Classification of aorta doppler signals using variable coded-hierarchical genetic fuzzy system

İnan Gülera,*, Fırat Hardalac, Uçman Ergünb, Necaat Barışçia

aDepartment of Electronics and Computer Education, Faculty of Technical Education, Gazi University, 06500 Teknikokullar, Ankara, Turkey
bDepartment of Biophysics, Faculty of Medicine, Fırat University, Elazıg, Turkey

Abstract

In this study, Doppler signals, recorded from the output of aorta valve of 80 patients, were transferred to personal computer via 16 bit sound card. The fast Fourier transform (FFT) method was applied to the recorded signal from each patient. Since FFT method inherently cannot offer a good spectral resolution at highly turbulent blood flows, it sometimes causes wrong interpretation of cardiac Doppler signals. In order to avoid this problem, two known diseased heart signals such as aorta stenosis and aorta insufficiency were introduced to two different genetic fuzzy systems. The disadvantages arise from these two different genetic fuzzy systems were eliminated by using the new genetic fuzzy system which is proposed in this study. The proposed genetic fuzzy system is called as variable coded-hierarchical genetic fuzzy system. As a result, it is shown that the proposed system decreases the computational time since it uses less genes.

Keywords: Fast Fourier transform method; Cardiac doppler; Aorta stenosis; Aorta insufficiency; Genetic fuzzy system; Hierarchical genetic fuzzy system; Variable coded-hierarchical genetic fuzzy system

1. Introduction

Ultrasonic Doppler blood flow measurement systems are widely used and highly successful non-invasive technique to detect heart diseases. Since the spectral analysis methods have become available; computing facilities provide the opportunity to extract parameters from the spectrograms with more sophistication (Evans, McDicken, Skidmore, & Woodcock, 1989).

Doppler signals obtained from heart valves have been analyzed using different spectral analysis methods such as fast Fourier transform (FFT), autoregressive, moving average, autoregressive-moving average and wavelet methods (Güler, Hardalaç, & Barışç, 2002; Güler, Hardalaç, & Müldür, 2001; Güler & Kara, 1995; Güler, Kara, Güler, & Kıymı, 1996; Serhatlıoglu, Hardalaç, & Güler, 2003). The output of a spectral analyzer is usually represented as a sonogram which illustrates the time variation of the spectral characteristics of the Doppler signal. A number of parameters related to the blood flow may be extracted from the sonogram and these are of high clinical value. In our previous study, the spectral analysis of stenosis heart valves were investigated using fuzzy algorithm (Güler et al., 2002).

Fuzzy systems are particularly suitable for modeling and classification of problems in form of linguistic variables and rules. Although fuzzy systems have been successfully applied in a large number of applications, they lack the ability to extract knowledge from a set of training data. Therefore, over the past years more researches have been devoted to augment the approximate reasoning method of fuzzy systems with the learning capabilities of neural networks and evolutionary algorithms. Over the past decade, there has been an increasing interest in genetic algorithms (GA) that adapt the knowledge base of a fuzzy system. These approaches are described by the general term genetic fuzzy rule-based systems (Cordon, Herrera, Hoffmann, & Magdalena, 2001). The role of the GA is to either tune the parameters of a Fuzzy Rule-Based Classifier System (FRBCS) or to completely automate the fuzzy knowledge base design (Cordon, del Jesus, & Herrera, 1998; Ishibuchi, Nozaki, Yamamoto, & Tanaka, 1995; Pena Reyes & Sipper, 1999).

Several researchers have used GA in the design of fuzzy membership functions and/or fuzzy rules previously for different areas (Chang & Wu, 1995; Karr, 1991;
Lee & Takagi, 1993; Pena Reyes & Sipper, 1999; Wu & Liu, 2000). There are several recent examples of the application of fuzzy systems and GA in the medical domain (Pena Reyes & Sipper, 2000), though only a few combine both methodologies in a hybrid way (Pena Reyes & Sipper, 1999)—as we do in this work. Up to now, there is no any comparative work on the application of Genetic Fuzzy System (GFS) to cardiac Doppler signal.

In this study, blood flow velocity variables have been extracted by applying FFT analysis to Doppler signals obtained from the aortas of 80 patients. In order to identify blood flow velocities of patients more accurately and faster, blood flow velocities of patients were applied to different GFS. Thus, a FRBCS is constructed to diagnose the aorta diseases.

After examining different genetic fuzzy systems such as GFS and Hierarchical Genetic Fuzzy System (HGFS), the Variable Coded-Hierarchical Genetic Fuzzy System (VC-HGFS) was proposed in order to eliminate the negative aspects of GFS and HGFS methods.

2. Material and methods

2.1. Feature extraction of doppler signals

Measurement system consists of five functional blocks as shown in Fig. 1. These are 2.50 and 3.75 MHz ultrasound transducers, an analog Doppler unit (Toshiba Sonolayer 140A-Echo equipment), Sony recorder, an analog/digital interface board (Sound Blaster Pro-16 bit), and a personal computer. The analog Doppler unit is capable of operating in both continuous and pulse wave modes. The Doppler unit is also equipped with an imaging facility that makes it possible to focus the sample volume to a desired location within the cardiac chambers. In this study, the Doppler unit is operated in the pulse mode and the sample volume is located at the aorta valve orifice. The measurement angle is taken as 60°.

The signal at the output of the analog Doppler unit is recorded by a Sony recorder. The recorded signal is then sampled and digitized into 16 bit data packets using an A/D interface board. The digital data are then stored on the hard disk of the PC. The interface board offers a broad range of sampling frequencies. The data stored as a sound file on the hard disk of the PC are converted to a text file by using a simple algorithm implemented in C programming language.

Spectral curves are then obtained from this text file using spectral analysis software developed in MATLAB. In this study, measurements were performed at the sampling frequency of 20.48 kHz, so that the frequency aliasing, in the case of stenosis, would be avoided. The frame length used in this study is 128. Systole and diastole values of Doppler signals are taken from the spectral curve that is formed according to the FFT analysis. These values are used as inputs to genetic fuzzy systems employed in this study.

2.2. Fuzzy rule-based classification system

A fuzzy approach is of particular value in domains for which it is important for a human expert to comprehend the classifier decision such as medical diagnosis and safety critical applications. Each fuzzy rule covers a particular region of the attribute space described by the rule antecedent, for which it proposes the classification specified in the rule consequent. Assume a training set of \( K \) instances \( T = \{(x^1, c^1),\ldots,(x^K, c^K)\} \) where \( x^k = \{x^k_1,\ldots,x^k_n\} \) is an instance taken from some attribute space \( \{X_1,\ldots,X_n\} \), and \( c^k \in \{C_1,\ldots,C_m\} \) is the class label associated with \( x^k \). We use upper indices \( k \) to denote the \( k \)th training examples, and lower indices \( n \) to denote the \( n \)th attribute \( x^k_n \) of a training example \( x^k \). Fuzzy rules are of the form

\[
R_i : \text{if } X_1 \text{ is } A_{i1} \text{ and } \ldots \text{ and } X_N \text{ is } A_{iN} \text{ then } Y = c_i M_i,
\]

in which \( X_n \) denotes the \( n \)th input variable, \( A_m \) the fuzzy set associated to \( X_n \) and \( c_i \in \{C_1,\ldots,C_m\} \) represents the class label of the rule. For a particular instance \( x^k = \{x^k_1,\ldots,x^k_N\} \), the rule activation

\[
\mu_{R_i}(x^k) = \min_{n=1}^{N} \mu_{A_{i_n}}(x^k_n),
\]

(1)

describes to what degree the rule matches the instance. For each possible classification \( C_m \) the degree of activation of fuzzy rules with a matching consequent \( c_i = C_m \) is aggregated. The instance \( x^k \) is classified by the class label

\[
C_{\text{max}}(x^k) = \arg \max_{R_i/C_m} \sum_{R_i/C_m} \mu_{R_i}(x^k).
\]

(2)


Fig. 1. Block diagram of measurement system.
2.3. Genetic fuzzy systems

GA are search and effective optimization techniques based on a formalization of natural genetics (Goldberg, 1989; Holland, 1992; Michalewicz, 1992). They are characterized by a parallel search of the state space as against a point-by-point search by the conventional optimization techniques. The parallel search is achieved by keeping a set of possible solutions to the optimization problem, called population. An individual in the population is a string of symbols and is an abstract representation of the solution. The symbols are called genes and each string of genes is termed as a chromosome.

In this study, fuzzy algorithm is chosen as the decision-making logic. In the design of a FRBCS, the membership functions are chosen at first. Personal experience is used for defining the output membership function during the fuzzification process. But, personal experience may lead some mistakes in diagnosis. On the other hand, the membership function may require reorganization in the variable systems. Having now learnt the complicated procedures of designing FRBCS, a practical realization of this system is not easy to determine. Manually operating procedures for these variables might not only yield a sub-optimal performance, but could also be dangerous if the complete fuzzy sets were wrongly augmented. For this reason, membership functions and rules are determined by using GA (also called GA based design process; see Fig. 2). So, fuzzy classification is constructed by membership functions. Accurate and quick diagnosis can be achieved by doing inference process from rule base (Cordon et al., 2001).

From the above description, it can be stated that the main objective of this paper is to design the membership functions and the fuzzy rules simultaneously based on a chosen criterion. Obviously, this is not an easy task since many parameters must be determined at the same time. This, in turn, explains why GAs are adopted since they provide a means to search poorly understood, highly complex spaces. However, before going further, several issues on GAs must be discussed first.

2.3.1. Chromosome representations

In this paper, since the parameters to be determined are all real, a real number representation is used, in which each chromosome vector is coded as a vector of real numbers of the same and variable lengths, respectively, as the solution vector. When considering two input and one output FRBCS, from the illustration in Fig. 3, one can find that there are $(3n_1 - 3)$, $(3n_2 - 3)$ and $(3n_3 - 3)$ parameters to be determined for membership functions of $A_{11}, A_{12}, \ldots, A_{1i}, A_{21}, A_{22}, \ldots, A_{2j}$, and $C_1, C_2, \ldots, C_m$, respectively. Moreover,
to determine fuzzy rules \( R_{ij} \), \( 1 \leq i \leq n_1, 1 \leq j \leq n_2, (n_1, n_2) \) parameters will be needed. Therefore, one can find easily that a chromosome will contain \((3n_1 + 3n_2 + 3n_o - 9)\) genes (Fig. 4). The first \((3n_1 - 3)\) genes, the next \((3n_2 - 3)\) genes, the next \((3n_o - 3)\) genes and the last \((n_1, n_2)\) genes will be used to determine the input membership functions of \(A_{11}, A_{12}, \ldots, A_{1n_1} \), the input membership functions of \(A_{21}, A_{22}, \ldots, A_{2n_2} \), the output membership functions of \(A_{31}, A_{32}, \ldots, A_{3n_o} \), and the rules \( R_{ij}, 1 \leq i \leq n_1, 1 \leq j \leq n_2, \) respectively (Man, Tang, & Kwong, 1999).

In determining the range for each gene of a chromosome

\[
x = [x_1, x_2, \ldots, x_{3n_1+3n_2+3n_o+n_1n_2-9}],
\]

(3)

since \( R_{ij} \in \{R_1, R_2, \ldots, R_{n_1}\} \) for \( 1 \leq i \leq n_1, 1 \leq j \leq n_2, \) (4)

one can find that \( x_{3n_1+3n_2+3n_o-8}, x_{3n_1+3n_2+3n_o-7}, \ldots, x_{3n_1+3n_2+3n_o+n_1n_2-9} \) are all within the range \([1, n_3]\). The values of 1 and \( n_3 \) will correspond to \( R_1 \) and \( R_2 \), respectively. For \( x_1, x_2, \ldots, x_{3n_1+3n_2+3n_o-9} \), however, their corresponding ranges are not easy to be determined since the following three equations must be satisfied at every time instant as shown in Fig. 3

\[
\sum_{i=1}^{3n_1-3} x_i = a_{\text{max}} - a_{\text{min}},
\]

(5)

\[
\sum_{i=3n_1+3n_2-6}^{3n_1+3n_2-6} x_i = b_{\text{max}} - b_{\text{min}},
\]

(6)

\[
\sum_{i=3n_1+3n_2+3n_o-9}^{3n_1+3n_2+3n_o-9} x_i = c_{\text{max}} - c_{\text{min}},
\]

(7)

where \([a_{\text{min}}, a_{\text{max}}], [b_{\text{min}}, b_{\text{max}}]\) and \([c_{\text{min}}, c_{\text{max}}]\) correspond to \([x_{\text{min}}, x_{\text{max}}]\) in Fig. 3, and are the actual physical domains over which the input linguistic variables \(a, b\) and \(c\) take their crisp values, respectively.

Since \( x_1, x_2, \ldots, x_{3n_1+3n_2+3n_o-9} \) are real numbers to be adjusted by GAs, their values will change from one generation to another in the evolutionary process. In this case, it is very possible that the constraints in Eqs. (5)–(7) will be violated. Therefore, a proportional scaling method will be used to solve this problem, in which the values of the \((3n_1 + 3n_2 + 3n_o - 9)\) genes are determined according to their proportions, not from the values themselves. This means the value of each gene is determined as follows:

\[
x_i' = (a_{\text{max}} - a_{\text{min}}) \frac{x_i}{\sum_{j=1}^{3n_1-3} x_j} \quad \text{for } 1 \leq i \leq 3n_1 - 3,
\]

(8)

\[
x_i' = (b_{\text{max}} - b_{\text{min}}) \frac{x_i}{\sum_{j=3n_1+3n_2-6}^{3n_1+3n_2-6} x_j} \quad \text{for } 3n_1 - 2 \leq i \leq 3n_1 + 3n_2 - 6,
\]

(9)

\[
x_i' = (c_{\text{max}} - c_{\text{min}}) \frac{x_i}{\sum_{j=3n_1+3n_2+3n_o-9}^{3n_1+3n_2+3n_o-9} x_j} \quad \text{for } 3n_1 + 3n_2 - 5 \leq i \leq 3n_1 + 3n_2 + 3n_o - 9.
\]

In this manner, since the constraints in Eqs. (5)–(7) will always be satisfied, the ranges of \( x_1, x_2, \ldots, x_{3n_1+3n_2+3n_o-6} \) can be chosen almost arbitrarily. However, if only two neighboring membership functions are assumed to have intersection as shown in Fig. 3, then the lower bounds of the ranges must be chosen to be greater than zero.

The systole and diastole values are used as input variables to FRBCS. The output variable is divided into three as normal, aorta insufficiency and aorta stenosis. If the number of the fuzzy sets in input variables is represented as \( n_1 \) (systole) and \( n_2 \) (diastole), and the number of the fuzzy sets in output variable is represented as \( n_o \), the required fuzzy system can be coded in a chromosome structure as in Fig. 5 by using

If : \( n_1 = n_2 = 7 \) and \( n_o = 3 \)

\[
3n_1 - 3 + 3n_2 - 3 + 3n_o - 3 + n_1 n_2
\]

\[
= 3n_1 + 3n_2 + 3n_o - 9 + n_1 n_2 = 91 \text{ genes}.
\]

2.3.2. Evaluation and selection

GA starts with a random set of population. An evaluation operator is then applied to evaluate the fitness of each
individual. Fitness values of each individual in a population are calculated via fitness function. These fitness values explain the affinity of the desired solution. The FRBCS output values of diagnosed diseases are applied to all individuals by GA, so that the error in the diagnosis process is determined. The fitness function is obtained by summing of the squared values of errors as in Eq. (11)

\[
\text{fitness} = \sum_{i=1}^{n} (y_i - y_0)^2,
\]

where \(y_i\) is the real diagnose, \(y_0\) is the FRBCS output values, and \(n\) is the patient number.

A selection operator is then applied to select the population members with higher fitness (so that they can be assigned higher probability for survival). Several selection operators are reported in the literature; the operators are proportionate reproduction, ranking selection, tournament selection, and steady state selection. Among the popular selection operators are ranking and tournament selection (Goldberg & Deb, 1991). Since the tournament selection operator requires lower computational overhead, we use tournament selection operator.

2.3.3. Genetic operators

Information exchange of the individuals is called crossover and is a basic property of GAs those produce a new chromosome from two parents. Crossover operation from a single point is defined as follows for two real-coded chromosomes \(x_1\) and \(x_2\)

\[
\begin{align*}
\text{crossover point:} & \quad \lambda \\
\text{offspring:} & \quad x'_i = (1 - \lambda)x_i + \lambda x'_i,
\end{align*}
\]

where \(\lambda \in (0, 1)\). Furthermore, the crossover operation is made using more than one crossover points (\(\lambda\)).

Mutation means a random change in the information of a chromosome that does not depend on a reason. A probability test determines whether a mutation will be carried out or not. During the use of mutation operator, one should be careful about the chromosome structure and constraints in Section 2.3.1.

2.4. Hierarchical genetic algorithm (HGA)

To generalize this architecture, a level of control genes are introduced in a hierarchical fashion as shown in Fig. 5. In this case, the activation of the parametric gene is governed by the value of the control gene. To indicate the activation of the control gene, an integer ‘1’ is assigned for each control gene that is being ignited where ‘0’ is for turning off. When ‘1’ is signaled, the associated parameter genes due to that particular active control gene are activated in the lower level structure. It should be noticed that the inactive genes always exist within the chromosome even when ‘0’ appears. This hierarchical architecture implies that the chromosome contains more information than that of the conventional GA structure. Hence, it is called Hierarchical Genetic Algorithm. Solutions of the divergent problems can be found by coding control genes with more than one level (Man et al., 1999).

The use of the HGA is particularly important for the structure or topology as well as the parametric optimization. Unlike the set-up of the conventional GA optimization, where the chromosome and the phenotype structure are assumed to be fixed or pre-defined, HGA operates without these constrains. The following application has been developed in accordance with the foundations set above.

2.5. Hierarchical genetic fuzzy system

Considering that the main attribute of the HGA is its ability to solve the topological structure of an unknown system, then the problem of determining the fuzzy membership functions and rules could also fail into this category. This approach has a number of advantages:

- an optimal and the least number of membership functions and rules are obtained;
- no pre-fixed fuzzy structure is necessary;
- simpler implementing procedures and less cost are involved (Man et al., 1999).

Hence, it is the purpose of this section to introduce the HGA for the designing of FRBCS. Similar to other uses of the HGA, the hierarchical chromosome for the FRBCS structure must be correctly formulated. In this case, the chromosome of a particular fuzzy set is shown in Fig. 6. The chromosome consists of the usual two types of genes, the control genes and parameter genes. The control genes, in the form of bits, determine the membership function activation, whereas the parameter genes are in the form of real numbers to represent the membership functions.

Before evaluating the fitness value of the FRBCS, their phenotype must be obtained. In some cases, a direct decoding of the membership chromosome may result in invalid membership functions. For example, Fig. 6 represents an invalid membership function for error fuzzy set because the ranges \((\alpha_2, \alpha_{32})\) and \((\alpha_3, \alpha_{32})\) are unclassified (only the error set is shown for clarity).

To ensure that there was no undefined region, a remedial procedure was operated to ensure validation. The decoded fuzzy membership functions were recovered as shown by the final membership characteristics in Fig. 6. It should be noted that the parameter gene remained unaltered but merely changed the interpretation of its form. In this way, the complexity of tuning the fuzzy memberships and rules can thus be optimized and the overall structure can be greatly reduced.

When the chromosome structure of HGFS is examined, \(n_1 + (3n_1 - 3)\) genes should be used in the coding of fuzzy sets in input variables as in Fig. 6. In this study, because the number of the fuzzy sets in output variables is fixed, \(3n_o - 3\)
genes can be used in the coding of the fuzzy sets in output variables. Consequently, when the chromosome structure in Section 2.3.1 is to be applied to HGFS, and when it is required to set up a FRBCS which have two inputs and one output, we need a chromosome structure as

\[ n_1 (3n_1 - 3) + n_2 (3n_2 - 3) + 3n_o - 3 + n_1 \cdot n_2 = 4n_1 + 4n_2 + 3n_o + n_1 \cdot n_2 - 9. \]

According to GFS system in Section 2.3., we should use \( n_1 + n_2 \) additional genes to form the chromosome structure. Extension of the chromosome structure will lead to increasing of the operation load.

In GFS, the number of the fuzzy sets in input variables should be predetermined. Thus, the number of the fuzzy sets to be used in the FRBCS will not be determined as optimum. In HGFS, coding in GA will optimize the number of the fuzzy sets in the FRBCS. Thus, even though the chromosome structure is longer, we will have a more optimum FRBCS. Since both GFS and HGFS have long chromosome structures there will be more operation load while obtaining FRBCS. Using the variable chromosome structures, the VC-HGFS is proposed so that less gene with less processing time will be needed.

### 3. Variable coded-hierarchical genetic fuzzy system

In the proposed system, we aim to use fewer genes by means of variable length code structure. On the other hand, if we would use the fixed code structure in HGFS, in this case, we would use more than fewer genes. Rather than coding fuzzy sets in the input variables in the FRBCS one by one as in Fig. 5, the number of the fuzzy sets will be expressed numerically. As in Fig. 7, the number of the fuzzy sets in each input variable is coded by means of only one gene. The number of the fuzzy sets to be used in the chromosome structure is determined by only control gene in Fig. 7. On the other hand, the number of the fuzzy sets to be used in chromosome structure is determined by using variable length code structure. We will use the model in Fig. 7. Furthermore, rules forming the third part of the chromosome structure are coded according to the control genes instead of maximum number of rules. Thus, when
the FRBCS is optimized, fewer genes will be used and the operation load will be lower, therefore we will reach the result in a shorter time than HGFS system.

3.1. Chromosome structure

The algorithm structure will be same, and when forming initial population in the applying stage, the coding is variable depending on the control genes $n_1$ and $n_2$ as in Fig. 7. In order to define the chromosome structure of the FRBCS that has two inputs and one output, the number of genes to be used is computed as follows:

$$1 + 3n_1 - 3 + 1 + 3n_2 - 3 + 3n_o - 3 + n_1n_2$$

$$= 3n_1 + 3n_2 + 3n_o - 7 + n_1n_2.$$  

3.2. Genetic operators

As in Fig. 8, according to the control genes including the numerical values $n_1$ and $n_2$, the gene information defining the fuzzy sets and rules in each individual is different. One should be careful using the crossover operator because of the variable length structure in VC-HGFS. Since the code structure in GFS and HGFS is fixed, crossover operation is adequate. However, the variable chromosome structure in HGFS does not allow making crossover from one point. Because according to control genes (including the numerical values $n_1$ and $n_2$) of two different individuals chosen from the crossover pool, each different individual’s gene information defining the fuzzy sets and rules is different. After determining the crossover point, different crossover strategies are used in line with the components in the chromosome structure:

(a) Genes defining the fuzzy sets in input variables:

- If the condition $(x_1(n_1) = x_2(n_1), x_1(n_2) = x_2(n_2))$ is provided, that is, the number of the fuzzy sets of two different individuals is same, crossover from one point is possible. Here, $x_1(n_1)$ represents the first control gene (the number of the fuzzy sets in the first input variable) in the first individual to be crossed over. $x_2(n_1)$ represents the first control gene (the number of the fuzzy sets in the first input variable) in the second individual to be crossed over. $x_1(n_2)$ represents the second control gene (the number of the fuzzy sets in the second input variable) in the first individual to be crossed over. $x_2(n_2)$ represents the second control gene (the number of the fuzzy sets in the second input variable) in the second individual to be crossed over.

(b) Genes defining the fuzzy sets in output variables:

As in Fig. 9, only the genes defining the fuzzy sets in output variables can be crossed over between themselves. The rules should not be crossed over.

(c) Rules:

Because the number and the arrangement of the rules depends on the number of the fuzzy sets in input variables,
there can be four possible situations depending on the number of the fuzzy sets in two individuals to be crossed over:

- If the number of the fuzzy sets of two different individuals is same, \((x_1(n_1) = x_2(n_1), x_1(n_2) = x_2(n_2))\), genes defining the rules can be crossed over from any points desired. Because the defined rule base in both individuals is same.
- If the number of the first fuzzy sets of two different individuals is same, \((x_1(n_1) = x_2(n_1), x_1(n_2) \neq x_2(n_2))\); since the number of rules of both individuals is different,
more than one crossover points should be determined, and the crossover operation should be done without destroying the rule base to be defined in each individual (Fig. 10). The fuzzy sets in input variables should also be to crossed over.

- If the number of the second fuzzy sets of two different individuals is same \( (x_1(n_1) \neq x_2(n_1), x_1(n_2) = x_2(n_2)) \) as in Fig. 11, to make the crossover operation from one crossover point in gene structure defining the rules is adequate. The fuzzy sets of input membership functions should also be crossover.

- If the number of the both fuzzy sets of both individuals is different, \( (x_1(n_1) \neq x_2(n_1), x_1(n_2) \neq x_2(n_2)) \) it is impossible to crossover the rule base of both individuals. Because the numbers of both row and column of matrixes formed by rule base will not be agreed, it is impossible that the new individuals to be formed after crossover define a FRBCS.

Three different genetic fuzzy systems examined in this study are compared in Table 1. During the genetic optimization operation, GFS accepts the fixed number of fuzzy sets. Therefore the number of fuzzy sets cannot be optimized for the FRBCS to be formed. This problem is solved in HGFS, and because fixed chromosome structure is used, for example, even though the number of fuzzy sets in optimum input variables is 3 and 4 as in Table 1, the number of genes used is 105. This leads to increasing of operation load in proportion to GFS, and the optimization operation takes more time. With the method of VC-HGFS, this problem is solved. The FRBCS can be coded by using 105 genes in HGFS method which uses variable length code structure, however, FRBCS can be coded by using 35 genes with VC-HGFS method.

### Table 1
Comparison of different genetic fuzzy systems

<table>
<thead>
<tr>
<th>System</th>
<th>Code structure</th>
<th>Fuzzy sets</th>
<th>Number of fuzzy sets</th>
<th>Number of genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFS</td>
<td>(3n_1 + 3n_2 + 3n_x + n_1n_2 - 9)</td>
<td>Fixed</td>
<td>7–7</td>
<td>91</td>
</tr>
<tr>
<td>HGFS</td>
<td>(4n_1 + 4n_2 + 3n_x + n_1n_2 - 9)</td>
<td>Variable</td>
<td>3–4</td>
<td>105</td>
</tr>
<tr>
<td>VC-HGFS</td>
<td>(3n_1 + 3n_2 + 3n_x + n_1n_2 - 7)</td>
<td>Variable</td>
<td>3–4</td>
<td>35</td>
</tr>
</tbody>
</table>

### 4. Results and discussions

FFT analysis was applied to recorded cardiac Doppler signals. The spectral curves are constructed as shown in Fig. 12. These results are obtained by taking 120,000 samples. Each spectral curve is constructed as off-line in 6 s. Sixty patients out of total of 80 were used in training purposes. Those 60 patients are categorized as follows: 20 of healthy subjects, 20 of aorta stenosis and 20 of aorta failure. The patients were diagnosed by the experts in cardiology department prior to this study.
Fig. 12. Blood flow velocity waveforms showing: (a) Normal, (b) aorta insufficiency, (c) aorta stenosis situations.

When the spectral curves given in Fig. 12 are examined, it is seen that the first peak (systole) occurs at just the mitral valve opens by forcing the pressure of left atrium. Since the blood flows from left atrium to left ventricle, the pressure of left atrium decreases so that the blood velocity decreases. This causes a decrease on the amplitude of spectral curve. This point is formed as a valley between two peaks. After all these events, the M-shaped spectral curve is obtained as shown in Fig. 12a. When the slope between systole and diastole is small on the envelope of spectral curve of patients, this means degeneration on the envelope of M-shaped spectral curve as shown in Fig. 12b,c. The mean value and standard deviation of systole and diastole, which forms the spectral curve and student-t test results, were grouped statistically in Table 2.

These spectral curves are then used for constructing a rule base, which involves systole and diastole situations. In order to define the positions of fuzzy sets of input and output variables, the solution space should be defined by GFS. These situations are grouped in the range of 80–650 cm/s for systole, 40–292 cm/s for diastole and 0–100 for output, then they are fuzzified as the fuzzy sets. The variables that are constituted by randomly generated individuals in initial population should be defined in the desired solution space.

For this purpose, the value of each variable is determined by Section 2.3.1.

The process of fitness computation, selection, crossover and mutation is continued for a fixed number of iterations is achieved. In order to arrive at safe conclusions, for each one of the above methods we performed 20 independent experiments (genetic runs). As in Fig. 13, the genetic fuzzy systems are evaluated on the basis of a normalized fitness value \( f \in [0,1] \) for 700 iteration.

As a result of several runs with different values, parameters were chosen as:

- Population size: 120
- Crossover probability \( (P_c): 80\% \)
- Mutation probability \( (P_m): 1\% \)

Methods used in this study have been examined using different maximum iteration numbers (300, 500, 700). Thus when the iteration numbers increase, it is observed how the results of the methods differ.

Several FRBCS were constructed using GFS, HGFS and VC-HGFS methods. Fig. 14 and Table 3 show fuzzy sets and rule bases that are components of the FRBCS constructed using GFS and VC-HGFS methods. Since the FRBCS which were constructed using HGFS and VC-HGFS methods are similar, the FRBCS constructed using HGFS method is not shown here. As there are differences between the operation load and time of two methods, these methods are compared in Tables 5 and 6.

In Table 3, \( C_0 \) represents normal, \( C_1 \) represents aorta failure, and \( C_2 \) represents aorta stenosis subject. During the application of each system examined in this study, input variables are represented as \( S_1,S_2,...,S_7 \) (systole); \( D_1,D_2,...,D_7 \) (diastole). They are grouped as seven sets. Input variables are used less than seven sets in HGFS and VC-HGFS systems were optimized as in Fig. 14.

There are two input variables (systole and diastole) which have seven sets in GFS, HGFS and VC-HGFS methods. Since the FRBCS which were constructed using HGFS and VC-HGFS methods are similar, the FRBCS constructed using HGFS method is not shown here. As there are differences between the operation load and time of two methods, these methods are compared in Tables 5 and 6.

In Table 3, \( C_0 \) represents normal, \( C_1 \) represents aorta failure, and \( C_2 \) represents aorta stenosis subject. During the application of each system examined in this study, input variables are represented as \( S_1,S_2,...,S_7 \) (systole); \( D_1,D_2,...,D_7 \) (diastole). They are grouped as seven sets. Input variables are used less than seven sets in HGFS and VC-HGFS systems were optimized as in Fig. 14.

There are two input variables (systole and diastole) which have seven sets in GFS, thus the rule table consisting of 7 * 7 = 49 rules is shown in Table 3a, and the decreased rule table consisting of 3 * 4 = 12 rules in VC-HGFS is shown in Table 3b. The first six rules in Table 3b are defined as follows:

- **Rule 1.** If Systole is \( S_1 \) and Diastole is \( D_1 \) then Class \( C_0 = \) normal,
- **Rule 2.** If Systole is \( S_1 \) and Diastole is \( D_2 \) then Class \( C_0 = \) normal,
- **Rule 3.** If Systole is \( S_1 \) and Diastole is \( D_3 \) then Class \( C_1 = \) aorta failure,
- **Rule 4.** If Systole is \( S_2 \) and Diastole is \( D_1 \) then Class \( C_1 = \) aorta failure,
- **Rule 5.** If Systole is \( S_2 \) and Diastole is \( D_2 \) then Class \( C_0 = \) normal,
- **Rule 6.** If Systole is \( S_2 \) and Diastole is \( D_3 \) then Class \( C_2 = \) aorta stenosis,
Upon examining data in Table 4, input fuzzy sets decreased 52.2–56.0%, and rules decreased 72.9% as the result of using the proposed VC-HGFS method.

As seen in Table 5, the gene numbers of individuals decreased 67.4–70.1% because of the proposed VC-HGFS method in this study. Thus, the operation load decreases during the genetic optimization. Since the operation load decreases, as it is shown in Table 6, VC-HGFS method yields a faster result at the ratio of 8.4–11% than HGFS method does. These results were obtained for a Pentium IV (Intel) with clock rate of 1.7 GHz and 512 MB of main memory.

Upon examining the results, VC-HGFS method that we proposed are more successful in terms of gene number and time. We aimed at measuring the classification success of the system by applying test data to FRBCS.

The first five test data classified by FRBCS are shown in Table 7. These test data were applied to FRBCS that was constructed using VC-HGFS method in Fig. 14b and
Table 3b. First patient (patient no. 1) has 167 cm/s blood velocity in systole belonging to $S_1$ set and 102 cm/s blood velocity in diastole belonging to $D_1$ set. These results emphasized that $C_0$ set in rule base is the active output, which represent normal. Second patient (patient no. 2) has 317 cm/s blood velocity in systole belonging to $S_2$ set and 68 cm/s blood velocity in diastole belonging to $D_1$ set. These results emphasized that $C_1$ set in rule base is the active output, which represent aorta failure. Fifth patient (patient no. 5) has 589 cm/s blood velocity in systole belonging to $S_4$ set and 201 cm/s blood velocity in diastole belonging to $D_3$ set. These results emphasized that $C_2$ set in rule base is the active output, which represent aorta stenosis. Similarly all the other data are tested.

Twenty patients among a group of 80 are investigated in the dedicated GFS for testing purposes. When the classification results are compared, as it is seen in Table 8, the success rate is 91.2–97.5%. As the iteration number increases, a more successful FRBCS is constructed in terms of classification.

5. Conclusion

In this study, it is seen that a faster and more accurate diagnosis is obtained classifying FFT analysis results of Doppler signals by genetic fuzzy systems. Therefore, it is aimed at eliminating the mistakes during the interpretation of FFT signals using an expert system.

In our previous study, FFT signals were classified by using only FRBCS. However, FRBCS, which was constructed for this aim, is developed through personal
experience. Thus, some difficulties and mistakes occur during the development of FRBCS. In this study, we aimed at designing FRBCS with an optimization operation by using GA. However, an important problem occurs in this stage. The number of fuzzy sets that are to be used for input variables should be determined by predicting the size of FRBCS that is to be designed, and rule base should be formed in compliance with these input variables. Therefore, determination of the number of fuzzy sets through personal experience impedes designing of an optimum FRBCS.

When HGFS method is used in order to overcome this problem, it is observed that too much genes were used in the structure of chromosome. Although a more optimum FRBCS has been obtained by using this method, the operation load increases and the process of optimization extends. Therefore the method of VC-HGFS was used in this study. By using variable length code structure and developing a crossover operation, a faster and more optimum FRBCS than other genetic fuzzy systems can be produced.

References


