

# Development of the Structure of the Knowledge Base for Neuro-Fuzzy Diagnostic System

A.V. Nikonov<sup>1</sup>, A.M. Vulfin<sup>1</sup>, M.M. Gayanova<sup>1</sup>, M.U. Sapozhnikova<sup>1</sup>

<sup>1</sup>Ufa State Aviation Technical University, K. Marks st. 12, Ufa, Russia, 450077

**Abstract.** Cardiovascular diseases are one of the leading causes of death worldwide. People suffering from or at high risk of such diseases need constant supervision, early diagnosis and timely assistance. It is shown that the achievement of high accuracy performance in real-time arrhythmias recognition is associated with significant hardware costs. Detection accuracy of arrhythmias recognition does not exceed 80%. An approach, which is offered to solve the problem of high-precision arrhythmia diagnosis on the basis of electrocardiosignal is based on the data mining methods. Application of such methods is necessary for processing large amounts of data with complex structure of the features. Determination of the arrhythmia type with the use of fuzzy inference tools needs to specify the technique of the original data preprocessing. Feature selection, formalization and coding is considered in this paper. The issue of the knowledge base construction – coding, generation and selection of the features (database) as well as the construction of the rules base – as the part of the neuro-fuzzy diagnostic system is also considered. The research goal is to improve the intelligent systems of arrhythmia diagnostics on the basis of neural network classifiers by developing a solution explanation subsystem based on neuro-fuzzy models.

**Keywords:** data mining, neural network, fuzzy logic, electrocardiosignal, arrhythmia.

## 1. Introduction

Cardiac arrhythmia as the most common disease is an irregular heart rhythm caused by improper operation of electrical impulses that regulate the heartbeat [1,2].

Arrhythmia can be detected by cardiac specialist using electrocardiogram (ECG) strip. Even for skillful cardiologist it may take several minutes to make a diagnosis; in some severe cases, this could be fatal for the patient. ECG analysis is too hard for the beginner or inexperienced staff [3,4].

The electrocardiosignal (ECS) is an electrical manifestation of the heart contractile activity and it characterizes the state of the cardiovascular system (CVS). Electrocardiosignal on the ECG looks like a set of alternating waves, flat intervals and segments located on the baseline. Each element contains information of the state of the heart and its components [2]. Parameters of the cardiac contraction pattern and individual features of the ECS are considered in [36].

Intelligent analysis of ECS data using neural network nonlinear models shows the highest results of recognition accuracy on model tasks [7,14-18]. Nevertheless, the distinctive feature of intelligent diagnostic systems based on the neural network technologies is the complexity of explaining the decision made. The reason is that signal processing algorithms and intermediate calculation results of

neural network models are uninterpreted by both the medical technique and the knowledge engineer [2,19].

Among the many models of expert knowledge representation, the production rules are widely accepted nowadays due to their flexibility and simplicity of describing the laws of the subject domain. With the help of products it is easier to model the thinking activity of an expert, his reasoning, conclusions, decision-making process [37].

The problem of diagnosing cardiac arrhythmias have to be solved in conditions of indistinctness, incompleteness and unreliability of the initial data. It is impossible to solve such problem using ordinary production rules. It is necessary to use fuzzy products and systems of fuzzy logic inference for this task.

Fuzzy neural networks are hybrid models that combine the main advantages of neural networks (possibility of adaptive learning) and fuzzy systems (interpretability of the result obtained with their help). The fuzzy neural network is analogous to the system of fuzzy logic inference. The system of fuzzy logic inference is able to make decisions only on the basis of a ready (expert-generated) knowledge base, while a fuzzy neural network as a result of training itself forms a system of rules.

Therefore, it is relevant to choose a flexible fuzzy-productive model of knowledge representation in the problem of arrhythmia diagnosis. Neuro-fuzzy system based on a fuzzy neural network allows, as a result of training, to form a knowledge base of the expert system like a set of rules in the form of fuzzy products.

The research goal is to improve the intelligent systems of arrhythmia diagnostics on the basis of neural network classifiers by developing a solution explanation subsystem based on neuro-fuzzy models.

To achieve the goal, the following tasks were set:

- Development of the structure of an intelligent system for diagnosing arrhythmias with a subsystem of explaining the solution based on neuro-fuzzy models;
- Construction of the knowledge base as well as construction of the rules base;
- Development of the solution explanation subsystem model.

## 2. Analysis of automatic arrhythmia diagnosis methods

Nowadays there are several methods of automatic arrhythmia diagnosis. Figure 2 shows proposed classification of methods currently used in automatic arrhythmia diagnosis systems [2-5,9,17,23,25,30]. General groups of methods are usually distinguished as linear, nonlinear and methods based on data mining (DM) [2, 10-13].

The introduction and widening application of automatic external defibrillators (AEDs) make strong demands for ECS analysis. It is required highly accurate discrimination between shockable and non-shockable rhythms, with sensitivity and specificity aimed to achieve the maximum values [23]. Also, a false positive detection will initiate a defibrillator to give improper therapeutic intervention [24,25].

Modern diagnostic systems are a compromise between algorithms that require significant computational costs to achieve high accuracy of diagnostic, and hardware with limited performance. It has been established that the accuracy of arrhythmia recognition in existing solutions does not often exceed 80% [2,5,10-13,17,20,25].

Other problems that common systems of arrhythmia diagnosis meet are the following:

### 1) The problem of analyzing large amounts of ECG data

As some arrhythmias appear irregularly, it is necessary recording ECG activity using, e.g., Holter monitor up to a week to successfully capture them [3]. The total number of such cardiac abnormalities reach up to 96 different categories [7]. Each category may contain about 28,800 beats, if 48 hours of single-channel ambulatory recording is considered, assuming an average heart-rate of about 60 beats-per-minute (BPM). The size of the database can be further increased if data is accumulated from multiple channels (up to 12) [7]. Considering the large computational costs of algorithms for processing ECS, processing large amounts of data is tending to take a long time.

### 2) The problem of generalization possibility

Existing approaches tend to suffer from high sensitivity to noise included in the ECG, and unreliability in dealing with new or ambiguous patterns from other patients. Classifiers of CAD systems must be developed that are able to deal with nonlinear discrimination between classes, incomplete or ambiguous input patterns, and suppression of false alarms [8].

The main difficulty in the task of automatic ECG analysis lies in the large scatter of the morphology of the ECG curves. Different patients have significantly changing parameters of the sequence of heartbeats (due to movements that produce high-frequency noises, or breathing, causing a baseline walk, or differences in the electrical characteristics of the body) [20]. Creation of a data set that would cover all possible ECG morphologies collected from different patients with various abnormalities in the functioning of the CVS is impossible. As a result, diagnostic systems based on classifiers of different types and trained on some collected data, show unsatisfactory diagnostic results of the ECG analysis of different patients [6,21,22] under real conditions of use.

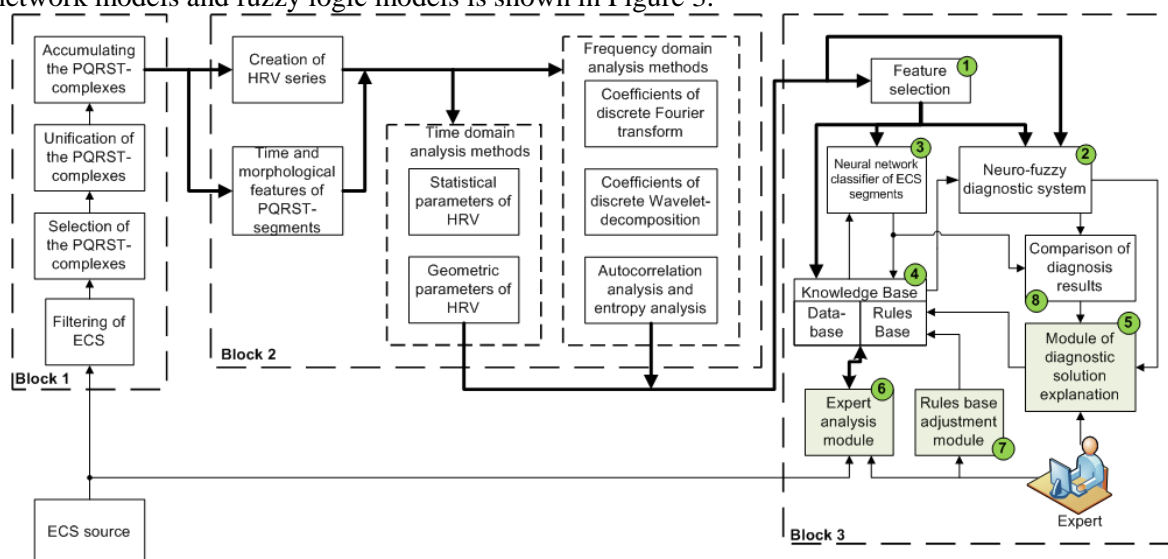
Thus, for the task of automatic arrhythmia diagnosis it is proposed to use data mining (DM) methods. The DM technologies are designed to search in large amounts of data of non-obvious, objective and useful regularities that reflect the objective internal data structure in comparison with the subjective opinion of the expert. The main purposes of the application of intelligent methods are:

- analysis of signals characterized by a high degree of uncertainty, e.g., “non-stochastic” type, which includes most biomedical signals, including ECS;
- increasing the level of intelligent assistance of medical specialists;
- revealing hidden regularities and extracting new knowledge from the accumulated data, which will allow to build production systems of explaining the diagnostic solutions.

To apply modern diagnostic methods to real ECS data, it is necessary to adapt existing methods of data mining, which allow to study complex nonlinear ECS more fully. CRISP-DM [27] (The Cross Industries Standard Process for Data Mining) is the most popular and widely used methodology. In accordance to the CRISP standard, data mining is a continuous process with many cycles and feedbacks.

### 3. Development of the structure of an intelligent system for arrhythmia diagnosis

To analyze complex nonlinear ECS signals it is proposed to use the diagnosis methods based on DM methods and algorithms. The structure of the proposed intelligent diagnostic system combining neural network models and fuzzy logic models is shown in Figure 3.



**Figure 3.** Structure of the neuro-fuzzy diagnostic system.

The system is divided into three blocks. The block 1 performs preprocessing of the raw ECS signal such as filtering the original signal and selecting the main segments describing the stages of cardiac contraction [13,18] – PQRST complex.

The procedure of generating and pre-selecting features is implemented in block 2. Time-stamps – durations of individual segments of the PQRST complex – are extracted from the signal and a time series of heart rate variability (HRV) is constructed, the analysis of which is widely used for the diagnosis of arrhythmias [32]. Feature generation methods perform signal processing both in the time and frequency domains. Statistical methods of analysis compute the basic statistical parameters of a HRV series [32], while geometric methods based on the analysis of the form of distribution function of the HRV series [33].

Since changes in the heart rate are difficult to analyze in the time domain [32], spectral analysis methods [2, 34-36] are used to investigate the frequency properties of HRV and the marked initial ECS such as discrete Fourier transform, wavelet-decomposition of the signal, and also autocorrelation analysis [33].

Most of arrhythmias can be analyzed due to the differences in the QRS part of the signal. The parameters of QRS complex can help to determine the differences in three of the five above-mentioned arrhythmias [4]. The most important function of the diagnostic system is the operative recognition of arrhythmias that directly threaten the patient's life.

It is proposed to consider the following types of arrhythmias and their characteristic features:

1. *Asystolia and expressed bradycardia*. Lack of or too low repetition rate QRS-complexes (less than 35 min<sup>-1</sup>). Effects cardiac arrest. For reliable recognition enough information about the durations of RR-intervals. RR-intervals are measured by QRS-complex detection algorithm.

2. *Ventricular fibrillation*. It begins with ventricular flutter (VF). Wave form similar to sine wave with frequency of up to 200 - 300 min<sup>-1</sup>; QRS and T waves are merged into a single wave of large amplitude without isoelectric intervals; Continuous wave form and amplitude varying with frequency of 400 to 600 min<sup>-1</sup> (ventricular fibrillation wave with small amplitude); small number of chaotic waves (150 - 300 min<sup>-1</sup>) of greater amplitude (ventricular fibrillation wave with large amplitude). There is no visible P wave. QRS complex and T wave are merged in regularly occurring undulatory waves with a frequency between 180 and 250 beats per minute. Effects termination of blood circulation and cardiac arrest. It requires analysis of wave structure of the ECS, featuring a variety of forms submitted from regular (sine) or irregular (variable amplitude and frequency) waves to chaotic oscillations expressed.

3. *Ventricular tachycardia*. Enhanced the QRS-complexes (longer than 120 ms), having the same or spindle-shaped, which is followed by a frequency of 120 - 220 min<sup>-1</sup>. Effects ventricular flutter, ventricular fibrillation. It may have three or more beats in a row that originate from the ventricle. A way that VT can be diagnosed is noticing that the P wave in the PQRST wave is no longer visible because the T wave from the previous beat has overtaken.

Based on this description, it is proposed to use the following ECS characteristics as the initial data for diagnostics:

1) Statistical:

- HR – heart rate;
- P-wave – moment of atrial contraction;
- PQ-segment – isoelectric interval, important for the recognition of the baseline;
- QRS-Complex – moment of ventricles contraction;
- ST-interval – isoelectric interval;
- T-wave – moment of the ventricles relaxation;
- RR-interval – defines HRV;
- PNN50 (%) – the proportion of adjacent R-R intervals that differ by more than 50 ms;
- RMSSD – the square root of the mean square of the difference in the values of adjacent R-R intervals. It is a measure of HRV with a short duration of cycles.
- SDANN – standard deviation from the mean of the R-R intervals computed for each 5-minute ECG recording section. Characterizes HRV with a long cycle duration.

2) Geometric (in time-domain):

- Mode ( $M_0$ ) – the most common value of the cardio interval in this dynamic series. Under normal distribution and high stationarity of the process under study,  $M_0$  is slightly different from the mathematical expectation;
- Amplitude of the mode – number of cardiointervals corresponding to the value of the mode;
- Variational range – reflects the degree of variability in the cardiointervals values in dynamic series. It is calculated from the difference between the maximum ( $M_x$ ) and the minimum ( $M_n$ ) values of the cardiointervals and, therefore, may be distorted in arrhythmias or artifacts.

### 3) Geometric (in frequency-domain):

The frequency composition of the heart rhythm is usually represented in the form of a graph of power distribution in frequency, according to which one can judge the expression of frequency components in different ranges:

- Total power in the range 0.15–0.40 Hz (period 6.5s–2.5s);
- Total power in the range 0.04–0.15 Hz (period 25s–6.5s);
- Total power in the range less than 0.04 Hz (period more than 25s);
- Total power in the range from 0.003 Hz to 0.40 Hz;
- Power Ratio.

### 4) Autocorrelation:

The calculation and construction of the autocorrelation function of the dynamic series of RR-intervals is aimed at studying the internal structure of this series as a random process. The autocorrelation function is a graph of the dynamics of the changes in the correlation coefficients obtained by sequentially shifting the analyzed dynamic series by one number relative to its own series:

- $C_1$  – value of the correlation coefficient after the first shift;
- $C_0$  – number of shifts, as a result of which the value of the correlation coefficient becomes negative.

Block 3 performs the diagnosis of arrhythmia. It consists of two complementary systems: a neural network classifier (NNC) and a neuro-fuzzy diagnostic solution (NFS). The neural network classifier was described in [10,13,32,36] and showed the results of a diagnostic accuracy of about 86% on the field data.

Features of marked signal go to modules 1 and 2. Module 1 implements feature selection procedure for NNC. Based on a pair of Hebb filters, this module isolates the most significant features of the time and frequency domains [10,13]. Then the set of selected features and the marked PQRST-segments are fed to the input of NNC (module 3) and NFS (module 2), and are also recorded in the database (module 4) [2,10,11,13,19].

The result of the comparison of the diagnostic solutions of the NNC and NFS, as well as the features identified by the NFS, are verified by the expert in module 5. The expert has access to the analysis of the original signal and the selected features through the interface module of expert analysis (module 6), as well as access to the rule base on which the NFS module operates via the interface module of the rule base adjustment (module 7). Analyzing the data at all stages of processing, the expert can make changes by manually marking PQRST segments, changing rules in the rule base or determining the type of arrhythmia if necessary.

The processed record is stored in the knowledge base. In the subsequent processing, this data will be used for the further NNC training.

## 4. Solution explanation subsystem model development and knowledge base construction

It is proposed to construct an explanation subsystem on the basis of neuro-fuzzy models. The advantage of such models is that it is a neural network implementation of fuzzy inference. On the one hand, parameters of membership functions act as weights and can be adjusted via training the network. On the other hand, each layer is a phase of fuzzy inference that makes transparent the functioning of network – it becomes easy to correlate outputs of certain layer with the corresponding phase of fuzzy inference.

Nowadays, such structures as the Takagi-Sugeno-Kang model (TSK) or the Mamdani model are distinguished as classifiers based on a fuzzy inference with the possibility of adjusting the parameters of membership functions [12,15,29,31-33]. However, these models in the classical version use a full rules base, which limits their application for problems with a large number of input variables and terms. So, it is proposed to use a structure based on the Mamdani model to classify the arrhythmia types based on the fuzzy inference, adapting the parameters of the membership functions of the terms of input and output linguistic variables and constructing a base of fuzzy rules as a result of training based on self-organization. Structure of the neuron used in the neural network interpretation of the fuzzy inference model is shown in Figure 5, structure of the neuro-fuzzy model is shown in Figure 6. Notations of layers of the network are given in Table 2 for the description.

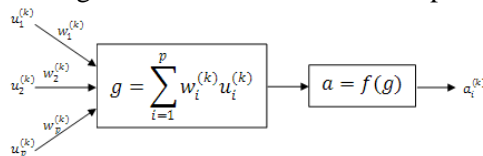


Figure 5. Neuron structure.

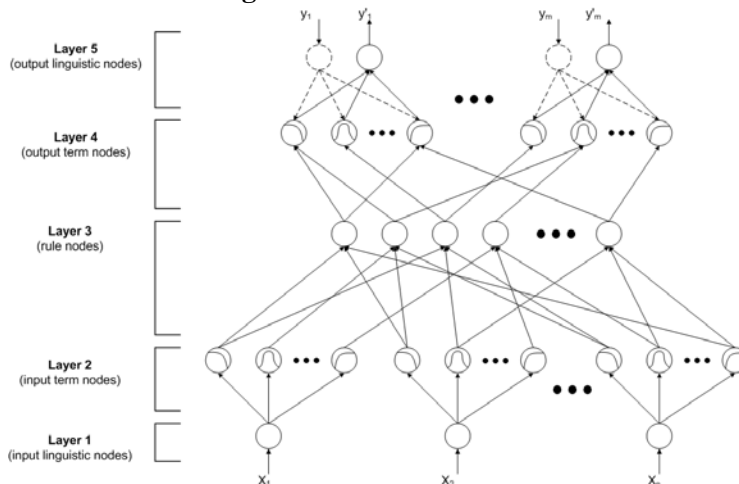


Figure 6. Structure of the neuro-fuzzy module.

Table 2. Notations of layers of the network.

Notation	Description
$k$	– number of the layer
$p$	– amount of node’s input links
$u_i^{(k)}$	– i-th input signal of k-th layer
$w_i^{(k)}$	– i-th link weight k-th layer
$g$	– integration function
$a$	– neuron activation function
$n$	– amount of input variables
$m$	– amount of output variables
$\sigma_{ij}, m_{ij}$	– width and center of the Gauss activation function of j-th term of i-th linguistic variable
$x_i$	– input variable
$y_i$	– training sample
$y'_i$	– output variable
$h_i$	– amount of terms of i-th linguistic variable

From the point of view of neural network architecture, rules base is a set of links and neurons of layer 3 in figure 5. Links are antecedents and neurons outputs are consequents of the rules. While constructing the network, each neuron of layer 2 connects to each neuron of layer 3 – this means building a full rules base. During the network training, most of the rules will be united or excluded.

Model of neural network functioning is the following. Nodes of first layer transmit input values directly to the next layer:

$$u_i^{(1)} = x_i, g = u_i^{(1)}, a = g, w_i^{(1)} = 1, i = \overline{1, n}. \quad (1)$$

Second layer performs a fuzzification operation. Activation function of the neurons of this layer is a membership function of the corresponding term of the linguistic variable. In this case, the Gaussian function is selected as the activation function:

$$g = M_{x_i}^j (m_{ij}, \sigma_{ij})^2 = -\frac{(u_i^{(2)} - m_{ij})^2}{\sigma_{ij}^2}, a = e^g, j = \overline{1, h_1}. \quad (2)$$

Number of neurons of the third layer is equal to the number of fuzzy rules, defined as

$$\text{numRules} = \prod_{i=1}^n h_i. \quad (3)$$

At the outputs of this layer, rules degree of activity is formed, defined as the minimum of the degrees of belonging calculated on the previous layer – the operation of the fuzzy “AND”:

$$g = \min_{i=\overline{1, p}} u_i^{(3)}, a = g, w_i^{(3)} = 1. \quad (4)$$

Neurons of the fourth layer operate in two modes: when transmitting a signal from the bottom to the top (operating mode) and vice versa (training mode). In the operating mode, neurons implement the composition of fuzzy subsets assigned to each output variable, using the sum of input signals with saturation – a fuzzy “OR” operation [28,33]:

$$g = \sum_{i=1}^p u_i^{(4)}, a = \min(1, g), w_i^{(4)} = 1. \quad (5)$$

In the training mode, the neurons of this layer operate similarly to the neurons of the second layer.

As shown in the figure 5, the fifth layer consists of two types of neurons. In the training mode, a training sample is fed to the input of the first type of neurons on the basis of which the parameters of the activation function of neurons of the fourth layer are adjusted. Thus, neurons of this type (marked with dashed lines) operate similarly to the neurons of the first layer:

$$g = y_i, a = g. \quad (6)$$

In the operating mode, second type of neurons operates (marked by a solid line). Neurons of this type perform the defuzzification operation using centroid method:

$$g = \sum w_{ij}^{(5)} u_i^{(5)} = \sum (m_{ij} \sigma_{ij}) u_i^{(5)}, a = \frac{f}{\sum \sigma_{ij} u_i^{(5)}}. \quad (7)$$

Learning algorithm [28,33] is divided into two phases:

- Training on the basis of self-organization to initialize the membership functions of terms of input and output linguistic variables and to construct fuzzy rule base;
- Supervised learning to adjust the parameters of the created membership functions and minimize the network error.

At the first stage of phase 1 expert manually determines the structure of the fuzzy system such as the number of terms of each input and output linguistic variable and initiates parameters of membership functions.

Second stage is responsible for primary selection of membership functions and their parameters: center and width. It is proposed to use the method of competitive training [28,33]. It involves independent iterative training of each neuron that implements the membership functions to the corresponding terms of linguistic variables. This training involves finding the minimum distance from the center of the membership function to the current input sample from the training sample. To initialize the value of the width of membership functions, first nearest neighbor heuristic can be used, since in the second phase of learning the optimal values of the parameters will be determined.

At the third stage, fuzzy rules are forming. First, a complete rules base is created, which is a combination of all conditions with all conclusions. Then, the competitive learning algorithm is used to determine the required number of rules from the full database. Here the fourth layer neurons operate in the training mode and receive the training sample signals from the fifth layer. Denoting the output of the  $j$ -th neuron of the fourth layer as  $o_j^{(4)}(t)$ , and the production part of the  $i$ -th rule (the weighted sum of the signals entering the neurons of the 4th layer) as  $o_i^{(3)}(t)$ , calculate the links weights  $w_{ij}$  as:

$$w_{ij}(t + 1) = o_i^{(4)}(t) \cdot [w_{ij}(t) - o_i^{(3)}(t)]. \quad (8)$$

At the fourth stage, some rules are deleted. If there is more than one output connection for the neuron of the third layer, the connection with the largest weight is selected, and the remaining ones are deleted. It is also possible to delete all links and the neuron itself if the weights of all the links are less than a given threshold  $\theta = 0.4$ .

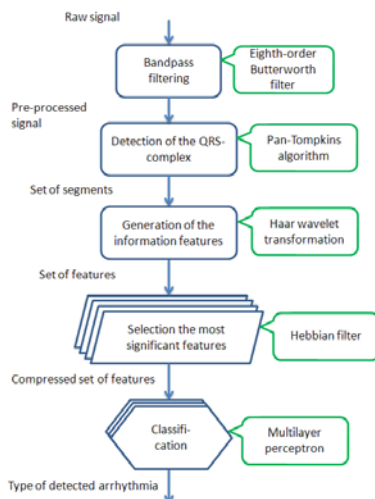
At the fifth stage, rules are combined to reduce their number. In Figure 5, the preconditional part of the rules is displayed by neurons of the third layer, and part of the consequent – by the connections of the fourth layer.

The sixth stage involves building a neural network with a complete structure and a set of elements and links.

At phase 2, optimization of activation function parameters values is done by supervised learning of the network. For example, a back propagation algorithm for can be used for training [19]. Thus, the reverse pass, which calculates the error for the parameters of the center and the width of the Gaussian activation functions of layers 5 and 2, is determined similarly as in [28,33].

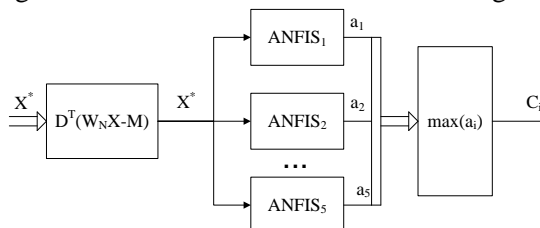
### 5. Experiment

Fig. 7 shows a generalized scheme of the diagnostic model of signal preprocessing, detection and identification the events on the basis of ECS analysis with the indicating of algorithms proposed for the implementation of each stage.



**Figure 7.** Generalized scheme of the diagnostic model of signal preprocessing.

Thus, after filtering and detection the PQRST-complex, wavelet decomposition is made on appropriate section of the original data vector and vector  $[sa_j, sd_j, sd_{j-1}, \dots, sd_1]$  is forming, where  $sa, sd$  are the approximation and detailing coefficients of  $j, j-1, \dots, 1$  levels respectively. Then, the 8 major components are allocated from resulting vector [10]. As shown in [32], a given number of main components is sufficient enough to save 95% of the variance of the original signal.



**Figure 8.** Fuzzy model committee of ANFIS networks.

In experiments [10, 32, 36], the resulting vector was fed to the input of a multilayer perceptron. The accuracy of arrhythmia classification was 86%. In this research, the vector is fed to the input of the fuzzy model committee described above. Each model determines the probability evaluation of having an appropriate arrhythmia class, as shown in Fig. 8.



In Fig. 8:

$X$  – source vector – 256...64 points of PQRTS-complex taken at regular time intervals;

$W_N = \begin{bmatrix} H_N \\ G_N \end{bmatrix}$  – matrix recording of successively applying the Haar decomposition, where:

$H_N$  – low-frequency components matrix allocation (approximation coefficients);

$G_N$  – high-frequency components matrix allocation (detailing coefficients);

$X^* = W_N X$  – direct Haar transformation;

$X = W_N^T X^*$  – reverse Haar transformation;

$Y = D^T (W_N X - M)$  – vector of extracted main components;

$D^T = (V_1, V_2, \dots, V_K)$  – eigenvectors corresponding to the first  $k$ -principal components, forming the projection matrix;

$M$  – the mathematical expectation of each component of all observation in vector  $X$ .

For each sample  $x_j$  from the source data after the training of the network committee, a module for explaining the decision can be obtained:

1. For the  $i$ -th ANFIS network, which demonstrates the greatest response to the input sample, a vector of activities of fuzzy rules can be constructed.
2. Sorting the rules in order of decreasing its activity, the contribution of the first  $M$  rules to the summary conclusions can be evaluated and the types of their conditional part can be traced: the corresponding values of linguistic variables terms.
3. With the help of the transformation (9) it is possible to move from the reduced feature space to the initial basis and to interpret the values of the initial characteristics typical for each type of arrhythmia:

$$\tilde{X} = W_N^T [D(D^T(W_N X' - M)) + M] \tag{9}$$

$\tilde{X}$  – values of the initial characteristics typical for  $i$ -th type of arrhythmia;

$X'$  – vector obtained as a result of the reverse output in ANFIS.

Described procedure of ECS processing was implemented in MATLAB and tested on a sample from the database of arrhythmias “MIT-BIH Arrhythmia Database” [2]. From this base records of diseases of 5 classes were selected, total amount of records that have been classified is 29,537. The choice of these classes is due to a sufficient number of records required for a complete training of the classifier. Selected classes are given in the Table 3.

**Table 3.** Notations of layers of the network.

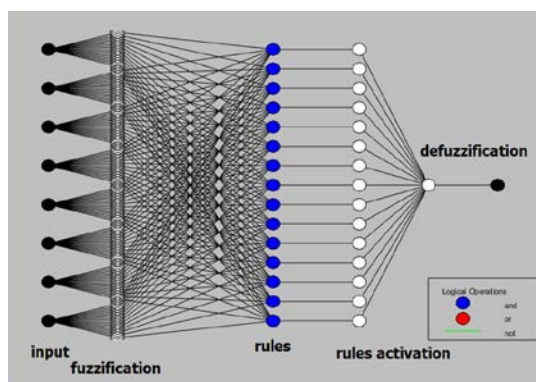
Class (type of arrhythmia)	Class code	Amount of records
<b>Paced beat</b>	C1	6977
<b>Atrial premature beat</b>	C2	2452
<b>Left bundle branch block beat</b>	C3	8018
<b>Right bundle branch block beat</b>	C4	7091
<b>Normal beat</b>	C5	4999

Arrhythmias classification was carried out by using a multilayer perceptron and a neuro-fuzzy system committee implemented as neural networks. Each network in the committee determines the probability evaluation of having an appropriate arrhythmia class in the sample of input data. The network structure is presented in the table 4 and in the Fig.9.

Table 5 contains the comparative results of the classification of the multilayer perceptron and the neuro-fuzzy system performed in the form of a neural network committee.

**Table 4.** Structure of the neuro-fuzzy system networks committee.

<b>Input vector dimension</b>	<b>(8,1)</b>
<b>Number of networks in the committee</b>	5
<b>Number of rules in one network</b>	15
<b>Number of terms of a fuzzy variable</b>	15
<b>Membership function</b>	Gaussian



**Figure 8.** Fuzzy model committee of ANFIS networks.

**Table 4.** Computation results.

Parameter	Value MLP	Value NFS
<b>Amount of records</b>	29537	29537
<b>Correctly recognized</b>	25565 (86,55%)	25704 (87.023%)
<b>Prediction error</b>	3972 (13,45%)	3833 (12.976%)

Judging by the number of correctly classified records, it can be argued that the neuro-fuzzy model is not inferior to the neural network in terms of the classification quality. At the same time, the neuro-fuzzy model allows to rank the fuzzy production rules according to the level of their activation, and to demonstrate by which elements of source feature vector a decision was made on whether the sample belongs to appropriate class of arrhythmia.

**6. Discussions**

When using a neural network classifier, various parameters can be used as characteristics of the sample of the analyzed area of the ECS, related to the frequency, time, and parametric forms. The key step in the analysis of the ECS in the diagnostic task is the construction of a classifier capable of deciding that the signal section belongs to a particular class of state in terms of the totality of the allocated features. Source space of characteristics can be transformed into one of the subspaces in which the separating hyperplane will have a simpler description. In this approach to the construction of an intelligent diagnostic system, it is difficult to implement a decision explanation subsystem.

With the use of a neuro-fuzzy classifier, it becomes possible to use the experience accumulated by cardiologists in the tasks of diagnosing certain types of arrhythmias. The existing medical methods can be adapted to a representation in a set of product rules of the form “IF-THEN”, taking into account the adjustable weights that allow us to assess the degree of importance of both the individual feature and the rule as a whole.

Formalized expert knowledge in this case is transferred to the structure of the neuro-fuzzy module and becomes the core of the classifier. At the next stages of the diagnostic system construction, with DM techniques it becomes possible to identify hidden and unobvious patterns that complement expert knowledge, which will improve the accuracy of diagnosis. A feature of the use of neuro-fuzzy networks is the possibility of supervised learning on the existing base of marked samples assigned to different classes, which will allow to expand the base of diagnostic rules.

**7. Conclusions**

In this paper an approach to improve the intelligent systems of arrhythmia diagnostics on the basis of neural network classifiers by developing a solution explaining subsystem based on Mamdani neuro-fuzzy model is proposed.

The analysis of modern methods of automatic diagnosis of cardiac arrhythmias is carried out. It is established that the correctness of diagnostics of the most common algorithms in automated diagnostic systems does not exceed 80%.

The structure of a diagnostic system for the detection of cardiac arrhythmias based on intellectual data mining technology is proposed.

The issue of the knowledge base construction – coding, generation and selection of the features (database) as well as the construction of the rules base – as the part of the neuro-fuzzy diagnostic system is considered.

It is shown that using the neuro-fuzzy classifier it becomes possible to use the experience accumulated by cardiologists in the tasks of diagnosing certain types of arrhythmias. A model of the neuro-fuzzy solution explaining module is proposed to formalize the expert knowledge.

## 8. References

- [1] Cardiovascular diseases (CVDs). Fact sheet. – [Electronic resource]. URL: available: <http://www.who.int/mediacentre/factsheets/fs317/en/>.
- [2] Rangayyan, R.M. Biomedical signal analysis. Second edition / R.M. Rangayyan // John Wiley & Sons, 2015. – 720 p.
- [3] De Chazal, P. Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features / P. de Chazal, M. O'Dwyer, R.B. Reilly // IEEE Trans. Biomed. Eng. – 2004. – Vol. 51(7). – P. 1196-1206.
- [4] Boonperm, P. ECG analyzing program for arrhythmia detection / P. Boonperm, D. Supakasemwong, P. Naiyanetr // 7th Biomedical Engineering International Conference, 2015. – P. 1-4.
- [5] Chakroborty, S. Accurate Arrhythmia classification using auto-associative neural network / S. Chakroborty // Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2013. – P. 4247-4250.
- [6] Ceylan, R. Comparison of FCM, PCA and WT techniques for classification ECG arrhythmias using artificial neural network / R. Ceylan, Y. Özbay // Expert Syst. Appl. – 2007. – Vol. 33(2). – P. 286-295.
- [7] Chakroborty, S. Real-time arrhythmia classification for large databases / S. Chakroborty, M.A. Patil // Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2014. – P. 1448-1451.
- [8] Gao, D. Bayesian ANN classifier for ECG arrhythmia diagnostic system: A comparison study / D. Gao // Proceedings of the International Joint Conference on Neural Networks. – 2005. – Vol. 4. – P. 2383-2388.
- [9] Amann, A. Reliability of old and new ventricular fibrillation detection algorithms for automated external defibrillators / A. Amann, R. Tratnig, K. Unterkofler // Biomed. Eng. Online. – 2005. – Vol. 4. – P. 60.
- [10] Nikonov, A.V. Data mining algorithms in the problem of cardiac arrhythmia diagnosis / A.V. Nikonov, A.M. Vulfin // Proceedings of a VI scientific conference with international participation “Information Technologies and Systems”, 2017. – P. 200-205.
- [11] Rodriguez, R. Adaptive threshold and principal component analysis for features extraction of electrocardiogram signals / R. Rodriguez // Proceedings International Symposium on Computer, Consumer and Control, 2014. – P. 1253-1258.
- [12] Kar, S. An Investigative Study on Early Diagnosis of Breast Cancer Using a New Approach of Mathematical Shape Theory and Neuro-Fuzzy Classification System / S. Kar, D.D. Majumder // Int. J. Fuzzy Syst. – 2016. – Vol. 18(3). – P. 349-366.
- [13] Nikonov, A.V. Neural network hardware-software complex for the cardiac arrhythmias diagnosis / A.V. Nikonov // Proceedings of the X All-Russian Youth Scientific Conference “Mavlyutovskie Readings”, 2017. – P. 1593-1597.
- [14] Ahmed, A.F. Novel Bayesian classifier discriminant function optimization strategies for arrhythmia classification / A.F. Ahmed, M.I. Owis, I.A. Yassine // IEEE-EMBS International Conference on Biomedical and Health Informatics, BHI, 2014. – P. 693-696.
- [15] Holimchayachotikul, P. Predictive performance measurement system for retail industry using

- neuro-fuzzy system based on swarm intelligence / P. Holimchayachotikul, K. Leksakul // *Soft Comput.* – 2017. – Vol. 21(7). – P. 1895-1912.
- [16] Dilmac, S. A new ECG arrhythmia clustering method based on Modified Artificial Bee Colony algorithm, comparison with GA and PSO classifiers / S. Dilmac, M. Korurek // *International Symposium on Innovations in Intelligent Systems and Applications*, 2013. – P. 1-5.
- [17] Faziludeen, S. ECG beat classification using wavelets and SVM / S. Faziludeen, P.V. Sabiq // *IEEE Conference on Information and Communication Technologies*, 2013. – P. 815-818.
- [18] Chetan, A. Cardiac arrhythmia classification from multilead ECG using multiscale non-linear analysis / A. Chetan, R.K. Tripathy, S. Dandapat // *IEEE UP Section Conference on Electrical Computer and Electronics, UPCON*, 2016. – P. 1-4.
- [19] Haykin, S. *Neural Networks and Learning Machines* / S. Haykin // Pearson Prentice Hall New Jersey USA 936 pLinks. – 2008. – Vol. 3. – 906 p.
- [20] Barro, S. Classifying multichannel ECG patterns with an adaptive neural network / S. Barro // *IEEE Engineering in Medicine and Biology Magazine.* – 1998. – Vol. 17(1). – P. 45-55.
- [21] Ghanwat, T.B. Arrhythmia and neural network / T.B. Ghanwat // *International Conference on Communication and Signal Processing*, 2015. – P. 1361-1365.
- [22] Osowski, S. Recognition and classification system of arrhythmia using ensemble of neural networks / S. Osowski, T. Markiewicz, L.T. Hoai // *Meas. J. Int. Meas. Confed.* – 2008. – Vol. 41(6). – P. 610-617.
- [23] Jekova, I. Comparison of five algorithms for the detection of ventricular fibrillation from the surface ECG / I. Jekova // *Physiol. Meas.* – 2000. – Vol. 21(4). – P. 429-439.
- [24] Arumugam, S.S. Wavelet based detection of ventricular arrhythmias with neural network classifier / S.S. Arumugam, G. Gurusamy // *J. Biomed. Sci. Eng.* – 2009. – Vol. 2(6). – P. 439-444.
- [25] Uus, A. Singular spectrum analysis for detection of abnormalities in periodic biosignals / A. Uus, P. Liatsis // *Int. Conf. Syst. Signals, Image Process.* – 2011. – P. 375-378.
- [26] Hadjem, M. An ECG T-wave anomalies detection using a lightweight classification model for wireless body sensors / M. Hadjem, F. Nait-Abdesselam // *IEEE International Conference on Communication Workshop*, 2015. – P. 278-283.
- [27] Chapman, P. *Crisp-Dm 1.0* / P. Chapman // *Cris. Consort.* – 2000. – P. 76.
- [28] Rutkovskaya, D. *Neural Networks, Genetic Algorithms and Fuzzy Systems* / D. Rutkovskaya, M. Pilihsjkij // *Hot line Telecom.* – Moscow, 2008.
- [29] Coy, C.G. A hybrid-genetic algorithm for training a Sugeno-type Fuzzy Inference System with a mutable rule base / C.G. Coy // *Theses and Dissertations.* – [Electronic resource]. URL: <http://utdr.utoledo.edu/theses-dissertations/814>.
- [30] Pan, J. A Real-Time QRS Detection Algorithm / J. Pan, W.J. Tompkins // *IEEE Trans. Biomed. Eng.* – 1985. – Vol. BME-32(3). – P. 230-236.
- [31] Wang, L. *Fuzzy systems and knowledge discovery : second international conference* / L. Wang, Y. Jin // *FSKD proceedings.* – Changsha, China, August 27-29, 2005. – P. 1354.
- [32] Nikonov, A.V. Neuro-fuzzy extractor in the problem of cardiac arrhythmia diagnosis / A.V. Nikonov, A.M. Vulfin, M.M. Gayanova // *Proceedings of the XI All-Russian Youth Scientific Conference “Mavlyutovskie Readings”*, 2017.
- [33] Lin, C. A Neural Fuzzy Control System with Structure and Parameter Learning / C. Lin // *Fuzzy Sets Syst.* – 1995. – Vol. 70. – P. 9-13.
- [34] Amin, A.-H.N. Analysis of bioelectrical signals for solving problems of medical diagnostics / A.-H.N. Amin // *System of processing and neural network*, 2014. – P. 421-424.
- [35] Fomasa, E. HRV spectral and fractal analysis in heart failure patients with different aetiologies / E. Fomasa // *Comput. Cardiol.* – 2014. – Vol. 4. – P. 421-424.
- [36] Nikonov, A.V. Cardiovascular diseases diagnosis on the basis of neural network analysis of the biomedical signals / A.V. Nikonov // *IEEE International Symposium on Signal Processing and Information Technology*, in publishing.
- [37] Ilyasov B.G. *Intelligent control systems. Theory and practice* / B.G. Ilyasov, V.I. Vasilyev // M: Radiotekhnica, 2009.