ABSTRACT
In this paper, the multiclass support vector machine (SVM) with the error correcting output codes (ECOC) was presented for the multiclass time-varying biomedical signals (electrocardiogram signals) classification problems. Decision making was performed in two stages: feature extraction by computing the wavelet coefficients and classification using the classifier trained on the extracted features. The purpose was to determine an optimum classification scheme for this problem and also to infer clues about the extracted features. The research demonstrated that the wavelet coefficients are the features which well represent the studied time-varying biomedical signals and the multiclass SVMs trained on these features achieved high classification accuracies.

KEY WORDS
Multiclass support vector machine (SVM), Wavelet coefficients, Time-varying biomedical signals

1. Introduction
The entire process of methodologies developed for automated diagnosis can generally be subdivided into a number of disjoint processing modules: preprocessing, feature extraction/selection, and classification (Fig. 1). Signal/image acquisition, artefact removing, averaging, thresholding, signal/image enhancement and edge detection are the main operations in the course of preprocessing. The accuracy of signal/image acquisition is of great importance since it contributes significantly to the overall classification result. The markers are subsequently processed by the feature extraction module. The module of feature selection is an optional stage, whereby the feature vector is reduced in size including only, from the classification viewpoint, what may be considered as the most relevant features required for discrimination. The classification module is the final stage in automated diagnosis. It examines the input feature vector and based on its algorithmic nature, produces a suggestive hypothesis [1,2].

In the feature extraction stage, numerous different methods can be used so that several diverse features can be extracted from the same raw data. The wavelet transform (WT) provides very general techniques which can be applied to many tasks in signal processing. Wavelets are ideally suited for the analysis of sudden short-duration signal changes. One very important application is the ability to compute and manipulate data in compressed parameters which are often called features [3]. Thus, the time-varying biomedical signal, consisting of many data points, can be compressed into a few parameters by the usage of the WT. These parameters characterize the behavior of the time-varying biomedical signal. This feature of using a smaller number of parameters to represent the time-varying biomedical signal is particularly important for recognition and diagnostic purposes [4-6]. Therefore, in order to discriminate the time-varying biomedical signals, multiclass support vector machine (SVM) with the error correcting output codes (ECOC) combined with wavelet preprocessing was implemented.

The objective in the field of automated detection of changes in time-varying biomedical signals (electrocardiogram signals – ECG) is to extract the representative features of the signals under study and to present the accurate classification model. As in traditional pattern recognition systems, the present model consists of three main modules: a feature extractor that generates a feature vector from the time-varying biomedical signals, feature selection (wavelet coefficients), and a feature classifier that outputs the class based on the features (Fig. 1). A significant contribution of this work was to examine the multiclass SVM with the ECOC on the ECG signals classification. The ability of WT to extract and localize specific transient patterns from the signal makes them a natural complement to the present applications of the SVMs. Each studied segment of the signals under study was wavelet decomposed into multi-level low- and high-pass subbands, which were then input into the SVMs for training and testing purposes. High accuracies were achieved by using the multiclass SVM trained on the wavelet coefficients.

The outline of this study is as follows. In section 2, the studied time-varying biomedical signals are briefly described. In section 3, the feature extraction by discrete wavelet transform (DWT) is explained. In section 4, brief review of the multiclass SVM with the ECOC is presented. In section 5, the results of application of the classifiers trained on wavelet coefficients to the studied...
time-varying biomedical signals are presented. Discussion of the presented results is performed in the light of existing studies in the literature. Finally, in section 6 the study is concluded.

Fig. 1. General structure of the implemented time-varying biomedical signals classifiers

2. Description of Studied Time-Varying Biomedical Signals

The ECG is the record of variation of bioelectric potential with respect to time as the human heart beats. Electrocardiography is an important tool in diagnosing the condition of the heart [4-8]. It provides valuable information about the functional aspects of the heart and cardiovascular system. Early detection of heart diseases/abnormalities can prolong life and enhance the quality of living through appropriate treatment. The ECG signals (normal beat, congestive heart failure beat, ventricular tachyarrhythmia beat) obtained from the Physiobank database [9] were classified. A rectangular window, which was formed by 256 discrete data, was selected so that it contained a single ECG beat.

3. Feature Extraction by Discrete Wavelet Transform

The WT can be thought of as an extension of the classic Fourier transform, except that, instead of working on a single scale (time or frequency), it works on a multi-scale basis. In the procedure of multiresolution decomposition of a signal \( x[n] \), each stage consists of two digital filters and two downsamplers by 2. The first filter, \( g[\cdot] \) is the discrete mother wavelet, high-pass in nature, and the second, \( h[\cdot] \) is its mirror version, low-pass in nature. The downsampled outputs of first high-pass and low-pass filters provide the detail, \( D_1 \) and the approximation, \( A_1 \), respectively. The first approximation, \( A_1 \) is further decomposed and this process is continued.

All wavelet transforms can be specified in terms of a low-pass filter \( h \), which satisfies the standard quadrature mirror filter condition:

\[
H(z)H(z^{-1}) + H(-z)H(-z^{-1}) = 1, \tag{1}
\]

where \( H(z) \) denotes the z-transform of the filter \( h \). Its complementary high-pass filter can be defined as

\[
G(z) = zH(-z^{-1}). \tag{2}
\]

A sequence of filters with increasing length (indexed by \( i \)) can be obtained:

\[
H_{i+1}(z) = H(z^{2^i})H_i(z), \quad i = 0, \ldots, l - 1 \tag{3}
\]

with the initial condition \( H_0(z) = 1 \). It is expressed as a two-scale relation in time domain

\[
h_{i+1}(k) = [h]_{2} \ast h_i(k) \quad \text{and} \quad g_{i+1}(k) = [g]_{2} \ast h_i(k), \tag{4}
\]

where the subscript \([\cdot]_m \) indicates the up-sampling by a factor of \( m \) and \( k \) is the equally sampled discrete time.

The normalized wavelet and scale basis functions \( \varphi_{i,j}(k) \), \( \psi_{i,j}(k) \) can be defined as

\[
\varphi_{i,j}(k) = 2^{i/2} h_i(k - 2^i l) \quad \psi_{i,j}(k) = 2^{i/2} g_i(k - 2^i l), \tag{5}
\]

where the factor \( 2^{i/2} \) is an inner product normalization, \( i \) and \( l \) are the scale parameter and the translation parameter, respectively. The DWT decomposition can be described as

\[
a_{i,j}(l) = x(k) \ast \varphi_{i,j}(k) \quad d_{i,j}(l) = x(k) \ast \psi_{i,j}(k), \tag{6}
\]

where \( a_{i,j}(l) \) and \( d_{i,j}(l) \) are the approximation coefficients and the detail coefficients at resolution \( i \), respectively [3].

4. Support Vector Machine

The SVM proposed by Vapnik [10] has been studied extensively for classification, regression and density estimation. Fig. 2 shows the architecture of the SVM. SVM maps the input patterns into a higher dimensional feature space through some nonlinear mapping chosen a priori. A linear decision surface is then constructed in this high dimensional feature space. Thus, SVM is a linear classifier in the parameter space, but it becomes a
nonlinear classifier as a result of the nonlinear mapping of
the space of the input patterns into the high dimensional
feature space. Training the SVM is a quadratic
optimization problem. The construction of a hyperplane
\( w^T x + b = 0 \) (\( w \) is the vector of hyperplane coefficients,
\( b \) is a bias term) so that the margin between the
hyperplane and the nearest point is maximized and can be
posed as the quadratic optimization problem. SVM has
been shown to provide high generalization ability. For a
two-class problem, assuming the optimal hyperplane in
the feature space is generated, the classification decision
of an unknown pattern \( y \) will be made based on
\[
    f(y) = \text{sgn} \left( \sum_{i=1}^{N} \alpha_i y_i K(x_i, y) + b \right)
\]
where \( \alpha_i \geq 0, \quad i = 1, 2, \ldots, N \) are nonnegative Lagrange
multipliers that satisfy \( \sum_{i=1}^{N} \alpha_i y_i = 0 \), \( \{y_i \mid y_i \in [-1, +1]\} \)
are class labels of training patterns \( \{x_i \mid x_i \in R^N\} \), and
\( K(x_i, y) \) for \( i = 1, 2, \ldots, N \) represents a symmetric positive
definite kernel function that defines an inner product in
the feature space. This shows that \( f(y) \) is a linear
combination of the inner products or kernels. The kernel
function enables the operations to be carried out in the
input space rather than in the high-dimensional feature
space. Some typical examples of kernel functions are
\( K(u, v) = v^T u \) (linear SVM); \( K(u, v) = (v^T u + 1)^n \)
(polynomial SVM of degree \( n \)); \( K(u, v) = \exp(-||u - v||^2 / 2\sigma^2) \) (radial basis function –
RBF SVM); \( K(u, v) = \tanh(\kappa v^T y + \theta) \) (two layer neural
SVM) where \( \sigma, \kappa, \theta \) are constants [10,11]. However, a
proper kernel function for a certain problem is dependent
on the specific data and till now there is no good method
on how to choose a kernel function. In this study, the
choice of the kernel functions was studied empirically and
optimal results were achieved using RBF kernel function.

The SVM is a binary classifier which can be extended by
fusing several of its kind into a multiclass classifier. In
this study, SVM decisions are fused using the ECOC
approach, adopted from the digital communication theory
[12]. In the ECOC approach, up to \( 2^n - 1 \) (where \( n \) is the number of classes) SVMs are trained, each of them
aimed at separating a different combination of classes.
For 3 classes (A, B, and C) 3 classifiers are necessary: one SVM classifies A from B and C, a second SVM
classifies B from A and C and a third SVM classifies C
from A and B. The multiclass classifier output code for a
pattern is a combination of targets of all the separate
SVMs. That is in the example, vectors from classes A, B,
and C have codes \((1, -1, -1), (-1, 1, -1), \) and \((-1, -1, 1)\),
respectively. If each of the separate SVMs classifies a
pattern correctly, the multiclass classifier target code is
met and the ECOC approach reports no error for that
pattern. However, if at least one of the SVMs
misclassifies the pattern, the class selected for this pattern
is the one its target code closest in the Hamming distance
sense to the actual output code and this may be an
erroneous decision.

Fig. 2. Architecture of the SVM (\( N \) is the number of support vectors)
5. Results and Discussion

5.1 Computing Feature Vectors

The WT is better suited to analyzing nonstationary signals, since it is well localized in time and frequency. Selection of appropriate wavelet and the number of decomposition levels is very important in analysis of signals using the WT. The number of decomposition levels is chosen based on the dominant frequency components of the signal. The levels are chosen such that those parts of the signal that correlate well with the frequencies required for classification of the signal are retained in the wavelet coefficients. The number of decomposition levels of the ECG signals was chosen to be 4. Thus, the ECG signals were decomposed into the details $D_1 - D_4$ and one final approximation, $A_4$. The smoothing feature of the Daubechies wavelet of order 2 (db2) made it more suitable to detect changes of the ECG signals. Therefore, the wavelet coefficients of the ECG signals were computed using the db2. Since the computed discrete wavelet coefficients provide a compact representation that shows the energy distribution of the signal in time and frequency, the discrete wavelet coefficients were used as the inputs of the SVMs. For each ECG beat (256 samples), the detail wavelet coefficients ($d^k$, $k = 1, 2, 3, 4$) at the first, second, third

and fourth levels ($129 + 66 + 34 + 18$ coefficients) and the approximation wavelet coefficients ($a^4$) at the fourth level (18 coefficients) were computed. Then 265 wavelet coefficients were obtained for each ECG beat.

High-dimension of feature vectors increased computational complexity and therefore, in order to reduce the dimensionality of the extracted feature vectors, statistics over the set of the wavelet coefficients were used. The following statistical features were used to represent the time-frequency distribution of the studied time-varying biomedical signals:

1. Maximum of the wavelet coefficients in each subband.
2. Mean of the wavelet coefficients in each subband.
3. Minimum of the wavelet coefficients in each subband.
4. Standard deviation of the wavelet coefficients in each subband.

Table 1 presents the extracted features of exemplary records from different classes of the signals under study. The wavelet coefficients were computed using the MATLAB software package.

<table>
<thead>
<tr>
<th>ECG beat types</th>
<th>Extracted Features</th>
<th>Wavelet Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Subbands</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$D_1$</td>
</tr>
<tr>
<td>Normal beat</td>
<td>Maximum</td>
<td>0.2062</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>-0.0003</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>-0.1814</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>0.0436</td>
</tr>
<tr>
<td>Congestive heart failure beat</td>
<td>Maximum</td>
<td>0.1316</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>-0.0003</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>-0.1119</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>0.0259</td>
</tr>
<tr>
<td>Ventricular tachyarrhythmia</td>
<td>Maximum</td>
<td>0.1568</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>-0.0001</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>-0.0839</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>0.0232</td>
</tr>
</tbody>
</table>

5.2 Experiments for Implementation of SVMs

The key design decisions for the neural networks used in classification are the architecture and the training process. The adequate functioning of neural networks depends on the sizes of the training set and test set. In the classification of the ECG signals, the 270 vectors (90 vectors from each class) were used for training and the 270 vectors (90 vectors from each class) were used for testing.

The generalization ability of the SVM is controlled by two different factors: the training error rate and the capacity of the learning machine measured by its Vapnik-Chervonenkis (VC) dimension [10]. The smaller the VC dimension of the function set of the learning machine, the larger the value of training error rate. The tradeoff between the complexity of decision rule and training error rate can be controlled by changing a parameter $C$ [11] in the SVM. In order to have the best result, the SVMs were trained for different $C$ values. The best result was obtained for $C = 70$ in the testing procedure. Moreover, when $C = 70$, the number of support vectors in the SVMs training was found. Training algorithm of the SVM, based on quadratic programming, incorporates several optimization techniques such as decomposition and caching. The quadratic programming problem in the SVM was solved by using the MATLAB optimization toolbox. Multiclass SVM and the ECOC algorithm was employed to classify the studied time-varying biomedical signals.
As mentioned earlier, each of the SVMs of the three-class classifier used the RBF kernel functions. For the implementation of the SVMs with the RBF kernel functions, one has to assume a value for $\sigma$. The optimal $\sigma$ can only be found by systematically varying its value in the different training sessions. To do this, the support vectors were extracted from the training data file with an assumed $\sigma$ value. After the support vectors have been found and SVM constructed, the model was applied to 1/3 of the training data set to compute the misclassification rate. The $\sigma$ value was varied between 0.1 and 0.6, at interval of 0.1. The $\sigma = 0.2$ resulted in the minimum misclassification rate was thus chosen. Table 2 defines the network parameters of the classifier implemented in this research.

<table>
<thead>
<tr>
<th>Table 2. Network parameters of the classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifier</td>
</tr>
<tr>
<td>SVM</td>
</tr>
</tbody>
</table>

*Design of SVMs: Number of input neurons · support vectors · output neurons, respectively.

5.3 Classification Errors of SVMs

Classification results of the classifiers were displayed by a confusion matrix. In a confusion matrix, each cell contains the raw number of exemplars classified for the corresponding combination of desired and actual network outputs. The confusion matrix showing the classification results of the SVMs used for classification of the ECG signals is given in Table 3. From this matrix, one can tell the frequency with which a segment is misclassified as another. As it is seen from Table 3, normal beats are most often confused with congestive heart failure beats, likewise congestive heart failure beats with ventricular tachyarrhythmia beats.

The test performance of the SVMs can be determined by the computation of specificity, sensitivity and total classification accuracy. The specificity, sensitivity and total classification accuracy are defined as:

- **Specificity**: number of true negative decisions / number of actually negative cases
- **Sensitivity**: number of true positive decisions / number of actually positive cases
- **Total classification accuracy**: number of correct decisions / total number of cases

A true negative decision occurs when both the classifier and the physician suggested the absence of a positive detection. A true positive decision occurs when the positive detection of the classifier coincided with a positive detection of the physician.

In order to determine the performance of the SVM used for classification of the studied time-varying biomedical signals, the classification accuracies (specificity, sensitivity, total classification accuracy) on the test sets are presented in Table 4.

<table>
<thead>
<tr>
<th>Table 3. Confusion matrix of the SVM used for classification of the ECG signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desired Result</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Normal beat</td>
</tr>
<tr>
<td>Congestive heart failure beat</td>
</tr>
<tr>
<td>Ventricular tachyarrhythmia beat</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4. The classification accuracies of the SVMs used for classification of the ECG signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifiers</td>
</tr>
<tr>
<td>SVM used for ECG beats</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

5.4 Discussion

Based on the results of the present study and experience in the ECG signals classification problems, the followings can be emphasized:

1. Güler and Übeyli [5] used CNN to guide model selection for classification of four types of ECG beats (normal beat, congestive heart failure beat, ventricular tachyarrhythmia beat, atrial fibrillation beat) obtained from the Physiobank database. The ECG signals were decomposed into time-frequency representations using DWT and statistical features were calculated to depict their distribution. The ECG beats were classified with the accuracy of 96.94% by the CNN. The results of the present study indicated that usage of SVMs improve the classification accuracy of ECG beats.
2. Güler and Übeyli [13] used modified mixture of experts (MME) network structure to guide model selection for classification of five types of ECG beats (normal beat, congestive heart failure beat, ventricular tachyarrhythmia beat, atrial fibrillation beat, partial epilepsy beat) with diverse features (wavelet coefficients and Lyapunov exponents). The ECG beats were classified with the accuracy of 97.78% by the MME. The results of the present study showed that SVMs slightly improve the classification accuracy of ECG beats.

3. Osowski et al. [8] presented the application of SVM for reliable heartbeat recognition on the basis of the ECG waveform. They applied two different preprocessing methods for generation of features. One method involved the higher order statistics (HOS) while the second the Hermite characterization of QRS complex of the registered ECG waveform. The SVM had the same number of inputs and one output. In learning multiclass recognition problem, they applied the one-against-one strategy leading to many network structures adapted for the recognition between two classes at one time. The classification accuracy of their model was 95.77% for Hermite preprocessing and 94.26% for HOS preprocessing. The multiclass SVM and the ECOC algorithm used in the present study to classify the ECG signals indicated higher performance than that of the SVM presented by Osowski et al. [8].

4. Acdr [14] used six fast least square support vector machines (LSSVMs) for classification of six types of ECG beats obtained from the MIT-BIH database. The classification accuracy was 95.2% by the proposed fast LSSVMs together with discrete cosine transform. The results of the present study indicated that the usage of multiclass SVM with the ECOC significantly improve the classification accuracy of ECG beats.

6. Conclusion

The purpose of the present research was to investigate the accuracy of multiclass SVM with the ECOC trained on the wavelet coefficients for classification of the time-varying biomedical signals. The multiclass SVM showed a great performance since it maps the features to a higher dimensional space. This may be attributed to several factors including the training algorithms, estimation of the network parameters and the scattered and mixed nature of the features. The results of the present study demonstrated that the multiclass SVM can be used in classification of the ECG signals by taking into consideration the misclassification rates.

Acknowledgement

This study has been supported by Scientific Research Project of TOBB Economics and Technology University (Project No: ETU-BAP-2006/06).

References