Feature saliency using signal-to-noise ratios in automated diagnostic systems developed for ECG beats

İnan Gülera,*, Elif Derya Übeylib

aDepartment of Electronics and Computer Education, Faculty of Technical Education, Gazi University, 06500 Teknikokullar, Ankara, Turkey
bDepartment of Electrical and Electronics Engineering, Faculty of Engineering, TOBB Ekonomi ve Teknoloji Üniversitesi, Sıdıkoğlu, Ankara, Turkey

Abstract

Artificial neural networks (ANNs) have been used in a great number of medical diagnostic decision support system applications and within feedforward ANNs framework there are a number of established measures such as saliency measures for identifying important input features. By identifying a set of salient features, the noise in a classification model can be reduced, resulting in more accurate classification. In this study, a signal-to-noise ratio (SNR) saliency measure was employed to determine saliency of input features of multilayer perceptron neural networks (MLPNNs) used in classification of electrocardiogram (ECG) beats (normal beat, congestive heart failure beat, ventricular tachyarrhythmia beat, atrial fibrillation beat) obtained from the Physiobank database. The SNR saliency measure determines the saliency of a feature by comparing it to that of an injected noise feature and the SNR screening method utilizes the SNR saliency measure to select a parsimonious set of salient features. ECG signals were decomposed into time–frequency representations using discrete wavelet transform. Input feature vectors were extracted using statistics over the set of the wavelet coefficients. The MLPNNs used in the ECG beats-classification were trained for the SNR screening method. The application results of the SNR screening method to the ECG signals demonstrated that classification accuracies of the MLPNNs with salient input features are higher than that of the MLPNNs with salient and non-salient input features.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Feature saliency; Signal-to-noise ratio; Classification accuracy; ECG beats classification

1. Introduction

The electrocardiogram (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. 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. Therefore, numerous research and work analyzing the ECG signals have been reported (Foo, Stuart, Harvey, & Meyer-Baese, 2002; Maglaveras, Stamkopoulos, Diamantaras, Pappas, & Strintzis, 1998; Saxena, Kumar, & Hamde, 2002).

ECG signals may contain important pointers to the nature of diseases afflicting the heart. Since non-stationarities of the ECG signals, for effective diagnostics the study of ECG pattern and heart rate variability signal may have to be carried out over several hours. Thus, the volume of the data being enormous, the study is tedious and time consuming. Naturally, the possibility of the analyst missing (or misreading) vital information is high. Therefore, computer-based analysis and classification of diseases can be very helpful in diagnostics (Dokur & Ölmez, 2001; Kundu, Nasipuri, & Basu, 2000; Li, Zheng, & Tai, 1995; Simon & Eswaran, 1997; Sternickel, 2002). A number of quantitative models including linear discriminant analysis, logistic regression, k nearest neighbor, kernel density, recursive partitioning, and neural networks are being used in medical diagnostic support systems to assist human decision-makers in disease diagnosis (Kundu et al., 2000). Artificial neural networks (ANNs) have been used in a great number of...
medical diagnostic decision support system applications because of the belief that they have greater predictive power (Kordylewski, Graupe, & Liu, 2001).

Various methodologies of automated diagnosis have been adopted, however, the entire process can generally be subdivided into a number of disjoint processing modules: beat detection, feature extraction/selection, and classification. Features are used to represent patterns with minimal loss of important information. The feature vector, which is comprised of the set of all features used to describe a pattern, is a reduced-dimensional representation of that pattern. Medical diagnostic accuracies can be improved when the pattern is simplified through representation by important features (Belue & Bauer, 1995; Kwak & Choi, 2002). Feedforward ANNs have received a great deal of attention for their application to pattern recognition and function prediction problems (Basheer & Hajmeer, 2000; Chaudhuri & Bhattacharya, 2000; Haykin, 1994). Within the feedforward ANNs framework, there are a number of established measures for identifying important input features. Such measures are known as saliency measures (Steppe & Bauer, 1997). In recent years, the general problem of selecting a parsimonious salient feature set for ANNs has generating a great deal of interest (Bauer, Alsing, & Greene, 2000; Laine, Bauer, Lanning, Russell, & Wilson, 2002; Polakowski et al., 1997; Steppe & Bauer, 1996, 1997; Steppe, Bauer, & Rogers, 1996). Non-salient input features to an ANN classifier can have negative results. First of all, insignificant input features may reduce classification accuracy. In addition, as the number of features grows, the number of training vectors required grows exponentially. The signal-to-noise ratio (SNR) saliency measure determines the saliency of a feature by comparing it to that of an injected noise feature (Bauer et al., 2000).

Until now, there has been no study in the literature relating to the SNR saliency measure for determining the saliency of input features of ANNs used in classification of ECG beats. In the present study, the ECG signals obtained from the Physiobank database (Goldberger et al., 2000) were classified using multilayer perceptron neural networks (MLPNNs). Feature extraction from the ECG signals for classification of ECG beats (normal beat, congestive heart failure beat, ventricular tachyarrhythmia beat, atrial fibrillation beat) was performed using discrete wavelet transform (DWT). In this study, in order to reduce the dimensionality of the extracted input feature vectors, statistics over the set of the wavelet coefficients were used. Then the SNR saliency measure was employed to identify salient input features of the proposed MLPNNs. The SNR saliency measure was chosen because of its versatility during the training cycle. In addition to this, the SNR saliency measure appears highly robust relative to the effects of the weight initialization, ANN architecture, and the selection of training and test sets. The SNR screening method, which utilizes the SNR saliency measure to select a parsimonious set of salient features, was applied to the ECG signals.

Some conclusions were drawn concerning the improvement of classification accuracies of the MLPNNs with salient input features determined by the SNR saliency measure.

The outline of this study is as follows. In Section 2, we explain SNR saliency measure. In Section 3, we describe the SNR screening method for the SNR saliency measure application. In Section 4, we present description of neural network models including MLPNN and neural network architecture used in this study. We also present Levenberg–Marquardt training algorithm used for training the MLPNNs. In Section 5, we perform spectral analysis of the ECG signals using DWT in order to extract features characterizing the behavior of the signals under study. We present the application results of the SNR screening method to the ECG signals. Finally, in Section 6 we conclude the study.

### 2. SNR Saliency measure

When employing neural networks of any type, one objective is to limit the number of input features. The dilemma is that as the number of features increases, the number of training vectors required in the training set also increases. Intuitively an analyst would like to include only those features that make a significant contribution to the network. Feature saliency measures provide a way to measure the relative usefulness of features and a means to rank order the features. A partial derivative based saliency measure calculates the impact of each feature on a single hidden layer ANN output by calculating the sum of the partial derivatives of the outputs to that feature. The derivative can be written as a function of the input vector and the weights as follows

\[
\Lambda_i = \frac{1}{K} \frac{1}{M_{\text{train}}} \sum_{k=1}^{K} \sum_{m=1}^{M_{\text{train}}} \frac{\partial z_{k,m}(x_m, W)}{\partial x_{i,m}},
\]

where \(\Lambda_i\) is the partial derivative-based saliency measure for feature \(i = 1, \ldots, I\), \(K\) is the total number of output nodes, \(M_{\text{train}}\) is the total number of exemplars in the training set, \(z_{k,m}(x_m, W)\) is the actual output of output node \(k\) with input vector \(x_m\) for exemplar \(m = 1, \ldots, M_{\text{train}}\). The trained ANN weight matrix \(W\). All feature inputs are normalized. The partial derivative-based saliency measure can be used to rank order the features from least salient to most salient where lower saliency measure values indicate lower relative saliency and higher values indicate higher relative saliency.

A weight-based saliency measure determines the saliency of a feature by summing the squared values of the weights connecting feature \(i\) to the hidden nodes and can be written as follows

\[
\tau_i = \sum_{j=1}^{I} (w_{i,j}^1)^2,
\]
where $\tau_i$ is the saliency measure for feature $i$, $J$ is the number of hidden nodes, $w_{ij}^1$ is the first layer weight between input node $i$ and hidden node $j$. The rationale behind this measure is that the square of the first layer weights associated with a salient feature in a trained ANN will, in general, be significantly larger than that of a non-salient feature. The SNR saliency measure is similar to the weight-based saliency measure in that both rely on the sum of squared first layer weights. However, this measure is different from both the partial derivative-based saliency measure and the weight-based saliency measure because it directly compares the saliency of a feature to that of an injected noise feature. The SNR saliency measure is computed using the following

$$SNR_i = 10 \log_{10} \left( \frac{\sum_{j=1}^{J} (w_{ij}^1)^2}{\sum_{j=1}^{J} (w_{Nj}^1)^2} \right),$$

where $SNR_i$ is the value of the SNR saliency measure for feature $i$, $J$ is the number of hidden nodes, $w_{ij}^1$ is the first layer weight from node $i$ to node $j$, and $w_{Nj}^1$ is the first layer weight from the injected noise node $N$ to node $j$. The weights, $w_{ij}^1$, for the noise feature are initialized and updated in the same fashion as the weights, $w_{ij}^1$, emanating from the other features in the first layer. The injected noise feature is created such that its distribution follows that of a Uniform $(0,1)$ random variable. The scaled logarithmic transformation of the ratio converts the saliency measure to a decibel scale. We wish to use the SNR saliency measure, which takes into consideration the saliency of a feature relative to the saliency of a known irrelevant feature. To establish a working procedure for determining which features are significant, a noise variable is included as a feature input along with the original inputs to represent an absolutely insignificant piece of information. Because all features were normalized between zero and one, the added noise was taken as random samples from a Uniform $(0,1)$ distribution. The procedure for determining significant feature inputs by including a noise feature is outlined in Section 3.

The idea behind the SNR saliency measure extends the concept of the weight-based saliency measure. If a given feature is relevant to ANN output, then the first layer weights emanating from that feature’s input node should be moved in the weight space in a constant direction until the error is minimized. If, on the other hand, a given feature is not relevant to an ANN output, the updates to the first layer weights emanating from that feature’s input node should be random and simply fluctuate around zero. As such, the SNR saliency measure should be significantly larger than 0.0 for salient features and close to or less than 0.0 for non-salient features. The SNR saliency measure can be used to rank order the saliency of features where higher SNR saliency measure values correspond to higher feature saliency (Bauer et al., 2000).

3. SNR screening method

To apply the SNR concept, a screening method that provides a mechanism to potentially identify a parsimonious set of salient features by removing non-salient features while striving to maintain good generalization is developed. The SNR screening method is summarized below (Bauer et al., 2000):

1. Introduce a Uniform $(0,1)$ noise feature $\chi_N$ to the original set of features.
2. Standardize all features to zero mean and unit variance.
3. Randomly initialize the weights between $-0.001$ and 0.001.
4. Randomly select the training and test sets.
5. Begin to train the ANN.
6. After each epoch, compute the SNR saliency measure for each input feature.
7. Interrupt training when the SNR saliency measures for all input features have stabilized.
8. Compute the test set classification error.
9. Identify the feature with the lowest SNR saliency measure and remove it from further training.
10. Continue training the ANN.
11. Repeat steps 6–9 until all the features (except the noise feature) in the original set are removed from training.
12. Compute the reaction of the test set classification error due to the removal of the individual features.
13. Retain the first feature whose removal caused a significant increase in the test set classification error, as well as all features, which were removed after that first salient feature.
14. Retrain the ANN with only the parsimonious set of salient input features.

4. Description of artificial neural network models

4.1. Multilayer perceptron neural network

ANNs may be defined as structures comprised of densely interconnected adaptive simple processing elements (neurons) that are capable of performing massively parallel computations for data processing and knowledge representation. ANNs can be trained to recognize patterns and the non-linear models developed during training allow neural networks to generalize their conclusions and to make application to patterns not previously encountered (Bashier & Hajmeer, 2000; Chaudhuri & Bhattacharya, 2000; Haykin, 1994). The MLPNN, which has features such as the ability to learn and generalize, smaller training set requirements, fast operation, ease of implementation and is therefore the most commonly used neural network architecture, is shown in Fig. 1. As shown in Fig. 1, a MLPNN consists of (i) an input layer with neurons representing input variables to the problem; (ii) an output layer with neurons
representing the dependent variables (what is being modeled); and (iii) one or more hidden layers containing neurons to help capture the non-linearity in the data.

The MLPNN is a non-parametric technique for performing a wide variety of detection and estimation tasks (Basheer & Hajmeer, 2000; Chaudhuri & Bhattacharya, 2000). In the MLPNN, each neuron in the hidden layer sums its input signals \( x_i \) after multiplying them by the strengths of the respective connection weights \( w_{ji} \) and computes its output \( y_j \) as a function of the sum:

\[
y_j = f \left( \sum w_{ji} x_i \right),
\]

where \( f \) is activation function that is necessary to transform the weighted sum of all signals impinging onto a neuron. The activation function \( f \) can be a simple threshold function, or a sigmoidal, hyperbolic tangent, or radial basis function. In the present study, in the hidden layer and the output layer, the activation function \( f \) was the sigmoidal function:

\[
f(\xi) = \frac{1}{1 + e^{-\xi}}.
\]

The sum of squared differences between the desired and actual values of the output neurons \( E \) is defined as

\[
E = \frac{1}{2} \sum_j (d_j - y_j)^2,
\]

where \( d_j \) is the desired value of output neuron \( j \) and \( y_j \) is the actual output of that neuron. Each weight \( w_{ji} \) is adjusted to reduce \( E \) as rapidly as possible. How \( w_{ji} \) is adjusted depends on the training algorithm adopted (Basheer & Hajmeer, 2000; Chaudhuri & Bhattacharya, 2000; Haykin, 1994).

Training algorithms are an integral part of ANN model development. An appropriate architecture may still fail to give a better model, unless trained by a suitable training algorithm. A good training algorithm will shorten the training time, while achieving a better accuracy. Therefore, training process is an important characteristic of the ANNs, whereby representative examples of the knowledge are iteratively presented to the network, so that it can integrate this knowledge within its structure. There are a number of training algorithms used to train a MLPNN and a frequently used one is called the backpropagation training algorithm (Basheer & Hajmeer, 2000; Chaudhuri & Bhattacharya, 2000; Haykin, 1994). The backpropagation algorithm, which is based on searching an error surface using gradient descent for points with minimum error, is relatively easy to implement. However, backpropagation has some problems for many applications. The algorithm is not guaranteed to find the global minimum of the error function since gradient descent may get stuck in local minima, where it may remain indefinitely. In addition to this, long training sessions are often required in order to find an acceptable weight solution because of the well-known difficulties inherent in gradient descent optimization. Therefore, a lot of variations to improve the convergence of the backpropagation have been proposed. Optimization methods such as second-order methods (conjugate gradient, quasi-Newton, Levenberg–Marquardt) have also been used for ANN training in recent years. The Levenberg–Marquardt algorithm combines the best features of the Gauss–Newton technique and the steepest-descent algorithm, but avoids many of their limitations. In particular, it generally does not suffer from the problem of slow convergence (Battiti, 1992; Hagan & Menhaj, 1994). Therefore, in this study the MLPNNs were trained with the Levenberg–Marquardt algorithm.

4.2. Levenberg–Marquardt algorithm

ANN training is usually formulated as a non-linear least-squares problem. Essentially, the Levenberg–Marquardt algorithm is a least-squares estimation algorithm based on the maximum neighborhood idea. Let \( E(w) \) be an objective error function made up of \( m \) individual error terms \( e_i^2(w) \) as follows

\[
E(w) = \sum_{i=1}^{m} e_i^2(w) = \|f(w)\|^2,
\]

where \( e_i^2(w) = (y_{di} - y_i)^2 \) and \( y_{di} \) is the desired value of output neuron \( i \), \( y_i \) is the actual output of that neuron.

It is assumed that function \( f(\cdot) \) and its Jacobian \( J \) are known at point \( w \). The aim of the Levenberg–Marquardt algorithm is to compute the weight vector \( w \) such that \( E(w) \) is minimum. Using the Levenberg–Marquardt algorithm, a new weight vector \( w_{k+1} \) can be obtained from the previous weight vector \( w_k \) as follows

\[
w_{k+1} = w_k + \delta w_k,
\]
5. Experimental results

5.1. Feature extraction using discrete wavelet transform

The ECG signals are considered as representative signals of cardiac physiology, useful in diagnosing cardiac disorders. The most complete way to display this information is to perform spectral analysis. The wavelet transform (WT) provides very general techniques, which can be applied to many tasks in signal processing. One very important application is the ability to compute and manipulate data in compressed parameters which are often called features (Akay, 1997; Daubechies, 1990; Saxena et al., 2002; Sternickel, 2002). Thus, the ECG signal, consisting of many data points, can be compressed into a few parameters. These parameters characterize the behavior of the ECG signal. This feature of using a smaller number of parameters to represent the ECG signal is particularly important for recognition and diagnostic purposes. 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. This multi-scale feature of the WT allows the decomposition of a signal into a number of scales, each scale representing a particular coarseness of the signal under study. The procedure of multi-resolution decomposition of a signal \(x[n]\) is schematically shown in Fig. 2. Each stage of this scheme 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 as shown in Fig. 2.

All wavelet transforms can be specified in terms of a low-pass filter \(h\), which satisfies the standard quadrature mirror filter condition
\[
H(z)H(z^{-1}) + H(-z)H(-z^{-1}) = 1,
\]

where \(H(z)\) denotes the \(z\)-transform of the filter \(h\). Its complementary high-pass filter can be defined as
\[
G(z) = zH(-z^{-1}).
\]

A sequence of filters with increasing length (indexed by \(i\)) can be obtained
\[
H_{i+1}(z) = H(z^{2i})H_i(z);
\]
\[
G_{i+1}(z) = G(z^{2i})H_i(z), \quad i = 0, \ldots, I - 1
\]

with the initial condition \(H_0(z) = 1\). It is expressed as a two-scale relation in time domain
\[
h_{i+1}(k) = [h]_{2^i} h_i(k); \quad g_{i+1}(k) = [g]_{2^i} g_i(k),
\]

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 \(\psi_{i,l}(k)\), \(\psi_{i,l}(k)\) can be defined as
\[
\psi_{i,l}(k) = 2^{i/2} h_i(k - 2^l l); \quad \psi_{i,l}(k) = 2^{i/2} g_i(k - 2^l l),
\]

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
\[ a_{i0}(l) = x(k) \varphi_{j0}(k); \quad d_{i0}(l) = x(k) \psi_{j0}(k), \]
where \(a_{i0}(l)\) and \(d_{i0}(l)\) are the approximation coefficients and the detail coefficients at resolution \(i\), respectively (Akay, 1997; Daubechies, 1990).

The ECG signals can be considered as a superposition of different structures occurring on different time scales at different times. One purpose of wavelet analysis is to separate and sort these underlying structures of different time scales. It is known that the WT is better suited to analyzing non-stationary signals, since it is well localized in time and frequency. The property of time and frequency localization is known as compact support and is one of the most attractive features of the WT. The main advantage of the WT is that it has a varying window size, being broad at low frequencies and narrow at high frequencies, thus leading to an optimal time–frequency resolution in all frequency ranges. Therefore, spectral analysis of the ECG signals was performed using the DWT as described above.

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. In the present study, the number of decomposition levels was chosen to be 4. Thus, the ECG signals were decomposed into the details \(D_1-D_4\) and one final approximation, \(A_4\). Usually, tests are performed with different types of wavelets and the one which gives maximum efficiency is selected for the particular application. The smoothing feature of the Daubechies wavelet of order 2 made it more suitable to detect changes of the ECG signals. Therefore, the wavelet coefficients were computed using the Daubechies wavelet of order 2 in the present study. The wavelet coefficients were computed using the MATLAB software package.

Feature selection is an important component of designing the neural network based on pattern classification since even the best classifier will perform poorly if the features used as inputs are not selected well. The computed discrete wavelet coefficients provide a compact representation that shows the energy distribution of the signal in time and frequency. Therefore, the computed detail and approximation wavelet coefficients of the ECG signals were used as the feature vectors representing the signals. 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 ECG signals:

1. Mean of the absolute values of the coefficients in each subband.
2. Maximum of the absolute values of the coefficients in each subband.
3. Average power of the wavelet coefficients in each subband.
4. Standard deviation of the coefficients in each subband.
5. Ratio of the absolute mean values of adjacent subbands.
6. Distribution distortion of the coefficients in each subband.

Features 1–3 represent the frequency distribution of the signal and the features 4–6 the amount of changes in frequency distribution. In some applications, in order to further reduce the dimensionality of the extracted feature vectors, only some of the statistical features given in this section can be used to represent the time–frequency distribution of the signal under study. In our application, only the maximum absolute values of the coefficients in each subband was used to represent the ECG signals.

5.2. Application of the SNR screening method to ECG signals

The waveforms of four different ECG beats classified in the present study are shown in Fig. 3(a)–(d). A rectangular window, which was formed by 256 discrete data, was selected so that it contained a single ECG beat. For the four diagnostic classes (normal beat, congestive heart failure beat, ventricular tachyarrhythmia beat, atrial fibrillation beat) training and test sets were formed by 720 vectors (180 vectors from each class). The adequate functioning of neural networks depends on the sizes of the training set and test set. The 360 vectors (90 vectors from each class) were used for training and the 360 vectors (90 vectors from each class) were used for testing. A practical way to find a point of better generalization is to use a small percentage (around 20%) of the training set for cross validation. For obtaining a better network generalization, 72 vectors (18 vectors from each class) of training set, which were selected randomly, were used as cross validation set. Beside this, in order to enhance the generalization capability of the MLPNNs, the training and the test sets were formed by data obtained from different patients. For other beat types in the Physiobank database, waveform variations were observed among the vectors belonging to the same class. Therefore, we chose the mentioned four types of ECG beats for classification.

The computed discrete wavelet coefficients were used as the inputs of the MLPNNs. In order to extract features, the wavelet coefficients corresponding to the \(D_1-D_4\) and \(A_4\) frequency bands of the ECG signals were computed. It was observed that the values of the coefficients are very close to zero in \(A_4\) frequency band. So the coefficients corresponding to the frequency band, \(A_4\) were discarded, thus reducing the number of feature vectors representing the signal. For each ECG beat, the detail wavelet coefficients \(d^k, k=1,2,3,4\) at the first, second, third and fourth levels \((129 + 66 + 34 + 18)\) coefficients were computed. Then 247 detail wavelet coefficients were obtained for each ECG beat. The detail wavelet coefficients at the first decomposition level of the four types of ECG beats are given in Fig. 4(a)–(d),
Fig. 3. Waveforms of the ECG beats (a) normal beat; (b) congestive heart failure beat; (c) ventricular tachyarrhythmia beat; (d) atrial fibrillation beat.

Fig. 4. The detail wavelet coefficients at the first decomposition level of the ECG beats: (a) normal beat; (b) congestive heart failure beat; (c) ventricular tachyarrhythmia beat; (d) atrial fibrillation beat.
respectively. It can be noted that the detail wavelet coefficients of the four types of ECG beats are different from each other. In order to reduce the dimensionality of the extracted feature vectors, the maximum absolute values of the coefficients in each subband (variables 1–4) were computed. Then non-salient features called noise 1, noise 2, noise 3, and noise 4 were generated for applying the SNR screening method which is used to reduce the dimension of feature vectors as explained in Section 3. Thus, the MLPNNs had eight inputs, equal to the number of input feature vectors.

Fig. 5 presents the rank distributions derived from the SNR saliency measure for each of the eight features of the noise corrupted ECG signals classification problem over 400 epochs. A rank distribution for a given feature represents the frequency at which that feature was ranked in comparison to the other features using the SNR saliency measure. Higher rankings correspond to higher SNR saliency measure values. A ranking of up to 10 is possible because the eight features were compared in addition to the Uniform (0,1) inserted noise feature and the bias. A ranking of 1 corresponds to the least salient feature and a ranking of 10 corresponds to the highest salient feature.

The MLPNNs were trained for the SNR screening method. The total classification accuracy which is defined as the percentage of beats correctly classified to the total number of beats considered for classification depends on the features used as inputs of the MLPNNs. The total classification accuracies of the MLPNNs used for classification of the ECG beats are presented in Table 1. The SNR screening method selected only two salient features as inputs (variables 1 and 3) of the MLPNN to achieve 97.78% accuracy, which is higher than that of the MLPNN with
eight features (salient and non-salient input features). The success of a method usually depends on the type of data used for its training and testing. The detection rates on the MIT/BIH database obtained by the existing studies in the literature have been varied from 80.0 to 99.8% (Dokur and Ölmaz, 2001; Foo et al., 2002; Li et al., 1995; Maglaveras et al., 1998; Saxena et al., 2002; Simon & Eswaran, 1997). Different methods claim satisfactory results but methods based on ANN require exhaustive training, settings and estimation of model parameters (features used as inputs) and, hence, are computationally complex and time consuming. To overcome the limitations of existing methods, new methods are being developed. One such powerful technique is the WT, which has been used by Li et al. (1995) and gave a detection rate of QRS complexes of 99.8% on the MIT/BIH database. Better results are possible due to a specific feature of the WT, which characterizes the local regularity of ECG signal. This feature is used to distinguish ECG waves from noise, artefacts and baseline drift. In principle, the WT cuts up data into different components that are well localized in time and frequency. In our study, DWT was used for feature extraction from the ECG signals obtained from the Physiobank database. The classification accuracy of our approach confirmed that the proposed MLPNN with two salient input features has potential in classifying four different categories of ECG beats.

### Table 1

<table>
<thead>
<tr>
<th>MLPNN inputs</th>
<th>Total classification accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All features (8 features)</td>
<td>95.83</td>
</tr>
<tr>
<td>Selected features (2 salient features; variables 1 and 3)</td>
<td>97.78</td>
</tr>
</tbody>
</table>

accuracy which is higher than that of the MLPNN with eight features (salient and non-salient input features). The results support our confidence in the SNR saliency measure and the screening method as a useful tool in selecting a parsimonious set of features for ANNs used in automated diagnostic systems developed for ECG beats.

## 6. Conclusion

This paper presented the SNR saliency measure for identifying the saliency of input features of the MLPNNs used in the ECG beats classification. A significant advantage of the SNR saliency measure is that the saliency of each feature is compared to that of a known non-salient noise feature. Feature extraction from the ECG signals for classification of ECG beats was performed using DWT. In order to reduce the dimensionality of the extracted input feature vectors, statistics over the set of the wavelet coefficients were used. The MLPNNs used in classification of the ECG beats were trained for the SNR screening method. The SNR screening method utilizes the SNR saliency measure to select a parsimonious set of salient features. The SNR screening method selected only two salient features as inputs (variables 1 and 3) of the MLPNN used for the ECG beats classification to achieve 97.78% accuracy which is higher than that of the MLPNN with eight features (salient and non-salient input features).

### Acknowledgement

This study has been supported by Scientific Research Project of Gazi University (Project No: 07/2004-04).

### References


