Hybrid Two Stage Neuro Genetic System for Arrhythmia Diagnosis

Hela Lassoued and Raouf Ketata,

National Institute of Applied Science and Technology (INSAT), University of Carthage, Tunis, Tunisia.
Electronic, Robotic, Control and Optimization (E.R.C.O) research laboratory

Summary
This work plans to design an intelligent Electrocardiogram (ECG) diagnosis support system that can identify heart abnormalities with high accuracy (ACC), low normalized mean square error (NMSE), and fast classification time response (TR). Therefore, 12 proposed Multilayer Perceptron networks (MLP) architectures, based on Single Stage MLP and Two Stage MLP, were evaluated and compared for their ability to classify five cardiac classes. The training and testing ECG signals were obtained from MIT-BIH database. The network inputs are either the entire feature dataset, containing 62 features based on morphological and Discrete Wavelet Transformer (DWT) coefficients, or a selected feature dataset, which is acquired after applying the Genetic Algorithm (GA). Concerning the number of hidden neurons, the test and error method and the GA designed for the network optimization were evaluated. The obtained performances were compared and discussed. Among the different networks, the proposed two-stage MLP network, called Net6, which uses the selected ECG features and applies the GA to fix its hidden neurons number, possesses the highest ACC, a lower NMSE and an acceptable classification time. Therefore, this network proves to be a suitable classifier in ECG diagnosis system especially when the analysis requires an accurate result and no matter of the classification time response.

Key words:
Arrhythmia diagnosis, hybrid, neuro, genetic, Two Stage MLP.

1. Introduction

The cardiologists use the Electrocardiogram (ECG) waveform mainly for identifying various cardiac diseases and thus saving patients from one of the important reason of death worldwide [1]. ECG waveform is the graphical recording of the electrical activity of the human heart [2]. It represents a single heartbeat which consists of three distinct wave shapes: the P wave, the QRS complex, and the T wave, as they are indicated in Fig.1. In fact, the three shapes involve three successive waves: atrial depolarization, ventricular depolarization and ventricular repolarization, respectively [2]. Consequently, when the ECG recording becomes irregular, cardiologists considered the ECG signal in an arrhythmia case. Unfortunately, diagnosing manually the ECG signal is time consuming task which requires a carefully control in detail. Thus, in order to overcome this confront, cardiologists are presently relying to the intelligent Cardiac Diagnosis Support Systems (CDSS). However, they are focusing frequently on the most advanced technologies of these intelligent systems. Consequently, the need to improve the CDSS performances, especially the accuracy, the classification time response and the memory computation, becomes a challenge. Thus, in order to cope with such an issue, the emergency of machine learning approaches in the soft computing field, seem to be an achievable alternative [3]. Indeed, presently, new perspectives are investigated with different methodologies on the arrhythmia classification field, which they have already been tested and shown good results [4]. These methodologies emphasize the application of various approaches, including the Genetic Algorithm (GA) [5], the Artificial Neural Networks (ANN)[6], the Fuzzy logic network [7], the Neuro-fuzzy logic Networks [8] and many others [9].
Otherwise, we sign the big number of such applications which use the Artificial Neural Networks (ANN) [10]. However, one of the causes to achieve an accurate and fast result is a good configuration of the neural classifier. Unfortunately, configuring correctly an ANN necessitates a careful choice of its topology, including the number of hidden layers and neurons in each layer, the type of the training algorithm, the type of the activation function, the size and the quality of the input feature vector and many other constraints [11]. In fact, in order to overcome this issue, the cross-validation or the test and error technique, is the more used approach in the literature [12]. In fact, it is mentioned that this approach revealed good results which can be enhanced. Hence, to reach this goal, plenty of papers applied the optimization approaches instead of this classical methodology [13]. In fact, among them, we found that the synergy between ANN and GA was a good methodology in several studies [14]. Indeed, the GA was used mostly for the ANN structure optimization as well as for the ECG feature selection. These studies have been applied both of the ANN and GA for the ECG arrhythmia classification task, but with different manners. For example, in [15], authors proposed a neural classifier of premature ventricular contraction beats that uses a GA for the determination of the optimal connections between neurons.

Manuscript received September 5, 2018
Manuscript revised September 20, 2018

The proposed classifier achieved a higher recognition rate and sensitivity than classical classifiers. However, in [16], authors tested the genetic optimization of parameters and the genetic selection of features, for a problem of classification of 17 cardiac disorders. Author deduced that GA, as a feature selection method, increased the recognition sensitivity of the heart pathology and the classification times were significantly shortened

In this paper, we investigate a number of different feed forward neural network architectures for classifying five different ECG waveforms. They are divided into two main categories, Single Stage ANN and Two Stage ANN. Regarding the structural configuration, especially the determination of the number of hidden neurons, two methods are applied: the test and error method and the GA for network optimization. Whereas, for the feature selection, a GA based for feature dataset selection is evaluated. Then, a comparative analysis, between the obtained results, is done and discussed in order to select the most appropriate proposed network for this arrhythmia classification problem.

The following parts of this paper are organized as follows. Section2 presents the proposed machine learning approaches. Section3 describes the proposed methodology. Section4 describes the experimental work. Section5 shows the obtained results and Section6 summarizes the main deductions. At the end a conclusion is achieved.

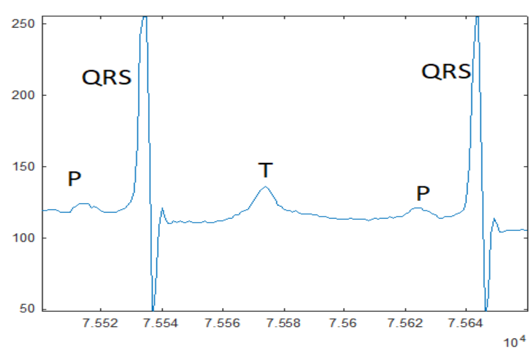

Fig. 1 Normal ECG signal.

2. Machine learning Approaches

In this study, we focus on the combination of the GA and the multilayer perceptron neural network (MLP), for the feature selection as well as for the neural network optimization problem, especially the determination of the hidden neurons number. The MLP and the GA are defined in the following sub sections.

2.1 Multilayer Perceptron Neural Networks (MLP)

In this work, all the networks have a MLP feed forward back propagation architecture, which has three layers (input layer, one hidden layer and an output layer) [17]. Each layer has a different number of neurons. Indeed, the input layer represents the feature vector where each component is attached to a neuron through a weight value. Then, features are multiplied by their corresponding weight. After that, they are summed in the hidden layer. Then, in order to give each neuron, its suitable output, they are directed to a transfer activation function (purelin, hardlim, sigmoid, logistic) in the output layer [17].

The Fig.2 defines the structure of a MLP network with one hidden layer, where X_n is the input feature vector and Y_p is the output feature vector.

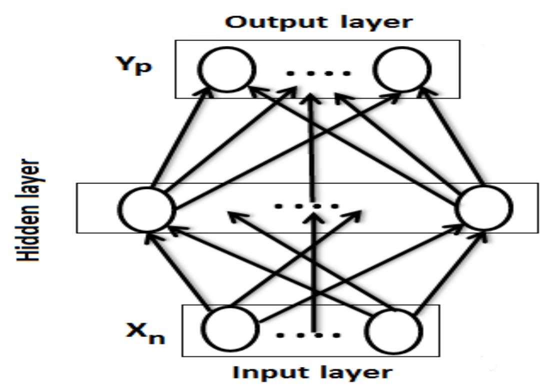

Fig. 2 MLP structure.

2.2 Genetic Algorithm (GA)

GAs were successfully applied for a number of problems in the field of arrhythmia classification [18]. They are mostly used to assist the ANN, into various schemes [19]: (1) using GA for feature selection, (2) using GA to select the learning rule or the learning parameters (the learning rate and the momentum), (3) using GA for the weight optimization, (4) using GA to define the ANN topology (the activation function, the hidden layer number, the hidden layer size.) and many other reasons.

In this investigation, GA is adopted into two different sections. It is applied in the preprocessing section, in order to select the most accurate feature subset and in the classification section for determining the optimal number of neurons in the hidden layer of the MLP network.

The Fig. 3 defines the standard genetic process as follow. Firstly, an initial population of n chromosomes containing binary bits, is generated randomly. Each chromosome represents a proposed solution to the problem. After decoding the vector of chromosomes, the population is evaluated using the appropriate fitness function defined for the optimization subject. The best solution is a vector that reaches the optimal value of the fitness function. Then, this vector is selected to the next generation. After that, in order to create the diversity of the population, crossover and mutation operators are applied, and new population is evaluated [19].

This process is iterated until finding the best solution provided before reaching the stopping criteria (the maximum generation number, the load limit...).

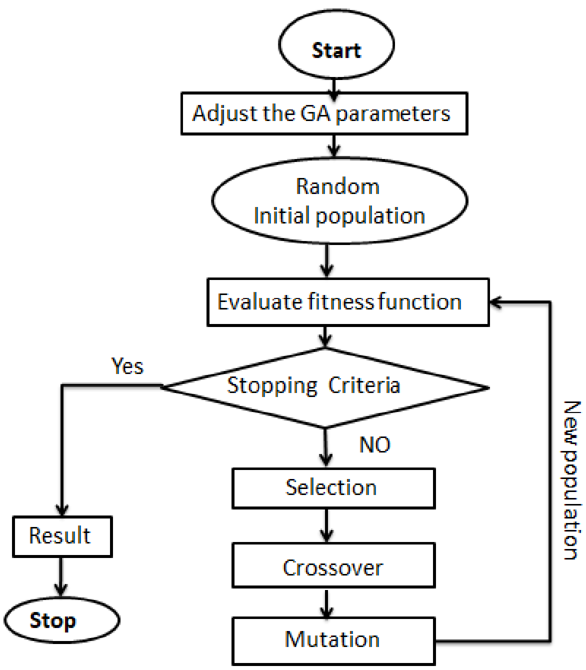

Fig. 3 Genetic Algorithm process.

2.2.1 Genetic algorithm for feature selection

Often, a successful classification task lies mostly with how the input feature vector is presented [20-21]. Therefore, by reducing the features which are redundant and erroneous, the performance of an ANN classifier is improved as well as its computation requirements is reduced.

In this research, among the entire features number; called N, the GA has to choose the most appropriate features according to the user desired feature dataset size.

Firstly, the user fixes the GA parameters and the desired number of the selected features; called m. Then, a large population of random chromosomes is created. Indeed, each chromosome in the population represents a proposed solution to the feature subset problem. Each one contains N genes of binary encoding. Every gene codes the feature index number as follows. The gene which designs a selected feature index takes a number of ones of binary bits and the non-selected one takes a number of binary bits of zeros.

Fig. 4 shows an example of a chromosome of N genes. Every gene is encoded into 6 binary bits. The first gene represents a selected feature and the last gene represents a non-selected one.

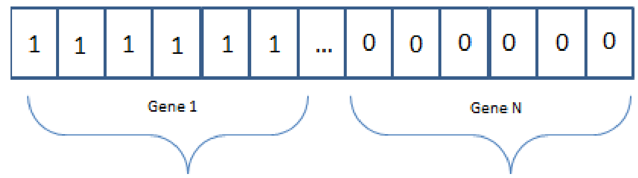

Fig. 4 Example of a chromosome in the population vector.

Then, each chromosome is tested in order to evaluate the fitness function and accordingly a fitness score is assigned to every evaluated chromosome. Otherwise, the elitist-strategy was used for the chromosome selection, where the best chromosomes are guaranteed to be selected for the next generation [22]. Besides, crossover and mutation events generate a new population and the process is repeated until the stopping criteria are reached.

In this investigation, the fitness function is represented by the entropy formula, as it is shown in equation (1), in order to determine the most valuable feature subset [23].

$$Entropy = -\sum p(x_j)\log 2\, p(x_j) \qquad (1)$$

According to equation (1), the entropy is minimal, equals 0, if the feature x_j appears only in one ECG class. We consider that this feature might has a good discriminatory power in the classification task. Conversely, the entropy is maximal, equals E_{max}, if x_j is not a good feature to represent the ECG class.

2.2.2 Genetic Algorithm for Network Optimization

In this research, a GA was used for the neural network optimization topology, especially for determining the optimal number of neurons in the hidden layer [24]. In fact, a number of chromosomes are randomly created to form the first population. Indeed, each chromosome proposes a number of hidden neurons for the ANN.

Often, this number belongs to an interval of decimal numbers, where the ANN escapes the over-fitting limit [25]. Then, each number is coded into a number of binary bits in order to form the chromosomes.

After that, by applying the MATLAB function, called newff, an ANN is created, with the GA proposed hidden neurons number and it is trained, accordingly [26]. Then, the fitness function for each chromosome is evaluated, where those which have the best fitness function are saved.

In this experiment, the fitness function is the computation of the accuracy (ACC) and the normalized mean square error (NMSE) coefficients of the neural classifier, as it is defined in the following equations.

- Accuracy

$$ACC = 100x \frac{TP + TN}{TP + FN + TN + FP} \quad (2)$$

Where TP: True Positive, TN: True Negative, FN: False Negative and FP: False positive. All of them are determined from the confusion matrix.

- Normalized Mean Square Error

$$NMSE = 1 - \frac{\left\| x_{ref}(:,i) - x(:,i) \right\|^2}{\left\| x_{ref}(:,i) - mean(x_{ref}(:,i) \right\|^2} \quad (3)$$

Where :

- ✓ x is the test data which is an N_S by N matrix, where N_S is the number of samples and N is the number of features.
- ✓ x_{ref} is the desired output (target) and it must have the same size as x .
- ✓ || Indicates the 2-norm of a vector.
- ✓ $NMSE$ is a row vector of length N .
- ✓ i = 1,...,N, where N is the number of features.

Fig.5 shows an example of a chromosome used for this subject, which contains the binary code of a proposed hidden neurons number.

1	1	0	1	1	1

Fig. 5 Example of a chromosome of 6 genes in the population.

Next, among a number of chromosomes, the GA selects a number of them, according to the wheel selection method, for the crossover and the mutation operators [22]. As a result, new chromosomes are produced and evaluated until the stopping criteria are reached.

3. Proposed Methodology

In this work, our goal is to select the more appropriate network which reaches the highest accuracy (ACC), the lowest Normalized Mean Square Error (NMSE), the smallest number of hidden neurons and an acceptable classification time response, for classifying 48 ECG signals into 5 cardiac classes, including: Normal beat (N), Paced beat (P), right Bundle Branch Block (RBBB), Left Bundle Branch Block (LBBB) and Premature Ventricular Contraction (PVC). As it is mentioned in Table1, a total of 32 signals were designed for the (N) class and 16 signals were divided equally between the other classes [27].

Table 1: Distribution of MIT-BIH database recordings.

ECG class	MIT_BIH ECG Recordings	Number
N	100, 101, 103, 105,106,108, 112, 113,114, 115, 116, 117, 121,122, 123, 201, 202, 203, 205, 209, 210, 213, 215, 219, 220, 221, 222, 223, 228, 230, 232, 234.	32
P	102, 104, 107, 217.	4
LBBB	109, 111, 207, 214.	4
RBBB	118, 124, 212, 231.	4
PVC	119, 200, 208, 233.	4
Total		48

In fact, 12 proposed ANN architectures are investigated. However, they are different from each other regarding the neural category and also the approaches adopted for the feature subset as well as for determining the optimal hidden neurons number.

In this investigation, the implemented arrhythmia classification methodology is as described in the Fig.6. It is decomposed into three main steps, as follows:

- ECG preprocessing which includes ECG feature extraction and selection sections: for part of the proposed networks, the feature selection is not applicate. Whereas, for the other, the GA is applied for this issue.
- Neural configuration regarding:
  - ✓ The topology of the neural classifier: either the Single Stage MLP or the Two Stage MLP category.
  - ✓ The approach for the determination of the number of hidden neurons in the single hidden layer: either the test and error method or the GA based network optimization.
- Classification which consists of training and testing the feature dataset for classifying the ECG signal into the five cardiac classes listed previously.

Subsequently, a comparison studies between the proposed neural networks results are done and discussed in order to select the more appropriate one for this arrhythmia task.

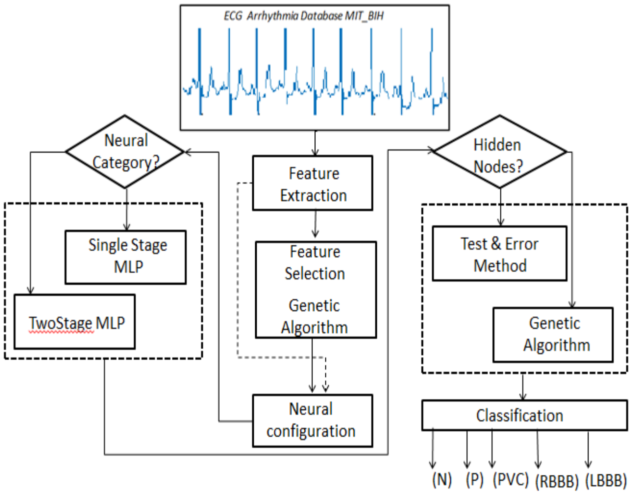

Fig. 6 The proposed methodology flow chart.

3.1. ECG Preprocessing

In this study, ECG waveforms from the MIT-BIH Arrhythmia database were used [27]. In fact, 48 ECG signals of one minute recordings were selected and categorized into 5 classes, listed previously.

3.1.1 Feature Extraction

Each ECG signal was presented as a feature vector. This vector is composed of two feature types [28]. The first one includes 48 features extracted from the Discrete Wavelet Transformer (DWT) with 8 decompositions of the mother wavelet Daubechies6. In fact, for each ECG waveform, 8 approximations and 8 details are saved. Then, for both of them, three statistical tools are computed, which are the mean, the standard deviation and the variance. However, the second one includes 14 morphological features containing ECG peaks (P, Q, R, S, T), time duration between waves (PR, QT, PT, ST, QRS and RR) and number of beat per minute (bpm). Hence the summation between the two feature types (48+14), gives a number of 62 features in the feature vector of each ECG signal [28].

3.1.2 Feature Selection

Regarding the number of features in the input feature vector, it depends on the feature selection method. Thus, it could be 62, if no feature selection method is applied or a number mine to 62, if a feature selection approach is applied. In this study, we have chosen the GA to do this subject, as it is described in section 2.

3.2 Neural Networks Classification

In this study, among the big number of possible ANN topologies, 12 topologies were proposed, as they are defined in Table 2. All of them are based on a feed-forward back-propagation MLP neural network and belong to two different ANN topologies: the Single Stage MLP and the Two Stage MLP.
The following flow chart shown in Fig.7, defines the methodology to differentiate the proposed networks whatever the ANN topology.

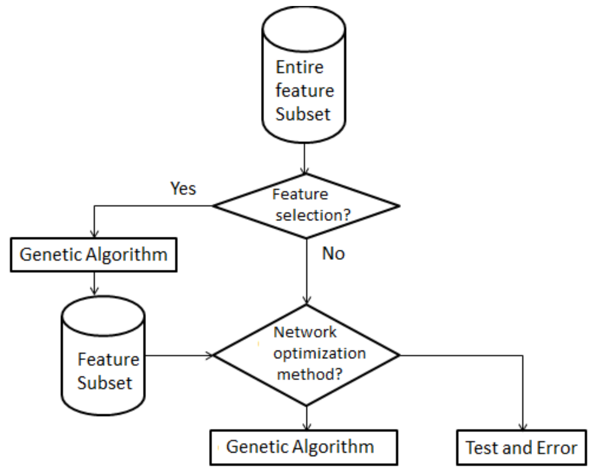

Fig. 7 Network identification methodology.

3.2.1 Single Stage MLP

In this category, we have proposed four networks. In fact, according to Table 2, both of Net1 and Net3 have 62 inputs, one hidden layer with X hidden neurons, where X is higher than 1 and minor to 35. For Net1 and Net4, a test and error method is applied to fix the more appropriate hidden node. Whereas for Net2 and Net4, a GA is applied to reduce the input vector and select the most suitable feature dataset. In addition, for Net2 and Net3, an optimization approach, based on the GA, is applied to choose autonomy the number of hidden neurons.
Fig.8 shows the Single Stage MLP architecture where Xi is the input feature vector.

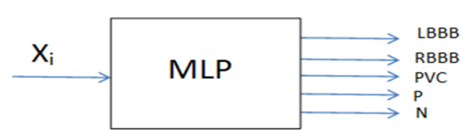

Fig.8 Single Stage MLP architecture.

3.2.2 Two Stage MLP

As it is discussed in [4], the unbalance between classes may damage the recognition efficiency of the cardiac abnormalities. As a result, we have opted for topology of the Two Stage MLP neural networks, in order to create the appropriate balance between classes [29]. Indeed, it consists of two MLPs networks which are similar to the Single Stage MLP category. However, they just have 2 outputs for the first MLP and 4 outputs for the second MLP. Therefore, as it is shown in Fig.9, MLP1 classifies the ECG waveforms into two classes: (N) and others unclassified (Y_i). Then MLP2 classifies (Y_i) into four classes: (P), (PVC), (LBBB) and (RBBB).
In this category, 6 topologies (Net5, Net6, Net7 Net8, Net9 and Net10) were proposed.
As it is described in Table2, the proposed networks are different from each other, since each one has its method for the feature selection task as well as for the determination of the hidden neurons number.

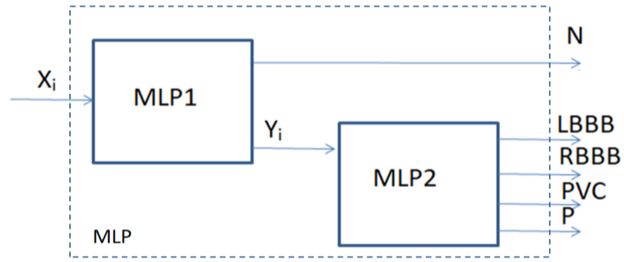

Fig. 9 Two Stage MLP architecture.

4. Experimental work

This work was done using the MATLAB R2017a environment. The computations were performed on an Intel Core i7, 4.0 GHz machine with 4GB of RAM.
In this study, four experiments were conducted for classifying the ECG signal into five cardiac classes: 1) Single Stage MLP networks, 2) Two Stage MLP networks, 3) test and error and genetic optimization for the determination of the number of hidden neurons and 4) GA based feature subset selection.
After normalization, data is divided into two datasets of training and testing. In this study 70% of the data was randomly assigned for ANN training and 30% for the test.
Concerning the target vector, it is composed of binary combination of 1 and 0 in order to distinct the five classes.

Table 2: The 12 proposed networks.

Network design	Topology		Input layer		Hidden layer		Output layer
			With FS	Without FS	Test and error	Genetic Optimization	
Single-Stage MLP	Net1			X	X		5
	Net2		X			X	
	Net3			X		X	
	Net4		X		X		
Two-Stage MLP	Net5	MPL1		X	X		2
		MLP2		X	X		4
	Net6	MLP1	X			X	2
		MLP2	X			X	4
	Net7	MLP1		X		X	2
		MLP2	X		X		4
	Net8	MLP1	X		X		2
		MLP2		X		X	4
	Net9	MLP1		X	X		2
		MLP2	X			X	4
	Net10	MLP1	X			X	2
		MLP2		X	X		4
	Net11	MLP1		X		X	2
		MLP2		X		X	4
	Net12	MLP1	X		X		2
		MLP2	X		X		4

4.1 Evaluation criteria

To evaluate the proposed networks, the following performances were determined and they are considered in the following priority: 1) the ACC (see equation (2)), 2) the NMSE (see equation (2)), and 3) the classification time (TR) as it is defined in equation (4).

$$TR = training\_time + testing\_time \quad (4)$$

Where training_time is the time depended for the training stage and testing_time is the time depended for the test stage. If all these performances are similar between two networks, we consider the number of hidden neurons (no/HL) in order to compare them.

4.2 GA Configuration

In this study, the GA deals with two problems: the feature dataset selection and the network optimization. For both of them, the GA parameter settings were adjusted as they are summarized in the Table3 and Table4, respectively.

Table 3: GA parameters for the feature dataset selection algorithm.

Parameters	
Number of selected features(m)	m<62
Population size	50
Number of generation	80
crossover	0.6
Probability of mutation	0.00001
Fitness function	Entropy

Table 4: GA parameters for the network optimization algorithm.

Parameters	
Number of variables	1 (one hidden layer)
Minimal number of hidden neurons	1
Maximal number of hidden neurons	35
Population size	50
Number of generation	25
Crossover rate	0.7
Mutation function	0.001
Fitness function	NMSE and ACC

4.3 Neural network Configuration

For the 12 proposed networks, one hidden layer is considered, a log-sigmoid and purlin transfer functions are selected for the hidden and the output layers, respectively and a Levenberg Marquard (trainLM) algorithm is applied for training the dataset.

5. Results

In this section, we present the obtained results which deals with: 1) the use of GA for the feature dataset selection, and 2) the networks comparison.

5.1 Feature Selection Results

By applying the GA based feature subset selection and according to the desired feature dataset size, we have obtained a number of selected datasets. In fact, the GA acquires each one after computing the entropy of its features. Thus, an evaluation of those feature datasets should be achieved by analyzing the performance criteria of the ANN which uses the selected feature dataset as the training set. In other words, if a feature dataset has n features, the ANN has n input neurons. In this experiment, we have used a Single Stage MLP network which its hidden neurons are obtained by the test and error method.

In Fig.10, we present the variation of the network performance criteria (ACC, NMSE, TR and no/HL) with respect to the number of the selected features.

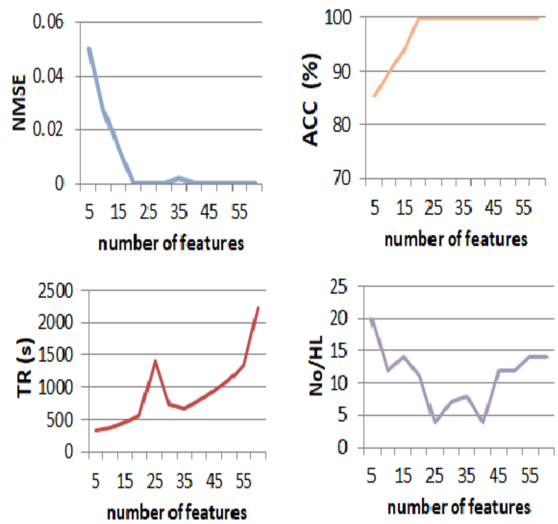

Fig. 10 Feature selection results.

Thus, it can be observed that the highest ACC (99.9%) and the lowest NMSE (2.07E-01) were obtained when the number of features >= 15. Afterwards, they are remained constant. However, the TR and the no/HL were fluctuating. Accordingly, we have chosen 30 as the number of selected features since it spends the lowest TR and uses an acceptable hidden neurons number.

As a result, it is found that the GA reduces the number of features required for classification by roughly 48.38%, since the length of the feature vector was reduced from 62 to 30 features.

5.2 Networks comparison

The initial experiment was performed using Single Stage MLP networks. The second experiment was executed using Two Stage MLP. Table5 and Table6 present the obtained results for the two networks categories, respectively. For the evaluation of each proposed network, the classification process was repeated 10 times. Then the average of each performance was saved. According to Table5 and Table6, two comparative studies are investigated. The first one is regarding the use of the GA for the feature selection task. However, the second one is about the impact of the test &

error and GA methods for determining the hidden neurons number.

5.3 Single Stage MLP Results

As it is defined in Table 2, Net1 and Net3 use the entire feature dataset (62 features). However, Net2 and Net4 use a selected one with 30 features. In addition, both of Net2 and Net3 apply the GA to optimize their topologies. Conversely, Net1 and Net4 apply the test and error method to choose the best values of their hidden neurons.

The results are summarized in Table 5. Referring to this table, two comparison studies are achieved.

The first comparison deals with Net2 and Net3. Indeed, Table 5 reveals that both of them use less number of hidden neurons (7 and 9 respectively) than Net1 and Net4 (25 and 15 respectively). In addition, they achieve the highest ACC (99.9%) and also the least NMSE values (2.88E-10 and 3.76E-10, respectively). However, regarding the NMSE and the TR results, Net2 (2.88E-10 and 7.39E+02s, respectively) outperforms Net3 (3.76E-10 and 1.78E+03s, respectively). As a result, when the GA is applied for optimizing the neural network topology, using the selected features dataset does not significantly affect the following performances: ACC, NMSE and number of hidden neurons. However, it affects mostly the classification time (TR).

The second comparison deals with Net1 and Net4. In fact, it is found that they achieve less accurate results, (91.7% and 93.7%, respectively) and consume less TRs (1.21E+02s and 7.95E+01s respectively) than Net2 and Net3.

Table 5: Single MLP networks test results.

Nets	Input	no/HL	NMSE	ACC (%)	TR(s)
Net1	62	25	3.43E-01	91.7	1.21E+02
Net2	30	7	2.88E-10	99.9	7.39E+02
Net3	62	9	3.76E-10	99.9	1.78E+03
Net4	30	15	2.07E-01	93.7	7.95E+01

However, Net4 uses less hidden neurons number (15), reaches less NMSE (2.07E-01) and less TR (7.95E+01s) than Net1 (25, 3.43E-01 and 1.21E+02s, respectively). Accordingly, results indicate that when the test and error method is applied to fix the neural network hidden neurons number, the use of selected input features significantly affects the ACC, the NMSE and the classification time.

Hence, referring to Fig.11, which shows the evolution of the Single Stage MLP networks performances (ACC, NMSE, TR and no/HL), it is found that among four proposed single MLP networks, the Net2 is the most appropriate network. Indeed, this network applies simultaneously the GA for fixing the hidden neurons number and also for the feature dataset selection task. It has better performances regarding the ACC (99.9%), the NMSE (2.88E-10) and an acceptable classification time (7.39E+02s), with 30 neurons in the input layer, 7 neurons in the hidden layer and 5 neurons in the output layer.

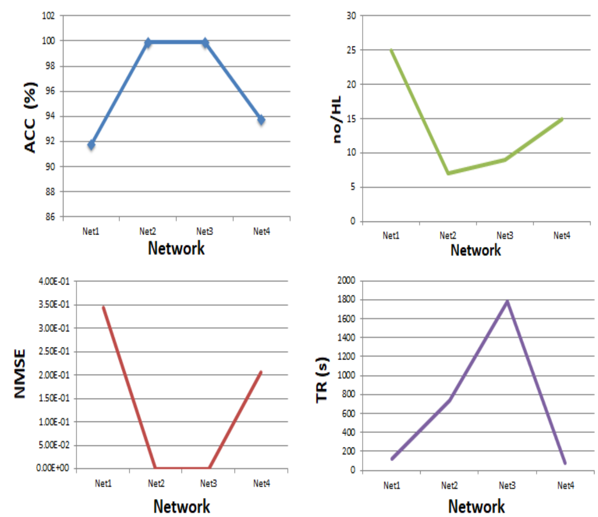

Fig. 11 Single Stage MLP results.

5.4 Two Stage MLP Results

As it is defined in Table2, Net5 and Net11 use the entire feature dataset (62 features). However, Net6 and Net12 use the selected feature dataset with 30 features. Conversely, for the determination of the hidden neurons number, Net6 and Net11 apply the GA while Net5 and Net12 apply the test and error method. Table2 indicates also that Net7, Net8, Net9 and Net10 are conversely configured. In other words, for each one, the first and the second MLP are conversely configured regarding the input dataset and the method to fix the hidden neurons number.

The results for the Two Stage MLP networks (Net5 until Net12) are summarized in Table 6. Three main comparisons studies are prepared.

The first comparison deals with Net6 and Net11. Indeed, Table 6 reveals that although they achieve the same average ACC (AVG_ACC) (99.9%), Net6 outperforms Net11 regarding the sum of the number of neurons (Sum_no/HL) (12), the average NMSE (AVG_NMSE) (4.33E-13) and the sum of the classification TR (Sum_TR) (7.94E+02s). Therefore, we can deduce that when the GA is applied for fixing the hidden neurons number, the use of the selected feature dataset does not affect the ACC, but it improves mostly the overall of the NMSE, the number of no/HL and the classification TR.

The second comparison deals with Net5 and Net12. In fact, it is found that they achieve less accurate results, (AVG_ACC= 92.7% for each one) and less AVG_NMSE values (1.41E-01 and 2.84E-01, respectively).

Appropriately, NET12 consumes less classification TRs (Sum_TR =6.97E+01s) than Net5 (Sum_TR=1.85E+02s). Therefore, when the test and error method is applied for the determination of the number of no/HL, the selected feature dataset affects mostly the total classification TR.
The third comparison deals with the networks conversely configured (Net7, Net8, Net9 and Net10). Accordingly, they achieve high values of AVG_ACC (96.8%, 95.8%, 95.8% and 96.8%, respectively) which outperform the AVG_ACC of Net5 and Net12 (92.7%). However, they underperform the AVG_ACC of Net6 and Net11 (99.9%). Furthermore, they reach acceptable AVG_NMSE values (1.48 E-01, 1.95E-01, 1.09E-01and 1.30E-01, respectively), acceptable Sum_TR values (3.79E+02 s, 1.17E+03s, 5.51E+02 and 6.99E+02, respectively) and also acceptable number of Sum_no/HL (28, 37, 44 and 23, respectively). But, their performances still minor to those of Net6. Consequently, using different methods and different feature datasets in the same Two Stage MLP, does not show an improvement in the obtained performances.
Hence, referring to Fig.12, it is clear that among the 8 proposed Two Stage MLP networks, it is indicated that Net6 produces the best overall performances including: the ACC (99.9%), the no/HL (12), the NMSE (4.33E-13) and the TR (7.94E+02s). For that reason, Net6 is considered the most appropriate network for this arrhythmia classification task.

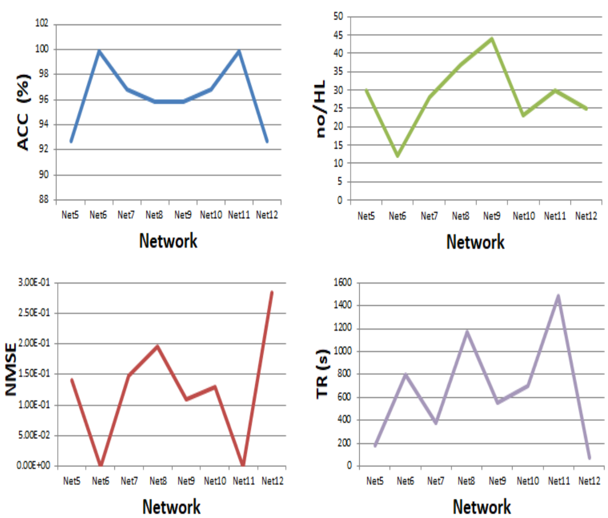

Fig. 12 Two Stage MLP results.

6. Discussion

The results of our investigation show the following deductions:
On the one side, regarding the Single Stage MLP networks, the number of neurons in the input layer have an effect on the ACC by applying the test & error method to fix the hidden nodes number (Net1, Net4), but the network with the entire input feature vector (Net1) requires the most number of hidden nodes (25 hidden nodes). This result indicates that a large number of features necessitates a large number of neurons in the hidden layer, when we apply the test& error method to configure the network. However, when the GA is applied for fixing the hidden neurons number (Net2, Net3), it was found that the feature selection task has not a significant impact on the ACC and the NMSE but it affects mostly the classification time.
On the second side, regarding the Two Stage MLP category, it is revealed that by breaking down the process into two classification stages, the evaluated performances (NMSE and ACC) are enhanced, especially when the selected features are used and the GA, for the determination of the hidden neurons number, is applied.

Table 6: Two Stage MLP networks test results.

Network		Input	no/HL	NMSE	ACC (%)	TR(s)	Sum_ no/HL	AVG_ NMSE	AVG _ ACC (%)	Sum _TR(s)
Net5	MLP1	62	10	2.19E-02	91.7	1.07E+02	30	1.41E-01	92.7	1.85E+02
	MLP2	62	20	2.61E-01	93.7	7.83E+01				
Net6	MLP1	30	3	8.66E-13	99.9	6.21E+02	12	4.33E-13	99.9	7.94E+02
	MLP2	30	9	9.39E-24	99.9	4.44E+02				
Net7	MLP1	62	13	5.77E-10	99.9	3.50E+02	28	1.48E-01	96.8	3.79E+02
	MLP2	30	15	2.96E-01	93.7	2.98E+01				
Net8	MLP1	30	20	3.89E-01	91.7	3.99E+01	37	1.95E-01	95.8	1.17E+03
	MLP2	62	17	2.70E-22	99.9	1.13E+03				
Net9	MLP1	62	10	2.19E-02	91.7	1.07E+02	44	1.09E-01	95.8	5.51E+02
	MLP2	30	9	9.39E-24	99.9	4.44E+02				
Net10	MLP1	30	3	8.66E-13	99.9	6.21E+02	23	1.30E-01	96.8	6.99E+02
	MLP2	62	20	2.61E-01	93.7	7.83E+01				
Net11	MLP1	62	13	5.77E-10	99.9	3.50E+02	30	1.72E-10	99.9	1.48E+03
	MLP2	62	17	2.71E-22	99.9	1.13E+03				
Net12	MLP1	30	10	3.89E-01	91.7	3.99E+01	25	2.84E-01	92.7	6.97E+01
	MLP2	30	15	2.19E-02	93.7	2.98E+01				

On the third side, the two ANN topologies (Single Stage MLP and Two Stage MLP) were compared, in order to find the impact of the neural network topology on the evaluation criteria. Therefore, by comparing Net3 and Net11, which have in common the use of non-selected features and the use of GA to fix the hidden neurons number, it is shown that both of them achieve the same ACC (99.9%). However, Net11 outperforms Net3, regarding the average of the NMSE (1.72E-10, 3.76E-10, respectively) and the TR value (1.48E+03s, 1.78E+03s). Whereas, concerning the number of hidden nodes, Net11 uses too many hidden nodes (no/HL=30) than Net3 (no/HL=9). Therefore, we can deduce that by using Two Stage MLP topology with non-selected features and applying the GA for the network optimization, the number of hidden nodes is affected badly, but the TR as well as the NMSE, are improved.

On the fourth side, we find that whatever the number of classes rises, the classification time rises also (see Table 5 and Table6). So, it is recommended to break down the task of 4 classes in to Two Stage MLP architectures.

On the last side, it is revealed that when the entire input feature vector is used, the number of hidden nodes is highly dependent on the number of classes in the data. In fact, with five classes (Net1), 25 hidden nodes are able to provide an ACC equals to 91.7% .However, as the number of classes decreases, less hidden nodes are employed. In fact for Net5, the first MLP, which classifies the input feature vector into 2 classes, needs only 10 hidden nodes to reach an ACC of 91.7%. Whereas the second MLP, which classifies ECG features into four classes, needs 20 hidden nodes to achieve an ACC of 93.7%.

Hence, among 12 proposed feed forward networks, it was found that the Two Stage MLP network (Net6), which applies the GA for the ECG feature subset selection and for determining the hidden neurons number, produces the best performance including the ACC and the NMSE. However, concerning classification TR, Net6 spends an acceptable classification time but not the best one. Accordingly, the selection of the most appropriate network depends on the priority criteria demanded by the users. Indeed, if users need a rapid classification response and less computation memory, they have to choose Net2, which achieves the same ACC (99.9%) as Net6, depends a less classification time (7.39E+02s) and uses less hidden neurons number(no/HL=7) than Net6(no/HL=12). However, when users demand only an accurate result and no matter how rapid the network and how large the memory, Net6 seems to be the best network, since it achieves the least NMSE (4.33 E-13).

7. Conclusion

In this paper, we compare two MLP topologies, including: the Single Stage MLP and the Two Stage MLP neural networks. They are evaluated in order to classify ECG signals into five classes. Whatever the MLP topology, we focus on the impact of using the GA for the feature subset selection as well as for the determination of the hidden neurons number. In fact, 12 different networks are investigated. Consequently, among the reached results, we cite that for both of them, the more accurate results are attained when the GA is executed for fixing the hidden neurons number. However, the rapider results are achieved when simultaneously; the GA is applied for the feature dataset selection and the test and error method is executed for the determination of the hidden neurons number.

Therefore, it is found that the Single stage MLP (Net2) and the Two Stage MLP (Net6), which use selected feature dataset and apply the GA for fixing the hidden neurons number, are the best networks. They achieve the same ACC (99.9%). However, Net6 is considered the most accurate network, since it achieves the least NMSE (4.33 E-13). While Net2 is considered as the rapidest network, since it spends less classification time (7.39E+02s). Thus, the choice of the most appropriate network depends mainly on the users' demanded criteria, regarding: the accuracy, the classification time and the computation memory. Thus, Net6 proves to be a suitable classifier especially when the analysis requires an accurate result and no matter of the classification time response. Conversely, Net2 proves to be the more appropriate one, if the classification time is the first requirement.

At the end, in this study, it is deduced that the use of the hybrid Two stage Neuro Genetic system, influences efficiently the accuracy of the neural classifier but it affects badly the classification TR and the number of hidden neurons. Hence, it is recommended to break down the second MLP in to different stages of different numbers of classes.

References

[1] Abdelghani, Samy A., Todd M. Rosenthal, and Daniel P. Morin, "Surface electrocardiogram predictors of sudden cardiac arrest," The Ochsner Journal, vol. 16, no.3, pp. 280-289, 2016.

[2] Haddadi, Rachid, et al., "ST Segment Analysis Using Wavelet Transform." International journal of computer science and network security, vol. 17.9, pp. 102-107, 2017.

[3] Rathi, M., and B. Narasimhan., "Data Mining, Soft Computing, Machine Learning and Bio-Inspired Computing for Heart Disease Classification/Prediction–A Review.", International Journal Advanced Research in Computer Science and Software Engineering, vol.7, no.4, April 2017.

[4] Luz, Eduardo José da S., et al. "ECG-based heartbeat classification for arrhythmia detection: A survey." Computer methods and programs in biomedicine, vol.127, pp. 144-164, April 2016.

[5] Li, Hongqiang, et al. "Genetic algorithm for the optimization of features and neural networks in ECG signals classification." Scientific reports, vol.7, 2017.

[6] Pławiak, Paweł. "Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system." Expert Systems with Applications, vol. 92, pp. 334-349, 2018.

[7] Elhaj, Fatin A., et al., "Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals." Computer methods and programs in biomedicine, vol. 127, pp. 52-63, April 2016.

[8] Mahapatra, Sakuntala, et al., "A neuro-fuzzy based model for analysis of an ECG signal using wavelet packet tree." Procedia Computer Science, vol. 92, pp. 175-180, 2016.

[9] Khalaf, Aya F., Mohamed I. Owis, and Inas A. Yassine., "A novel technique for cardiac arrhythmia classification using spectral correlation and support vector machines." Expert Systems with Applications, vol.42, no.21, pp. 8361-8368, 2015.

[10] Savalia, Shalin, Eder Acosta, and Vahid Emamian., "Classification of Cardiovascular Disease Using Feature Extraction and Artificial Neural Networks." Journal of Biosciences and Medicines, vol. 5, no.11 , pp 64, 2017.

[11] Alshamrani, Barakat Saeed, and Ahmed Hamza Osman., "Investigation of Hepatitis Disease Diagnosis using Different Types of Neural Network Algorithms." International Journal of Computer Science and Network Security (IJCSNS), vol.17, no.2, pp 242, 2017.

[12] Šter, Branko, and Andrej Dobnikar., "Neural networks in medical diagnosis: Comparison with other methods." International Conference on Engineering Applications of Neural Networks, pp. 427-457, 1996.

[13] Syahputra, Ramadoni., "Distribution Network Optimization Based on Genetic Algorithm." Journal of Electrical Technology UMY, vol. 1, no.1, pp. 1-9, 2017.

[14] Arabasadi, Zeinab, et al., "Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm." Computer methods and programs in biomedicine, vol. 141, pp. 19-26, 2017.

[15] Rahman, Md Mijanur, and T. Akter Setu., "An implementation for combining neural networks and genetic algorithms." Int. J. Comput. Sci. Technol, vol.6, no.3, 2015.

[16] Sekkal, Mansouria, Mohamed Amine Chikh, and Nesma Settouti., "Evolving neural networks using a genetic algorithm for heartbeat classification." Journal of medical engineering & technology, vol. 35, no.5, pp. 215-223, 2011.

[17] Varoquaux, Gaël, et al., "Assessing and tuning brain decoders: cross-validation, caveats, and guidelines." NeuroImage , vol.145, pp. 166-179, 2017.

[18] Li, Hongqiang, et al.,, "Genetic algorithm for the optimization of features and neural networks in ECG signals classification." Scientific reports, vol. 7, 2017.

[19] Whitley, Darrell., "Genetic algorithms and neural networks." Genetic algorithms in engineering and computer science, vol. 3, pp. 203-216, 1995.

[20] Saeys, Yvan, Iñaki Inza, and Pedro Larrañaga. "A review of feature selection techniques in bioinformatics." Bioinformatics, vol.23, no.19, pp. 2507-2517, 2017.

[21] Leardi, Riccardo., "Genetic algorithms in feature selection." Genetic algorithms in molecular modeling, pp. 67-86, 1996.

[22] Deb, Kalyanmoy, et al., "A fast and elitist multiobjective genetic algorithm: NSGA-II." IEEE transactions on evolutionary computation , no.2 , pp. 182-197, 2002.

[23] Liu, Huiqing, et al., “Using amino acid patterns to accurately predict translation initiation sites.” In silico biology, vol.4, no.3, pp. 255-269, 2004.
[24] MirRokni, Mahshid Kaviani1 Seyed Majid. “Applying Genetic Algorithm in Architecture and Neural Network Training.” International Journal of Computer Science and Network Security IJCSN, vol.17, no.6, p.118, June 2017.
[25] Tetko, Igor V., David J. Livingstone, and Alexander I. Luik., “Neural network studies. 1. Comparison of overfitting and overtraining.” Journal of chemical information and computer sciences , vol.35, no.5,pp. 826-833, 1995.
[26] Zhao, Zhizhong, et al., “Application and comparison of BP neural network algorithm in MATLAB.” IEEE International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), vol. 1, 2010.
[27] Moody, and R. G. Mark, “The impact of the MIT-BIH arrhythmia database,” IEEE Engineering in Medicine and Biology Magazine, vol. 20, no.3, pp. 45-50, 2001.
[28] H.Lassoued,,and R. Ketata, “Artificial neural network classifier for heartbeat detection,” In Proceedings of the International Conference on Automatic and Signal Processing (ATS), Mars 2017,Tunisia.PET Journal:2017.
[29] Hosseini, H. Gholam, Dehan Luo, and Karen Jane Reynolds. "The comparison of different feed forward neural network architectures for ECG signal diagnosis." Medical Engineering and Physics, vol. 28, no.4, pp. 372-378, 2006.

Hela Lassoued was born on Mai 1987 in Tunisia. She received her degree in industrial computing and automatic engineer in 2013, from the “Applied Science and Technology Institute (INSAT), Carthage University, Tunisia”. She is currently a PhD Student in the “Electronic, Robotic, Control and Optimization research laboratory” (E.R.C.O) in the “Applied Science and Technology Institute ” (INSAT), Tunisia. Her current research interests are in the machine learning fields especially the artificial neural networks, the fuzzy logic systems and the neuro fuzzy systems.

Raouf Ketata was born in Sfax, Tunisia in 1961. He received the Engineering Diploma in Electrical Engineering and Instrumentation from the National School of Engineer in Sfax-Tunisia in 1987, the Master degree in Signal Treatment, Computer Science and Automatic in 1988, the Ph.D. degree in Industrial Computer Science and Automatic in 1992 from National Institute of Applied Science in Toulouse-France and the University Habilitation in Industrial Computer Science from the National Institute of Applied Science and Technology - Tunisia. From 1988 to 1992, he had a Contract Assistant Post in National Institute of Applied Science in Toulouse-France. From 1992 to 1994, he had a Conference Professor Post in National Atlantic School of Engineer in Genie Electric-France. Between 2010 and 2012, he was appointed Director of the department of Physic and Instrumentation Engineering at the National Institute of Applied Science and Technology – Tunisia. He is currently a Conference Professor in National Institute of Applied Science and Technology - Tunisia and member of the Research Laboratory on Energy, Robotic, Control and Optimization (ERCO). His current research interests lie in fuzzy logic, neural networks, genetic algorithms, intelligent systems and their applications in machine learning, classification, intelligent control.