Advances in Automated Diagnostic Systems

ELİF DERYA ÜBEYLİ
Department of Electrical and Electronics Engineering
TOBB University of Economics and Technology
06530 Sogutozu, Ankara
TURKEY
edubeyli@etu.edu.tr  http://edubeyli.etu.edu.tr/

Abstract: This paper intends to an integrated view of the advances in automated diagnostic systems. The paper includes illustrative and detailed information about automated diagnostic systems and feature extraction/selection from biomedical signals. In this respect, the paper satisfies the automated diagnostic systems, which includes the spectral analysis techniques, feature extraction and/or selection methods and decision support systems. The paper includes spectral analysis techniques used in the feature extraction stage, experiments for implementation of automated diagnostic systems and measuring performance of automated diagnostic systems. The objective of the paper is to be a guide for the readers, who want to develop an automated diagnostic systems for detection of any diseases.

Key-Words: Automated diagnostic systems, Spectral analysis techniques, Feature extraction/selection, Performance measure

1 Introduction
Automated diagnostic systems are important applications of pattern recognition, aiming at assisting doctors in making diagnostic decisions. Automated diagnostic systems have been applied to and are of interest for a variety of medical data, such as electrocardiograms (ECGs), electroencephalograms (EEGs), ultrasound signals/images, X-rays, and computed tomographic images [1-12]. Conventional methods of monitoring and diagnosing the diseases rely on detecting the presence of particular signal features by a human observer. Due to large number of patients in intensive care units and the need for continuous observation of such conditions, several techniques for automated diagnostic systems have been developed in the past ten years to attempt to solve this problem. Such techniques work by transforming the mostly qualitative diagnostic criteria into a more objective quantitative signal feature classification problem [1-12]. Figure 1 shows the various stages followed for the design of a classification system. As it is apparent from the feedback arrows, these stages are not independent. On the contrary, they are interrelated and, depending on the results, one may go back to redesign earlier stages in order to improve the overall performance.

Spectral analysis is performed during feature extraction from the time-varying biomedical signals. Spectral analysis considers the problem of determining the spectral content (distribution of power over frequency) of a time series from a finite set of measurements, by means of various spectral analysis techniques. Spectral analysis finds applications in many diverse fields. In different fields, the spectral analysis may reveal “hidden periodicities” in the studied data, which are to be associated with cyclic behavior or recurring processes [3,4]. Spectral analysis techniques have traditionally been based on Fourier transform and filtering theory. Spectral analysis techniques can be found at the heart of many biomedical signal processing systems designed to extract information. In medicine, spectral analysis of various signals recorded from a subject, such as ECGs, EEGs, ultrasound signals, can provide useful information for diagnosis [1-12]. In this paper, the characteristics of each spectral estimate have been presented with the related references for further reading [1-12]. It is hoped that this paper will serve as a guide in helping the reader to make intelligent choices for analysis of biomedical signals recorded from of healthy subjects (control group) and subjects suffering from diseases.

Figure 1. The basic stages involved in the design of a classification system
Medical diagnostic decision support systems have become an established component of medical technology. The main concept of the medical technology is an inductive engine that learns the decision characteristics of the diseases and can then be used to diagnose future patients with uncertain disease states. A number of quantitative models including multilayer perceptron neural networks (MLPNNs), combined neural networks (CNNs), mixture of experts (MEs), modified mixture of experts (MMEs), probabilistic neural networks (PNNs), recurrent neural networks (RNNs), and support vector machines (SVMs) are being used in medical diagnostic support systems to assist human decision-makers in disease diagnosis [1-12]. 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. Unfortunately, there is no theory available to guide an intelligent choice of model based on the complexity of the diagnostic task. In most situations, developers are simply picking a single model that yields satisfactory results, or they are benchmarking a small subset of models with cross validation estimates on test sets [1-12].

ANNs are computational architectures composed of interconnected units (neurons). Its name reflects its initial inspiration from biological neural systems, though the functioning of today’s ANNs may be quite different from that of the biological ones. Sometimes the term neural network also refers to the corresponding mathematical model, but properly speaking a network is an architecture. It is difficult to give a clear definition of ANNs, due to their variety. However, at least the following two particularities distinguish them from other computational architectures or mathematical models.

Neural networks are naturally massively parallel: This is the structural similarity of ANNs to biological ones. Though in some cases neural network models are implemented in software on ordinary digital computers, they are naturally suitable for parallel implementations.

Neural networks are adaptive: A neural network is composed of “living” units or neurons. It can learn or memorize information from data. Learning is the most fascinating feature of neural networks [1-15].

ANNs are computational modeling tools that have recently emerged and found extensive acceptance in many disciplines for modeling complex real-world problems. ANN-based models are empirical in nature, however they can provide practically accurate solutions for precisely or imprecisely formulated problems and for phenomena that are only understood through experimental data and field observations. ANNs produce complicated nonlinear models relating the inputs (the independent variables of a system) to the outputs (the dependent predictive variables). ANNs have been widely used for various tasks, such as pattern classification, time series prediction, nonlinear control, function approximation, and telecommunications. ANNs are desirable because (i) nonlinearity allows better fit to the data, (ii) noise-insensitivity provides accurate prediction in the presence of uncertain data and measurement errors, (iii) high parallelism implies fast processing and hardware failure-tolerance, (iv) learning and adaptivity allow the system to modify its internal structure in response to changing environment, and (v) generalization enables application of the model to unlearned data. Neural networks can be trained to recognize patterns. Also the nonlinear models developed during training allow neural networks to generalize their conclusions and to make application to patterns not previously encountered [1-15].

Recent developments showed that the trend is to develop new methods for computer decision-making in medicine and to evaluate critically these methods in clinical practice. Diagnosis of diseases may be considered as a pattern classification task. If the inputs are ambiguous and possess variability, the conventional pattern classification system may not work. Two patients may not have similar signs and symptoms resulting in the same disease. The diseases of the patients cannot be classified into a single class unless some more measurements and tests are made to resolve the ambiguity. ANN is capable of classifying patterns under variability and ambiguity [1-15].

2 Spectral Analysis Techniques

The basic problem that we consider in this paper is the estimation of the power spectral density (PSD) of a signal from the observation of the signal over a finite time interval. The basic problem that we consider in this paper is the estimation of the PSD of a signal from the observation of the signal over a finite time interval. The biomedical signals are conventionally interpreted by analyzing their spectral content. Diagnosis and disease monitoring are assessed by analysis of spectral shape and parameters [3,4]. In order to determine the degree of diseases, biomedical signals are processed by spectral analysis methods to achieve PSD estimates.

In order to obtain PSD estimates which represent the changes in frequency with respect to time, the classical methods (nonparametric or fast Fourier transform-based methods), model-based methods (autoregressive, moving average, and autoregressive moving average methods), time-frequency methods (short-time Fourier transform, Wigner-Ville distribution, wavelet transform), eigenvector methods (Pisarenko, multiple signal classification, Minimum-Norm) can be used.
3 Feature Extraction/Selection

Feature is a distinctive (sets it appart) or characteristic (its make-up) measurement, transform, structural component made on a segment of a pattern. 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. This, in effect, means that the set of all features that could be used to describe a given pattern (large and in fact infinite infinitesimal changes in some parameter are allowed to separate different features) is limited to those actually stated in the feature vector. One purpose of the dimensionality reduction is to meet engineering constraints in software and hardware complexity, the computing cost, and the desirability of compressing pattern information. In addition, classification is often more accurate when the pattern is simplified through representation by important features or properties only [1-12] (Figure 2).

Figure 2. Functional modules in a typical automated diagnostic system

Feature extraction is the determination of a feature or a feature vector from a pattern vector. For pattern processing problems to be tractable requires the conversion of patterns to features, which are condensed representations of patterns, ideally containing only salient information. Feature extraction methods are subdivided into: 1) statistical characteristics and 2) syntactic descriptions. Spectral analysis techniques can be used for extraction of features characterizing the signals under study [1-12].

Feature selection provides a means for choosing the features which are best for classification, based on various criteria. The feature selection process performed on a set of predetermined features. Features are selected based on either 1) best representation of a given class of signals, or 2) best distinction between classes. Therefore, feature selection plays an important role in classifying systems such as neural networks. For the purpose of classification problems, the classifying system has usually been implemented with rules using if-then clauses, which state the conditions of certain attributes and resulting rules. However, it has proven to be a difficult and time consuming method. From the viewpoint of managing large quantities of data, it would still be most useful if irrelevant or redundant attributes could be segregated from relevant and important ones, although the exact governing rules may not be known. In this case, the process of extracting useful information from a large data set can be greatly facilitated [1-12].

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 features can be used. The maximum, mean, minimum, standard deviation of the features can be computed to represent the signals.

There are numerous methods to represent patterns as a grouping of features. The choice of methods appropriate for a given pattern analysis task is rarely obvious. At each level (feature extraction, feature selection, classification) many methods exist. Since the architecture of the decision support system can be compatible with different types of features, it is necessary to know how to fuse different types of features. Fusion of features for some types of decision support systems can increase the accuracy of the system. In this respect, feature extraction/selection is important in dealing with the accuracy of the developed decision support system [1-12]. In the following, a brief explanation about diverse and composite features is presented.

In the feature extraction stage, numerous different methods can be used so that several diverse features can be extracted from the same raw data. To a large extent, each feature can independently represent the original data, but none of them is totally perfect for practical applications. Moreover, there seems to be no simple way to measure relevance of the features for a pattern classification task. For this kind of pattern classification tasks, diverse features often need to be jointly used in order to achieve robust performance. This kind of pattern classification tasks are called as classification with diverse features. In order to perform a classification, two different methods are used. One is the use of a composite
feature formed by lumping diverse features together and the other is combination of multiple classifiers that have been already trained on diverse feature sets. Several problems given as follows occur with the usage of composite feature:

1) Its dimension is higher than that of any component feature and it is well known that high-dimension vectors will not only increase computational complexity but will also produce implementation problems and accuracy problems.

2) It is difficult to lump several features together due to their diversified forms, e.g., they may be continuous variables, binary values, discrete labels, structural primitives.

3) Those component features are usually not independent.

In general, therefore, the use of a composite feature does not provide a significantly improved performance. However, the combination of multiple classifiers is a good solution for the problem involving a variety of features [1-12].

4 Experiments for Implementation of Automated Diagnostic Systems

The key design decisions for the neural networks used in classification are the architecture and the training process. The architectures of the MLPNN, CNN, ME, MME, PNN, RNN, SVM used for classification of the signals are shown in the reference [3]. The adequate functioning of neural networks depends on the sizes of the training set and test set. To comparatively evaluate the performance of the classifiers, all the classifiers can be trained by the same training data set and tested with the evaluation data set. The explanations about the training algorithms of the classifiers are presented in reference [3] with the related references for further reading.

Different experiments are performed during implementation of these classifiers and the number of support vectors in the SVMs, pattern layer neurons in the PNNs, expert networks in the MEs and MMEs, recurrent neurons in the RNNs, hidden layers and hidden neurons in the MLPNNs are determined by taking into consideration the classification accuracies. In the hidden layers and the output layers, sigmoid, tan-sigmoid, linear functions can be used as the activation function. The sigmoidal function with the range between zero and one introduces two important properties. First, the sigmoid is nonlinear, allowing the network to perform complex mappings of input to output vector spaces, and secondly it is continuous and differentiable, which allows the gradient of the error to be used in updating the weights.

Table 1 defines examples of the network parameters of the classifiers [3].

<table>
<thead>
<tr>
<th>Classifier (features)</th>
<th>Neural network parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM (composite feature)</td>
<td>41:9:3(^a)</td>
</tr>
<tr>
<td>RNN (composite feature)</td>
<td>34:30r:25r:4(^b) \n 600(^c)</td>
</tr>
<tr>
<td>PNN (composite feature)</td>
<td>41:21:3:1(^d)</td>
</tr>
<tr>
<td>MME (diverse features)</td>
<td>5:25:3(^e), 4:25:3(^e), 28:25:3(^e), 4:25:3(^e), \n 5:25:3(^e), 4:25:3(^e), 28:25:3(^e), 4:25:3(^e), \n 500(^e)</td>
</tr>
<tr>
<td>ME (composite feature)</td>
<td>41:25:3(^f), 41:25:3(^f), \n 700(^f)</td>
</tr>
<tr>
<td>CNN (composite feature)</td>
<td>41:25:9(^g), 9:30:3(^h), \n 1200(^g)</td>
</tr>
<tr>
<td>MLPNN (composite feature)</td>
<td>41:25:3(^i), \n 1900(^i)</td>
</tr>
</tbody>
</table>

\(^a\)Design of SVMs: Number of input neurons \cdot support vectors \cdot output neurons, respectively.
\(^b\)Design of RNNs: Number of input neurons \cdot recurrent neurons in the first hidden layer \cdot recurrent neurons in the second hidden layer \cdot output neurons, respectively.
\(^c\)Number of training epochs.
\(^d\)Design of PNNs: Number of input neurons \cdot pattern layer neurons \cdot summation layer neurons \cdot output layer neurons, respectively.
\(^e\)Design of expert networks: Number of input \cdot hidden \cdot output neurons, respectively.
\(^f\)Design of gating networks in gate-bank: Number of input \cdot hidden \cdot output neurons, respectively.
\(^g\)Design of gating network: Number of input \cdot hidden \cdot output neurons, respectively.
\(^h\)Design of first level network: Number of input \cdot hidden \cdot output neurons, respectively.
\(^i\)Design of second level network: Number of input \cdot hidden \cdot output neurons, respectively.
5 Measuring Performance of Automated Diagnostic Systems

Given a random set of initial weights, the outputs of the network will be very different from the desired classifications. As the network is trained, the weights of the system are continually adjusted to reduce the difference between the output of the system and the desired response. The difference is referred to as the error and can be measured in different ways. The most common measurement is the mean square error (MSE). The MSE is the average of the squares of the difference between each output and the desired output. In addition to MSE, normalized mean squared error (NMSE), mean absolute error (MAE), minimum absolute error and maximum absolute error can be used for the measuring error of the neural network [3,13-15].

The training holds the key to an accurate solution, so the criterion to stop training must be very well described. In general, it is known that a network with enough weights will always learn the training set better as the number of iterations is increased. However, neural network researchers have found that this decrease in the training set error was not always coupled to better performance in the test. When the network is trained too much, the network memorizes the training patterns and does not generalize well. The aim of the stop criterion is to maximize the network’s generalization [3,13-15].

The size of MSE can be used to determine how well the network output fits the desired output, but it may not reflect whether the two sets of data move in the same direction. The correlation coefficient (r) solves this problem. The correlation coefficient is limited with the range [-1,1]. When $r = 1$ there is a perfect positive linear correlation between network output and desired output, which means that they vary by the same amount. When $r = -1$ there is a perfectly linear negative correlation between network output and desired output, that means they vary in opposite ways (when network output increases, desired output decreases by the same amount). When $r = 0$ there is no correlation between network output and desired output (the variables are called uncorrelated). Intermediate values describe partial correlations [3,13-15].

Neural networks are used for both classification and regression. In classification, the aim is to assign the input patterns to one of several classes, usually represented by outputs restricted to lie in the range from 0 to 1, so that they represent the probability of class membership. While the classification is carried out, a specific pattern is assigned to a specific class according to the characteristic features selected for it. In regression, desired output and actual network output results can be shown on the same graph and performance of network can be evaluated in this way. Classification results of the classifiers are 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 [3-15]. From the confusion matrices one can tell the frequency with which a signal is misclassified as another. Table 2 is presenting the confusion matrices for the classification of the coronary arterial signals [3].

<table>
<thead>
<tr>
<th>Classifiers (features)</th>
<th>Desired Result</th>
<th>Output Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM (composite feature)</td>
<td>Healthy</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Coronary artery stenosis</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Healthy</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Coronary artery stenosis</td>
<td>32</td>
</tr>
<tr>
<td>RNN (composite feature)</td>
<td>Healthy</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>Coronary artery stenosis</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Healthy</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Coronary artery stenosis</td>
<td>32</td>
</tr>
<tr>
<td>PNN (composite feature)</td>
<td>Healthy</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>Coronary artery stenosis</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Healthy</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Coronary artery stenosis</td>
<td>31</td>
</tr>
<tr>
<td>MME (diverse features)</td>
<td>Healthy</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Coronary artery stenosis</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Healthy</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Coronary artery stenosis</td>
<td>32</td>
</tr>
<tr>
<td>ME (composite feature)</td>
<td>Healthy</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Coronary artery stenosis</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Healthy</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Coronary artery stenosis</td>
<td>31</td>
</tr>
<tr>
<td>CNN (composite feature)</td>
<td>Healthy</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>Coronary artery stenosis</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Healthy</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Coronary artery stenosis</td>
<td>31</td>
</tr>
<tr>
<td>MLPNN (composite feature)</td>
<td>Healthy</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Coronary artery stenosis</td>
<td>3</td>
</tr>
</tbody>
</table>

The test performance of the classifiers 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. Receiver operating characteristic (ROC) plots provide a view of the whole spectrum of sensitivities and specificities because all possible sensitivity/specificity pairs for a particular test are graphed. The performance of a test can be evaluated by plotting a ROC curve for the test and therefore, ROC curves are used to describe the performance of the classifiers [3-12]. A good test is one for which sensitivity rises rapidly and 1-specificity hardly increases at all until sensitivity becomes high (Figure 3).

![Figure 3. ROC curve](image)

### 6 Conclusion

The automated diagnostic systems trained on diverse or composite features for classification of the signals are presented. The signals classification is considered as a typical problem of classification with diverse features since the methods used for feature extraction have different performance and no unique robust feature has been found. The inputs (diverse or composite features) of the automated diagnostic systems are obtained by preprocessing of the signals with various spectral analysis methods. The performance measuring can be performed in order to demonstrate the accuracy of the implemented automated diagnostic systems.

### References:


