IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 7, 2019, accepted February 26, 2019, date of publication March 13, 2019, date of current version April 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2904800

An Automated Diagnostic System for Heart Disease Prediction Based on χ^2 Statistical Model and Optimally Configured Deep Neural Network

LIAQAT ALI[1,2], ATIQUR RAHMAN[3], AURANGZEB KHAN[3], MINGYI ZHOU[1], ASHIR JAVEED[4], AND JAVED ALI KHAN[5,6]

[1]School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]Department of Electrical Engineering, University of Science and Technology Bannu, Bannu 28100, Pakistan
[3]Department of Computer Science, University of Science and Technology Bannu, Bannu 28100, Pakistan
[4]School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[5]School of Software Engineering, Tsinghua University, Beijing 100084, China
[6]Department of Software Engineering, University of Science and Technology Bannu, Bannu 28100, Pakistan

Corresponding author: Liaqat Ali (engr_liaqat183@yahoo.com)

ABSTRACT Different automated decision support systems based on artificial neural network (ANN) have been widely proposed for the detection of heart disease in previous studies. However, most of these techniques focus on the preprocessing of features only. In this paper, we focus on both, i.e., refinement of features and elimination of the problems posed by the predictive model, i.e., the problems of underfitting and overfitting. By avoiding the model from overfitting and underfitting, it can show good performance on both the datasets, i.e., training data and testing data. Inappropriate network configuration and irrelevant features often result in overfitting the training data. To eliminate irrelevant features, we propose to use χ^2 statistical model while the optimally configured deep neural network (DNN) is searched by using exhaustive search strategy. The strength of the proposed hybrid model named χ^2-DNN is evaluated by comparing its performance with conventional ANN and DNN models, another state of the art machine learning models and previously reported methods for heart disease prediction. The proposed model achieves the prediction accuracy of 93.33%. The obtained results are promising compared to the previously reported methods. The findings of the study suggest that the proposed diagnostic system can be used by physicians to accurately predict heart disease.

INDEX TERMS Deep neural network, heart disease, hyperparameters optimization, overfitting, underfitting.

I. INTRODUCTION

A. INTRODUCTION

Any disorder that affects the heart ability to function normal is called heart disease. Narrowing or blockage of the coronary arteries, which supply blood to the heart itself, is considered to be the most common cause of heart failure. Heart disease is also known as cardiovascular disease (CVD). It is the most widespread cause of death world over. Coronary Artery Disease (CAD) is the most common form of heart disease and the leading cause of heart attacks in the United States [1], [2]. Patients with CAD has at least one of the left anterior descending (LAD), left circumflex (LCX), and right coronary arteries(RCA) stenotic [3].

Clinical research has pointed out different factors that increase the risk of CAD and heart attack. These factors can be categorized into two types i.e. risk factors that cannot be changed and those that can be changed. Sex, age and family history are those factors that cannot be changed while factors that are related to lifestyle of a subject e.g., smoking, high cholesterol, high blood pressure and physical inactivity can be changed. The later are risk factors that can be modified and in some cases can be eliminated with lifestyle changes and medication [2].

Currently, the most commonly used method for diagnosis of CAD by physicians is angiography which is considered the most precise method [1]. But, it has major side effects

The associate editor coordinating the review of this manuscript and approving it for publication was Ganesh Naik.

and high cost is associated with it. Moreover, analyzing too many factors, as discussed above, to diagnose a patient, makes the physician's job difficult. These problems provoke the need for the development of non-invasive methods for heart disease detection. Furthermore, conventional methods for the diagnosis of heart disease are mainly based on analysis of patients medical history, review of relevant symptoms by a medical practitioner and physical examination report. Hence, these methods often lead to imprecise diagnosis due to human errors [4], [5]. Thus, there is a need of development of an automated diagnostic system based on machine learning for heart disease diagnosis that can resolve these problems.

In the last decade, different hybrid diagnostic systems based on features preprocessing and ANN have been developed with an objective to improve the classification accuracy. These diagnostic systems have improved the quality of decision making during diagnosis of patients by physicians. The study of these automated diagnostic systems also motivated us to develop an automated diagnostic system based on χ^2 statistical model and DNN for the improved diagnosis of heart disease.

Our proposed method targets two main problems i.e., the problem of underfitting and overfitting. The main objective is to develop such a diagnostic system that neither underfits nor overfits to the training data. Generally, the goal of every predictive model (in supervised learning) is to generate a hypothesis i.e., a fitting function from the training instances. The hypothesis if does not fit well the training data, the model is underfitted. Thus, the model is expected to show poor performance on both training and testing data. On the other hand if the hypothesis does fit very accurately (more than necessary) the training data but it does not perform well on the testing data, it means the model is overfitted [24], [25]. This may be due to learning irrelevant or noisy features in the training data or due to inappropriate network configuration with too many parameters. Thus, we need to search optimal subset of features and optimally configured neural network. Hence, to eliminate irrelevant or noisy features from feature space, we propose to use χ^2 statistical model. And to avoid the network from overfitting and underfitting, we search optimal configuration of the network by using exhaustive grid search strategy. The working of the proposed model is more clearly depicted in FIGURE 1.

In FIGURE 1, the preprocessing block denotes the normalization process performed on feature vectors. After normalization, data is partitioned into training and testing sets. To avoid bias, feature ranking and selection is performed on training data using χ^2 statistical model. The same subset of features is used for testing data which is selected in training data by the *chi*^{2} model. The training data with reduced features is applied to neural network for training purpose. The performance of the trained neural network is evaluated using testing data.

The main contributions of this study are summarized as follows:

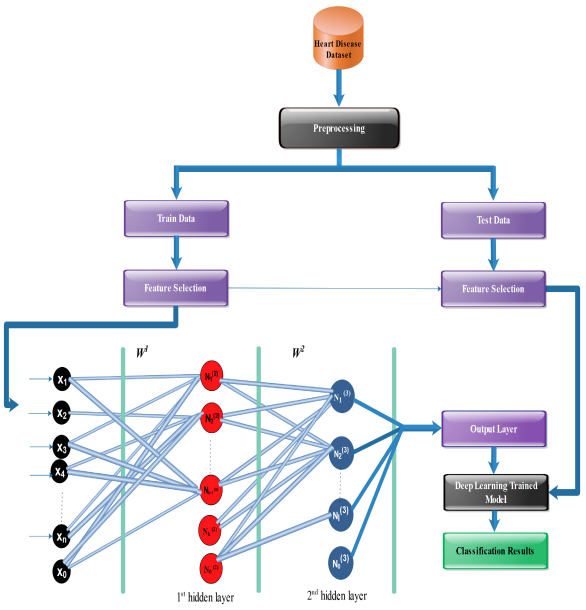

FIGURE 1. Block diagram of the proposed diagnostic system.

1. An intelligent hybrid system named χ^2-DNN has been developed to improve the classification accuracy of heart disease detection.
2. In this paper, χ^2-DNN model is developed and investigated for the first time for heart disease detection problem.
3. This paper investigates the impact of depth of neural network on the accuracy. Generally, it is perceived that shallow neural network performs better than DNN on small scale datasets. But, in this paper, we show that DNN with multiple hidden layers can show good performance than ANN on heart disease dataset provided irrelevant features are eliminated and the network is carefully tuned i.e., not made too deep.

B. STATE OF THE ART

Literatures show that different automated diagnostic systems based on machine learning models such as naive Bayes (NB), k-nearest neighbour (KNN), support vector machine (SVM), fuzzy logic, artificial neural network (ANN) and ensembles of ANN have been proposed in previous studies for heart disease predictions [2], [6]–[19]. A detailed survey of these studies shows that ANN-based methods have been widely adopted in medical diagnosis due to their capability in handling complex linear and non-linear problems [20]. Most of the studies which applied ANN for heart disease detection used Levenberg Marquardt (LM), scaled conjugate gradient (SCG) and Pola-Ribiere conjugate gradient (CGP) algorithms for learning the values or weights of parameters from training data. However, in this study we used recently proposed optimization algorithms known as IBFGS and Adam [21]. Moreover, the earlier studies used ANN which is a neural network with only one hidden layer while

TABLE 1. Description of features of the dataset.

Feature No	Feature Description	Feature Abbreviations	Features Code
1	Age (Years)	AGE	F_1
2	Sex	SEX	F_2
3	Chest Pain Type	CPT	F_3
4	Resting Blood Pressure	RBP	F_4
5	Serum Cholesterol	SCH	F_5
6	Fasting Blood Sugar	FBS	F_6
7	Resting Electrocardiographic Results	RES	F_7
8	Maximum Heart Rate achieved	MHR	F_8
9	Exercise Induced Angina	EIA	F_9
10	Old Peak	OPK	F_{10}
11	Peak Exercise Slope	PES	F_{11}
12	Number of Major Vessels Colored by Fluoroscopy	VCA	F_{12}
13	Thallium Scan	THA	F_{13}

in this paper we used a deep neural network with more than one hidden layer. Deep neural networks are neural networks that use multiple hidden layers and are trained using new methods [22].

Recently, Resul *et al.* proposed an ensemble of neural networks model for the diagnosis of heart disease and achieved 89.01% accuracy, 80.95% sensitivity, and 95.91% specificity [2]. Samuel *et al.* developed a novel hybrid system based on ANN and fuzzy analytic hierarchy process i.e., Fuzzy-AHP technique for the diagnosis of heart disease [20]. The ANN and Fuzzy-AHP based system achieved prediction accuracy of 91.10%. Most recently, Paul *et al.* checked the feasibility of adaptive weighted fuzzy system ensemble method for heart disease detection problem and achieved classification accuracy of 92.31% [23].

The rest of the paper is organized as follows: In section II the dataset and the proposed methods are described. Section III deals with evaluation metrics and validation schemes. While section IV is about experimental results and discussion. The last section is conclusion and future work.

II. MATERIALS AND METHODS

A. DATASET DESCRIPTION

In this paper, an online heart disease dataset, known as Cleveland heart disease dataset, was used which is online available on UCI machine learning repository. The dataset contains 303 instances out of which 297 instances have no missing data while six instances have missing attributes. The dataset has 76 raw features in its original form. However, all the published experiments only refer to 13 of them, and these 13 features are tabulated in TABLE 1. Additionally, only those instances have been used in the experiments that have no missing value as most of the published work have used the same practice. As an example two samples, one belonging to patient and another belonging to healthy subject, are depicted in FIGURE 2.

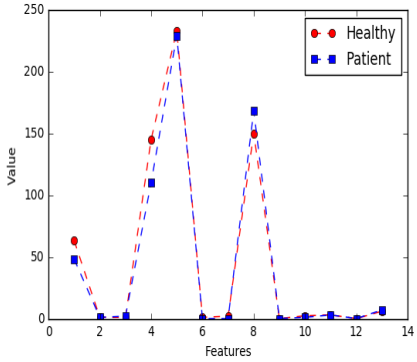

FIGURE 2. Samples of a patient and a healthy subject.

B. PROBLEM FORMULATION AND PROPOSED SOLUTION

In machine learning, the main goal of a predictive model is to generate a hypothesis i.e., $h_\beta(x)$ by using a learning algorithm (or optimization algorithm) on the training data. That is the model learns a fitting function by analyzing the behavior of training data. The hypothesis is generated by minimizing the error achieved on all the training instances.

In general, the hypothesis can be represented as

$$h_\beta(x) = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots\ldots + \beta_n x_n \quad (1)$$

The behavior of a neural network is dependent on the capacity of the network. Capacity of a network is linked with the number of parameters used in hypothesis. Model capacity more than required i.e., excessive number of parameters in hypothesis will lead to overfitting while model with less capacity than required i.e., insufficient number of parameters will result in underfitting. Thus, we need to search a model that generates a hypothesis with optimum capacity i.e., optimum number of parameters or optimum hypothesis. To link the number of parameters with the neural network configuration, it is important to understand the formulation of neural network. For a neural network model, the hypothesis can be formulated using forward propagation process as follows:

Every neuron in the network takes some inputs, applies some process on them and generates an output. The process is described by its activation function. In the below given formulation, a_i^j denotes activation or output of i^{th} neuron present in j^{th} layer while f and g denote the activation functions of the neurons. The activations of the neurons present at 2^{nd} layer i.e., hidden layer of the network are formulated as follows:

$$a_1^2 = f(\beta_{10}^1 x_0 + \beta_{11}^1 x_1 + \ldots\ldots\ldots\ldots\ldots + \beta_{1n}^1 x_n) \quad (2)$$

$$a_2^2 = f(\beta_{20}^1 x_0 + \beta_{21}^1 x_1 + \ldots\ldots\ldots\ldots\ldots + \beta_{2n}^1 x_n) \quad (3)$$

$$\vdots$$

$$a_n^2 = f(\beta_{n0}^1 x_0 + \beta_{n1}^1 x_1 + \ldots\ldots\ldots\ldots\ldots + \beta_{nn}^1 x_n) \quad (4)$$

These activations act as input to the neurons at the third layer. If the network use only one hidden layer then the third layer is the output layer and it will have only one neuron when dealing with binary classification. The output neuron applies some process, described by its activation function, to its input and generates the predicted label for the input feature vector applied to the network. The activation of the output neuron is formulated as follows:

$$h_\beta(x) = a_1^3 = g(\beta_{10}^2 a_0^2 + \beta_{11}^2 a_1^2 + \ldots\ldots\ldots\ldots\ldots\ldots + \beta_{1n}^2 a_n^2) \tag{5}$$

In the above equations, β_{ij}^1 represent the parameters or weights associated with inputs of the neurons present in the hidden layer and β_{ij}^2 represent the parameters associated with inputs of the output neuron while x denotes the feature vector of a subject. From (5), it is clear that the performance of neural network is dependent on the input features i.e., x and the configuration of the network. Thus, we need to find out optimal subset of features and an optimally configured network. The problem of avoiding our model from underfitting and overfitting, and searching optimally configured model is an optimization problem. In this paper, we use exhaustive search for searching optimal subset of features and optimized hyperparameters of the neural network.

It is important to note that DNN itself is capable of features extraction. However, feature extraction is different from features selection. In feature extraction, all the features present in feature space are processed and new features are extracted from them. While in feature selection, irrelevant features are eliminated and the most important or relevant features are applied to the network. Thus, if we supply all features to DNN, some features may be noisy and if they are learned in the training process, they may degrade generalization of the network although the network will show good performance on the training data. That is why large number of features are also considered one of the main causes of overfitting. Thus, searching out optimal subset of features by eliminating noisy features can help DNN to show good performance on both training and testing data.

In this paper, we use χ^2 statistical model to eliminate irrelevant features. In the features elimination process, we compute χ^2 statistics between each non-negative feature F_i and class i.e., y. The χ^2 model performs χ^2 test that measures dependence between the features and class. Hence, the model is capable of eliminating those features which are more likely to be independent of class. Because, these features can be regarded as irrelevant for classification. In the first step based on the χ^2 test score, features are ranked. Next, we search n number of optimal features from the ranked features. For more details on feature selection and discretization using χ^2 statistics, readers are refereed to [26]. The main process of features ranking based on χ^2 test is formulated as follows:

TABLE 2. Table for calculating χ^2 score.

	Positive Class	Negative Class	Total
Feature F occurs	a	b	$a+b=m$
Feature F does not occur	c	d	$g+h=t-m$
Total	a+c = p	b+d = t-p	t

For heart disease binary classification problem, let there are t instances and two classes i.e., positive and negative. To calculate χ^2 test score, we can build TABLE 2.

where m denotes the number of instances that contain feature F, $t-m$ denotes the number of instances that do not contain feature F, p denotes the number of positive instances and $t-p$ denotes the number of negative instances.

χ^2 test is actually the measure of how much the expected count i.e., E and the observed count i.e., O derivate from each other. Let a, b, c and d denote the observed values, and E_a, E_b, E_c and E_d denote the expected values then the expected values based on the null hypothesis that the two events are independent can be calculated as

$$E_a = (a+b)\frac{a+b}{t} \tag{6}$$

Similar to (6), E_b, E_c and E_d can also be calculated. Using the general formulation of χ^2 test, we have

$$\chi^2 = \frac{1}{d}\sum_{k=1}^{n}\frac{(O_k - E_k)^2}{E_k} \tag{7}$$

$$\chi^2 = \frac{(a-E_a)^2}{E_a} + \frac{(b-E_b)^2}{E_b} + \frac{(c-E_c)^2}{E_c} + \frac{(d-E_d)^2}{E_d} \tag{8}$$

After simplification, (8) gets the form

$$\chi^2 = \frac{t(at-mp)^2}{pm(t-p)(t-m)} \tag{9}$$

After feature ranking based on (9), in next step we need to decide a threshold for the number of features denoted by n i.e., what optimum number of features with highest χ^2 test score to use. In this paper, we search the optimal number of features by using an exhaustive search. During the search process, we start with subset of features with $n=1$ i.e., only one feature with highest χ^2 test score. The subset of features is applied to DNN and performance of DNN is evaluated by searching its optimized hyperparameters through grid search algorithm. The results for the subset of features are saved. In second iteration, another subset of features with $n=2$ is constructed by selecting the first two features with highest χ^2 score. The subset of features is applied and neural network configuration is optimized for that subset of features. The results for this subset of features are also saved. The same process is repeated until all the ranked features are added into the subset of features. Finally, that subset of features is declared as optimal which gives best performance results.

III. VALIDATION SCHEMES AND EVALUATION METRICS

A. VALIDATION SCHEMES

In this paper, we perform experiments using train-test holdout validation scheme. The 70-30 train-test data partioning scheme is used. That is, we split our dataset into two parts using 70-30% split following the approach of Das *et al.* in [2] and Paul *et al.* in [23]. Thus, the model is trained on 207 samples and tested on the remaining 90 samples. The main reason for using the same 70-30% data partitioning is to better compare our proposed model with the previous studies conducted on the heart disease dataset.

B. EVALUATION METRICS

To evaluate the performance of the proposed predictive model, different evaluation metrics including accuracy, sensitivity, specificity and Matthews correlation coefficient (*MCC*) are used. Accuracy is the percentage of correctly classified subjects in the test dataset. Sensitivity conveys information about the percentage of correctly classified patients while specificity conveys information about correctly classified healthy subjects. The formulation of these evaluation metrics is given as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

where TP denotes number of true positives, FP denotes number of false positives, TN denotes number of true negatives and FN denotes number of false negatives.

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

$$Specificity = \frac{TN}{TN + FP} \tag{12}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{13}$$

MCC is used in statistical analysis of binary classification. It is a measure of test's accuracy. *MCC* returns value between -1 and 1 where 1 indicates perfect predictions and -1 means worst predictions.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In order to evaluate the effectiveness of the proposed method rigorously, four different types of experiments are performed on the heart disease dataset. All computations were performed on Intel (R) Core (TM) i3-2330M CPU @2.20GHz with 64bit windows 7 as the operating system. All the experiments are performed using Python software package.

A. EXPERIMENT NO 1: RESULTS OF CONVENTIONAL DNN AND THE PROPOSED χ^2-DNN MODEL

In this experiment, the proposed predictive model is trained on 207 samples and tested on the remaining 90 samples. The training samples are applied to χ^2 statistical model to generate refine features of better quality. Next, the same features are also selected in the testing data. The testing data samples are applied to the trained DNN model. We achieved classification accuracy of 93.33% on 11 features using neural network with two hidden layers. The number of neurons used in the first layer are 50 while 2 neurons were used in the second layer. This optimized configuration is searched by using exhaustive grid search algorithm. Moreover, all the neurons use relu activation function while the neuron at the output layer uses sigmoid activation function. In all the three experiments that uses 70-30 holdout validation, to optimize the weights or parameters of the neural network during training process, IBFGS algorithm is used with $\alpha = 0.00001$. Moreover, the optimal subset of features includes $F_1, F_2, F_3, F_4, F_7, F_8, F_9, F_{10}, F_{11}, F_{12}$ and F_{13}.

To validate the fact that the proposed method improves the performance of conventional DNN, we also simulated DNN model using full features set by optimizing it using the same search algorithm. The experimental results revealed that best accuracy of 90% could be achieved using DNN on full features. The optimized configuration of DNN contains 2 neurons in the first hidden layer and 4 neurons in the second hidden layer. Additionally, 91.83%, 87.80% and 0.798 score was achieved for specificity, sensitivity and MCC, respectively. From the experimental results, it is evidently clear that the proposed χ^2 model for the feature selection improves the performance of conventional DNN model by 3.33%.

To further evaluate the effectiveness of the proposed hybrid model, ROC charts and area under the curve (AUC) are utilized. ROC chart is used to investigate the classifier output quality. ROC chart is basically the plot of true positive rate on the Y axis and true negative rate on the X axis. It means that an ROC curve in the top left corner of the plot is an ideal one. Because such ROC chart indicates a true positive rate of 1 and a false positive rate of 0. In other way, we can say that an ROC chart with larger area under the curve (AUC) is better [27]. The ROC charts for the proposed χ^2-DNN model and conventional DNN model are given in Figure 3 (a) and Figure 3 (b), respectively. As depicted in the figure, the AUC for the proposed hybrid model is 0.940 while the AUC for conventional DNN is 0.934. Hence, from the ROC charts, it is clear that the proposed χ^2-DNN model shows better performance than conventional DNN model owing to its more area under the curve.

B. EXPERIMENT NO 2: RESULTS OF CONVENTIONAL ANN AND χ^2-ANN MODEL

In this experiment, under the same data partitioning scheme, the feasibility of a conventional ANN model and ANN model hybridized with χ^2 statistical model is checked. It is important to note that the term ANN is used in literature for neural network model with only one hidden layer. In case of ANN there is no concept of depth i.e., we cannot add more layers during optimization or tuning process. The model is optimized by tunning its width i.e., by increasing or decreasing the number of neurons in the hidden layer of the network. In the first step ANN was optimized using the same exhaustive search strategy on all features of the dataset.

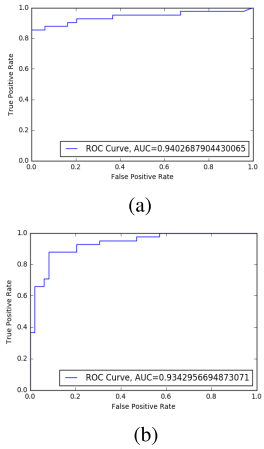

FIGURE 3. ROC charts of conventional DNN and χ^2-DNN. (a) ROC chart obtained for the proposed χ^2-DNN model. (b) ROC chart obtained for the conventional DNN model.

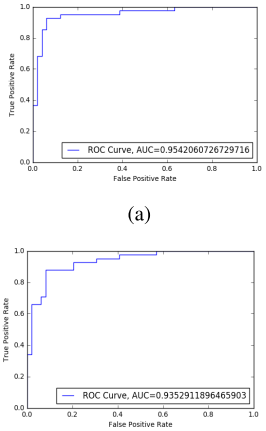

FIGURE 4. ROC charts of conventional ANN and χ^2-ANN. (a) ROC chart obtained for the χ^2-ANN model. (b) ROC chart obtained for the conventional ANN model.

The ANN model could achieve 90% accuracy, 91.83% specificity, 87.80% sensitivity and MCC value of 0.798. Furthermore, the feasibility of the proposed two stage method was also checked by replacing the DNN model by ANN i.e., χ^2-ANN was designed and applied to the heart disease data. The results show an accuracy rate of 91.11% on two types of subsets, one with $n = 9$ i.e., 9 features and other with $n = 11$ i.e., 11 features. The optimal subset of features with $n = 9$ includes $F_2, F_3, F_7, F_8, F_9, F_{10}, F_{11}, F_{12}$ and F_{13} while the optimal subset of features for $n = 11$ includes $F_1, F_2, F_3, F_4, F_7, F_8, F_9, F_{10}, F_{11}, F_{12}$ and F_{13}. Hence, it is evidently clear that the proposed χ^2 model for the feature selection improves the performance of ANN by 1.11%.

To further investigate the effectiveness of the χ^2-ANN model, ROC and AUC metrics are utilized. The ROC charts for the χ^2-ANN model and conventional ANN model are given in Figure 4 (a) and Figure 4 (b), respectively. The AUC for the proposed χ^2-ANN model is 0.954 while AUC for conventional ANN model is 0.935. From the ROC charts it is evident that χ^2 model also improves the performance of ANN as the AUC in case of χ^2-ANN model is better than the AUC of ROC chart for conventional ANN.

C. EXPERIMENT NO 3: VALIDATING GENERALIZATION CAPABILITIES OF THE PROPOSED χ^2-DNN MODEL UNDER k-FOLD CROSS VALIDATION

In this experiment, we validate the generalization capabilities of the proposed χ^2-DNN model. In machine learning, different validation schemes can be used to validate the performance of a learning model. Among them, k-fold cross validation is considered a practical method to check the generalization capabilities of a model. Hence, we simulated the proposed model for 20-fold cross validation using the optimal subset of features with $n = 11$ and ADAM algorithm to optimize or learn the weights of the parameters during training process. We divided the whole dataset into 20 folds. Each fold was tested while the model was trained on the data of the remaining folds. The final accuracy is the average of the accuracies achieved on all the 20 folds. This method is good for checking the generalization capabilities of a model because it checks the performance of the model on k different distributions or populations of data. It was observed that the proposed method achieved 91.57% of accuracy, 89.78% of sensitivity, 93.12% of specificity and 0.830 value of MCC. The ROC chart for this experiment is depicted in Figure 5. In the figure, 20 ROC charts for 20 folds are drawn in different colors while the blue curve represents the mean of all the 20 charts. Experimental results show that the proposed method outperformed previous methods discussed in literature, in terms of cross validation accuracy.

TABLE 3. Performance of other well known predictive models on the heart disease dataset.

Model	Hyperparameters	$\mathbf{Acc_{test}}$	$\mathbf{Acc_{train}}$	Spec.	Sens.	MCC
Adaboost	$N_e = 25$	85.55	90.33	87.75	82.92	0.708
Adaboost	$N_e = 50$	86.66	94.68	87.75	85.36	0.731
Adaboost	$N_e = 75$	84.44	94.68	85.71	82.92	0.686
Adaboost	$N_e = 100$	82.22	97.58	81.63	82.92	0.643
Random Forest	$N_e = 10$	81.11	98.06	87.75	73.17	0.619
Extra Tree	$N_e = 10$	68.88	100.0	65.30	73.17	0.383
SVM (Linear)	$C = 0.055$	90.00	84.05	93.87	85.36	0.799
SVM (RBF)	$C = 5, G = 0.2$	90.00	84.54	93.87	85.36	0.799
Proposed	**(50, 2)**	**93.33**	**84.05**	**100**	**85.36**	**0.872**

D. EXPERIMENT NO 4: COMPARATIVE STUDY WITH OTHER WELL KNOWN MACHINE LEARNING MODELS

In this section, to further validate the improved performance of the proposed model, comparative study is conducted with other well known machine learning models that have shown state of the art performance on many biomedical datasets. The machine learning models chosen for this purpose consist of Adaboost ensemble classifier, random forest (RF) classifier, randomized decision tree classifier, linear support vector machine (SVM) and SVM with radial basis function (RBF)

TABLE 4. Classification accuracies of the proposed method and other methods in literature that used the heart disease dataset.

Study (Year)	Method	Accuracy(%)
ToolDiag, RA [28]	IB1-4	50.00
WEKA, RA [28]	InductH	58.50
ToolDiag, RA [28]	RBF	60.00
WEKA, RA [28]	FOIL	64.00
ToolDiag, RA [28]	MLP + BP	65.60
WEKA, RA [28]	T2	68.10
WEKA, RA [28]	1R	71.40
WEKA, RA [28]	IB1c	74.00
WEKA, RA [28]	K^*	76.70
Robert Detrano [28]	Logistic regression	77.00
Newton Cheung (2001) [29]	C4.5	81.11
Newton Cheung (2001) [29]	Naive Bayes	81.48
Newton Cheung (2001) [29]	BNND	81.11
Newton Cheung (2001) [29]	BNNF	80.96
WEKA, RA [28]	Naive Bayes	83.60
Ster and Dobnikar [30]	Fisher discriminant Analysis	84.2
Ster and Dobnikar [30]	Linear discriminant Analysis	84.5
Ster and Dobnikar [30]	Naive Bayes	82.5-83.4
K. Polat et al. (2005) [31]	AIRS	84.50
Ozsen et al. (2005) [32]	Kernel functions with AIS	85.93 (k-fold)
K. Polat et al. (2006) [10]	Fuzzy-AIRS-KNN based system	87.00 (k-fold)
Kahramanli and Allahverdi (2008) [34]	Hybrid neural network system	86.80 (k-fold)
Resul et al. (2009) [2]	Neural network ensembles	89.01 (holdout)
Jankowski and Kadirkamanathan (1997) [35]	IncNet	90.00 (holdout)
Senthil Kumar (2011) [36]	ANFIS	91.18 (holdout)
Samuel et al. (2017) [20]	ANN-Fuzzy-AHP	91.10 (holdout)
Senthil Kumar (2012) [37]	Fuzzy resolution mechanism	91.83 (holdout)
Paul et al. (2018) [23]	Adaptive Weighted Fuzzy system ensemble	92.31 (holdout)
Proposed Method (2019)	**χ^2 Statistical Model + DNN**	**91.57 (k-fold)**
Proposed Method (2019)	**χ^2 Statistical Model + DNN**	**93.33 (holdout)**

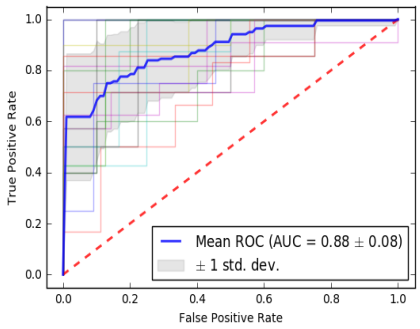

FIGURE 5. ROC chart for 20-fold cross validation.

kernel. The experimental results for these models are reported in TABLE 3.

In TABLE 3, the performance of each model at mentioned values of hyperparameters is reported. For Adaboost classifier, the hyperparameter N_e represents maximum number of estimators at which boosting is terminated. In case of RF classifier, the hyperparameter N_e represents number of trees in the forest. Extra tree also known as randomized decision tree, is an ensemble model that fits a number of randomized decision trees and uses averaging to improve the prediction accuracy. In case of extra tree, the hyperparameter N_e denotes the number of trees used by the ensemble model. For SVM, C is the soft margin constant and G is the width of Gaussian kernel. Finally, in the last row of the table (50, 2) denote the number of neurons in the first hidden layer of DNN i.e., $N_1 = 50$ and the number of neurons in the second hidden layer of the DNN i.e., $N_2 = 2$. From the table, it is evidently clear that the proposed method shows better performance than different state-of-the-art ensemble models and SVM models.

E. COMPARATIVE STUDY WITH OTHER METHODS PRESENTED IN LITERATURE

In this section, comparative study of the proposed method is conducted with other methods presented in literature. The classification accuracies and a brief description of these methods are reported in Table 4.

V. CONCLUSION AND FUTURE WORK

In this paper, we have developed an automated diagnostic system for the diagnosis of heart disease. The proposed diagnostic system used χ^2 statistical model for features refinement and DNN for classification. The strength of the proposed diagnostic system was evaluated using six different evaluation metrics including accuracy, sensitivity, specificity, MCC, AUC and ROC charts. Moreover, the performance of the proposed method was compared with other well known machine learning models and with other methods discussed in the literature. From the experimental results, we can safely conclude that the proposed diagnostic system can improve the quality of decision making during the diagnosis process of heart disease.

The proposed method achieved higher detection accuracy for HF disease, but the current study did not investigate the time complexity of the proposed hybrid diagnostic system. In future studies, it will be investigated as it is considered an important factor in clinical application. Another limitation of the current study is that the optimal width of each hidden layer in ANN and DNN model is searched using grid search algorithm. In future studies, more sophisticated and fast algorithms like genetic algorithm will be utilized.

REFERENCES

[1] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Yarifard, "Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm," *Comput. Methods Programs Biomed.*, vol. 141, pp. 19–26, Apr. 2017.

[2] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7675–7680, 2009.

[3] R. O. Bonow, D. L. Mann, D. P. Zipes, and P. Libby, *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*. Amsterdam, The Netherlands: Elsevier, 2011.

[4] H. Yan, Y. Jiang, J. Zheng, C. Peng, and Q. Li, "A multilayer perceptron-based medical decision support system for heart disease diagnosis," *Expert Syst. Appl.*, vol. 30, no. 2, pp. 272–281, 2006.

[5] K. Vanisree and J. Singaraju, "Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks," *Int. J. Comput. Appl.*, vol. 19, no. 6, pp. 6–12, 2011.

[6] O. W. Samuel, M. O. Omisore, and B. A. Ojokoh, "A Web based decision support system driven by fuzzy logic for the diagnosis of typhoid fever," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4164–4171, 2013.

[7] P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 24, no. 1, pp. 27–40, 2012.

[8] I. Babaoglu, O. Findik, and E. Ülker, "A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine," *Expert Syst. Appl.*, vol. 37, no. 4, pp. 3177–3183, 2010.

[9] S. B. Patil and Y. Kumaraswamy, "Intelligent and effective heart attack prediction system using data mining and artificial neural network," *Eur. J. Sci. Res.*, vol. 31, no. 4, pp. 642–656, 2009.

[10] K. Polat, S. Şahan, and S. Güneş, "Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and *k*-nn (nearest neighbour) based weighting preprocessing," *Expert Syst. Appl.*, vol. 32, no. 2, pp. 625–631, 2007.

[11] E. O. Olaniyi, O. K. Oyedotun, and A. Khashman, "Heart diseases diagnosis using neural networks arbitration," *Int. J. Intell. Syst. Appl.*, vol. 7, no. 12, pp. 75–82, 2015.

[12] N.-S. Tomov and S. Tomov. (2018). "On deep neural networks for detecting heart disease." [Online]. Available: https://arxiv.org/abs/1808.07168

[13] G. Manogaran, R. Varatharajan, and M. K. Priyan, "Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system," *Multimedia Tools Appl.*, vol. 77, no. 4, pp. 4379–4399, 2018.

[14] R. Alizadehsani *et al.*, "Non-invasive detection of coronary artery disease in high-risk patients based on the stenosis prediction of separate coronary arteries," *Comput. Methods Programs Biomed.*, vol. 162, pp. 119–127, Aug. 2018.

[15] A. D. Dolatabadi, S. E. Z. Khadem, and B. M. Asl, "Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM," *Comput. Methods Programs Biomed.*, vol. 138, pp. 117–126, Jan. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169260716306605

[16] U. R. Acharya, H. Fujita, O. S. Lih, M. Adam, J. H. Tan, and C. K. Chua, "Automated detection of coronary artery disease using different durations of ecg segments with convolutional neural network," *Knowl.-Based Syst.*, vol. 132, pp. 62–71, Sep. 2017.

[17] R. Alizadehsani *et al.*, "Coronary artery disease detection using computational intelligence methods," *Knowl.-Based Syst.*, vol. 109, pp. 187–197, Oct. 2016.

[18] I. Paschalidis, "How machine learning is helping us predict heart disease and diabetes," *Harvard Bus. Rev.*, 2017. [Online]. Available: https://hbr.org/2017/05/how-machine-learning-is-helping-us-predict-heart-disease-and-diabetes

[19] M. A. M. Abushariah, A. A. M. Alqudah, O. Y. Adwan, and R. M. M. Yousef, "Automatic heart disease diagnosis system based on artificial neural network (ANN) and adaptive neuro-fuzzy inference systems (ANFIS) approaches," *J. Softw. Eng. Appl.*, vol. 7, no. 12, pp. 1055–1064, 2014.

[20] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, "An integrated decision support system based on ANN and fuzzy_AHP for heart failure risk prediction," *Expert Syst. Appl.*, vol. 68, pp. 163–172, Feb. 2017.

[21] D. P. Kingma and J. Ba. (2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980

[22] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[23] A. K. Paul, P. C. Shill, M. R. I. Rabin, and K. Murase, "Adaptive weighted fuzzy rule-based system for the risk level assessment of heart disease," *Appl. Intell.*, vol. 48, no. 7, pp. 1739–1756, 2017.

[24] J. Loughrey and P. Cunningham, "Overfitting in wrapper-based feature subset selection: The harder you try the worse it gets," in *Research and Development in Intelligent Systems XXI*. London, U.K.: Springer, 2005, pp. 33–43.

[25] Z. K. Pourtaheri and S. H. Zahiri, "Ensemble classifiers with improved overfitting," in *Proc. IEEE 1st Conf. Swarm Intell. Evol. Comput. (CSIEC)*, Mar. 2016, pp. 93–97.

[26] H. Liu and R. Setiono, "Chi2: Feature selection and discretization of numeric attributes," in *Proc. 7th IEEE Int. Conf. Tools Artif. Intell.*, Nov. 1995, pp. 388–391.

[27] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

[28] (2007). *Datasets Used for Classification: Comparison of Results*. [Online]. Available: http://www.is.umk.pl/projects/datasets.html

[29] N. Cheung, "Machine learning techniques for medical analysis," Ph.D. dissertation, School Inf. Technol. Elect. Eng., Univ. Queenland, Brisbane, QLD, Australia 2001.

[30] B. Şter and A. Dobnikar, "Neural networks in medical diagnosis: Comparison with other methods," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, 1996, pp. 427–430.

[31] K. Polat, S. Sahan, H. Kodaz, and S. Günes, "A new classification method to diagnosis heart disease: supervised artificial immune system (AIRS)," in *Proc. Turkish Symp. Artif. Intell. Neural Netw. (TAINN)*, 2005.

[32] S. Ozsen, S. Gunes, S. Kara, and F. Latifoglu, "Use of kernel functions in artificial immune systems for the nonlinear classification problems," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 4, pp. 621–628, Jul. 2009.

[33] H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," *Expert Syst. Appl.*, vol. 35, nos. 1–2, pp. 82–89, 2008.

[34] N. Jankowski and V. Kadirkamanathan, "Statistical control of RBF-like networks for classification," in *Proc. Int. Conf. Artif. Neural Netw.* Berlin, Germany: Springer, 1997, pp. 385–390.

[35] A. V. S. Kumar, "Adaptive neuro-fuzzy inference system for heart disease diagnosis," in *Proc. Int. Conf. Inf. Syst., Comput. Eng. Appl. (ICISCEA)*, Singapore, 2011, pp. 91–99.

[36] A. V. S. Kumar, "Diagnosis of heart disease using fuzzy resolution mechanism," *J. Artif. Intell.*, vol. 5, no. 1, pp. 47–55, 2012.

Authors' photographs and biographies not available at the time of publication.