

Biogeography-Based Optimization of Neuro-Fuzzy System Parameters for Diagnosis of Cardiac Disease

Mirela Ovreiu
The Cleveland Clinic
Cleveland, Ohio
ovreium@ccf.org

Dan Simon
Cleveland State University
Cleveland, Ohio
d.j.simon@csuohio.edu

ABSTRACT

Cardiomyopathy refers to diseases of the heart muscle that becomes enlarged, thick, or rigid. These changes affect the electrical stability of the myocardial cells, which in turn predisposes the heart to failure or arrhythmias. Cardiomyopathy in its two common forms, dilated and hypertrophic, implies enlargement of the atria; therefore, we investigate its diagnosis through P wave features. In particular, we design a neuro-fuzzy network trained with a new evolutionary algorithm called biogeography-based optimization (BBO). The neuro-fuzzy network recognizes and classifies P wave features for the diagnosis of cardiomyopathy. In addition, we incorporate opposition-based learning in the BBO algorithm for improved training. First we develop a neuro-fuzzy model structure to diagnose cardiomyopathy using P wave features. Next we train the network using BBO and a clinical database of ECG signals. Preliminary results indicate that cardiomyopathy can be reliably diagnosed with these techniques.

Categories and Subject Descriptors

J.3 [Life and Medical Sciences]: Medical Information Systems

General Terms

Algorithms

Keywords

Evolutionary algorithm, biogeography-based optimization, neuro-fuzzy system, electrocardiogram

1. INTRODUCTION

Cardiovascular diseases are the major cause of death in the western world, resulting in more than 800,000 deaths per year in the United States alone [1]. One in five Americans has some form of cardiovascular disease [20]. Cardiomyopathy is a significant clinical problem which is mainly generated by volume/diastolic overload. To accommodate the increased volume of blood, the heart chambers may stretch or dilate. Valvular regurgitation and congestive heart failure are two conditions that contribute to chamber dilation.

Cardiomyopathy is generally diagnosed by an echocardiograph investigation, which is a sonogram of the heart. But an electro-

cardiographic (ECG) investigation is always part of a cardiologic work-up. The ECG represents the deflection of ionic current across myocardial cell membranes and through the extra-cellular space of the tissues of the thoracic cavity. The ECG, in competition to many other techniques, retains an important role in diagnosis and prognosis of cardiovascular diseases.

It has been suggested that cardiomyopathy is reflected in modification of ECG characteristics such as P wave morphology. Previous statistics-based attempts to classify cardiomyopathy from ECG data have been underwhelming [16], [17], but we hypothesize that these limitations can be overcome using a hybrid neuro-fuzzy classification model. To test this hypothesis and direct the results to patient care, we report the following results in this paper: first we design a neuro-fuzzy model to diagnose cardiomyopathy, and then we train the network using an acquired clinical database of ECG signals.

Neuro-fuzzy systems can be trained with derivative-based methods like gradient descent [5], [13] or with evolutionary algorithms (EAs) such as genetic algorithms and swarm intelligence [11]. EAs have the advantage of not requiring derivative information, and have less likelihood of getting stuck in a local optimum. Here we use a new biologically motivated optimization algorithm called biogeography-based optimization (BBO) [30] to train the neuro-fuzzy ECG classification network. We also incorporate opposition-based learning in the BBO algorithm [8] for better classification.

Section 2 provides preliminary background information about this research, including discussions of cardiomyopathy, neuro-fuzzy networks, BBO, and opposition-based learning. Section 3 discusses ECG data collection, the neuro-fuzzy classification network, the BBO training algorithm, experimental results, and suggestions for neuro-fuzzy model refinement. Section 4 provides conclusions and directions for future work.

2. BACKGROUND

2.1 Cardiomyopathy

The term “cardiomyopathy” defines a group of diseases primarily affecting the cardiac muscle by weakening it or changing its structure. Cardiomyopathy can be acquired or inherited, and in many cases its cause is unknown. Hypertrophic cardiomyopathy is inherited and is supposed to be a result of defects of the genes that regulate heart muscle growth. In dilated cardiomyopathy, which is the acquired type, two common conditions contributing to chamber dilation are valvular regurgitation and congestive heart failure. Studies have shown that the number of myocardial fibers in an adult human heart is essentially constant and equal across the population. Abnormal cardiac enlargement can thus be achieved only by an increase in the length or diameter of existing cardiac muscle cells [6]. Cardiomyopathy, through electrical

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
GECCO'10, July 7–11, 2010, Portland, Oregon, USA.
Copyright 2010 ACM 978-1-4503-0073-5/10/07...\$10.00.

instability of myocardial cells due to structural or metabolic integrity loss, is associated with cardiac conduction abnormalities that can degenerate to arrhythmia or heart failure [18].

The history of cardiomyopathy research reveals the evolution of analytic and diagnostic capabilities. Due to the critical role of the left ventricle, initial studies focused only on the electrocardiographic features of the hypertrophic left ventricle [32]. The QRS complexes and T waves, as the reflection of ventricular depolarization and repolarization respectively, were properly and precisely analyzed [39]. In the study by Sox et al., citing the Framingham Study, the left ventricular hypertrophy is defined by a prolonged ventricular activation period of 0.05 seconds, tall R waves, depressed ST segments, and inverted T waves [32]. Ziegler was the first to introduce T wave analysis to the characterization of left ventricular hypertrophy; that study presented different patterns of the QRS and T configurations into left or right precordial limb leads [39].

Other studies extend the analysis of other cardiac chambers to cardiomyopathy and also relate P wave changes to this disease. The P wave portrays atrial electrical activity and it is natural to assume that changes in the atrial action potential and substrate will be reflected in changes in P wave timing or morphology [4]. Bahl et al. presented the P wave changes associated with the type and stage of the disease [2]. Analyzing each of the four chamber enlargements, Johnson et al. emphasized P wave changes of duration and morphology for an enlarged left atrium as well as amplitude for an enlarged right atrium [10].

The atria, as chambers with relatively thin walls, respond to volume and pressure overload due to dilatation. Moreover, the enlargement of the associated ventricle is recognized as the cause of the enlargement of the atrium [16], [17]. The right atrium enlargement is recognized by the increased amplitude of the P wave (0.25 mV) while the left atrial abnormality is reflected by lengthened P wave duration (greater than 120 ms) as well as a notched P wave.

In a recent document, the American Heart Association, the American College of Cardiology Foundation, and the Heart Rhythm Society, review the current literature and conclude on recommendations and standards to be used when interpreting ECG data related to cardiomyopathy [9]. The changes of ECG associated with cardiac chamber hypertrophy are delineated and discussed for each disease localization. In left ventricular hypertrophy, the P wave shape is mentioned as a criterion. In right ventricular hypertrophy, a P wave amplitude larger than 0.25 mV in lead II is presented as a classification index. Left atrial abnormality implies a prolongation of the total atrial activation time (P wave duration longer than 120 ms), widely notched P wave, and possibly P wave area. The right atrial abnormality list includes an increase in the amplitude of the P wave (greater than 0.25 mV) and a prolongation of the P wave after cardiac surgery, which is the case for our patients.

2.2 Neuro-Fuzzy Networks

Consider a multi-input, single-output fuzzy logic system. Our discussion can be easily generalized to multiple output systems, but restricting our discussion to single-output systems simplifies the notation considerably. In addition, the ECG classification system that we consider in this paper is single-output. The i th rule R_i of the fuzzy system can be written as follows [5].

$$R_i : \text{If } x_1 \text{ is } A_{i1} \text{ and } \dots \text{ and } x_m \text{ is } A_{im} \text{ then } \quad (1)$$

$$y = z_i(x), \quad (i = 1, \dots, p).$$

The inputs x_i and the output y are linguistic variables, A_{ij} are fuzzy sets, and $z_i(x)$ is a function of the input $x = [x_1 \dots x_m]^T$. The output function $z_i(x)$ typically takes one of the following forms: (1) singleton, (2) fuzzy set, (3) linear function. If the fuzzy system uses center average defuzzification, product inference, and singleton fuzzification, then $z_i(x) = z_i$ (a singleton) and the fuzzy system output can be written as

$$y = \frac{\sum_{i=1}^p z_i \prod_{j=1}^m \mu_{ij}(x_j)}{\sum_{i=1}^p \prod_{j=1}^m \mu_{ij}(x_j)} \quad (2)$$

where $\mu_{ij}(x_j)$ denotes the degree of membership of x_j in R_i . As in many neuro-fuzzy networks, we use a Gaussian form for μ_{ij} :

$$\mu_{ij}(x_j) = \exp\left(\frac{-(x_j - c_{ij})^2}{\sigma_{ij}^2}\right) \quad (3)$$

where c_{ij} is the j th element of the center of the i th rule, and σ_{ij} is its standard deviation. In this case, Eq. (2) becomes

$$y = \frac{w}{\sum_{i=1}^p m_i(x)} \quad (4)$$

$$w = \sum_{i=1}^p z_i m_i(x) \quad (5)$$

$$m_i(x) = \exp[-(x - c_i)^T P^{-2} (x - c_i)] \quad (6)$$

where $c_i = [c_{i1} \dots c_{im}]^T$ and $P = \text{diag}(\sigma_1, \dots, \sigma_m)$. Eq. (5) is in the form of a radial basis function, which is a type of neural network [5]. The system of Eqs. (5) and (6) is therefore called a neuro-fuzzy system. It can be depicted as shown in Figure 1.

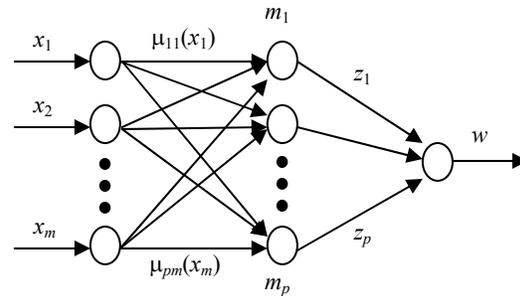


Figure 1. Multi-input single-output neuro-fuzzy system architecture.

The neuro-fuzzy system in Figure 1 is a function of the $p \times m$ elements of the membership centers c_{ij} , the $p \times m$ elements of the membership standard deviations σ_{ij} , and the p elements of the singleton outputs z_i . There are thus $p(2m+1)$ parameters that define the neuro-fuzzy system. For a given neuro-fuzzy system architecture and a given training set of input/output data, the neuro-fuzzy system parameters can be optimized with respect to these $p(2m+1)$ parameters.

2.3 Biogeography-Based Optimization

Biogeography-based optimization (BBO) is a recently-developed EA [30]. As its name implies, BBO is motivated by biogeography, which is the study of the distribution of species over time and space [14], [38]. BBO has demonstrated good performance on various benchmark functions [15], [30]. It has also been successfully applied to several real-world optimization problems, including sensor selection [30], power system optimization [22], groundwater detection [12], and satellite image classification [21].

Given an optimization problem and a population of candidate solutions (individuals), a BBO individual with high fitness is likely to share features with other individuals, and an individual with low fitness is unlikely to share features. Conversely, an individual with high fitness is unlikely to accept features from other individuals, while an individual with low fitness is likely to accept features. Solution feature sharing, which is called immigration and emigration, tends to improve the individuals and thus evolve a good solution to the problem.

Each BBO individual has an immigration rate λ_i and emigration rate μ_i . A fit individual has high μ and low λ , and vice versa for a poor solution. The immigration and emigration rates are functions of the individual's fitness. They are often calculated as

$$\begin{aligned} \lambda_i &= f_i/n \\ \mu_i &= 1 - \lambda_i \end{aligned} \quad (7)$$

where n is the population size and f_i is the fitness rank of the i th individual (the most fit individual has a rank $f_i = 1$). The immigration rates λ_i are used as immigration probabilities. The emigration rates μ_i are proportional to fitness and are used in a roulette-wheel type of algorithm to determine the emigrating solution in case immigration is selected for a solution.

Although the migration rates in Eq. (7) are linear with respect to fitness rank [30], nature-inspired migration rates which are sigmoid with respect to fitness rank generally seem to give better optimization performance [15]. In this paper we retain the original linear migration rates for the sake of simplicity.

As with other EAs, mutation is typically implemented to increase exploration, and elitism is often used to retain highly fit solutions. The standard BBO algorithm is shown in Figure 2.

```

For each solution  $H_i$ 
  For each solution feature  $s$ 
    Select solution  $H_i$  with probability proportional to  $\lambda_i$ 
    If solution  $H_i$  is selected then
      Select  $H_j$  with probability proportional to  $\mu_j$ 
      If  $H_j$  is selected then
         $H_i(s) \leftarrow H_j(s)$ 
      end
    end
  next solution feature
  Probabilistically mutate  $H_i$ 
next solution

```

Figure 2. One generation of the standard BBO algorithm.

2.4 Oppositional BBO

Opposition-based learning (OBL) has been introduced as a method that can be used by EAs to accelerate convergence speed

by comparing the fitness of an individual to its opposite and retaining the fitter one in the population [25], [34]. The “opposite” of an individual is defined as the reflection of that individual’s features across the midpoint of the search space. Opposition-based differential evolution (ODE) [24] was the first application of OBL to EAs. OBL was first incorporated in BBO in [8] and was shown to improve BBO by a significant amount on standard optimization benchmarks.

Given an EA population member x , there are at least three different types of oppositional points that can be defined. These oppositional points are referred to as the opposite x_o , the quasi-opposite x_q , and the quasi-reflected-opposite x_r . Figure 3 illustrates these points for an arbitrary x in a one-dimensional domain. The point c is the center of the domain, x_o the reflection of x across c , x_q is a randomly generated point from a uniform distribution between c and x_o , and x_r is a randomly generated point from a uniform distribution between x and c .

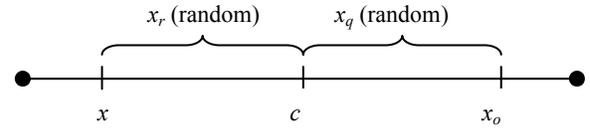


Figure 3. One-dimensional illustration of an arbitrary EA individual x , its opposite x_o , its quasi-opposite x_q (randomly chosen in $[c, x_o]$), and its quasi-reflected-opposite x_r (randomly chosen in $[x, c]$).

OBL is essentially a more intelligent way of implementing exploration instead of generating random mutations. Another way of viewing OBL is from the perspective of social revolutions in human society. Society often progresses on the basis of a few individuals who embrace philosophies that are not just random, but that are deliberately contrary to accepted norms. Given that an EA individual is described by the vector x , and that the solution to the optimization problem is uniformly distributed in the search domain, it is shown in [25] that x_q is probably closer to the solution than x or x_o . Further, it is shown in [8] that x_r is probably closer to the solution than x_q . These results are nonintuitive, but probability is often nonintuitive, and the OBL results are derived not only analytically but also using simulation.

In this paper we use oppositional BBO (OBBO) to train the neuro-fuzzy ECG classification network. Suppose that the population size is N . OBBO works by generating a population of N opposite individuals which are the opposite of the current population. Then, given the entire $2N$ individuals comprised of both the original and the opposite populations, the best N individuals are retained for the next population. However, this does not occur at each generation. Instead it occurs randomly with a probability of J_r at each generation. J_r is called the jump rate. Based on [23] we use $J_r = 0.3$ in this paper.

In order to increase the likelihood of improvement at each generation we implement OBBO as follows. At each generation, we save the original population of N individuals before creating a population of N new individuals via migration. We then create an opposite population of N additional individuals if indicated by the jump rate. Of the total $2N$ or $3N$ individuals, we finally select the best N for the next generation. Note that this approach guarantees that the best individual in each generation is at least as good as that of the previous generation. This is similar to a $(\mu+\lambda)$

evolutionary strategy [7], whose parameters are not to be confused with the μ and λ migration parameters in BBO. The resulting OBBO algorithm is summarized in Figure 4.

```

 $H^{(1)} \leftarrow H$  (make a copy of the population  $H$ )
For each solution  $H_i \in H^{(1)}$  ( $i = 1, \dots, N$ )
  For each solution feature  $s$ 
    Select solution  $H_i$  with probability proportional to  $\lambda_i$ 
    If solution  $H_i$  is selected then
      Select  $H_j$  with probability proportional to  $\mu_j$ 
      If  $H_j$  is selected then
         $H_i(s) \leftarrow H_j^{(1)}$ 
      end
    end
  end
  next solution feature
  Probabilistically mutate  $H_i$ 
next solution
if  $\text{rand}(0,1) < J_r$  then
  Use  $H$  to create an  $N$ -member opposite population  $H^{(2)}$ 
else
   $H^{(2)} = \emptyset$ 
end
Copy the best  $N$  individuals from  $\{H, H^{(1)}, H^{(2)}\}$  to  $H$ 

```

Figure 4. One generation of oppositional BBO (OBBO).

3. EXPERIMENTAL RESULTS

3.1 Data Collection

In preparation for the testing of a cardiomyopathy diagnosis model, a database of long-duration ECG signals was collected. The database includes signals from 55 subjects, 18 of them with cardiomyopathy. Not all subjects experienced chronic or paroxysmal atrial fibrillation. The cardiomyopathy group contained 10 males and 8 females with a mean age of 54 (range 23–88) years. The control group contained 22 males and 15 females with a mean age of 60 (range 27–77) yrs. The inclusion criteria were the same for both groups: no chronic or paroxysmal atrial fibrillation and no perioperative pacing.

ECG parameters describing P wave morphology were computed for each minute of data recording for all 55 patients in the training data set. This set of ECG parameter values constitutes the input component of the training data set for neuro-fuzzy model development. For additional details of ECG parameter computation algorithms see [3], [35], [36], [37].

The ECG P wave reflects the electrical activity of the atria and may indicate the existence of irregularities in electrical conduction. Using a previously developed P wave detection method, the starting, ending, and maximum points of the P wave were determined [36]. The average P wave morphology parameters were computed once per minute. The P wave morphology parameters included the following:

- Duration;
- Amplitude;
- A shape parameter which represents mono- or bi-phasicity;
- Inflection point, which is the duration of the P wave between the onset and the peak points;
- Energy ratio, defined as the fraction between the right atrial excitation energy and the total atrial excitation energy.

Initial investigation revealed that the monophasicity / biphasicity parameter did not vary appreciably between cardiomyopathy and control patients. We therefore discarded the monophasicity / biphasicity parameter from our data set. Differences between the remaining P wave morphology parameters for cardiomyopathy and control patients in the training database are presented in Figure 5. Based on the standard deviation bars, there is apparently important information included in these parameters. Their usefulness in identifying the patients with cardiomyopathy is determined by the proposed neuro-fuzzy model as discussed in the following section.

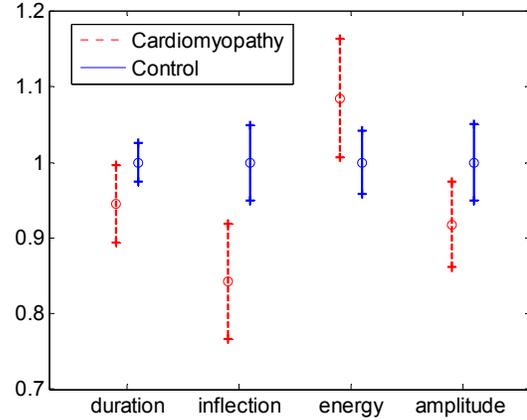


Figure 5. P wave characteristics of cardiomyopathy and control patients, normalized to the mean values of the control patients. Error bars show one standard deviation.

3.2 Experimental Setup

Cardiomyopathy diagnosis is performed by a multivariate, neuro-fuzzy classification model that uses current values of ECG P wave parameters to generate a cardiomyopathy classification index. The initial model is a multi-input single-output fuzzy inference system with a three-layer architecture (Figure 1). The fuzzification layer takes crisp parameter values and determines their memberships in linguistic categories (low, medium, high, etc.). Each of these fuzzy variables are then input to each node of the fuzzy rule layer (i.e., the middle layer shown in Figure 1). The model output, which is the cardiomyopathy classification index, is the weighted average of the output rules.

Since we have four inputs (see Figure 5), we have $m = 4$ in Figure 1. The number of middle-layer neurons is equal to p and should be chosen as a tradeoff between good training performance and good generalization. If p is too small then training performance will be poor because we will not have enough degrees of freedom in the neuro-fuzzy network. If p is too large then test performance will be poor because the training algorithm will tend to “memorize” the training inputs rather than obtaining a good generalization for test data.

The output y shown in Eq. (4) is chosen to be +1 for cardiomyopathy patients and -1 for control patients. The ECG database is used for training and the output of the neuro-fuzzy system is compared to the known classification of the ECG patient. The RMS training error is defined as

$$E = \sqrt{\frac{1}{N} \sum_{i=1}^N (d_i - y_i)^2} \quad (8)$$

where N is the number of training inputs, d_i is the desired output of the i th training datum (+1 or -1), and y_i is the corresponding neuro-fuzzy output. In order to determine the best value of p (the number of middle-layer neurons) we run 10 Monte Carlo simulations with various values for p and compare training and testing errors. The BBO parameters that we use are as follows:

- Population size = 200
- Mutation rate = 2% per solution feature
- Generation limit = 50

Mutation is implemented by randomly generating a new parameter from a uniform distribution between the minimum and maximum parameter bounds. The parameter bounds that we use are as follows:

- Output singletons $z_i \in [-10, +10]$
- Membership centroids $c_{ij} \in [0, \pi]$
- Membership standard deviations $\sigma_{ij} \in [0.01, 5]$

We use ECG data from 55 test subjects as described in Section 3, which includes 37 control patients and 18 cardiomyopathy patients. We randomly divide the patients into approximately equal numbers of training patients and test patients. We therefore have 9 cardiomyopathy patients and 19 control patients for training the network, and 9 cardiomyopathy patients and 18 control patients for testing the network. We randomly choose 200 ECG data points from a 700-minute time interval for each patient for both training and testing. Therefore, we have $200 \times (9+19) = 5600$ data points for training, and $200 \times (9+18) = 5400$ data points for testing.

3.3 Parameter Tuning and Test Results

Table 1 shows the minimum training error attained as specified in Eq. (8) for various numbers of middle-layer neurons, along with the resulting correct classification rate for training and testing. An ECG data point is classified as cardiomyopathy if the neuro-fuzzy output $y > 0$, and control if the neuro-fuzzy output $y < 0$. The quantity of primary interest is the correct classification rate for the test data, and Table 1 shows that this is attained with 3 middle-layer neurons. Fewer neurons gives too few degrees of freedom, and more neurons results in a tendency of the neuro-fuzzy system to overfit the training data and hence not provide adequate generalization for the test data.

Table 1. Training error and correct classification rate (CCR) for training and testing as a function of the number of middle layer neurons p .

p	Training Error		Train CCR (%)		Test CCR (%)	
	Best	Mean	Best	Mean	Best	Mean
2	0.85	0.88	76	72	66	58
3	0.77	0.84	82	77	75	62
4	0.78	0.83	84	77	65	55
5	0.78	0.83	82	76	63	58

Next we implement OBBO to explore the effect of OBL on classification performance. Table 2 shows results for three different OBL options: standard BBO, quasi-opposition BBO (Q-BBO), and quasi-reflected BBO (R-BBO). We use the same

population size, mutation rate, and generation limit as discussed earlier. We use 3 middle-layer neurons as indicated by Table 1. Table 2 shows that OBBO using quasi-opposition provides the best neuro-fuzzy classification performance when test performance is used as the criterion.

Table 2. Training error and correct classification rate (CCR) for training and testing for alternative implementations oppositional BBO.

	Training Error		Train CCR (%)		Test CCR (%)	
	Best	Mean	Best	Mean	Best	Mean
BBO	0.77	0.86	84	76	66	58
Q-BBO	0.83	0.86	79	74	69	62
R-BBO	0.80	0.85	81	75	65	60

Note that the numbers in Tables 1 and 2 do not match exactly because they are the results of different sets of Monte Carlo simulations. Note that the $p=3$ row in Table 1 is the same experiment as the BBO row in Table 2. However, different random number seeds were used for the two experiments. Ideally the two rows should be identical, but due to different random number seeds in the 10 Monte Carlo runs, the two rows are somewhat different. The fact that two different sets of experiments performed under the same conditions lead to somewhat different results means either that we did not use enough Monte Carlo runs of the experiments, or the variance of results is relatively high. In future work we will analyze variance to allow for a better assessment of the significance of the results.

After settling on Q-BBO with 3 middle-layer neurons, we explore the effect of mutation rate on Q-BBO performance. Table 3 shows neuro-fuzzy results for various mutation rates. We use the same population size and generation limit as before. Table 3 shows that mutation rate does not have a strong effect on neuro-fuzzy system results, but based on test data performance, a low mutation rate generally gives better results than a high mutation rate.

Table 3. Training error and correct classification rate (CCR) for different mutation rates using Q-BBO.

Mutation rate (%)	Training Error		Train CCR (%)		Test CCR (%)	
	Best	Mean	Best	Mean	Best	Mean
0.1	0.79	0.85	81	76	71	61
0.2	0.82	0.86	80	75	72	59
0.5	0.77	0.85	82	76	69	62
1.0	0.80	0.85	80	74	67	57
2.0	0.83	0.86	79	74	69	62
5.0	0.82	0.87	81	74	68	58
10.0	0.80	0.87	78	73	65	59

Figure 6 shows the progress for a typical Q-BBO training simulation. Note that the minimum training error in the top plot is monotonically nonincreasing due to the inherent elitism of the algorithm (see Figure 3). However, the average cost in the top plot, along with the success rates in the bottom plot, sometimes increases and sometimes decreases from one generation to the next. The results shown in Figure 6 also indicate that better results might be obtained if the generation limit were increased. However, care must be taken when increasing the generation limit. As the generation count increases, the training error will continue to decrease but the test error will eventually begin to increase due to overtraining [33].

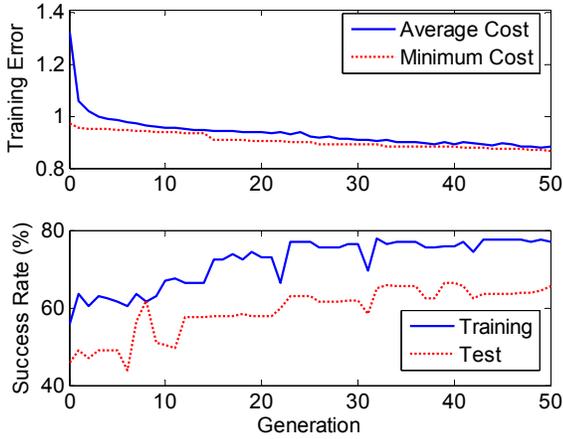


Figure 6. Typical Q-BBO training results.

The Q-BBO training run illustrated in Figure 6 resulted in the following neuro-fuzzy parameters:

$$c = \begin{bmatrix} 0.513 & 0.116 & 0.981 & 0.065 \\ 0.316 & 0.930 & 0.138 & 0.214 \\ 0.899 & 0.235 & 0.041 & 0.613 \end{bmatrix} \quad (9)$$

$$\sigma = \begin{bmatrix} 1.119 & 0.409 & 0.133 & 0.101 \\ 0.326 & 0.805 & 1.963 & 1.529 \\ 1.825 & 0.356 & 0.858 & 0.438 \end{bmatrix} \quad (10)$$

$$z = [1.641 \quad -0.967 \quad 0.779]. \quad (11)$$

Recall that we used a c range of $[0, \pi]$, but from Eq. (9) the highest membership centroid was less than 1 after Q-BBO training. This indicates that we could decrease the c range in order to get better resolution during training.

Similarly, recall that we used a σ range of $[0.01, 5]$, but from Eq. (10) the highest standard deviation was less than 2 after Q-BBO training. This indicates that we could decrease the σ range in order to get better resolution during training.

Finally, recall that we used an output singleton z range of $[-10, +10]$, but from Eq. (11) the output singletons were between -1 and 2 after Q-BBO training. This indicates that we could decrease the z range to get better resolution during training.

3.4 Clustering and Pruning

The appropriate number of clusters in the neuro-fuzzy system is equivalent to the number of middle-layer neurons p shown in Figure 1. Determination of the optimal number of fuzzy rules is equivalent to finding a suitable number of clusters for the given data set. This can also be performed using fuzzy c-means clustering [5], [13]. Clustering is itself a multiobjective optimization problem that maximizes compactness within clusters, maximizes separation between clusters, and maximizes neuro-fuzzy system performance.

In the previous section we solved for cluster count using manual tuning (see Table 1). However, we could also solve for cluster count by observing the output singletons z_i after training, discarding those that are significantly smaller than the others, and

then retraining the network. For example, when using BBO to train the neuro-fuzzy system with 5 middle-layer neurons, a typical result for the output singletons after convergence is

$$z = [1.766 \quad 1.880 \quad -1.712 \quad \mathbf{0.392} \quad -1.542].$$

The magnitude of z_4 (0.392) is smaller by a factor of 4 than any other element of z . This indicates that the corresponding fuzzy rule could be removed from the system without sacrificing performance. Retuning should then be performed to adjust the neuro-fuzzy parameters for the network size reduction.

Another way to check if we are using too many middle-layer neurons is by looking at the distance between fuzzy membership function centers. If, after training, two membership function centers are very close to each other, that indicates that those two fuzzy sets could be combined. For example, the matrix of fuzzy centroids after a typical training run with 5 middle-layer neurons (i.e., 5 fuzzy membership sets) is given by

$$c = \begin{bmatrix} 0.5587 & 0.0046 & 0.9480 & 0.6628 \\ 0.4908 & 0.3719 & 0.4274 & 0.2847 \\ 0.5534 & 0.9005 & 0.9880 & 0.2659 \\ 0.9839 & 0.7428 & 0.3904 & 0.2067 \\ 0.9992 & 0.6061 & 0.2754 & 0.2185 \end{bmatrix}.$$

Each row of c corresponds to a fuzzy set centroid, and each column of c corresponds to one dimension of the input data. A cursory look at the c matrix shows that rows 4 and 5 are similar to each other. A matrix of Euclidean distances between centroids (i.e., between columns of c) can be derived as

$$\Delta c = \begin{bmatrix} 0 & 0.7439 & 0.9807 & 1.1157 & 1.0980 \\ 0.7439 & 0 & 0.7732 & 0.6231 & 0.5838 \\ 0.9807 & 0.7732 & 0 & 0.7556 & 0.8919 \\ 1.1157 & 0.6231 & 0.7556 & 0 & \mathbf{0.1797} \\ 1.0980 & 0.5838 & 0.7556 & \mathbf{0.1797} & 0 \end{bmatrix}$$

where Δc_{ij} is the Euclidean distance between centroids i and j . The Δc matrix indicates that fuzzy centroids 4 and 5 are much closer to each other than the other centroids, which implies that the corresponding membership functions overlap, and so they could be combined. Afterward, the neuro-fuzzy system should be retrained to compensate for the change in its structure.

3.5 Fine Tuning Using Gradient Information

The BBO algorithm that we used, like other EAs, does not depend on gradient information. Therein lies its strength relative to gradient-based optimization methods. EAs can be used for global optimization since they do not rely on local gradient information. Since the neuro-fuzzy system shown in Figure 1 may have multiple optima, BBO training is less likely to get stuck in a local optima compared to gradient-based optimization.

However, additional performance improvement could be obtained in the neuro-fuzzy classifier by using gradient information in conjunction with EA-based optimization. Gradient-based methods can be combined with EAs in order to take advantage of the strengths of each method. First we can use BBO, as above, in order to find neuro-fuzzy parameters that are in the neighborhood of the global optimum. Then we can use a gradient-based method to fine tune the BBO result. The most commonly-used gradient-

based method is gradient descent [5], [13]. Gradient descent can be further improved by using an adaptive learning rate and momentum term [19].

Kalman filtering is a gradient-based method that can give better fuzzy system and neural network training results than gradient descent [26], [27]. Constrained Kalman filtering can further improve fuzzy system results by optimally constraining the network parameters [28]. H-infinity estimation is another gradient-based method that can be used for fuzzy system training to improve robustness to data errors [29].

3.6 Training Criterion

The ultimate goal of the neuro-fuzzy network is to maximize correct classification percentage. If the neuro-fuzzy output is greater than 0, then the ECG is classified as cardiomyopathy; otherwise, the ECG is classified as non-cardiomyopathy. The bottom plot in Figure 6 shows that while RMS training error is monotonically nondecreasing, the success rate for the training data is non-monotonic. We could more directly address the problem of ECG data classification by using classification success rate as our fitness function rather than trying to minimize the RMS error of Eq. (8). That is, in fact, one of the advantages of EA training relative to gradient-based methods – the fitness function does not have to be differentiable. However, if we use classification success rate as our fitness function, and then try to use a gradient-based method for fine-tuning, the cost functions of the two training methods would be inconsistent.

4. CONCLUSIONS

We have shown that clinical ECG data can be correctly classified as cardiomyopathy or non-cardiomyopathy using a neuro-fuzzy network trained by BBO. Our results show a correct classification rate on test data of over 60%. Better results can undoubtedly be attained with further training, but the main goal of this initial research was to demonstrate feasibility and establish a framework for further refinement.

Although preliminary results are good, there are enhancements that need to be made to improve performance and incorporate this work into a commercial product. For example, demographic information needs to be included with the ECG data. Some of the test ECGs were correctly classified 100% of the time, while others had a low success rate. Figure 7 shows the classification success rate for the test data sorted by patient ID. Some patient's ECG data were successfully classified only 2% of the time, while others were successfully classified 100% of the time. This indicates that demographic data is important and that we should group patients into similar groups for testing and training. Some of these data include gender, race, medication usage, and age. This will become feasible as we perform more clinical studies and collect data from more patients.

We note that our results are based on data snapshots at single instants of time. We could presumably get better results by using a "majority rules" strategy for data over several minutes. For example, suppose test accuracy is 60% for a given patient. We could use ECG data at three separate time instants and diagnose cardiomyopathy if the network predicts cardiomyopathy for two or more of the data. This would boost test accuracy from 60% to 65%, assuming that the probability of correct classification is independent from one time instant to the next. We could then further improve accuracy by using more time instants.

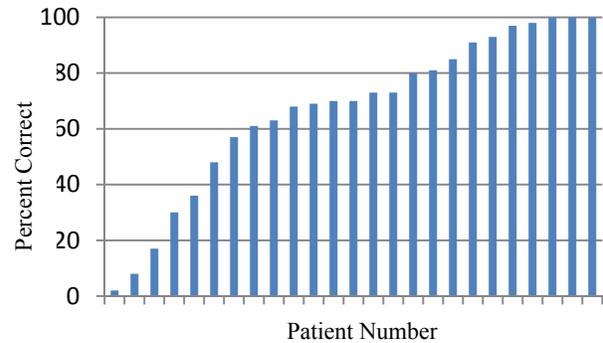


Figure 7. ECG classification success rate for test patients.

5. ACKNOWLEDGMENT

This work was supported by Grant 0826124 from the National Science Foundation.

6. REFERENCES

- [1] American Heart Association, 2009. Heart Disease and Stroke Statistics – 2009 Update, www.americanheart.org/presenter.jhtml?identifier=3018163
- [2] Bahl, O. 1972. Electrocardiographic and vector-cardiographic pattern in cardiomyopathy, *Cardiovascular Clinics*, 4, 1 (1972), 95–112.
- [3] Bashour, C., Visinescu, M., Gopakumaran, B., Wazni, O., Carragio, F., Yared, J., and Starr, N. 2004. Characterization of premature atrial contraction activity prior to the onset of postoperative atrial fibrillation in cardiac surgery patients, *Chest*, 126, 4 (2004), 831S.
- [4] Chandy, J., Nakai, T., Lee, R., Bellows, W., Dzankic, S., and Leung, J. 2004. Increases in P-wave dispersion predict postoperative atrial fibrillation after coronary artery bypass graft surgery, *Anesthesia & Analgesia*, 98 (2004), 303–310.
- [5] Chen, M. and Linkens, D. 2001. A systematic neuro-fuzzy modelling framework with application to material property prediction, *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 31, 5 (2001), 781–790.
- [6] Dische, M. 1972. Observations on the morphological changes of the developing heart. *Cardiovascular Clinics*, 4, 3 (1972), 175–191.
- [7] Du, D., Simon, D., and Ergezer, M. 2009. Biogeography-based optimization combined with evolutionary strategy and immigration refusal. In *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics (San Antonio, Texas, October 2009)*. 1023–1028.
- [8] Ergezer, M., Simon, D., and Du, D. 2009. Oppositional biogeography-based optimization. In *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics (San Antonio, Texas, October 2009)*. 1035–1040.
- [9] Hancock, E., Mirvis, D., Okin, P., et al. 2001. AHA/ACC/HRS recommendations for the standardization of interpretation of the electrocardiogram: Part V: Electrocardiogram changes associated with cardiac chamber hypertrophy: A scientific statement from the American Heart Association,

- Council on Clinical Cardiology, The American College of Cardiology Foundation, and the Heart Rhythm Society, *J. Amer. Coll. Cardiology*, 53 (2009), 992–1002.
- [10] Johnson, J., Horan, L., and Flowers, N. 1977. Diagnostic accuracy of the electrocardiogram. *Cardiovascular Clinics*, 8, 3 (1977), 25–40.
- [11] Kennedy, J., Eberhart, R., and Shi, Y. 2001. *Swarm Intelligence*. Morgan Kaufmann,.
- [12] Kundra, H., Kaur, A., and Panchal, V. 2009. An integrated approach to biogeography based optimization with case based reasoning for retrieving groundwater possibility. In *Proceedings of the 8th Annual Asian Conference and Exhibition on Geospatial Information, Technology and Applications* (Singapore, August 2009).
- [13] Linkens, D. and Chen, M. 1999. Input selection and partition validation for fuzzy modelling using neural network. *Fuzzy Sets and Systems*, 107, 3 (1999), 299–308.
- [14] Lomolino, M., Riddle, B., and Brown, J. 2009. *Biogeography*. Sinauer Associates.
- [15] Ma, H., Ni, S., and Sun, M. 2009. Equilibrium species counts and migration model tradeoffs for biogeography-based optimization. 2009. In *Proceedings of the IEEE Conference on Decision and Control* (Shanghai, Dec. 2009).
- [16] Macfarlane, P. 2006. Is electrocardiography still useful in the diagnosis of cardiac chamber hypertrophy and dilatation? *Cardiology Clinics*, 24, 3 (2006), 401–411.
- [17] Magdic, K. and Saul, L. 1997. ECG interpretation of chamber enlargement. *Critical Care Nurse*, 17, 1 (1997), 13, 16–25.
- [18] McKenna, W., Krikler, D., and Goodwin, J. 1984. Arrhythmias in dilated and hypertrophic cardiomyopathy. *Medical Clinics of North America*, 68, 4 (1984), 983–1000.
- [19] Nauck, D., Klawonn, F., and Kruse, R. 1997. *Foundations of Neuro-Fuzzy Systems*. John Wiley & Sons.
- [20] Olson, E. 2004. A decade of discoveries in cardiac biology. *Nature Medicine*, 10 (2004). 467–474.
- [21] Panchal, V., Singh, P., Kaur, N., and Kundra, H. 2009. Biogeography based satellite image classification. *International Journal of Computer Science and Information Security*. 6, 2 (Nov. 2009), 269–274.
- [22] Rarick, R., Simon, D., Villaseca, F. E., and Vyakaranam, B. 2009. Biogeography-based optimization and the solution of the power flow problem, In *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics* (San Antonio, Texas, October 2009). 1029–1034.
- [23] Rahnamayan, S., Tizhoosh, H., and Salama, M. 2006. Opposition-based differential evolution for optimization of noisy problems, in *Proc. IEEE Congress on Evolutionary Computation* (Vancouver, Canada, July 2006). 1865–1872.
- [24] Rahnamayan, S., Tizhoosh, H., and Salama, M. 2008. Opposition-based differential evolution. *IEEE Transactions on Evolutionary Computation*, 12, 1 (2008), 64–79.
- [25] Rahnamayan, S., Tizhoosh, H., and Salama, M. 2007. Quasioppositional differential evolution, in *Proc. IEEE Congress on Evolutionary Computation* (Singapore, September 2007). 2229–2236.
- [26] Simon, D. 2002. Training fuzzy systems with the extended Kalman filter. *Fuzzy Sets and Systems*, 132, 2 (December 2002), 189–199.
- [27] Simon, D. 2002. Training radial basis neural networks with the extended Kalman filter. *Neurocomputing*, 48, 1 (October 2002), 455–475.
- [28] Simon, D. 2002. Sum normal optimization of fuzzy membership functions. *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 10, 4 (August 2002), 363–384.
- [29] Simon, D. 2005. H-infinity estimation for fuzzy membership function optimization, *International Journal of Approximate Reasoning*, 40, 3 (November 2005), 224–242.
- [30] Simon, D. 2008. Biogeography-Based Optimization, *IEEE Transactions on Evolutionary Computation*, 12, 6 (December 2008), 702–713.
- [31] Simon, D., Ergezer, M., and Du, D. 2009. Population distributions in biogeography-based optimization algorithms with elitism, In *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics* (San Antonio, Texas, October 2009). 1017–1022.
- [32] Sox, H., Garber, A., and Littenberg, B. 1989. The resting electrocardiogram as a screening test: A clinical analysis, *Annals of Internal Medicine*, 111, 6 (Sep. 1989), 489–502.
- [33] Tetko, I., Livingstone, D., and Luik, A. 1995. Neural network studies. 1. Comparison of overfitting and overtraining. *Journal of Chemical Information and Modeling*, 35, 5 (September 1995), 826–833.
- [34] Tizhoosh, H. 2005. Opposition-based learning: A new scheme for machine intelligence, in *Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation* (Vienna, Austria, November 2005). 695–701.
- [35] Visinescu, M., Bashour, A., Wazni, O., and Gopakumaran, B. 2004. Automatic detection of conducted premature atrial contractions to predict atrial fibrillation in patients after cardiac surgery, in *Proceedings of the 31st Annual International Conference on Computing in Cardiology* (Chicago, Illinois, September 2004). 429–432.
- [36] Visinescu, M. 2005. Analysis of ECG to predict atrial fibrillation in post-operative cardiac surgical patients. Doctoral Dissertation, Cleveland State University, Cleveland, Ohio.
- [37] Visinescu, M., Bashour, A., Bakri, M., and Nair, B. 2006. Automatic detection of QRS complexes in ECG signals collected from patients after cardiac surgery, in *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (New York, August 2006). 3724–3727.
- [38] Whittaker, R. 1998. *Island Biogeography*. Oxford University Press.
- [39] Ziegler, R. 1970. Electrocardiographic clues in the diagnosis of congenital heart disease, *Cardiovascular Clinics*, 2, 1 (1970). 97–114.