neural networks, Kohonen network,
neurological diseases diagnosis,
Parkinson disease,
hemiparesis after ischemic stroke

Sławomir CHANDZLIK[*]

THE METHOD OF NEURON WEIGHT VECTOR INITIAL VALUES SELECTION IN KOHONEN NETWORK

Diagnosing of morbid conditions by means of automatic tools supported by computers is a significant and often used element in modern medicine. Some examples of these tools are automatic conclusion-making units of Parotec System for Windows (PSW). In the initial period of PSW system implementation, the units were used for recognition of orthopaedic diseases on the basis of the patient's walk and posture [15,17]. Subsequently, many additional options have been implemented, which have been used for purposes of diagnosing neurological diseases [1,2,3,9,12]. During automatic classification of diseases the additional units use elements of neural networks. The vectors based on normalised diagnostic measures [3] are inputs of the units. The measurements describe a patient's posture condition, his walk and overloads occurring on his feet. The Counter-Propagation (CP), two-layer network has been used in one of the automatic conclusion-making units. During CP network activity, we can see not only supervised but unsupervised learning processes as well. This is a characteristic feature of the CP network. The initial steps of the CP network learning process are very important, because the success of the network training process depends on them to a great extent. Therefore, a new method of weight vector initial values selection was proposed. The efficiency of the method was compared with classical methods. The results were very satisfactory. Owing to the proposed method, the time of the network training process as well as the mean-square error and the classification error was reduced. The research has been carried out using clinical cases of some neurological diseases: Parkinson's Disease, left-lateral hemiparesis and right-lateral hemiparesis after ischemic stroke. The measurements, which were made on a control group of patients without any neurological diseases, were the reference for these diagnostic classes.

1. INTRODUCTION

Computer data processing concerns many fields of scientific research. Modern medicine is one of the application areas especially connected with data processing and data analysis. An example of the application of advanced computer systems used in medicine is Parotec System for Windows (PSW), which uses measuring units and new technologies based on microprocessors. PSW was implemented during the years of 1991 to 2002 by the Polish-German research group [13,14,15]. Currently a German company: Paromed Medizitechnik produces the system in accordance with the European Union medicine directives. The additional options of PSW have been implemented since the beginning of production of PSW. The main aim of these options is helping doctors during diagnosing orthopaedic and neurological diseases [1,3,9,12]. During automatic conclusion these modules use components of artificial neural networks [1,9]. A Counter-Propagation (CP)

[*] University of Silesia, Institute of Informatics, Dept. of Computer Systems, Sosnowiec, Poland

network was implemented in one of the units. The CP network consists of two layers of neurons: the input layer, where the unsupervised training process is executed, and the output layer, where the supervised training process is performed. The described unit concludes automatically about the state of two diseases, which are difficult to diagnose: Parkinson's Disease and hemiparesis after ischemic stroke. Input data for the unit are input vectors based on normalized diagnostic measures [3]. The diagnostic measures describe the condition of the diseases. The diagnostic measures have been selected on the basis of pressure data recorded during static and dynamic measurements conducted with use of PSW hardware. An example of data visualisation has been presented in Fig. 1 and Fig. 2.

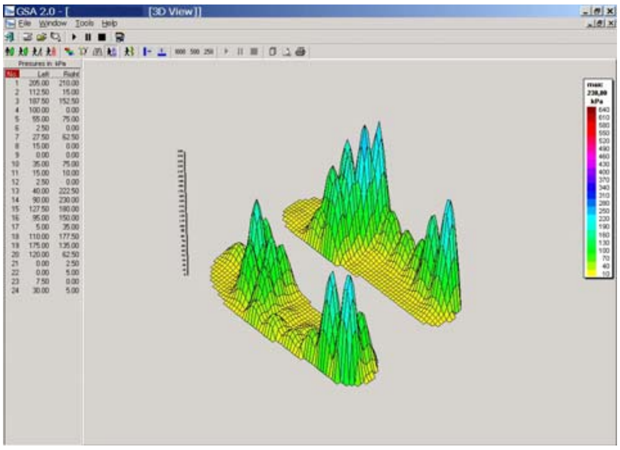

Fig. 1. 3D pressure map

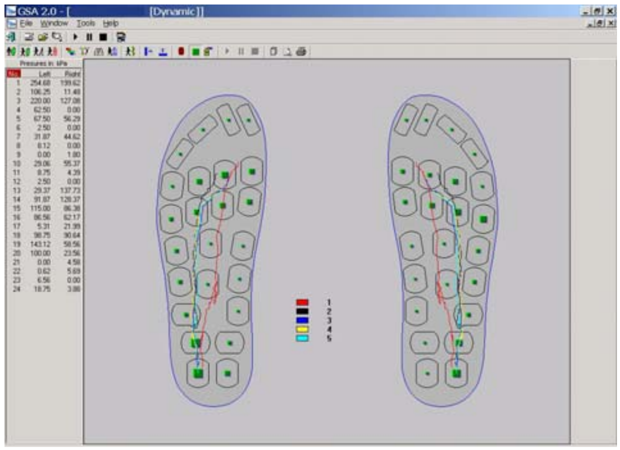

Fig. 2. The walk body balance trajectories recorded during dynamic measure

The investigations of using CP networks for purposes of diagnosis assistance have been carried out based on the analysis of four diagnostic classes: the control group, left-lateral hemiparesis, right-lateral hemiparesis and Parkinson's Disease (PD). In co-operation with medical personnel, 94 well-diagnosed data records have been selected, which could be used for diagnostic measure calculations. This set of well-diagnosed data records consists of: 30 data records of the control group, 28 data records of left-lateral hemiparesis, 26 data records of right-lateral hemiparesis and 10 data records of PD. The virtual data records have been also used during the neural network training process [4]. They have been generated on the basis of the clinical data records.

2. THE METHOD DESCRIPTION

A CP network consists of two layers of neurons. In the first input-adaptation layer, also called the Kohonen layer, an unsupervised training process using the Winner Takes All (WTA) algorithm was carried out. In the second output layer a supervised training process using the delta rule is executed. The initial steps of the unsupervised training process are very important, because the success of the network training process depends on the initial training steps. Inappropriate selection of initial values of weight vector in the Kohonen layer can cause termination of the training process of the non-linear neurons in local minima of the objective function [11,16]. In the consequence of these cases the training process fails, which is connected with absence of global minima of the objective function.

The basic assumption concerning normalization of all initial values of weight vectors has to be fulfilled in order to get an appropriate learning process in Kohonen networks [6,7]:

$$\left\| W_j^{(1)} \right\| = 1. \qquad (1)$$

Moreover, it is advisable, in the first steps of the algorithm, that the weight vectors are spread on a surface of a unitary sphere placed in m-dimensional space. This assumption is hard to obtain [16]. Moreover, it does not guarantee good terms for a network unsupervised training process [6,7,11].

This problem is often described in the literature concerning unsupervised training processes and many times the Convex Combination Method (CCM) is quoted as a solution for the problem [5,10,11]. The weight vectors generated by the CCM are initialised by the same values, which are obtained by the following formula:

$$w_{ji}^{(1)} = \rho \qquad (2)$$

where $\rho = \sqrt{\frac{1}{m}}$ is a constant defined by the number m of components of the weight vectors W_j.

the CCM fulfils a condition of vector normalisation (1), but it also causes that all weight vectors are covered by themselves, whereas the proposed new method gives vectors, which are more spread out over the unit sphere.

Moreover, the CCM adaptation requires, in the first steps of the training process, to pass the modified vector X' of the input vector X with components $x'^{(k)}_i$ determined according to the following formula [16]:

$$x'^{(k)}_i = \xi(k) x_i^{(k)} + \rho\left(1 - \xi(k)\right) \qquad (3)$$

where $\xi(k)$ is an adaptation function dependent on the k^{th} iteration step and ρ is the constant determined according to the formula (2).

The adaptation of the input vector X is necessary in the starting period of the training process. It makes possibility for expansion of co-linear – in the first steps – W_j weight vectors, in direction arisen from natural trends that appears in a set of input vectors X. The absence of adaptation or too short time of its use causes learning of insignificant number of neurons of the Kohonen network. The experiments with CP networks showed, that in some cases only 5% structure of input layer was well-learned. Wrong selection of the adaptation function $\xi(k)$ was the reason for the failures [1,8].

The adaptation function $\xi(k)$ values should be low at the beginning of training process and they should increase up till "1". After that, the $\xi(k)$ function becomes constant [11,16]. The experiments described in [1] showed that the minimum length of range in that the linear adaptations function $\xi(k)$ increasing is equal to $3m$, where m is the number of components of the input vector X.

2.1. MATHEMATICAL MODEL

The CCM is not optimum, though it keeps the learning process assumptions. So, a new method of initial values selection of weight vector was proposed. The new method also keeps assumptions of the CP network classical learning process. Let us define the components isolation function of a vector.

Definition 1.

A nonperiodic function $f_i(x)$, which projects the set of $X=\{x\colon 0\le x\le n\}$ on the set of $Y=\{y\colon 0\le y\le 1\}$, where: $n\in N^+$ is a natural number, and $\forall i, j\in N^+\ \exists x\in X\ f_i(x)\neq f_j(x)$, is called *the vector's components isolation function* (VCI). The natural number $i\in N^+$ is called *the number of the VCI function*.

According to Def. 1 we can define the i-th VCI function as:

$$f_i(x)=\frac{1}{2}\left(1-\cos\left(2\pi\frac{(x-1)}{n}-(m-1)\pi\frac{(i-1)}{m}\right)\right) \tag{4}$$

where: $x\in[1,\ n]$, $i=1,2,\dots,m$.

A set of VCI function graphs has been illustrated by Fig. 3. Five VCI functions are given by the number i (here from 1 to 5), and $x\in[1,100]$.

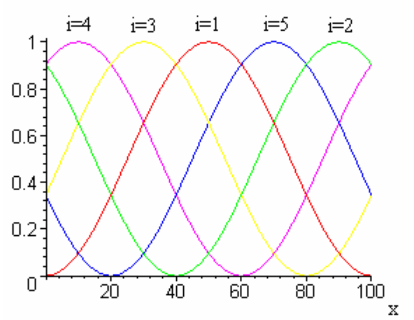

Fig. 3 A graph of five VCI functions

Let us assume that n is the number of $Z_j=[z_{ji}]$ m-dimensional vectors $j=1,2,\dots,n$. The values of z_{ji} are given by VCI functions (Equ. 4) according to the following formula:

$$z_{ji}=f_i(j) \tag{5}$$

where: $j=1,2,\dots,n$ is the number of the vector Z_j, $i=1,2,\dots,m$ is the component number of the vector Z_j (also the number of the VCI function).

The Z_j vectors have been shown in Table 1. Components of these vectors were chosen by VCI functions with following parameters: $n=4$, $m=5$.

Table 1. An example of 5-component vectors determined by VCI functions

vector no. \ component no.	1	2	3	4	5
1	[0.0000,	0.9045,	0.3456,	0.3456,	0.9045]
2	[0.5000,	0.2062,	0.9755,	0.0245,	0.7938]
3	[1.0000,	0.0955,	0.6544,	0.6544,	0.0955]
4	[0.5000,	0.7938,	0.0245,	0.9755,	0.2062]

The described method of the weight vector initial value selection takes on the normalization assumption of all initial weight vectors (1), according to the Kohonen networks' learning process. In this connection, input layer vectors were selected according to the normalization process, as:

$$W_j^{(1)} = \frac{1}{\|Z_j\|} Z_j; \quad j = 1,2,\ldots,n; \quad i = 1,2,...,m \tag{6}$$

where: n is the number of input layer neurons, m is the dimension of the Z_j vectors.

An example of $W_j^{(1)}$ weight vectors created by Z_j vectors (Equ. 6) is presented in Table 2.

Table 2. An example of weight vectors created by Z_j vectors and shown in Table 1

vector no. \ component no.	1	2	3	4	5
1	[0.0000,	0.6604,	0.2524,	0.2524,	0.6604]
2	[0.3651,	0.1506,	0.7124,	0.0178,	0.5796]
3	[0.7304,	0.0697,	0.4780,	0.4780,	0.0697]
4	[0.3651,	0.5796,	0.0178,	0.7124,	0.1506]

2.2. THE CHARACTERISTICS OF VCI FUNCTIONS

The vectors W_j chosen by Equ. 6, contrary to vectors selected by the CCM method, are distinct and distance between them has the same value provided that m is odd. They have also accomplished normalization assumption (1) and have decomposed the n–dimension unit surface evenly in a range defined by the input vectors' values. An example of vectors distribution on the unit sphere in 2-D and 3-D was illustrated by Fig. 4.

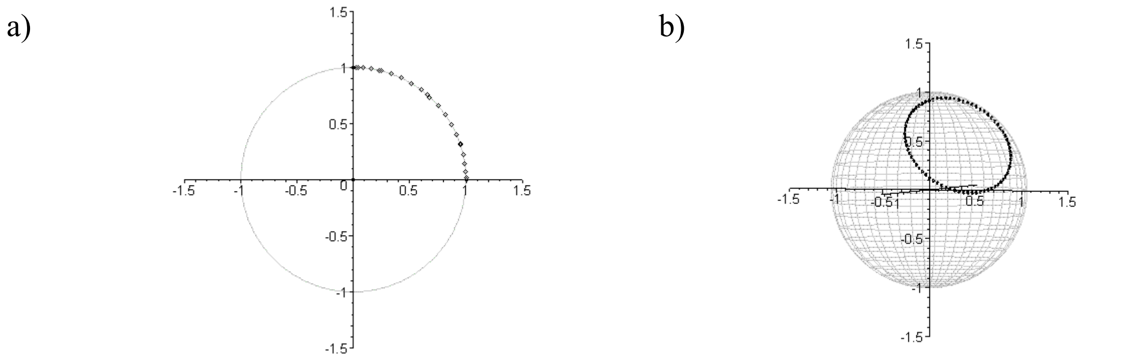

Fig. 4. An initial weight vector's dimension:
a) for 25 neurons in 2-D, b) for 75 neurons in 3-D.

An advantage of using the described method of initial values selection of weight vectors with using VCI functions is that it is no longer necessary to modify the input vectors X' (according to Equ. 3) for the input layer during the initial steps of the CP network learning process, so the VCI functions were used for establishing the initial values of the weight vectors $W_j^{(1)}$.

3. THE METHOD ANALYSIS AND COMPARISON WITH CCM

The method of the weight vector initial value selection with the use of VCI functions was implemented into the neural classifier unit. This unit is used for classifying four classes: Parkinson's Disease, left-lateral and right-lateral hemiparesis, and the control group (patients without diseases). The input data were the diagnostic measures [3] extracted from PWS data record. The diagnostic measures have been normalised. This means that they were independent of "external" elements such as insole size, patient's weight or measurement time. Extraction and aggregation of the diagnostic measures have been made by a pre-processor [4]. The pre-processor module was included in the PSW software package.

During method analysis, CP networks were used with following setup:

- input vector size $|X| = 67$,
- constant values of learning rates η, α: for the input layer: $\eta = 0.4$ and for the output layer: $\alpha = 0.15$,
- number of neurons of output layer: 4,
- number of neurons of Kohonen (input) layer: from 50 to 225,
- learning series size: N=51750.

The method was compared with the CCM method. During that, training process was carried out until two conditions occurred, defined in the following *finish learning criteria*.

Definition 2. ***Finish learning criteria***

One of three following conditions defines the finish learning criterion:

1. an arithmetic average of the mean-square error obtained during the W_S time period reaches value less than 0.02,

2. a learning process stabilisation appears; changes of an arithmetic average of the mean-square error are less than 2% in $3W_S$ time period,
3. a training process increases up to $N{\cdot}10^4$ iterations, where N is a learning series size.

The W_S time period was defined by following formula:

$$W_S = 5N \tag{7}$$

where N is a learning series size.

The training process was carried out ten times with use of 51750 elements of the learning series. Finally, the obtained results were averaged. The time of the finish learning criteria occurring was the primary factor estimated during the training process. An example of averaged mean-square errors obtained in the W_S=335 time period is illustrated by Fig. 5

3.1. THE RESEARCH RESULTS AND EFFICIENCY ESTIMATION OF THE METHOD

360 training tests were made for estimation of the weight vector initial value selection. The results, compared with the CCM method, have been presented below:

1. for 210 cases a shorter approach time of average value of mean-square error below 0.02 was obtained during the W_S=335 time period; this composes 58.3% of all tests,
2. for cases described in point 1 the approach time (described in point 1) was shorter from 0.1% to 98.5% (average 12.1%),
3. for 103 cases a longer approach time of average value of mean-square error below 0.02 was obtained in W_S=335 time period; this composes 28.6% of all tests,
4. for cases described in point 3 the approach time (described in point 3) was longer from 0.2% to 78.1% (average 4.6%),
5. for 47 cases the same approach time of average value of mean-square error below 0.02 was obtained in W_S=335 time period; this composes 13.0% of all tests,
6. for 211 cases a shorter time of stabilisation of training process was obtained with average value of mean-square error equal 0.018; this composes 58.6% of all tests,
7. for 142 cases a longer time of stabilisation of training process was obtained with average value of mean-square error equal 0.017; this composes 39.4% of all tests,
8. for 7 cases there was no stabilisation of training process; this composes 1.9% of all tests (the same result was obtained for CCM method),
9. for 73 cases a lower value of average value of mean-square error was obtained in the W_S=335 time period after the stabilisation process obtaining; this composes 20.3% of all tests,
10. for 24 cases a higher value of average value of mean-square error was obtained in the W_S=335 time period after the stabilisation process obtaining; this composes 6.7% of all tests,
11. for 263 cases the same value of average value of mean-square error was obtained in the W_S=335 time period after the stabilisation process obtaining; this composes 73.1% of all tests.

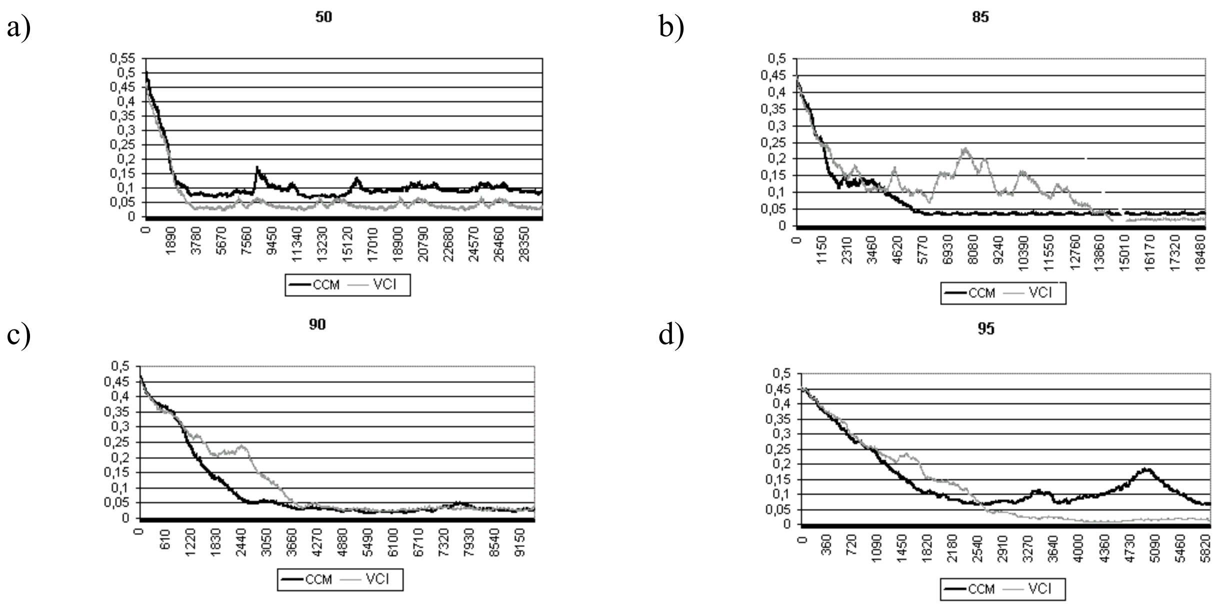

Fig. 5 The diagrams of average values of mean-square error appointed during time W_S=335 during the CP network training process – comparison of weight vector initial value selection in Kohonen layer: CCM method and VCI method (neuron numbers describe Kohonen layer):

a) 50 neurons – there was no stabilisation for each method,
b) 85 neurons – there was shorter time for stabilisation of the CCM method, but lower mean-square error was obtained for the VCI method at later times,
c) 90 neurons – there was stabilisation for both methods, but shorter time for stabilisation of the CCM method was obtained,
d) 95 neurons – there was stabilisation for the VCI method at a shorter time.

As we can see, the method described in the paper resulted in shortening the time for which low mean-square error below 0.02 was obtained, average about 12.1% tests in over half the cases. The comparison was performed by means of the CCM method. This fact directly influences the shortening of the training process. However, for less than 30% tests the time was longer but these cases occurred during testing of the Kohonen layer with a large number of the layer's neurons.

The stabilisation training process obtaining in shorter time than with use of CCM method appeared the additional advantage of the method in over half the number of cases. Moreover, after stabilisation the value of the mean-square error value was lower by about 24.4% on average (even about 66% in some cases).

4. CONCLUSIONS

A Counter-Propagation neural network was used for implementation of a neural classifier for neurological diseases. During the network activity, an unsupervised learning process has been performed in the Kohonen layer. The neural network was selected from three analysed networks: back-propagation, CP and ART (Adaptive Resonance Theory). An efficiency estimation of classification and training time determined the CP network selection. A new method for selection of initial values of weight vectors was used during the implementation of the CP network. The method allowed decreasing of classification error and training time reduction in comparison with the classic method for selection of initial conditions.

The neural classifier founded on CP network theory was implemented for neurological disease recognition based on the patient's gait and posture disturbances. However, the results obtained for diagnoses about strokes and Parkinson's Disease can be transposed on other diseases as well. However, the main problem is making accurate clinical measurements and selecting adequate characteristic features for these diseases. Selection of the main parameters of the neural classifier is also necessary; it depends on input vector size and number of output classes. For each case, the network parameters selection, such as: initial values of weight vectors, learning rates, etc., is very important. Unsuitable selection of these parameters can be the reason for learning process termination at local minima of the objective function and for failure of the classifier selection process in consequence.

BIBLIOGRAPHY

[1] Chandzlik S., Kopicera K.: Experiments with neural network parameters – selection for foot abnormalities recognition, Journal of Medical Informatics & Technologies. Vol. 5, pp: CS-71 – CS-78. ISBN 83-909517-2-7, 2000.

[2] Chandzlik S., Piecha J.: A patient walk-data-record modelling using a spline interpolation method. Journal of Medical Informatics & Technologies. Vol. 3, pp: MIT-153 – MIT-160. ISSN 1642-6037, 2002.

[3] Chandzlik S., Piecha J.: The body balance measures for neurological disease estimation and classification. Journal of Medical Informatics & Technologies, Vol. 6, pp: IT-87 – IT-94, 2003.

[4] Chandzlik S., Piecha J.: the gait characteristic data spectrum extraction. Proc. 4th Inter. Conf. on Computer Recognition System CORES'05, Vol. 18, pp. 493 – 501, 2005.

[5] Floater M.S.: Parameterization and smooth approximation of surface triangulations. Comp. Aided Geom. Des., Vol. 14, pp: 231-250, 1997.

[6] Hecht-Nielsen R.: Counterpropagation networks. Applied Optics, Vol. 26, pp: 4979-4984, 1987.

[7] Hecht-Nielsen R.: Applications of counterpropagation networks. Neural Networks, Vol. 1, pp: 131-139, 1988.

[8] Kopicera K., Piecha J.: The fault analysis made by PSW data recorder for neurological disease classification, Journal of Medical Informatics & Technologies, Vol. 4, pp: SN-10 – SN-13, 2002.

[9] Kopicera K., Piecha J., Zyguła J.: The neural networks in diagnostics support for PSW system. Proc. of Int. Conference ASIS'99, pp. 113-118, Krnov, 1999.

[10] Levin D., Nadler E.: Convexity preserving interpolation by algebraic curves and surfaces. Numerical Algorithms, Vol. 9, pp: 113-139, 1995.

[11] Osowski S.: Sieci neuronowe w ujęciu algorytmicznym. WNT, Warszawa, 1996.

[12] Piecha J.: The neural network conclusion-making system for foot abnormality recognition. Proceedings of IMACS World Congress, Lausanne, Switzerland, August 2000.

[13] Piecha J., Kopicera K.: The conclusion making method using pathology classifiers, Proc. on KOSYR'01, pp: 29-33, 2001.

[14] Piecha J., Zyguła J.: PC visual interface for orthopaedic expertise, Proc. of Int. Conference, pp. 162-167, 1995.

[15] Piecha J., Zyguła J., Łyczak J., Gaździk T., Proksa J.: The advanced measuring system for orthopaedic pathologies diagnostics using a static and dynamic footprints, Chirurgia narządów ruchu i ortopedia polska vol. LXI 1996, suplement 3B, pp.119-124. (in Polish)

[16] Tadeusiewicz R.: Sieci neuronowe. Akademicka Oficyna Wydawnicza RM, Warszawa, 1993.

[17] Zyguła J.: Przetwarzanie danych pomiarowych dla systemu wnioskowania o patologiach w obszarze stopy. Rozprawa doktorska, Gliwice 1997. (in Polish)