Detection of Heart Disease using Binary Particle Swarm Optimization

Mona Nagy Elbedwehy
Mathematics department, Faculty of Science, Damietta Branch, Mansoura University Egypt
Scientific Research Group in Egypt (SRGE)
Email: monanagyelbedwehy@ymail.com

Hossam M. Zawbaa†, Neveen Ghali‡, and Aboul Ella Hassanien§
†Faculty of Computers and Information Cairo University, Egypt
‡Faculty of Science, Al-Azhar University, Egypt
Scientific Research Group in Egypt (SRGE)
http://www.egyptscience.net

Abstract—This article introduces a computer-aided diagnosis system of the heart valve disease using binary particle swarm optimization and support vector machine, in conjunction with K-nearest neighbor and with leave-one-out cross-validation. The system was applied in a representative heart dataset of 198 heart sound signals, which come both from healthy medical cases and from cases suffering from the four most usual heart valve diseases: aortic stenosis (AS), aortic regurgitation (AR), mitral stenosis (MS) and mitral regurgitation (MR). The introduced approach starts with an algorithm based on binary particle swarm optimization to select the most weighted features. This is followed by performing support vector machine to classify the heart signals into two outcomes: healthy or having a heart valve disease, then its classified the having a heart valve disease into four outcomes: aortic stenosis (AS), aortic regurgitation (AR), mitral stenosis (MS) and mitral regurgitation (MR). The experimental results obtained, show that the overall accuracy offered by the employed approach is high compared with other techniques

Keywords—Binary particle swarm optimization, Support vector machine, Heart valve diseases, Heart sounds.

I. INTRODUCTION

Every human body and its physiological processes show some symptoms of a diseased condition. The proposed model in this paper used for identification of the heart valve diseases using heart sounds. The signal of heart sound carries important physiological and pathological information, Its about the general state of contractile activity of the cardiovascular system. The heart murmurs caused by turbulent blood flow and the incomplete opening or closing of the valves, could be heard clearly sounding like whistling, swishing or humming [10].

The feature selection process can be considered as a problem of global combinatorial optimization in machine learning and statistics. Feature selection, also known as variable selection, feature reduction, attribute selection or variable subset selection. Therefore, a good feature selection is which speeds up the processing rate, predictive accuracy, and avoids incomprehensibility. Several methods have been previously used to perform feature selection on training and testing data, for example genetic algorithms, catfish binary PSO, improved binary PSO, and support vector machine(SVM) [1]–[3].

In this paper, binary particle swarm optimization (BPSO) is used to implement feature selection, and the K-nearest neighbor (KNN) method with leave-one-out-cross-validation (LOOCV) based on Euclidean distance calculations serves as fitness function of the BPSO for a classification problem. BPSO increases either the classification accuracy, reduce the number of necessarily selected features for classification, or does both. This proposed method is applied on heart sound data. Also, in this paper, support vector machine (SVM) is used to identify on the heart valve diseases using the reduced heart sound data.

The rest of this paper is organized as follows. Section (II) gives an overview of BPSO, KNN, and SVM machine learning techniques. Section (III) presents the dataset used for constructing and testing the BPSO-KNN-SVM. Section (IV) shows the proposed system for identification of heart valve diseases. Finally, Section (V) addresses conclusions and future work.

II. BACKGROUND AND AN OVERVIEW

A. Binary particle swarm optimization (BPSO)

The particle swarm optimization (PSO) is a robust stochastic optimization technique based on the movement and intelligence of swarms. PSO is a population-based search algorithm based on simulation of the social behavior of birds within a flock [5]. PSO is widely used to solve the optimization problems and also the feature selection problem [3].

PSO uses a number of particles that constitute for a swarm moving around in the search space looking for the best initialized randomly. A swarm consists of N of particles, where each particle represents a candidate solution moving around D-dimensional search space.

All of the particles have fitness values, which are evaluated by a fitness function to be optimized, and have velocities which direct the movement of the particles. During movement, the changes to a particle within the swarm are influenced by the experience, or knowledge, of its neighbors and each particle adjusts its position according to two fitness value, pbest and gbest, to avoid being trapped in a local optimum by fine-tuning the inertia weight. pbest is a local fitness value, whereas gbest constitutes a global fitness value. If the gbest value is itself trapped in a local optimum, a search of each particle limit in the same area will occur, thereby preventing superior results of classification. The modification of the particle’s position can be mathematically modeled according the following equations
(1), (2). Algorithm 1 gives the pseudo code of the BPSO procedure as given in [3].

Initialize particles with random position and velocity vectors.

while The number of generations (N), or the stopping criterion doesn’t meet do
  Evaluate the fitness of particle swarm.
  for p = 1 to N (number of particles) do
    if The fitness of \( x_p \) > the fitness of \( p_{best} \) then
      update \( p_{best} = x_p \).
    end if
  end for
  for D = 1 to maximum dimension do
    update particle’s velocity and position using equations 1 and 2
  end for
  go to next generation until stopping criterion.
end while

Algorithm 1: PSO pseudo-code

\[
v_{i}^{k+1} = w v_{i}^{k} + c_{1} \cdot rand_{1} (p_{best_{i}} - x_{i}^{k}) + c_{2} \cdot rand_{2} (g_{best} - x_{i}^{k}).
\]  
(1)

\[
x_{i}^{k+1} = x_{i}^{k} + v_{i}^{k+1}.
\]  
(2)

Where \( v_{i}^{k+1} \) is the modified velocity of a particle \( i \), \( v_{i}^{k} \) is the velocity of particle \( i \) at iteration \( k \), \( w \) is the weight function, \( c_{j} \) is the weighting factor, \( rand_{j} \) is uniformly distributed random number between 0 and 1, \( x_{i}^{k+1} \) is the modified position of particle \( i \), and \( x_{i}^{k} \) is the current position of particle \( i \) at iteration \( k \).

The following weighting function is described in [7]:

\[
w = \frac{w_{max} - w_{min}}{iter_{max}} \cdot iter.
\]  
(3)

Where \( w_{max} \) is initial weight, \( w_{min} \) is final weight, \( iter_{max} \) is maximum iteration number and \( iter \) is current iteration number.

Many optimization problems occur in a space featuring discrete, qualitative distinctions between variables and between levels of variables. For the reason, Kennedy and Eberhart introduced binary particle swarm optimization (BPSO), which can be applied to discrete binary variables. In BPSO, the elements only take the values 0 or 1 and the velocity \( V_{ij} \) is interpreted as a probability to change a bit from 0 to 1, or from 1 to 0. Updating the position of particles by sigmoid function as in the equation 4 and 5. The updating particle’s positions is done by the following probabilistic update equation 6.

\[
V_{ij}^{k+1} = w_{ij}^{k} + c_{1} \cdot rand_{1} (p_{best_{ij}} - x_{ij}^{k}) + c_{2} \cdot rand_{2} (g_{best} - x_{ij}^{k}).
\]  
(4)

Where \( w_{ij}^{k} \) is the weighting factor, \( rand_{1} \) and \( rand_{2} \) are uniformly distributed random numbers ranging from 0 to 1.

\[
Sig(V_{ij}^{k+1}) = \frac{1}{1 + e^{-V_{ij}^{k+1}}}. \tag{5}
\]

\[
x_{ij}(k+1) = \begin{cases} 1, & \text{if} rand_{n} < Sig(V_{ij}^{k+1}) \\ 0, & \text{if} rand_{n} \geq Sig(V_{ij}^{k+1}). \end{cases} \tag{6}
\]

Where 1 means this feature is selected as a required feature for the next renewal, 0 means this feature is not selected as a required feature for the next renewal, and \( rand_{n}(t) \) is a random value in the range [0; 1].

B. K-Nearest Neighbor (KNN)

The K-nearest neighbor (K-NN) method is one of the most popular nonparametric methods used for classification of new object based on attributes and training samples. The K-NN consists of a supervised learning algorithm where the result of a new instance query is classified based on the majority of the K-nearest neighbor category [3]. In this paper, the feature subset was measured by LOOCV of one nearest neighbor (1-NN). Neighbors are calculated using their Euclidean distance measuring as distance function as the following equation (7). The 1-NN classifier is simple and provides a reasonable classification performance in most applications. As the 1-NN classifier does not require any user-specified parameters, its classification results are implementation independent [2], [3].

\[
D(x, y) = \sum_{i=0}^{N} \sqrt{x_{i}^{2} - y_{i}^{2}}. \tag{7}
\]

Where \( D(x, y) \) is the distance function, \( x \) is the query sample, \( y \) is the sample from the training set, \( N \) is feature dimension. The advantages of K-nearest neighbors are robust to noisy training data especially in inverse square of weighted distance. And disadvantages of K-nearest neighbors are need to determine value of parameter K (number of nearest neighbors), and computation cost is quite high because we need to compute distance of each query instance to all training samples.

C. Support Vector Machine (SVM)

The support vector machine (SVM) exhibit great performance in pairwise classification and regression problems. While recently efficient algorithms have been developed that extended their applicability to multi-class classification problems [9]. Classification is achieved by a linear or nonlinear separating surface in the input space of the data set. The goal of SVM is to minimize the expectation of the output of sample error. SVM map a given set of binary labeled training data to a high dimensional feature space and separate the two classes of data with a maximum margin of hyperplane. SVM algorithm seeks to maximize the margin around a hyperplane that separates a positive class from a negative class [11] as shown in equations 8 and 9.

\[
\max \sum_{i=1}^{n} \alpha_{i} - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_{i} \alpha_{j} y_{i} y_{j} K(x_{i}, x_{j}). \tag{8}
\]
Subject to: \[ \sum_{i=1}^{n} \alpha_i y_i = 0, 0 \leq \alpha_i \leq C. \] (9)

Where \( \alpha_i \) is the weight assigned to the training sample \( x_i \), \( \alpha_i \) is called a support vector, \( C \) is a regulation parameter, \( K \) is a kernel function.

III. BINARY PARTICLE SWARM OPTIMIZATION - K-NEAREST NEIGHBOR PROCEDURE (BPSO-KNN)

Feature selection was implemented using BPSO, and a KNN served as an evaluator for the classification obtained by BPSO. The predicative accuracy of a 1-NN determined by the LOOCV method is used to measure the fitness of an individual. The pseudo code for BPSO-KNN is given in Algorithm 2.

Initialize particles with random position and velocity vectors.

while number of generations, or the stopping criterion isn’t meet do
  Evaluate fitness of particle swarm by 1-Nearest Neighbor().
  for p = 1 to number of particles do
    if the fitness of \( x_p \) is greater than the fitness of \( pbest_p \) then
      Update \( pbest_p = x_p \).
    end if
    if fitness of any particle of the particle swarm is greater than \( gbest \) then
      \( gbest = \) position of particle
    end if
  end for
  for \( d=1 \) to number of dimension of particle do
    Update particles velocity and particles position as in equations 4, 5, and 6 respectively.
  end for
  go to next generation until stopping criterion.
end while

Algorithm 2: BPSO-KNN pseudo-code

IV. HEART SOUND SIGNALS PRE-PROCESSING

The heart sound signal from a healthy heart has the form shown in the upper part of Fig. 1. Its basic components are: S1 which is generated by the nearly simultaneous closure of the mitral and the tricuspid valve being followed by the systolic phase. S2 which is generated by the nearly simultaneous closure of the aortic and the pulmonic valve being followed by the diastolic phase. Most heart diseases generate additional components in the heart sound, such as murmurs in the systolic or diastolic phase (see lower part of Fig. 1). S3, S4, clicks, snaps, etc. concerning the heart valve diseases dealt with in this paper, aortic stenosis and mitral regurgitation generate systolic murmurs, while aortic regurgitation and mitral stenosis generate diastolic murmurs.

V. THE PROPOSED METHOD AND EXPERIMENTAL RESULTS

A. BPSO-KNN-SVM for identification the heart valve diseases

BPSO-KNN-SVM is used to identify the heart valve diseases using heart sounds. First, we reduce the number of features by using BPSO-KNN to remove irrelevant data, and the results in an acceptable classification accuracy. Second, we use the feature selection after removing irrelevant/redundant features to identify of heart valve diseases using heart sounds by using SVM. The heart sound dataset format is shown in Table I.

The 1-NN method with LOOCV was used to evaluate all data sets. During the experiment, the parameters of BPSO-KNN-SVM are \( rand_1, rand_2 \) and \( rand_3 \) are random numbers between \((0,1)\), \( c_1 \) and \( c_2 \) are learning factors \( c_1 = 2 \) (Kennedy et al., 2001), maximum generation=100, \( w_{max} = 0.9 \) and \( w_{min} = 0.4 \) are inertia weight, \( V_{max} = 6 \) and \( V_{min} = -6 \) are user specified parameters.

B. BPSO-KNN-SVM for diagnosis of healthy vs. pathological cases

The development and implementation of BPSO-KNN-SVM based automatic diagnostic system for cardiac sounds launches by studying patient cases of pathological heart murmurs and cases where no heart disease is diagnosed. The pathological heart murmurs are characterized as "unhealthy" cases in contrast with the "healthy" cases. The performance indices calculated were accuracy, specificity and sensitivity. The heart sound dataset used consisted of 38 healthy cases and 32 unhealthy cases. Several polynomial, gaussian and linear kernel functions have been used comparatively in order to select the model with the best performance. The values of order of polynomial function used providing different hyperplanes for the classification of data during the SVM calculations.
The kernel functions we examine during the development of the SVM models are polynomial and linear. The highest performances correspond to the linear kernel function and number of features selection to BPSO-KNN-SVM are presented in table II and the highest performances to SVM are presented in table III. The highest performances correspond to the polynomial kernel function and number of features selection to BPSO-KNN-SVM are presented in table IV and the highest performances to SVM are presented in table V. The accuracy of the BPSO-KNN-SVM classifier and SVM for feature selection and then the classification between healthy and unhealthy classification using different kernel functions and parameters are depicted in fig. 2 and fig. 3.

In tables the \( TP \) indicates the number of correctly classified positive samples, \( FP \) indicates the number of incorrectly classified positive samples, \( TN \) indicates the number of correctly classified negative samples, and \( FN \) indicates the number of incorrectly classified negative samples. Sens. and spec. are statistical measures of the performance of a binary classification test, also known in statistics as classification function. 

\[
\text{Sens.} = \frac{TP}{TP + FN} \quad \text{and} \quad \text{Spec.} = \frac{TN}{TN + FP}
\]

Also, \( FS \) means feature selection, \( kernel \) means Kernel function handle specifying the kernel function that maps the training data into kernel space. We use two types of kernel functions, linear kernel and polynomial kernel with different orders (1,2,3,5).

### C. BPSO-KNN-SVM for diagnosis of aortic regurgitation vs. mitral stenosis

In the second stage, an additional BPSO-KNN-SVM classifier has been developed for the discrimination of aortic regurgitation vs. mitral stenosis.

#### TABLE II

<table>
<thead>
<tr>
<th>Iteration</th>
<th>FS</th>
<th>Accuracy</th>
<th>Error</th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>Sens.</th>
<th>Spec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF = 1</td>
<td>39</td>
<td>82.56%</td>
<td>6</td>
<td>14</td>
<td>15</td>
<td>4</td>
<td>2</td>
<td>87.30%</td>
<td>75.93%</td>
</tr>
<tr>
<td>PF = 2</td>
<td>41</td>
<td>82.86%</td>
<td>6</td>
<td>11</td>
<td>18</td>
<td>1</td>
<td>5</td>
<td>68.75%</td>
<td>94.74%</td>
</tr>
<tr>
<td>PF = 3</td>
<td>44</td>
<td>82.86%</td>
<td>6</td>
<td>14</td>
<td>15</td>
<td>4</td>
<td>2</td>
<td>87.50%</td>
<td>78.95%</td>
</tr>
<tr>
<td>PF = 5</td>
<td>37</td>
<td>82.86%</td>
<td>6</td>
<td>16</td>
<td>13</td>
<td>6</td>
<td>0</td>
<td>100.0%</td>
<td>68.42%</td>
</tr>
</tbody>
</table>

#### TABLE III

<table>
<thead>
<tr>
<th>Iteration</th>
<th>FS</th>
<th>Accuracy</th>
<th>Error</th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>Sens.</th>
<th>Spec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF = 1</td>
<td>39</td>
<td>74.29%</td>
<td>9</td>
<td>10</td>
<td>16</td>
<td>3</td>
<td>6</td>
<td>62.50%</td>
<td>88.21%</td>
</tr>
<tr>
<td>PF = 2</td>
<td>41</td>
<td>74.29%</td>
<td>9</td>
<td>9</td>
<td>17</td>
<td>2</td>
<td>7</td>
<td>56.25%</td>
<td>89.47%</td>
</tr>
<tr>
<td>PF = 3</td>
<td>44</td>
<td>85.71%</td>
<td>5</td>
<td>14</td>
<td>16</td>
<td>3</td>
<td>2</td>
<td>87.50%</td>
<td>84.21%</td>
</tr>
<tr>
<td>PF = 5</td>
<td>37</td>
<td>74.29%</td>
<td>9</td>
<td>10</td>
<td>16</td>
<td>3</td>
<td>6</td>
<td>62.50%</td>
<td>88.21%</td>
</tr>
</tbody>
</table>

#### TABLE IV

<table>
<thead>
<tr>
<th>Iteration</th>
<th>FS</th>
<th>Accuracy</th>
<th>Error</th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>Sens.</th>
<th>Spec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>60</td>
<td>82.56%</td>
<td>6</td>
<td>12</td>
<td>17</td>
<td>2</td>
<td>4</td>
<td>75.00%</td>
<td>89.47%</td>
</tr>
<tr>
<td>20</td>
<td>43</td>
<td>74.28%</td>
<td>8</td>
<td>14</td>
<td>13</td>
<td>6</td>
<td>2</td>
<td>87.00%</td>
<td>68.42%</td>
</tr>
<tr>
<td>100</td>
<td>48</td>
<td>74.28%</td>
<td>9</td>
<td>12</td>
<td>14</td>
<td>5</td>
<td>4</td>
<td>75.00%</td>
<td>73.68%</td>
</tr>
</tbody>
</table>
SVM are presented in table X and the highest performances to SVM are presented in table XI. The highest performances correspond to the polynomial kernel function and number of features selection to BPSO-KNN-SVM are presented in table XII and the highest performances to SVM are presented in table XIII. The accuracy of the BPSO-KNN-SVM classifier and SVM for feature selection and then the classification between aortic stenosis (AS) and mitral regurgitation (MR) with different kernel functions and parameters are depicted in fig. 6 and fig. 7.

VI. CONCLUSIONS

The main task is the selection of the features that give a high accuracy to identify on the heart valve diseases using heart sounds. In this paper, we used binary particle swarm optimization (BPSO) to perform feature selection. The K-nearest neighbor (KNN) method with leave-one-out-cross-validation (LOOCV) based on Euclidean distance calculations serves as fitness function of the BPSO for classification problem and then used support vector machine (SVM) to identify on the heart valve diseases using heart sounds. The proposed methodology includes initially a preprocessing step of the heart sounds, then aortic regurgitation and mitral stenosis classification using polynomial kernel function

TABLE VIII
RESULTS OF BPSO-KNN-SVM FOR FEATURE SELECTION, THEN AORTIC REGURGITATION AND MTRAL STENOSIS CLASSIFICATION USING LINEAR KERNEL FUNCTION

<table>
<thead>
<tr>
<th>Kernel</th>
<th>FS</th>
<th>Accuracy</th>
<th>Error</th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>Sens.</th>
<th>Spec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>40</td>
<td>94.74%</td>
<td></td>
<td>2</td>
<td>18</td>
<td>18</td>
<td>1</td>
<td>1</td>
<td>94.74%</td>
</tr>
<tr>
<td>100</td>
<td>43</td>
<td>92.11%</td>
<td></td>
<td>3</td>
<td>17</td>
<td>17</td>
<td>2</td>
<td>0</td>
<td>94.74%</td>
</tr>
<tr>
<td>20</td>
<td>45</td>
<td>86.84%</td>
<td></td>
<td>4</td>
<td>18</td>
<td>16</td>
<td>3</td>
<td>1</td>
<td>94.74%</td>
</tr>
</tbody>
</table>

TABLE IX
RESULTS OF SVM FOR FEATURE SELECTION, THEN AORTIC REGURGITATION AND MTRAL STENOSIS CLASSIFICATION USING LINEAR KERNEL FUNCTION

<table>
<thead>
<tr>
<th>Kernel</th>
<th>FS</th>
<th>Accuracy</th>
<th>Error</th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>Sens.</th>
<th>Spec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>40</td>
<td>94.74%</td>
<td></td>
<td>2</td>
<td>18</td>
<td>18</td>
<td>1</td>
<td>1</td>
<td>94.74%</td>
</tr>
<tr>
<td>100</td>
<td>43</td>
<td>92.11%</td>
<td></td>
<td>3</td>
<td>17</td>
<td>17</td>
<td>2</td>
<td>0</td>
<td>94.74%</td>
</tr>
<tr>
<td>20</td>
<td>45</td>
<td>86.84%</td>
<td></td>
<td>4</td>
<td>18</td>
<td>16</td>
<td>3</td>
<td>1</td>
<td>94.74%</td>
</tr>
</tbody>
</table>

D. BPSO-KNN-SVM for diagnosis of aortic stenosis vs. mitral regurgitation

Finally, an additional SVM classifier has been developed for the discrimination of aortic stenosis (AS) vs. mitral regurgitation (MR) for the cases diagnosed by the BPSO-KNN-SVM classifier. The heart sound dataset used consisted of 41 aortic stenosis cases and 43 mitral regurgitation cases.

The highest performances correspond to the linear kernel function and number of features selection to BPSO-KNN-
phase based on SVM classifiers: in the first step the heart sound signal is classified as normal or pathological, in the second step the type of heart systolic murmur diseases (aortic stenosis, mitral regurgitation), while in the third step it is decided the type of heart diastolic murmur diseases (aortic regurgitation, mitral stenosis). The proposed method can serve as an ideal pre-processing tool to help optimize the feature selection process, since it increase the classification accuracy.

REFERENCES