Coronary Heart Disease Diagnosis using Deep Neural Networks

*Kathleen H. Miao[a, b], *Julia H. Miao[a]
[a]Cornell University, Ithaca, NY 14853, USA
[b]New York University School of Medicine, New York, NY 10016, USA

***Abstract*—According to the World Health Organization, cardiovascular disease (CVD) is the top cause of death worldwide. In 2015, over 30% of global deaths was due to CVD, leading to over 17 million deaths, a global health burden. Of those deaths, over 7 million were caused by heart disease, and greater than 75% of deaths due to CVD were in developing countries. In the United States alone, 25% of deaths is attributed to heart disease, killing over 630,000 Americans annually. Among heart disease conditions, coronary heart disease is the most common, causing over 360,000 American deaths due to heart attacks in 2015. Thus, coronary heart disease is a public health issue. In this research paper, an enhanced deep neural network (DNN) learning was developed to aid patients and healthcare professionals and to increase the accuracy and reliability of heart disease diagnosis and prognosis in patients. The developed DNN learning model is based on a deeper multilayer perceptron architecture with regularization and dropout using deep learning. The developed DNN learning model includes a classification model based on training data and a prediction model for diagnosing new patient cases using a data set of 303 clinical instances from patients diagnosed with coronary heart disease at the Cleveland Clinic Foundation. The testing results showed that the DNN classification and prediction model achieved the following results: diagnostic accuracy of 83.67%, sensitivity of 93.51%, specificity of 72.86%, precision of 79.12%, *F*-Score of 0.8571, area under the ROC curve of 0.8922, Kolmogorov-Smirnov (K-S) test of 66.62%, diagnostic odds ratio (DOR) of 38.65, and 95% confidence interval for the DOR test of [38.65, 110.28]. Therefore, clinical diagnoses of coronary heart disease were reliably and accurately derived from the developed DNN classification and prediction models. Thus, the models can be used to aid healthcare professionals and patients throughout the world to advance both public health and global health, especially in developing countries and resource-limited areas with fewer cardiac specialists available.**

Keywords*—*Cardiovascular disease; heart disease; coronary artery disease; classification; accuracy; diagnosis; diagnostic odds ratio; deep learning; deep neural network; machine learning;* F-*score; global health; public health;* K-S *test; precision; prediction; prognosis; ROC curve; specificity; sensitivity

I. Introduction

According to the World Health Organization, cardiovascular disease (CVD) is the top cause of mortality worldwide. In 2015, over 30% of global deaths was due to CVD, leading to over 17 million deaths, a global health burden [1]. Of those deaths, over 7 million were caused by heart disease, and greater than 75% of deaths due to CVD were in developing countries [1]. Over 50% of the deaths caused by heart disease were in men [1]. In the United States alone, 25% of deaths is attributed to heart disease, killing over 630,000 Americans annually [2].

Heart disease is an umbrella term that includes many types, including congenital, coronary, and rheumatic heart diseases. Among those conditions, coronary heart disease is the most common, causing over 360,000 American deaths due to heart attacks in 2015 [2]. According to the Centers for Disease Control and Prevention, it is estimated that approximately every 40 seconds, an American experiences a heart attack [2]. Consequently, heart disease expenditures have risen to over $200 billion annually in the United States alone [2]. Furthermore, by 2030, health care costs due to heart disease are expected to double according to the American Heart Association [3].

Coronary heart disease occurs due to atherosclerosis, long-term buildup of plaque in the arteries of a patient caused by the elevation of low-density lipoprotein (LDL) cholesterol in plasma [4]. As the walls of the coronary arteries of the patient accumulate plaque, the arteries narrow over time. Subsequently, this results in reduced blood flow to the muscles of the heart leading to decreased heart movement. Once the artery is partially or completely blocked, there is an increased risk for a heart attack, also known as a myocardial infarction.

Risk factors of coronary heart disease include family history, smoking, high LDL cholesterol levels, high blood pressure, age, and uncontrolled diabetes [5]. Lifestyle and medical factors that can lead to greater risk of heart disease include sedentary lifestyle, unhealthy diet, obesity, and excess alcohol consumption. Symptoms of coronary artery disease include chest pain, pressure, shortness of breath, sweating, heart palpitations, dizziness, weakness, and nausea [4].

To diagnose heart disease severity in patients, current methods that are used include exercise stress tests, chest X-rays, heart scans (CT), cardiac magnetic resonance imaging (MRI), coronary angiograms, and electrocardiograms (EKG) [5]. Early, accurate diagnoses of coronary heart disease in patients are essential for administering early and optimal treatments in order to increase their chances of long-term survival. However, in many resource-limited areas throughout the world, cardiovascular specialists may not be available to perform these diagnostic tests. Furthermore, in many cases, missed diagnoses as well as erroneous diagnoses and treatments place the health of patients at risk. Early detection

*Both authors contributed equally to this work.

of heart disease can lead to preventative measures, such as medications, lifestyle changes, angioplasty, and/or surgery, to reduce disease progression and morbidity [5]. Therefore, accurate and early heart disease diagnoses in patients are crucial for lowering mortality rate and improving their long-term survival rates.

Early diagnosis of coronary heart disease can be challenging, and computer-aided methods have been created to detect and diagnose heart disease in patients. Increasingly utilized among computer-aided detection methods in medical institutions is machine learning, a technology that analyzes clinical data, processes it, and provides diagnoses for medical conditions. Research reports have utilized the following computer-aided detection methods to diagnose heart disease in patients based on their clinical data: decision trees [6], artificial neural network [7], support vector machine learning [8], fuzzy neural network [9], ensemble machine learning [10], binary particle swarm optimization [11], rotation forest classifier [12], principal component analysis-based evolution classifier [13], K-star algorithm [14], Bayesian algorithm [15], rule organization method [16], and neuro fuzzy classifier [17].

In this research, an advanced deep neural network approach is developed and utilized to predict coronary heart disease in patients and increase diagnostic accuracy using classification and prediction models based on deep learning [18, 19]. For this research, the developed classification and diagnosis models contain two parts: a deep neural network learning-based training model and a prediction model for the presence of heart disease. The training model is first created using deep learning algorithms based on a deeper multilayer perceptron with regularization and dropout in system and architecture. Based on the training model, the diagnosis model is then utilized to predict whether or not patients have coronary heart disease. The subsequent performance of the deep learning model for heart disease diagnosis is evaluated in terms of the performance measure parameters, including diagnostic accuracy, probability of misclassification error, sensitivity, specificity, precision, area under the ROC curve (AUC), Kolmogorov-Smirnov (K-S) measure, receiver operating characteristic (ROC), and *F*-score.

II. Materials and Methods

In Section A, the clinical data for coronary heart disease are described. Then, in Section B, the classification and prediction models for coronary heart disease prognosis based on the deep neural network system and architecture are described. Sections C and D present the theory of the deep neural network classification and prediction models. Finally, in Section E, the methods for evaluating the performance of the deep neural network model are discussed.

A. Heart Disease Data

Used in this research, the clinical heart disease data were from 303 patients at the Cleveland Clinic Foundation (CCF) located in Cleveland, Ohio in the United States. The dataset was obtained from the Heart Disease Database made available in the UCI Machine Learning Repository [20]. Each of the 303 clinical instances contained 75 attributes and a target attribute. The target attribute represented an integer valued from 0 to 4, signifying absence [0] or presence [1, 2, 3] of heart disease in patients. For this research, binary values of 0 and 1 were reassigned to the target attributes for the absence or presence of heart disease in patients, respectively. The dataset included 91 female patients (30.03%) and 212 male patients (69.97%), and their ages ranged from 29 to 77 years with the average being 54 years old of the 303 clinical instances from the Cleveland Clinic Dataset, 282 clinical cases were utilized and the remainder were excluded from the research due to missing data values. Of the 282 total clinical instances, 125 of the cases (44.33%) had heart disease while 157 were cases (55.67%) that were absent of heart disease. Each clinical instance was described with 76 raw attributes. However, only 29 of the raw attributes were utilized in the development of the deep neural network models due to missing values among the other raw attributes. Details regarding the 29 raw attributes are listed in Table 1. In the development of the deep neural network model, the entire data set of 282 total clinical instances was randomly separated into a training data set of 135 clinical instances (47.87%) and testing data set of 147 clinical instances (52.13%).

TABLE I. THE 29 ATTRIBUTES AND DESCRIPTIONS USED IN DEVELOPING THE DEEP NEURAL NETWORK MODEL

Variable	Attribute Description	Variable	Attribute Description
Age	Years	**Htn**	Hypertension
Sex	1= male 0 = female	**Tpeakbp**	Peak exercise blood pressure (part 2)
CP	Chest pain rating: 1 = typical angina 2 = atypical angina 3 = non-anginal pain 4 = asymptomatic	**Restecg**	Resting ECG 0 = normal, 1 = ST-T wave abnormality (> 0.05 mV), 2 = left ventricular hypertrophy
Tpeakbps	Peak exercise blood pressure (part 1)	**Tresrbp**	Blood pressure at rest (mm Hg)
Chol	Serum cholesterol (mg/dl)	**Exang**	Exercise induced angina 1 = yes, 0 = no
Ekgmo	Month of exercise ECG reading	**Lvf**	Left ventricular failure
Ekgday	Day of exercise ECG reading	**Oldpeak**	Exercise-induced ST depression
Ekgyr	Year of exercise ECG reading	**Cmo**	Cardiac cath: month
Dummy	Dummy variable	**Cday**	Cardiac cath: day
Xhypo	1 = yes, 0 = no	**Cyr**	Cardiac cath: year
Prop	Beta blocker used during exercise ECG 1 = yes, 0 = no	**Nitr**	Nitrates used during exercise ECG 1 = yes, 0 = no
Thaldur	Exercise test duration (min)	**Thalach**	Maximum heart rate achieved
Xhypo	1 = yes, 0 = no	**Thalrest**	Resting heart rate
Pro	Calcium channel blocker used during exercise ECG 1 = yes, 0 = no	**Diag**	Heart disease diagnosis: angiographic 0 (< 50% diameter) 1 (> 50% diameter)

B. Deep Neural Network System and Architecture

In this section, the deep neural network system and architecture are presented for coronary heart disease diagnosis based on the CCF dataset using deep learning algorithms, hyper-parameters, and turning and controls block.

Displayed in Figure 1 are the system and architecture of the designed deep learning models, which includes two subsystems: (1) a deep neural network training classification model, and (2) a deep neural network diagnosis model for heart disease.

The deep learning training classification model as illustrated in Figure 1 is based on a deeper multilayer perceptron employing more deeper number of hidden layers with linear and non-linear transfer functions, regularization and dropout, a sigmoid function for binary classification using deep learning technologies. An input data matrix, which included N clinical instances and R attributes of heart disease, where $R = 28$, was simultaneously fed into the deep neural network training model, which propagated all input patterns of coronary heart disease to determine all unit outputs of linear and non-linear transfer functions. A hyper-parameter turning and control block not only enabled adjustment of a set of hyper parameters but also controlled the batch size and the number of epochs during the training of the deep neural network classification model. The batch size was used to determine the input data matrix with N clinical instances, where, in this research, $N = 80$. The number of epochs in deep learning represents the number of times in which all training data pass through the learning algorithm to adjust the deep neural network weights.

While training the deep neural network classification model, comparisons of all unit outputs with the desired pattern responses of the coronary heart diseases from the target variable class were used to determine the errors. The error was further multiplied by the hyper-parameter, which was subsequently adjusted by learning algorithm block. Then, weights in the deep neural network classification model were updated after the minimization of error at each stage through the unit weight adjustment where input data and the corresponding target variable data were used to train the deep neural network model until it approximated a function within a prior defined error value. This training process was repeated until the sum of squared errors was minimized to the smallest possible below the prior defined error value, or repeated until the total number of epochs was utilized.

After completion of the training for the DNN classification model, the final weights were fed into the deep neural network prediction (also known as diagnostic) model. Then, the DNN prediction model was used to detect and diagnose coronary heart disease patterns for outcome predictions of future patients during the testing process.

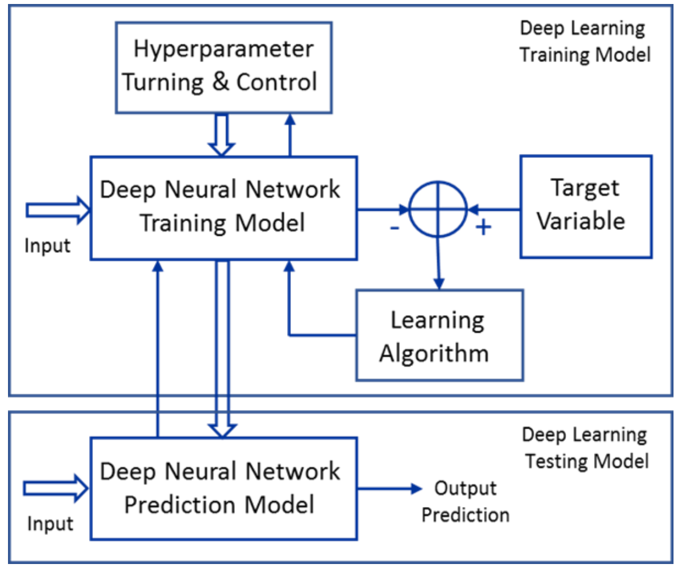

Fig. 1. System and Architecture of the Designed Deep Neural Network Models, Including the Deep Learning Training and Testing Models for Heart Disease Diagnosis used in this Research.

C. Deep Neural Network Classification Model

The deep neural network classification model is like a massive parallel distributed processor that learns and stores experiential knowledge [21]. Knowledge is acquired by the DNN system and architecture through a learning process [22], which is referred to as machine self-learning capabilities. Interneuron connection strengths, which is also known as weights in the network system and architecture, are used to store the knowledge [23, 24]. During training, a learning algorithm is utilized to modify the weights of linear and non-linear transfer functions within all of the neuron units in the network for the DNN classification model, thereby achieving a desired design objective usually in terms of the minimum mean squared error (MMSE) in an optimal sense. In other words, the learning algorithm adjusts all the weights of the DNN classification model based on the input data and output data of the target variable to obtain the best or optimal performance during each iterative process of the training sessions.

Structurally, the DNN model differs from a traditional multilayer perceptron neural network classification model. One of the key differences is regarding the network depth, which depends on the number of hidden layers in the network. Generally, if a neural network classification model has three or more hidden layers in the network architecture, it is considered a deep neural network classification model [19]. Thus, the higher hidden layers in the deep neural network build new abstractions on top of previous hidden layers. This thereby allows the DNN architecture to learn better solutions for the DNN classification model.

On the other hand, for DNN classification models, a key issue is overfitting [19, 25], which is the central problem in the field of deep learning because the DNN classification model performs effectively using a training dataset, but the model may perform less well with a test dataset using novel data [19]. This is because the DNN classification model is a complex system and architecture due to the large number of parameters based on a training dataset. However, in most cases, training data size is often not large enough. Therefore, overfitting can result in a high diagnostic accuracy based on the training dataset but a relatively less high diagnostic accuracy with the testing dataset when presented with novel cases. Thus, to prevent overfitting, the DNN classification model usually utilizes a regularization algorithm, which would decrease the complexity of the DNN model while maintaining the same number of large parameters.

The most common regularization algorithms is L_2 regularization [19, 22, 26] or *weight decay* and is a direct function approach that can regularize the DNN classification model. However, in order to minimize a L_2 norm, the regularization algorithm is used to penalize large weights by using a hyper-parameter λ, which is used to define the relative importance of the L_2 norm for decreasing loss on the training dataset. Thus, training a DNN classification model $f_\theta(x_i)$ involves finding a weight function θ (***w***, ***b)*** where ***w*** and ***b*** represent weights and bias, so that expected regularization loss is minimized and given by [19]:

$$E(\theta, D) = arg\min_\theta \left\{\frac{1}{D}\sum_{(x_i,t_i)\in D} E(f_\theta(x_i), t_i)\right\} + \lambda\|\theta\|_p, \quad (1)$$

Where D represents a training dataset and (x_i, T_i) represent samples in the training dataset; x_i represent inputs, and T_i are target variable data. The hyper-parameter λ is then utilized to control the regularization algorithm. The first function in the regularization algorithm in Eq. (1) represents an error function while the second function is known as a regularization error given by,

$$\|\theta\|_p = \left(\sum_{j=0}^{N} |\theta_j|^p\right)^{\frac{1}{p}}. \quad (2)$$

Note that Eq. (2) is the L_p norm in terms of the parameters θ and p. As can be seen, when p = 1, Eq. (2) is the L_1 regularization. Eq. (2) is the L_2 regularization when p = 2. On the other hand, the error function in Eq. (1) depends on output data of the target variable and assigns a penalty to predictions according to their consistency with the targets. In addition, the regularization error in Eq. (2) allocates a penalty depending on factors besides those of the targets' variable data.

Equally important, dropout is an effective regularization technique utilized to prevent overfitting in the DNN architecture [27]. Dropout is an alternative technology, which is used during each iterative training process to randomly eliminate neural network units and their connections in the DNN model system. The dropout technology enables us to train only smaller subsets of the deep neural network system and architecture in the deep neural network classification model [19]. As a result, dropout can effectively prevent single neural node units from dominating the deep neural network and causing over-adaptation.

D. Deep Neural Network Prediction Model

Performance quality of the deep neural network prediction (or diagnostic) model is heavily dependent on the DNN classification model during training. In this research, final weights of the deep neural network prediction model were loaded from the deep learning training model subsystem after the training processing was completed.

Generally, the DNN prediction model with (*L-1*) hidden layers has an output function [19, 24]:

$$\boldsymbol{Y} = \Phi_L((\cdots\Phi_3(\Phi_2(\Phi_1(\boldsymbol{XW}_1 + \boldsymbol{B}_1)\boldsymbol{W}_2 + \boldsymbol{B}_2)\boldsymbol{W}_3 + \boldsymbol{B}_3)\cdots)\boldsymbol{W}_L + \boldsymbol{B}_L), \quad (3)$$

Where input matrix data ***X*** is fed into the layer for input; $\boldsymbol{W}_n$ and $\boldsymbol{B}_n$, $n = 1, 2, \ldots, L$, are weight matrix and bias vectors, respectively, for one of the n hidden layers; the transfer function Φ_n, $n = 1, 2, \ldots, L$, is either linear or nonlinear. The last layer at n = L is known as an output layer, and the other layers are hidden layers in the DNN system and architecture. As a result, the DNN prediction model in Eq. (3) can be used to detect coronary heart disease in future patients with novel clinical data during a diagnostic process.

E. Evaluation Methods of Deep Neural Network Model

Performances of deep learning models in diagnoses are often evaluated using the following measures: diagnostic accuracy, misclassification error, specificity, sensitivity, precision, *F*-score, AUC, and *K-S* test [10, 19, 24, 30-34]. Thus, in order to evaluate the diagnostic performances of the deep neural network models presented in this research, the above measures, displayed in Table 2, will be used based on both the training and testing datasets, where true positive = TP, false positive = FP, true negative = TN, and false negative = FN.

TABLE II. EVALUATION METHODS AND EQUATIONS FOR PERFORMANCE OF THE DEEP NEURAL NETWORK MODEL

Evaluation Methods	**Equations**
diagnostic accuracy	(TP + TN)/(TP + FN + FP + TN)
probability of misclassification error (PME)	(FN + FP) / (TP + FN + FP + TN) where diagnostic accuracy = (1 - PME)
sensitivity (recall)	TP / (TP + FN)
specificity	TN / (FP + TN)
precision	TP / (TP + FP)
F- Score	$\frac{(1+\beta^2)(\text{Precision} \times \text{Recall})}{\beta^2 \cdot Precision + Recall}$ In this research, β = 1ıs used; *F*-score is the harmonic mean of precision and recall.
area under ROC curve (AUC)	Graphical plot of sensitivity and (1-specificity); AUC = 0.5 is random chance; AUC = 1.0 is 100% diagnostic accuracy
K-S test	Model output probabilities; *K-S* test ranges between 0% and 100%; The *K-S* test measures degree of separation of distribution of the diagnostic results; *K-S* test = 100% represents perfect separation.

Additionally, a diagnostic odds ratio (DOR) is often used to evaluate effectiveness of a diagnostic test in medical testing. The DOR can be defined in a mathematical formula given by:

$$DOR = \frac{TP/FP}{FN/TN} = \frac{Sensitivity \times Specificity}{(1-Sensitivity)\times(1-Specificity)}\,. \quad (4)$$

The logarithm of the DOR is close to a normal distribution [35]. The corresponding standard error (SE) of the ln{DOR} distribution is approximately obtained by,

$$SE\{\ln(DOR)\} = \sqrt{\frac{1}{TP}+\frac{1}{TN}+\frac{1}{FP}+\frac{1}{FN}}. \quad (5)$$

This allows us to derive an approximate 95% confidence interval of the ln{DOR} in the following:

$$\ln(DOR) \pm 1.96 \times SE\{\ln(DOR)\}\,. \quad (6)$$

Using a method of back-transformation, computing the anti-log of this expression in Eq. (6) provides the 95% confidence interval of the DOR in the following:

$$e^{\ln(DOR)\pm 1.96\times SE\{\ln(DOR)\}}. \quad (7)$$

Note that a value of the DOR in Eq. (4) usually ranges from 0 to ∞. A higher DOR value represents a more optimal performance. In order to be considered useful tests, the value of the DOR should be at least greater than one. Thus, the prediction (or diagnosis) model test is distinguishing correctly using the DNN model to identify whether or not heart disease is present in new patient cases.

III. Results

In this research, an enhanced deep neural network system and architecture are proposed for increasing the accuracy of heart disease diagnoses. The developed DNN model contains classification and prediction models, which are based on a deep multilayer perceptron with linear and non-linear transfer functions, regularization and dropout, and a binary sigmoid classification using deep learning technologies, thereby creating strong and enhanced classification and prediction models.

The proposed DNN classification model has a DNN architecture, including 28 input units, first and second hidden layers, and a binary output unit. 105 neurons in the first layer and 42 neurons in the second layer are connected with each Rectified linear unit activation functions [19, 24] along with a 50% dropout. The output unit in the final stage of the DNN architecture is connected to a sigmoid activation function [19, 24]. During training of the DNN, dropout rates in both hidden layers are randomly applied, resulting in random connections within the DNN architecture. Thus, the dropout technology is capable of reducing overfitting problems in the DNN model.

For evaluating the performances of the developed classification and prediction models based on the deep neural network algorithm, using the holdout method [36], a nonparametric approach was utilized to calculate the diagnostic accuracy and probability of misclassification error. The entire CCF dataset was randomly separated into two mutually exclusive data sets. In this research, 135 clinical instances were allocated to the training data set while 147 clinical instances were included in the testing data set. Both the datasets were normalized using the function [x – min(x)] divided by [max(x) – min(x)], in which *x* is an input data.

Training the DNN classification model using the training dataset was based on a batch size of 80, number of epochs of 5000, and learning rate of 0.00005 with a root mean square error (RMSE) in an optimization sense. Then, the designed DNN model was tested using the testing dataset.

Based on the results, Figure 2 displays the accuracy curve of a cutoff point at the output layer during testing process. As can be seen, when the cutoff point increases from 0 to 0.5, the accuracy of the designed DNN model increases; when the cutoff point increases from 0.5 to 1, the model accuracy decreases accordingly. Thus, to achieve maximum accuracy, the best cutoff point is 0.5 optimally.

In Figure 3, the ROC curve of the testing dataset is displayed, where AUC for detection of heart disease in the CCF dataset was 0.8922. AUC is utilized to rank the quality of the diagnostic model performance by graphing a curve from a series of trade-off points for sensitivity and (1-specificity) results, which were classified for presence of heart disease [30, 32]. Generally, the closer the AUC value is to 1, the higher the diagnostic accuracy in correctly detecting for the presence or absence of heart disease.

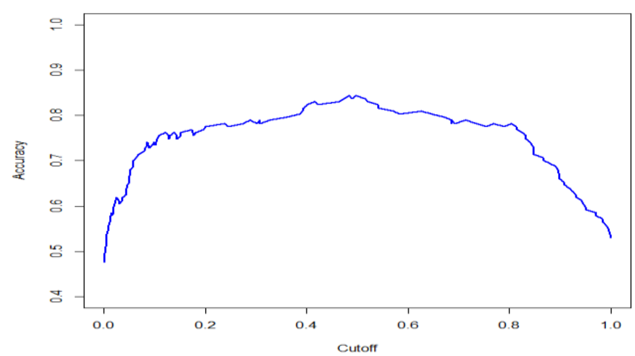

Fig. 2. An Accuracy Curve associated with a Cutoff Point at the Output Layer during the Testing. This Chart can be used to Determine the Cutoff Point in Terms of Probability, Thereby Ensuring the Maximum Value of Accuracy for the Developed DNN Model.

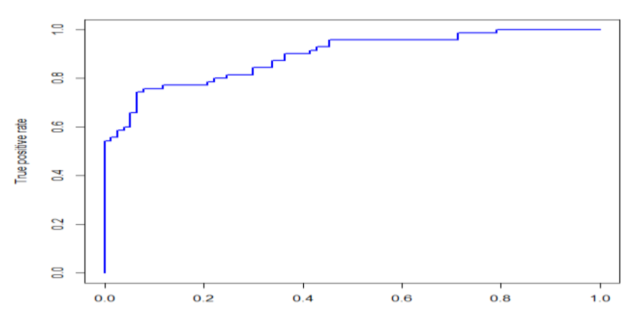

Fig. 3. In this ROC Curve, True Positive Rate Represents Sensitivity and False Positive Rate is the Difference (1 – Specificity). In this Research, AUC for the DNN Model was 0.8922 using the Testing Dataset.

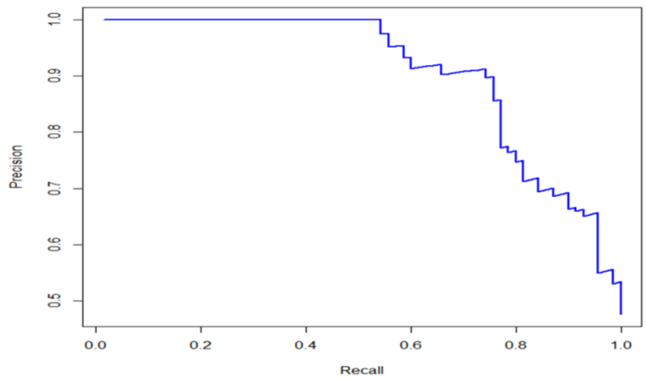

Fig. 4. This Shows a Relationship Curve between the Recall (or Sensitivity) vs. the Precision. When the Recall Increases, the Precision Decreases. When Precision Increases, Recall Decreases. This Chart can be used to Determine the Balance between Recall and Precision, Leading to *F*-Score Calculation.

Figure 4 displays a relationship curve between recall (or sensitivity) at the *x*-axis and precision at the *y*-axis, which is used to determine balance points between recall and precision as well as to calculate the *F*-score for the DNN classification and prediction model. As can be seen, when the recall increases from 0 to 0.55, the precision maintains a constant of 1; when the recall continues to increase from 0.55 to 1, the precision starts decreasing from 1 to 0 accordingly. On the other hand, when the precision increases from 0 to 1, the recall decreases from 1 to 0. Thus, in order to make ensure that the model can operate in an optimal sense, one of the common methods is to maximize a value of the *F*-score calculation using the equations displayed in Table 2.

In Figure 5, the *K-S* chart was created by using the model output probabilities of the DNN prediction model based on the testing dataset. The *K-S* test is nonparametric and measures degree of separation for the distribution of diagnostic results for heart disease in patients. With *K-S* test results ranging between 0% and 100%, a higher *K-S* test result represents a better diagnostic model in detecting disease in patients. In our research, the highest *K-S* value is 66.62%, located at the 4th decile population.

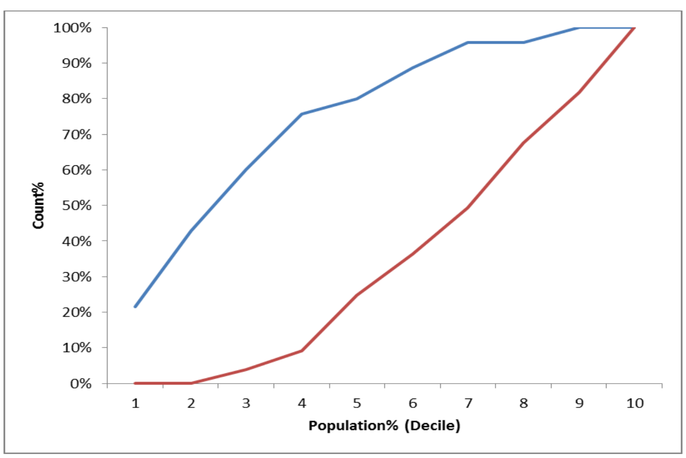

Fig. 5. The Highest *K-S* Value Achieved by the DNN Model is 66.62%, Located at the 4th Decile Population.

TABLE III. TESTING RESULTS FOR THE MODEL PERFORMANCE BASED ON THE TESTING DATA OF 147 CLINICAL INSTANCES USING THE DNN PREDICTION MODEL

	Heart Disease Presence	**Heart Disease Absence**	**Total**
Predicted Presence	TP = 51	FP = 5	56
Predicted Absence	FN = 19	TN = 72	91
Total	70	77	147

TABLE IV. TESTING RESULTS OF MODEL PERFORMANCE FOR THE TESTING DATA SET OF 147 CLINICAL CASES USING THE DEEP NEURAL NETWORK PREDICTION MODEL

Evaluation Methods	**Testing Results**
Diagnostic Accuracy	83.67%
Probability of Misclassification Error	16.33%
Sensitivity	93.51%
Specificity	72.86%
Precision	79.12%
F-Score	0.8571
Area under the ROC curve (AUC)	0.8922
K-S Test (highest at the 4th decile)	66.62%
Diagnostic Odds Ratio (DOR)	38.65
95% confidence interval of the DOR	[13.55, 110.28]

Displayed in Table 3 are the testing results for the performance of the DNN models. In Table 4, testing results of the model performances for the CCF testing data set of 147 clinical cases were calculated based on the above values, resulting in the diagnostic accuracy of 83.67%, probability of misclassification error of 16.33%, sensitivity of 93.51%, specificity of 72.86%, precision of 79.12%, *F*-score of 0.8571, AUC of 0.8922, the highest *K-S* test of 66.62% at the 4th decile population, DOR of 38.65, and 95% confidence interval for the DOR of this test of [38.65, 110.28].

Based on the testing results, medical application of the deep neural network learning model can be reliably and clinically useful in diagnosing patients with chest pain and ranges of heart disease presentation with high accuracy. The results can also be used to aid both healthcare professionals and patients, especially those in low-resource areas and developing countries where there are fewer cardiac specialists.

IV. Discussion

In this research, the deep neural network classification and prediction models were created based on a deep learning algorithm. The DNN models were used to diagnose coronary heart disease and were applied to dataset of 303 clinical instances from the Cleveland Clinic Foundation. The models were trained and tested using randomly generated training and testing datasets, respectively. The performances of the developed DNN models were evaluated using diagnostic accuracy, probability of misclassification error, specificity, precision, AUC, sensitivity, *F*-score, and *K-S* test.

Displayed in Table 4, testing results of the developed DNN model performances for the CCF testing data set of 147 clinical cases were calculated based on a set of measurement parameters, including diagnostic accuracy of 83.67%, probability of misclassification error of 16.33%, sensitivity of 93.51%, specificity of 72.86%, precision of 79.12%, *F*-score of 0.8571, AUC of 0.8922, *K-S* test of 66.62%, DOR of 38.65, and 95% confidence interval for the DOR of this test of [38.65, 110.28]. Therefore, based on these testing results, the developed deep neural network classification and prediction models would be 83.67% accurate in diagnosing coronary heart disease in a new patient. The high sensitivity result of 93.51% is essential because it represents the probability of positive test result detection among those with the heart disease, and this indicates that given a new patient comes to the clinic with undiagnosed heart disease, the model is 93.51% accurate in detecting and making an accurate diagnosis. Since early and accurate detection of heart disease is essential for early intervention and prolonging chances of long-term survival, this high sensitivity score along with the relatively high *F*-score of 0.8571 and AUC of 0.8922 demonstrate that the developed DNN models were able to achieve a high accuracy in coronary heart disease diagnosis in patients.

Compared to related papers, several different methods were developed using the CCF heart disease data set. These included decision trees [6], SVM learning [8], a new type of discriminant function model using Bayesian algorithm [15], Bagging approach [37], and ensemble machine learning [10]. However, the majority of the methods associated with these developed models only used 13 input attributes compared to the 28 input attributes used in this research, enhancing the advanced deep neural network. These previous methods had model performance accuracies within a range of 61.93% to 80.14% using CCF clinical data. Furthermore, the sensitivity for detecting heart disease in patients using the above methods ranged from 70.97% to 77.9%. Our research using the deep neural network method exceeded both diagnostic accuracy and sensitivity of the above approaches with results of 83.67% and 93.51%, respectively. Furthermore, in the previously published method using ensemble machine learning [10], which also used 28 input attributes for developing the model, had sensitivity value of 70.97%, *F*-Score of 0.76, probability of misclassification error of 19.86%, and accuracy of 80.14% with *K-S* test value of 58.66%. In comparison, the developed deep neural network learning model in this research achieved a higher value of sensitivity at 93.51%, lower probability of misclassification error of 16.33%, higher diagnostic accuracy of 83.67%, higher *F*-Score of 0.8571, and further enhanced the *K-S* test value by an increase of 13.57% to 66.62%.

Our testing results indicate that the accuracy of the developed deep neural network classification and prediction models is relatively higher than most of those of previously published methods. Our testing results also used a relatively larger testing dataset of 147 novel clinical instances than the training dataset size of 131 clinical instances compared to the other methods using the holdout method. This suggests that if given a new patient with unique clinical data, the developed model would be able to diagnose with high reliability and accuracy. Therefore, the deep neural network learning models would be effective in reducing the number of inaccurate diagnoses, subsequently prevent erroneous treatments, and thereby enhance the health of patients. Furthermore, the developed DNN models provided a more reliable and greater diagnostic accuracy in detecting coronary heart disease in patients and improving chances of long-term survival.

V. Conclusion and Future Work

In this paper, the deep neural network learning classification and prediction models were developed and evaluated based on diagnostic performance of coronary heart disease in patients using sensitivity, specificity, precision, *F*-score, AUC, DOR, 95% confidence interval for DOR, and *K-S* test. The developed deep learning classification and prediction models were built with a deep multilayer perceptron eqipped with linear and non-linear transfer functions, regularization and dropout, and a binary sigmoid classification using deep learning technologies to create a strong and enhanced classification and prediction model.

The developed deep neural network classification and prediction models were trained and tested using the holdout method and 28 input attributes based on the clinical dataset from patients at the Cleveland Clinic. Based on the testing results, the developed deep learning models achieved diagnostic accuracy for heart disease of 83.67%, probability of misclassification error of 16.33%, sensitivity of 93.51%, specificity of 72.86%, precision of 79.12%, *F*-score of 0.8571, AUC of 0.8922, the *K-S* test of 66.62%, DOR of 38.65, and 95% confidence interval for the DOR of this test of [38.65, 110.28]. These results exceed those of currently published research. Therefore, the developed deep learning classification and prediction models can provide highly reliable and accurate diagnoses for coronary heart disease and reduce the number of erroneous diagnoses that potentially harm patients. Thus, the models can be used to aid healthcare professionals and patients throughout the world to advance both public health and global health, especially in developing countries and resource-limited areas where there are fewer cardiac specialists available.

In future research, we would investigate other enhanced methods that would further raise the diagnostic accuracy of the deep learning model by utilizing deep learning based on morphologic class pattern predictions [19] in order to further enhance the performances of the DNN models for heart disease diagnoses in patients worldwide. Furthermore, current advances in deep learning, including recurrent neural network, deep convolutional neural network, long short-term memory neural network, deep brief network based on restricted Boltzmann machines [38], and deep auto-encoder [39] may be utilized to further increase the accuracy of heart disease diagnoses for patients [40].

Acknowledgment

The authors would like to thank Robert Dotrano, MD, PhD, from the Cleveland Clinic Foundation and V.A. Medical Center, Long Beach for contributing the clinical heart disease datasets, which are available at the Heart Disease Databases in the UCI Machine Learning Repository. The data included clinical information about heart disease diagnoses in patients.

REFERENCES

[1] World Health Organization (WHO), "Cardiovascular diseases (CVD)," http://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), 2017.

[2] Centers for Disease Control and Prevention (CDC), "Heart disease fact sheet,"https://www.cdc.gov/dhdsp/data_statistics/fact_sheets/fs_heart_disease.htm, 2015.

[3] E. J. Benjamin, S. S. Virani, C. W. Callaway, A. R. Chang, S. Cheng, S. E. Chiuve, M. Cushman, and F. N. Delling, et. al., "Heart disease and stroke statistics-2018 update: a report from the American Heart Association," *Circulation,* Vol. 137, No. 12, pp. 67- 492, 2018.

[4] Centers for Disease Control and Prevention, "Facts about heart disease," http://www.cdc.gov/heartdisease/docs/consumered_heartdisease.pdf, 2013.

[5] E. G. Nabel and E. Braunwald, "A tale of coronary artery disease and myocardial infarction," *New England Journal of Medicine,* Vol. 366, pp. 54-63, 2012.

[6] M. Shouman, T. Turner, and R. Stocker, "Using decision tree for diagnosing heart disease patients," *Proceedings of the 9th Australasian Data Mining Conference*, pp. 23–29, Ballarat, Australia, 2011.

[7] O. Y. Atkov, S. G. Gorokhova, A. G. Sboev, E. V. Generozov, E. V. Muraseyeva, S. Y. Moroshkina, and N. N. Cherniy, "Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters," *Journal of Cardiology*, Vol. 59, Issue 2, pp. 190-194, 2012.

[8] A. D. Dolatabadi, S. E. Khadem, B. M. Asl, "Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM," *Computer Methods and Programs in Biomedicine,* Vol. 138, Issue C, pp. 117-126, 2017.

[9] 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, 2013.

[10] K. H. Miao, J. H. Miao, and G. J. Miao, "Diagnosing heart disease using ensemble machine learning," *International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 10, pp. 30-39, 2016.

[11] H. M. Sawbwa, N. Ghali, and A. E. Hassanien, "Detection of heart disease using binary particle swarm optimization," *Proceedings of the Federated Conference on Computer Science and Information Systems*, pp. 177-182, 2012.

[12] A. Ozcift and A. Gulten, "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.

[13] 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*, Vol. 7, pp. 263-283, 2010.

[14] W. Wiharto, H. Kusnanto, and H. Herianto, "Intelligence system for diagnosis level of coronary heart disease with K-star algorithm," *Healthcare Informatics Research*, Vol. 22, No. 1, pp. 30-38, 2016.

[15] R. Detrano, A. Janosi, A. Steinbrunn, W. Pfisterer, M. Schmid, J. Sandhu, S. Guppy, K. Lee, and V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease," *American Journal of Cardiology*, Vol. 64, pp. 304-310, 1989.

[16] D. Pal, K. Mandana, S. Pal, D. Sarkar, and C. Chakraborty, "Fuzzy expert system approach for coronary artery disease screening using clinical parameters," *Knowledge-Based System*, Vol. 36, pp. 162-174, 2012.

[17] H. R. Marateb and S. Goudarzi, "A noninvasive method for coronary artery diseases diagnosis using a clinically-interpretable fuzzy rule based system," *Journal of Research in Medical Sciences*, Vol. 20, No. 3, pp. 214-223, 2015.

[18] J. Schmidhuber, "Deep learning in neural networks: An Overview." *Neural Networks*, Vol. 61, pp. 85-117, 2015.

[19] J. H. Miao and K. H. Miao, "Cardiotocographic diagnosis of fetal health based on multiclass morphologic pattern predictions using deep learning classification," *International Journal of Advanced Computer Science and Applications*, Vol. 9, No. 5, pp. 1-11, 2018.

[20] D. J. Newman, S. Hettich, C. L. Blake, and C. J. Merz, *UCI Repository of Machine Learning Databases*, University California Irvine, Department of Information and Computer Science, 2018.

[21] Y. Bengio, "Learning deep architectures for AI", *Foundations and Trends in Machine Learning*, Vol. 2, No. 1, pp 1-127, 2009.

[22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, The MIT Press, 2016.

[23] F. Q. Lauzon, "An introduction to deep learning," The *11th International Conference on Information Science, Signal Processing and their Applications*, Montreal, QC, pp. 1438-1439, 2012.

[24] G. J. Miao, K. H. Miao, and J. H. Miao, "Neural pattern recognition model for breast cancer diagnosis," *Journal of Selected Areas in Bioinformatics*, Vol. 2, Issue 8, pp. 1–8, 2012.

[25] C. Gavin, and T. L. C. Nicola, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *Journal of Machine Learning Research*, Vol. 11, pp. 2079−2107, 2010.

[26] J. Kukačka, V. Golkov, and D. Cremers, "Regularization for deep learning: a taxonomy," *Artificial Intelligence*, pp 1-27, 2017.

[27] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, Vol. 15, pp. 1929-1958, 2014.

[28] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *Neural and Evolutionary Computing*, pp. 1-18, 2012.

[29] D. Warde-Farley, I. J. Goodfellow, A. Courville, and Y. Bengio, "An empirical analysis of dropout in piecewise linear networks," *Machine Learning*, pp. 1-10, 2014.

[30] 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*, Vol. 3, Issue 9, pp. 1–10, 2013.

[31] 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.

[32] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation.," A. Sattar and B. Kang (eds), Advances in Artificial Intelligence, Lecture Notes in Computer Science, Vol. 4304, Berlin, Heidelberg, 2006.

[33] A. M. Law, "A tutorial on how to select simulation input probability distributions," *The IEEE Proceedings of the 2013 Winter Simulation Conference*, pp. 306–320, 2013.

[34] 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, 2011.

[35] A. S. Glas, J. G. Lijmer, M. H. Prins, G. J. Bonsel, and P. M. M. Bossuyt, "The diagnostic odds ratio: a single indicator of test performance". *Journal of Clinical Epidemiology*, Vol. 56, No. 11, pp. 1129–1135, 2003.

[36] G. J. Miao and M. A. Clements, *Digital Signal Processing and Statistical Classification*, Artech House, Inc., 2002.

[37] M. C. Tu and D. Shin, "Effective diagnosis of heart disease through Bagging approach," *The IEEE 2nd International Conference on Biomedical Engineering and Informatics*, pp. 1-4, 2009.

[38] G. Hinton, "A practical guide to training Restricted Boltzmann Machines," *University of Toronto, Department of Computer Science*, Version 1, pp. 1-21, 2010.

[39] Q. V. Le, "A tutorial on deep learning part 2: autoencoders, convolutional neural networks and recurrent neural networks," *Google Inc*, Mountain View, California, pp. 1-20, 2015.

[40] Y. LeCun, Y. Bengio, and G. Hinton. "Deep Learning" *Nature*, Vol. 521, pp 436-444, 2015.