Int. J. Advance Soft Compu. Appl, Vol. 10, No. 2, July 2018
ISSN 2074-8523

Cardiac Arrhythmias Classification Using Deep Neural Networks and Principle Component Analysis Algorithm

Siti Nurmaini[1,*], Radiyati Umi P[2], Muhammad Naufal R[1], Ali Gani[2]

[1,*]Intelligent System Research Group
Faculty of Computer Science
Universitas Sriwijaya, Palembang, Indonesia
e-mail: sitinurmaini@gmail.com, naufalrachmatullah@gmail.com

[2]Faculty of Medicine
Universitas Sriwijaya, Palembang, Indonesia
e-mail: radiandinadr@yahoo.co.id, alighanie@yahoo.com

Abstract

A primary diagnostic tool for cardiovascular diseases is electrocardiogram (ECG). The ECG evaluation for predicting cardiac arrhythmias from databases, resulting in a comparable or even higher accuracy than experienced examiners. In this paper, automatic classification for cardiac arrhythmia is developed for producing higher accuracy by using combination between DNNs and PCA technique. The algorithm developed based on the following stages such as, data preparation, data reduction, feature extraction and classification of rhythms. DNNs structure utilize Soft-max regression on the top of output layer and Cross-entropy as a cost function. To validate the method on the well-known MIT-BIH arrhythmia database is used. In the experiment 18 classifier structure with several activation functions are created, to analyze the classifier performance. To benchmark, the performance of DNNs algorithm is compared to SVM algorithm in terms of accuracy. The result obtained show that the proposed method provides good accuracy against to MLP and SVM, about 97.7%, 95.56% and 79.51% respectively with less expert interaction.

Keywords: *Cardiac arrhythmias, deep neural network, ECG signal, classifier, feature learning*

1 Introduction

The Electrocardiogram (ECG) signal is a noninvasive test widely used for reflecting the underlying heart conditions. A careful inspection of its behavior is essential for predicting cardiac arrhythmias, particularly in long-term recordings. Therefore, the utilization of computer-based methods represents an important solution that can benefit cardiologists in the diagnosis. Several pattern recognition methods were developed for arrhythmia detection and classification [1][2][3][4][5]. Usually, these approaches are based on three main steps which are preprocessing, feature extraction; and classification. The ECG signals are enhanced by eliminating various kinds of noise and artifacts [6][7][8] to produce ECG waveforms series. This series of waves consists of six different waveforms, are discernible, and are differentiated as P, Q, R, S, T and U. The first, P wave associated with right and left atrial depolarization. The second wave is the QRS complex. It has a series of 3 deflections that reflect the current associated with right and left ventricular depolarization. Following the QRS complex is the ST segment, extending from where the QRS ends (irrespective of what the last wave in the complex is) to where the T wave begins. In general, it consisting mainly of P wave, QRS complex and T wave are extracted by means of segmentation [9][10][11].

Basically, the available feature representation methods include, but are not limited to, morphology [12][13], temporal information [14][15], wavelet transform [16][17], high-order statistics [15], Hermite basis function [18], and hidden Markov modeling [13]. Moreover, principal component analysis [19], independent component analysis (ICA) [20], and linear discriminant analysis [17] are usually applied to reduce the dimensionality of the feature representation. Finally, the obtained features are used to learn the decision function of a classifier such as neural networks [21][22][23], support vector machines [1][5][24]], path forest [25] and Gaussian processes [1][12]. Despite these great efforts, it has been shown recently [26][5] that automatic methods do not perform well and the results obtained by such methods remain up till now unsatisfactory.

To improve the machine learning methods in cardiac arrhythmias classification, this paper proposes Deep Learning method [27], due to such method produce good feature representation automatically from the input data [28][29][30][31][32]. Furthermore, general deep framework usually used for classification with many hidden layers (more than 2), and it allows complex hypotheses to be expressed, named Deep Neural Networks (DNNs). Such algorithm is used with successes in many areas. Unfortunately, the training process is not trivial because once the errors are back-propagated to the first few layers they become teensy and the learning process can be very slow. Consequently, this work proposes to use DNNs to efficiently solve feature generation and classification Arrhythmias database, for increasing the accuracy with simple data classified. However, the data base is imbalanced, therefore to

reduction such imbalanced data Principal Component Analysis (PCA) is combined.

This paper present classification cardiac arrhythmias from databases MIT-BIH by using Deep Neural Network and Principal Component Analysis. The remainder of the paper is organized as follows. Section 2 gives some explanation about deep neural network algorithm. Section 3 describe method and design of this research. Section 4 discussed the simulation results and discussion about the classifier performance. Finally, section 5 present our conclusions and point the idea for future extension of this work.

2 Deep Neural Networks

The neural network has a very simple architecture and concept. One of the neural networks (NNs) technique is Feed-forward neural networks with many hidden layers, which are often referred to as deep neural networks (DNNs). In such network, Back-propagation (BP) algorithm is used for learning the parameters of these networks. The first generation of the NNs work with BP algorithm, to reveal its fundamental limitations when solving the practical problems that machine learning faced and its performance on practical problems did not meet the expectations, poor performance and always gets trapped in "local minima" [36][37]. DNNs with deep structure to provide a solution to this problem and could improve it.

However, DNNs has the same weakness as NNs, with BP training often resulted in poor performance, due to network was not properly trained [36], and the local optimum happens along with the increase of hidden layer [36]. If learning parameters are trapped into the local optimum, the network can still work well because the probability of having a low local optimum is lower than when a small number of neurons are used in the network [37][38]. Moreover, three primary difficulties in the learning process of DNNs technique, such as vanishing gradient, overfitting and computational load [36]. To improve such method, many types of weight adjustments are proposed to find the best learning technique. The development of various weight adjustment approaches is due to the pursuit of a more stable and faster learning of the network.

Neural Network can be thought of as a function $f_\theta: x \to y$ which takes an input $x \in \mathbb{R}^n$, and produces an output $y \in \mathbb{R}^n$ and whose behavior is parameter by $\theta \in \mathbb{R}^n$. Therefore, for instance, f_θ could be simply $y = f_\theta(x) = \theta.x$. A unit is parameter by a weight vector w and a bias term denoted by b. The first layer has p^1 units then each of the units has $w \in \mathbb{R}^n$ weights associated with them. The first layer produces an output $o_1 \in \mathbb{R}^{p1}$. The output of the unit can be described as, $o_i = f[\sum_{i=1}^{n} x_k . w_k + b_i]$. The index k corresponds to each of the inputs/weights (from1 to n) and the index i corresponds to the unit in the first layer (from 1 to p_1).

We can assume that our data has the form $D = \{(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)\}$. For a single data point, the output of the NNs is computed, which denote as $\hat{y}$.

In the BP algorithm is needed to compute how good the prediction of Neural Network $\hat{y}$ is as compared to y, in the notion of a loss function. A loss functions measures the disagreement between $\hat{y}$ and y which denote by l. There are a number of loss functions appropriate for the task at hand: binary classification, multi-classification, or regression, (typically derived using Maximum Likelihood) as a function,

$$l(\hat{y}, y) = l(f_{NN}(x, \theta), y) \tag{1}$$

and the gradient of the loss function,

$$\nabla l(f_{NN}(x, \theta), y) \tag{2}$$

in this case that $y \in \{0,1, \ldots k\}$ which are the classes, Maximum Likelihood which is to find a θ that maximizes $P(D|\theta)$ where θ represents the parameters of the model. Assuming a Multinomial distribution and given that each of the examples $\{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\}$ are independent, using the following expression:

$$P(D|\theta) = \frac{n!}{n_{1!}.n_{2!} \ldots n_{k!}} \prod_{i=1}^{n} f(x_i, \theta)^{yi} \tag{3}$$

If θ that maximizes in $P(D|\theta)$, thus equation (3) become,

$$\log P(D|\theta) = \sum_{i=1}^{n} y_i log\, f(x_i, \theta) \tag{4}$$

maximizing the RHS we minimize its negative value as follows:

$$-\log P(D|\theta) = -\sum_{i=1}^{n} y_i log\, f(x_i, \theta) \tag{5}$$

Therefore, cross entropy function is given by the expression,

$$-\sum_{i=1}^{n} y_i log\, f(x_i, \theta) \tag{6}$$

Most of the DNNs training approach employ the cross entropy-driven learning rules. This is due to their superior learning rate and performance [39][36].

3 Method and Design

In this paper the DNNs flowchart to process ECG Arrhythmia data classification use six stages such as, data collection, data pre-processing including data preparation, feature extraction, feature classification, learning process, validation and analysis. All stages can be described in Fig. 1. The implementation of classification is conducted by using python programming language which is

combined with some additional libraries including: Numpy libraries, WFDB libraries, Keras libraries and Scious-Learn libraries.

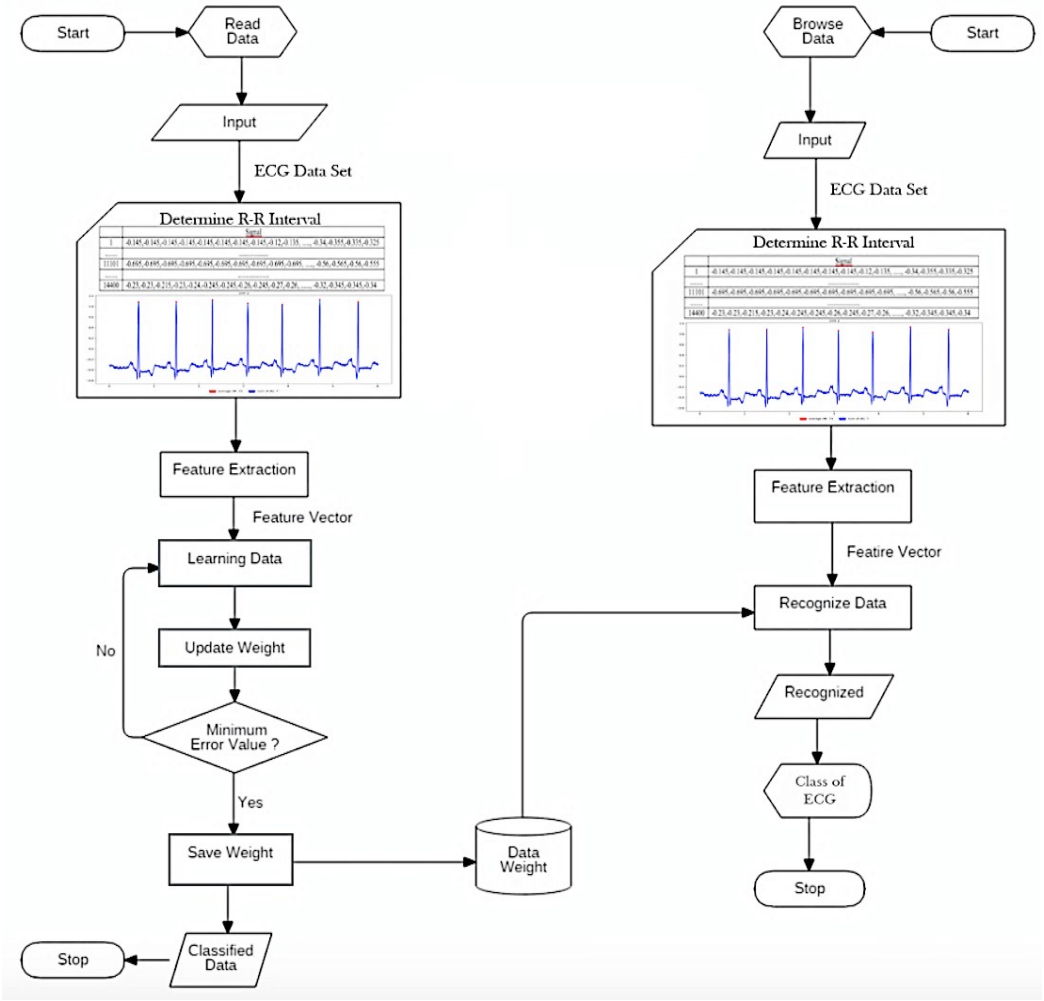


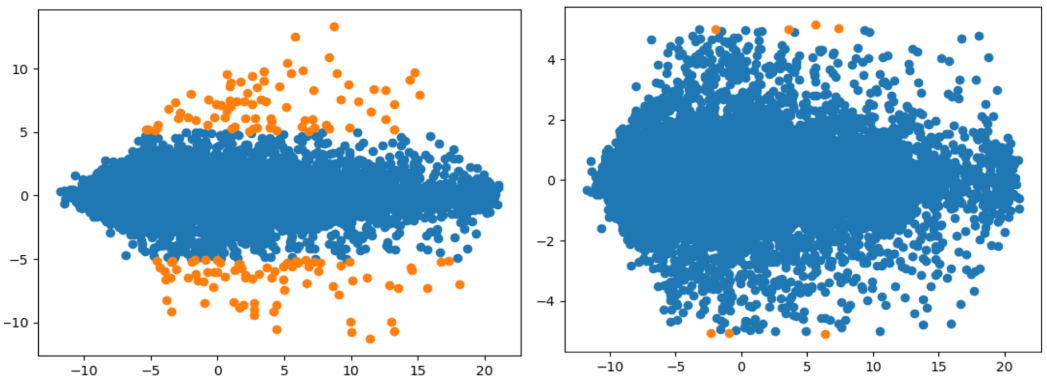
Fig. 1 DNNs classifier flowchart

3.1. Data Preparation

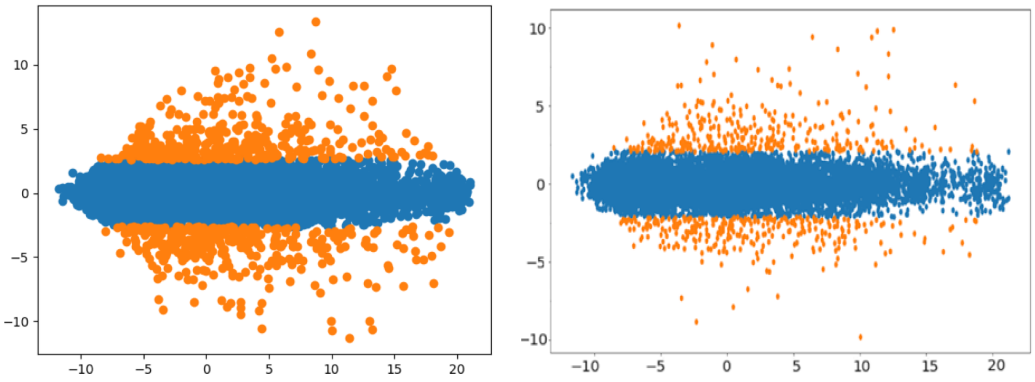
This study uses ECG signal database published by Harvard-MIT Division of Health Sciences and Technology [37]. The number of data about 48 instances consists of 25 data from male patients with age ranges between 32-89 years and 23 data from female patients with age range 23-89 years. The Arrhythmia databases have three file extensions namely .dat, .atr, and .hea. The data with

exr .dat contains ECG signals taken from 2 modified limb lead II (MLII) signal types for upper signal and modified lead V1 (sometimes V2 or V5, and there is 1 instance V4) for lower signal. The data with ext .atr contains the peak position (R) of the ECG signal. In this study, the files used are formatted .dat and .atr and the data with etx .dat contains the signals to be processed, while the data with ext .ann contains peak data of each signal. The signal contains 650,000 records, taken for 30 minutes, all data divided into 2167 sample rate. The duration of each sample records taken about 6 seconds, therefore each instance has sampled each 300 part of signal. The number of ECG data in the classification process about 14,352, with the data distribution such as 7826 data for sinus normal, 2101 data for sinus bradycardia condition, 1694 data for ventricular tachycardia condition and 3271 data for sinus Arrhythmia (irregular) condition.

Before feature extraction process, filtering data is needed to separate the out-layer data from the data used in the training and testing process. It is done by using principal components analysis (PCA) algorithm. The PCA is selected, due to it has target function in the form of minimizing reconstruction error by selecting N Eigen-vector with the highest value. The selected Eigen-vector is a projection of the data attribute that has a good feature discrimination level. In such process, data selection performs based on fix threshold and automatic threshold for determining the reference value (See Fig. 2). The PCA process is described as follows: (1). Create an "X" matrix of size N x d (with N is the total number of rhythm data/6 sec and d is the dimension of rhythm data/6 sec). Each row contains the data point value of each instance; (2). Calculate the average () value of each row in the matrix "X"; (3) Calculate the covariance matrix "X"; (4) Calculate the eigenvalue and eigenvector of the Covariance-matrix; (5). Determine the M of the Eigen vector with the highest Eigen value; (6). Transform all data into the selected Eigen vector; and (7). The results are plotted into the diagram 2 D (x, y) and perform the data selection.



(a) Fix threshold



(b) Automatic threshold

Fig. 2 Data reduction based-on PCA approach

3.2. Feature Extraction

In this research DNNs with BP algorithm is used to process classification of beats rate database. Simple feature extraction on rhythm for data classification is selected by using R-R Interval approach, its use to calculate the distance between peak R to peak R of ECG signal. In order to evaluate the performance of the algorithm used for R-peak detection, taking advantage of the beat annotations present. The Waveform Database (WFDB) Software Package is used for validation [40]. Standard comparison options are used: the comparison started five minutes after the beginning of the record and the match window, maximum absolute difference in annotation times allowable for matching annotations, is set about 0.15 second. Fig. 3, presents the pre-processing stage for feature extraction and labeling. The input signal is fed into the system for extracting, therefore it produces the following vector values. To find the R-R interval, all vectors are transformed in ECG graphic to assign the mark in peak of R graph. From the interval value of R-R features are plotted into a graph. It performs by using software for viewing, analyzing, and creating recording of physiologic signal or WFDB tools.

In practice, the extraction features process of R-R intervals encountered several constraints and one of them is the number of unequal peaks between each signal. Thus, the R-R interval can't be used as an input feature for the classification process, due to the number of R-R intervals are diverse. The problem of differences in the number of R peaks is overcome by conducting data acquisition process named "Zero Padding," to find the highest number of R-R intervals and to produce the number of R-R interval peaks being the same. The zero-padding process gives a value of 0 to the signal having an R-R interval of < 22 features. In addition, the feature also performed an absolute process to eliminate the negative value of the extraction process. The results obtained that the number of R-R Interval largest amounted to 22 features and it's ready for the training and testing process. A sample of ECG data can be separated into a series of waves known as

the P, Q, R, S, and T waves [39]. All ECG signals with noise are not included in the process, it is reduced by using Principal Component Analysis (PCA) algorithm.

(a) Sinus normal (b) Sinus bradycardia

(c) Ventricular tachycardia (d) Sinus Arrhythmia

Fig. 3 The sample of ECG Signal from MIT-BIH data set

The number of peaks in each signal is strongly influenced by the type of signal. If the signal is sinus normal, then the number of peaks ranges from 6-10 records. In sinus bradycardia type signal the number of R peaks < 6, whereas the signal belonging to ventricular tachycardia has number of R peak > 10. Rhythms classified based on data Beats per Minute (BPM) from each record. BPM is the duration between two identical points of consecutive ECG waveforms such as the R-R interval. The BPM value generated between 60-100 is assigned as the sinus normal rate; the BPM value generated is less than 60 or BPM < 60 is assigned as sinus Bradycardia; The BPM value generated more than 100 or BPM > 100 is assigned as ventricular Tachycardia; and the last one if the resulting BPM belongs to a normal signal, but one of the R-R Intervals is <600 is assigned as sinus Arrhythmia (irregular).

3.3. Deep Neural Networks Method

A DNNs comprises of multiple nodes connected with each other, each node representing the activation function. The simple structure of proposed DNNs can be seen in Fig. 4, that is the best model with highest accuracy.

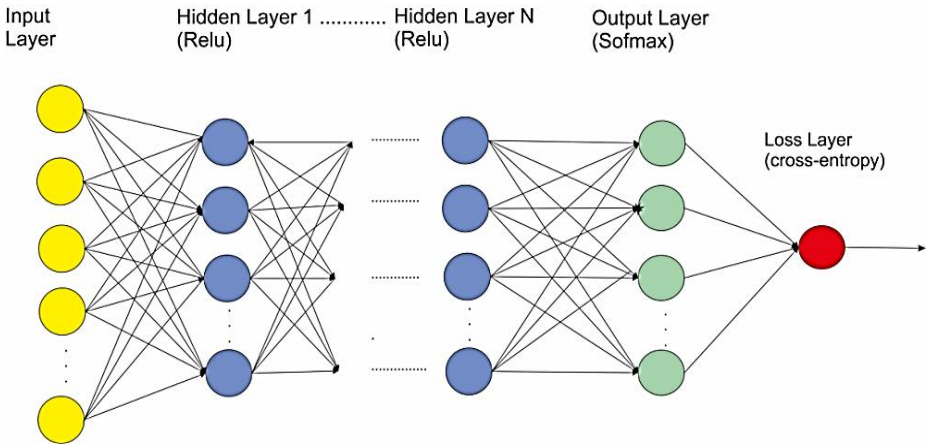


Fig.4 The proposed of DNNs structure

The classification of cardiac arrhythmias based on DNNs use BP algorithm for training and testing. Each step of the classification process can be described as follows;

1. Create input layer (X_i), hidden layer unit (Z_{in_j}), weight (V_{ij}) and activation function f (Relu),

$$\begin{aligned} Z_{in\_j}[j] &= X_i[i] * V_{ij}[i,j] \\ Z_{in\_j}[j] &= Z_{in\_j}[j] + V_{ij}[i,j] \\ Z[j] &= f(Z_{in\_j}[j]) \\ f\ (relu) &= \max\ (Z_{in\_j}[j], 0) \end{aligned} \tag{7}$$

2. Create output layer (Y_{ink}), hidden layer output value (Z_j), weight (W_{jk}) and activation function Y_k

$$\begin{aligned} Y_{ink}[k] &= Z_z[j] * W_{jk}[j,k] \\ Y_{ink}[k] &= Y_{ink}[k] + W_{jk}[0,k] \\ Y_k[k] &= f(Y_{ink}[k]) \end{aligned} \tag{8}$$

3. Determine error in the output layer by using soft-max, and to obtain a global error used cross-entropy method,

$$Error = \frac{1}{N} * \sum_N \sum_I (target_{n,i} * \ln(Yk_{n,i}) \tag{9}$$

Error calculation in derivative of activation function as,

$$delta_k[k] = \text{Error Cross Entropy} * f'(Y_{ink}[k]) \quad (10)$$

4. Calculate the gradient of the loss function against all parameters by finding a partial derivative of the function, then use delta rule method. The derivative of the activation function on the output layer (soft-max function).

$$f'(softmax) = \begin{cases} y_i * (1 - y_i); \ for \ i = j \\ -y_i * y_j; for \ i \neq j \end{cases} \quad (11)$$

The cross-entropy error derivation using the Soft-max activation function as follows:

$$\begin{aligned} \frac{\partial Error}{\partial zi} &= -\sum_{j=1}^{C} t_j \frac{1}{y_j} \frac{\partial y_j}{\partial z_i} \\ &= -\frac{t_i}{y_i} y_i * (1 - y_i) - \sum_{j \neq i}^{C} \frac{t_j}{y_j} - y_i * y_j \\ &= -t_i + y_i \sum_{j=i}^{C} t_j \\ &= y_i - t_i \end{aligned}$$

$$delta_k[k] = (y_i - t_i) \quad (12)$$

5. Update the weight between output and hidden layer

$$W_{jk}[j,k] = e_{rate} * delta_k[k] * Z_j[j] + (\text{update}\_W_{jk}[j,k] * e_{momentum})$$
$$W_{jk}[0,k] = e_{rate} * delta_k[k] \quad (13)$$

6. Error calculation in each weight

$$delta_{in\_j}[j] = delta_{in\_j}[j] + delta_k[k] * W_{jk}[j,k]$$
$$delta_j[j] = delta_{in\_j}[j] * f'(Z_{in\_j}[j] \quad (14)$$

7. Update the weight between hidden and input layer

$$W_{jk}[j,k] = e_{rate} * delta_k[k] * Z_j[j] + (\text{update}\_W_{jk}[j,k] * e_{momentum}) \quad (15)$$

$$W_{jk}[0,k] = e_{rate} * delta_k[k] \ (30)$$

8. Update all parameters (weight and bias) using Stochastic Gradient Descent by reducing or adding the old weight value or the "partial" (learning rate) of the gradient values by using the equation as:

$$W_{jk}[j,k] = W_{jk}[j,k] + \text{update}\_W_{jk}[j,k]$$
$$V_{ij}[i,j] = V_{ij}[i,j] + \text{update}\_V_{ij}[i,j] \quad (16)$$

4 The Results and Discussion

In this paper, the data are sharing about 70% for training and 30% for testing with 10 cross fold. The performance of DNNs classifier is compared to, MLP and Support Vector Machine (SVM) in terms of accuracy, sensitivity, precision and F1 measure. The experiments are conducted in several multilayer's structures, from 1 hidden layer until 10 hidden layers. In the neural network structure, the activation function in hidden layer and output layer also lost function are changed to find the good performance. In the hidden layer the activation function uses Sigmoid and Relu, in the output layer, the activation function use Sigmoid, Soft-sign and Soft-max and lost function use Categorical-Cross-entropy, and Mean squared error. Other classifiers including SVM with linear, RBF and Polyline kernel create to complete the analysis of the classifier structure.

The experiment on several NNs structures aims to analyze the effect of many layers on the formation of robust models and changes in accuracy. Table 2 shows that the validation model and highest accuracy are obtained by case 1 with the number of hidden layers is 1. In contrast, the smallest accuracy is obtained by case 6 with the number of hidden layers is 6. Furthermore, the training process on the case 7 to 9 (with the number of hidden layers 7 to 9) have not changed. In such condition, the number of layers that are overloaded, the error generated in the feed-forward process is insufficient to make any weight changes on each layer.

From Table 2, describes the confusion matrix for DNNs cases 2-8 and SVM (1 vs 1) it can be seen that the largest classification error is obtained in normal and irregular data. This is probably because both types of signals (normal and irregular) have nearly the same number of R-R intervals. Therefore, the features generated by both types of signals are mutually exclusive. The highest accuracy of testing data classification is obtained by DNN algorithm of 2-8 cases with an accuracy of 96.7%. While the highest accuracy of testing data classification use SVM algorithm is 79.51% with configuration of linear SVM (1 vs 1). However, SVM with the polyline kernel was not successfully modeled because of the long-lasting training process (more than eight hours of training process not completed). From the experimental results, the greatest misclassification occurs in normal sinus and sinus arrhythmia classes, due to both signals have overlapping features. All result of 18 cases of experiment can be seen in Table1, and Table 2 respectively.

From the all experimental results are obtained that the highest accuracy in DNN structure with eight layers. The parameters of DNNs such as activation function at hidden layers is Relu function, activation function in output layer is Soft-max function and loss function is categorical-cross entropy, it produces accuracy about 96.7%. But, in general, the change number of hidden layers not very influential only + - 2%. In addition, the time required for training tends to increase as the hidden layer increases. Therefore, the computational cost is increased too.

Table 1: Classifier accuracy without PCA for four classes

Case	Sinus Normal	Sinus Bradycardia	Ventricular Tachycardia	Sinus Arrhythmia
MLP 1	94.37	99.86	91.13	98.88
MLP 2	94.90	92.45	90.96	66.87
MLP 3	95.59	99.66	94.11	98.72
MLP 4	97.86	86.01	97.95	95.34
MLP 5	94.15	92.73	91.47	65.19
MLP 6	97.84	95.38	94.97	96.11
DNNs 2 layers	97.21	99.59	97.86	94.49
DNNs 3 layers	95.12	99.46	99.49	97.48
DNNs 4 layers	97.66	99.73	98.55	93.89
DNNs 5 layers	96.06	99.86	99.57	93.33
DNNs 6 layers	96.68	99.73	94.45	90.85
DNNs 7 layers	97.04	99.79	97.53	93.72
DNNs 8 layers	98.44	99.66	98.12	94.27
DNNs 9 layers	97.47	99.18	99.65	94.06
DNNs 10 layers	97.47	90.22	99.66	93.97
SVM (1 vs 1)	92.12	94.16	93.34	34.72
SVM (1vs All)	82.86	99.86	89.16	76.96

Table 2: Average accuracy for NNs, DNNs, and SVM without PCA for several activation functions and loss function

Classifier Model	Performance	
	Training (%)	Testing (%)
(HN= 100); (HL= 1), (E = 100); (AH = Sigm; AO = Sigm); (LF = CCE)	95.58	95.56
(HN = 100); (HL = 1), (E = 100); (AH = Sigm; AO = Sigm); (LF = MSE)	87.38	87.67
(HN = 100); (HL = 1), (E = 100); (AH = Relu; AO = Soft-max); (LF = CCE)	97.28	96.24
(HN = 100); (HL = 1), (E = 100); (AH = Relu; AO = Soft-sign); (LF = MSE)	97.24	94.24

(HN = 100); (HL = 1), (E = 100); (AH =Sigm; AO= Soft-sign); (LF =MSE)	86.18	86.95
(HN = 100); (HL = 2), (E = 100); (AH = Relu; AO = Soft-max); (LF = CCE)	97.00	96.38
(HN = 100); (HL = 3), (E = 100); (AH = Relu; AO = Soft-max); (LF = CCE)	96.18	95.59
(HN = 100); (HL = 4), (E = 100); (AH = Relu; AO = Soft-max); (LF = CCE)	97.19	96.68
(HN = 100); (HL= 5), (E = 100); (AH = Relu; AO = Soft-max); (LF = CCE)	97.41	94.96
HN = 100); (HL = 6), (E = 100); (AH = Relu; AO = Soft-max); (LF = CCE)	96.48	95.38
(HN = 100); (HL = 7), (E = 100); (AH = Relu; AO = Soft-max); (LF = CCE)	96.76	96.01
(HN = 100); (HL = 8), (E = 100); (AH = Relu; AO = Soft-max); (LF = CCE)	97.67	96.70
(HN =100); (HL = 9), (E = 100); (AH = Relu; AO = Soft-max); (LF = CCE)	97.18	95.73
(HN =100); (HL = 10), (E = 100); (AH = Relu; AO = Soft-max); (LF = CCE)	96.73	94.33
SVM with Linear Kernel (One vs One)	79.19	79.51
SVM with Linear Kernel (One vs Rest)	68.58	69.53
SVM with RBF Kernel	100	51.6
SVM with Polyline Kernel	Fail	Fail

(Number of Hidden Node = HN, Number of Hidden Layer = HL, Epoch = E, Activation Hidden = AH, Activation Output = AO, Loss Function = LF, Sigmoid =Sigm, Categorical-Cross-entropy= CCE, Mean-Square error= MSE)

Table 2 and table, shows the training and testing results obtained for each configuration on DNN structure against SVM. Here, we clearly notice that using a DNN representation (case 2.4) leads to better classification results compared to others structure. The number of hidden layers is 8, by using 100 hidden nodes. All the parameters are a reasonable choice for our initial DNNs model. Confirm clearly the superiority of DNNs over standard MLP and SVM. From the

classification result using SVM algorithm, the best data testing accuracy is obtained by using SVM -Linear with one-against-one setting with an accuracy of 79.19. Therefore, this SVM model takes a long time in training around 2176 seconds. RBF kernel has 100% accuracy in training by using SVM with, but it decreases to 51.6 % in testing. However, the SVM performance is not good when compared to NNs 1 hidden layer because with NNs 1 layer accuracy reach 95.56%. The biggest classification error is obtained in both normal and irregular data, due to both types of signals have nearly the same number of R-R intervals, it affects the accuracy of the beat rate.

After obtaining the best model in terms of accuracy data from all experiments without noise data reduction, furthermore, the PCA algorithm is implemented. Such algorithm is used to reduce all data ECG which contain the noise or unknown beats elements for comparing the classifier performance. The experiment was conducted in 12 cases with different structures. Table 3 shows the average result of accuracy during training and testing. The results show that an unfiltered MLP algorithm provides a good accuracy value because the amount of data becomes significantly reduced and decreases the performance of DNNs.

Table 3: Classification accuracy in MLP, DNNs, SVM with PCA

Description of NNs Structure	Training Accuracy	Testing Accuracy	Average Accuracy
MLP (No. Filter)	97.57	96.40	96.17
MLP with Fix Threshold	97.62	97.13	95.92
MLP with Auto Threshold	97.21	96.77	95.72
DNN without Filter	97.14	96.47	95.30
DNN with Fix Threshold	96.97	96.15	95.03
DNN with Auto Threshold	96.98	95.66	94.89
SVM RBF without Filter	95.05	51.63	50.39
SVM RBF with Fix Filter	95.08	51.54	50.54
SVM RBF with Auto Filter	95.11	50.75	50.86
SVM Linear without Filter	79.28	79.28	79.03
SVM Linear with Fix Filter	79.56	79.75	79.57
SVM Linear with Auto Filter	79.77	79.30	79.49

Analysis of the overall accuracy of the DNNs implementation in this study can be seen in Fig. 5. The parameters assessed are sensitivity, precision, F measure, and accuracy. It has been obtained from the results, that the highest average sensitivity value is the sinus normal data, and the lowest average value is the sinus arrhythmia data. While the highest average precision value is Ventricular tachycardia data while the lowest average value is sinus arrhythmia data. In addition, the highest F measure value is Ventricular tachycardia data while the lowest value on the sinus arrhythmia data. Therefore, it can be concluded that the

lowest classification parameter is on the sinus arrhythmia data because there is still a lot of data that has not been clearly classified due to noise. While the lowest classification parameter that is on the ventricular tachycardia data, because the amount of data is less and the value is detailed classified.

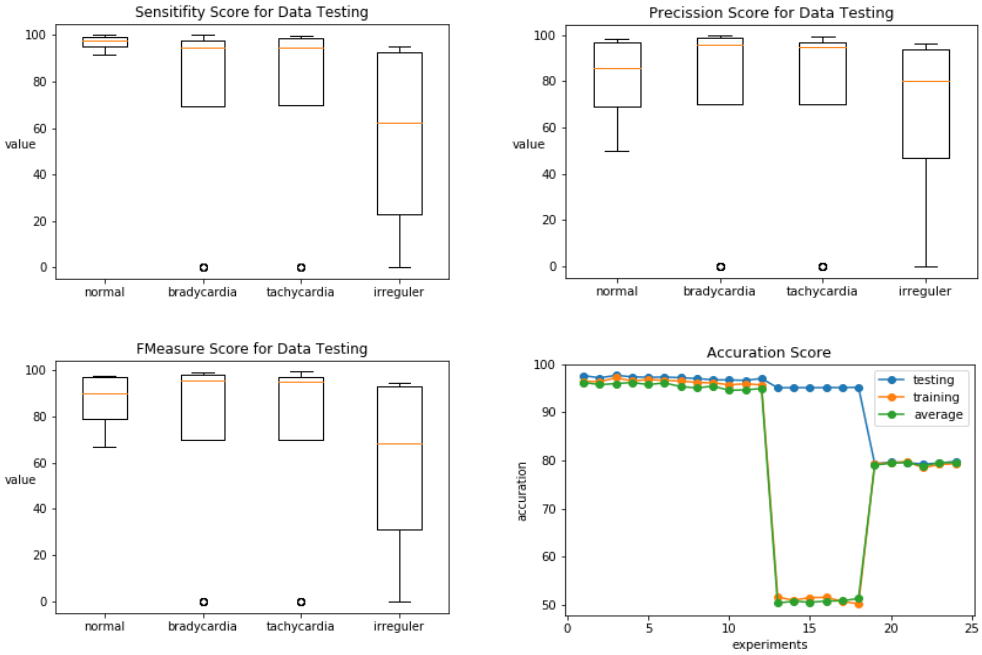


Fig. 5 Performances of all classifier

5 Conclusion

We have proposed and evaluated an intelligent system for the cardiac activity. This system classifies an ECG waveform in terms of rhythm features using a Deep Neural Networks with 18 cases against NNs and SVM algorithm. The proposed DNNs achieved a high-level of accuracy compare to NNs and SVM, about 96.7 %, 95.56%, and 79.51% respectively. However, the noise data is included in the classification process without process noise cancelation in pre-processing stage. This research is appropriate for a project at an early stage. In the future, we intend to implement and evaluate our system based on unsupervised approach, with noise cancelation processing using Denoising Auto-Encoder (DAE) with AAMI class and the data obtained directly from a patient, by adding a "live mode" to our system, which will allow it to be combined with Holter monitoring. We also plan to improve the accuracy of classification by reformulating the method of P wave detection and QRS complex analysis.

ACKNOWLEDGMENT
Authors would like to thank Universitas Sriwijaya and the Ministry of Research Technology and Higher Education of Indonesia, for their support in our research work, with the funded of Hibah Profesi 2018.

References

[1] N. Alajlan, Y. Bazi, F. Melgani, S. Malek, and M. A. Bencherif, "Detection of premature ventricular contraction arrhythmias in electrocardiogram signals with kernel methods," *Signal, Image Video Process.*, vol. 8, no. 5, pp. 931–942, 2014.

[2] R. Alcaraz, F. Sandberg, L. Sörnmo, and J. J. Rieta, "Classification of paroxysmal and persistent atrial fibrillation in ambulatory ECG recordings," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 5, pp. 1441–1449, 2011.

[3] F. Alonso-Atienza, E. Morgado, L. Fernandez-Martinez, A. Garc'\ia-Alberola, and J. L. Rojo-Alvarez, "Detection of life-threatening arrhythmias using feature selection and support vector machines," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 3, pp. 832–840, 2014.

[4] A. S. Alvarado, C. Lakshminarayan, and J. C. Principe, "Time-based compression and classification of heartbeats," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 6, pp. 1641–1648, 2012.

[5] M. R. Homaeinezhad, S. A. Atyabi, E. Tavakkoli, H. N. Toosi, A. Ghaffari, and R. Ebrahimpour, "ECG arrhythmia recognition via a neuro-SVM--KNN hybrid classifier with virtual QRS image-based geometrical features," *Expert Syst. Appl.*, vol. 39, no. 2, pp. 2047–2058, 2012.

[6] M. Längkvist, L. Karlsson, and A. Loutfi, "Sleep stage classification using unsupervised feature learning," *Adv. Artif. Neural Syst.*, vol. 2012, 2012.

[7] R. Sameni, M. B. Shamsollahi, C. Jutten, and G. D. Clifford, "A nonlinear Bayesian filtering framework for ECG denoising," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2172–2185, 2007.

[8] B. H. Tracey and E. L. Miller, "Nonlocal means denoising of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2383–2386, 2012.

[9] V. Bono *et al.*, "Development of an automated updated selvester QRS scoring system using SWT-based QRS fractionation detection and classification," *IEEE J. Biomed. Heal. informatics*, vol. 18, no. 1, pp. 193–204, 2014.

[10] X. Ning and I. W. Selesnick, "ECG enhancement and QRS detection based on sparse derivatives," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 713–723, 2013.

[11] P. Phukpattaranont, "QRS detection algorithm based on the quadratic filter," *Expert Syst. Appl.*, vol. 42, no. 11, pp. 4867–4877, 2015.

[12] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, 2004.

[13] S.-M. Dima *et al.*, "On the detection of myocadial scar based on ECG/VCG analysis," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3399–3409, 2013.

[14] P. de Chazal and R. B. Reilly, "A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2535–2543, 2006.

[15] Y. Kutlu and D. Kuntalp, "Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients," *Comput. Methods Programs Biomed.*, vol. 105, no. 3, pp. 257–267, 2012.

[16] H. Yang, C. Kan, G. Liu, and Y. Chen, "Spatiotemporal differentiation of myocardial infarctions," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 4, pp. 938–947, 2013.

[17] S.-N. Yu and K.-T. Chou, "Integration of independent component analysis and neural networks for ECG beat classification," *Expert Syst. Appl.*, vol. 34, no. 4, pp. 2841–2846, 2008.

[18] G. De Lannoy, D. François, J. Delbeke, and M. Verleysen, "Weighted conditional random fields for supervised interpatient heartbeat classification," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 1, pp. 241–247, 2012.

[19] R. J. Martis, U. R. Acharya, and L. C. Min, "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform," *Biomed. Signal Process. Control*, vol. 8, no. 5, pp. 437–448, 2013.

[20] C. Ye, B. V. K. V. Kumar, and M. T. Coimbra, "Heartbeat classification using morphological and dynamic features of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2930–2941, 2012.

[21] W. Jiang and S. G. Kong, "Block-based neural networks for personalized ECG signal classification," *IEEE Trans. Neural Networks*, vol. 18, no. 6, pp. 1750–1761, 2007.

[22] 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.

[23] E. D. Übeyli, "Recurrent neural networks employing Lyapunov exponents for analysis of ECG signals," *Expert Syst. Appl.*, vol. 37, no. 2, pp. 1192–1199, 2010.

[24] S. Dutta, A. Chatterjee, and S. Munshi, “Correlation technique and least square support vector machine combine for frequency domain based ECG beat classification,” *Med. Eng. Phys.*, vol. 32, no. 10, pp. 1161–1169, 2010.

[25] E. J. da S. Luz, T. M. Nunes, V. H. C. De Albuquerque, J. P. Papa, and D. Menotti, “ECG arrhythmia classification based on optimum-path forest,” *Expert Syst. Appl.*, vol. 40, no. 9, pp. 3561–3573, 2013.

[26] F. Melgani and Y. Bazi, “Detecting premature ventricular contractions in ECG signals with Gaussian processes,” in *Computers in Cardiology, 2008*, 2008, pp. 237–240.

[27] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, 2013.

[28] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science (80-.).*, vol. 313, no. 5786, pp. 504–507, 2006.

[29] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[30] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, “Extracting and composing robust features with denoising autoencoders,” in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 1096–1103.

[31] P. Swietojanski, A. Ghoshal, and S. Renals, “Convolutional neural networks for distant speech recognition,” *IEEE Signal Process. Lett.*, vol. 21, no. 9, pp. 1120–1124, 2014.

[32] J.-C. Li, W. W. Y. Ng, D. S. Yeung, and P. P. K. Chan, “Bi-firing deep neural networks,” *Int. J. Mach. Learn. Cybern.*, vol. 5, no. 1, pp. 73–83, 2014.

[33] P. W. Mirowski, Y. LeCun, D. Madhavan, and R. Kuzniecky, “Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG,” in *Machine Learning for Signal Processing, 2008. MLSP 2008. IEEE Workshop on*, 2008, pp. 244–249.

[34] D. Wang and Y. Shang, “Modeling physiological data with deep belief networks,” *Int. J. Inf. Educ. Technol.*, vol. 3, no. 5, p. 505, 2013.

[35] M. Huanhuan and Z. Yue, “Classification of electrocardiogram signals with deep belief networks,” in *Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on*, 2014, pp. 7–12.

[36] S. Min, B. Lee, and S. Yoon, “Deep learning in bioinformatics,” *Brief. Bioinform.*, vol. 18, no. 5, pp. 851–869, 2017.

[37] G. B. Moody and R. G. Mark, “The impact of the MIT-BIH arrhythmia database,” *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001.

[38] B. M. Asl, S. K. Setarehdan, and M. Mohebbi, “Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal,” *Artif. Intell. Med.*, vol. 44, no. 1, pp. 51–64, 2008.

[39] R. Arsanjani *et al.*, “Prediction of revascularization after myocardial perfusion SPECT by machine learning in a large population,” *J. Nucl. Cardiol.*, vol. 22, no. 5, pp. 877–884, 2015.

[40] P. S. Kostka and E. J. Tkacz, “Feature extraction for improving the support vector machine biomedical data classifier performance,” in *Information Technology and Applications in Biomedicine, 2008. ITAB 2008. International Conference on*, 2008, pp. 362–365.

[41] M. M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. R. Yager, “Deep learning approach for active classification of electrocardiogram signals,” *Inf. Sci. (Ny).*, vol. 345, pp. 340–354, 2016.