

Diagnosing Coronary Heart Disease Using Ensemble Machine Learning

Kathleen H. Miao¹, Julia H. Miao¹, and George J. Miao²

¹Cornell University, Ithaca, NY 14850, USA

²Flezi, LLC, San Jose, CA 95134, USA

Abstract—Globally, heart disease is the leading cause of death for both men and women. One in every four people is afflicted with and dies of heart disease. Early and accurate diagnoses of heart disease thus are crucial in improving the chances of long-term survival for patients and saving millions of lives. In this research, an advanced ensemble machine learning technology, utilizing an adaptive Boosting algorithm, is developed for accurate coronary heart disease diagnosis and outcome predictions. The developed ensemble learning classification and prediction models were applied to 4 different data sets for coronary heart disease diagnosis, including patients diagnosed with heart disease from Cleveland Clinic Foundation (CCF), Hungarian Institute of Cardiology (HIC), Long Beach Medical Center (LBMC), and Switzerland University Hospital (SUH). The testing results showed that the developed ensemble learning classification and prediction models achieved model accuracies of 80.14% for CCF, 89.12% for HIC, 77.78% for LBMC, and 96.72% for SUH, exceeding the accuracies of previously published research. Therefore, coronary heart disease diagnoses derived from the developed ensemble learning classification and prediction models are reliable and clinically useful, and can aid patients globally, especially those from developing countries and areas where there are few heart disease diagnostic specialists.

Keywords—accuracy; adaptive Boosting algorithm; AUC; classifier; classification error; coronary heart disease; diagnosis; ensemble learning; F-score; K-S measure; machine learning; precision; prediction; recall; ROC; sensitivity; specificity

I. INTRODUCTION

Globally, heart disease is the leading cause of death for both men and women [1], with more than half of the deaths occurring in men. One in every four people is afflicted with and dies of heart disease, and in the United States, over 610,000 afflicted Americans lose their lives annually [2].

Heart disease encompasses several types of heart conditions. The most common type of heart condition is coronary heart disease [3], which can cause heart attacks that kill more than 370,000 people every year. In the United States, for every 43 seconds, one person suffers a heart attack, and for every minute one person dies of heart disease [2]. As a result, the total annual cost of coronary heart disease, including health care services, medications, and lost productivity, is about \$108.9 billion in the United States.

Coronary heart disease occurs when plaque builds up in a patient's arteries [4]. As plaque continues to accumulate, the patient's coronary arteries detrimentally narrow over time and

reduce blood flow to the heart, thus increasing the risk of heart attack or stroke.

High blood pressure, high cholesterol, and smoking are three key risk factors for heart disease. Several other medical conditions and lifestyle choices, including diabetes, obesity, poor diet, physical inactivity, and excessive alcohol use, can also place people at a higher risk for heart disease.

Currently, there are four main methods that are utilized to diagnose the severity of heart disease in patients. They include chest X-rays, coronary angiograms, electrocardiograms, also known as ECG or EKG, and exercise stress tests [3]. In terms of diagnosing heart disease and saving the lives of patients, time and diagnostic accuracy at early stages are very crucial. Early detection of coronary heart disease aids physicians in determining the most appropriate treatment and enhances the chances of survival for patients. In many developing countries and areas, however, specialists are not widely available to perform these diagnostic tests. Additionally, for many cases, inaccurate diagnoses and erroneously conducted medical procedures could lead to compromises in the patients' health. Thus, early and accurate diagnoses of heart disease have become immensely important in improving the chances of long-term survival for patients.

Diagnosing coronary heart disease is a challenging task, but computer-aided detection (CAD) have been developed to provide automated predictions for heart disease in patients. As one of the modern computer-aided detection methods, machine learning is an emerging technology for analyzing medical data and providing prognosis on early detection outcomes. One research report used CAD approaches to diagnose heart disease patients based on a method of integrating multiple different types of decision trees [5]. In other research reports, methods include support vector machine (SVM) learning [6]-[8], principal component analysis (PCA)-based evolution classifier [9], rotation forest (RF) classifier [10], artificial neural network (ANN) and fuzzy neural network (FNN) [11], and particle swarm optimization [12]. These methods were developed using the medical data of patients to classify and predict heart disease outcomes.

In this research, an alternative and enhanced machine learning approach is proposed for coronary heart disease prediction based on classification and prediction models utilizing an adaptive Boosting algorithm that combines a set of weak classifiers into a strong ensemble learning prediction model. The developed classification and prediction models

contain two components: an ensemble learning-based training model and a prediction model (also called a diagnosis model). The training model is based on the adaptive Boosting algorithm to form ensemble learning consisting of an optimally weighted majority vote on a number of individual classifiers. On the other hand, the diagnosis model is used to distinguish and classify the presence or absence of coronary heart disease for heart disease outcome predictions. The classification and prediction model for diagnosing coronary heart disease was evaluated using the model sensitivity (or recall), specificity, precision, *F*-score, probability of the model misclassification error and the model accuracy, receiver operating characteristic (ROC) curve, area under the ROC (AUC), and Kolmogorov-Smirnov (K-S) measure.

II. MATERIALS AND METHODS

In this section, the coronary heart disease data sets are introduced. The classification and prediction models for the coronary heart disease prediction based on the ensemble learning using the adaptive Boosting algorithm are presented. Lastly, the evaluation methods of the ensemble learning model are discussed in detail.

A. Heart Disease Dataset

The heart disease data sets, which were used in this research, were obtained from the Heart Disease Databases available in the UCI Machine Learning Repository [13]. These databases contain data information on heart disease clinical instances, contributed by the Cleveland Clinic Foundation (CCF), Hungarian Institute of Cardiology (HIC), Long Beach Medical Center (LBMC), and University Hospital in Switzerland (SUH), respectively.

There are 4 different heart disease databases contributed by 4 different medical institutions, including CCF, HIC, LBMC, and SUH. The databases contain 303 clinical instances, 294 clinical instances, 200 clinical instances, and 123 clinical instances in each data set, respectively. This results in a total combination of 920 clinical instances.

Each heart disease database has the same clinical instance format for each patient. Each clinical instance contains a total of 75 attributes and one target attribute. The target attribute refers to the status of the presence of heart disease in the patients. It is represented by an integer valued from “0” to “4,” where “0” signifies absence and the values (“1,” “2,” “3,” and “4”) signify the presence and severity of heart disease. In this research, the target attribute is reclassified into a binary value of “0” or “1,” indicating the diagnoses of absence or presence of coronary heart disease in the patients, respectively.

B. Adaptive Boosting Algorithm and its Classifiers

In this section, the diagnostic method for predicting and classifying the presence or absence of coronary heart disease is designed and developed based on ensemble learning classification and prediction models using an adaptive Boosting algorithm. The developed ensemble learning classification and prediction (or diagnostic) models, associated with their algorithms and methods, are presented in detail.

The adaptive Boosting algorithm, also known as “AdaBoost,” is a machine learning meta-algorithm [14]. This

algorithm is adaptive because it runs multiple iterations to generate a strong composite ensemble learning method by using an optimally weighted majority vote of a number of weak classifiers. While the individual weak classifiers are only slightly correlated to the true classifier, the adaptive Boosting algorithm creates a strong ensemble learning classifier, which is well-correlated with the resulting true classifier by iteratively adding the weak classifiers.

Given M training data $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_M, y_M)\}$, \mathbf{x}_i is a vector corresponding to an input sample data, associated with P input attributes, and y_i is a target variable with a class label of either 1 or -1. In this research, the P input attributes are represented by the 75 input attributes in the heart disease data sets that can be utilized to build classification and prediction models.

The adaptive Boosting algorithm can be stated and described in the following [14]-[16]:

Initialize weights $D_1[i] = \frac{1}{M}$ for $i = 1, \dots, M$.

For each iteration, $t = 1, \dots, T$:

- Train a weak classifier using distribution D_t .

- Select a weak classifier with low weighted error:

$$\varepsilon_t = \text{Pr}_{D_t}[h_t[\mathbf{x}_i] \neq y_i]. \quad (1)$$

- Calculate a new component β_t based on its error:

$$\beta_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right). \quad (2)$$

- Update distribution $D_t[i]$ for $i = 1, \dots, M$:

$$D_{t+1}[i] = \frac{D_t[i] e^{-\beta_t y_i h_t[\mathbf{x}_i]}}{Z_t}. \quad (3)$$

where Z_t is a normalization constant such that the weights $D_{t+1}[i]$ sum to one.

After all of the boosting iterations, a final ensemble learning classifier, which has a weighted error that is better than chance, is obtained by combining all weak classifiers with an optimal weight,

$$H[\mathbf{x}] = \text{sign}(\sum_{t=1}^T \beta_t h_t[\mathbf{x}]). \quad (4)$$

Eq. (4) is guaranteed to have a lower exponential loss over the training samples. This is equivalent to say that the final classifier $H[\mathbf{x}]$ is computed as a weighted majority vote of the weak classifiers $h_t[\mathbf{x}]$, where each classifier is assigned by weighting β_t .

During the training, the adaptive Boosting iterations also decrease the classification error of the ensemble learning classifier over the training samples. In addition, the classification error must quickly decrease exponentially if the weighted errors of the component classifiers, ε_t , are better than chance, that is, $\varepsilon_t < 0.5$. The ensemble learning-based classification error is bound by

$$\text{err}(H[\mathbf{x}]) \leq \prod_i^T 2\sqrt{\varepsilon_t(1 - \varepsilon_t)}. \quad (5)$$

Furthermore, the weighted error of each new component classifier, ε_t , in Eq. (5) can be expressed:

$$\varepsilon_t = 0.5 - \frac{1}{2} (\sum_i^M D_t[i] y_i h_t[\mathbf{x}_i]). \quad (6)$$

Eq. (6) shows that the weighted error of each new component classifier tends to increase in association with a function of adaptive Boosting iterations.

During each training round, a new weak classifier is added to the ensemble learning process, and a weighting vector is adjusted to focus on training samples that were misclassified in previous rounds. As a result, the final model $H[\mathbf{x}]$ is a classifier that has a higher accuracy than those of the weak classifiers.

C. The Methods of Adaptive Boosting Model Evaluations

In order to evaluate the performances of the adaptive Boosting algorithm-based ensemble learning classification and prediction models, one of the best methods is to analyze the model's accuracy and misclassification error, sensitivity (also known as recall), specificity, precision, F -score, ROC, AUC, and K -S measure using the training and testing data sets. In this research, these analyses depend on the number of false positive and false negative instances of the heart disease data according to the references [17]-[21]. The diagnostic results, associated with the positive or negative results for distinguishing between presence and absence of coronary heart disease from the ensemble learning classification and prediction model, are shown in Table 1.

TABLE I. A MATRIX OF THE DEVELOPED ENSEMBLE LEARNING CLASSIFICATION AND PREDICTION MODELS' DIAGNOSTIC RESULTS FOR DISTINGUISHING BETWEEN PRESENCE AND ABSENCE OF CORONARY HEART DISEASE

	Actual Heart Disease	Actual No Heart Disease	Total Number
Predicted heart disease patients	True Positive (TP)	False Positive (FP)	TP + FP
Predicted no heart disease patients	False Negative (FN)	True Negative (TN)	FN + TN
Total Number	TP + FN	FP + TN	TP + FP + FN + TN

The sensitivity is defined as the probability of correctly identifying the presence of heart disease in patients given by [18],

$$Sensitivity = \frac{TP}{TP+FN}. \quad (7)$$

The sensitivity is also referred to as the *true positive rate* or *recall* in the field of machine learning.

The *specificity* is defined as the probability of correctly identifying the absence of heart disease in patients given by,

$$Specificity = \frac{TN}{FP+TN}. \quad (8)$$

The specificity is sometimes called the *true negative rate*. The difference of $(1 - specificity)$ is known as the *false positive rate*.

The *precision* or the *positive predictive value* is defined as

$$Precision = \frac{TP}{TP+FP}. \quad (9)$$

Thus, the probability of the misclassification error (PME) is obtained by

$$PME = \frac{FN+FP}{TP+FN+FP+TN}, \quad (10)$$

and the model's accuracy is defined by

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN}, \quad (11)$$

where the model's accuracy = $(1 - PME)$.

Notice that both the recall in Eq. (7) and precision in Eq. (9) are in a mutual relationship based on the understanding and measure of relevance. The recall is a measure of quantity, while the precision is a measure of quality. Thus, based on the harmonic mean of recall and precision, the relationship between the recall and precision definitions is given by a F -score, which is defined as

$$F_score = 2 \left(\frac{Precision \times Recall}{Precision + Recall} \right), \quad (12)$$

where a F -score of 1 would signify the best score in terms of accuracy of the classification and prediction model, and a F -score of 0 would be the worst score.

Thus, the F -score in Eq. (12) is used to measure the model performances and likewise can be used as a single measure of a model's accuracy during the testing. In addition, the F -score can also be interpreted as a weighted average of the recall and precision.

A ROC curve for classification and prediction models is a graph plot, which is obtained by using a set of trade-off points between the sensitivity and the difference of $(1 - specificity)$ for cases classified as presence of heart disease. The corresponding AUC under the ROC curve can be used to evaluate and rank the quality of the performance of classification and prediction models [18]. To estimate the AUC, a trapezoidal approximation formula is given by [18], [22],

$$\int_0^1 f(x)dx \cong \sum_{i=0}^N \left(\frac{y_i + y_{i+1}}{2} \right) (x_{i+1} - x_i), \quad (13)$$

where $f(x)$ denoted the function of the ROC curve analysis, y_i and x_i represented the sensitivity and $(1 - specificity)$ at the i th ($i = 0, 1, 2, \dots, M$) point, respectively. An AUC of 1 represents that the classification and prediction model is a perfect model in terms of diagnostic accuracy in distinguishing the presence of heart disease from absence of heart disease. On the other hand, an AUC of 0.5 indicates that the model is simply based on chance and is unmeaningful. Thus, the higher the AUC is, the better the classification and prediction model performs.

The AUC under the ROC curve is one of the most important parameters to evaluate and rank the quality of the performance of classification and prediction models under a condition of balance samples; that is, the number of presence and the number of absence of heart disease cases are approximately equal in the training and testing data sets. However, if unbalanced samples are represented in the data set, the F -score in Eq. (12) is the most important parameter for quality evaluation of the classification and prediction models. In that case, the AUC would not be an effective method of ranking the quality of the performance of the classification and prediction models.

In this research, the K -S measure [20], [21] will also be used to measure performance of the ensemble learning classification and prediction models. More accurately, in our

research, the *K-S* measure is used to determine the degree of separation between the distributions of the presence and absence of heart disease in patients. The *K-S* measure can achieve a value of 100% if the scores of the model partition the population into two separate groups, in which one group contains all clinical instances classified with presence of coronary heart disease and the other consists of all clinical instances classified with absence of heart disease. In other words, the *K-S* measure results in 100% if output probabilities (or model scores) of the developed ensemble learning classification and prediction model allow the results of the presence and absence of heart disease in patients to be perfectly separated. In an unusual case, the *K-S* measure would be 0 if the developed ensemble learning classification and prediction model cannot differentiate between presence and absence of coronary heart disease. However, in most cases for classification models, the *K-S* measure will fall in a range between 0% and 100%. Thus, the higher the *K-S* measure value is, the better the developed ensemble learning classification and prediction model is at diagnosing the presence or absence of coronary heart disease in patients.

III. RESULTS

In this paper, an advanced ensemble machine learning technology, utilizing an adaptive Boosting algorithm, is proposed for accurate heart disease diagnosis and outcome predictions. The proposed adaptive Boosting model is an ensemble machine learning meta-algorithm, which combines a set of outputs from other learning algorithms into a weighted sum, thereby converging multiple mathematical models into a strong and enhanced classification and prediction model.

The proposed ensemble learning classification and prediction models were applied to 4 different data sets for coronary heart disease diagnosis. With data collected from four different medical institutions, these 4 data sets contain clinical instances of patients diagnosed with heart disease: 303 instances from the CCF, 294 instances from the HIC, 200 instances from the LBMC, and 123 instances from the SUH. Table 2 shows the details of the clinical instances in terms of the number of cases with the presence or absence of coronary heart disease in each of the 4 data sets, after the removal of clinical instances with missing values.

As can be seen in Table 2, there are large differences in terms of the percentage of the presence of coronary heart disease in patients, with the lowest at 36.18% and the highest at 93.44% in the data sets.

In each data set, each clinical instance consists of 76 raw attributes. Among all of the raw attributes, only 29 of them were used for developing the ensemble learning classification and prediction models due to a large number of missing values. Table 3 lists the detailed 29 raw attributes, which had been used for the model development in this research.

To evaluate the performances of the developed ensemble learning classification and prediction models based on the adaptive Boosting algorithm, the probabilities of the model misclassification error and the model accuracy were estimated using a nonparametric approach based on a *holdout* method [23]. The holdout method is also known as the *H* method. The

TABLE II. THE CLINICAL INSTANCES IN TERMS OF THE PRESENCE AND ABSENCE OF HEART DISEASE IN EACH DATA SET

	Cases with Presence of Heart Disease	Cases with Absence of Heart Disease	Total Clinical Instance Cases	Percentage of the Presence of Heart Disease in Data Set
CCF	125	157	282	44.33%
HIC	106	187	293	36.18%
LBMC	113	32	145	77.93%
SUH	114	8	122	93.44%

TABLE III. THE 29 RAW ATTRIBUTE NAMES AND THEIR DESCRIPTIONS

Variable Name	Descriptions	Variable Name	Descriptions
Age	Age in years	Tpeakbps	Peak exercise blood pressure (first of 2 parts)
Sex	Sex (1 = male; 0 = female)	Tpeakbpd	Peak exercise blood pressure (second of 2 parts)
CP	Chest pain type (1 = typical angina; 2 = atypical angina; 3 = non-anginal pain; 4 = asymptomatic)	dummy	Integer, from 94 to 200
Htn	Binary, 0 and 1	tresrbpd	Resting blood pressure
Chol	Serum cholesterol in mg/dl	exang	Exercise induced angina (1 = yes; 0 = no)
Restecg	Resting electrocardiographic results (0 = normal; 1 = having ST-T wave abnormality where T wave inversions and/or ST elevation or depression of > 0.05 mV; 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria)	xhypo	(1 = yes; 0 = no)
Ekgmo	Month of exercise ECG reading	oldpeak	ST depression induced by exercise relative to rest
Ekgday	Day of exercise ECG reading	cmo	Month of cardiac cath
Ekgyr	Year of exercise ECG reading	cday	Day of cardiac cath
Prop	Beta blocker used during exercise ECG (1 = yes; 0 = no)	cyr	Year of cardiac cath
Nitr	Nitrates used during exercise ECG (1 = yes; 0 = no)	Lvx3	Integer, from 0 to 8
Pro	Calcium channel blocker used during exercise ECG (1 = yes; 0 = no)	Lvx4	Integer, from 0 to 8
Thaldur	Duration of exercise test in minutes	lvf	Integer, from 1 to 4
Thalach	Maximum heart rate achieved	num	Diagnosis of heart disease (angiographic disease status) 0 (< 50% diameter narrowing) 1 (> 50% diameter narrowing)
Thalrest	Resting heart rate		

pattern data $\{X, \theta\}$ are partitioned into two mutually exclusive data sets $\{X, \theta\}_1$ and $\{X, \theta\}_2$. For the holdout method, the ensemble learning classification and prediction models were trained using the training data $\{X, \theta\}_1$, and then the ensemble learning classification and prediction models were tested using the testing data $\{X, \theta\}_2$.

Each data set was separated into equally sized training and testing data sets. The ensemble learning classification and prediction models were trained and tested by using the training and testing data sets, respectively. To train the classification and prediction models, the adaptive Boosting

algorithm parameters were set to 100 iterations for the CCF, HIC, LBMC, and SUH.

Fig. 1 displays the trained and tested classification error curves at each iteration using the CCF training and testing data sets. Fig. 2 also shows the trained and tested classification error curves at each iteration using the HIC training and testing data sets. For the LBMC, the trained and tested classification error curves at each iteration are shown in Fig. 3, using the LBMC training and testing data sets. Finally, the trained and tested classification error curves using the SUH training and testing data sets are shown in Fig. 4.



Fig. 1. The model classification error plot at each iteration, where the red curve represents the training error and the green curve represents the testing error by using the CCF training and testing data sets, respectively.

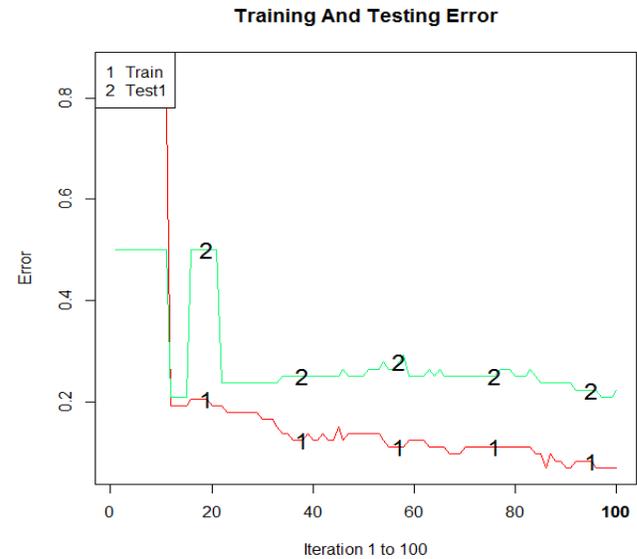


Fig. 3. The model classification error plot at each iteration, where the red curve represents the training error and the green curve represents the testing error by using the LBMC training and testing data sets, respectively.

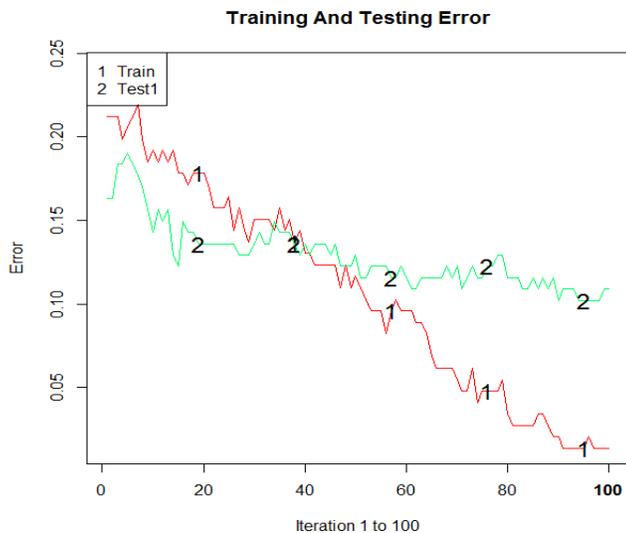


Fig. 2. The model classification error plot at each iteration, where the red curve represents the training error and the green curve represents the testing error by using the HIC training and testing data sets, respectively.

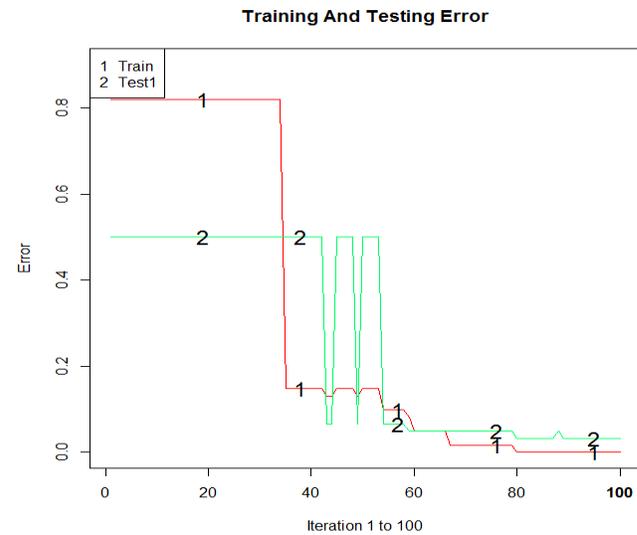


Fig. 4. The model classification error plot at each iteration, where the red curve represents the training error and the green curve represents the testing error by using the SUH training and testing data sets, respectively.

TABLE IV. TRAINING RESULTS OF THE MODEL PERFORMANCES FOR THE CCF, HIC, LBMC, AND SUH USING THE TRAINING DATA SETS

	TP	FP	FN	TN	Total	Sensitivity (Recall)	Specificity	Precision	F-Score	Error	Accuracy
CCF	60	1	3	77	141	95.24%	98.72%	98.36%	0.97	2.84%	97.16%
HIC	54	0	2	90	146	96.43%	100%	100%	0.98	1.37%	98.63%
LBMC	57	5	0	11	73	100%	68.75%	91.94%	0.96	6.85%	93.15%
SUH	50	0	0	11	61	100%	100%	100%	1	0%	100%

TABLE V. TESTING RESULTS OF THE MODEL PERFORMANCES FOR THE CCF, HIC, LBMC, AND SUH USING THE TESTING DATA SETS

	TP	FP	FN	TN	Total	Sensitivity (Recall)	Specificity	Precision	F-Score	Error	Accuracy
CCF	44	10	18	69	141	70.97%	87.34%	81.48%	0.76	19.86%	80.14%
HIC	40	6	10	91	147	80.00%	93.81%	86.96%	0.83	10.88%	89.12%
LBMC	54	14	2	2	72	96.43%	12.50%	79.41%	0.87	22.22%	77.78%
SUH	56	2	0	3	61	100%	60.00%	96.55%	0.98	3.28%	96.72%

Table 4 displays the detailed training model performances of the developed ensemble learning classification and prediction models in predicting the presence and absence of coronary heart disease using the training data sets. The training results of the model accuracies of the developed ensemble learning classification and prediction models were the following: 97.16% for CCF, 98.63% for HIC, 93.15% for LBMC, and 100% for SUH. The corresponding *F*-score for the trained ensemble learning classification and prediction models were 0.97 for CCF, 0.98 for HIC, 0.96 for LBMC and 1 for SUH.

Table 5 shows the detailed testing model performances of the developed ensemble learning classification and prediction models in predicting the presence and absence of coronary heart disease using the testing data sets. As shown, the testing results of the model accuracies of the developed ensemble learning classification and prediction models were the following: 80.14% for CCF, 89.12% for HIC, 77.78% for LBMC, and 96.72% for SUH. The corresponding *F*-scores for

the tested ensemble learning classification and prediction models were 0.76 for CCF, 0.83 for HIC, 0.87 for LBMC and 0.98 for SUH.

The ROC curve results of the developed ensemble learning classification and prediction models were produced by varying a set of trade-off points between the model sensitivity on the y-axis and the difference value (1 – specificity) on the x-axis for CCF, HIC, LBMC, and SUH as shown in Figures 5, 6, 7, and 8, respectively. The corresponding estimated AUCs under the ROC curve for CCF, HIC, LBMC, and SUH were 0.8526, 0.9212, 0.6864, and 0.6357, respectively. The estimated AUCs of the ROC curves based on CCF and HIC implied that the proposed ensemble learning classification and prediction models can provide a consistently high accuracy in diagnosing and classifying presence of heart disease and absence of heart disease for predicting coronary heart disease outcome. Additionally, because the samples are approximately balanced in terms of the presence and absence heart disease cases in both of the CCF and HIC data sets, the AUCs under the ROC curves

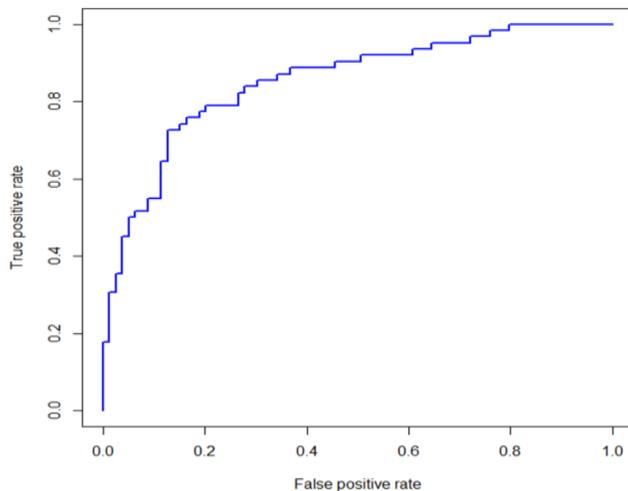


Fig. 5. An estimated AUC under the ROC curve of the developed ensemble learning classification and prediction model for cases classified as presence of heart disease and absence of heart disease in the CCF dataset, where the true positive rate is sensitivity on the y-axis and the false positive rate is the difference (1 – specificity) on the x-axis.

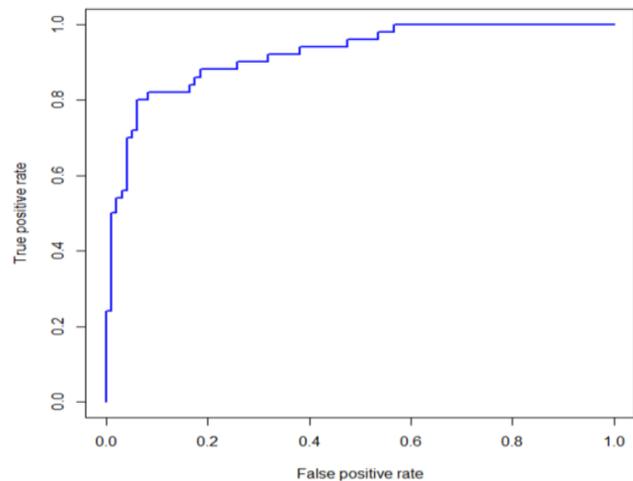


Fig. 6. An estimated AUC under the ROC curve of the developed ensemble learning classification and prediction model for cases classified as presence of heart disease and absence of heart disease in the HIC dataset, where the true positive rate is sensitivity on the y-axis and the false positive rate is the difference (1 – specificity) on the x-axis.

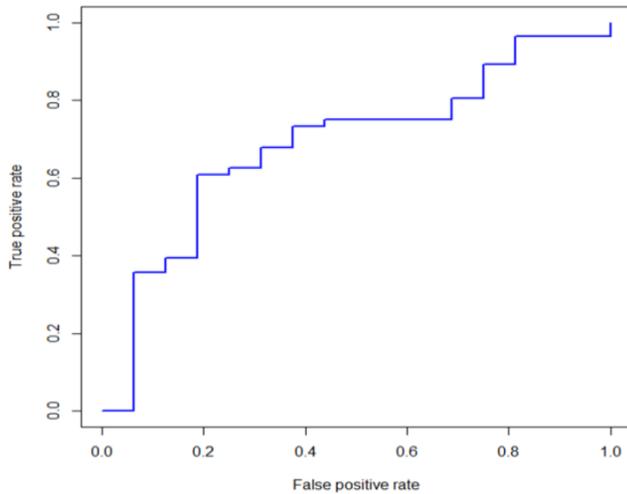


Fig. 7. An estimated AUC under the ROC curve of the developed ensemble learning classification and prediction model for cases classified as presence of heart disease and absence of heart disease in the LBMC dataset, where the true positive rate is sensitivity on the y-axis and the false positive rate is the difference $(1 - \text{specificity})$ on the x-axis.

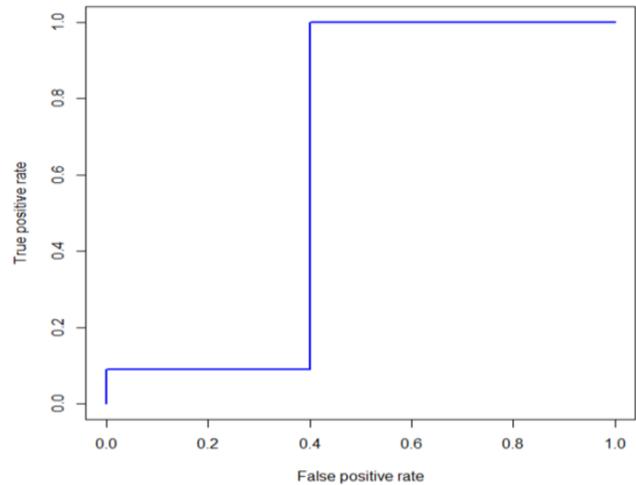


Fig. 8. An estimated AUC under the ROC curve of the developed ensemble learning classification and prediction model for cases classified as presence of heart disease and absence of heart disease in the SUH dataset, where the true positive rate is sensitivity on the y-axis and the false positive rate is the difference $(1 - \text{specificity})$ on the x-axis.

can be used to evaluate and rank the quality of the performances of the ensemble learning classification and prediction models.

The developed ensemble learning classification and prediction models also enabled the production of a set of model probabilities (also called the model scores), which were associated with the presence and absence of coronary heart disease in the cases crossing over the 4 datasets. By sorting the model scores, the *K-S* charts were generated according to the cumulative counts of instances of the presence and absence of coronary heart disease cases.

As a result, Fig. 9 shows a *K-S* chart of the CCF with the highest *K-S* value of 58.66% at the 4th decile population. Fig. 10 is a *K-S* chart of the HIC with the highest *K-S* value of 66.54% located at the 4th decile population. Fig. 11 is the *K-S* chart of the LBMC with the highest *K-S* value of 41.96% at the 5th decile population. For the SUH, the *K-S* chart is shown in Fig. 12 with the highest *K-S* value of 52.86% located at the 9th decile population. As shown in the charts from Fig. 9 to Fig. 12, the highest *K-S* values are consistently associated with the tested model accuracies as listed in Table 5. Likewise, the higher the highest *K-S* test value is, the better and more accurate the developed ensemble learning classification and prediction model is in distinguishing between the presence and absence of coronary heart disease in patients.

Therefore, when applied to patients with chest pain syndromes and intermediate disease prevalence, the diagnostic results of coronary heart disease diagnoses derived from the ensemble learning classification and prediction models are reliable and clinically useful. The results can be used to aid patients, especially those in developing countries and areas where there are few heart disease diagnostic specialists

available.

IV. DISCUSSION

In this research, the ensemble learning classification and prediction models were designed and developed based on an adaptive Boosting algorithm. The developed classification and prediction models were utilized to diagnose and classify the presence and absence of coronary heart disease in diagnostic outcome predictions. The developed ensemble learning classification and prediction models were applied to 4 different coronary heart disease databases, where data sets were collected from 4 different medical institutions at the CCF, HIC, LBMC, and SUH. The performances of the developed ensemble learning classification and prediction models were tested and measured by using the training and testing data sets. Based on these testing results, the developed ensemble learning classification and prediction models were further evaluated by using the model accuracy and misclassification error, sensitivity (or recall), precision, specificity, *F*-score, ROC curve, AUC, and the *K-S* measure.

As shown in Table 5, the tested model accuracies of the developed ensemble learning classification and prediction models, utilizing the 28 input attributes, were the following: 80.14% for the CCF, 89.12% for the HIC, 77.78% for the LBMC, and 96.72% for the SUH using the testing data sets. Furthermore, the *F*-scores of the developed ensemble learning classification and prediction models were 0.76 for the CCF, 0.83 for the HIC, 0.87 for the LBMC, and 0.98 for the SUH. The corresponding AUCs under the ROC curves were 0.8526 for the CCF, 0.9212 for the HIC, 0.6864 for the LBMC, and 0.6357 for the SUH. In addition, the highest *K-S* values of the developed ensemble learning classification and prediction model were 58.66% for the CCF, 66.54% for the HIC, 41.96%

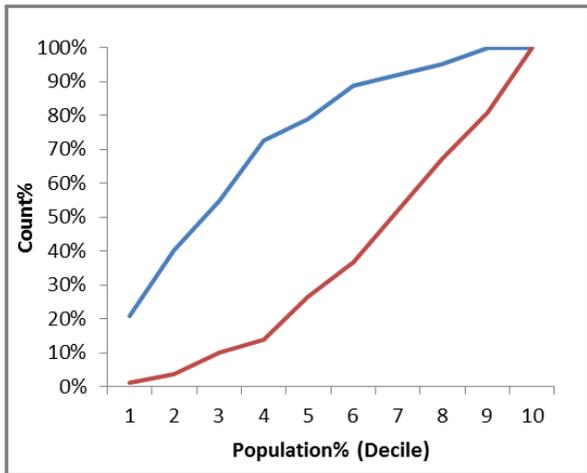


Fig. 9. The K-S chart for the CCF was generated by using the model output probabilities. The highest K-S value is 58.66%, located at the 4th decile population.

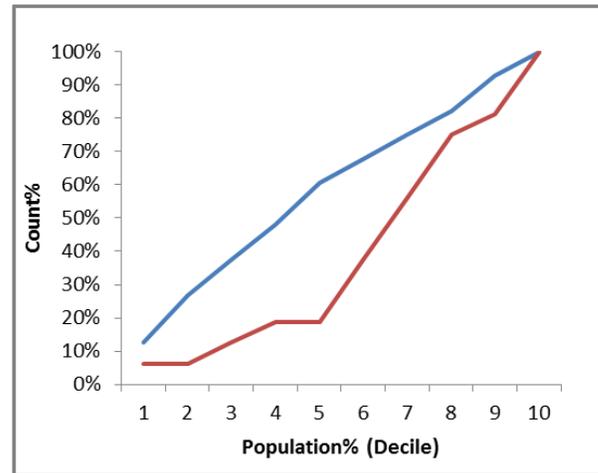


Fig. 11. The K-S chart for the LBMC was generated by using the model output probabilities. The highest K-S value is 41.96%, located at the 5th decile population.

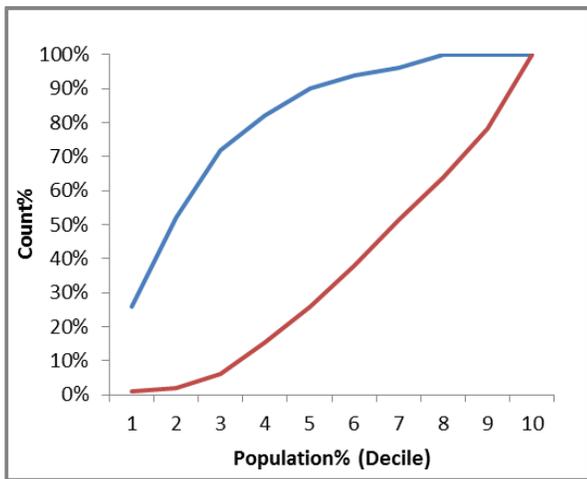


Fig. 10. The K-S chart for the HIC was generated by using the model output probabilities. The highest K-S value is 66.54%, located at the 4th decile population.

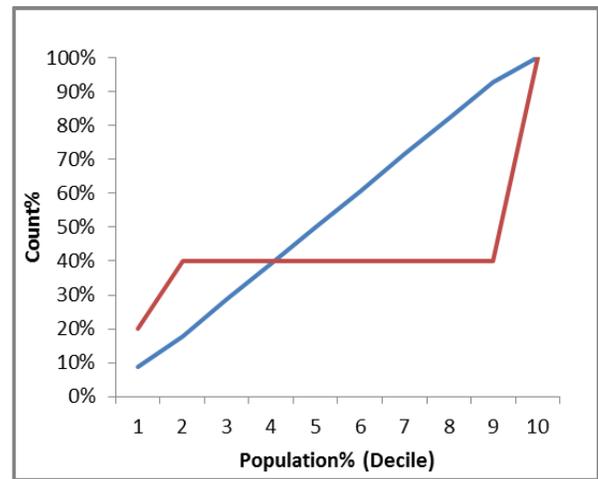


Fig. 12. The K-S chart for the SUH was generated by using the model output probabilities. The highest K-S value is 52.86%, located at the 9th decile population.

for the LBMC, and 52.86% for the SUH. Thus, based on the testing results, the average diagnostic accuracy of the developed ensemble learning classification and prediction model would be 85.27% accurate in distinguishing between presence and absence of coronary heart disease in a new patient with clinical heart disease data, crossing over the 4 different locations in the CCF, HIC, LBMC, and SUH overall. Additionally, the average developed model sensitivity (or recall) was 86.61%; the average specificity was 83.76%; the average model precision was 85.84%; the average model F-score 0.86; and the average highest K-S value was 55.01%.

Thus, the developed ensemble learning classification and prediction models were able to achieve a consistently high accuracy in diagnosing the presence and absence of coronary heart disease for heart disease patient outcome predictions.

In comparison to related papers, there were several different methods developed using the same heart disease data sets. However, the methods associated with these developed models only considered 13 input attributes and in most cases were developed to classify and predict heart disease outcomes using only one of the 4 data sets. In general, these previous methods showed different performances in terms of the model

accuracies within a range of approximately 77% to 85%. The previous model accuracies of using a new probability algorithm [24] showed 77% for the HIC, 79% for the LBMC, and 81% for the SUH. The classification accuracy was 77% for the CCF based on the instance-based prediction model [25]. The conceptual clustering model [26] achieved 78.9% accuracy on the CCF data set. A decision tree (J4.8) was 78.9% accuracy and a Bagging algorithm [27] achieved 81.41% accuracy in diagnosing heart disease for the CCF data set. Recently, the data mining approaches [28], including Naïve Bayes, J48 decision tree, and Bagging algorithm, achieved the model accuracies of 82.31%, 84.35%, and 85.03% for the HIC data, respectively.

On the other hand, in this research, the developed ensemble learning classification and prediction models based on the 28 input attributes were not only applied to the CCF data set but also applied to the HIC, LBMC, and SUH data sets. The testing results, as shown in Table 5 and Figures from 5 to 12, also indicate that the model accuracy of the developed ensemble learning classification and prediction models is comparably higher than most of those of the previously published methods. In addition, the developed ensemble learning classification and prediction models had more flexibility due to its use of the adaptive Boosting algorithm, regardless of whether or not there were overlapping data (or clusters) between the presence and absence of heart disease cases. The developed ensemble learning classification and prediction models moreover provided a more reliable and greater percentage of accuracy in distinguishing between the presence and absence of coronary heart disease in the patient outcome predictions.

Therefore, the proposed ensemble learning classification and prediction models achieve significant potential in reducing the number of unnecessary, inaccurate diagnoses and erroneously conducted medical procedures that have compromised patients' health. The proposed ensemble learning classification and prediction models enable early and accurate heart disease diagnose and thus help improve chances of long-term survival for heart disease patients and save millions of lives.

V. CONCLUSION AND FUTURE WORK

In this paper, ensemble learning classification and prediction models have been developed to diagnose and classify the presence and absence of coronary heart disease in patient outcome predictions; additionally, the model accuracies, sensitivities (or recalls), precisions, specificities, F -scores, ROC curves, AUCs, and K -S measures have been evaluated. The developed classification and prediction models, based on the adaptive Boosting algorithm, were ensemble learning classifiers that had high flexibility in adjusting a weighting vector to generate a strong, single composite ensemble learning classification and prediction model by using an optimally weighted majority vote of a number of weak classifiers.

The developed ensemble learning classification and prediction models were trained and tested using the holdout method based on 4 different data sets from 4 different medical institutions. The testing results showed that the developed ensemble learning classification and prediction models had an

average sensitivity (or recall) of 86.61% in diagnosing the presence of heart disease, an average specificity of 83.76% in diagnosing the absence of coronary heart disease, an average model precision of 85.84%, an average model F -score of 0.86, and an average model accuracy of 85.27% in diagnosing both the presence and absence of coronary heart disease. In each data set, the accuracies of the testing results of the ensemble machine learning models were the following: 80.14% for CCF, 89.12% for HIC, 77.78% for LBMC, and 96.72% for SUH. Therefore, the developed ensemble learning classification and prediction models using the 28 input attributes can provide highly accurate and consistent diagnoses for coronary heart disease patient outcome predictions, thereby allowing patients to bypass unnecessary, inaccurate diagnoses and erroneously conducted medical procedures.

From Fig. 1 to Fig. 4, the classification errors based on the testing data sets are higher than the classification errors based on the training data sets at the 100th iterations, where data sets were collected from the 4 different medical institutions. This phenomenon involving the differences of the classification errors between the model training and testing processes is an expected encounter, known as an over-fitting problem in the field of machine learning during model development. Minimizing training error will often result in the over-fitting problem during each iteration in the adaptive Boosting algorithm since the Boosting algorithm is sensitive to noise and/or outlier samples. Thus, in future research, other enhanced methods that prevent and/or reduce the over-fitting problem associated with the adaptive Boosting algorithm during a training process would be investigated, thereby further enhancing the performances of the ensemble learning classification and prediction model and coronary heart disease diagnosis.

ACKNOWLEDGMENT

The authors would like to thank Dr. Andras Janosi from the Hungarian Institute of Cardiology, Budapest; Dr. William Steinbrunn from University Hospital, Zurich, Switzerland; Dr. Matthias Pfisterer from University Hospital, Basel, Switzerland; and Dr. Robert Detrano from V.A. Medical Center, Long Beach and the Cleveland Clinic Foundation, whose heart disease datasets of clinical instances were contributed to and made available from the Heart Disease Databases in the UCI Machine Learning Repository. These datasets contain the clinical instances and information regarding the diagnoses of the presence and absence of heart disease in clinical cases.

REFERENCES

- [1] Centers for Disease Control and Prevention (CDC), Deaths: leading causes for 2008, National Vital Statistics Reports, Vol. 60, No. 6, June 6, 2012.
- [2] Centers for Disease Control and Prevention (CDC), Heart disease in the United States, Available <http://www.cdc.gov/heartdisease/facts.htm>.
- [3] Centers for Disease Control and Prevention (CDC), Know the facts about heart disease, Available http://www.cdc.gov/heartdisease/docs/consumered_heartdisease.pdf
- [4] American Heart Association, What is a heart attack? Available http://www.heart.org/HEARTORG/Conditions/Conditions_UCM_001087_SubHomePage.jsp

- [5] M. Shouman, T. Turner, and R. Stocker, "Using decision tree for diagnosing heart disease patients," *Proceedings of the 9th Australasian Data Mining Conference*, Ballarat, Australia, pp. 23–29, 2011.
- [6] A. F. Ootom, E. E. Abdallah, Y. Kilani, A. Kefaye, and M. Ashour, "Effective diagnosis and monitoring of heart disease," *International Journal of Software Engineering and Its Applications*, Vol. 9, No. 1, pp. 143-156, 2015.
- [7] S. Ghumbre, C. Patil, and A. Ghatol, "Heart disease diagnosis using Support Vector Machine," *International Conference on Computer Science and Information Technology*, Pattaya, pp. 84-88, December 2011.
- [8] S. Bhatia, P. Prakash, and G. N. Pillai, "SVM based decision support system for heart disease classification with integer-coded genetic algorithm to select critical features," *Proceedings of the World Congress on Engineering and Computer Science*, San Francisco, California, October 22-24, 2008.
- [9] L. Pasi and L. Jouni, "A classification method based on principal component analysis and differential evolution algorithm applied for prediction diagnosis from clinical EMR heart data sets," *Computational Intelligence in Optimization*, Springer-Verlag, Vol. 7, pp. 263-283, 2010.
- [10] A. Ozcift and A. Gulden, "Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms," *Journal of Computer Methods and Programs in Biomedicine*, Vol. 104, pp. 443-451, 2011.
- [11] H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," *Journal of Expert Systems with Applications*, Vol. 35, pp. 82-89, 2008.
- [12] P. Gayathri and N. Jaisankar, "Comprehensive study of heart disease diagnosis using data mining and soft computing techniques," *International Journal of Engineering and Technology*, Vol. 5, No. 3, pp. 2947-2957, July 2013.
- [13] UCI Machine Learning Repository, Heart disease data set, available <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>.
- [14] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Springer*, Vol. 37, Issue 3, pp 297-336, December 1999.
- [15] Y. Freund and R. E. Schapire, "Experiments with a new Boosting algorithm," AT&T Research, Murray Hill, New Jersey, pp. 1-15, January 22, 1996, available <http://www.research.att.com/orgs/ssr/people/{yoav.schapire/>
- [16] R. E. Schapire, "Explaining AdaBoost," in *Empirical Inference*, B. Scholkopf, Z. Luo and V. Vovk, Eds., Springer Berlin Heidelberg, Chapter 5, pp 37-52, October 2013.
- [17] G. J. Miao, K. H. Miao, and J. H. Miao, "Neural pattern recognition model for breast cancer diagnosis," *Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Bioinformatics*, August Edition, pp. 1–8, September 2012.
- [18] K. H. Miao and G. J. Miao, "Mammographic diagnosis for breast cancer biopsy predictions using neural network classification model and receiver operating characteristic (ROC) curve evaluation," *Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Bioinformatics*, September Edition, Vol. 3, Issue 9, pp. 1–10, October 2013.
- [19] J. H. Miao, K. H. Miao, and G. J. Miao, "Breast cancer biopsy predictions based on mammographic diagnosis using Support Vector Machine learning," *Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Bioinformatics*, Vol. 5, No. 4, pp. 1–9, 2015.
- [20] A. M. Law, "A tutorial on how to select simulation input probability distributions," in *IEEE Proceedings of the 2013 Winter Simulation Conference*, R. Pasupathy, S. H. Kim, A. Tolk, et al., Eds., pp. 306–320, 2013.
- [21] T. B. Arnold and J. W. Emerson, "Nonparametric goodness-of-fit tests for discrete null distributions," *The R Journal*, Vol. 3/2, pp. 34-39, December 2011.
- [22] R. L. Finney, F. D. Demana, B. K. Waits, and D. Kennedy, *Calculus: A Complete Course*, Second Edition, Addison Wesley Longman, Inc., 2000.
- [23] G. J. Miao and M. A. Clements, *Digital Signal Processing and Statistical Classification*, Artech House, Inc., 2002.
- [24] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, and J. J. Schmid, et al., "International application of a new probability algorithm for the diagnosis of coronary artery disease," *American Journal of Cardiology*, Vol. 64, pp. 304-310, 1989.
- [25] D. W. Aha and D. Kibler, "Instance-based prediction of heart-disease presence with the Cleveland database," *Technical Report*, University of California, Irvine, Department of Information and Computer Science, Number ICS-TR-88-07, 1988.
- [26] J. H. Gennari, P. Langley, and D. Fisher, "Models of incremental concept formation," *Artificial Intelligence*, Vol. 40, Issues 1-3, pp. 11-6, September 1989.
- [27] M. C. Tu, D. Shin, and D. K. Shin, "Effective diagnosis of heart disease through Bagging approach," *IEEE 2nd International Conference on Biomedical Engineering and Informatics*, October 17-19, pp. 1-4, 2009.
- [28] V. Chaurasia and S. Pal, "Data mining approach to detect heart disease," *International Journal of Advanced Computer Science and Information Technology*, Vol. 2, No. 4, pp. 56-66, 2013.