Expert Systems with Applications 37 (2010) 5110–5115

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Evaluation of ensemble methods for diagnosing of valvular heart disease

Resul Das [a], Abdulkadir Sengur [b,*]

[a] Department of Informatics, Firat University, 23119 Elazig, Turkey
[b] Department of Electronics and Computer Science, Firat University, 23119 Elazig, Turkey

ARTICLE INFO

Keywords:
Doppler heart sounds
Heart valves
Ensemble methods
Bagging
Boosting
Random subspaces

ABSTRACT

In this work, we investigate the use of ensemble learning for improving classifiers which is one of the important directions in the current research of machine learning, in which bagging, boosting and random subspace are three powerful and popular representatives. Researchers have so far shown the efficacies of ensemble methods in many practical classification problems. However, for valvular heart disease detection, there are almost no studies investigating their feasibilities. Thus, in this study, we evaluate the performance of three popular ensemble methods for the diagnosis of the valvular heart disorders. To evaluate the performance of investigated ensemble methodology, a comparative study is realized by using a data set containing 215 samples. Moreover, to achieve a comprehensive comparison, we consider the previous results reported by earlier methods (Çomak, Arslan, & Turkoglu, 2007; Sengur, 2008a,b; Sengur & Turkoglu, 2008; Turkoglu, Arslan, & Ilkay, 2002, 2003; Uguz, Arslan, & Turkoglu, 2007). Experimental results suggest the feasibilities of ensemble classification methods, and we also derive some valuable conclusions on the performance of ensemble methods for valvular heart disease detection.

1. Introduction

Valvular heart disease is the name given to any dysfunction or abnormality of one or more of the heart's four valves, including the mitral valve and aortic valve on the left side, and the tricuspid valve and pulmonic valve on the right side (http://yourtotalhealth.ivillage.com/heart-disease-fast-facts.html Accessed 13.02.08). In a normally functioning heart, the four valves keep blood flowing in one direction and only at the right time. They act as gates that swing open to allow the blood to flow through and then tightly shut until the next cycle begins.

According to the American Heart Association's 2006 Heart and Stroke Statistical Update, valvular heart disease is responsible for nearly 20,000 deaths each year in the United States and is a contributing factor in about 42,000 deaths (http://www.americanheart.org Accessed 13.02.08). The majority of these cases involve disorders of the aortic valve (63%) and the mitral valve (14%). Deaths due to pulmonic and tricuspid valve disorders are rarer (0.06% and 0.01%, respectively).

There are a number of types of valvular heart disease, including valvular stenosis, valvular regurgitation, atresia of one of the valves and mitral valve prolapse (http://yourtotalhealth.ivillage.com/heart-disease-fast-facts.html Accessed 13.02.08). The diagnosis of valvular heart disease is usually performed by one of the following tests: Physical examination may reveal a murmur, evidence of heart enlargement and fluids within the lungs. An electrocardiogram (EKG) may reveal arrhythmias and chamber enlargement. Echocardiography and Doppler ultrasound are the most widely used methods and are very useful in the assessment of presence and severity of valve disease. MRI can provide clear three-dimensional images of the heart and its valves.

Doppler technique has gained much more interest since Satomura first demonstrated the application of Doppler Effect in the measurement of blood velocity in 1959 (Keeton & Schlindwein, 1997). Doppler heart sounds (DHS) are one of the most important sounds produced by blood flow, valves motion and vibration of the other cardiovascular components (Wright, Gough, Rakebrandt, Wahab, & Woodcock, 1997). However, the factors such as calcified disease or obesity often result in a diagnostically unsatisfactory Doppler techniques assessment and, therefore, it is sometimes necessary to assess the spectrogram of the Doppler shift signals to elucidate the degree of the disease (Wright et al., 1997). In addition to Doppler techniques, the techniques that are more complex have also been developed (Laplace Transform and Principal Components Analysis). Many studies have been implemented to classify Doppler signals in the pattern recognition field (Chan, Chan, Lam, Lui, & Poon, 1997; Guler, Kiymik, Kara, & Yuksel, 1992).

Recent advances in the field of artificial intelligence have led to the emergence of expert systems for medical applications. Moreover, in the last few decades, computational tools have been

* Corresponding author. Address: Department of Electronics and Computer Science, Firat University, Technical Education Faculty, 23119 Elazig, Turkey.
E-mail addresses: rdas@firat.edu.tr (R. Das), ksengur@firat.edu.tr (A. Sengur).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.12.085

designed to improve the experiences and abilities of physicians for making decisions about their patients. Motivated by the need of such an expert system, in this study, we investigate the use of ensemble learning for improving the performance of classifiers, therefore bagging, boosting and random subspace, which are three powerful and popular representatives, are used in this direction. The performance of the proposed methodology was evaluated on data set containing 215 samples with several statistical validation methods. Moreover, we compared our results with the previous methods where the same data set have been used. Our system obtained 97.3% sensitivity and 100% specificity rate. This classification accuracy is the highest so far.

2. Background

2.1. A brief overview on related works

So far, various classification algorithms have been employed on Turkoglu's valvular heart disease data set, and high classification accuracies have been reported in the last decade (Çomok, Arslan, & Turkoglu, 2007; Sengur, 2008a, 2008b; Sengur & Turkoglu, 2008; Turkoglu, Arslan, & Ilkay, 2002, 2003; Uguz, Arslan, & Turkoglu, 2007). Turkoglu's valvular heart disease data set was obtained from Firat Medical Center. A detailed description for the data set will be given in the next section.

The valvular heart disease data set was firstly utilized in Turkoglu et al. (2002) where they proposed an expert diagnosis system which uses back-propagation artificial neural networks (BP-ANN) classifier. The performance evaluation of the proposed system was evaluated by classification accuracy, and the correct classification rate was about 94% for normal subjects and 95.9% for abnormal subjects. Later, Turkoglu et al. (2003) proposed an intelligent system for the detection of heart valve disease based on wavelet packet neural networks (WPNN) Turkoglu et al., 2003. The reported correct classification rate was about 94% for abnormal and normal subjects. Recently, Comak et al. (2006) investigated the use of least-square support vector machines (LS-SVM) classifier for improving the performance of the Turkoglu's (2002) proposal Çomak et al., 2007. Moreover, they intended to realize a comparative study. Classification rates of the examined classifiers were evaluated by ROC curves based on the terms of sensitivity and specificity. The application results showed that according to the ROC curves, the LS-SVM classifier performance was almost the same with BP-ANN. It was reported that LS-SVM was more suitable than BP-ANN since it has some advantages against BP-ANN. Another reported advantage of LS-SVM was its shorter training time. It was also reported that ANN's training time was about 13 times longer than LS-SVM's training time according to the experimental results. Furthermore, this property was the only advantage of LS-SVM against BP-ANN. More recently, Uguz et al. (2006) proposed a biomedical system based on Hidden Markov Model for clinical diagnosis and recognition of heart valve disorders (Uguz et al., 2007). The proposed methodology also used the database of Turkoglu et al. (2002). In the present study, continuous HMM (CHMM) classifier system was used. Single Gaussian model was preferred to determine emission probability. The proposed methodology was composed of two stages. At the first stage, the initial values of average and standard deviation were calculated by separating observation symbols into equal segments according to the state number, and using observation symbols appropriate to each segment. At the second stage, the initial values of average and standard deviation were calculated by separating observation symbols into the clusters (FCM or K means algorithms) that have equal number of states and using observation symbols appropriate to the separated clusters. The implementations of the experimental studies were carried out on three different classification systems such as CHMM, FCM–K-means/CHMM and ANN. These experimental results were obtained for specificity and sensitivity rates of 92% and 94% for CHMM, 92% and 97.26% for FCM–K-means/CHMM), respectively. Finally, Sengur (2008b) investigated the use of principal component analysis (PCA), artificial immune system (AIS) and fuzzy *k*-NN to determine the normal and abnormal heart valves from the Doppler heart sounds (Sengur, 2008a). For reducing the complexity, PCA was used. In the classification stage, AIS and fuzzy *k*-NN were used. To evaluate the performance of the proposed methodology, a comparative study was realized using a data set containing 215 samples. The validation of the proposed method was done by using the sensitivity and specificity parameters; 95.9% sensitivity and 96% specificity rate was obtained. Sengur (2008b) also investigated the use of Linear Discriminant Analysis (LDA) and Adaptive neuro-fuzzy inference system (ANFIS) for clinical diagnosis and recognition of heart valve disorders (Sengur, 2008b). The validation of the proposed method is done using the sensitivity and specificity parameters; 95.9% sensitivity and 94% specificity rate was obtained.

2.2. Feature extraction and construction of the database

All the original audio Doppler Heart Sound (DHS) signals were acquired from the Acuson Sequoia 512 Model Doppler Ultrasound system in the Cardiology Department of the Firat Medical Center. DHS signals were sampled at 20 kHz for 5 s, and signal to noise ratio of 0 dB using a sound card which has 16-bit A/D conversion resolution, and computer software prepared by us in the MATLAB. The Doppler ultrasonic flow transducer used (Model 3V2c) was run in a continuous operating mode 2 MHz. The Doppler signals of the heart valves were obtained by placing the transducer over the chest of the patient with the aid of ultrasonic image. Pre-processing to obtain the feature vector was performed on the digitized signal in the following order:

i. Filtering: The stored DHS signals were high-pass filtered to remove unwanted low-frequency components, because the DHS signals is generally in the range of 0.5–10 kHz. The filter is a digital FIR, which is a fiftieth-order filter with a cut-off frequency equal to 500 Hz, and window type is the 51-point symmetric Hamming window.
ii. White de-noising: White noise is a random signal that contains equal amounts of every possible frequency, i.e., its FFT has a flat spectrum (Şengür & ve Türkoğlu, 2008). The DHS signals were filtered from removing the white noise by using wavelet packet. The white de-noising procedure contains three steps (Daş, Türkoğlu, & Şengür, 2009):
   1. Decomposition: Computing the wavelet packet decomposition of the DHS signal at level 4, and using the Daubechies wavelet of order 4.
   2. Detail coefficient thresholding: For each level from 1 to 4, soft thresholding is applied to the detail coefficients.
   3. Reconstruction: Computing wavelet packet reconstruction based on the original approximation coefficients of level 4, and the modified detail coefficients of levels from 1 to 4.
iii. Normalization: The DHS signals in this study were normalized using Eq. (1) so that the expected amplitude of the signal is not affected from the rib cage structure of the patient.

$$\mathrm{DHS_{signal}} = \frac{\mathrm{DHS_{signal}}}{|(\mathrm{DHS_{signal}})_{\max}|} \tag{1}$$

The DHS waveform patterns from heart valves are rich in detail and highly non-stationary. After the data pre-processing has been realized, feature extraction process was carried out. The feature extraction process has two stages (Sengur & Turkoglu, 2008):

Stage 1 – wavelet packet decomposition: For wavelet packet decomposition of the DHS waveforms, the tree structure used a binary tree at depth $m = 8$. Wavelet packet decomposition was applied to the DHS signal using the Daubechies-1 wavelet packet filters Ψ with the Shannon entropy (Coifman & Wickerhauser, 1992) as defined in Eq. (2). In the equation, s is the DHS signal, and (s_i) is the coefficients of wavelet packet decomposition of s, thus obtaining $2^8 = 256$ terminal node signals.

$$E(s) = -\sum_i s_i^2 . \log(s_i^2) \tag{2}$$

Stage 2 – entropy: An Entropy-based criterion describes information-related properties for an accurate representation of a given signal. Entropy is a common concept in many fields, mainly in signal processing (Quiroga, Roso, & Basar, 1999). A method for measuring the entropy appears as an ideal tool for quantifying the ordering of non-stationary signals. The norm entropy was calculated as defined in Eq. (3) of the waveforms at the terminal node signals obtained from wavelet packet decomposition.

$$E(s) = \frac{\sum_i |s_i|^p}{N} \tag{3}$$

where the entropy E is a real number, s is the terminal node signal, and (s_i) is the waveform of terminal node signals. In norm entropy, P is the power and must be such that $1 \leqslant P < 2$. The resultant entropy data were normalized with $N = 20$. Thus, the feature vector was extracted by computing the 256-entropy values per DHS signal. For more information, please refer to Turkoglu et al. (2002).

3. Methods

To evaluate the ensemble methods for diagnosing the valvular heart disease, three widely used classifiers k-Nearest Neighbors (k-NN), Multi Layer Perceptron (MLP) and Support Vector Machine (SVM) are determined as base classifiers (Duda & Hart, 1973). In this section, after briefly reviewing the three ensemble classification methods bagging, boosting and random subspace, we provide short descriptions for the aforementioned base classifiers.

3.1. Ensemble methods

An ensemble of classifiers is a collection of several classifiers whose individual decisions are combined in some way to classify the test examples (Skurichina & Duin, 2002). It is known that an ensemble often shows much better performance than the individual classifiers that make it up.

3.1.1. Bagging

In bagging, each classifier is adjusted on a randomly drawn training set with the probability of drawing any given example being equal (Breiman, 1996). Samples are drawn with replacement, so that some examples may be selected multiple times while others may not be selected at all. As a result, each classifier could return a higher test set error than a classifier using all of the data. However, when these classifiers are combined, the resulting ensemble produces lower test set error than a single classifier. The diversity among individual classifiers compensates for the increase in error rate of any individual classifier and improves prediction performance.

3.1.2. Boosting (AdaBoost)

The AdaBoost family of algorithms, also known as boosting, is another category of powerful ensemble methods (Freund & Shapire, 1997). It explicitly alters the distribution of training data fed to every individual classifier, specifically weight so each training sample. Initially the weights are uniform for all the training samples. During the boosting procedure, they are adjusted after the training of each classifier is completed. For misclassified samples, the weights are increased, while for correctly classified samples they are decreased. The final ensemble is constructed by combining individual classifiers according to their own accuracies.

3.1.3. Random subspaces

The random subspace ensemble method, proposed by Ho, applies the random selection of feature subspaces to construct individual classifiers (Ho, 1998). This method can take the advantage of high dimensionality and is an effective countermeasure for the traditional problem of the curse of dimensionality. Its merit can be attributed to the high ensemble diversity, which compensates for the possible deficiency of accuracies in individual classifiers. In random subspace, feature subspaces are picked at random from the original feature space, and individual classifiers are created only based on those attributes in the chosen feature subspaces using the original training set. The outputs from different individual classifiers are combined by the uniform majority voting to give the final prediction. For random subspace, how to select the optimum dimensionality of feature subspaces is still an open problem.

3.2. Base classifiers

3.2.1. k-NN

The k-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms (Duda & Hart, 1973). An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors. k is a positive integer, typically small. If $k = 1$, then the object is simply assigned to the class of its nearest neighbor. In binary (two class) classification problems, it is helpful to choose k to be an odd number as this avoids tied votes.

3.2.2. MLP

An MLP is a network of simple neurons called perceptrons (Bishop, 1995). The perceptron computes a single output from multiple real-valued inputs by forming a linear combination according to its input weights and then possibly putting the output through some non-linear activation function. In other words, MLPs are feedforward neural networks trained with the standard backpropagation algorithm (Bishop, 1995). They are supervised networks, so they require a desired response to be trained. They learn how to transform input data into a desired response, so they are widely used for pattern classification. With one or two hidden layers, they can approximate virtually any input–output map. They have been shown to approximate the performance of optimal statistical classifiers in difficult problems.

3.2.3. SVM

Support Vector Machines (SVMs) are learning machines that can perform binary classification and real-valued function approximation tasks (Vapnik, 1998). SVMs non-linearly map their n-dimensional input space into a high-dimensional feature space. Depending on the choice of kernel functions, different classifiers including linear and non-linear classifiers can be constructed.

4. Performance evaluation methods

Different evaluation methods were used for calculating the performance of the proposed system. These methods are classification accuracy, sensitivity and specificity measures and confusion matrix. The description of these methods will be given in the following sub sections.

4.1. Classification accuracy

The classification accuracy is the common method that is used in the pattern recognition applications. The classification accuracy for the experiment is taken as the ratio of the number of samples correctly classified to the total number of samples.

4.2. Sensitivity and specificity

For sensitivity and specificity analysis, we use the following expressions.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}\% \tag{4}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}}\% \tag{5}$$

where TP, TN, FP and FN denote true positives, true negatives, false positives and false negatives, respectively (http://www.bmj.com/cgi/reprint/308/6943/1552 Accessed 01.05.08). True positive (TP): An input is detected as a patient with atherosclerosis diagnosed by the expert clinicians. True negative (TN): An input is detected as normal that is labeled as a healthy person by the expert clinicians. False positive (FP): An input is detected as a patient that is labeled as healthy by the expert clinicians. False negative (FN): An input is detected as normal with atherosclerosis diagnosed by the expert clinicians.

4.3. Confusion matrix

A confusion matrix is composed of the actual and the predicted classifications done by a classification system. Confusion matrix identifies the common misclassifications of the proposed classification schema. Performance of such a system is commonly evaluated using the data in the matrix (http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion_matrix/confusion_matrix.html Accessed 20.04.08).

5. Experimental classification results

The aortic and mitral valves of 132 men and 83 women, ages 15 through 80 (mean age = 48 years), were studied. A total of 215 valvular classifications were identified, including 56 normal and 54 abnormal aortic valves, and 39 normal and 66 abnormal mitral valves. The classification of abnormal heart valves included both stenosis and insufficiency. A normal valve had no such stenosis or insufficiency. Of the 110 aortic valve patients studied, 14 abnormal and 25 normal subjects were selected for the training process. Of the 105 mitral valve patients studied, 33 abnormal and 20 normal subjects were selected for the training process. The remaining patients were used as the test set. The true diagnoses were identified under the supervision of expert physicians using known Doppler and clinical observations (Turkoglu et al., 2002). In the training processing, 14 abnormal and 25 normal subjects were selected for diagnosis of the aortic heart valve, and 33 abnormal and 20 normal subjects were chosen for diagnosis of the mitral heart valve. Remaining of the data set was used as the test data set.

As we mentioned earlier, three types of base classifiers were used: *k*-NN, MLP and SVM. *k*-NN is a type of instance-based learning or lazy learning where the function is only approximated locally, and all computation is deferred until classification. We use 7-nearest neighbor. MLPs have been known to be very accurate and robust to noise in data sets with excellent performance. For this study, a standard three-layered back-propagation network with the tangent-sigmoid transfer function is considered. The weights and biases of the neural networks are initialized randomly, and the number of neurons in the hidden node is determined heuristically as $inputs + outputs$. A small value of the learning rate (0.15) and a large value of the momentum rate (0.8) are chosen to avoid local minima. The number of training epochs was 500. To implement the principles of SVMs, we used the LS-SVM, modifying it for our study. The two most important steps in implementation of SVM is scaling and kernel selection; for scaling, we linearly scaled the values of all features to the range [1, +1] to prevent the cases that features great numeric ranges dominating those in smaller numeric ranges. Among many available kernel functions (linear, polynomial and radial basis function (RBF)), we used linear kernel. The ensemble size is taken as 25, since it has been shown that for many ensemble problems, the biggest profit in accuracy is already made with this number of individual classifiers (Sun, Zhang, & Zhan, 2007).

The experimental results of ensemble classification with different base classifiers for valvular heart disease data set are given in Table 1. Therein, 'single' means an individual base classifier trained using all the intact training data.

From the experimental results that are given in Table 1, we draw the following conclusions about the performance of the ensemble methods. First, all the ensemble methods produced successful results than a single classifier. With the base classifier *k*-NN, only random subspace brings significant performance improvements compared to a single classifier. The bagging and boosting ensemble does almost not change the performance. This should happen due to the stability of *k*-NN with respect to changes in training sets (Skurichina and Duin, 2002). *k*-NN is a stable classifier, while ensemble methods constructed through sub sampling the training examples (e.g., bagging and boosting) do not work well for stable classifiers (Skurichina and Duin, 2002). On the other hand, as it was noted in Sun et al. (2007), *k*-NN benefits random subspace because, random subspace carries out classification in subspaces of much lower dimensionality and this can reduce the negative influence of noises for the calculation of neighbors.

Table 1
Test set performance evaluation rates (%) of ensemble methods with base classifiers *k*-NN, MLP and SVM.

Base classifier and method	Performance evaluation methods		
	Sensitivity (%)	Specificity (%)	Accuracy (%)
(k-NN)			
Single	91.8	90.0	91.1
Bagging	93.1	92.0	92.6
Adaboost	93.1	92.0	92.6
Random subspaces	94.3	94.0	94.2
(MLP)			
Single	95.9	94.0	95.1
Bagging	95.9	94.0	95.1
Adaboost	97.3	94.0	95.9
Random subspaces	95.9	94.0	95.1
(SVM)			
Single	94.5	90.0	92.7
Bagging	95.9	94.0	95.1
Adaboost	97.3	100	98.4
Random subspaces	94.5	90.0	92.7

Table 2
Obtained performance parameters with our proposed system and other classifiers from literature.

Method and classifier	Type	No. of patients	Detected as abnormal	Detected as normal	SN = Sensitivity SP = Specificity
ANN (Turkoglu et al., 2002)	Abnormal	73	70	3	SN = 95.9%
	Normal	50	3	47	SP = 94%
WPNN (Turkoglu et al., 2003)	Abnormal	73	69	4	SN = 94.5%
	Normal	50	3	47	SP = 94%
SVM (Comak, 2006)	Abnormal	73	69	4	SN = 94.5%
	Normal	50	5	45	SP = 90%
FCM/CHMM (Uguz, 2006)	Abnormal	73	71	2	SN = 97.3%
	Normal	50	4	46	SP = 92%
LDA, ANFIS (Sengur, 2008b)	Abnormal	73	70	3	SN = 95.9%
	Normal	50	3	47	SP = 94%
PCA, AIS and Fuzzy *k*-NN (Sengur, 2008a)	Abnormal	73	70	3	SN = 95.9%
	Normal	50	2	48	SP = 96%
The best ensemble method	Abnormal	73	71	2	SN = 97.3%
	Normal	50	0	50	SP = 100%

With the base classifier MLP, only boosting method gets better performance than a single classifier and further bagging and random subspace do not bring any performance improvements.

With base classifier SVM, both bagging and boosting methods improve the performance. Moreover, boosting slightly makes great performance improvements compared to a single classifier. However, random subspace ensemble does not bring any performance improvements. The performance of the bagging, boosting and random subspace ensemble methods when combining linear classifiers also influenced training sample size (Sun et al., 2007). Therefore, we validate one of their conclusions, namely, boosting is the best for large training sample sizes. Since bagging and random subspace are useful for critical training sample sizes (Skurichina and Duin, 2002), their performances are inferior to that of boosting for the considered valvular heart disease diagnosing problem in this paper.

Moreover, a confusion matrix and the calculated sensitivity and specificity rates are given in Table 2. The performance comparison of the best ensemble method with the other classifiers from literature is also demonstrated in the Table 2. According to these results, higher sensitivity rate (97.3%) is obtained using the FCM–CHMM and the best ensemble method. But, while FCM–CHMM produced the 92% specificity value, ensemble method produced the highest specificity rate (100%). SVM method gains the worst specificity rate (90%), while ANN, WPNN and LDA–ANFIS methods obtain the same 94% specificity rate, and ANN, LDA–ANFIS and PCA–AIS and fuzzy *k*-NN produce the same sensitivity rate (95.9%). Finally, the SVM and WPNN produce the lower sensitivity rate (94.5%).

6. Discussion and conclusion

In this study, we investigated the usage feasibility of three popular ensemble methods, bagging, boosting and random subspace in the context of valvular heart disease classification. The efficiency of ensemble methods over a single base classifier is shown. Experimental results also indicate that the capability of ensemble methods is subject to the type of base classifiers. The findings of the present study are helpful in guiding the choice of classification algorithms for future applications.

Since ensemble methods have more individual classifiers than a single classifier does, they would require higher computational burden, especially for classifier training. This is the disadvantage of the ensemble method. In addition, the selection of base classifiers influences the performance of ensemble methods greatly. Depending on the choice of base classifiers, the performance of an ensemble method can outperform a single classifier.

Generally speaking, using ensemble methods for diagnosing valvular heart disease is effective, though different ensemble methods would display different performance, and even some combination of base classifiers and ensemble methods would deteriorate the performance.

References

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford University Press.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123–140.

Chan, B. C. B., Chan, F. H. Y., Lam, F. K., Lui, P. W., & Poon, P. W. F. (1997). Fast detection of venous air embolism is Doppler heart sound using the wavelet transform. *IEEE Transactions on Biomedical Engineering, 44*(4), 237–245.

Coifman, R. R., & Wickerhauser, M. V. (1992). Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory, 38*(2), 713–718.

Comak, E., Arslan, A., & Turkoglu, I. (2007). A decision support system based on support vector machines for diagnosis of the heart valve diseases. *Computers in Biology and Medicine, 37*, 21–27.

Daş, R., Türkoğlu, İ., & Şengür, A. (2009). Diagnosis of valvular heart disease through neural networks ensembles. *Computer Methods and Programs in Biomedicine, 93*(2), 185–191.

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York: Wiley.

Freund, Y., & Shapire, R. E. (1997). A decision-theoretic generalization of on line learning and an application to boosting. *Journal of Computer and System Sciences, 55*(1), 119–139.

Guler, I., Kiymik, M. K., Kara, S., & Yuksel, M. E. (1992). Application of autoregressive analysis to 20 MHz pulsed Doppler data in real time. *International Journal of Bio-Medical Computing, 31*(3–4), 247–256.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(8), 832–844.

Keeton, P. I. J., & Schlindwein, F. S. (1997). Application of wavelets in Doppler ultrasound. *MCB University Press, 17*(1), 38–45.

Quiroga, R. Q., Roso, O. A., & Basar, E. (1999). Wavelet entropy: a measure of order in evoked potentials. *Elsevier Science, Evoked Potentials and Magnetic Fields, 49*(September), 298–302.

Sengur, A. (2008a). An expert system based on principal component analysis, artificial immune system and fuzzy *k*-NN for diagnosis of valvular heart diseases. *Computers in Biology and Medicine, 38*(3), 329–338.

Sengur, A. (2008b). An expert system based on linear discriminant analysis and adaptive neuro-fuzzy inference system to diagnosis heart valve diseases. *Expert Systems with Applications, 35*(1–2), 214–222.

Sengur, A., & Turkoglu, I. (2008). A hybrid method based on artificial immune system and fuzzy *k*-NN algorithm for diagnosis of heart valve diseases. *Expert Systems with Applications, 35*(3), 1011–1020.

Skurichina, M., & Duin, R. P. W. (2002). Bagging, boosting and the random subspace method for linear classifiers. *Pattern Analysis and Applications, 5*(2), 121–135.

Sun, S., Zhang, C., & Zhan, D. (2007). An experimental evaluation of ensemble methods for EEG signal classification. *Pattern Recognition Letters, 28*, 2157–2163.

Turkoglu, I., Arslan, A., & Ilkay, E. (2002). An expert system for diagnosis of the heart valve diseases. *Expert Systems with Applications, 23*, 229–236.

Turkoglu, I., Arslan, A., & Ilkay, E. (2003). An intelligent system for diagnosis of heart valve diseases with wavelet packet neural networks. *Computer in Biology and Medicine, 33*(4), 319–331.

Uguz, H., Arslan, A., & Turkoglu, I. (2007). A biomedical system based on hidden Markov model for diagnosis of the heart valve diseases. *Pattern Recognition Letters*, Available online 11 October 2006.

Vapnik, V. (1998). *Statistical learning theory*. New York: John Wiley and Sons.

Wright, I. A., Gough, N. A. J., Rakebrandt, F., Wahab, M., & Woodcock, J. P. (1997). Neural network analysis of Doppler ultrasound blood flow signals: A pilot study. *Ultrasound in Medicine and Biology, 23*(5), 683–690.