default parameters this approach was developed and classification established based on the hidden data test set. The prediction accuracy get out to be 81.46%. Fig. 15 depicts the confusion matrix obtained from this model.

We have noticed that Random Forest and K-NN are constantly provide better prediction accuracy as compared to other classification models. The performance of the each model in accuracy prediction of Heart-Disease as shown in the Fig. 16.

Certainly, the overall prediction accuracy of this study after organizing the Hard Voting ensemble-method get out to

be 85.71% which is treated a fairly required accuracy that can be further developed upon in future.
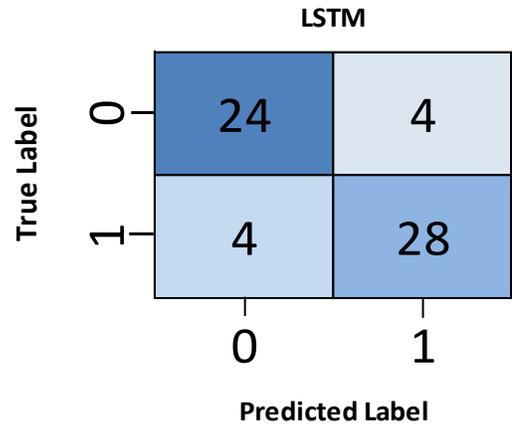
LSTM

Fig. 14. LSTM Model Confusion-Matrix.

GRU
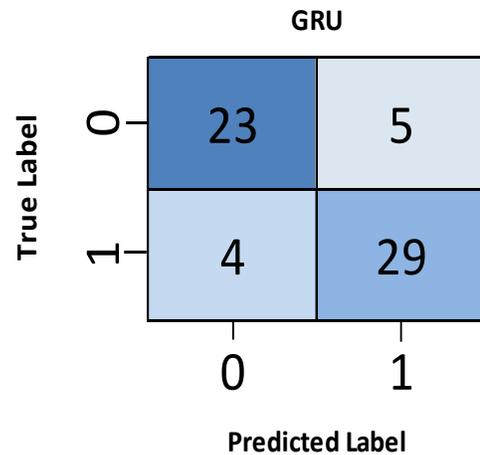
Fig. 15. GRU Model Confusion-Matrix.
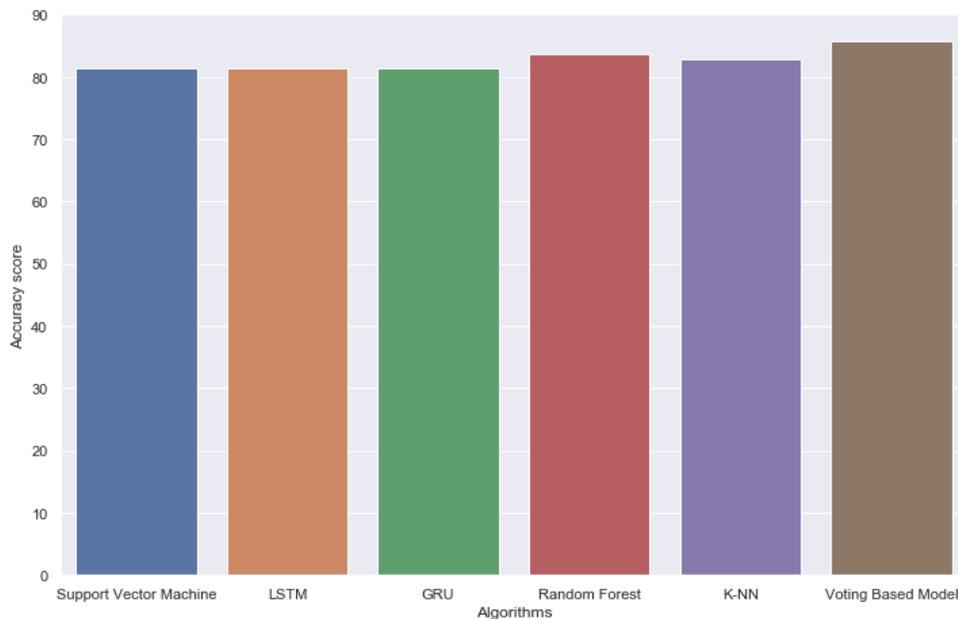
Fig. 16. The Performance Study of different ML and DL Models.

VI. CONCLUSION

To save the life of the human beings, early prediction of heart disease plays significant role. Here, in this paper we presented a ML and DL ensemble models that united multiple ML and DL models in order to give a maximum accuracy and vigorous model for the prediction of any possibility of having heart disease. Table III depicts the prediction accuracy comparison of Machine learning techniques (i.e. RF, SVM and KNN), deep learning models (i.e. LSTM and GRU) and proposed methodology. This Ensemble approach retained 85.71% accuracy, which surpass the prediction accuracy of every particular model. This approach may be very useful to assist the doctors to investigate the patient cases in order to legitimize their prescription. The future work of this study can be performed with different mixtures of ML and DL models to better prediction.

TABLE. III.   PREDICTION ACCURACY COMPARISON OF THE MODELS

Model Name	Accuracy
Random Forest	83.6%
Support Vector Machine	81.3%
K-NN	82.8%
LSTM	81.31%
GRU	81.46%
Hard Voting Ensemble Model	85.71%

REFERENCES

[1] "Cardiovascular diseases (CVDs)," World Health Organization, 26-Sep2018.[Online]. Available: https://www.who.int/cardiovascular_diseases /en/. [Accessed: 27-Apr-2019].

[2] Bache K, Lichman M (2013) UCI machine learning repository [http://archive.ics.uci.edu/ml]. University of California, School of Information and Computer Science. Irvine, CA.

[3] Bansal A, Agarwal R, Sharma R (2015) Determining diabetes using iris recognition system. Int J Diabetes Dev Ctries 35(4):432–438.

[4] Kalaiselvi C, Nasira GM (2015) Classification and prediction ofheart disease from diabetes patients using hybrid particle swarm optimization and library support vector machine algorithm. Int J Comput Algorithm 4(1):2278–2397.

[5] Bhramaramba R, Allam AR, Kumar VV, Sridhar G (2011) Application of data mining techniques on diabetes related proteins. Int J Diabetes Dev Ctries 31(1):22–25.

[6] King RD (1992) Statlog databases. Department of Statistics and Modelling Science, University of Strathclyde, Glasgow.

[7] Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[8] Vanisree K, JyothiSingaraju. Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks. Int J Comput Appl April 2011;19(6). (0975 8887).

[9] L. Baccour, ``Amended fused TOPSIS-VIKOR for classication (ATOVIC) applied to some UCI data sets," Expert Syst. Appl., vol. 99, pp. 115125, Jun. 2018. doi: 10.1016/j.eswa.2018.01.025.

[10] D. Chaki, A. Das, and M. Zaber, "A comparison of three discrete methods for classification of heart disease data," Bangladesh Journal of Scientific and Industrial Research, vol. 50, no. 4, pp. 293–296, 2015.

[11] R. G. Saboji, "A scalable solution for heart disease prediction using classification mining technique," 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), 2017.

[12] S. Palaniappan, R. Awang, in: Intelligent heart disease prediction system using data mining techniques, 2008 IEEE/ACS International Conference on Computer Systems and Applications, 2008, pp. 108–115.

[13] J. Soni, U. Ansari, D. Sharma, S. Soni, Predictive data mining for medical diagnosis: an overview of heart disease prediction, Int. J. Comput. Appl. 17 (8) (2011) 43–48.

[14] I. Kadi, A. Idri, J.L. Fernandez-Aleman, Knowledge discovery in cardiology: a systematic literature review, Int. J. Med. Inform. 97 (2017) 12–32.

[15] S. Ismaeel, A. Miri, D. Chourishi, in: Using the Extreme Learning Machine (ELM) technique for heart disease diagnosis, IEEE Canada International Humanitarian Technology Conference, 2015, pp. 1–3.

[16] M. Raihan, S. Mondal, A. More, and M. Sagor et al., "Smartphonebased ischemic heart disease (heart attack) risk prediction using clinical data and data mining approaches, a prototype design," in International Conference on Computer and Information Technology (ICCIT), pp. 299-303, 2016.

[17] G. Shanmugasundaram, V. M. Selvam, R. Saravanan, and S. Balaji, "An Investigation of Heart Disease Prediction Techniques," in IEEE International Conference on System, Computation, Automation and Networking (ICSCA), pp. 1-6, 2018.

[18] P. Umasankar and V. Thiagarasu, "Decision Support System for Heart Disease Diagnosis Using Interval Vague Set and Fuzzy Association Rule Mining," in International Conference on Devices, Circuits and Systems (ICDCS), pp. 223-227, 2018.

[19] K. G. Dinesh, K. Arumugaraj, K. D. Santhosh, and V. Mareeswari, "Prediction of Cardiovascular Disease Using Machine Learning Algorithms," in International Conference on Current Trends towards Converging Technologies (ICCTCT), pp. 1-7, 2018.

[20] I. K. A. Enriko, M. Suryanegara, and D. Gunawan, "Heart Disease Diagnosis System with k-Nearest Neighbors Method Using Real Clinical Medical Records," in International Conference on Frontiers of Educational Technologies, pp. 127-131, 2018.

[21] H. Kahtan, K. Z. Zamli, W. N. A. W. A. Fatthi, and A. Abdullah et al., "Heart Disease Diagnosis System Using Fuzzy Logic," in International Conference on Software and Computer Applications, pp. 297-301, 2018.

[22] T. Ergen and S. S. Kozat, "Online Training of LSTM Networks in Distributed Systems for Variable Length Data Sequences," IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 10, pp. 5159-5165, 2018.

[23] C. Lin, Y. Zhangy, J. Ivy, and M. Capan et al., "Early Diagnosis and Prediction of Sepsis Shock by Combining Static and Dynamic Information Using Convolutional-LSTM," in IEEE International Conference on Healthcare Informatics (ICHI), pp. 219-228, 2018.

[24] El-Bialy, R., Salamay, M., Karam, O. and Khalifa, M. (2015). Feature Analysis of Coronary Artery Heart Disease Data Sets. Procedia Computer Science, 65, pp.459-468.

[25] Vapnik VN, Vapnik V (1998) Statistical learning theory, vol 2.Wiley, New York.

[26] Vapnik V (2000) The nature of statistical learning theory.Springer, Berlin

[27] Burges CJ (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Disc 2(2):121–167.

[28] S. Hochreiter and J. Schmidhuber, ``Long short-term memory,'' Neural Comput., vol. 9, no. 8, pp. 17351780, 1997.

[29] F. A. Gers and J. Schmidhuber, ``Recurrent nets that time and count,'' in Proc. IEEE-INNS-ENNS Int. Joint Conf. Neural Netw. (IJCNN), vol. 3. Jul. 2000, pp. 189194.

[30] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. (2014). "Empirical evalua-tion of gated recurrent neural networks on sequence modeling.'' [Online]. Available: https://arxiv.org/abs/1412.3555.

[31] Chung J, Gulcehre C, Cho K, Bengio Y. arXiv preprint arXiv:1412.3555.

[32] Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks. In ICASSP2013, pages 66456649. IEEE, 2013

[33] J. Kittler, M. Hatef, R.P. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.

[34] J. Kittler, M. Hatef, R.P. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.

[35] L.I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, John Wiley & Sons, New York, 2004.