

WAVELET-NEURAL SYSTEMS AS APPROXIMATORS OF AN UNKNOWN FUNCTION – A COMPARISON OF BIOMEDICAL SIGNAL CLASSIFIERS

PAWEŁ KOSTKA AND EWARYST TKACZ

*Silesian University of Technology, Institute of Electronics,
Division of Biomedical Electronics,
Akademicka 16, 44-100 Gliwice, Poland
{kostka, et}@boss.iel.polsl.gliwice.pl*

(Received 12 September 2003; revised manuscript received 14 February 2004)

Abstract: Wavelet-neural systems (WNS) presented in this work, inheriting the properties of neural networks, belong to the class of universal approximators of unknown functions, F , describing the relationship between input $\mathbf{X} \in \mathbb{R}^N$ and output $\mathbf{Y} \in \mathbb{R}^M$ of a process or object. Classifier structures described in this work fulfil the role of approximators of functions, which are able to assign the input signal to a particular class with a given accuracy. A performance comparison of elaborated classifier structures with preliminary time-frequency analysis in the wavelet layer has been made for different types of the neural part. A feed forward multi-layer perceptron and a neural net with radial basic functions are analysed theoretically and practically. Results included in this paper present a comparison of the learning and verification stages of a classifier, tested on the basis of non-stationary signals of heart rate variability. Despite the fact that a WNS with the Morlet basic function gives the best results for the learning phase of WNS, the other tested wavelets used in the preliminary layer, Db4, allow us to obtain the best system performance during its verification.

Keywords: wavelets, neural networks, biomedical signal classifiers

1. Introduction

The learning process of artificial neural networks can be interpreted as an expansion of an unknown relation, $f(x)$, between the input, $\mathbf{X} \in \mathbb{R}^N$, and output, $\mathbf{Y} \in \mathbb{R}^M$, of the analysed object or phenomenon in space $F(x)$, determined by the set $\phi_l(x, k)$ of activation functions of the neurons:

$$F(x) = \sum_{l=1}^m c_l \phi_l(x, k), \quad (1)$$

where k is a parameter chosen during the learning process.

Coefficients $c_1 \dots c_m$ are the parameters of neuron activation functions, computed on the basis of experimental data to minimize the approximation error.

Poggio, Girosi and Hornik have shown [1, 2] that the learning of feed-forward neural networks without feedback connections can be treated as a projection of a multi-dimensional function, $F(x)$, which corresponds to the reconstruction of a hyper-plane representing the unknown relation, $f(x)$.

Approximation abilities of artificial neural networks are expressed in the thesis of Cybenko, who has proven that, if $\sigma(\cdot)$ is a continuous discriminating function, then the sum $g(x)$ expressed by:

$$g(x) = \sum_{i=1}^N w_i \sigma(a_i^T x + b_i), \quad (2)$$

can approximate the unknown continuous function, f , defined on the $[0, 1]^n$ interval with a given accuracy, ε :

$$|g(x) - f(x)| < \varepsilon, \quad \text{for } x \in [0, 1]^n; \quad \varepsilon > 0, \quad (3)$$

where $w_i, b_i \in \mathbb{R}$, $a_i \in \mathbb{R}^n$.

Equation (2) can express the action of a multilayer perceptron with sigmoidal activation functions, $\sigma(\cdot)$, which fulfil the conditions of the discriminating function.

In classifier structures presented in this paper, the approximation abilities of ANN, well-known for years, have been used the influence of an additional layer with preliminary time-frequency analysis on system performance has also been studied. The learning and verification stages of WNS have been evaluated to find the optimal wavelet function for the preliminary layer, where time-frequency signal decomposition is performed, and the best neural network structure for the classifier task.

All classifiers have been tested on the basis of heart rate variability signals (HRV) obtained from patients with coronary diseases. The wavelet transform used in the preliminary layer seems to be a suitable tool for HRV signals, which are non-stationary in nature. Information included in time and frequency domain analysis of these signals reflect the activity of the human control nervous system [3]. The main task is to extract data which can distinguish pathological and psychological cases.

2. Wavelet neural systems in the structure of a biomedical signal classifier

2.1. General structure of WNS

The idea of using wavelets in neural networks was first proposed first by Zhang and Benveniste [4], Pati and Krishnaprasad [5]. The output of the so-called wavelet networks, presented in Figure 1, can be expressed as:

$$y(x) = \sum_{i=1}^N w_i \psi\left(\frac{x - b_i}{a_i}\right) + w_0. \quad (4)$$

The structural and analytical description of the wavelet network was analogous to that of the neural network perceptron type, with neurons replaced with pre-scaled and shifted basic wavelet functions, $\psi_{a,b}(t)$. There are two approaches to the construction of such wavelet networks presented in literature. They are based on: the construction of a network based on signal decomposition, as presented in [6], or extensions of the radial basic function approach shown in [7].

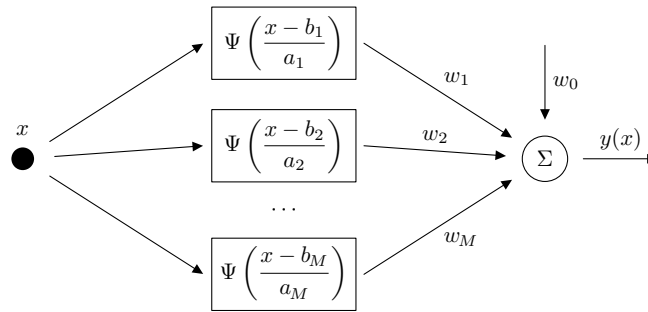


Figure 1. The structure of a wavelet network proposed by Q. Zhang and A. Benveniste [4]

Wavelet neural systems (WNS) described in this paper have a different structure (Figure 2). In the first stage, a continuous representation of a naturally unevenly spaced HRV signal is performed using Derivative Cubic Spline interpolation. Then, resampling with frequency $f_s = 5$ [Hz] is done. These preliminary steps are necessary to obtain a signal in the form required by the bank of wavelet filters, which is applied in the next stage of the system.

Before insertion into the neural net, the vector \mathbf{X}_2 of a discrete input signal is processed in an additional layer with wavelet nodes, where wavelet time-frequency decomposition is performed. The output of this layer is a new vector, \mathbf{X}_3 , which contains the most characteristic part of the input signal for further analysis in the neural network. Adaptive wavelet transform signal analysis, which takes place in WNS, relies on the fact, that, apart from the neural net parameters (weights and biases), also wavelet scale and shift factors are subjected to the learning process. In the approach presented in this paper, two types of learning algorithms are described, which have been determined by the features of the basic wavelet function used in the wavelet layer of WNS. One algorithm, based on a modified error back propagation scheme was applied for a WNN structure using the Morlet wavelet, which is given by a differentiable analytical formula. A pre-determined wavelet decomposition dyadic

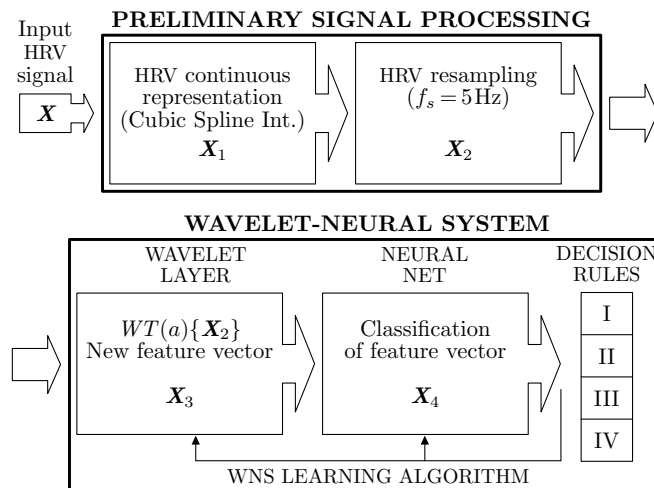


Figure 2. General block diagram of a wavelet neural system with preliminary HRV signal processing

grid and entropy analysis were used in the other algorithm, designed for orthogonal Daubechies (Db4) and bi-orthogonal (Bior 2.4) mother wavelets. The way of WNS parameter initialization, a part of the learning procedure, is also highly conditioned by the specific features of the problem solved with the usage of WNS, as will be shown in the following sections.

2.2. Learning algorithms of wavelet neural systems

As distinct from Zhang wavelet networks introduced in Section 1, where both a and b wavelet parameters were subjected to the learning process, in the wavelet neural system (WNS) presented in this paper only the value of scale factor a is optimised. This results from the nature of the analysed problems, where the basic goal is to extract the most characteristic signal component, given over the whole original signal time range and not in one time instant.

The common main stages of WNS learning procedures can be stated as follows [8]:

1. Definition of the WNS structure and the type of the used mother wavelet, which is the most important stage in the construction of a learning algorithm.
2. Initialization of the wavelet layer and the neural net part of WNS parameters, based on specific features of the problem under consideration.
3. Definition of the quality index (cost function) – the measure of WNS performance, which will be optimised during the learning process.
4. Updating the optimised parameters according to the learning rules until meeting the end conditions set for the quality index or the maximum value of learning iterations.

2.2.1. Parameter initialization in wavelet neural systems (WNS)

As will be shown in the results section of this paper, properly setting the initial scale parameter a of the wavelet layer is crucial for speed and convergence of a WNS learning process. This initial common step of the described learning algorithms consists in the determination in the first approach of the range of the possible scale a value changes. Subsequently, the range of a initial values which guarantee the system's parameter convergence should be defined.

In contrast to the first step, which is performed using the characteristic information included in the time or frequency domain of the analysed signals, the second stage of wavelet scale parameter definition is based on experimental results obtained for different set values.

Two learning algorithms meant for different types of wavelet basic functions used in WNS are presented below.

2.2.2. The modified error back-propagation algorithm

A modification of the multi-layer perceptron error back-propagation learning algorithm to adapt it for WNS consists in the introduction of a scale parameter, a , of the wavelet layer into the optimisation procedure. A new value of a is computed at every learning iteration on the basis of the WNS output error gradient, similarly to the neural net weight and bias parameters. A wavelet layer structure corresponding

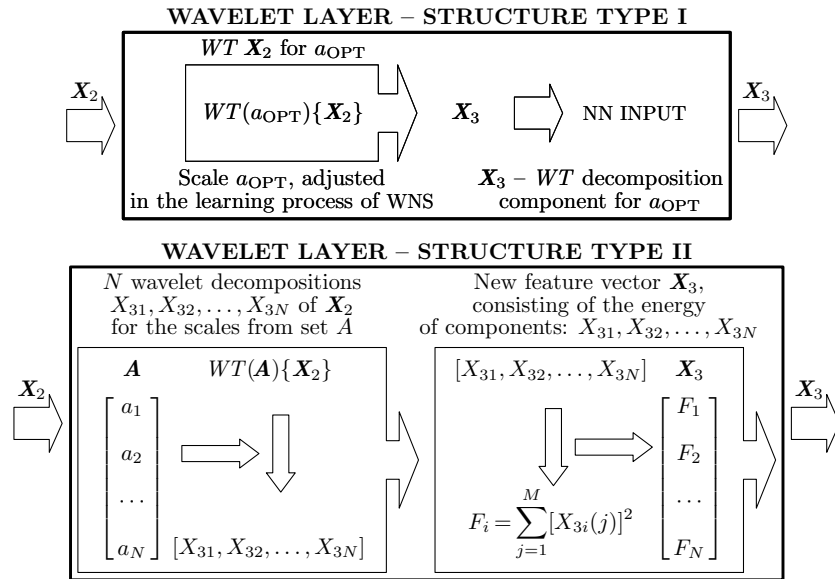


Figure 3. Two structures of the preliminary wavelet layer: (top) optimal component extraction for scale a_{OPT} computed during learning process; (bottom) feature vector created on the basis of energies of wavelet decompositions for set A of wavelet scale values; WT – wavelet transform, NN – the neural network part of system

to this algorithm is presented in Figure 3 (top). As a product of the wavelet layer, one-signal decomposition for the optimal scale, a_{OPT} , is obtained.

Algorithms based on the steepest gradient descent require an activation function of the neural nodes to be expressed by an analytic, differentiable formula. These conditions are fulfilled by the Morlet mother wavelet, which has been used in WNS, learnt according to a modified error back propagation algorithm.

The real Morlet wavelet function is expressed by the following analytical formula:

$$\psi_{\text{Morlet}}(t) = \cos(\omega_0 \tau) \exp(-0.5\tau^2), \quad (5)$$

where $\tau = \frac{t-b}{a}$, $\omega_0 = 2\pi f_0$.

According to the delta rule, the increase of scale a is computed as a negative value of the error gradient:

$$\Delta a = -\frac{\partial E}{\partial a}, \quad (6)$$

where

$$E = \sum_{i=1}^{Ns} (\varepsilon_i^2). \quad (7)$$

The error of k^{th} of L layers is expressed as follows:

$$\varepsilon_i^{(k)}(n) = \begin{cases} d_i^{(L)}(n) - y_i^{(L)}(n) & \text{for } k = L, \\ \sum_{m=1}^{N_{k+1}} \delta_m^{(k+1)}(n) w_m^{(k+1)}(n) & \text{for } k = 1, \dots, L-1. \end{cases} \quad (8)$$

As this layer is the realization of a continuous wavelet transform, its output is given by:

$$y^{(TW)} = \langle s(t), \psi_{\text{Morlet}}(t) \rangle = \frac{1}{\sqrt{a}} \cdot \int_{-\infty}^{\infty} s(t) \cdot \psi\left(\frac{t-b}{a}\right) dt. \quad (9)$$

From formula (6), after replacing integration with summation due to the discrete nature of the analysed signals, the value of increase Δa can be expressed by formula (10), while formula (11) describes the new, modified wavelet scale parameter value. As in neural net structures, the learning rate factor μ_a decides the extent of influence of the computed Δa on the new value of parameter a :

$$\frac{\partial E}{\partial a} = -2 \sum_{i=1}^{Ns} \varepsilon(i) s(i) \tau_i \frac{\partial \psi_{\text{Morlet}}(\tau_i)}{\partial b}, \quad (10)$$

$$a(n+1) = a(n) + \mu_a \Delta a. \quad (11)$$

The required value of the first derivative of the Morlet wavelet mother function is given by the following equation:

$$\frac{\partial \psi_{\text{Morlet}}(\tau_i)}{\partial b} = \frac{1}{a} [\omega_0 \sin(\omega_0 \tau) \exp(-0.5\tau^2) + \tau \psi(\tau)]. \quad (12)$$

The presented algorithm belongs to the group of local optimisation. To improve its performance, mainly in reducing the possibility of settling down at a local minimum, procedures of momentum term introduction and an adaptive learning factor were applied as suggested in literature [9]. Additionally, the learning process of WNS was repeated in the presented application fields with random initial NN parameters to find their best values.

2.2.3. A WNS learning algorithm based on a pre-determined discrete dyadic grid of WT scale parameter

For the remaining basic wavelets, Db4 and Bior 2.4, which are represented by the corresponding filters, the wavelet frames theory [10] and multilevel Mallat decomposition [11] have been applied. Figure 3 (bottom) presents the structure of the wavelet layer for this learning algorithm. The proposed learning procedure can be described by following steps:

- A. An analysis of the initialization stage results, which contain the information about the range of possible wavelet scale a changes, determined by the specific features of the studied problem.

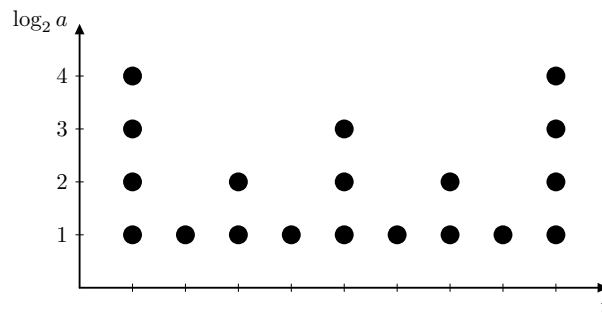


Figure 4. Dyadic grid of WT scale, a values distribution

Table 1. Location of frequency sub-bands corresponding to wavelet the decomposition levels according to the Mallat multi-resolution algorithm

Wavelet decomposition component	Scale a_i	Borders of i^{th} sub-band [Hz]
d_3	8	0.3125 – 0.6250
d_4	16	0.1563 – 0.3125
d_5	32	0.0781 – 0.1563
d_6	64	0.0391 – 0.0781
d_7	128	0.0195 – 0.0391
d_8	256	0.0098 – 0.0195

- B. Definition of a dyadic grid (Figure 4) of discrete a values in the appropriate range (see Table 1).
- C. Multi-level signal decomposition according to the Mallat algorithm for the established discrete values of the optimised scale parameter, a . It corresponds to signal filtering into:

$$f_{C,k} = f_0/a_k, \quad (13)$$

the frequency band, the central frequency of which is connected with the decomposition level and scale a by formula (13).

- D. Choice of the optimal scale a (decomposition level) based on the measure of energy and the amount of information included in the consecutive components obtained in multilevel signal decomposition. Energy (14) and two other entropy-based criteria (15), (16) which match this condition and describe information-related properties for a given signal have been tested (where $s(n)$ is the analysed signal, $c_{i,j}$ – i^{th} WT coefficient on the j^{th} decomposition level):

- energy:

$$E_{3,j}\{c_{i,j}\} = \log(c_{i,j})^2 \Rightarrow E_{3,j}\{s(n)\} = \sum_i \log(c_{i,j})^2; \quad (14)$$

- the (non-normalized) Shannon entropy:

$$E_{1,j}\{c_{i,j}\} = -(c_{i,j})^2 \log(c_{i,j})^2 \Rightarrow E_{1,j}\{s(n)\} = -\sum_i [(c_{i,j})^2 \log(c_{i,j})^2]; \quad (15)$$

- the concentration in l^P norm with $1 \leq p < 2$:

$$E_{2,j}\{c_{i,j}\} = |c_{i,j}|^P \Rightarrow E_{2,j}\{s(n)\} = -\sum_i |c_{i,j}|^P = \|s(n)\|_P^P. \quad (16)$$

An evaluation of the effect of the various criteria has allowed us to take the energy of a chosen wavelet decomposition component (16) for use in the learning algorithm of WNNs.

3. Materials

3.1. Application of wavelet neural networks

Considering the characteristic features of non-stationary heart rate variability (HRV) signals, time-frequency analysis performed by a wavelet transform seems to be

a very effective tool to reveal the useful diagnostic information included therein [12]. Regarding the problem of non-invasive coronary heart disease detection, a review of time-frequency methods used in ECG and EEG signal analysis has been presented in [3]. The results obtained by Thakor indicate the existence of a relationship between pathologies of coronary arteries and ECG parameters, especially the change of the S-T interval caused by oxygen deficiency in the heart muscle with atherosclerosis in coronary arteries. These pathologies, difficult to detect in the ECG shape only, can be detected by time-frequency methods, which allow changing the domain of analysis.

3.2. Database of signals

For the classification procedure, a HRV data set was collected with the help of a notebook PC-based electrocardiograph system from 62 patients clinically characterized as cases of coronary artery diseases of various levels. Additionally, a similar control group of healthy patients was analyzed. The whole database was divided into a learning and a verifying set.

4. Results

To find the optimal structure and parameters of WNS, the WT and NN layers were created, based on the following type structures [13]:

the WT layer:

- Morlet WBF,
- Db4 WBF,
- Bior 2.4 WBF,

the NN subsystem:

- 1LLinAF – 1-layer NN with a linear activation function (AF),
- 2LtangLinAF – 2-layer NN with 1st tangsoidal AF and 2nd linear AF,
- RBFLinFA I, II – a neural net with a radial basic function learnt according to a simple matching algorithm (I) and using the Gramm-Schmidt orthogonalization method (II) to find the optimal node number in the hidden layer.

4.1. WNNs parameter initialization

Formula (17) expresses the ranges of the WT layer scale parameter a and the corresponding frequency values obtained for the WNS used as a HRV signal classifier (a_{HRV}) structure. They were estimated on basis of an analysis of significant components included in the frequency spectrum of HRV signals characterizing the studied issues:

$$a_{\text{HRV}} \in \langle 5; 250 \rangle \equiv f_{\text{HRV}} \in \langle 0.01; 0.5 \rangle [\text{Hz}]. \quad (17)$$

Figure 5 presents the changes of scale a_{HRV} during the learning process as a function of epochs for various initial values. The results show that, in spite of taking the initial a_{HRV} values in the predefined range (17), they fall outside this bracket after the learning procedure, which disqualifies the learning process. Based on experimental results obtained for various initialization values taken from the ranges defined above,

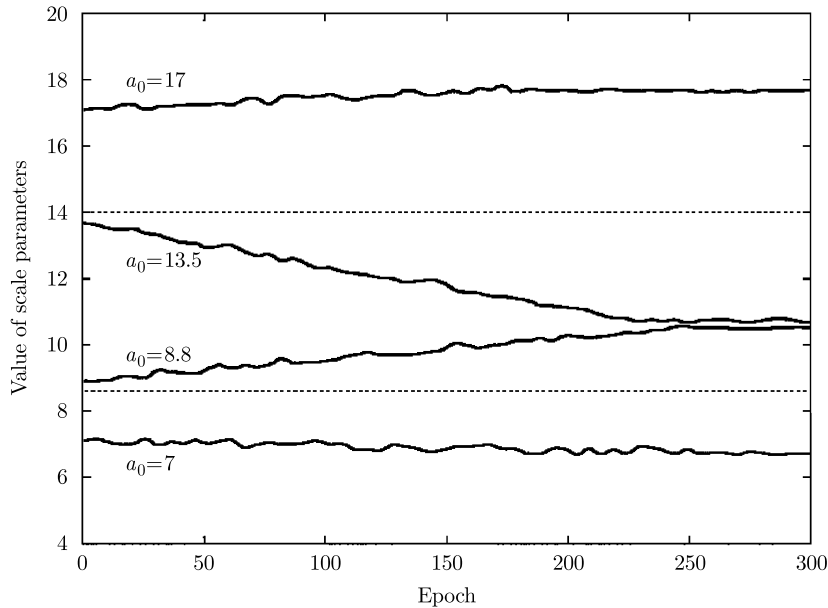


Figure 5. Changes of WT scale parameter, a , during the learning process for various initial values. WNS in HRV classifier structure

a narrowed range of a_{HRV} initialization values which guarantee convergence during the learning process was estimated as follows:

$$a_{HRV0} \in \langle 8.6; 14 \rangle \equiv f_{VP0} \in \langle 0.18; 29 \rangle [\text{Hz}]. \quad (18)$$

The remaining WNS parameters, weights $w_{i,j}$ of the NN structure part, were initialized according to the error back-propagation algorithm with random values.

4.2. Learning process evaluation

The discussed learning process was characterized by the value of normalized mean square error (NMSE) of the tested WNS structures.

Figure 6 presents the results of speed and convergence analysis of WNS during the learning process. The influence of the neural part's structure on the WNS's performance was analysed. A two-layer perceptron with a wavelet layer (2LtangLinAF) and a neural net with a radial basic function were given in the test.

Additionally, a typical perceptron without WT analysis was tested to verify the usefulness of introducing a wavelet layer (2LtangLinAF-no WT). Figure 6 shows the results obtained for the basic wavelet (Morlet function), application of which in the wavelet layer gave the best results.

4.3. Verification of wavelet neural system classifiers' performance

The described systems of classification of coronary artery diseases have been verified to find the optimal structure of the wavelet layers and the neural part of WNS. The measures of sensitivity (S) and specificity (SP), computed from the results obtained for HRV signals taken from the testing set, were assumed as the indexes of classifier evaluation. Figure 7 presents the results of this classification for various

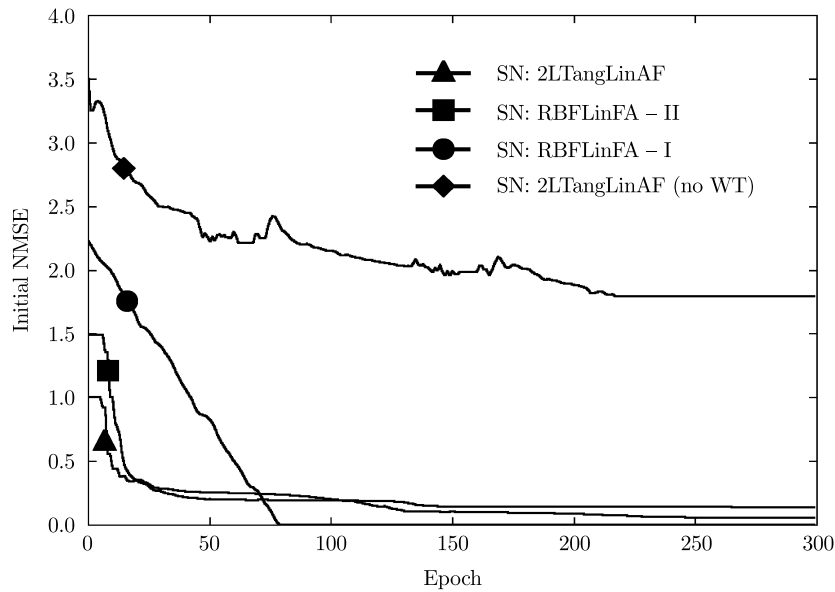


Figure 6. Normalised Mean Square Error of WNS for different type of neural part structure.
Wavelet function: Morlet

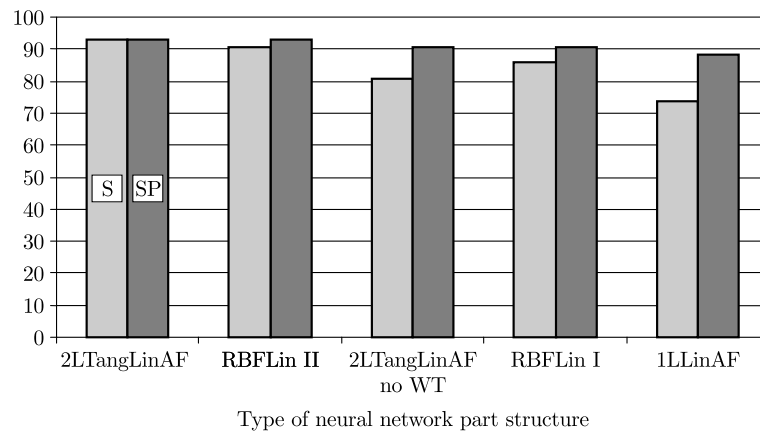


Figure 7. The measure of sensitivity (S) and specificity (SP) [%] of classifier for different type of WNS neural part, (mother wavelet of WT layer: Db4)

types of the neural part's structure, with a DB4 wavelet used in the time-frequency layer, for which the best results have been obtained.

5. Conclusions

The obtained results have shown that the described wavelet neural systems with different structures fulfil the role of unknown function approximators. On the example of clinical cases, the unknown relation between pathology of coronary arteries and heart rate variability signals taken from patients' ECG was mapped into the structure of wavelet neural systems.

The application of time-frequency decomposition in the preliminary stage of non-stationary heart rate variability signal analysis has improved the quality of

classifiers based on the neural network structure. The comparison study has allowed us to optimize the structure of the wavelet-neural system for the specific classification task. One of the most important stages of WNS learning is the initialization of the wavelet scale parameter, a . As distinct from randomly initialized neural network weights, this parameter should be chosen at the beginning of the learning process on the basis of the knowledge of the analysed problem (range of frequencies components) and experiments results.

The comparison of learning performance (Figure 6) has shown that the best quality of the learning process is achieved for a structure with a Morlet wavelet basic function used in the preliminary system layer.

The stage of WNS verification, presented in Figure 7, has allowed to characterize the generalization of classifier structures, which is one of the most important features of neural net-like systems. Although the systems with the Morlet wavelet achieved the smallest error during the learning process, our verification has yielded the highest values of sensitivity and specificity for systems using Db4 and Bior 2.4 wavelets in preliminary analysis.

The presented WNS cannot be treated as a universal structure ready to analyse a wide spectrum of issues. To achieve satisfactory results, the appropriate structure and the way of its parameter initialization should be worked out properly. Our WT scale a initialization results have shown that a correctly determined initial value is necessary to assure good convergence of the learning process. This stage of our WNS learning algorithms has been based on the analysis of the specific features of the analysed problem. The structure and learning process of a WNS should be adjusted according to the character of the considered problem.

References

- [1] Poggio T and Girosi F 1990 *Proc. IEEE* **78** (9) 1481
- [2] Hornik K, Stinchcombe M and White H 1989 *Neural Networks* **2** 359
- [3] Thakor N V and Sherman D L 1994 *Proc. IEEE-SP*, pp. 536–539
- [4] Zhang Q and Benveniste A 1992 *IEEE Trans. on Neural Networks* **3** (6) 889
- [5] Pati Y C and Krishnaprasad P S 1993 *IEEE Trans. on Neural Networks* **4** 73
- [6] Bakshi B R and Stephanopoulos G 1993 *AIChE J.* **39** 57
- [7] Chen J and Bruns D D 1995 *Ind. Eng. Chem. Res.* **34** 4420
- [8] Kostka P, Tkacz E, Nawrat Z, Wrzeźniowski A and Domider T 2001 *Proc. 6th Int. Conf. SYMBIOSIS*, Szczyrk, Poland, pp. 38–44
- [9] Tadeusiewicz R 1993 *Neural Networks*, Akademicka Oficyna Wydawnicza (in Polish)
- [10] Lawton W M 1990 *J. Math. Phys.* **31** (8) 1898
- [11] Mallat S 1989 *IEEE Trans. on Pattern Analysis* **7** (11) 674
- [12] Tkacz E J and Kostka P 2000 *Proc. World Congress on Medical Physics and Biomedical Engineering*, Chicago, CD-ROM
- [13] Kostka P and Tkacz E 2002 *Proc. 16th Int. Eurasip Conf. Biosignal*, Brno, Czech Republic, pp. 46–48

