Support Vector Machine based expert system for reliable heart beat recognition

Stanislaw Osowski, Member, IEEE, Linh Tran Hoai, and Tomasz Markiewicz

Abstract— The paper presents a new solution to the expert system for reliable heart beat recognition. The recognition system uses the Support Vector Machine (SVM) working in the classification mode. Two different preprocessing methods for generation of features are applied. One method involves the higher order statistics (HOS) while the second the Hermite characterization of QRS complex of the registered ECG waveform. Combining the SVM network with these preprocessing methods yields two neural classifiers, which have been combined into one final expert system. The combination of classifiers utilizes the least mean square method to optimize the weights of the weighted voting integrating scheme. The results of the performed numerical experiments for the recognition of 13 heart rhythm types on the basis of ECG waveforms confirmed the reliability and advantage of the proposed approach.

Index Terms— Heart beat recognition, Expert system, Combination of classifiers, Support vector machine.

I. INTRODUCTION

The electrocardiogram (ECG) is the electrical manifestation of the contractile activity of the heart that can be recorded fairly easily. The ECG waveform is well characterized by the so called waves: the P, QRS and T one [2,8]. The most important wave is the QRS complex, a sharp biphasic or triphasic wave of about 1mV amplitude, and duration of approximately 80 - 100ms. The peak point of QRC part of ECG is denoted by R. Any disturbance in the regular rhythmic activity of the heart (amplitude, duration and the shape of rhythm) is termed arrhythmia. The electrocardiogram (ECG), especially its QRS complex, remains the simplest non-invasive diagnostic method for various heart diseases. Physicians interpret the shapes (morphology) of the ECG waveform and decide, whether the heart beat belongs to the normal (healthy) sinus rhythm or to the appropriate class of arrhythmia.

Computerized electrocardiography is currently a well-established practice, supporting human diagnosis. Many algorithms have been proposed over past years for developing the automated systems to accurately classify the electrocardiographic signals in a real time [1-7]. Depending on the type of the applied method of signal processing techniques and their formal description, we distinguish statistical, syntactic or artificial intelligence methods [11].

Presently particularly interesting applications of artificial neural networks are in the area of data processing. Many different neural solutions have been proposed [1,2,4,5,6,7]. The best known include the multilayer perceptron, Kohonen self-organizing network, the fuzzy or neuro-fuzzy systems and the combination of different neural networks within a hybrid system. The typical heart beat recognition system applying neural classifier trains many neural networks and chooses the best one, while discarding the rest. The more efficient approach is based on the combination of many classifiers utilizing either different classifier network structures or different data preprocessing methods [11,20,21].

One may observe, that for a specific recognition problem individual classifiers, relying on different feature sets, may attain different degrees of success. Typically none of them is perfect or as good as one would want. Thus there is a need to combine different classifiers so that a better result could be obtained. Combining the trained networks rather than discarding them, helps to integrate the knowledge acquired by the individual classifiers and thus improves the accuracy of the final classification.

In this paper we propose the combination of multiple classifiers by the weighted voting principle. Each classifier influences the final decision according to its performance on the training data. The weights are adjusted on the basis of an individual classifier’s performance on the training data by applying the pseudo-inverse technique. The proposed approach is validated in the MIT BIH Arrhythmia Database [8] heart beat recognition problems.

The Support Vector Machine (SVM) classifier and two different preprocessing techniques of the ECG waveform are applied. One of these techniques uses Hermite basis functions expansion while the second characterization of the ECG by the cumulants of the second, third and fourth orders. The results of numerical experiments on the recognition of normal sinus beat and 12 other types of arrhythmias are presented and discussed. They demonstrate the strength of the proposed expert system relying on the weighted voting integration principle of SVM classifiers.

II. SUPPORT VECTOR MACHINE CLASSIFIER

The SVM, pioneered by Vapnik [13,14], is known as an excellent tool for classification and regression problems with a good generalization performance. Unlike the classical neural networks approach the SVM formulation of the learning problem leads to quadratic programming with linear constraints.

Basically, the SVM is a linear machine working in the highly dimensional feature space formed by the nonlinear mapping of the n-dimensional input vector $x$ into a $K$-dimensional feature space ($K>n$) through the use of a mapping $\phi(x)$. The equation of the hyperplane separating two different classes is given by the relation
yp(x) = w^T \phi(x) = \sum_{j=1}^{K} w_j \phi_j(x) + w_0 = 0 \), where the vector \( \phi(x) \) is equal \( \phi(x) = [\phi_1(x), \phi_2(x), \ldots, \phi_K(x)]^T \) with \( \phi_0(x) = 1 \) and \( w = [w_0, w_1, \ldots, w_K]^T \) is the weight vector of the network. Fulfillment of condition \( y(x) > 0 \) means one class and \( y(x) < 0 \) the opposite one.

The most distinctive fact about SVM is that the learning task is reduced to quadratic programming by introducing the so-called Lagrange multipliers \( \alpha_i \). All operations in learning and testing modes are done in SVM using so-called kernel functions satisfying Mercer conditions [13]. The kernel is defined as \( K(x_i, x_j) = \phi^T(x_i) \phi(x_j) \). The best known kernels include radial Gaussian, polynomial and sigmoidal functions.

The problem of learning SVM, formulated as the task of separating learning vectors \( x_i \) into two classes of the destination values either \( d_i = 1 \) or \( d_i = -1 \), with maximal separation margin is reduced to the dual maximization problem of the function \( Q(\alpha) \), defined as follows [13, 14, 15],

\[
Q(\alpha) = \sum_{i=1}^{p} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{p} \alpha_i \alpha_j d_i d_j K(x_i, x_j)
\]

with the constraints

\[
\sum_{i=1}^{p} \alpha_i d_i = 0
\]

\[
0 \leq \alpha_i \leq C
\]

where \( C \) is a user-defined constant and \( p \) is the number of learning data pairs \((x_i, d_i)\). \( C \) is the regularizing parameter and determines the balance between the complexity of the network, characterized by the weight vector \( w \) and the error of classification of data. For the normalized input signals the value of \( C \) is usually much bigger than 1 and adjusted by cross validation.

The solution with respect to the Lagrange multipliers gives the optimal weight vector \( w_{opt} \), as \( w_{opt} = \sum_{i=1}^{p} \alpha_i d_i \phi(x_i) \). In this equation index \( s \) points to the set of \( N \) support vectors, i.e. the learning vectors \( x_s \), for which the relations

\[
d_i = \sum_{j=1}^{K} w_j \phi_j(x_i) + w_0 \geq 1 - \xi_i \quad (\xi_i \geq 0 \ - \ the \ slack \ variables)
\]

are fulfilled with the equality sign [12, 13]. The output signal \( y(x) \) of the SVM network in the retrieval mode (after learning) is determined as the function of kernels

\[
y(x) = \sum_{i=1}^{N} \alpha_i d_i K(x_i, x) + w_0
\]

and the explicit form of the nonlinear function \( \phi(x) \) need not be known. The value of \( y(x) \) greater than 0 is associated with 1 (membership of the particular class) and the negative one with \( -1 \) (membership of the opposite class). Although SVM separates the data only into two classes, the recognition of more classes is straightforward by applying either “one against one” or “one against all” methods [18, 19].

The important advantage of SVM approach is the transformation of the learning task to the quadratic programming problem. For this type of optimization there exist many highly effective learning algorithms [14, 16, 17], leading in almost all cases to the global minimum of the cost function and to the best possible choice of the parameters of the neural network.

III. GENERATION OF FEATURES

The recognition of the heart rhythms by neural network classifier requires generation of the input signals, representative for the task. A good recognition system should depend on the features representing the ECG signals in such a way, that the differences among the ECG waveforms are suppressed for the waveforms of the same type but are emphasized for the waveforms belonging to different types of beats. This is a very important item, since we observe high variation of signals among the same types of beats. Moreover all types of beats occupy similar range of amplitudes and frequencies. It is thus difficult to separate one from the other on the basis of only time or frequency representation.

We rely the recognition process of heart rhythm on the QRS complex of the ECG, proposing the description of it by higher order statistics (HOS) and the Hermite basis functions expansion.

A. HOS description of the ECG features

Three types of statistics have been taken into account: the second-, third- and fourth-order cumulants. The cumulants are the coefficients of the Taylor expansion of the cumulant generating function [9]. They can be expressed in terms of the well-known statistical moments as their linear or non-linear combinations. For zero mean statistical process \( x(t) \) the 2nd and 3rd order cumulants are equal to their corresponding moments \( c_2 = \mu_2 = m_2, c_3 = \mu_3 = m_3 \). The 4th order cumulant uses the information of the 4th and 2nd-order moments [9]

\[
c_4 = \mu_4 - \mu_2 \mu_2 - \mu_2 \mu_2 - m_4
\]

In these expressions \( c_n \) means the \( n \)th order cumulant and \( m_n \) the \( n \)th order statistical moment of the process \( x(k) \), while \( \tau_1, \tau_2, \tau_3 \) are the time lags.

![Fig. 1 The normal sinus rhythm (a) and its cumulant representations of the second (b), third (c) and fourth (d) orders](image-url)
same type of heart rhythm. It is well visible on the set of QRS complexes belonging to the normal sinus rhythm registered at different time for the same healthy person (Fig. 1). The cumulant characteristics are shown for different time lags (the diagonal slices). It is evident that the relative spread of the cumulant characteristics has been significantly reduced. For example the variance of the original normalized N-type beats presented in Fig. 1a was 0.0353. The same values for the second, third and fourth-order cumulants (Fig. 1b,c,d) are 0.0121, 0.0129 and 0.00528, respectively. By comparing HOS characteristics for different rhythm types we have observed that the differences among them have been increased and thanks to this are more visible and easier to distinguish.

As the features of the heart rhythm we have chosen the values of five points of QRS complex, distributed evenly within the QRS length and represented by the cumulants of the 2nd, 3rd and 4th orders (for the 3rd and 4th order cumulants the diagonal slices have been calculated). For 91-element vector representing QRS complex the cumulants corresponding to the time lags of 15, 30, 45, 60 and 75 have been chosen (see Fig. 1b, c, d). Additionally we have added two classical temporary features: one element corresponding to the instantaneous RR interval of the beat (the time span between two consecutive R points), calculated as the distance between the QRS peaks of the present and previous beat, and one element representing the average RR interval of 10 last beats. In this way each beat is represented by the 17-element vector, with the first 15 coordinates corresponding to higher order statistics of QRS complex (the second, third and fourth order cumulants, each represented by 5 values) and the last two - the temporary features of the actual QRS.

For the purpose of standardization of different input signals, some normalization of them was necessary. For this purpose each component of cumulant vector has been divided by the same value, chosen in a way to transform the maximum value of the cumulants to the number close to one. The other two temporary features have been normalized in a similar way.

B. Hermite basis functions expansion

Another representation of the ECG waveform, proposed in the paper [5] is via Hermite basis functions. It exploits the similarity of the shapes of these functions and QRS complexes of the ECG curves. The features characterizing the shape of the ECG beat are formed by the coefficients of the Hermite basis functions expansion. Let us denote by \( x(t) \) the QRS complex of the ECG curve. Its expansion into Hermite series yields

\[
x(t) = \sum_{n=0}^{N-1} c_n \phi_n(t, \sigma)
\]

where \( c_n \) are the expansion coefficients, \( \sigma \) is the width parameter and \( \phi_n(t, \sigma) \) - the Hermite basis functions of \( n \)th order defined as follows [5,22]

\[
\phi_n(t, \sigma) = \frac{1}{\sqrt{\pi} \sigma^n n!} e^{-t^2/2\sigma^2} H_n(t/\sigma)
\]

and \( H_n(t/\sigma) \) is the Hermite polynomial of \( n \)th order. The Hermite polynomials are generated using the recurrence relation

\[
H_n(x) = 2xH_{n-1}(x) - (n-1)H_{n-2}(x)
\]

with \( H_0(x) = 1 \) and \( H_1(x) = 2x \), for \( n=2, 3, ..., \) [22]. The higher the order of the Hermite polynomial, the higher is its frequency of changes within the time domain and the better is its capability to reconstruct quick changes of the ECG paradigms.

In our approach, we present the QRS segment of ECG signal as 91 data points around the R peak (45 points before and 45 ones after). The data sample rate of 360 Hz yields window of 250 ms, which is long enough to cover most QRS signals. The data has been also expanded by adding 45 zeros to each end of the beats. Subtracting the mean level of the first and the last points normalizes the ECG signals. The width \( \sigma \) has been proportional to the width of the QRS complex. These modified QRS complexes are decomposed into a linear combination of Hermite basis functions. The expansion coefficients \( c_n \) are obtained by minimizing the sum of squared errors, defined in the following way

\[
E = \sum_{j=0}^{N-1} \left| x(t_j) - \sum_{n=0}^{N-1} c_n \phi_n(t_j, \sigma) \right|^2
\]

This error function represents the set of linear equations with respect to the coefficients \( c_n \). They have been solved by using SVD decomposition and pseudo-inverse technique [10]. Different numbers of Hermite basis functions have been used to extract the features of QRS complexes. The experiments have shown that with 15 Hermite coefficients, including width coefficient \( \sigma \), we obtained quite good reconstruction of the most important details of the curve. These coefficients together with 2 classical features (the instantaneous RR interval of the beat and the average RR interval of 10 last beats), form the 17-element feature vector \( x \) inputted to the SVM neural classifier.

IV. COMBINATION OF CLASSIFIERS BY WEIGHTED VOTING

The general classification system applying the combination of many individual classifiers by weighted voting is presented in Fig. 2. It contains \( M \) channels of individual classifiers combined into one classifying system by the integrating part of the network.

The sampled signals of the process form the \( n \)-dimensional vector \( \mathbf{x}_m \). This vector is transformed into different feature vectors by appropriate preprocessing blocks. Each classifier has \( N \) outputs corresponding to \( N \) classes and the output signals of each classifier form the vector \( \mathbf{y}_j \), for \( j=1, 2, ..., M \). These vectors, representing the classification results of different classifiers, are combined through the integrating matrix to form one output vector \( \mathbf{z} \) of the classifying system. The position of the highest value element of \( \mathbf{z} \) indicates the membership with the appropriate class.

In our solution we have applied the weighted voting of the individual classifiers for the final decision. The outputs of individual classifiers denoted by vectors \( \mathbf{y}_j \) are combined using the mixing matrix \( \mathbf{W} \). The result in the form of vector \( \mathbf{z} \) gives the final decision of classification. The input vector \( \mathbf{x}_m \) characterizing the process is assigned to the class indicated by the position of the element \( z_i \) of the
highest value.

The most important step in this approach is an adjustment of the values of elements of the mixing matrix W. We have proposed the approach based on the minimization of the sum of squared errors of the classification results of each classifier measured on the training data, using the pseudo-inverse technique and SVD technique. [10].

We assume that all M classifiers applied in the system may point to any of N-classes. Each classifier is excited by the corresponding feature vector \( \mathbf{x}_i \) \((i=1, 2, ..., p)\) where \( p \) denotes the number of learning pairs. Let us assume that \( j \)th classifier excited by \( \mathbf{x}_j \) generates the vector response, \( \mathbf{y}_j = \mathbf{y}_j(\mathbf{x}_j) = \left[ y_{j1}^{(i)}, y_{j2}^{(i)}, ..., y_{jN}^{(i)} \right] \) for \( j=1, 2, ..., M \). These vectors form matrix \( \mathbf{Y} \in \mathbb{R}^{p \times MN} \), defined as follows

\[
\mathbf{Y} = \begin{bmatrix}
\mathbf{y}_1^{(1)} & \mathbf{y}_2^{(1)} & \cdots & \mathbf{y}_N^{(1)} \\
\mathbf{y}_1^{(2)} & \mathbf{y}_2^{(2)} & \cdots & \mathbf{y}_N^{(2)} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{y}_1^{(p)} & \mathbf{y}_2^{(p)} & \cdots & \mathbf{y}_N^{(p)}
\end{bmatrix}
\]

(9)

Let us assume that the target vector \( \mathbf{d}^{(i)} = \left[ d_1^{(i)}, d_2^{(i)}, ..., d_{MN}^{(i)} \right] \) for \( i=1, 2, ..., p \), represents the membership of the beat associated with the feature vectors \( \mathbf{x}_i \) to the appropriate class. It is the vector composed of zeros and only one unity element, pointing to the appropriate class. These vectors form \( p \times N \) matrix \( \mathbf{D} \)

\[
\mathbf{D} = \begin{bmatrix}
d_1^{(1)} & d_2^{(1)} & \cdots & d_{MN}^{(1)} \\
d_1^{(2)} & d_2^{(2)} & \cdots & d_{MN}^{(2)} \\
\vdots & \vdots & \ddots & \vdots \\
d_1^{(p)} & d_2^{(p)} & \cdots & d_{MN}^{(p)}
\end{bmatrix}
\]

(10)

Let \( \mathbf{W} \in \mathbb{R}^{MN \times N} \) be the mixing matrix defined as follows

\[
\mathbf{W} = \begin{bmatrix}
\mathbf{w}_1^{(1)} & \mathbf{w}_2^{(1)} & \cdots & \mathbf{w}_N^{(1)} \\
\mathbf{w}_1^{(2)} & \mathbf{w}_2^{(2)} & \cdots & \mathbf{w}_N^{(2)} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{w}_1^{(N)} & \mathbf{w}_2^{(N)} & \cdots & \mathbf{w}_N^{(N)}
\end{bmatrix}
\]

(11)

where \( \mathbf{w}_i^{(k)} = \left[ w_{i1}^{(k)}, w_{i2}^{(k)}, ..., w_{iMN}^{(k)} \right] \) denotes an initially unknown weighting vector of \( i \)th classifier \((i=1, 2, ..., M)\) for the data belonging to the \( k \)th class. All weights forming matrix \( \mathbf{W} \) will be determined through minimization of the mean squared errors over all training data points, that is, by solving the following optimization problem

\[
\min_{\mathbf{W}} \left\| \mathbf{YW} - \mathbf{D} \right\|^2
\]

(12)

Such formulation of problem means, that response of the network to all learning data is equally important for the choice of the mixing weights. The solution of it can be found in one step by applying the pseudo-inverse technique, where

\[
\mathbf{W} = \mathbf{Y}^{+} \mathbf{D}
\]

(13)

The pseudo-inverse matrix \( \mathbf{Y}^{+} \) can be easily calculated by using the singular value decomposition (SVD) of the matrix \( \mathbf{Y} \) [10].

In the solution of the heart beat recognition problem we have applied the system combining two classifiers presented in the previous section. Both use SVM as the classifier. They differ by the feature vectors used as the input to the SVM. In the first approach the input vector is formed by the HOS features (named HOS) while in the second by the Hermite coefficients (named HER).

In the numerical experiments we have used the ECG data from the MIT BIH Arrhythmia Database [8] corresponding to the normal sinus rhythm (N) and 12 types of arrhythmias. The following arrhythmias have been considered in recognition: left bundle branch block (L), right bundle branch block (R), atrial premature beat (A), aberrated atrial premature beat (a), nodal (junctional) premature beat (J), ventricular premature beat (V), fusion of ventricular and normal beat (F), ventricular flutter wave (I), nodal (junctional) escape beat (i), ventricular escape beat (E), supraventricular premature beat (S) and fusion of paced and normal beat (f). These beats come from 52 patients of the database. Due to the scarcity of data corresponding to some beat types the number of data belonging to each beat type was variable. Limiting the number of some beats we tried to provide the minimal balance for the number of beats among all classes under consideration. As a result of such strategy for data selection, the total number of data used in learning neural networks was equal 6690. Another 6095 data points have been left for testing.

The SVM neural network in both cases had the same number of inputs (17) and one output. In learning multi-class recognition problem we have applied the one-against-one strategy leading to many network structures adapted for the recognition between two classes at one time [18,19]. All networks have been trained using Mangasarian algorithm [16], adjusting the support vectors and their position in multidimensional space.

We have used Gaussian radial kernels \( K(\mathbf{x}, \mathbf{x}_i) = \exp(-\gamma \mathbf{x}^T \mathbf{x}_i) \) with \( \gamma=2 \). The parameter C used in experiments was set to \( C=100 \). These values have been selected after series of numerical experiments on the learning and testing data to get the best generalization ability. Generally they should be chosen in such a way that at smallest possible number of support vectors the best performance of the network on the testing data is observed. As a result of such a learning procedure we have got the SVM networks with different number of hidden units (support vectors) changing from network to network. The total number of support vectors for all the networks was close to 1200. The parameters of the networks were then fixed and used in the retrieval mode on the data set, not used in learning.

Table I and II present the results of recognition of 13 classes of heart rhythms by applying individual classifiers. The succeeding columns give the number of learning and testing beats, number of errors in learning and testing mode, as well as relative average errors in both modes, expressed in percentages. Table I presents the classification results for HOS preprocessing of ECG waveforms while Table II for Hermite basis functions representation of the ECG.

For the purpose of simple comparison of both systems we have calculated the average of all mean errors in learning and testing modes. To get reliable results independent of the number of beats belonging to different classes we have used a simple average of all mean errors. The total average error for all data, counted in this way was 4.23% for Hermite preprocessing and 5.74% for HOS.
Table I

<table>
<thead>
<tr>
<th>Beat type</th>
<th>Number of learning beats</th>
<th>Total number of learning errors</th>
<th>Average relative learning error</th>
<th>Number of testing beats</th>
<th>Total number of testing errors</th>
<th>Average relative testing error</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>700</td>
<td>23</td>
<td>3.29%</td>
<td>500</td>
<td>16</td>
<td>3.20%</td>
</tr>
<tr>
<td>R</td>
<td>600</td>
<td>10</td>
<td>1.67%</td>
<td>400</td>
<td>5</td>
<td>1.25%</td>
</tr>
<tr>
<td>A</td>
<td>484</td>
<td>25</td>
<td>5.17%</td>
<td>418</td>
<td>24</td>
<td>5.74%</td>
</tr>
<tr>
<td>V</td>
<td>1272</td>
<td>26</td>
<td>2.04%</td>
<td>1237</td>
<td>34</td>
<td>2.75%</td>
</tr>
<tr>
<td>I</td>
<td>272</td>
<td>12</td>
<td>4.41%</td>
<td>200</td>
<td>13</td>
<td>6.50%</td>
</tr>
<tr>
<td>E</td>
<td>55</td>
<td>2</td>
<td>3.64%</td>
<td>50</td>
<td>3</td>
<td>6.00%</td>
</tr>
<tr>
<td>N</td>
<td>2000</td>
<td>17</td>
<td>0.85%</td>
<td>2000</td>
<td>57</td>
<td>2.85%</td>
</tr>
<tr>
<td>a</td>
<td>67</td>
<td>6</td>
<td>8.96%</td>
<td>64</td>
<td>13</td>
<td>20.31%</td>
</tr>
<tr>
<td>f</td>
<td>200</td>
<td>14</td>
<td>7.00%</td>
<td>200</td>
<td>13</td>
<td>6.50%</td>
</tr>
<tr>
<td>F</td>
<td>371</td>
<td>9</td>
<td>2.43%</td>
<td>370</td>
<td>19</td>
<td>5.14%</td>
</tr>
<tr>
<td>j</td>
<td>117</td>
<td>20</td>
<td>17.09%</td>
<td>105</td>
<td>14</td>
<td>13.33%</td>
</tr>
<tr>
<td>J</td>
<td>40</td>
<td>4</td>
<td>10.00%</td>
<td>39</td>
<td>3</td>
<td>7.69%</td>
</tr>
<tr>
<td>S</td>
<td>512</td>
<td>5</td>
<td>0.98%</td>
<td>512</td>
<td>2</td>
<td>0.39%</td>
</tr>
<tr>
<td>Total</td>
<td>6690</td>
<td>173</td>
<td>5.19%</td>
<td>6095</td>
<td>216</td>
<td>6.28%</td>
</tr>
</tbody>
</table>

preprocessing. Counting them as the ratio of all misclassification numbers to the number of beats we would get 2.15% (Hermite) and 3.04% (HOS) average error values.

Generally the worst results have been observed for the beat types of smaller representations in the MIT BIH database. Both classifiers are fairly sensitive to this number and their efficiencies depend on it significantly. It is also seen that the accuracy of recognition changes greatly for different beat types. The important fact is that the efficiency of both classifiers on the same rhythm type is different, leaving space for further improvement through their combination into one expert system. The combination of networks should be made in a way to take advantage of better performing classifier. We have applied the weighted voting of the classifiers, according to the procedure described in section 4.

Table III presents the results of classification for different heartbeats, obtained by combining classifiers. As it is seen the results of classification have been improved significantly. This is clearly visible for the data used only in testing, where the highest improvement has been obtained. This validates our claim that combination of a few classifiers into one classification system increases the generalization ability of the whole system.

Table IV summarizes recognition results for all classes of heart beats by individual classifiers and by our expert system combining both classifiers. Column 1 represents the classification method, column 2 – the total number of learning errors, column 3 – the average relative learning errors, column 4 – the total number of testing errors on the data not used in learning and column 5 – the average relative testing errors. The average relative errors given in the table have been calculated in the same way as it was done in the previous cases (the simple mean of the relative errors of the recognition for all beat types). Thanks to such definition of mean, the values of final results shown in the table, are least dependent on the number of beats, belonging to different classes, that have taken part in experiments.

The results are given for individual classifiers named after the preprocessing methods (HOS, HER) and for combination of individual classifiers. It is evident that the best results have been obtained for weighted voting strategy.
the final result of 1.82% of misclassifications. This is of course overly optimistic for the data presently available in the MIT BIH Arrhythmia database.

However, on the basis of all experimental results it is evident that good recognition rates have been achieved for those beats, for which the large number of cases were available in the database. Hence, increasing the number of all beats types above some level (for example over 500) should also increase significantly the overall accuracy of recognition.

### Table III

**THE DETAILED RESULTS OF DIFFERENT HEARTBEAT TYPE RECOGNITION AT APPLICATION OF COMBINATION OF BOTH CLASSIFIERS**

<table>
<thead>
<tr>
<th>Beat type</th>
<th>Number of learning beats</th>
<th>Total number of learning errors</th>
<th>Average relative learning error</th>
<th>Number of testing beats</th>
<th>Total number of testing errors</th>
<th>Average relative testing error</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>700</td>
<td>7</td>
<td>1.00%</td>
<td>500</td>
<td>17</td>
<td>3.40%</td>
</tr>
<tr>
<td>R</td>
<td>600</td>
<td>2</td>
<td>0.33%</td>
<td>400</td>
<td>4</td>
<td>1.00%</td>
</tr>
<tr>
<td>A</td>
<td>484</td>
<td>20</td>
<td>4.13%</td>
<td>418</td>
<td>29</td>
<td>6.94%</td>
</tr>
<tr>
<td>V</td>
<td>1272</td>
<td>14</td>
<td>1.10%</td>
<td>1237</td>
<td>23</td>
<td>1.86%</td>
</tr>
<tr>
<td>I</td>
<td>272</td>
<td>5</td>
<td>1.84%</td>
<td>200</td>
<td>6</td>
<td>3.00%</td>
</tr>
<tr>
<td>E</td>
<td>55</td>
<td>0</td>
<td>0%</td>
<td>50</td>
<td>2</td>
<td>4.00%</td>
</tr>
<tr>
<td>N</td>
<td>2000</td>
<td>12</td>
<td>0.60%</td>
<td>2000</td>
<td>44</td>
<td>2.20%</td>
</tr>
<tr>
<td>a</td>
<td>67</td>
<td>2</td>
<td>2.99%</td>
<td>64</td>
<td>4</td>
<td>6.25%</td>
</tr>
<tr>
<td>f</td>
<td>200</td>
<td>1</td>
<td>0.50%</td>
<td>200</td>
<td>3</td>
<td>1.50%</td>
</tr>
<tr>
<td>F</td>
<td>371</td>
<td>3</td>
<td>0.81%</td>
<td>370</td>
<td>10</td>
<td>2.70%</td>
</tr>
<tr>
<td>j</td>
<td>117</td>
<td>3</td>
<td>2.56%</td>
<td>105</td>
<td>10</td>
<td>9.52%</td>
</tr>
<tr>
<td>J</td>
<td>40</td>
<td>3</td>
<td>7.50%</td>
<td>39</td>
<td>4</td>
<td>10.26%</td>
</tr>
<tr>
<td>S</td>
<td>512</td>
<td>2</td>
<td>0.39%</td>
<td>512</td>
<td>3</td>
<td>0.59%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>6690</strong></td>
<td><strong>74</strong></td>
<td><strong>1.83%</strong></td>
<td><strong>6095</strong></td>
<td><strong>159</strong></td>
<td><strong>4.09%</strong></td>
</tr>
</tbody>
</table>

Some additional observation on the combination methods is necessary in this place. The other possible solution seems to be simple concatenation of both feature vectors (based on HOS and Hermite) into the longer one (in our case it would be 32-element vector) and adaptation of the proper SVM network as the classifier. However the numerical experiments of such combination have shown worse accuracy of classification in comparison to the integration approach presented in the paper.

It is also interesting to compare our results to the others published in the literature. However it should be stressed that most of the papers considered only a small number of chosen rhythm types. Applying HOS preprocessing and hybrid neural classifier, the paper [6] has reported the mean error of 3.94% for 7 rhythm types. Application of multilayer perceptron and Fourier preprocessing [4] for 3-rhythm types recognition, resulted in 2% of the mean error. Application of LVQ and multilayer perceptron reported in the paper [7] for 2 rhythm types has resulted in 3.2% mean error. Lagerholm at al [5] have presented the most complete recognition system among all beat types existed in the MIT BIH AD and demonstrated the most integrated recognition procedure. The misclassification ratio has been presented on different platforms. The most meaningful seems to be the confusion matrix method. For all 13 beat types, present in the MIT BIH AD they got the average error rate of 7.12%, calculated as the ordinary mean of all errors.

The mixed expert approach (MEA) presented in [1] is the most similar in principle to our method. The paper used so-called Global Expert (GE), Local Expert (LE) and the mixture of them (MEA). The experts used SOM and LVQ methods and have been tested on beat types of MIT BIH AD, classified into 4 categories. The average classification rate obtained by MEA has been greatly improved over the worst (GE) and reached 94%, which corresponds to 6% of the mean error. Even better results have been obtained by applying LE - the average error was only 4.1%.

This comparison indicates high efficiency of the proposed method. However it should be noted, that in each case different number of beat types belonging to different patients have been recognized, thus it is really difficult to compare the results in a fair, objective way.

### VI. Conclusions

The paper has presented the expert system, based on the application of Support Vector Machine for reliable heart beat recognition on the basis of the ECG waveform. Two different preprocessing methods of the data, cooperating with SVM classifier, have been integrated into one expert system to improve the overall accuracy of heart beat recognition. The classifiers have been combined using weighted voting principle.

The experiments of the recognition of 12 types of arrhythmias and normal sinus rhythm were carried out on MIT BIH Arrhythmia Database. They have shown that the performance of individual classifiers could be significantly improved by the combination of classifiers proposed in the paper. This confirms our conjecture that a highly reliable classifier can be obtained by combining a number of classifiers which exhibit an average performance.

### REFERENCES


*Stanislaw Osowski* was born in Poland in 1948. He received the M.Sc., Ph.D., and Dr.Sc. degrees from the Warsaw University of Technology, Warsaw, Poland, in 1972, 1975, and 1981, respectively, all in electrical engineering. Currently he is a professor of electrical engineering at the Institute of the Theory of Electrical Engineering and Electrical Measurements, Warsaw University of Technology. His research and teaching interest are in the areas of neural networks, optimization techniques, and digital signal processing. He is an author or co-author of more than 200 scientific papers and ten books.

*Tran Hoai Linh* was born in Vietnam in 1974. He received the M.Sc., and Ph.D. degrees from Warsaw University of Technology, Warsaw, Poland, in 1997 and 2000, respectively, in computer science and electrical engineering. Currently he conducts post-doctoral studies in the Institute of the Theory of Electrical Engineering and Electrical Measurements, Warsaw University of Technology and is employed at Hanoi University of Technology (Vietnam). His research interest is in neural networks and biomedical signal processing. He is an author or co-author of 30 scientific papers.

*Tomasz Markiewicz* was born in 1976 in Warsaw, Poland. He received his M.Sc. degree from Warsaw University of Technology in 2001 in electrical engineering. Currently he is working towards Ph.D degree at Warsaw University of Technology, Warsaw, Poland. His major interest is in Support Vector Machine and its application in biomedical signal and image processing.