

Coronary Heart Disease Interpretation Based on Deep Neural Network

Annisa Darmawahyuni, Siti Nurmaini*, Firdaus

Intelligent System Research Group, Universitas Sriwijaya, Indonesia

**sitinurmaini@gmail.com*

ABSTRACT

Coronary heart disease (CHD) population increases every year with a significant number of deaths. Moreover, the mortality from coronary heart disease gets the highest prevalence in Indonesia at 1.5 percent. The misdiagnosis of coronary heart disease is a crucial fundamental that is the major factor that caused death. To prevent misdiagnosis of CHD, an intelligent system has been designed. This paper proposed a simulation which can be used to diagnose the coronary heart disease in better performance than the traditional diagnostic methods. Some researchers have developed a system using conventional neural network or other machine learning algorithm, but the results are not a good performance. Based on a conventional neural network, deeper neural network (DNN) is proposed to our model in this work. As known as, the neural network is a supervised learning algorithm that good in the classification task. In DNN model, the implementation of binary classification was implemented to diagnose CHD present (representative “1”) or CHD absent (representative “0”). To help performance analysis using the UCI machine learning repository heart disease dataset, ROC Curve and its confusion matrix were implemented in this work. The overall predictive accuracy, sensitivity, and specificity acquired was 96%, 99%, 92%, respectively.

Keywords: Coronary Heart Disease, UCI Machine Learning Repository, Feature Scaling, Deep Neural Network

1. INTRODUCTION

Cardiovascular disease is the number one cause of death at 17.7 million deaths, accounting for 45 percent of the total mortality of non-communicable diseases in the world [1]. In Indonesia, the highest prevalence of cardiovascular disease is coronary heart disease (CHD) at 1.5 percent [2]. The Indonesia 2014 Sample Registration System Survey (SRSS) showed that CHD was the highest cause of death at 12.9 percent at all ages after stroke. CHD is most prevalent in the age range 65-74, in the total of 3.6 percent.

The National Heart, Lung, and Blood Institute (NHLBI) describes angina is a common symptom of CHD [3]. Angina is chest pain or discomfort that occurs if an area of the heart muscle doesn't get enough oxygen-rich blood [3]. CHD also known as coronary artery disease or ischaemic heart disease. A silent CHD is a person who has CHD but have no symptoms. The disease might not be diagnosed until a person has symptoms of a heart failure, heart attack, or an arrhythmia [3].

In this work, UCI Machine Learning Repository Heart Disease Dataset is used for classification normal and abnormal (CHD) with 14 attributes belong to it. UCI heart disease dataset provides some information about the chest pain type (angina) that being significant symptoms of CHD. In previous work, some machine learning algorithms used in CHD interpretation. Chowdhury et. al proposed artificial neural network (ANN) and the overall predictive accuracy acquired was 75%, with one hidden layer [4]. Subbalakshmi et. al proposed Naïve Bayes and the increased accuracy achieved was 82.31% [5]. Nahar et. al proposed techniques of computational intelligence for the prediction of heart disease and get an accuracy of 86.77% [6]. Helwan et. al proposed ANN with one hidden layer and the system has 85% accuracy [7].

Due to the number of CHD population increases every year with a significant number of deaths, we proposed a learning algorithm to get better performance in accuracy, sensitivity, and specificity in CHD interpretation. CHD interpretation minimizes the diagnosis time of experts and increases the accuracy of diagnosis. As known as, the traditional feature-based classification is less effective since their performance is usually dependent on the quality of handcrafted features [8]. Conquering those drawbacks, the exploration of feature learning is also used to improve the performance of traditional feature-based approaches. We propose the deep neural network (DNN) for CHD interpretation. As is known to all, DNN can learn a hierarchical feature representation from raw data automatically, so it does need any handcrafted features by experts [9].

This paper is organized as follows. Section 2 presents the of all related works that related in our proposed method and classification problem. Section 3 describes the material and method of DNN model. Section 4 explains the experimentation and the results of this work. The last, section 5 shows the conclusions of this work.

2. RELATED WORKS

In the recent research works, Chowdhury et. al (2011) represents the use of artificial neural networks in predicting neonatal disease diagnosis with a backpropagation learning algorithm. The dataset was collected from 94 instances and improves the diagnostic accuracy of 75% with higher stability [4]. Subbalakshmi et. al (2011) using the data mining modeling technique, namely, Naïve Bayes to predict heart disease from 303 instances (UCI Machine Learning Repository) and accuracy achieved was 82.31% [5].

Based on Subbalakshmi et. al, Sundar et. al (2012) using Naïve Bayes to detect the likelihood of a patient getting a heart disease [10]. The recognition rate of a weighted associative classifier and Naïve Bayes achieved was 84% and 78%, respectively. Nahar et. al (2013) presents a number of computational intelligence techniques in the detection of heart disease using UCI Machine Learning Repository. The medical knowledge based feature selection method has shown promise for use in heart disease diagnostics achieved was 86.77% in accuracy [6]. Vikar et. al (2013) also implement the data mining technique to investigate heart disease using 11 attributes from the UCI Machine Learning Repository. They modelled on Naïve Bayes and J48 decision tree with the recognition rate of 82.31% and 84.35%, respectively [11].

In 2015, Helwan et. al proposed a backpropagation neural network (BPNN) to diagnose heart disease with sharing data ratio 60:40 using six hidden layers. This shows that BPNN is more effective and accurate in the diagnosis of heart disease, the accuracy achieved was 85% [7]. Shinde et. al (2017) proposes a multilayered feedforward neural network and backpropagation neural network with three hidden layers consist of a total 18 nodes to diagnose heart disease. The accuracy achieved through the proposed system is 92% [12].

3. MATERIAL AND METHOD

3.1 DATA PREPARATION

UCI Machine Learning Repository Heart Disease Dataset contains 76 attributes, but all published experiments refer to using a subset of 14 (age, sex, chest pain (cp), trestbps, chol, fasting blood sugar (fbs), restecg, thalach, exang, oldpeak, slope, ca, thal, and including the predicted attribute) of them [13]. This file describes the contents of the CHD directory. The data were collected from the Cleveland Clinic Foundation, with the total number is 303 instances (241 of males, 62 of females, mean age: 51.25 ± 9.74 years). Several missing values and dummy variables belong to Cleveland dataset. In this work, the diagram of the proposed scheme has shown in Figure 1 from data preprocessing and supervised learning based on a neural network.

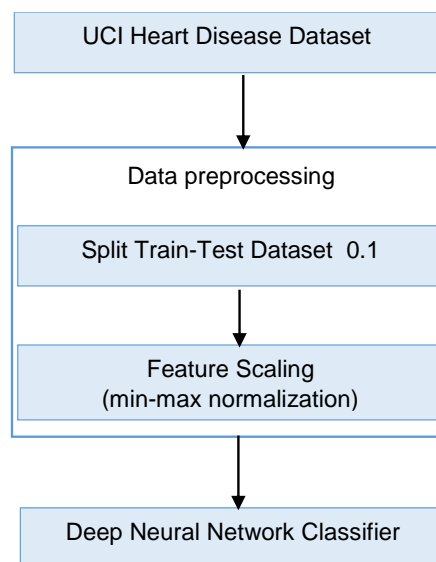


FIGURE 1. The diagram of a proposed scheme

3.2 FEATURE SCALING

The objective functions will not work properly without normalization due to a range of values of raw data widely in the learning algorithm. Before raw data implemented in neural network classifiers, preprocessing data known as normalization used to increase the stability and performance of the network [7]. Feature scaling, a part of normalization is a method used to standardize the range of independent variables or features of data. The simplest method of this feature scaling is a min-max normalization. It consists in rescaling the range of features to scale the range in [0, 1] or [-1, 1]. The general formula is given as Equation (1) [14]:

$$a' = \frac{a - \min(a)}{\max(a) - \min(a)} \quad (1)$$

where; a is original value, a' is normalized value, $\min(a)$ is minimum value of attribute, and $\max(a)$ is maximum value of attribute.

3.3 DEEP NEURAL NETWORK

A neural network is inspired by human neural architecture, reflecting its learning and generalization abilities [15]. A neural network is formed by a series of the neuron (or nodes) and its structure formed by the input layer, hidden layer, and the output layer. Each neuron receives and processes inputs from other neurons and information passed through until the output layers [16]. The network can be viewed as a *black box* or hidden that receives a vector with $X_{ia} = X_{i1}, X_{i2}, \dots, X_{im}$ of inputs and provides a vector with $Y_{ib} = Y_{i1}, Y_{i2}, \dots, Y_{in}$ of outputs. In this work, the black box can be represented as shown in Figure 2.

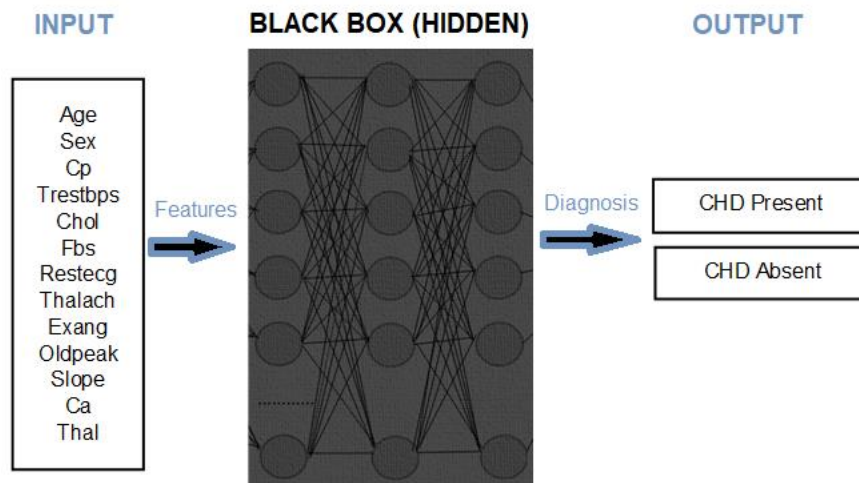


FIGURE 2. *Black Box* of the neural network in CHD interpretation [15]

In network learning, a suitable database must be trained. The database concerning patients for whom the diagnosis CHD present or CHD absent about a certain disease

is already known. The objective of the training process is to approximate the function f between the vector X_{ia} and Y_{ib} [15]:

$$Y_{ib} = f(X_{ia}) \quad (2)$$

Deep neural network (DNN) is a standard feed-forward neural network that is both much larger and much deeper than traditional neural networks [17]. The general deep framework usually used for classification with many hidden layers (more than 2), and it allows complex hypotheses to be expressed [18]. Each layer only receives connections from its previous layer. The network is trained using a supervised learning algorithm DNN. DNN model creates a one input layer, more than one hidden layer, and one output layer. One input layer used in the network model is the number of the feature of the heart disease dataset. Each hidden layer consists of nodes which have activation function. The output layer contains one node which gives the output, whether the CHD present or CHD absent. The sigmoid activation function in an output layer to predict the probability as an output due to its soft switching nature and non-linear characteristics [7]. Since the probability of anything exists only between the range of 0 and 1 [15]:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

In an output layer, a loss function in the binary distribution, we use binary cross-entropy to quantify the difference between two probability distributions (in this work; CHD present and CHD absent). Cross-entropy method provides a simple, efficient and general method for solving such problems [19]. In binary classification, where the total number of classes CHD equals 2, cross-entropy can be calculated as [20]:

$$-(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) \quad (4)$$

where; CHD is classes (CHD present, CHD absent), \log is the natural log, y is true value / binary indicator (CHD present, CHD absent), and \hat{y} is predicted probability (predicted value).

3.4 PERFORMANCE OF CLASSIFICATION ALGORITHM

Classification is a process of determining the class label that a sample belongs to [21]. In a classification problem, the heart disease interpretation based on the confusion matrix has been studied. The confusion matrix contains an information about actual and predicted classifications done by a classification system, it has a two-dimensions; one dimension is indexed by the actual class of an object, and the

other is indexed by the class that the classifier predicts [21]. Assume that a valid procedure of discriminating between CHD present and CHD absent. We differentiate two groups of individuals: with and without a CHD. In real life, a perfect separation between the two groups is quite rare. Most of the times, the distribution of test results will overlap, as shown in Figure 3.

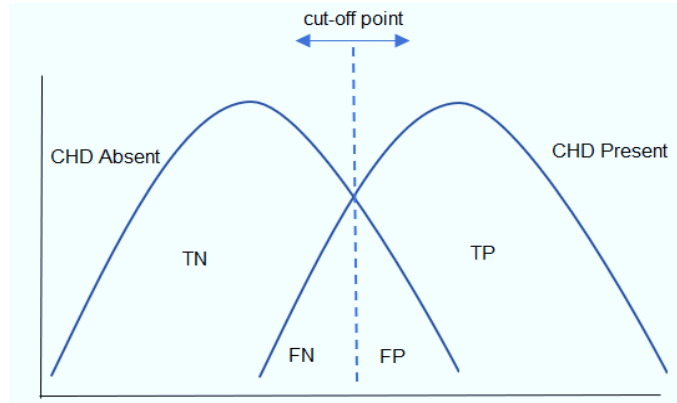


FIGURE 3. Possibility outcomes when intersecting a valid diagnosis with a DNN classifier [22]

In [22][23], from Figure 3, the valid diagnosis is CHD present and it is correctly classified as positive, the outcome is counted as true positive (TP); if the same outcome is incorrectly classified as negative, it is counted as a false negative (FN). If the valid diagnosis is CHD absent and it is correctly classified as negative, the outcome is counted as true negative (TN); if the same outcome is incorrectly classified as positive, it is counted as a false positive (FP). Figure 4 shows a confusion matrix in a binary classification with four possible outcomes.

		TRUE CLASS		
		Positive	Negative	
PREDICT CLASS	Positive	TP	FP	TP : True Positive TN : True Negative FP : False Positive FN : False Negative
	Negative	FN	TN	

FIGURE 4. Confusion Matrix [23]

A number of measures of classification performance can be defined based on the confusion matrix. Some common measures are given as follows [24]:

$$Accuracy (\%) = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (5)$$

$$Sensitivity (\%) = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

$$\text{Specificity (\%)} = \frac{TN}{TN+FP} \times 100\% \quad (7)$$

The total number of true positives (TP) and true negatives (TN) in testing dataset, a confusion matrix can reconstruct from any point on the ROC curve. Receiver Operating Characteristics (ROC) is a plot that displays the full picture of the trade-off between the sensitivity and (1- specificity) across a series of cutoff points [25]. ROC curve is commonly used in the decision making of medical. The medical decision making has an extensive literature on the use of ROC curves for diagnostic testing [26]. ROC curve represents a curveical display of sensitivity (TPR) on y-axis and (1 – specificity) (FPR) on x-axis for varying cut-off points of test values [25]. Two popular indicators; sensitivity and specificity, a statistical validity of medical diagnostic test are the probabilities for interpreting correct diagnosis by test among the true diseased subjects (D+) and the true non-diseased subjects (D-) [25]. Table 1 represents dichotomous response, the results in terms of test positive (T+) or test negative (T-).

TABLE 1.
Diagnostic Test Results [25]

Diagnostic Test Result	Disease Status		Total
	Positive	Negative	
Positive	True Positive (TP)	False Positive (FP)	All tests positive (T+)
Negative	False Negative (FN)	True Negative (TN)	All tests negative (T-)
Total	Total with disease (D+)	Total without disease (D-)	Total Sample Size

4. EXPERIMENTATION AND RESULTS

The recognition rate of this work was compared with previous work on the CHD interpretation [4][5][6][7]. In this paper, the data are sharing about 90% for training and 10% for testing. One input layer used in the network model is the number of the feature of heart disease dataset with a total of 13 nodes. The simulation takes 13 parameters as an input to generate the input layer. Furthermore, the number of hidden layers of this research work was obtained by experimenting which the number of hidden layers will best represent the DNN model. The number of hidden layer was an experiment from 1 layer until 5 layers, and also the number of the epoch was an experiment from 200 until 400 epochs. Each hidden layer consists of total 100 nodes. Each of the hidden layers, the rectified linear unit (ReLU) activation function was used due to its popularity. Since it was used in almost all neural networks or deep learning. ReLU refers to a unit in a neural network that uses the activation function $\max(0, x)$ [27]. The output layer contains one node which gives the output, whether the CHD present or CHD absent. Figure 5 presents of DNN architecture of CHD interpretation of this work.

**Annisa Darmawahyuni, Siti Nurmaini, Firdaus
Coronary Heart Disease Interpretation
Based on Deep Neural Network**

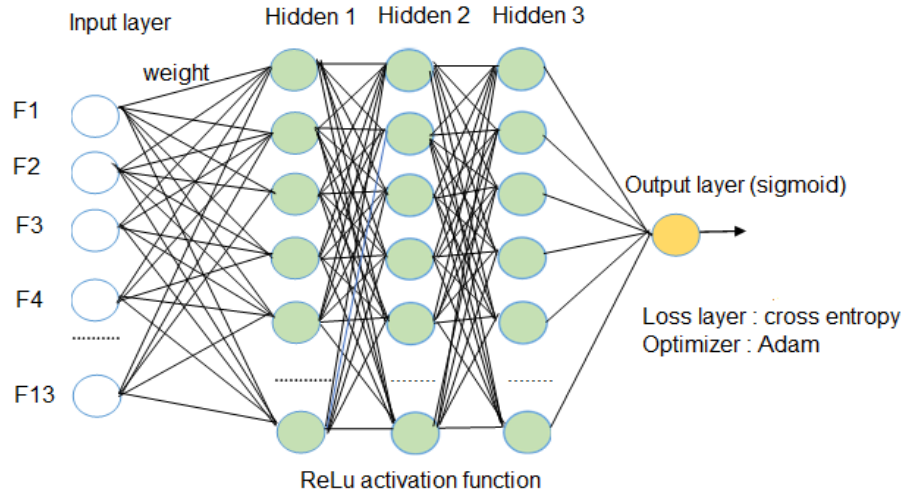


FIGURE 5. DNN architecture of coronary heart disease interpretation

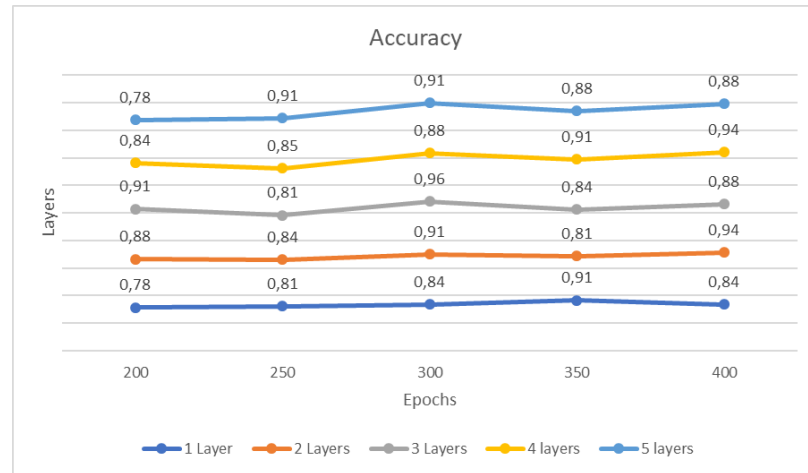
From the experimentation in 1 layer to 5 hidden layers, the best performance of DNN model was using 3 hidden layers with 300 epoch number in the same number of nodes. Table 2 shows the results of performance in our experiment, then the best performance of each layer compared to the other layers with different epoch numbers.

TABLE 2.

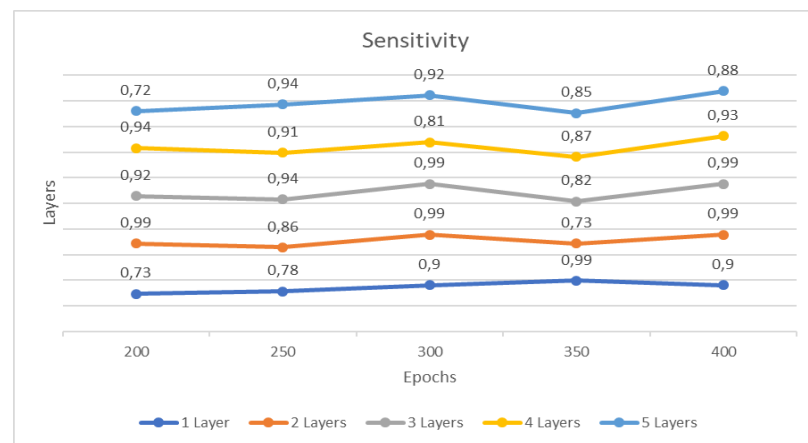
The performance of DNN using one to five hidden layers with different epoch numbers

Hidden Layer	Epoch	Accuracy	Sensitivity	Specificity
1 Layer	200	0.78	0.73	0.82
	250	0.81	0.78	0.83
	300	0.84	0.90	0.76
	350	0.91	0.99	0.80
	400	0.84	0.90	0.80
2 Layers	200	0.88	0.99	0.77
	250	0.84	0.86	0.82
	300	0.91	0.99	0.82
	350	0.81	0.73	0.88
	400	0.94	0.99	0.89
3 Layers	200	0.91	0.92	0.88
	250	0.81	0.94	0.69
	300	0.96	0.99	0.92
	350	0.84	0.82	0.86
	400	0.88	0.99	0.80
4 Layers	200	0.84	0.94	0.73
	250	0.85	0.91	0.75
	300	0.88	0.81	0.94
	350	0.91	0.87	0.94
	400	0.94	0.93	0.93
5 Layers	200	0.78	0.72	0.85
	250	0.91	0.94	0.88
	300	0.91	0.92	0.88
	350	0.88	0.85	0.88
	400	0.88	0.88	0.88

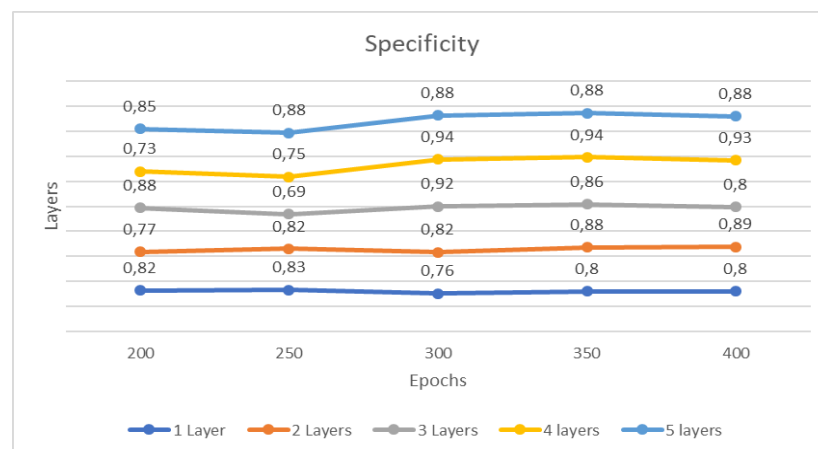
From Table 2, the results can be represented in Figure 6 in accuracy, sensitivity, and specificity, respectively.



(a)



(b)



(c)

FIGURE 6. The performance of DNN using 1 to 5 hidden layers with different epoch numbers, (a) accuracy, (b) sensitivity, (c) specificity

As our proposed model, we achieve an excellent learning process that can implement to predict the testing dataset. In our experiment to achieve a better performance of DNN model based on Table 2, the highest predictive (or testing) accuracy, sensitivity, and specificity acquired was 96%, 99%, 92%, respectively based on Equation (5), (6), and (7). The loss function, binary cross-entropy achieved was 0.3. Represented in Figure 4 in section 3.4 above, the confusion matrix was created to shows diagnostic test that results shown in Table 3. The confusion matrix was achieved from the 0.1 testing dataset. Table 3 shows true positive (TP) means the patient has CHD and positive test result is 23, true negative (TN) means the healthy patient and negative test result is 23, false positive (FP) means the healthy patient and positive test result is 2, and false negative (FN) means the patient has CHD and negative test result is 0.

TABLE 3.
The results of a confusion matrix in binary classification

	CHD Present	CHD Absent
CHD Present	23	2
CHD Absent	0	23

The total number of CHD present (true positive, TP) and CHD absent (true negative, TN) in testing dataset, a confusion matrix can reconstruct from any point on the ROC curve has shown in Figure 7. Figure 7 shows the good performance in ROC curve due to near to the point (0, 1) represents a perfect classification.

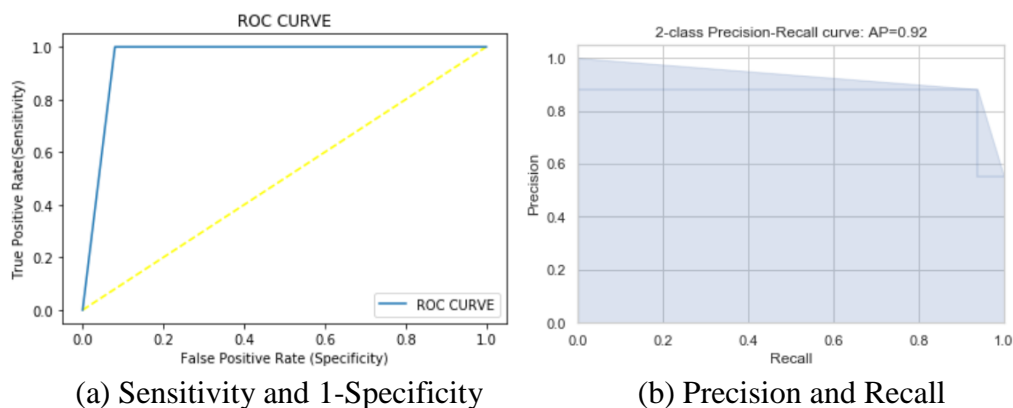


FIGURE 7. ROC curve performance

5. CONCLUSION

The developed coronary heart disease interpretation simulation can be used by experts to automatically diagnose by entering the basic information, especially chest pain type to give the angina information. As we explained above, angina is a common symptom of CHD. The misdiagnosis is a crucial fundamental that is the

major factor that caused death. Therefore, an intelligent system has been designed which will prevent misdiagnosis of CHD. This work presents a simulation which can be used to diagnose the coronary heart disease in better performance than the traditional diagnostic methods. The overall predictive accuracy, sensitivity, and specificity acquired was 96%, 99%, 92%, respectively.

REFERENCES

- [1] W. H. Organization and others, “World health statistics 2017: monitoring health for the SDGs, sustainable development goals,” 2017.
- [2] R. I. Kementrian Kesehatan, “Riset kesehatan dasar,” *Jakarta Kementerian. Kesehatan. RI*, 2013.
- [3] NHLBI, “Coronary Heart Disease.” [Online]. Available: <https://www.nhlbi.nih.gov/health-topics/coronary-heart-disease>.
- [4] D. R. Chowdhury, M. Chatterjee, and R. K. Samanta, “An artificial neural network model for neonatal disease diagnosis,” *Int. J. Artif. Intell. Expert Syst.*, vol. 2, no. 3, pp. 96–106, 2011.
- [5] G. Subbalakshmi, K. Ramesh, and M. C. Rao, “Decision support in heart disease prediction system using naive bayes,” *Indian J. Comput. Sci. Eng.*, vol. 2, no. 2, pp. 170–176, 2011.
- [6] J. Nahar, T. Imam, K. S. Tickle, and Y.-P. P. Chen, “Computational intelligence for heart disease diagnosis: A medical knowledge driven approach,” *Expert Syst. Appl.*, vol. 40, no. 1, pp. 96–104, 2013.
- [7] E. O. Olaniyi, O. K. Oyedotun, A. Helwan, and K. Adnan, “Neural network diagnosis of heart disease,” in *Advances in Biomedical Engineering (ICABME), 2015 International Conference on*, 2015, pp. 21–24.
- [8] Z. Xing, J. Pei, and E. Keogh, “A brief survey on sequence classification,” *ACM Sigkdd Explor. Newsl.*, vol. 12, no. 1, pp. 40–48, 2010.
- [9] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. L. Zhao, “Time series classification using multi-channels deep convolutional neural networks,” in *International Conference on Web-Age Information Management*, 2014, pp. 298–310.
- [10] N. A. Sundar, P. P. Latha, and M. R. Chandra, “Performance analysis of classification data mining techniques over heart disease database,” *IJESAT/ Int. J. Eng. Sci. Adv. Technol. ISSN*, pp. 2250–3676, 2012.
- [11] V. Chaurasia and S. Pal, “Data mining approach to detect heart diseases,” 2014.
- [12] A. A. Shinde, “Heart Disease Prediction System using Multilayered Feed Forward Neural Network and Back Propagation Neural Network,” 2017.
- [13] A. Janosi, “UCI Machine Learning Repository Heart Disease Dataset.” [Online]. Available: <http://archive.ics.uci.edu/ml/datasets/Heart+Disease>. [Accessed: 06-Dec-2018].
- [14] Y. K. Jain and S. K. Bhandare, “Min max normalization based data perturbation method for privacy protection,” *Int. J. Comput. Commun. Technol.*, vol. 2, no. 8, pp. 45–50, 2011.
- [15] F. Amato, A. López, E. M. Peña-Méndez, P. Va\vnhara, A. Hampl, and J. Havel, “Artificial neural networks in medical diagnosis.” Elsevier, 2013.
- [16] N. Larasati, T. Dewi, and Y. Oktarina, “Object following design for a mobile

- robot using neural network,” *Comput. Eng. Appl. J.*, vol. 6, no. 1, pp. 5–14, 2017.
- [17] Y. Lei, N. Scheffer, L. Ferrer, and M. McLaren, “A novel scheme for speaker recognition using a phonetically-aware deep neural network,” in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, 2014, pp. 1695–1699.
- [18] S. Nurmaini, A. Gani, and others, “Cardiac Arrhythmias Classification Using Deep Neural Networks and Principle Component Analysis Algorithm.,” *Int. J. Adv. Soft Comput. Its Appl.*, vol. 10, no. 2, 2018.
- [19] P.-T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, “A tutorial on the cross-entropy method,” *Ann. Oper. Res.*, vol. 134, no. 1, pp. 19–67, 2005.
- [20] S. Ronaghan, “Deep Learning: Which Loss and Activation Functions should I use?” [Online]. Available: <https://towardsdatascience.com/deep-learning-which-loss-and-activation-functions-should-i-use-ac02f1c56aa8>.
- [21] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, “An improved method to construct basic probability assignment based on the confusion matrix for classification problem,” *Inf. Sci. (Ny)*, vol. 340, pp. 250–261, 2016.
- [22] S. Pintea and R. Moldovan, “The receiver-operating characteristic (ROC) analysis: Fundamentals and applications in clinical psychology.,” *J. Cogn. Behav. Psychother.*, vol. 9, no. 1, 2009.
- [23] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.
- [24] J.-S. Wang, W.-C. Chiang, Y.-L. Hsu, and Y.-T. C. Yang, “ECG arrhythmia classification using a probabilistic neural network with a feature reduction method,” *Neurocomputing*, vol. 116, pp. 38–45, 2013.
- [25] R. Kumar and A. Indrayan, “Receiver operating characteristic (ROC) curve for medical researchers,” *Indian Pediatr.*, vol. 48, no. 4, pp. 277–287, 2011.
- [26] K. H. Zou, “Receiver operating characteristic (ROC) literature research,” *Online Bibliogr. available from* <http://splweb.bwh.harvard.edu>, vol. 8000, 2002.
- [27] G. E. Dahl, T. N. Sainath, and G. E. Hinton, “Improving deep neural networks for LVCSR using rectified linear units and dropout,” in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 8609–8613.