ELSEVIER

Computers in Biology and Medicine 38 (2008) 329–338

Computers in Biology and Medicine

www.intl.elsevierhealth.com/journals/cobm

An expert system based on principal component analysis, artificial immune system and fuzzy *k*-NN for diagnosis of valvular heart diseases

Abdulkadir Sengur

Department of Electronics and Computer Science, Technical Education Faculty, Firat University, 23119, Elazig, Turkey

Received 30 December 2006; accepted 24 November 2007

Abstract

In the last two decades, the use of artificial intelligence methods in medical analysis is increasing. This is mainly because the effectiveness of classification and detection systems have improved a great deal to help the medical experts in diagnosing. In this work, we investigate 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. The proposed heart valve disorder detection system is composed of three stages. The first stage is the pre-processing stage. Filtering, normalization and white de-noising are the processes that were used in this stage. The feature extraction is the second stage. During feature extraction stage, wavelet packet decomposition was used. As a next step, wavelet entropy was considered as features. For reducing the complexity of the system, PCA was used for feature reduction. In the classification stage, AIS and fuzzy *k*-NN were used. To evaluate the performance of the proposed methodology, a comparative study is realized by using a data set containing 215 samples. The validation of the proposed method is measured by using the sensitivity and specificity parameters; 95.9% sensitivity and 96% specificity rate was obtained.

Keywords: Doppler heart sounds; Heart valves; Wavelet packet decomposition; Principal component analysis; Artificial immune system; Fuzzy *k*-NN

1. Introduction

According to the researches most of human deaths in the world are due to heart diseases. Valvular heart disease (VHD) affects nearly 200,000 patients per year world-wide. As recently as 40 years ago, the disease was untreatable, leading to cardiopulmonary failure and eventual death [1]. Early detection of heart valve disorders is necessary in the medical research areas [2].

The heart consists of four chambers, two atria and two ventricles. There is a valve through which blood passes before leaving each chamber of the heart. These valves are actual flaps that prevent backward flow of blood. VHD occurs when one or more valves in the heart are not working properly and severe VHD can cause the heart to pump less efficiently leading to symptoms such as breathlessness and ankle swelling [3]. VHD may be suspected if heart sounds heard through a stethoscope are abnormal. Auscultation with a stethoscope is usually the first step in diagnosing VHD. A characteristic heart murmur (abnormal sounds in the heart due to turbulent blood flow) can often indicate valve regurgitation or stenosis. To further define the type of valve disease and the extent of the valve damage, physicians may use any of the following diagnostic procedures: electrocardiogram (ECG or EKG), chest X-ray, cardiac catheterization, transesophageal echo (TEE), radionuclide scans and magnetic resonance imaging (MRI) [4].

In the last decade, Doppler technique has gained much more interest since Satomura first demonstrated the application of the Doppler effect to the measurement of blood velocity in 1959 [5]. Doppler heart sounds (DHS) are one of the most important sounds produced by blood flow, valve motion and vibration of the other cardiovascular components [6]. 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 [6]. In addition to Doppler techniques, the techniques that are more complex have also been developed (laplace transform

E-mail address: ksengur@firat.edu.tr.

0010-4825/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2007.11.004

and principal components analysis (PCA)) [13]. Many studies have been implemented to classify Doppler signals in the pattern recognition field [7,8].

In this study, an expert system has been proposed that has three stages. The first stage presents several pre-processing units for DHS signals. White de-noising, normalization and filtering are the components of the pre-processing unit. The second stage includes wavelet package decomposition and wavelet entropy to extract features from the pre-processed Doppler signals to distinguish normal and abnormal heart valves. Feature reduction using PCA was also carried out in this stage [9]. The third stage is the classifier stage where an artificial immune system and fuzzy k-nearest neighbor (k-NN) structure were hybridized.

2. Background

2.1. Previous research

Up to now, several papers have been proposed to determine Doppler signals by using pattern recognition approaches [9–12]. Turkoglu et al. [9] proposed an expert diagnosis system for interpretation the Doppler signals of the VHDs based on the pattern recognition. The proposed methodology was composed of a feature extractor and a back-propagation artificial neural networks (BP-ANN) classifier. Wavelet transforms and short time Fourier transform methods were used for feature extraction from the Doppler signals on the time-frequency domain. Wavelet entropy method was applied to these features. The back-propagation neural network was used to classify the extracted features. The performance of the developed system has been evaluated in 215 samples. The test results showed that this system was effective to detect Doppler heart sounds. The correct classification rate was about 94% for normal subjects and 95.9% for abnormal subjects.

Later, Turkoglu et al. [10] proposed an intelligent system for detection of VHD-based on wavelet packet neural networks (WPNNs). In this work, a methodology was considered which was a combination of the feature extraction and classification from measured Doppler signal waveforms at the heart valve using the Doppler ultrasound. The proposed methodology was composed of two layers: wavelet layer and multi-layer perceptron layer. The wavelet layer was used for adaptive feature extraction in the time-frequency domain and was composed of wavelet packet decomposition and wavelet packet entropy. The multi-layer perceptron used for classification is a feed-forward neural network. The reported correct classification rate was about 94% for abnormal and normal subjects.

The data set (215 samples) which was obtained by Turkoglu et al. [9] was later used by Çomak et al. [11]. Çomak et al. [11] investigated the use of least-square support vector machines (LS-SVM) classifier for improving the performance of Turkoglu's [9] proposal. 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 same as BP-ANN. It was reported that LS-SVM was more suitable than BP-ANN since it has some advantages over 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.

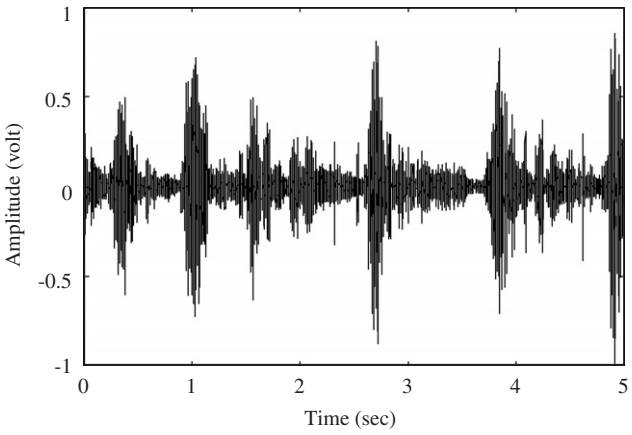

Fig. 1. The waveform pattern of the Doppler heart sound.

Later, Uguz et al. [12] proposed a biomedical system based on hidden Markov model for clinical diagnosis and recognition of VHDs. The proposed methodology was also used for the database of Turkoglu et al. [9]. 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 92% and 94% for CHMM, 92% and 97.26% for FCM–K-means/CHMM), respectively.

2.2. DHS signals

The DHS can be obtained directly by placing the Doppler ultrasonic flow transducer over the chest of the patient [7]. A DHS signal from aortic heart valve is shown in Fig. 1. The DHS produced from echoes backscattered by moving blood cells is generally in the range of 0.5–10 kHz [13]. DHS signal spectral estimation is now commonly used to evaluate blood flow parameters in order to diagnose cardiovascular diseases. Spectral estimation methods are particularly used in Doppler ultrasound cardiovascular disease detection. Clinical diagnosis procedures generally include analysis of a graphical display and parameter

measurements, produced by blood flow spectral evaluation. Ultrasonic instrumentation typically employ Fourier-based methods to obtain the blood flow spectra and blood flow measurements [14].

A Doppler signal is not a simple signal. It includes random characteristics due to the random phases of scattering particles present in the sample volume. Other effects such as geometric broadening and spatially varying velocity also affect the signal [15]. The following is Doppler equation:

$$\Delta f = \frac{2vf\cos\theta}{c}, \tag{1}$$

where v equals the velocity of the blood flow, f equals the frequency of the emitted ultrasonic signal, c equals the velocity of sound in tissue (approximately 1540 m/s), Δf equals the measured Doppler frequency shift, and θ equals the angle of incidence between the direction of blood flow and the direction of the emitted ultrasonic beam [13].

2.3. Wavelet packet decomposition

Wavelet transforms are finding inversed use in fields as diverse as telecommunications and biology. Because of their suitability for analyzing non-stationary signals, they have become a powerful alternative to Fourier methods in many medical applications, where such signals abound [5,16–18].

The main advantages of wavelets are that they have a varying window size, being wide for slow frequencies and narrow for the fast ones, thus leading to an optimal time-frequency resolution in all the frequency ranges. Furthermore, owing to the fact that windows are adapted to the transients of each scale, wavelets lack the requirement of stationarity [19].

Wavelet decomposition uses the fact that it is possible to resolve high frequency components within a small time window, while only low frequency components need large time windows. This is because a low frequency component completes a cycle in a large time interval, whereas a high frequency component completes a cycle in a much shorter interval. Therefore, slow varying components can only be identified over long time intervals but fast varying components can be identified over short time intervals. Wavelet decomposition can be regarded as a continuous time wavelet decomposition sampled at different frequencies at every level or scale. The wavelet decomposition functions at level m and time location t_m can be expressed as follows:

$$d_m(t_m) = x(t) * \Psi_m\left(\frac{t - t_m}{2^m}\right), \tag{2}$$

where Ψ_m is the decomposition filter at frequency level m. The effect of the decomposition filter is scaled by the factor 2^m at stage m, but otherwise the shape is the same at all scales [17].

Wavelet packet analysis is an extension of the discrete wavelet transform (DWT) [20] and it turns out that the DWT is only one of the much possible decompositions that could be performed on the signal. Instead of just decomposing the low frequency component, it is therefore possible to subdivide the whole time-frequency plane into different time-frequency pieces as can be seen from Fig. 2. The advantage of the wavelet packet analysis is that it is possible to combine the different levels of decomposition in order to achieve the optimum time-frequency representation of the original [5].

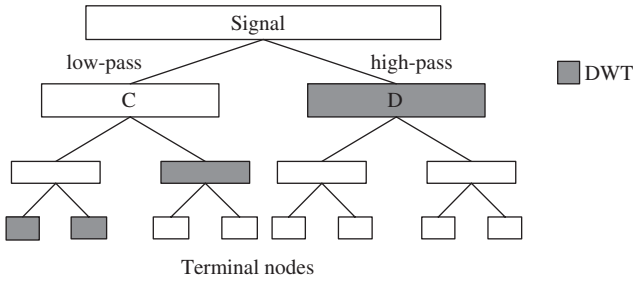

Fig. 2. Total decomposition tree of a time varying signal using wavelet packet analysis.

2.4. Wavelet entropy

Entropy-based criteria describe information-related properties for an accurate representation of a given signal. Entropy is a common concept in many fields, mainly in signal processing [18]. A method for measuring the entropy appears as an ideal tool for quantifying the ordering of non-stationary signals. An ordered activity (i.e. a sinusoidal signal) is manifested as a narrow peak in the frequency domain, thus having low entropy. On the other hand, random activity has a wide band response in the frequency domain, reflected in a high entropy value [19]. The types of entropy computing are Shannon, threshold, norm, log energy and sure [18].

2.5. Natural and artificial immune system

"The immune system is a complex of cells, molecules and organs with the primary role of limiting damage to the host organism by pathogens, which elicit an immune response and thus are called antigens (Ag)" [21,22]. Immune system constitutes the defense mechanism of the body by means of innate and adaptive immune responses. Between these, adaptive immune response is more important for human being because it contains metaphors like recognition, memory acquisition, and etc. The main component of adaptive immune response is lymphocytes, which can be divided into two classes as T and B lymphocytes (cells), each having its own function. B cells have a great functionality because of their secreted antibodies (Ab) that take very critical roles in adaptive immune response. The illustration of the working procedure of the immune system is shown in Fig. 3 [21,22].

Specialized antigen presenting cells (APCs) called macrophages circulate through the body and if they encounter an antigen, they ingest and fragment them into antigenic peptides (I). The pieces of these peptides are displayed on the cell surface by major histocompatibility complex (MHC) molecules existing in the digesting APC. The presented MHC–peptide combination on the cell surface is recognized by the T cells

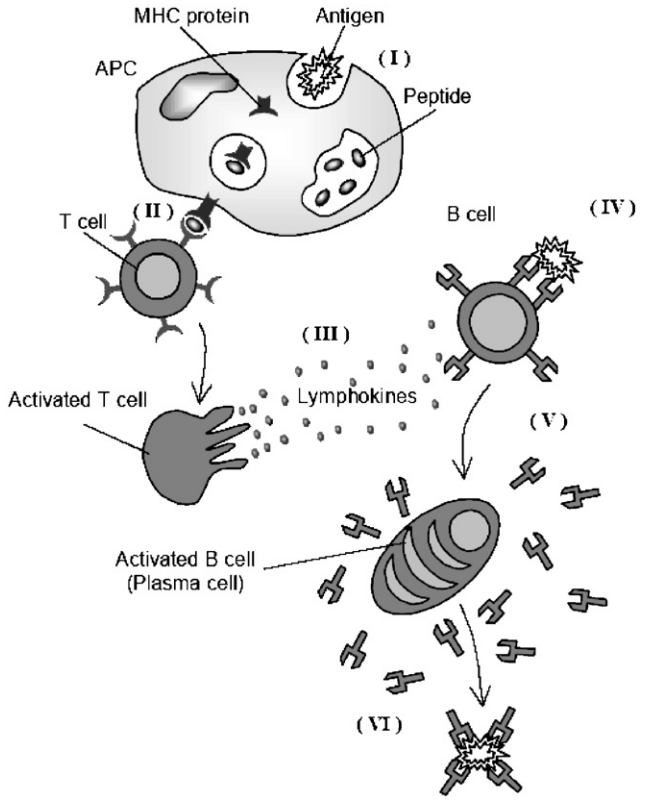

Fig. 3. Human immune system [21].

causing them to be activated (II). Activated T cells secrete some chemicals as alert signals to other units in response to this recognition. B cells, one of the units that take these signals from the T cells, become activated with the recognition of antigen by their antibodies occurring at the same time (IV). When activated, B cells turn into plasma cells that secrete bound antibodies on their surfaces (V). Secreted Abs bind the existing antigens and neutralize them signaling other components of immune system to destruct the antigen–antibody complex (VI) [21,22]. Artificial immune systems (AISs) emerged in the 1990s as a new computational research area. AISs link several emerging computational fields inspired by biological behaviors such as artificial neural networks and artificial life.

Among the AIS methodologies, B cell modeling is the commonly used method. Different representation methods have been proposed in that modeling. Among these, shape-space representation is the most commonly used one [22]. The shape-space model (S) aims at quantitatively describing the interactions among antigens (Ags), the foreign elements that enter the body like microbe, etc., and antibodies (Ag–Ab). The set of features that characterize a molecule is called its *generalized shape*. The Ag–Ab representation (binary or realvalued) determines a distance measure to be used to calculate the degree of interaction between these molecules. Mathematically, the generalized shape of a molecule (m), either an antibody or an antigen, can be represented by a set of coordinates $m = \{m1, m2, \ldots, mL\}$, which can be regarded as a point in an L-dimensional realvalued shape space $m \in S^L$. In this work, we used real strings to represent the molecules. Ags and Abs were considered of the same length L. The length and cell representation depend upon the problem [30].

2.6. Fuzzy k-nearest neighbor classification

k-NN is a supervised learning algorithm where the result of the new sample is classified based on majority of k-NN category [23]. The purpose of this algorithm is to classify a new object based on features and training patterns. The k-NN classifiers do not use any model to fit only based on memory. Given a sample point, a K number of objects closest to the samples point are found. The classification uses majority vote among the classification of the K objects. There are some problems with k-NN algorithms [23]. First, each of the neighbors is considered equally important in determining the classification of the input data. This causes difficulty in those places where the sample sets overlap. Moreover, a far neighbor to the input is given the same weight as a close neighbor. Second, the algorithm only assigns a class label to the input data; it does not determine the strength of membership in the class. The fuzzy k-NN classifier, designed by Keller et al. [24,25], is a fuzzy classification technique which overcomes the limitations of the k-NN classifier. The fuzzy k-NN algorithm assigns to pattern x a membership vector $(\mu_{\omega 1}(x), \mu_{\omega 2}(x), \ldots, \mu_{\omega M}(x))$ as a function of the pattern's distance from its k-NNs. This ensures that no arbitrary assignments are made. The class membership of the input pattern x is calculated based on the following formula:

$$\mu_{\omega i}(x) = \frac{\sum_{j=1}^{K} \mu_{\omega i}(x_j) d_j^{-2/(m-1)}}{\sum_{j=1}^{K} d_j^{-2/(m-1)}}, \tag{3}$$

where $x_1, x_2, \ldots, x_k$ denote the k-NN labeled reference patterns of x, and $d_j = \|x - x_j\|$ is the distance between x and its jth nearest neighbor x_j. The pattern x is assigned to the class given by

$$\omega(x) = \arg \max_{i=1}^{M} (\mu_{\omega i}(x)). \tag{4}$$

The fuzzification parameter m determines how heavily the distance is weighted when calculating the class membership. The fuzzy k-NN classifier is based on the estimation of the membership functions for the labeled reference patterns. Methods for automatic estimation of membership functions have been summarized in [24,25]. We are interested in the fuzzy nearest neighbor labeling techniques rather than crisp labeling. In crisp labeling, each labeled reference pattern is assigned complete membership in its class and zero membership in all other classes, i.e., $l(v_i) \in [0, 1]^M$. The fuzzy nearest neighbor labeling, known as soft labeling, assigns memberships to labeled reference patterns according to the k-NNs rule. It is required to estimate M degrees of membership $(\mu_{\omega 1}(vi), \mu_{\omega 2}(vi), \ldots, \mu_{\omega M}(vi))$ for any $v_i \in V$ by first finding the k patterns in V closest to each labeled reference pattern vi and then calculating the membership functions. The

soft labeling of reference patterns was performed using the technique proposed by Keller et al. [24]:

$$\mu_{\omega i}(v_j) = \begin{cases} 0.51 + \left(0.49\frac{k_i}{k}\right) & \text{if } j = i, \\ \left(0.49\frac{k_i}{k}\right) & \text{if } j \neq i, \end{cases} \quad (5)$$

where k_i is the number of labeled reference patterns amongst the k closest labeled reference patterns which are labeled in class ω_i, and j ranges from 1 to n. Notice that the classes share the membership, i.e.,

$$\sum_{i=1}^{M} \mu_{\omega i}(v_j) = 1, \quad \forall v_j \in V. \quad (6)$$

3. Methodology

The proposed methodology for detection of VHDs is illustrated in Fig. 4. It consists of three parts: (a) data acquisition and pre-processing, (b) feature extraction and feature reduction, and (c) classification using AIS and fuzzy k-NN.

3.1. Data acquisition and pre-processing

All the original audio 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 by using a sound card which has 16-bit A/D conversion resolution and computer software prepared by us in the MATLAB version 5.3) (The MathWorks Inc., USA). The Doppler ultrasonic flow transducer used (Model 3V2c) was run in a continuous operating mode at 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:

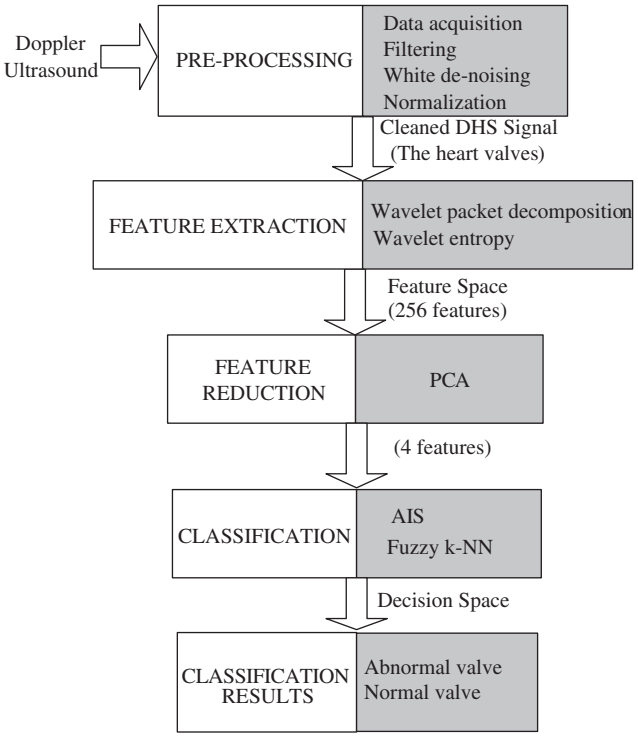

Fig. 4. The algorithm of the expert diagnostic system.

(i) Filtering: The stored DHS signals were high-pass filtered to remove unwanted low frequency components, because the DHS signals are generally in the range of 0.5–10 kHz. The filter is a digital FIR, which is a 50th-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 [17]. The DHS signals were filtered from removing the white noise by using wavelet packet. The white de-noising procedure contains three steps [26]:
   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. (7) so that the expected amplitude of the signal is not affected from the rib cage structure of the patient.

$$\mathrm{DHS}_{\mathrm{signal}} = \frac{\mathrm{DHS}_{\mathrm{signal}}}{|(\mathrm{DHS}_{\mathrm{signal}})_{\max}|}. \quad (7)$$

3.2. Feature extraction

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:

Stage 1—wavelet packet decomposition: For wavelet packet decomposition of the DHS waveforms, the tree structure used was 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 [18] as defined in Eq. (8). In the equation s is the DHS signal and (s_i) i the coefficients of wavelet packet decomposition of s, thus obtaining $2^8 = 256$-terminal node signals.

$$E(s) = -\sum_i s_i^2 \log(s_i^2). \quad (8)$$

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 [19]. 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. (9) of the waveforms at the terminal node signals obtained from wavelet packet decomposition.

$$E(s) = \frac{\sum_i |s_i|^p}{N}, \tag{9}$$

where the entropy E is a real number, s is the terminal node signal and (s_i) i 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.

3.3. PCA for feature reduction

The number of extracted features for each DHS signal is 256. In other words, the classifier which will be used for this system (VHD diagnosis system) has 256 inputs. Because of the complexity problem, the dimension of the input vector has to be reduced. Therefore, PCA was used for feature reduction. The dimension of VHD that had 256 features was reduced to 4 using PCA.

PCA is based on the assumption that most information about classes is contained in the directions along which the variations are the largest. The derivation of PCA is in terms of a linear projection, which maximizes the variance in the projected space [27,28]. For a given p-dimensional data set X, the m principal axes $T_1, T_2, \ldots, T_m$, where $1 \leqslant m \leqslant p$, are orthonormal axes onto which the retained variance is maximum in the projected space. Generally, $T_1, T_2, \ldots, T_m$ can be given by the m leading eigenvectors of the sample covariance matrix $S = (1/N)\sum_{i=1}^{N}(x_i - \mu)^{\mathrm{T}}(x_i - \mu)$, where $x_i \in X$, μ is the sample mean, and N is the number of samples, so that

$$ST_i = \lambda_i T_i, \quad i \in 1, \ldots, m, \tag{10}$$

where λ_i is the ith largest eigenvalue of S. The m principal components of a given observation vector $x_i \in X$ are given by

$$y = [y_1, y_2, \ldots, y_m] = [T_1^{\mathrm{T}}x, T_2^{\mathrm{T}}, \ldots, T_m^{\mathrm{T}}] = T^{\mathrm{T}}x. \tag{11}$$

The m principal components of x are decorrelated in the projected space. In multiclass problems, the variations of data are determined on a global basis, that is, the principal axes are derived from a global covariance matrix:

$$\widehat{S} = \frac{1}{N}\sum_{j=1}^{K}\sum_{i=1}^{N_j}(x_j - \widehat{\mu})(x_j - \widehat{\mu})^{\mathrm{T}}, \tag{12}$$

where $\widehat{\mu}$ is the global mean of all the samples, K is the number of classes, N_j is the number of samples in class j; $N = \sum_{j=1}^{K} N_j$ and x_{ji} represents the ith observation from class j. The principal axes $T_1, T_2, \ldots, T_m$ are therefore the m leading eigenvectors of $\widehat{S}$:

$$\widehat{S}\,T_i = \widehat{\lambda}_i T_i, \quad i \in 1, 2, \ldots, m, \tag{13}$$

where $\widehat{\lambda}_i$ is the ith largest eigenvalue of $\widehat{S}$. An assumption made for feature extraction and dimensionality reduction by PCA is that most information of the observation vectors is contained in the subspace spanned by the first m principal axes, where $m < p$. Therefore, each original data vector can be represented by its principal component vector with dimensionality m [28].

3.4. Classification with AIS and fuzzy k-NN

Throughout a human's life, the body is subjected to pathogenic materials. The immune system contains lymphocyte cells known as B and T cells, each of which has a unique type of molecular receptor. These receptors in the shape space allow for binding of the pathogenic materials known as antigens with the higher affinity between the receptor and the antigen indicating a stronger bind [29]. Immune cells have the ability of grouping foreign invaders with regard to their origin, and immune responses are produced according to this grouping. This clustering capability of immune system has been used as an inspiration source to design the data reduction algorithm in this study. The aiNET algorithm, originally proposed by De Castro et al. [30] is a discrete immune network algorithm that was developed for data compression and clustering and was also extended slightly and applied to optimization to create the algorithm opt-aiNET [31]. Actually, aiNET is AIS based unsupervised clustering algorithm and here used for data reduction purposes in a supervised manner. At first sight, this situation can be seen as a conflict because using an unsupervised methodology in a supervised manner is not sensible. But in this study, the data set for each class is known and each class data set can be used via aiNET for constituting the antibodies. Thus, aiNET is employed for each class and the obtained antibodies are saved and later used for classification purposes. The presented training samples are named as Ag while formed memory units which will be then used as classifying units for k-NN algorithm are called Ab in the algorithm.

The following notation was adopted in Ref. [30] and also used here:

- X data set composed of N_p patterns (feature vectors) of dimension p;
- C matrix containing all the N_t network cells ($C \in R^{(Nt \times p)}$);
- M matrix of the N memory cells, each of dimension p ($M \subseteq C$);
- Nc total number of clones generated by the stimulated cells at each iteration;
- D (dissimilarity) matrix with elements d_{ij} of Ag–Ab affinity;
- S (similarity) matrix with elements s_{ij} of Ab–Ab affinity;
- N n highest affinity cells selected for cloning and mutation;
- ζ percentage of the matured cells to be selected; and
- $\sigma_{d,s}$ natural death and suppression threshold, respectively.

The training procedure of the algorithm conducts the following steps as reported in Ref. [26]:

1. At each iteration step, do:
    1.1. For each antigen i, do:
        1.1.1. Determine the class of the Ag_i and call the memory cell Ab of that class and calculate the affinity of the given antigen to all the network cells according to the Euclidean distance metric in shape space, d_{ij};
        1.1.2. Select the n (or $n\%$ of the) highest affinity network cells;
        1.1.3. Clone these n selected cells. The number of progeny of each cell, Nc, being proportional to their affinity: the higher the affinity, the larger the clone size;
        1.1.4. Increase the affinity of these Nc cells to antigen i, by reducing the distance between them (see Eq. (1));
        1.1.5. Calculate the affinity of these improved cells via Euclidean distance metric with antigen i;
        1.1.6. Re-select $\zeta\%$ of the most improved cells and put them into a partial matrix Mp of memory cells;
        1.1.7. Eliminate those cells whose affinity is inferior to threshold σ_d (affinity threshold), yielding a reduction in the size of the Mp matrix;
        1.1.8. Determine the network cell–cell (Ab–Ab) affinity, s_{ij};
        1.1.9. Eliminate those cells whose affinity s_{ij} is inferior to threshold σ_s, leading to another possible reduction in Mp (clonal suppression);
        1.1.10. Concatenate the original network cell matrix with the partial matrix of memory cells;
    1.2. Determine the whole network inter-cell affinities and eliminate those cells whose affinity with each other is inferior to threshold σ_s (network suppression);
    1.3. Replace $r\%$ of the worst individuals by novel randomly generated ones;
2. Test the stopping criterion. If the stopping criterion is reached, save the memory cell Ab for that class.

As it was noted in [18], steps 1.1.1–1.1.7 describe the clonal selection and affinity maturation processes. Steps 1.1.8–1.1.10 and 1.2–1.3 simulate the activity of immune network. The affinity of the cells with the given antigen i can be improved as follows:

$$C = C - \alpha(C - X), \tag{14}$$

where C is the matrix of network cells, X the matrix of antigens and α is the learning rate or mutation rate. The learning algorithm aims at building a memory set of network cells that recognize the data and hence represent their structural information. The resulting memory cells (Ab) constitute the training samples in the algorithm of fuzzy k-NN. These memory cells carry the same class information in training samples if their number and location in sample space are well adjusted in the above training algorithm. This adjustment is realized through σ_s parameter and mutation mechanism, respectively. Especially, the number of memory Ab affects the classification performance a great deal since with a few number of memory units it is hard to represent the class information hidden in training data while a high number of these memory units prevent us to reduce the data.

4. Performance evaluation methods

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

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:

$$\text{Sensitivity}\% = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{15}$$

$$\text{Specificity}\% = \frac{\text{TN}}{\text{FP} + \text{TN}}, \tag{16}$$

where TP, TN, FP and FN denote true positives, true negatives, false positives and false negatives, respectively [32]. TP represents an input detected as a patient labeled with disease that was also diagnosed by the expert clinicians. TN represents an input detected as a patient who is labeled as healthy and was also labeled as healthy by the expert clinicians. FP represents an input detected as a patient labeled with disease but diagnosed as healthy by the expert clinicians. FN represents an input detected as a patient labeled as healthy but diagnosed with disease 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 [33].

5. Experimental classification results

The aortic and mitral valves of 132 men and 83 women, ages 15–80 (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 [7]. 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. The remaining data set was used as the test data set.

As it was mentioned in the earlier section, the number of the features that was used for characterizing the normal and abnormal mitral and aortic heart valves was 256. The dimension of the feature vector was reduced to 4 with the PCA algorithm which was described in Section 3.3. Moreover, experiments have been done with constant value of the fuzzy strength parameter m and the number of nearest neighbor k. A value of 7 was assigned to k value and the m value was 2 for the experimental study.

The experiments were conducted to generate the optimum number of memory cells (Ab) that were used for classification purposes. σ_s is the parameter which determines the number of Ab memory cells. The value of this parameter was selected in the [0 1] range for each class. When this value is selected too low, more memory cells are generated. On the contrary, if we selected this value too high the number of memory cells will be too low. The number of memory cells highly affects the classification accuracy. It is worth nothing that the algorithm of the aiNET does not train for each class of the given data set. Thus, memory cells for each class which will be used later in fuzzy k-NN algorithm are obtained and saved.

The training parameters of aiNET were chosen as follows:

- The aiNET structure was initialized with a random number of 10 cells because as it was noted in Ref. [30], aiNET is not sensitive to the initial configurations.
- Euclidean distance was employed for affinity calculation.
- A fixed number of $n = 4$ cells with the highest affinity Ab were selected to proliferate.
- Thus, the clone generated size Nc was calculated according to the following equation:

$$\mathrm{Nc} = \sum_{j} \mathrm{round}(N - D_{j,k}N),$$

  where the sum is employed for all antibodies selected that identify the antigen k and N is the total number of antibodies in the network.
- $\zeta = 10\%$ was selected to be maintained as memory cells.
- The number of generation was 200.
- The network suppression parameter was adjusted manually for generating a fixed number of memory cells.
- The mutation process was done as it was defined in Section 3.3.

Table 1
Testing results of the proposed methodology

	Mitral valve				Aortic valve				All valves			
	Disease	Healthy	Total	Accuracy (%)	Disease	Healthy	Total	Accuracy (%)	Disease	Healthy	Totals	Accuracy (%)
Test results (predictions)	TP = 30	FP = 3	33	90.9	TP = 40	FP = 0	40	100.0	TP = 70	FP = 3	73	95.9
	FN = 0	TN = 19	19	100.0	FN = 2	TN = 29	31	93.5	FN = 2	TN = 48	50	96.0
Total	30	22	52	94.2	42	29	71	97.2	72	51	123	95.9

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	Sens = sensitivity, Spec = Specificity (%)
PCA, AIS and Fuzzy k-NN	Abnormal	73	70	3	Sens = 95.9
	Normal	50	48	2	Spec = 96
ANN [9]	Abnormal	73	70	3	Sens = 95.9
	Normal	50	47	3	Spec = 94
WPNN [10]	Abnormal	73	69	4	Sens = 94.5
	Normal	50	47	3	Spec = 94
SVM [11]	Abnormal	73	69	4	Sens = 94.5
	Normal	50	45	5	Spec = 90
FCM/CHMM [12]	Abnormal	73	71	2	Sens = 97.3
	Normal	50	46	4	Spec = 92

The classification accuracy of the proposed PCA, AIS and fuzzy k-NN algorithm's testing results are given in Table 1. Three abnormal heart mitral valve patterns were classified incorrectly, whereas two normal heart aortic valve patterns were classified incorrectly by the proposed methodology.

The confusion matrix and the calculated sensitivity and specificity rates are given in Table 2. The performance comparison of the proposed system with the other classifiers from literature was also demonstrated in Table 2. According to these results, higher sensitivity rate (97.3%) was obtained by using the FCM–CHMM [12]. Our proposal and ANN methods produced the same and the second higher sensitivity rate (95.9%). SVM method gained the worst sensitivity rate (94.5%). On the other hand, the higher specificity rate (96%) was obtained with our proposal. FCM–CHMM produced the 92% specificity value. ANN and WPNN obtained the same 94% specificity rate. And finally, the SVM produced the lower specificity rate (90%).

6. Discussion and conclusion

In this study, a medical decision support system with normal and abnormal classes has been developed. In the related previous work, a feature vector which has 256 features was considered as input to the classifiers [11]. The complexity of the previous work was quite high due to the dimension of the input feature vector. Therefore, a data reduction process was emerged. Data reduction with PCA and AIS with fuzzy k-NN was used in this work for diagnosis of aortic and mitral VHDs from DHS signals. For this purpose, computer simulations were carried out and statistical validation indexes were used for determining the performance of the proposed methodology. The previous reported results were also compared with our proposal.

The proposed methodology was composed of three main stages. These stages are pre-processing, feature extraction, and feature reduction with PCA and classification. The wavelet packet decomposition for multi-scale analysis and the wavelet entropy was used for feature extraction, whereas PCA was used for feature reduction and AIS with fuzzy k-NN classifier was adopted for efficient recognition.

According to the experimental results, the proposed method is efficient for interpretation of the disease. Our proposal produced 95.9% sensitivity rate and 96% specificity rate. This specificity rate is the highest rate. ANN and WPNN almost produced the similar results. SVM method gained the worst sensitivity and specificity rate. FCM–CHMM produced 97.3% sensitivity and 92% specificity values; 97.3% sensitivity rate is the highest value among the proposed previous works.

The proposed system has some advantages of automation. It is rapid, easy to process, non-invasive and cheap for clinical application. This system is of better clinical application especially for earlier survey of population. However, the position of the ultrasound probe, which is used for data acquisition from the heart valves, must be taken into consideration by physician.

Conflict of interest statement

None declared.

Acknowledgment

The author would like to thank Dr. Ibrahim Turkoglu for providing the DHS signals and for his valuable suggestion in improving the technical presentation of this paper.

References

[1] American Heart Association, ⟨http://www.americanheart.org/presenter.jhtml?identifier = 3000333⟩ (accessed: 10.09.2006).
[2] M. Akay, Y.M. Akay, W. Welkowitz, Neural networks for the diagnosis of coronary artery disease, in: International Joint Conference on Neural Networks, IJCNN, vol. 2, 1992, pp. 419–424.
[3] E. Braunwald, D.P. Zipes, P. Libby, R. Bonow, Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine, Seventh ed., Saunders, London, 2004.
[4] N.C. Nanda, Doppler Echocardiography, second ed., Lea & Febiger, London, 1993.
[5] P.I.J. Keeton, F.S. Schlindwein, Application of Wavelets in Doppler Ultrasound, vol. 17(1), MCB University Press, 1997, pp. 38–45.
[6] I.A. Wright, N.A.J. Gough, F. Rakebrandt, M. Wahab, J.P. Woodcock, Neural network analysis of Doppler ultrasound blood flow signals: a pilot study, Ultrasound Med. Biol. 23 (5) (1997) 683–690.
[7] B.C.B. Chan, F.H.Y. Chan, F.K. Lam, P.W. Lui, P.W.F. Poon, Fast detection of venous air embolism is Doppler heart sound using the wavelet transform, IEEE Trans. Biomed. Eng. 44 (4) (1997) 237–245.
[8] I. Guler, M.K. Kiymik, S. Kara, M.E. Yuksel, Application of autoregressive analysis to 20 MHz pulsed Doppler data in real time, Int. J. Biomed. Comput. 31 (3–4) (1992) 247–256.
[9] I. Turkoglu, A. Arslan, E. Ilkay, An expert system for diagnosis of the heart valve diseases, Expert Syst. Appl. 23 (3) (2002) 229–236.
[10] I. Turkoglu, A. Arslan, E. Ilkay, An intelligent system for diagnosis of heart valve diseases with wavelet packet neural networks, Comput. Biol. Med. 33 (4) (2003) 319–331.
[11] E. Çomak, A. Arslan, I. Türkoğlu, A decision support system based on support vector machines for diagnosis of the heart valve diseases, Comput. Biol. Med. 37 (2007) 21–27.
[12] H. Uguz, A. Arslan, I. Türkoğlu, A biomedical system based on hidden Markov model for diagnosis of the heart valve diseases, Pattern Recognition Lett. 28 (4) (2007) 395–404.
[13] V.D. Saini, N.C. Nanda, D. Maulik, Basic principles of ultrasound and Doppler effect, in: Doppler Echocardiography, Lea & Febiger Philadelphia, London, 1993.
[14] M.M. Madeira, M.O. Tokhi, M.G. Ruano, Real-time implementation of a Doppler signal spectral estimator using sequential and parallel processing techniques, Microprocessors Microsyst. 24 (2000) 153–167.
[15] E. Karabetsos, C. Papaodysseus, D. Koutsouris, Design and development of a new ultrasonic doppler technique for estimation of the aggregation of red blood cells, Measurement 24 (1998) 207–215.
[16] R.Q. Quiroga, Quantitative Analysis of EEG Signals: Time-Frequency Methods and Chaos Theory, Institute of Physiology, Medical University Lübeck, 1998.
[17] S.R. Devasahayam, Signals and Systems in Biomedical Engineering, Kluwer Academic Publishers, Dordrecht, MA, 2000.
[18] R.R. Coifman, M.V. Wickerhauser, Entropy-based algorithms for best basis selection, IEEE Trans. Inf. Theory 38 (2) (1992) 13–718.
[19] R.Q. Quiroga, O.A. Roso, E. Basar, Wavelet entropy: a measure of order in evoked potentials, in: Evoked Potentials and Magnetic Fields, vol. 49, Elsevier, Amsterdam, 1999, pp. 298–302.
[20] C.S. Burrus, R.A. Gopinath, H. Guo, Introduction to Wavelet and Wavelet Transforms, Prentice-Hall, Englewood Cliffs, NJ, 1998.
[21] L.N. De Castro, J. Timmis, Artificial Immune Systems: A New Computational Intelligence Approach, Springer, UK, 2002.
[22] S. Şahan, K. Polat, H. Kodaz, S. Güneş, A new hybrid method based on fuzzy-artificial immune system and k-NN algorithm for breast cancer diagnosis, Comput. Biol. Med. 37 (3) (2007) 415–423.

[23] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.
[24] J.M. Keller, M.R. Gray, J.A. Givens Jr., A fuzzy *k*-nearest neighbor algorithm, IEEE Trans. Syst. Man Cybern 15 (4) (1985) 580–585.
[25] S. Rasheed, D. Stashuk, M. Kamel, Adaptive fuzzy *k*-NN classifier for EMG signal decomposition, Med. Eng. Phys. 28 (2006) 694–709.
[26] A. Bakhtazad, A. Palazoglu, J.A. Romagnoli, Process data de-noising using wavelet transform, Intell. Data Anal. 3 (1999) 267–285.
[27] X. Wang, K.K. Paliwal, Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition, Pattern Recognition 36 (2003) 2429–2439.
[28] K. Polat, S. Güneş, Automatic determination of diseases related to lymph system from lymphography data using principles component analysis (PCA), fuzzy weighting preprocessing and ANFIS, Expert. Syst. Appl. 33 (3) (2007) 636–641.
[29] A. Watkins, J. Timmis, L. Boggess, Artificial immune recognition system: an immune-inspired supervised learning algorithm, Genet. Programming Evolvable Mach. 5 (3) (2004) 291–317.
[30] L.N. De Castro, F.J. Von Zuben, aiNet: an artificial immune network for data analysis, Draft, 2001.
[31] L.N. De Castro, J. Timmis, An artificial immune network for multimodal function optimisation, in: Proceedings of IEEE World Congress on Evolutionary Computation, 2002, pp. 669–674.
[32] T.J. Jordan, Understanding Medical Information: A User's Guide to Informatics and Decision Making, McGraw-Hill, New York, 2002.
[33] ⟨http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion_matrix/confusion_matrix.html⟩, (last accessed: 01.12.2006).

Abdulkadir Sengur graduated from the department of Electronics and Computer Education at Firat University in 1999. He obtained his M.S. degree from the same department and the same university in 2003. His Ph.D. degree was from the department of Electronic Engineering at Firat University in 2006. He is a research assistant at Firat University. His interest areas include pattern recognition, machine learning and image processing.