Article

Data Preprocessing and Neural Network Architecture Selection Algorithms in Cases of Limited Training Sets—On an Example of Diagnosing Alzheimer's Disease

Aleksandr Alekseev [1,2,*], Leonid Kozhemyakin [2,3], Vladislav Nikitin [2,4] and Julia Bolshakova [4,5]

1   Administrative Directorate for Organization of Scientific Researches, Perm National Research Polytechnic University, Perm 614990, Russia
2   Perm Decision Making Support Center, Perm 614107, Russia
3   Department of Economics and Finance, Perm National Research Polytechnic University, Perm 614990, Russia
4   Department of Computational Mathematics, Mechanics and Biomechanics, Perm National Research Polytechnic University, Perm 614990, Russia; ybolshakova@pstu.ru
5   Diagnosing Systems, Perm 614101, Russia
*   Correspondence: aoalekseev@pstu.ru

Abstract: This paper aimed to increase accuracy of an Alzheimer's disease diagnosing function that was obtained in a previous study devoted to application of decision roots to the diagnosis of Alzheimer's disease. The obtained decision root is a discrete switching function of several variables applicated to aggregation of a few indicators to one integrated assessment presents as a superposition of few functions of two variables. Magnetic susceptibility values of the basal veins and veins of the thalamus were used as indicators. Two categories of patients were used as function values. To increase accuracy, the idea of using artificial neural networks was suggested, but a feature of medical data is its limitation. Therefore, neural networks based on limited training datasets may be inefficient. The solution to this problem is proposed to preprocess initial datasets to determine the parameters of the neural networks based on decisions' roots, because it is known that any can be represented in the incompletely connected neural network form with a cascade structure. There are no publicly available specialized software products allowing the user to set the complex structure of a neural network, which is why the number of synaptic coefficients of an incompletely connected neural network has been determined. This made it possible to predefine fully connected neural networks, comparable in terms of the number of unknown parameters. Acceptable accuracy was obtained in cases of one-layer and two-layer fully connected neural networks trained on limited training sets on an example of diagnosing Alzheimer's disease. Thus, the scientific hypothesis on preprocessing initial datasets and neural network architecture selection using special methods and algorithms was confirmed.

Keywords: integrated rating mechanisms; decisions' roots; criteria trees; convolution matrices; data preprocessing; system-cognitive analysis; neural network structure selection; neural network training

1. Introduction

Data mining and machine learning methods are used in various areas including engineering, business and management, economics and finance, medicine, etc.

Artificial neural networks are actively used in the medical studies [1,2], etc., for example in diagnosing cardiovascular diseases [3–5], intestinal dysbacteriosis [6], oncology [7–12], in modeling of the prognosis of the outcome of the disease, and the likelihood of tumor recurrence [13–15]. In [16], it is proposed to use convolutional neural networks for automatic segmentation of edema in patients with an intracerebral hemorrhage. In [17], deep learning models are used to analyze the complication of cerebral edema induced by radiation therapy in patients with an intracranial tumor.

A feature of medical data is their limitation. This fact is partly due to ethical requirements, i.e., patients or their legal representatives must consent to the use of their medical

data for scientific purposes and publication in open press. For example, researchers [18] took 5 years to collect data about 81 subjects, including 59 subjects with clinically diagnosed Alzheimer's disease, and 22 elderly subjects did not have problems in cognitive abilities. This limited dataset was chosen for performance verification of neural network training after preprocessing initial datasets and neural network architecture selection using special methods and algorithms on an example of diagnosing Alzheimer's disease. The key problem of Alzheimer's disease is an accurate diagnosis in the early stages, where cognitive impairments do not yet appear at all or are not noticeable to the patient and others.

1.1. Related Papers

The key conclusions of some works of the last two decades [19–22] devoted to the study of the brain are the growth of diseases associated with Alzheimer's disease, which does not immediately have distinctive signs for its identification. As shown in [23–25], there is a connection between oxygen saturation of brain tissues and the level of its blood circulation, qualitatively expressed through cerebral metabolism, with the symptoms of Alzheimer's disease.

In [26], decrease of the dopaminergic activity of the ventral tegmental area (located in the midbrain area) may be of decisive importance for the earliest pathological symptoms of Alzheimer's disease. This was verified by data analysis of resting structural and functional magnetic resonance imaging and neuropsychological assessments in 51 healthy adults. Mild cognitive impairments were diagnosed for 30 patients, and Alzheimer's disease was diagnosed for 29 patients. A significant functional relationship was also noted between the ventral tegmental area and the hippocampus in pathological symptoms of Alzheimer's disease. Changes in the dopaminergic system are often noted among patients with Alzheimer's disease and are usually associated with cognitive and noncognitive symptoms [27].

In [28], it is shown that most dopamine-producing neurons are localized in the midbrain. Their loss is associated with some of the most famous human neurological disorders—Parkinson's disease and cognitive impairments. In [29,30], the importance of dopamine is noted in participation in the mediated reaction of the body to the environment on different time scales (from fast impulse responses associated with rewards to slower changes with uncertainty, punishment, etc.). In [29], it is shown that the human brain contains about 200,000 dopamine neurons on each side of the brain. In [31–33], it is noted that cerebrovascular diseases largely stimulate the more rapid development of amnestic moderate cognitive impairments and Alzheimer's disease and are also characterized by more severe damage to neuronal connections. In [34], the fact of increased cerebral blood flow without changing the oxygen consumption of the brain is described with the use of acetazolamide. This changes the signal observed on MRI images with a gradient echo from areas with increased cerebral blood flow. There is also a relationship between the oxygen content in the blood and the release of dopamine and neurotransmitters in [35,36]. In [37], the dependence of the level of dopamine and the oxygen content into the blood are noted. Since the neuron activity depends on the degree of oxygen uptake, the assessment of the degree of venous blood saturation can become an indicator of the functional state of neurons. Accordingly, we chose the following veins of all involved in the brain blood supply: the basal veins involved in the blood circulation of the striatum and veins of the thalamus (these veins drain the blood from the thalamus, which is closely connected with the midbrain [38]).

In [39], there is a fairly pronounced relationship between quantitative magnetic resonance imaging (MRI) and spectroscopy (MRS) measurements of energy metabolism (the rate of oxygen metabolism in the brain, CMR (O_2)), the blood circulation (i.e., cerebral blood flow (CBF) and volume (CBV)) and functional MRI signal over a wide range of neuronal activity. Pharmacological treatments are used to interpret the neurophysiological basis of levels of the blood oxygenation dependent on image contrast at 7 T in glutamatergic neurons in the rat cerebral cortex.

1.2. Research Statement

The motivation for the present paper was continuation of the previous study [40] on the application of decisions' roots to diagnosis of Alzheimer's disease. "Decision root" is a new name for the integrated rating mechanisms traditionally applied to aggregation of a few indicators into one integrated assessment [41]. The new term was proposed by Dr. Habil. Korgin, Nikolay A. and his Ph.D. student Sergeev, Vladimir A. [42] in November 2021. Previously, the term "integrated rating mechanisms" was used.

Historically, "the integrated rating mechanisms were introduced as multidimensional assessment and ranking systems for management and control in organizational and manufacturing systems in the Soviet Union in earlier 80th of the previous century" [43] (p. 610). Nowadays, these systems are used in different areas, including medicine [40]. Decision support system for diagnosis of Alzheimer's disease is developed at the LLC "Diagnosing systems".

Mathematically, the integrated rating mechanism is a discrete switching function of several variables presenting as a superposition of a few functions of two variables. Therefore, it can be represented in a hierarchical form, which determines the sequence of operations on variables. Each discrete function of two variables can be represented in matrix form. The tuple of criteria, binary trees and convolution matrices, located in its nodes, gives a graphical representation of such a function. In [44] (p. 1) it was said "integrated rating mechanism belongs to the class of so-called verbal decision analysis approaches". Verbal decision analysis has been actively used for unstructured decision-making problems [45].

In 2020, two approaches to identification of integrated rating mechanisms based on a training set were suggested. The first one was proposed by Professor, Dr. Habil Burkov, Vladimir N., Dr. Habil Korgin, Nikolay A. and Sergeev, Vladimir A. [44]. This one-hot encoding approach considered the integrated rating mechanism as a sequence of matrix operations with one-hot encoded indicator values.

The second approach was proposed by one of the authors of this paper—Alekseev, Aleksandr O. and was based on truth table transformation [46], therefore, it was called the tabular algorithm for the identification of an integrated rating mechanism [47] (p. 599).

In 2021, Sergeev, Vladimir A. and Korgin, Nikolay A. [48] proposed to investigate the integrated rating mechanism identification problem on an incomplete dataset using equivalence groups. Approaches [46,47] and [48] are similar, but not the same.

In paper [46] (p. 402) it was shown that any integrated rating mechanism can be represented in the form of an artificial neural network. An identified neural network reproduces the training data with discrete values.

The obtained function, by using a tabular algorithm for identification [40], establishes the relationship between Alzheimer's disease and the magnetic susceptibility value of the basal veins and veins of the thalamus. However, the accuracy of the function obtained in a previous study [40] was near 84%, which is not enough for its application in medical practice.

To increase accuracy, the idea was suggested to apply an artificial neural network to approximate the relationship between the magnetic susceptibility value of cerebral veins and Alzheimer's disease. However, as previously mentioned, a feature of medical data is their limitation. In such cases, neural networks based on limited training datasets may be inefficient.

To solve this problem, it is proposed to preprocess initial datasets to determine the parameters of the neural networks based on the integrated rating mechanism (decision's root). The scientific hypothesis of this paper is that it will be possible to obtain the acceptable accuracy of the neural network by training a neural network on the initial data in a continuous form.

Thus, the purpose of this work is performance verification of neural network training in cases of limited training sets on the example of diagnosing Alzheimer's disease.

The paper is organized as follows: Section 2 provides initial data, methods and algorithms used in this study according to suggested methodology; Section 3 provides

results of preliminary data analysis based on interval coding and identifies significant indicators of the study area. The architecture of the neural network based on the found decision's root is proposed, which is the key result in this study, and the significance of the results is beyond doubt for development and discussion in future works; Section 4 is devoted to a discussion of the obtained results and the insights of this work. Lastly, the main results of this paper and future research directions are briefly discussed in Section 5.

2. Materials and Methods

2.1. Initial Data

In reference [49], it was shown that synthetic data can be used for neural network training, but in this paper, real medical data were taken for performance verification of neural network training based on preprocessing initial datasets and neural network architecture selection using special methods and algorithms.

The initial data were tomographic images of 81 patients examined using the minimental state examination (MMSE), Montreal cognitive assessment (MoCA), clock drawing task (CDT), and activity of daily living scale (ADL). Alzheimer's disease was clinically diagnosed in 59 subjects (21 males and 38 females). These patients are hereinafter defined as in [18] as the AD group. Twenty-two patients (12 males, 10 females) did not have problems in cognitive abilities and were placed in the control group (defined as CON group).

The magnetic susceptibility values (MSV) of the following cerebral veins were quantified based on tomographic images: left and right basal veins (L_BV and R_BV, respectively), left and right internal cerebral veins (L_ICV, R_ICV), left and right veins of the thalamus (L_TV, R_TV), left and right septal veins (L_SV, R_SV), left and right veins of the dentate nucleus (L_DNV, R_DNV).

In this study, we exclusively used numerical data given in the appendix to the reference [18]. A fragment of the initial data is presented in Table 1.

Table 1. The fragment of the initial data with magnetic susceptibility values of the cerebral veins [18].

Subject	Group [1]	L_BV	R_BV	L_ICV	R_ICV	L_TV	R_TV	L_SV	R_SV	L_DNV	R_DNV
sub001	1	279	288	255	263	140	138	131	131	165	185
sub002	1	274	247	223	243	239	262	190	222	204	102
sub003	1	259	333	236	243	172	159	135	145	152	153
...	...	...	...	...	...	...	...	...	...	...	...
sub079	0	279	288	179	189	141	153	245	214	177	149
sub080	0	249	232	221	232	154	142	165	151	161	143
sub081	0	300	259	295	299	216	197	163	131	162	129

[1] In this Table in column "Group" 1 is the AD group, and 0—the CON group.

2.2. Methodology and Methods

Data preprocessing consists in interval coding of initial data and search of integrated rating mechanisms (decisions' roots). In paper [46] (p. 402), it is shown that any integrated rating mechanism can be represented in the form of an artificial neural network. Thus, it is possible to determine some of the parameters of the required neural network based on the integrated rating mechanism (decision's root). This mechanism reproduces the encoded data using discrete values. This idea corresponds to the approach presented in reference [50]. The scientific hypothesis lies in the assumption that it will be possible to obtain the acceptable accuracy of the neural network by training a predefined neural network on the initial data in a continuous form.

According to [46] (p. 398) the integrated rating mechanism is determined as a tuple (1).

$$< G, M, X, P >, \tag{1}$$

where G is a graph describing a sequence of convolution of particular criteria, M is a set of convolution matrices corresponding to the nodes of a criteria tree, X is a set of the scales for

the scoring of the particular criteria, and P is the procedure of aggregation. In this study, discrete procedure is used, but continuous, interval, fuzzy and F–fuzzy procedures are also noted [51].

One of the problems of identifying integrated rating mechanisms (decisions' roots) is the choice of the structure of the graph G, which is a full binary tree with labeled leaves [52], because their total number is determined according to the equation [43,48]:

$$|G| = (2l - 3)!!, \qquad (2)$$

where *l* is the total number of considered variables (leaves in terms of tree graph).

The number of G structures for 4 variables is 15, for 5–105, for 6–945 as can be seen from Equation (2), and there are already more than 34 million, for example, for 10 variables. This is a key limitation of the proposed algorithms, as can be seen from the above example.

To reduce the number of variables, it is suggested to use methods of the system-cognitive analysis (hereinafter–SCA) [53,54] using the computer program RU 2022615135 "Personal intelligent online development environment "EIDOS-X Professional" (System "EIDOS-Xpro")" [55], developed by professor Lutsenko, Evgenue V. In this computer program, to solve the classification problem, two integrated similarity criteria are used, known as "Sum of knowledge" (3) and "Semantic resonance of knowledge" (4).

$$I_j = \sum_{i=1}^{I} L_i I_{ij}, \quad L_i = \{0;1\}, \qquad (3)$$

where I_j—is the amount of information about class *j*; *I*—the amount of features (signs, attributes); L_i—a variable describing the presence ($L_i = 1$) or absence ($L_i = 0$) of the feature *i*; I_{ij}—assessment of informational importance of feature *i* for class *j* (amount of information about feature *i* for class *j*).

$$I_j = \frac{1}{\sigma_j \sigma_l M} \sum_{i=1}^{I} \left(I_{ij} - \overline{I_j}\right)\left(L_i - \overline{L}\right), \quad L_i = \{0;1\} \qquad (4)$$

where *M*—number of gradations of descriptive scales (features); σ_j—standard deviation particular criteria for class vector knowledge; σ_l—standard deviation along the vector of the recognized object; $\overline{I_j}$—average informativeness for class vector; *L*—average for object vector.

In the case of diagnosing Alzheimer's disease, index *j* is the ordinary number of patient group {group AD, group CON}, *j* = 1 correspond to group AD, *j* = 2 correspond to group CON. Index *i* is the ordinary number of the intervals of MSV of cerebral veins. We have 10 veins, each of which is divided into 3 equal intervals, so the total amount of MSV intervals is 30 (*I* = 30).

As a result of the implementation of the SC-analysis methods, the following model of determination informational importance (5) were selected as the most reliable.

$$I_{ij} = N_{ij} - \frac{N_i N_j}{N}, \qquad (5)$$

where N_{ij}—the number of observations of the *i*-th feature (sign) for objects of the *j*-th class in the training sample; N_i—the total number of observations of the *i*-th feature over the entire training sample; N_j—the total number of features for objects of the *j*-th class in the training set; *N*—the volume of the training sample (number of observations).

These methods have been used to identify the most significant brain veins and subsequent reduction of the number of indicators.

The algorithm for identification of integrated rating mechanisms [47] has been used for data structuring and representation as a decision root. These algorithms were suggested previously by one of the authors of the present study–Alekseev, Aleksandr O.

A specialized computer program called "Neurosimulator 5.0" was used for neural network modeling [56]. It was developed by professor Yasnitsky, Leonid N. and his Ph.D student, Cherepanov, Fedor M.

For training the neural network, the computer program "Software package that implements the operation of incompletely connected neural networks" [57] was tested, but this program does not implement the conjugate gradient method. The application of this program is shown in the work [58].

Through inputs, the mathematical neuron receives input signals, which it sums up by multiplying each input signal by the appropriate weighting factor $w_{i1}, w_{i2}, \ldots, w_{ij}$:

$$s_i = \sum_{j=1}^{Nn} w_{ij} \cdot x_{ij}, \tag{6}$$

where Nn—the number inputs of i-th neuron; x_{ij}—signals on the input of the i-th neuron; w_{ij}—weights obtained in each iteration when performing neural network training.

Next, the activation function acts, which is usually an arbitrary monotonically increasing function that takes real values in the range from 0 to 1. The activation threshold is a number from the same interval.

$$y_i = f(s_i) = f\left(\sum_{j=1}^{Nn} w_{ij} \cdot x_{ij} + \theta_i\right) = f\left(\mathbf{w}^T \cdot \mathbf{x}\right), \tag{7}$$

where y_i—output signal; θ_i—neural bias b is interpreted as the weight of an additional input with a synaptic link strength w_0, whose signal $x_0 = 1$; f—activation function.

The general task of training a neural network is as follows:

Let X be a set of objects; R the solution of the algorithm, and then there is an unknown objective function $G' : X \rightarrow R$. It is necessary to find such a primary function $G : X \rightarrow R$ that restores the assessment G'. According to the generally accepted neural network design technology, the entire general population is divided into training, testing and confirming sets in the ratio of 70%: 20%: 10% [59].

Based on a set of logical pairs $\mathbf{d}_n' = (\mathbf{x}_n, \mathbf{r}_n)$, where $\mathbf{d}_n'$ is n-th precedent, we will distinguish three relevant subsets of precedents $D^{Tr} = \{\mathbf{d}_n'\}_{n=1}^{N_{Tr}}$—Train Set, $D^{Ts} = \{\mathbf{d}_n'\}_{n=1}^{N_{Ts}}$—Test Set, $D^{Tv} = \{\mathbf{d}_n'\}_{n=1}^{N_{Tv}}$—Validation Set, provided that data loss is excluded $D^{Tr} \cap D^{Ts} \cap D^{Tv} \equiv \varnothing$ [60] (p. 426).

According to the methodology presented in Figure 1, the training of a non-fully connected neural network will be carried out by the method of conjugate gradients [61], which makes it possible to find the extrema of the function by iterative calculation. The conjugate gradient vector is determined by the following formula:

$$\rho_k = -\nabla E(w_k) + \beta_k \rho_{k-1}, \tag{8}$$

where the choice of the parameter β_k responsible for the conjugate direction can be determined according to the Fletcher and Reeves algorithm [62] (10) or algorithm by Polak and Ribiere (11).

$$\beta_k = \frac{\nabla E(w_k)^T \nabla E(w_k)}{\nabla E(w_{k-1})^T \nabla E(w_{k-1})}, \tag{9}$$

$$\beta_k = \frac{(\nabla E(w_k) - \nabla E(w_k))^T \nabla E(w_k)}{\nabla E(w_{k-1})^T \nabla E(w_{k-1})}, \tag{10}$$
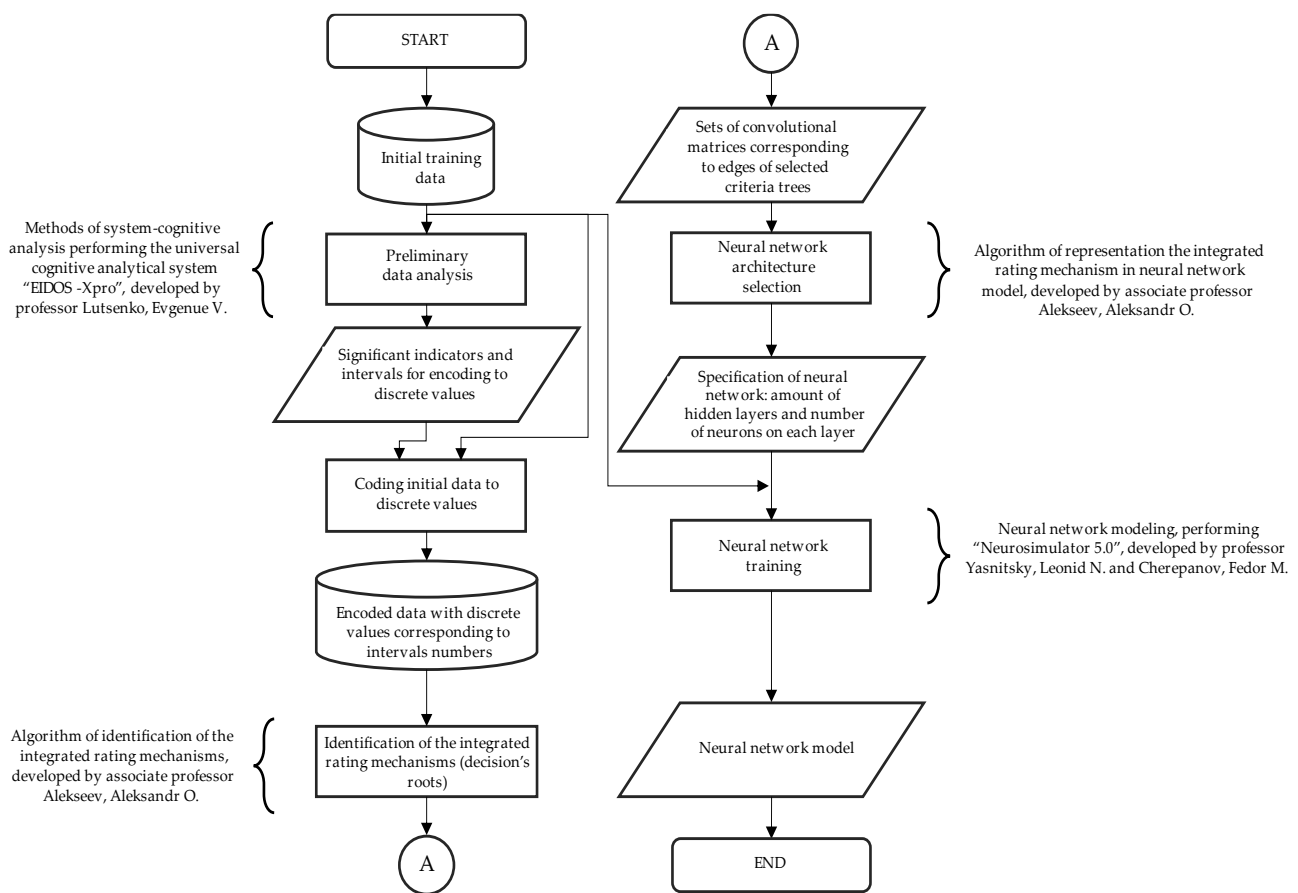
Figure 1. Suggested algorithm of data preprocessing and neural network architecture selection for neural network training in limited initial dataset cases.

Two directions ρ_k and ρ_{k-1} are defined as conjugate if the following condition is met:

$$\rho_k{}^T H \rho_{k-1} = 0, \tag{11}$$

In other words, conjugate directions are orthogonal directions in the space of an identity Hessian matrix.

We chose the algorithm of elastic backpropagation (RPROP—Resilent back PROPagation) to train classic single-layer two-layer hidden artificial neural networks with a different number of neurons on each hidden layer. According to this algorithm, when correcting the weight coefficients, only the sign of the gradient matching [63]:

$$\Delta w_{ij}(t) = -\eta(t) sign\left(\frac{\partial \varepsilon(t)}{\partial w_{ij}}\right), \tag{12}$$

The methodology of this study is presented below (Figure 1).

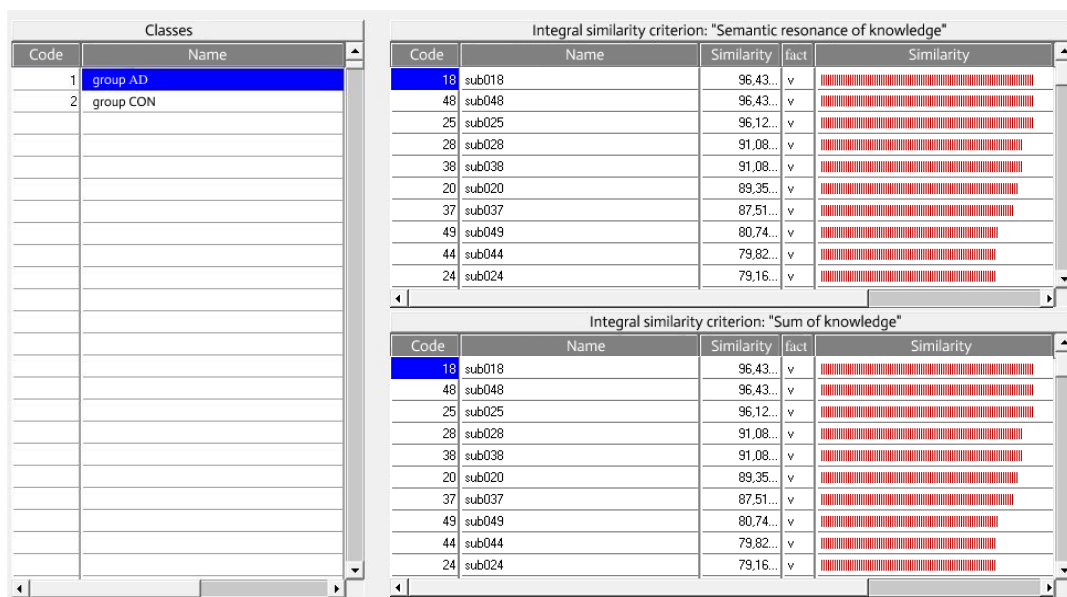## 3. Results

3.1. Results of the Preliminary Data Analysis

The initial data were processed using SCA methods. The intervals of magnetic susceptibility values of the above-described veins L_BV, R_BV, L_ICV, R_ICV, L_TV, R_TV, L_SV, R_SV, L_DNV, R_DNV were used as features. The state of the patient was used as a class: AD group or CON group, and the domains of observed values were divided into Three equal intervals for each vein (Table 2). Such a division of the initial dataset into three intervals was due to expert opinion, since when choosing two intervals, we get the classical problem of finding a Boolean function. The choice of four intervals is

inappropriate, since when using 10 variables in the initial data, 80 features will be formed, and the initial dataset, which must be divided into training and testing, contains only 81 examples. Of interest is the division of the observation area into different intervals with an almost uniform number of examples in each of these intervals. In view of the multivariate results of possible partitions, this approach will be considered in a future study.

Table 2. The intervals for discrete coding.

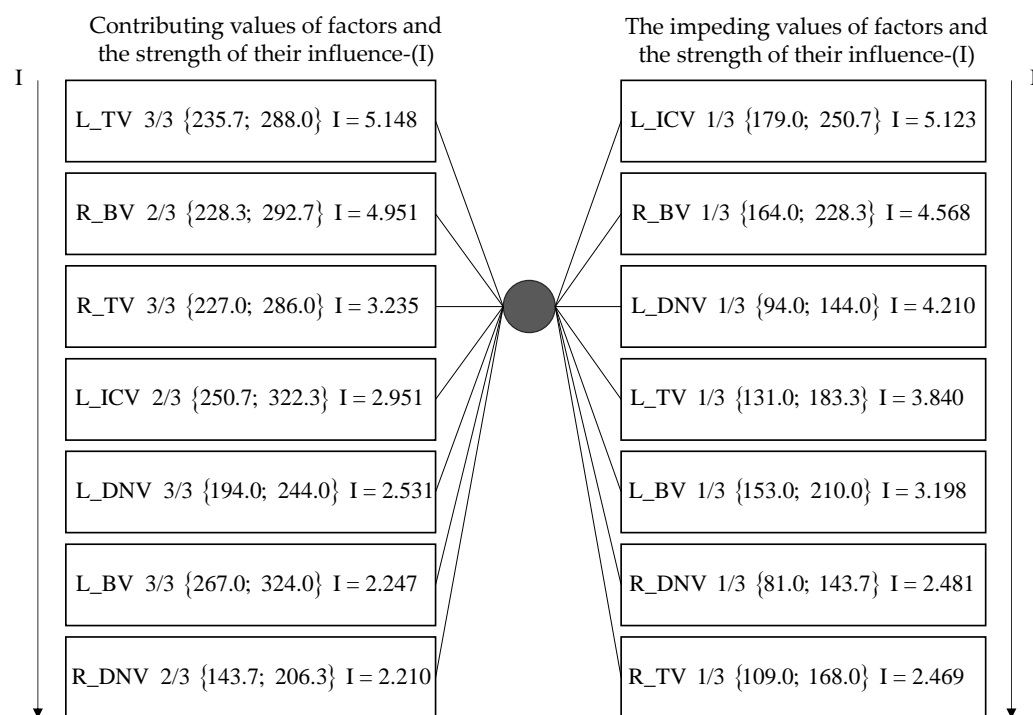| Cerebral Veins	Variable Abbrev.	Domain of MSV	1st Interval	2nd Interval	3rd Interval
left basal vein	L_BV	{153.0; 324.0}	{153.0; 210.0}	{210.0; 267.0}	{267.0; 324.0}
right basal vein	R_BV	{164.0; 357.0}	{164.0; 228.3}	{228.3; 292.7}	{292.7; 357.0}
left internal cerebral vein	L_ICV	{179.0; 394.0}	{179.0; 250.7}	{250.7; 322.3}	{322.3; 394.0}
right internal cerebral vein	R_ICV	{163.0; 411.0}	{163.0; 245.7}	{245.7; 328.3}	{328.3; 411.0}
left vein of the thalamus	L_TV	{131.0; 288.0}	{131.0; 183.3}	{183.3; 235.7}	{235.7; 288.0}
right vein of the thalamus	R_TV	{109.0; 286.0}	{109.0; 168.0}	{168.0; 227.0}	{227.0; 286.0}
left septal vein	L_SV	{83.0; 310.0}	{83.0; 158.7}	{158.7; 234.3}	{234.3; 310.0}
right septal vein	R_SV	{73.0; 287.0}	{73.0; 144.3}	{144.3; 215.7}	{215.7; 287.0}
left vein of the dentate nucleus	L_DNV	{94.0; 244.0}	{94.0; 144.0}	{144.0; 194.0}	{194.0; 244.0}
right vein of the dentate nucleus	R_DNV	{81.0; 269.0}	{81.0; 143.7}	{143.7; 206.3}	{206.3; 269.0}

With the help of the intelligent analytical system "EIDOS-Xpro", all patients were distributed based on the level of progression of Alzheimer's disease by applying two integral similarity criteria: "Semantic resonance of knowledge" (4) and "Sum of knowledge" (3) (Figure 2).

Figure 2. Recognition results for the AD group of patients in the "Class-objects" mode in the system "EIDOS-Xpro".

The intervals of MSV were determined as a result of the application of SCA. These values had the greatest informational significance for classification of a patient in the AD group (Figure 3, see "contributing"), as well as signs that did not characterize this class to the greatest extent (Figure 3, see "impending"). Since patients belong to only two classes (AD and CON) in the research dataset, the insignificance for the AD group meant significance for the CON group, respectively.

Contributing values of factors and the strength of their influence-(I)

The impeding values of factors and the strength of their influence-(I)

L_TV 3/3 {235.7; 288.0} I = 5.148

R_BV 2/3 {228.3; 292.7} I = 4.951

R_TV 3/3 {227.0; 286.0} I = 3.235

L_ICV 2/3 {250.7; 322.3} I = 2.951

L_DNV 3/3 {194.0; 244.0} I = 2.531

L_BV 3/3 {267.0; 324.0} I = 2.247

R_DNV 2/3 {143.7; 206.3} I = 2.210

L_ICV 1/3 {179.0; 250.7} I = 5.123

R_BV 1/3 {164.0; 228.3} I = 4.568

L_DNV 1/3 {94.0; 144.0} I = 4.210

L_TV 1/3 {131.0; 183.3} I = 3.840

L_BV 1/3 {153.0; 210.0} I = 3.198

R_DNV 1/3 {81.0; 143.7} I = 2.481

R_TV 1/3 {109.0; 168.0} I = 2.469

Figure 3. The most significant factors influencing a possible brain pathology in the form of Alzheimer's disease (after the name of the variable corresponding to a certain vein, the serial number of the interval of values out of 3 is indicated, and the interval of MSV values are directly indicated in curly brackets) [40] (p. 426).

It can be seen from the left part of Figure 3, that the 3rd interval of MSV in the veins of the thalamus (L_TV, R_TV) are among the first 3 most contributing factors, and the 1st interval of MSV in the same veins are impending factors (see right part of Figure 3). Second and third intervals of MSV for both the basal veins (R_BV, L_BV) were included in seven contributing factors for classification of a patient in the AD group (see Figure 3), and the first interval of MSV in both left and right basal veins are considered impending factors (see Figure 3) [40]. Thus, four indicators (L_BV, R_BV, L_TV, R_TV) were selected as the most significant from among all available indicators.

3.2. Results of Interval Coding Initial Data to Discrete Values

Veins of the thalamus L_TV, R_TV and in the basal veins L_BV, R_BV have intervals performed for coding to discrete values (see, Table 2).

The initial data were quantized by replacing the MSV values for a specific patient in a specific vein (L_BV, R_BV, L_TV, R_TV) with the value of the interval obtained using SCA methods after reducing the number of analyzed parameters. In other words, the original initial data set [18] was encoded (Table 3) to discrete values from one to three with correspondence to the number of MSV intervals (see, Table 2).

Table 3. The fragment of the encoded data.

Subject	Group	Encoded L_BV	Encoded L_TV	Encoded R_BV	Encoded R_TV
sub001	AD	3	1	2	1
sub002	AD	3	3	2	3
sub003	AD	2	1	3	1
. . .	. . .	. . .	. . .	. . .	. . .
sub079	CON	3	1	2	1
sub080	CON	2	1	2	1
sub081	CON	3	2	2	2

Subjects with the same vectors of encoded veins were united into 42 subgroups. For example, subjects 003 has a vector of values of the MSV {259.0; 172.0; 333.0; 159.0}, which was encoded to vector {2; 1; 3; 1}, and subjects 057 has a vector of values of the MSV {216.0; 147.0; 301.0; 140.0}, which was encoded to the same vector {2; 1; 3; 1}. Moreover, there were 8 contradictory examples in the encoded dataset. For example, from Table 3 we can see that subjects 001 and 079 have the same vector of discrete values {3; 1; 2; 1}, but they belong to different groups: AD and CON. To reduce conflict examples, we excluded those subgroups or subjects that had high level of similarity to one group but in fact belonged to another group. The similarity was measured in the System "EIDOS-Xpro" by values of the integrated similarity criteria using information model INF3.

The "EIDOS-Xpro" system, according to both integral similarity criteria "Semantic stability of knowledge" and "Sum of knowledge" with a value of 99.69% similarity, refers the patient sub007 to the CON groups, i.e., is not ill, while this patient is clinically diagnosed with Alzheimer's disease (Figure 4).
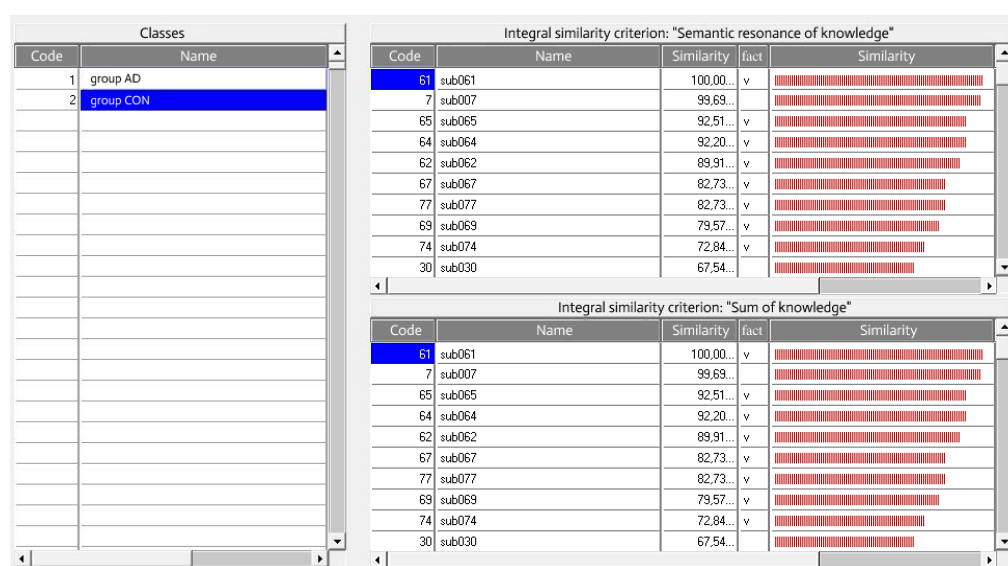
Figure 4. Recognition results for CON group of patients in the "Class-objects" mode in the system "EIDOS-Xpro".

Other subjects (Table 4) were excluded in the same way.

Table 4. Subjects excluded from conflicting examples.

Subjects	Group	Encoded L_BV	Encoded L_TV	Encoded R_BV	Encoded R_TV	Comments
sub007	AD	1	1	1	2	The similarity of the subject 007 to CON group is 99.69%, but in fact is sick
sub004, sub006	AD	2	1	2	1	Subject 004 is sick, but the similarity to the CON group is 12.94%, also the subject 080, which has the same vector {2; 1; 2; 1} and is in fact in the CON group, has a similarity to the CON group value 38.61%
sub042, sub059	AD	2	1	2	2	The similarity of the subject 059 to the CON group is 12.94%, also the subjects 070 and 072 which have the same vector {2; 1; 2; 2} and are in fact in the CON group, have similarity to the CON group value 62.66% and 38.67%, respectively
sub060	CON	2	3	2	3	The similarity of the subject 060 to the AD group is 60.01%, but in fact is not sick
sub063, sub073	CON	2	2	3	2	The similarity of the subject 063 to the AD group is 19.15%, but in fact is not sick

Table 4. *Cont.*

Subjects	Group	Encoded L_BV	Encoded L_TV	Encoded R_BV	Encoded R_TV	Comments
sub031	AD	2	2	3	2	The similarity of the subject 031 to the CON group is 50.23%, but in fact is sick
sub068, sub078	CON	2	3	3	3	The similarity of the subjects 078 and 068 to the AD group are 63.63% and 39.33%, respectively, but in fact they are not sick
sub079	CON	3	1	2	1	The similarity of the subject 079 to the CON group is 10.03%, but the subject 001, which has the same vector {3; 1; 2; 1} and in the AD group, has similarity to the AD group value 28.73%
sub071, sub081	CON	3	2	2	2	The similarity of both subjects 071 and 081 to the AD group is 33.62%, but in fact they are not sick

Subjects 057, 003, 076 and 075 were excluded from the encoded dataset because the vectors {2; 1; 3; 1}, {2; 2; 1; 3} corresponding to them have uncertainty.

The training set, after excluding the conflict examples, includes 34 unique, encoded vectors. They corresponded to 63 subjects (patients), which accounted for 78% of the original number of subjects. This training set for identification of the integrated rating mechanism (decision's root) is presented in Table S1.

3.2.1. The Integrated Rating Mechanism (Decisions' Roots) Identified Based on the Encoded Training Set

In the previous study [40], we used the identification algorithm proposed in [47], which is applicable to any structure of criteria trees without restrictions on the alphabet used inside the convolution matrix. Removing the restriction on the alphabet used to encode the elements of the convolution matrices, it is possible to approximate the initial data with several functions due to the fact that when gluing equivalence groups, the variability of these gluings arises [48].

As was said above, 15 full binary trees with named leaves are possible for 4 factors. From all possible tree structures in [40], it was proposed to convolve to produce indicators characterizing the left veins with each other, as well as the right ones with each other. This is because veins are paired and participate in the process of blood circulation of the two hemispheres separately and because arteries and veins in each of the hemispheres, and in the brain as a whole, have many collaterals that redistribute blood flow both among themselves at the beginning, within one hemisphere, and then, with a lack of compensatory capabilities within the entire brain (Figure 5).
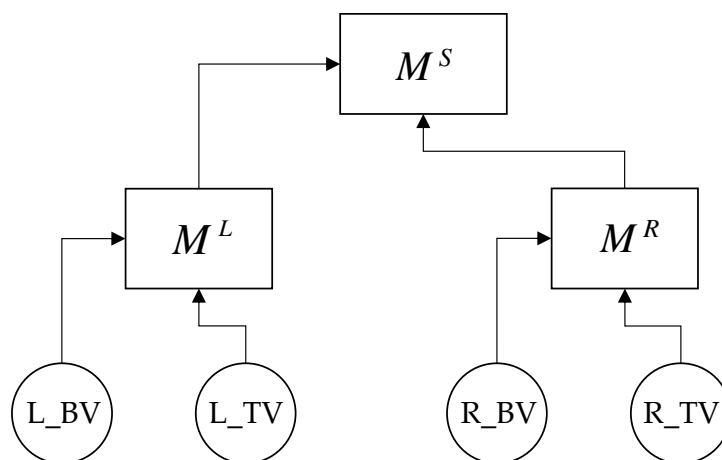
Figure 5. Criteria tree structure of convolution for the basal and thalamic.

For each criteria tree structure, a few decisions' roots can be identified; one example was presented in [40] (p. 426). The following decision's root (Figure 5) was identified based in the training set presented in Table S2.

The founded decision root (see Figure 6) can be used to predefine architecture of the desired neural network.
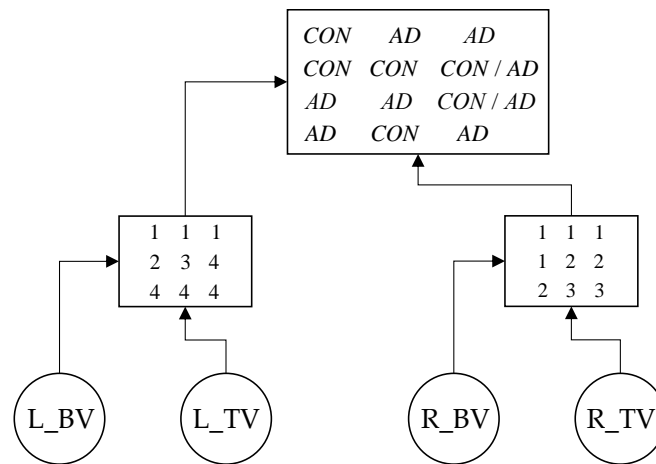
Figure 6. Decision's root for diagnosing Alzheimer's disease. The values in the left matrix at the lower level are the average numbers of the rows of the root matrix, the values in the right matrix are the numbers of the columns of the root matrix, respectively.

3.2.2. Neural Network Architecture Selection

Using algorithm [47] (p. 402), we can predefine the neurons number of the desired neural network based on the decisions' root (see, Figure 6). The first hidden layer has 12 neurons, because four input variables are indicated on Set 3 by encoded discrete values {1; 2; 3}. For both convolution matrices from the bottom level of the decisions' roots, there are nine neurons on the second hidden layer, because matrices have nine elements. Matrix M^L has four values, M^R has three values, so the third hidden layer has seven neurons. The fourth hidden layer has 12 neurons because root matrix M^L has 12 elements. The fifth hidden layer has two neurons because root matrix M^S has values from set {CON; AD}. Thus, we have obtained a neural network with five hidden layers and 56 neurons on all layers (including one input and one output layer). The neural network architecture is shown in Figure 7, and the parameters of this network are presented in Table 5.
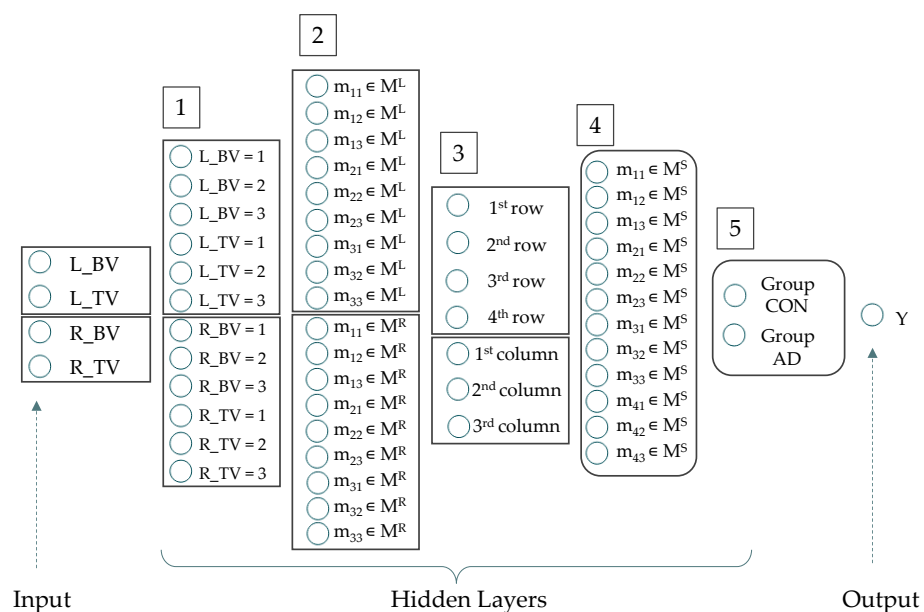
Figure 7. Neurons of artificial neural network (where: vector of neurons on the Input Layer $\in R^4$, vector of neurons on the Hidden Layer (1) $\in R^{12}$, vector of neurons on the Hidden Layer (2) $\in R^{18}$,

vector of neurons on the Hidden Layer (3) $\in R^7$, vector of neurons on the Hidden Layer (4) $\in R^{12}$, vector of neurons on the Hidden Layer (5) $\in R^2$, vector of neurons on the Output Layer $\in R^1$).

Table 5. Numbers of neurons in hidden layers of the desired artificial neural network.

	Hidden Layer	Number of Neurons	Group of Neurons	Comments
Input	1	12	6 and 6	Neurons corresponds to intervals on input signals
4 signals	2	18	9 and 9	Neurons corresponds to elements of matrices M^L and M^R
Output	3	7	4 and 3	Neurons corresponds to values of matrices M^L and M^R
1 signal	4	12	12	Neurons corresponds to elements of matrix M^S
	5	2	2	Neurons corresponds to values of matrix M^S

The synaptic connections (Figure 8) are determined in full accordance with the elements of convolutional matrices in the decisions' roots (see, Figure 6).
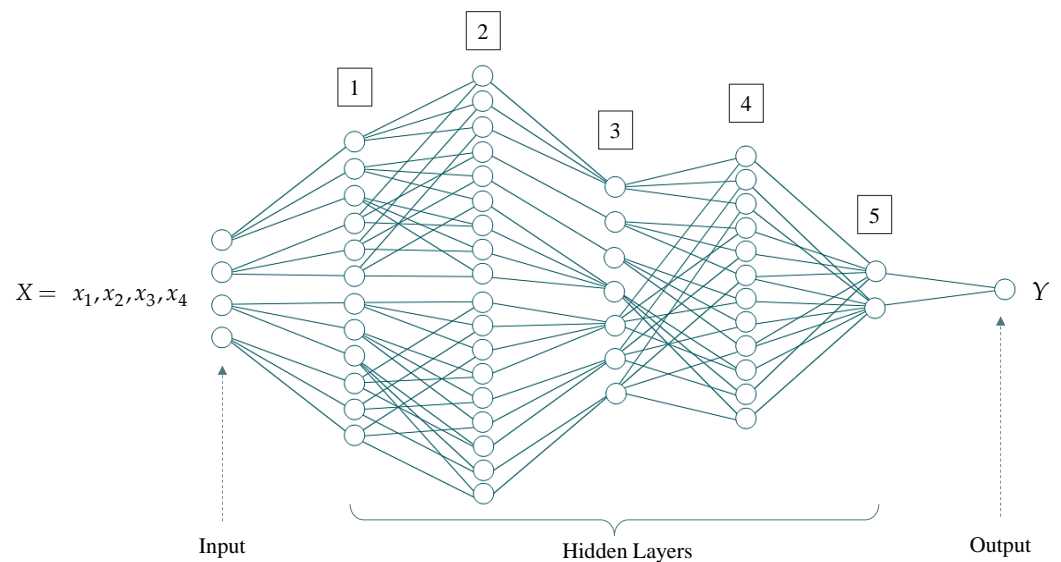
Figure 8. Incompletely connected artificial neural network identified based on the decision root. Numbers from 1 to 5 designations of serial numbers of hidden layers of the neural network.

The total number of unknown parameters (synaptic coefficients) of the incompletely connected neural network predefined, based on the decision's root is 158.

The structure of the predefined neural network model had 12 connections between the input signals and the first hidden layer (each connection generates two parameters of the mathematical model, which means that we have 24 parameters in total).

The next layer, in which the neurons correspond to the elements of the matrix, has three parameters each (two inputs from the previous layer and one unknown constant). Since we have two matrices of nine elements each at the bottom level of the criteria tree, this generates 18 neurons on the second layer, which means a total of 54 unknown parameters.

The next layer consists of the neurons corresponding to the matrix estimates, and there are exactly as many connections as there were neurons on the previous layer, i.e., in our case, 18, but we need to add seven parameters to them (determined by the number of neurons on the 3rd layer), which also means we have eighteen and seven parameters (25 parameters on the third layer).

Then, there is a matrix containing 12 elements, where each element corresponds to three parameters (two inputs and one unknown constant). It means 36 parameters on the 4th layer.

There are two neurons on the last hidden layer, where 14 input signals (10 neurons on the previous layer have one signal and two neurons have two signals; the last ones correspond to elements in the root matrix with value CON/AD, see Figure 6) plus two unknown constants (the number of neurons on the current layer) are taken.

There is only one signal at the output. This signal is associated with the last two neurons. It means there are three parameters (two inputs plus one unknown constant).

Thus, the total number of unknown parameters is determined by the sum (24 + 54 + 25 + 36 + 16 + 3 = 158).

Nowadays, it is worth noting that there are no publicly available specialized software products allowing the user to set the neural network structure, for example, to load the adjacency matrix or the incidence matrix of the graph corresponding to the neural network architecture. We tried to use the computer program "Software package that implements the operation of incompletely connected neural networks" [57]. It implemented incompletely connected neural networks only for one hidden layer, but as you can see from Figure 8, the required neural network has incompletely connected neurons on each hidden layer. Thus, this program is not applicable for this case.

Though there are no publicly available specialized software products allowing the user to set the structure of the neural network, we can find fully connected neural networks comparable in terms of the number of unknown parameters and compare the effectiveness of diagnosing Alzheimer's disease knowing the number of unknown parameters that will be determined in the process of training the neural network.

Therefore, a fully connected single-layer neural network (Figure 9) with four input signals and one output signal has the following rule—each neuron on the hidden layer corresponds to four input-unknown variables corresponding to the input signals and one unknown constant (i.e., $5x$ unknown parameters); the output signal receives signals from x neurons at the input and has one unknown constant.
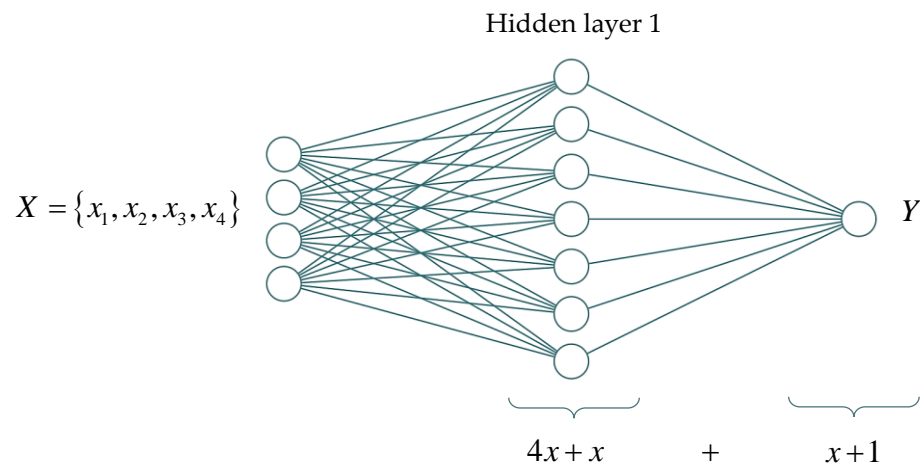
Figure 9. Schematic representation of a single-layer neural network structure with x neurons and numbers of unknown parameters.
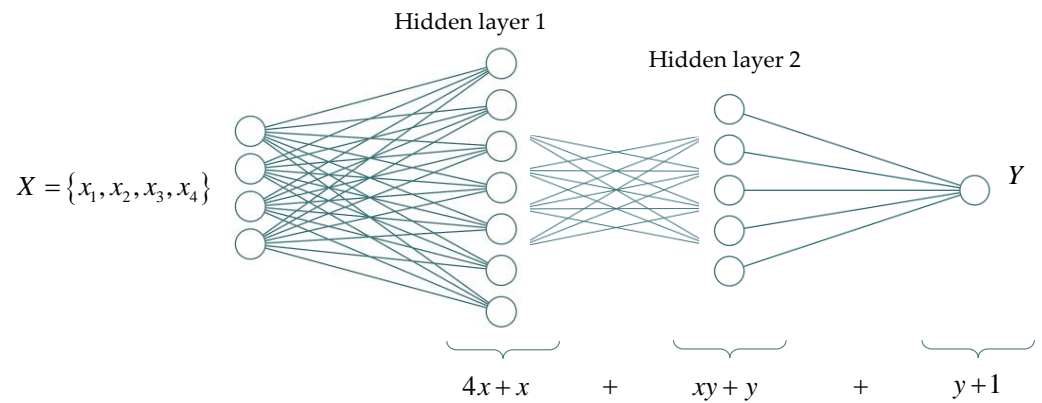
Thus, the number of unknown parameters in a fully connected network with one hidden layer can be expressed using the Equation (14)

$$4x + x + 1 \geq 158, \quad x \in N \tag{13}$$

According to (3), the minimum number of neurons on the hidden layer for a single-layer network is 27, where this number is comparable to the number of unknown parameters with the found incompletely connected neural network (see Figure 8).

The following equation will be obtained for a two-layer neural network structure (Figure 10):

$$5x + xy + 2y + 1, \quad x, y \in N \tag{14}$$

Figure 10. Schematic representation of the structure of a fully connected two-layer neural network and the number of unknown parameters.

These minimal integer values x were found by changing the number of neurons on the second hidden layer y, starting from one, subject to the fulfillment of inequality (4). These values determine the number of neurons on the first hidden layer (Figure 11). Thus, various two-layer neural networks were found, where they are comparable in terms of the number of unknown neural network parameters (this network is built by the decision's root) (see Figure 6).
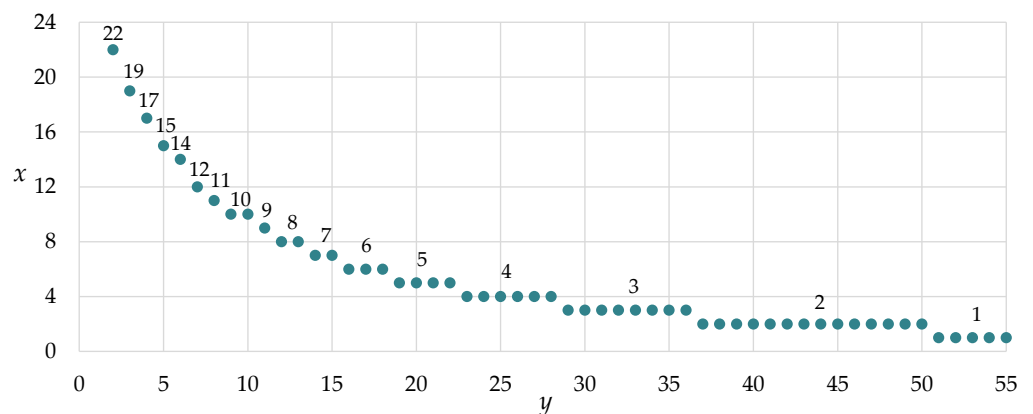
Figure 11. The dependence of the number of neurons between the two layers (x—first hidden layer, y—second hidden layer).

3.2.3. Results of Neural Network Training

According to the generally accepted neural network design technology, the entire general population is divided into training, testing and confirming sets in the ratio of 70:20:10% [52]. In our problem, we neglect a strict formal approach to determining the percentage ratio between training, testing and confirming sets due to a fairly small number of examples, which is acceptable [63], thereby dividing the entire initial set only into training and testing in the ratio of 90:10%.

We performed multiple training and testing of neural networks for diagnosing Alzheimer's disease on limited raw data using one-layer and two-layer neural networks. Several results of one-layer neural networks training and testing are shown below (Figure 12, Table 6).
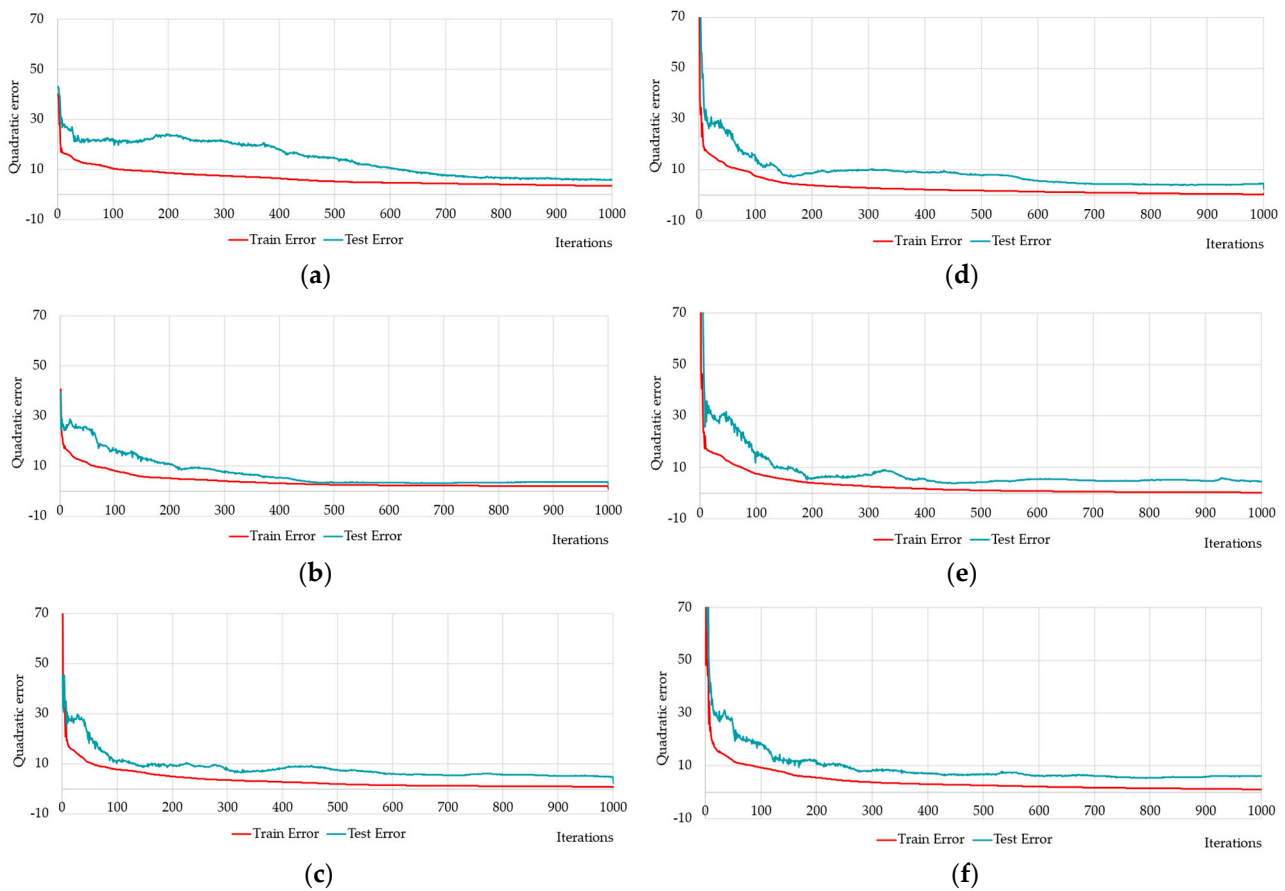
Figure 12. Changing of the square errors at training and testing depending on iterations of calculation synaptic coefficients at fully connected one-layer neural networks with different numbers of neurons on the hidden layer (hereinafter the following designation will be used HL1 stands for neural network with 1 hidden layer): HL1-5 has 5 neurons (**a**); HL1-10 has 10 neurons (**b**); HL1-15 has 15 neurons (**c**); HL1-20 has 20 neurons (**d**), HL1-25 has 25 neurons (**e**), and fully connected one-layer neural network with predefined number of neurons HL1-27 has 27 neurons (**f**).

Table 6. Results of a trained different fully connected one-layer neural networks.

|	Neural Network	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
HL1-5	1 layer with 5 neurons	17.068	12.582	24.517	21.047	2.913	6.011	0.849	0.744
HL1-10	1 layer with 10 neurons	11.557	8.718	17.329	13.899	1.334	3.003	0.930	0.872
HL1-15	1 layer with 15 neurons	9.268	7.000	21.897	18.348	0.859	4.795	0.955	0.795
HL1-20	1 layer with 20 neurons	5.829	4.422	21.262	18.988	0.034	4.521	0.982	0.807
HL1-25	1 layer with 25 neurons	2.047	1.440	21.827	16.692	0.042	4.764	0.997	0.797
HL1-27	1 layer with 27 neurons	1.987	1.392	21.138	14.462	0.039	4.468	0.997	0.809

Figure 12a–e shows five neural networks, distinguished by five neurons. Figure 12f shows the training and testing errors for the neural network HL1-27 with the 27 neurons satisfying inequality (14). Figure 12 shows that the quadratic error becomes less than 10% in all cases. Therefore, we should compare the results of the training with the other criteria: Q_1–average quadratic relative (%); Q_2–average relative error; Q_3–average quadratic relative (Test, %); Q_4–average relative error (Test); Q_5–quadratic train error; Q_6–quadratic test error; Q_7–R^2 train set; Q_8–R^2 test set (Table 6).

Comparing the performance by all the criteria in Table 6, it can be seen that the proposed neural network HL1-27 is better on almost all the criteria. At the same time, the best neural network training results from the multiple experiments we showed on both Figure 12a–e and the first seven rows of Table 6.

In our experiments, we noticed that the quality and speed of neural network learning could be improved by increasing the number of neurons. The neural network with pre-defined number of neurons HL1-27 (see Figure 12f) was trained rather quickly, and the results were more robust compared to variants HL1-5, HL1-10, HL1-15, HL1-20, which often over-trained the neural network.

We propose to consider nine variations of two-layer neural networks with different numbers of neurons on first and second hidden layers. The network structures are chosen so that the number of neurons on the first and the second hidden layer form a cross of gridlines with steps of five neurons—selected combinations of neuron number on the first and second layers are highlighted by black points (Figure 13, points 1–9).
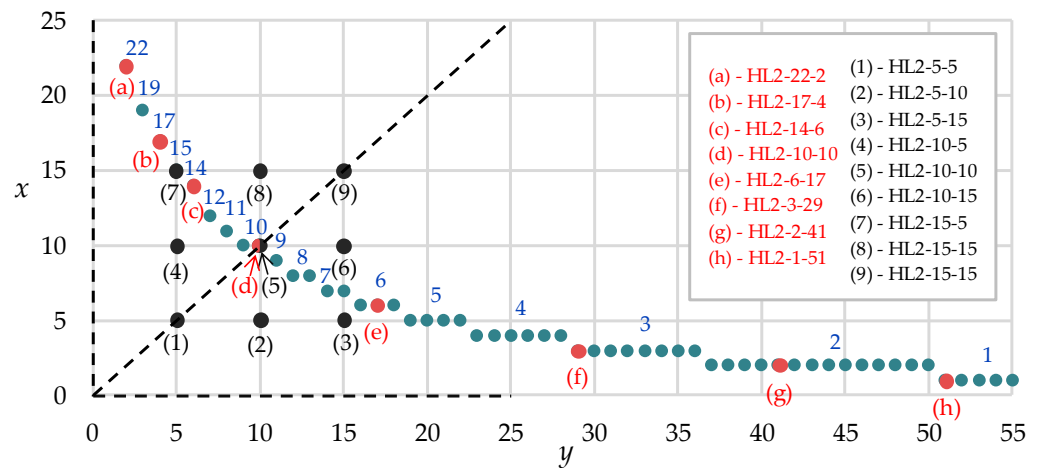
Figure 13. Selected two-layers neural networks with different numbers of neurons on first and second hidden layers (hereinafter the following designation will be used–HL2 stands for two hidden layers, the next number after the hyphen stands for the number of neurons on the first hidden layer, and the last number stands for the number of neurons on the second hidden layer, respectively).

We also propose to consider several variations of neural networks whose structures are comparable to dependence (see, Figure 11). Selected eight combinations of neuron number on first and second are highlighted by red points (Figure 13, points a–h). This is necessary to assess the quality and ability to recognize and predict our neural network, the structure of which is determined according to the decision's root and convolution matrices.

The dotted line in Figure 13 divides all neural networks into two types: in the upper left corner, there are neural networks with fewer neurons on the first layer than on the second layer. In the lower right area, on the contrary, there are neural networks with more neurons on the first layer than on the second layer.

The network HL2-10-10 refers to both a set of nine neural networks (see Figure 13, point 5) chosen by varying the number of variables on the first and second layer of five neurons (see Figure 13, black points 1–9), and a set of eight predefined neural networks (Figure 13, point d) that contain a number of synaptic coefficients comparable to a decision-root-based neural network (see Figure 8).

Therefore, out of nine neural networks, Figure 14 and Table 7 show the training and testing results of eight networks: HL2-5-5-5 (Figure 14a); HL2-5-10 (Figure 14b); HL2-5-15 (Figure 14c); HL2-10-5 (Figure 14d); HL2-10-15 (Figure 14e); HL2-15-5 (Figure 14f); HL2-15-10 (Figure 14g); and HL2-15-15 (Figure 14h).
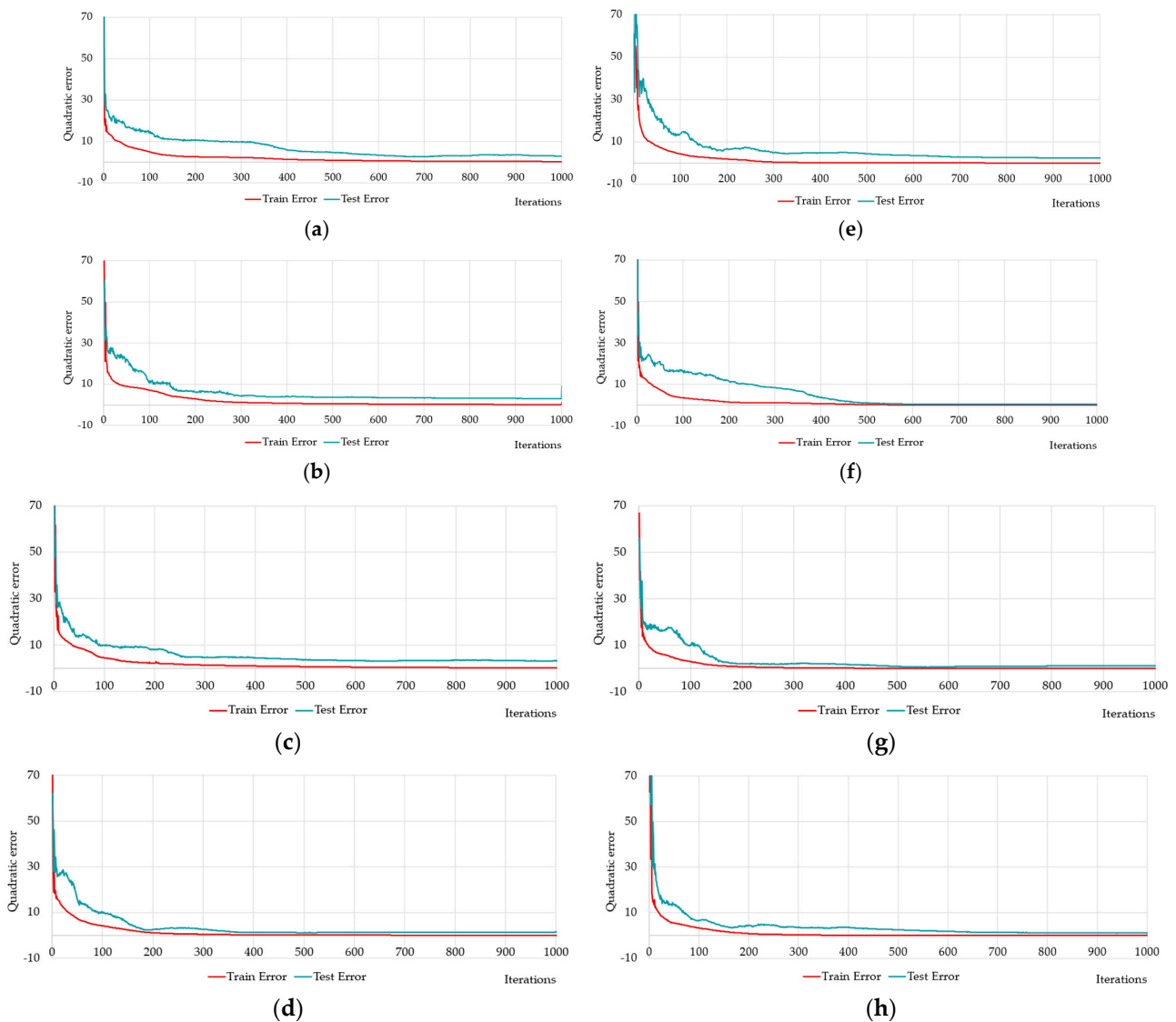
Figure 14. Changing of the square errors at training and testing depending on iterations of calculation synaptic coefficients at two-layers neural networks with different numbers of neurons on the hidden layers: HL2-5-5 (**a**); HL2-5-10 (**b**); HL2-5-15 (**c**); HL2-10-5 (**d**); HL2-10-15 (**e**); HL2-15-5 (**f**); HL2-15-10 (**g**); HL2-15-15 (**h**).

Table 7. Results of a trained different fully connected two-layer neural networks.

	Neural Network	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
HL2-5-5	2 layers with 5 neurons on first layer and 5 neurons on second layer	4.132	2.755	16.518	12.429	0.145	2.210	0.996	0.884
HL2-5-10	2 layers with 5 neurons on first layer and 10 neurons on second layer	4.003	2.922	17.234	11.922	0.160	2.970	0.992	0.873
HL2-5-15	2 layers with 5 neurons on first layer and 15 neurons on second layer	4.581	3.163	17.720	15.341	0.210	3.140	0.989	0.866
HL2-10-5	2 layers with 10 neurons on first layer and 5 neurons on second layer	0.715	0.480	13.094	8.362	0.005	1.714	0.999	0.934
HL2-10-15	2 layers with 10 neurons on first layer and 15 neurons on second layer	0.775	0.458	16.823	9.647	0.006	2.830	0.999	0.888
HL2-15-5	2 layers with 15 neurons on first layer and 5 neurons on second layer	0.665	0.479	9.918	7.110	0.004	0.797	0.999	0.958
HL2-15-10	2 layers with 15 neurons on first layer and 10 neurons on second layer	0.904	0.548	14.145	10.509	0.007	1.621	0.999	0.915
HL2-15-15	2 layers with 15 neurons on first layer and 15 neurons on second layer	1.010	0.654	12.276	9.647	0.008	1.221	0.999	0.936

Figure 14 shows the networks HL2-10-5 (Figure 14d); HL2-15-5 (Figure 14f) and HL2-15-10 (Figure 14g) required fewer iterations to training.

Table 7 shows that the HL2-15-5 network is better than the others by all criteria except Q_2–average relative error. The HL2-15-5 network is insignificantly inferior to the HL2-10-15 network in this criterion. At the same time, the HL2-10-15 network is inferior to other networks in many criteria.

It is of interest to compare the training results of randomly selected neural networks and predefined neural networks. Table 8 and Figure 15 show the results of training and testing of eight selected neural networks with predefined number of synaptic coefficients: HL2-22-2 (Figure 15a); HL2-17-4 (Figure 15b); HL2-14-6 (Figure 15c); HL2-10-10 (Figure 15d); HL2-6-17 (Figure 15e); HL2-3-29 (Figure 15f); HL2-2-41 (Figure 15g); HL2-1-51 (Figure 15h).

Table 8. Results of a trained predefined fully connected two-layer neural networks.

Neural Network		Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
HL2-22-2	2 layers with 22 neurons on first layer and 2 neurons on second layer	1.043	0.556	20.344	15.716	0.011	4.139	0.998	0.823
HL2-17-4	2 layers with 17 neurons on first layer and 4 neurons on second layer	0.774	0.477	16.971	11.897	0.006	2.888	0.999	0.877
HL2-14-6	2 layers with 14 neurons on first layer and 6 neurons on second layer	0.543	0.348	14.759	9.564	0.003	2.178	1.000	0.907
HL2-10-10	2 layers with 10 neurons on first layer and 10 neurons on second layer	0.809	0.507	14.293	12.834	0.005	1.655	0.999	0.913
HL2-6-17	2 layers with 6 neurons on first layer and 17 neurons on second layer	1.121	0.818	21.026	16.886	0.010	3.581	0.998	0.811
HL2-3-29	2 layers with 3 neurons on first layer and 29 neurons on second layer	6.006	3.763	14.313	12.314	0.292	1.659	0.981	0.913
HL2-2-41	2 layers with 2 neurons on first layer and 41 neurons on second layer	15.952	6.950	32.393	17.504	2.061	8.499	0.868	0.552
HL2-1-51	2 layers with 1 neuron on first layer and 51 neurons on second layer	27.104	16.408	34.519	23.725	5.951	9.652	0.618	0.492

The network HL2-15-5 is close to the predefined networks HL2-17-4 and HL2-14-6 by number of neurons, precisely these two networks with the HL2-10-10 network that have Q_1 values less than one (see Table 8). The HL2-14-6 network has the lowest Q_2, Q_4 and Q_5 values (see Table 8). HL2-10-10 has the lowest Q_3, Q_6 values (see Table 8). In terms of the coefficient of determination, the network HL2-14-6 has one, which means 100% of the variance of the resultant quantity in the training set is explained by the influence of selected variables.

Figure 15 shows that neural networks with a predefined number of neurons required significantly fewer iterations to train as compared to the brute-force selected networks. It is worth recalling again that Figures 14 and 15 show the best results from the number of multiple computational experiments. A smaller number of iterations was required to train the predefined neural networks.

Having learnt the results of training two-layer networks, we noticed the following— the quality of neural network models is better if the number of neurons on the first layer is greater than on the second (Figures 14d–g and 15a–d).

The results of training and testing two-layer neural networks are not entirely consistent with the observation in the case of single-layer neural networks. In the case of single layer neural networks, we observed the following—as the number of neurons on the hidden layer increases, the quality of neural network models becomes better. With two-layer networks, this is not true, e.g., HL2-22-2 with 24 neurons or HL2-17-4 with 21 neurons are in many ways (see Table 8) worse than the 20-neuron HL2-14-6 and HL2-10-10.
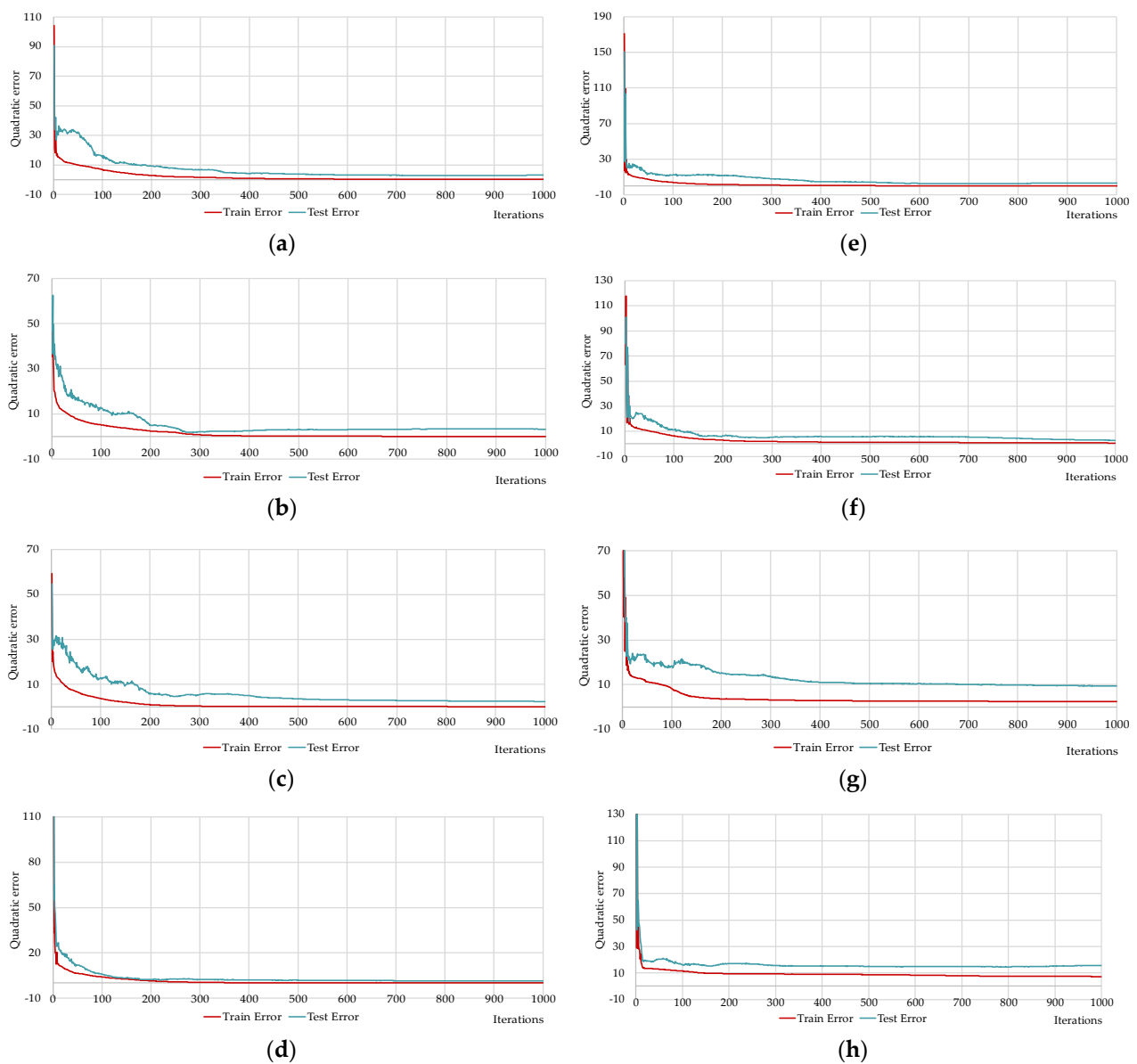
Figure 15. Changing of the square errors at training and testing depending on iterations of calculation synaptic coefficients at two-layers neural networks with different numbers of neurons on the hidden layers: HL2-22-2 (**a**); HL2-17-14 (**b**); HL2-14-6 (**c**); HL2-10-10 (**d**); HL2-6-17 (**e**); HL2-3-29 (**f**); HL2-2-41 (**g**); HL2-1-51 (**h**).

4. Discussion

Having compared all obtained results of training and testing of neural networks for diagnosing Alzheimer's disease on limited data, it is possible to make a general conclusion—the constructed neural networks whose structure is chosen according to a predefined number of synaptic coefficients according to inequalities (14) and (15) have acceptable accuracy in comparison with neural networks whose structure is chosen by brute force.

In our experiments, we noticed that neural networks with a predefined number of neurons needed less training to obtain acceptable accuracy, and the results were more robust compared to neural networks with a random number of neurons. In the latter case, if the number of neurons was very different from the predefined number of neurons, we quite often observed retraining of neural networks.

The key value of the approach proposed in this paper for analyzing data and determining the structure of a neural network is to reduce the time spent on finding the optimal

neural network, since based on the results of data preprocessing, we obtain an encoded data set, which is already a graph-matrix representation of the possible structure of the neural network. Moreover, the prospects for such an approach to data processing and analysis are reduced to improve the quality and accuracy of models on a limited data set, which we encountered in the study of brain disease in the form of Alzheimer's disease.

Thus, it is possible to determine some of the parameters of the incomplete neural network with cascade structure based on the integrated rating mechanism (decision root). In future studies, we will use the term decisions-root-based neural network (DRB NN).

We would also like to point out that the best distribution for two-layer neural networks appears to be in the upper left corner of the space, separated by a dotted line (see Figure 13), which means that there should be fewer neurons on the first layer than on the second in the case of limited sets of training examples. Based on this, we can assume that in the case of multilayer neural networks, a similar ratio of neurons on upstream and downstream layers will also affect the quality of neural networks. This is fundamentally important for the proposed approach of building incomplete neural networks. Figure 8 clearly shows that the decisions-root-based neural network on layers three and five has a greater number of neurons than layers two and four, respectively. This could potentially limit the learning effectiveness of decision-root-based neural networks. However, the pattern described above regarding the correlation of neurons on previous and subsequent layers is true for fully connected neural networks. We will study incompletely connected neural networks in the future. This is of interest, and we plan to explore such a pattern in more detail in the following studies.

There are no publicly available specialized software products allowing the user to get the incompletely connected neural network structure. For example, the computer program "Software package that implements the operation of incompletely connected neural networks" [57] can create incompletely connected neural networks, but with only one hidden layer with incompletely connected neurons. It is not enough. Therefore, the authors intend to create a special computer program on software as a service available online for all researchers. This work is performed at the LLC "Perm Decision Making Support Center".

5. Conclusions

It should be noted that the purpose of this study was achieved, and the hypothesis was confirmed. The neural network, preidentified, based on preprocessing initial datasets using suggested methods and algorithms, has acceptable accuracy in a case with a very limited training set. The suggested methodology consisting of interval coding of initial data and search of integrated rating mechanisms, following representation in the form of an artificial neural network and training using initial data, is effective.

Based on the results of this study, it should be noted that with the help of special methods based on the mechanism of complex assessment, it was possible to obtain the optimal structure of the neural network model capable of describing the study area with sufficient accuracy, using the example of Alzheimer's disease. This, once again, confirms that neural network technologies are quite successful in medicine, which is especially valuable with a limited set of initial data. Similar conclusions about the successful application of neural network modeling in medicine and, in particular, the study of the brain, can be found in references [16,17] etc., where it was proposed to use convolutional neural networks and deep learning models in medicine.

Author Contributions: Conceptualization, A.A. and L.K.; methodology, A.A.; software, L.K.; validation, A.A., L.K., V.N. and J.B.; formal analysis, A.A., L.K., V.N. and J.B.; investigation, A.A. and L.K.; resources, L.K.; data curation, A.A. and L.K.; writing—original draft preparation, A.A., L.K. and V.N.;

References

1. Yasnitsky, L.N.; Dumler, A.A.; Cherepanov, F.M.; Yasnitsky, V.L.; Uteva, N.A. Capabilities of neural network technologies for extracting new medical knowledge and enhancing precise decision making for patients. *Expert Rev. Precis. Med. Drug Dev.* **2022**, *7*, 70–78. [CrossRef]
2. Yasnitsky, L.N. Artificial intelligence and medicine: History, current state, and forecasts for the future. *Curr. Hypertens. Rev.* **2020**, *16*, 210–215. [CrossRef] [PubMed]
3. Pyatakovich, F.A.; Khlivnenko, L.V.; Mevsha, O.V.; Yakunchenko, T.I.; Makkonen, K.F. Development of a biotechnical system based on the functioning of neural networks for solving the problem of the scattergrams analysis. *Netw. Electron. Sci. Educ. J. Mod. Issues Biomed.* **2018**, *3*, 171–183.
4. Bogdanov, L.A.; Komossky, E.A.; Voronkova, V.V.; Tolstosheev, D.E.; Martsenyuk, G.V.; Agienko, A.S.; Indukaeva, E.V.; Kutikhin, A.G.; Tsygankova, D.P. Prototyping neural networks to evaluate the risk of adverse cardiovascular outcomes in the population. *Fundam. Clin. Med.* **2021**, *6*, 67–81. [CrossRef]
5. Kilic, A. Artificial Intelligence and Machine Learning in Cardiovascular Health Care. *Ann. Thorac. Surg.* **2020**, *109*, 1323–1329. [CrossRef]
6. Solodukha, T.V. Development of a Specialized Computer System Based on Neural Networks for Predicting the Consequences of Allergic Reactions. Master's Thesis, Donetsk National Technical University, Donetsk, Ukraine, 2002.
7. Carrara, M.; Bono, A.; Bartoli, C.; Colombo, A.; Lualdi, M.; Moglia, D.; Santoro, N.; Tolomio, E.; Tomatis, S.; Tragni, G.; et al. Multispectral imaging and artificial neural network: Mimicking the management decision of the clinician facing pigmented skin lesions. *Phys. Med. Biol.* **2007**, *52*, 2599–2613. [CrossRef] [PubMed]
8. Gruvberger-Saal, S.K.; Edén, P.; Ringnér, M.; Baldetorp, B.; Chebil, G.; Borg, A.; Fernö, M.; Peterson, C.; Meltzer, P.S. Predicting continuous values of prognostic markers in breast cancer from microarray gene expression profiles. *Mol. Cancer Ther.* **2004**, *3*, 161–168. [CrossRef]
9. Ercal, F.; Chawla, A.; Stoeker, W.V.; Lee, H.C.; Moss, R.H. Neural network diagnosis of malignant melanoma from color images. *IEEE Trans. Biomed. Eng.* **1994**, *41*, 837–845. [CrossRef]
10. Lundin, J.; Lundin, M.; Holli, K.; Kataja, V.; Elomaa, L.; Pylkkänen, L.; Turpeenniemi-Hujanen, T.; Joensuu, H. Omission of histologic grading from clinical decision making may result in overuse of adjuvant therapies in breast cancer: Results from a nationwide study. *J. Clin. Oncol.* **2001**, *19*, 28–36. [CrossRef]

11. Zheng, B.; Leader, J.K.; Abrams, G.S.; Lu, A.H.; Wallace, L.P.; Maitz, G.S.; Gur, D. Multiview based computer-aided detection scheme for breast masses. *Med. Phys.* **2006**, *33*, 3135–3143. [CrossRef]

12. Gavrilov, D.A.; Zakirov, E.I.; Gameeva, E.V.; Semenov, V.Y.; Aleksandrova, O.Y. Automated skin melanoma diagnostics based on mathematical model of artificial convolutional neural network. *Res. Pract. Med. J.* **2018**, *5*, 110–116. [CrossRef]

13. Jerez-Aragonés, J.M.; Gómez-Ruiz, J.A.; Ramos-Jiménez, G.; Muñoz-Pérez, J.; AlbaConejo, E.A. Combined neural network and decision trees model for prognosis of breast cancer relapse. *Artif. Intell. Med.* **2003**, *27*, 45–63. [CrossRef]

14. Lundin, J.; Burke, H.B.; Toikkanen, S.; Pylkkänen, L.; Joensuu, H. Artificial neural networks applied to survival prediction in breast cancer. *Oncology* **1999**, *57*, 281–286. [CrossRef]

15. Reed, T.R.; Reed, N.E.; Fritzson, P. Heart sound analysis for symptom detection and computer-aided diagnosis. *Simul. Model. Pract. Theory* **2004**, *12*, 129–146. [CrossRef]

16. Chang, P.D.; Kuoy, E.; Grinband, J.; Weinberg, B.D.; Thompson, M.; Homo, R.; Chen, J.; Abcede, H.; Shafie, M.; Sugrue, L.; et al. Hybrid 3D/2D convolutional neural network for hemorrhage evaluation on head CT. *Am. J. Neuroradiol.* **2018**, *39*, 1609–1616. [CrossRef] [PubMed]

17. Chao, P.-J.; Chang, L.; Kang, C.L.; Lin, C.-H.; Shieh, C.-S.; Wu, J.M.; Tseng, C.-D.; Tsai, I.-H.; Hsu, H.-C.; Huang, Y.-J.; et al. Using deep learning models to analyze the cerebral edema complication caused by radiotherapy in patients with intracranial tumor. *Sci. Rep.* **2022**, *12*, 1555. [CrossRef] [PubMed]

18. Liu, Y.; Dong, J.; Song, Q.; Zhang, N.; Wang, W.; Gao, B.; Tian, S.; Dong, C.; Liang, Z.; Xie, L.; et al. Correlation between cerebral venous oxygen level and cognitive status in patients with Alzheimer's disease using quantitative susceptibility mapping. *Front. Neurosci.* **2021**, *14*, 570848. [CrossRef]

19. Gribanov, A.V.; Dzhos, Y.S.; Deryabina, I.N.; Deputat, I.S.; Yemelianova, T.V. An aging of the human brain: Morpho functional aspects. *Zhurnal Nevrol. Psikhiatr. Im. S.S. Korsakova* **2017**, *117*, 3–7. [CrossRef]

20. Guseva, E.I.; Hecht, A.B. *Diseases of the Brain: Problems and Solutions*; Sam Polygraphist LLC: Moscow, Russia, 2021; 416p.

21. Perepelkina, O.V.; Tarasova, A.Y.; Poletaeva, I.I. Modeling of diseases of the human brain in experiments on rodents (brief review). *Mod. Foreign Psychol.* **2016**, *5*, 13–23. [CrossRef]

22. Bilkei-Gorzo, A. Genetic mouse models of brain aging and Alzheimer's disease. *Pharmacol. Ther.* **2014**, *142*, 244–257. [CrossRef]

23. Andjelkovic, A.V.; Stamatovic, S.M.; Phillips, C.M. Modeling blood–brain barrier pathology in cerebrovascular disease in vitro: Current and future paradigms. *Fluids Barriers CNS* **2020**, *17*, 44. [CrossRef] [PubMed]

24. Ago, T. The neurovascular unit in health and ischemic stroke. *Nihon Rinsho* **2016**, *74*, 583–588. [PubMed]

25. Cai, W.; Liu, H.; Zhao, J.; Chen, L.Y.; Chen, J.; Lu, Z.; Hu, X. Pericytes in brain injury and repair after ischemic stroke. *Transl. Stroke Res.* **2017**, *8*, 107–121. [CrossRef] [PubMed]

26. De Marco, M.; Venneri, A. Volume and connectivity of the ventral tegmental area are linked to neurocognitive signatures of Alzheimer's disease in humans. *J. Alzheimer's Dis.* **2018**, *63*, 167–180. [CrossRef]

27. Nobili, A.; Latagliata, E.C.; Viscomi, M.T.; Cavallucci, V.; Cutuli, D.; Giacovazzo, G.; Krashia, P.; Rizzo, F.R.; Marino, R.; Federici, M.; et al. Dopamine neuronal loss contributes to memory and reward dysfunction in a model of Alzheimer's disease. *Nat. Commun.* **2017**, *8*, 14727. [CrossRef]

28. Chinta, S.J.; Andersen, J.K. Dopaminergic neurons. *Int. J. Biochem. Cell Biol.* **2005**, *37*, 942–946. [CrossRef] [PubMed]

29. Schultz, W. Multiple dopamine functions at different time courses. *Annu. Rev. Neurosci.* **2007**, *30*, 259–288. [CrossRef]

30. Schultz, W. Behavioral dopamine signals. *Trends Neurosci.* **2007**, *30*, 203–210. [CrossRef]

31. Vipin, A.; Loke, Y.M.; Liu, S.; Hilal, S.; Shim, H.Y.; Xu, X.; Tan, B.Y.; Venketasubramanian, N.; Chen, C.L.; Zhou, J. Cerebrovascular disease influences functional and structural network connectivity in patients with amnestic mild cognitive impairment and Alzheimer's disease. *Alzheimer's Res. Ther.* **2018**, *10*, 82. [CrossRef]

32. Zhu, H.; Zhou, P.; Alcauter, S.; Chen, Y.; Cao, H.; Tian, M.; Ming, D.; Qi, H.; Wang, X.; Zhao, X.; et al. Changes of intranetwork and internetwork functional connectivity in Alzheimer's disease and mild cognitive impairment. *J. Neural Eng.* **2016**, *13*, 046008. [CrossRef]

33. Buckner, R.L.; Snyder, A.Z.; Shannon, B.J.; LaRossa, G.; Sachs, R.; Fotenos, A.F.; Sheline, Y.I.; Klunk, W.E.; Mathis, C.A.; Morris, J.C.; et al. Molecular, structural, and functional characterization of Alzheimer's disease: Evidence for a relationship between default activity, amyloid, and memory. *J. Neurosci.* **2005**, *25*, 7709–7717. [CrossRef] [PubMed]

34. Hedera, P.; Lai, S.; Lewin, J.S.; Haacke, E.M.; Wu, D.; Lerner, A.J.; Friedland, R.P. Assessment of cerebral blood flow reserve using functional magnetic resonance imaging. *J. Magn. Reason. Imaging* **1996**, *6*, 718–725. [CrossRef] [PubMed]

35. Choi, J.-K.; Chen, Y.I.; Hamel, E.; Jenkins, B.G. Brain hemodynamic changes mediated by dopamine receptors: Role of the cerebral microvasculature in dopamine-mediated neurovascular coupling. *NeuroImage* **2006**, *30*, 700–712. [CrossRef] [PubMed]

36. Walton, L.R.; Verber, M.; Lee, S.H.; Chao, T.H.; Wightman, R.M.; Shih, Y.I. Simultaneous fMRI and fast-scan cyclic voltammetry bridges evoked oxygen and neurotransmitter dynamics across spatiotemporal scales. *NeuroImage* **2021**, *244*, 118634. [CrossRef]

37. Knutson, B.; Gibbs, S.E.B. Linking nucleus accumbens dopamine and blood oxygenation. *Psychopharmacology* **2007**, *191*, 813–822. [CrossRef]

38. Leung, B.K.; Balleine, B.W. Ventral pallidal projections to mediodorsal thalamus and ventral tegmental area play distinct roles in outcome-specific Pavlovian-instrumental transfer. *J. Neurosci.* **2015**, *35*, 4953–4964. [CrossRef]

39. Bruinsma, T.J.; Sarma, V.V.; Oh, Y.; Jang, D.P.; Chang, S.Y.; Worrell, G.A.; Lowe, V.J.; Jo, H.J.; Min, H.K. The relationship between dopamine neurotransmitter dynamics and the blood-oxygen-level-dependent (BOLD) signal: A review of pharmacological functional magnetic resonance imaging. *Front. Neurosci.* **2018**, *12*, 238. [CrossRef]

40. Kozhemyakin, L.V.; Alekseev, A.O.; Nikitin, V.N. Application of Decisions' Roots for Data Analysis on Example of Dataset with Magnetic Susceptibility Values of the Brain Veins and the Alzheimer's Disease. In Proceedings of the 2022 4th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA), Lipetsk, Russia, 9–11 November 2022. [CrossRef]

41. Trapeznikov, V.; Gorelikov, N.; Burkov, V.; Zimokha, V.; Tolstykh, A.; Cherkashin, A.; Tsyganov, V. An integrated approach to managing scientific and technological progress in the industry. *Her. Acad. Sci. USSR* **1983**, *3*, 33–43.

42. Korgin, N.; Sergeev, V. The Art of Scientific Computingdesitions' Root—Yet Another Tool for Ordinal Data Analysis. Available online: https://youtu.be/b4dF7znmVyo (accessed on 27 November 2021).

43. Korgin, N.; Sergeev, V. Identification of integrated rating mechanisms on complete data sets. In *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems. In Proceedings of the APMS 2021. IFIP Advances in Information and Communication Technology, Nantes, France, 5–9 September 2021*; Springer: New York, NY, USA, 2021; pp. 610–616. [CrossRef]

44. Burkov, V.N.; Korgin, N.A.; Sergeev, V.A. Identification of Integrated Rating Mechanisms as Optimization Problem. In Proceedings of the 2020 13th International Conference Management of Large-Scale System Development, MLSD 2020, Moscow, Russia, 28–30 September 2020; p. 9247638. [CrossRef]

45. Larichev, O.I.; Moshkovich, H.M. *Verbal Decision Analysis for Unstructured Problems*; Springer: New York, NY, USA, 2013; 272p. [CrossRef]

46. Alekseev, A. Identification of integrated rating mechanisms based on training set. In Proceedings of the 2020 2nd International Conference on Control Systems, Mathematical Modeling, , Automation and Energy Efficiency (SUMMA), Lipetsk, Russia, 11–13 November 2020. [CrossRef]

47. Alekseev, A. Identification of integrated rating mechanisms with non-serial structures of criteria tree. In Proceedings of the 2021 3rd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA), Lipetsk, Russian, 10–12 November 2021. [CrossRef]

48. Sergeev, V.A.; Korgin, N.A. Identification of integrated rating mechanisms as an approach to discrete data analysis. *IFAC-PapersOnLine* **2021**, *54*, 134–139. [CrossRef]

49. Rabchevsky, A.N.; Yasnitsky, L.N. The Role of Synthetic Data in Improving Neural Network Algorithms. In Proceedings of the 2022 4th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA), Lipetsk, Russian, 9–11 November 2022. [CrossRef]

50. Vereskun, V.D.; Guda, A.N.; Butakova, M.A. Data mining: Discretization of attribute values using the theory of rough sets and clustering. *Vestn. Rostov. Gos. Univ. Putej Soobshcheniya* **2018**, *3*, 76–84.

51. Alekseev, A.; Salamatina, A.; Kataeva, T. Rating and Control Mechanisms Design in the Program "Research of Dynamic Systems". In Proceedings of the 2019 IEEE 21st Conference on Business Informatics (CBI), Moscow, Russia, 15–17 July 2019. [CrossRef]

52. Barnett, J.; Carreia, H.; Johnson, P.; Laughlin, M.; Willsom, K. Darwin Meets Graph Theory on a Strange Planet: Counting Full n-ary Trees with Labeled Leafs. *Ala. J. Math.* **2010**, *35*, 16–23.

53. Lutsenko, E.V. Conceptual principles of the system (emergent) information theory and its application for the cognitive modelling of the active objects (entities). In Proceedings of the 2002 IEEE International Conference on Artificial Intelligence Systems (ICAIS 2002), Divnomorskoe, Russia, 5–10 September 2002. [CrossRef]

54. Lutsenko, E.V.; Troshin, L.P.; Zviagin, A.S.; Milovanov, A.V. Application of the automated system-cognitive analysis for solving problems of genetics. *J. Mech. Eng. Res. Dev.* **2018**, *41*, 1–8. [CrossRef]

55. Lutsenko, E.V. Personal Intelligent Online Development Environment "Eidos-X Professional" (System "Eidos-Xpro"). Programmy Dlia Evm. Bazy Dannykh. Topologii Integral'Nykh Mikroskhem [Computer Programs. Database. Topologies of Integrated Circuits]. Computer Program RU 2022615135, 3 March 2022.

56. Cherepanov, F.M.; Yasnitsky, L.N. "Neurosimulator 5.0". Programmy Dlia Evm. Bazy Dannykh. Topologii Integral'Nykh Mikroskhem [Computer Programs. Da-Tabase. Topologies of Integrated Circuits]. Computer Program RU 2014618208, 12 July 2014.

57. Gusev, A.L.; Okunev, A.A. "Software Package That Implements the Operation of Incompleatly Connected Neural Networks". Programmy Dlia Evm. Bazy Dannykh. Topologii Integral'Nykh Mikroskhem [Computer Programs. Database. Topologies of Integrated Circuits]. Computer program RU 202166563, 20 September 2021.

58. Okunev, A.A. Functional data preprocessing application to oil-transfer pumps vibration parameters forecasting. *Appl. Math. Control Sci.* **2020**, *3*, 51–72. [CrossRef] [PubMed]

59. Kozhemyakin, L.V. Application of neural networks in simulation of cluster-network relations in oil and gas industry. *Appl. Math. Control Sci.* **2020**, *4*, 137–152. [CrossRef]

60. Makarenko, A.V. Control Theory (additional chapters). In *Artificial Neural Networks*; Novikov, D.A., Ed.; LENAND: Moscow, Russia, 2019; pp. 426–455.

61. LeCun, Y.; Bottou, L.; Orr, G.B.; Müller, K.-R. Efficient BackProp. *Lect. Notes Comput. Sci.* **2012**, *7700*, 9–48. [CrossRef]

62. Press, W.H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P. Numerical Recipes. In *The Art of Scientific Computing*, 3rd ed.; Cambridge University Press: Cambridge, UK, 2007.
63. Yasnitsky, L.N. *Artificial Intelligence. Elective Course: Textbook*; BINOM Knowledge Lab.: Moscow, Russia, 2011; 197p.