



# The examination of the effect of the criterion for neural network's learning on the effectiveness of the qualitative analysis of multidimensional data

Dariusz Jamróz<sup>1</sup> 

Received: 25 August 2018 / Revised: 8 January 2020 / Accepted: 12 January 2020 /  
Published online: 3 February 2020  
© The Author(s) 2020

## Abstract

A variety of multidimensional visualization methods are applied for the qualitative analysis of multidimensional data. One of the multidimensional data visualization methods is a method using autoassociative neural networks. In order to perform visualizations of  $n$ -dimensional data, such a network has  $n$  inputs,  $n$  outputs and one of the interlayers consisting of two outputs whose values represent coordinates of the analyzed sample's image on the screen. Such a criterion for the network's learning consists in that the same value as the one at the  $i$ th input appears at each  $i$ th output. If the network is trained in this way, the whole information from  $n$  inputs was compressed to two outputs of the interlayer and then decompressed to  $n$  network outputs. The paper shows the application of different learning criteria can be more beneficial from the point of view of the results' readability. Overall analysis was conducted on seven-dimensional real data representing three coal classes, five-dimensional data representing printed characters, 216-dimensional data representing hand-written digits and, additionally, in order to illustrate additional explanations using artificially generated seven-dimensional data. Readability of results of the qualitative analysis of these data was compared using the multidimensional visualization utilizing neural networks for different learning criteria. Also, the obtained results of applying all analyzed criteria on 20 randomly selected sets of multidimensional data obtained from one of the publicly available repositories are presented.

**Keywords** Multidimensional data analysis · Data mining · Multidimensional visualization · Self-organized neural network · Autoassociative neural network

## 1 Introduction

A qualitative analysis of multidimensional data constitutes an effective and increasingly used tool used for searching different real data properties. A variety of multidimensional visualization methods are applied for the qualitative analysis of multidimensional data. Thanks to

---

✉ Dariusz Jamróz  
jamroz@agh.edu.pl

<sup>1</sup> Department of Applied Computer Science, AGH University of Science and Technology, al. Mickiewicza 30, Kraków 30-059, Poland

this, we can observe some searched data properties in a way which is the most natural for us—through the sense of sight. In general, such a visualization consists in transformation of a multidimensional space containing data into a two-dimensional space representing a screen so as not to lose these data properties significant from the point of view of the conducted analysis. The multidimensional visualization and qualitative analysis utilizing such a visualization attracts a growing interest [1–3]. Various methods are used for such a visualization. A relatively new method is the perspective-based observational tunnels method [4]. It intuitively consists in the prospective parallel projection with the local orthogonal projection along the additional use of perspective. The PCA method [5,6] uses the orthogonal projection onto two eigenvectors of the data set covariance matrix corresponding to two eigenvalues largest in terms of module. Multidimensional scaling [7,8] constructs a representation of the multidimensional space into a two-dimensional space in such a way that for each pair of points, their mutual distance in the target space and in the source space is as close as possible. The method of relevance maps [9,10] performs the multidimensional space transformation into a two-dimensional image in such a way that the distance of the image of a given data vector from each of specially created points  $P_i$  is as close as possible to the  $i$ th coordinate of the data vector to which the image refers. In the method of parallel coordinates [11,12],  $n$ -coordinate axes are distributed in parallel next to each other. In this method, each point is represented by a polyline going through each of the axes in a point corresponding to the appropriate coordinate. Star graphs [13,14] constitute a similar method, except that in this method,  $n$ -axes go radially outward from one point. Visualizations of multidimensional data are also performed using tSNE and its variations [15,16]. The main advantage of this method is the fact that if points are close to each other in the multidimensional space, then they will also be close to each other after reducing the number of dimensions.

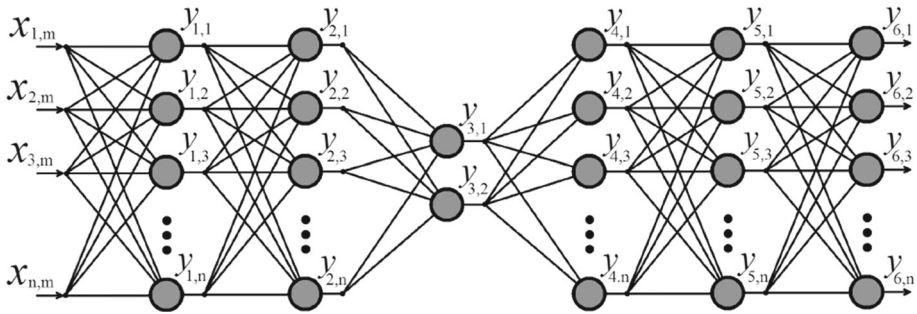
Neural networks are also used for multidimensional visualizations of data. In general, the initial data processing and choice of parameters significantly affect the operation of neural networks. Therefore, apart from the basic models of neural networks, many strategies of designing and improving these networks were developed. Many approaches to learning divide the set into two parts: a training set and validation set. By comparing the accuracy on a validation set with accuracy on a training set, it can be stated that the network requires further learning or is already over-learned [17]. The manual selection of the network parameters requires a great deal of experience, knowledge, effort and is complicated; therefore, guidebooks devoted to this subject were written [18,19]. The automatic selection of the network parameters is also possible [20]; however, it is computationally complex. Thus, from the practical point of view, the possibility to obtain visual information presenting how the network operates inside [21] is valuable during designing the network. This improves the network operation and even alter its structure [22] and conduct the analysis of the role of separate neurons [23]. The analysis through visualization is also carried out on convolutional neural networks (CNNs) [24,25]. The visualization of the network's hidden layers is performed also using PCA [26]. However, this type of visualization is different from the one performed in the paper, because its main goal is optimization of the network operation, not the data analysis. In the case of the visualization of multidimensional data, self-organizing neural networks (Kohonen maps, SOM), in which learning occurs without any trainer, are usually applied [27,28]. In their case, during learning weights are modified in such a manner to additionally enhance the response of the winning neuron and its neighbors. The winning neuron is the one whose response on a given sample is the strongest. Autoassociative neural networks, in which one of the interlayers comprises two neurons serving for the visualization [28–30], are also applied for the visualization. Researchers are still improving the approach utilizing networks for the visualization of multidimensional data [31–34]. There are also

attempts to significantly accelerate the learning process of such networks [35]. A relatively new approach is the use of restricted radius basis function networks [36,37]. In this approach, one part of the network is trained without any trainer and the other one with a trainer. The significance and popularity of neural network studies results from the fact that the general idea of networks and mechanisms similar to those in neural networks, including different variations of neural networks, are used for solving many problems in real world. They can be applied to search for parking spaces, where wireless sensor networks (WSNs) were used and gradient ascent method along with gradient field in the parking space structures [38]. Another example includes the use of self-learning interval type-2 fuzzy neural network (SLIT2FNN) for designating robot trajectories [39].

In this paper, criteria, which can be applied in any artificial neural network or its modification if only such a network can comprise many layers and can operate as a network with a trainer and (applying the criterion comparing the network's outputs with its inputs) as the autoassociative neural network, are applied. As a result, in each such network all analyzed criteria can be applied, so the analogical comparison of the criteria analyzed in the paper can be made. Therefore, the paper is focused on the comparison of the effectiveness of the one, adopted type of network with different criteria. Only the criterion of the autoassociative neural networks constitutes the classic use of neural networks to visualize multidimensional data. The remaining criteria constitute a new approach to the use of neural networks for the visualization of multidimensional data. From among the criteria presented in the paper, only the criterion of the autoassociative neural network was used in the previous papers by the author. Studies based on the remaining criteria used in the paper are new, and the idea to utilize such criteria results from the experience gained in the course of numerous studies. For the first time, the comparison of the effectiveness of methods utilizing the above criteria using the evaluating criterion of the qualitative analysis readability introduced in earlier papers was made in this paper.

## 2 Learning criteria

In the paper, neural networks applying different learning criteria were used for the qualitative analysis of multidimensional data through their visualization. Networks in the case of the analyzed  $n$ -dimensional data had  $n$  inputs and one of the interlayers consisting of two outputs used directly for the visualization. All used networks consisted of six layers in total. Three layers served to change  $n$  inputs into two outputs of the interlayer used for the visualization. Then, three layers served to change two outputs of the interlayer into the appropriate number of network outputs which depends on the assumed criterion. After the completion of learning, data visualization consisted in providing each  $n$ -dimensional data vector to  $n$  neural network inputs and displaying a point with coordinates equal to values obtained from two outputs of the interlayer used for the visualization on the screen. All networks were trained by the method of error back-propagation. And these networks were different in the assumed learning criterion and the number of outputs resulting from it. The method of error back-propagation has been used successfully for many years, and there are many publications devoted to this method which describe both its theoretical background and practical application [40,41].



**Fig. 1** A diagram of the autoassociative neural network used for the visualization of multidimensional data. Three layers served to change  $n$  inputs into two outputs of the interlayer used for the visualization. Then, three layers served to change two interlayer outputs into  $n$  network outputs. Signals obtained from the interlayer consisting of two neurons constitute coordinates of a screen for the image of the analyzed data vector

## 2.1 The criterion of the autoassociative neural network

It is a classic example of an autoassociative neural network used for the visualization of multidimensional data in which the number of outputs is equal to the number of inputs. In connection with the above, this network had  $n$  outputs for the analyzed data. It is an example of self-organized neural networks, learning of which occurs in an unsupervised way. The learning criterion is the same value as the one at the  $i$ th input appears at each  $i$ th output. If the network is trained in this way, this means that the whole information from  $n$  inputs was compressed to two outputs of the interlayer and then decompressed to  $n$  network outputs. This means that by utilizing the interlayer consisting of two outputs it is possible to present the whole information concerning the analyzed data on a two-dimensional computer screen. A diagram of such a network is presented in Fig. 1.

## 2.2 The criterion in which the network has outputs representing separate classes

With this criterion, the network has as many outputs as there are classes. The learning criterion is such that values at outputs represent separate classes. It is realized that one output is attributed to each class. If a sample belongs to a given class, then one should appear at the output attributed to this class and 0 at the remaining outputs. If the network is trained in this way, the whole information concerning the division into classes was compressed to two outputs of the interlayer and then processed to the appropriate number of network outputs. Such a network is an example of neural networks in the case of which learning occurs in a supervised way. A diagram of such a network differs from the one presented in Fig. 1 in only that the last layer of the network comprises as many neurons as there are classes.

## 2.3 The criterion in which the network has one output whose values represent separate classes

With this criterion, the network has one output. The learning criterion is such that values of the output represent separate classes. It is realized that the next index is attributed to each class occurring in the data. If a sample belongs to a given class, then the value equal to the index attributed to this class divided by the number of classes occurring in the data should appear

at the output. If the network is trained in this way, this means that the whole information concerning the division into classes was compressed to two outputs of the interlayer and then processed to 1 network output. Such a network is an example of neural networks in the case of which learning occurs in a supervised way. A diagram of such a network differs from the one presented in Fig. 1 that the last layer of the network comprises one neuron.

## 2.4 The criterion in which the network has one output representing a random value attributed to a given sample

The learning criterion is such that the value at the output is equal to a random value attributed to a given sample. Random values are attributed to each sample before learning starts. During training of such networks, the whole information about the dependence between samples and random values attributed to them was compressed to 2 outputs of the interlayer and then processed to 1 network output. A diagram of such a network differs from the one presented in Fig. 1 that the last layer of the network comprises one neuron.

## 3 The algorithm

The network's learning is the first thing which must be done in order to obtain possibilities of multidimensional data visualization. At this stage, weights of all neurons are counted. Before learning starts, input data must be scaled in such a way that it is contained within the range of neuron outputs' values. Because the hyperbolic tangent function was assumed at the output of each neuron, output values are contained within range  $(-1, 1)$ . Values of each data vector coordinates were thus scaled to the range  $(-0.9, 0.9)$ . Before learning starts, initial values of all weights attributed to all neurons must also be determined. Each weight was attributed a random value from range  $(-0.5, 0.5)$ . The network's learning procedure described in the form of pseudocode is presented as Algorithm 1. This whole procedure can be repeated many times in order to better train the neural network. As can be observed in Algorithm 1, at first, for the  $m$ th data vector, we calculate the output value of all neurons in the first layer:

$$y_{1,j} = g \left( w_{1,j,0} + \sum_{k=1}^n w_{1,j,k} x_{k,m} \right) \quad (1)$$

Let  $g$  denote the assumed nonlinear function (hyperbolic tangent was used in the conducted experiments),  $n$ —the number of network inputs, and  $y_{1,j}$ —the output value of a neuron placed in the first network layer at the  $j$ th position. Similarly,  $x_{k,m}$  denotes the  $k$ th coordinate of  $m$ th input data set vector,  $w_{1,j,k}$  denotes weight of the  $k$ th input of a neuron placed in the first network layer at the  $j$ th position, and the weight  $w_{1,j,0}$  for input number 0 denotes the additional constant component. Hence, the weight plays a very important role in increasing possibilities of each neuron by enabling the calculation of the value of its output on the basis of not only variables but also a constant. It plays an analogical role to, e.g., coefficient  $c$  in function  $y = ax^2 + bx + c$ .

Then, we calculate the output values of all neurons located in the remaining network layers. The calculation of neuron outputs can occur after the calculation of output values of the previous layer's neurons:

**Algorithm 1**

```

for all  $m$ -th input data vector do
  for  $j = 1$  to the number of neurons of the first layer do
     $y_{1,j} = g \left( w_{1,j,0} + \sum_{k=1}^n w_{1,j,k} x_{k,m} \right)$ 
  end for
  for  $i = 2$  to the number of layers of the network do
    for  $j = 1$  to the number of neurons of layer  $i$  do
       $y_{i,j} = g \left( w_{i,j,0} + \sum_{k=1}^{size(i-1)} w_{i,j,k} y_{i-1,k} \right)$ 
    end for
  end for
  for  $j = 1$  to the number of neurons of the last layer do
     $i =$  the number of the last network layer
     $z_{j,m} =$  the value resulting from the criterion related to  $j$ -th output of the  $m$ -th input data set vector
     $\delta_{i,j} = \left( 1 - y_{i,j}^2 \right) \left( z_{j,m} - y_{i,j} \right)$ 
  end for
  for  $i =$  the number of layers of the network - 1 downto 1 do
    for  $j = 1$  to the number of neurons of layer  $i$  do
       $\delta_{i,j} = \left( 1 - y_{i,j}^2 \right) \sum_{k=1}^{size(i+1)} \left( \delta_{i+1,k} w_{i+1,k,j} \right)$ 
    end for
  end for
  for  $i = 1$  to the number of layers of the network do
    for  $j = 1$  to the number of neurons of layer  $i$  do
      for  $k = 0$  to the number of neurons of layer  $i - 1$  do
         $\tilde{w}_{i,j,k} = w_{i,j,k} + \eta \delta_{i,j} y_{i-1,k}$ 
      end for
    end for
  end for
end for

```

$$y_{i,j} = g \left( w_{i,j,0} + \sum_{k=1}^{size(i-1)} w_{i,j,k} y_{i-1,k} \right) \tag{2}$$

$size(i - 1)$  denotes the number of neurons in layer number  $i - 1$ ,  $y_{i,j}$  denotes the output value of a neuron placed in the  $i$ th network layer at the  $j$ th position,  $w_{i,j,k}$  denotes the weight of the  $k$ th input of a neuron placed in the  $i$ th network layer at the  $j$ th position, the weight  $w_{i,j,0}$  denotes analogical meaning as in formula 1, and  $g$  denotes nonlinear function, the same as in formula 1.

In the next part of the algorithm, we calculate errors of the network output. For this purpose, we calculate the difference between the values we obtained at the outputs of the last network layer neurons and values we should obtain. Values we should obtain depend directly on the assumed criterion:

- for the autoassociative neural network, for each vector, they will be directly its coordinates, that is values provided to the network inputs.
- for the criterion in which the network has outputs representing separate classes, they will be combinations of zeros and ones resulting from the vector’s affinity to a given class.
- for the criterion in which the network has one output whose values represent separate classes, for each vector, the value equal to the index of a class to which it belongs divided by the number of classes.
- for the criterion in which the network has one output representing a random value attributed to a given sample, for each vector, this will be a random value attributed to this vector before the network’s learning starts.

We multiply the difference obtained depending on the selected criterion by the derivative of assumed function  $g$ , that is by the derivative of the hyperbolic tangent function, and we obtain:

$$\delta_{i,j} = \left(1 - y_{i,j}^2\right) \left(z_{j,m} - y_{i,j}\right) \tag{3}$$

$\delta_{i,j}$  denotes the value of the calculated error of the output of a neuron placed in the  $i$ th network layer at the  $j$ th position (in this formula  $i$  denotes the number of the last network layer),  $y_{i,j}$  denotes the value of the output of the  $j$ th neuron from the  $i$ th layer, and  $z_{j,m}$  denotes the value resulting from the criterion related to  $j$ th output of the  $m$ th input data set vector. In the order from the penultimate layer to the first layer, we calculate errors of outputs of neurons from the remaining network layers.

$$\delta_{i,j} = \left(1 - y_{i,j}^2\right) \sum_{k=1}^{size(i+1)} \left(\delta_{i+1,k} w_{i+1,k,j}\right) \tag{4}$$

where  $\delta_{i,j}$ —the value of the calculated error of the output of a neuron placed in the  $i$ th network layer at the  $j$ th position,  $w_{i+1,k,j}$ —the weight of the  $j$ th input of the  $k$ th neuron from layer  $i + 1$ ,  $size(i + 1)$ —the number of neurons in layer number  $i + 1$ , and  $y_{i,j}$ —the output value of a neuron placed in the  $i$ th network layer at the  $j$ th position. Based on the previously calculated errors, we modify weights of all network neurons:

$$\tilde{w}_{i,j,k} = w_{i,j,k} + \eta \delta_{i,j} y_{i-1,k} \tag{5}$$

where  $w_{i,j,k}$  denotes the weight of the  $k$ th input of the  $j$ th neuron from the  $i$ th layer,  $\delta_{i,j}$  denotes the value of the error of the output of the  $j$ th neuron from the  $i$ th layer,  $y_{i-1,k}$  denotes the output value of the  $k$ th neuron from layer  $i - 1$ ,  $y_{0,k}$  denotes the  $k$ th coordinate of  $m$ th input data set vector  $x_{k,m}$ , and  $\eta$  denotes the parameter specifying the learning rate.

After the completion of learning, we may proceed to obtain the view of multidimensional data sets. For each  $m$ th data vector, we calculate values of outputs of subsequent neuron layers using formulas 1 and 2. We conduct these calculations up to the moment we obtain values of two outputs of neurons belonging to the interlayer used for the visualization. These two values constitute directly coordinates of the location of a screen on which the image of the  $m$ th data vector should be drawn. In this way, we can draw images representing all vectors belonging to the multidimensional data set on the screen. It can be observed that in the presented algorithm and equations occurring in it, the applied neural networks can be very easily paralleled because the whole network layer can be calculated in parallel. In order to accelerate the operation of the visualization of multidimensional spaces, the author repeatedly applied both parallel programming, writing programmes using CUDA (Compute Unified Device Architecture) under GPU (graphics processing unit) and the equipment-based execution of the algorithm (the parallel stream processors specialized for visualization), creating hardware structures in VHDL (Very High Speed Integrated Circuits Hardware Description Language) and executing them on FPGA (field-programmable gate array). However, in the case of networks described in the paper and used for the visualization, in the course of the conducted studies, there was no situation in which accelerating their operation would be purposeful. Learning parameters were always selected in such a manner that time needed for the sufficient training of the network was acceptable.



## 4 Experiments' results

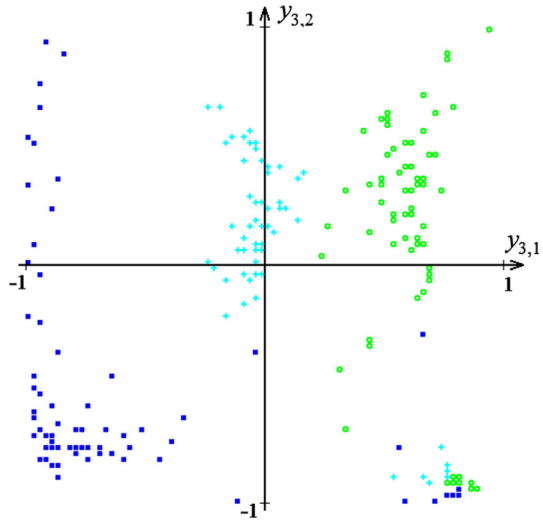
Applying the above-mentioned criteria for neural networks' learning, at first, seven-dimensional real data describing different energy classes of coal are presented on the computer screen. These data were already used multiple times before, also during the examination of numerous methods of qualitative analysis of multidimensional data through their visualization [42]. The complete data set was published earlier in [43]. According to the Polish classification, the data set consisted of 205 samples, out of which 72 samples represented class of coal 31 (energetic coal), 61 samples represented class of coal 34.2 (semi-coking coal) and 72 samples represented class of coal 35 (coking coal). Each sample was described by seven features: density, mass, combustion heat, ash contents, sulfur contents, volatile matter contents and analytical moisture. Therefore, these data can be interpreted as a set placed in the seven-dimensional space of features. A special system was created to obtain the results presented in the paper. This system was created based on the algorithm presented in Sect. 3 using programming language C++. The purpose of the qualitative analysis of the presented data was to state whether samples belonging to different classes of coal occupy separate subareas of the multidimensional space of features. This allows to state whether selected features are sufficient for the correct recognition of the class of coal.

Figures 2, 3, 4, 5, 6 and 7 present the obtained views of the analyzed seven-dimensional data using different criteria for the neural network's learning. In each of these figures, signals representing samples of coal of class 31 are marked with a symbol of a square ( $\square$ ), samples of coal of class 34.2—plus (+), samples of coal of class 35—circle (o). Figure 2 presents the result obtained using the criterion for the autoassociative neural network's learning. Parameter ITER denotes the number of repetitions of the network's learning. It can be observed in the figure that signals ( $y_{3,1}$ ,  $y_{3,2}$ ) of the interlayer used for the visualization, being the response to data representing samples of coal of a given class accumulate in aggregations. It can be seen that these aggregations can be separated from each other. Classes 34.2 and 35 formed two aggregations of points each, while class 31 formed two aggregations of points in the figure and the third subarea of the figure occupied by two points. As a result, by using the analyzed criterion, we can indicate the possibility to divide the space of features into subareas occupied by different classes. It must be noted that in the case of autoassociative neural networks, information on belonging of data vectors to specific classes is not used during learning. In this situation, the way in which signals going through the layer consisting of two neurons, representing a given class, will be grouped depends only on some properties of this data observed by the network.

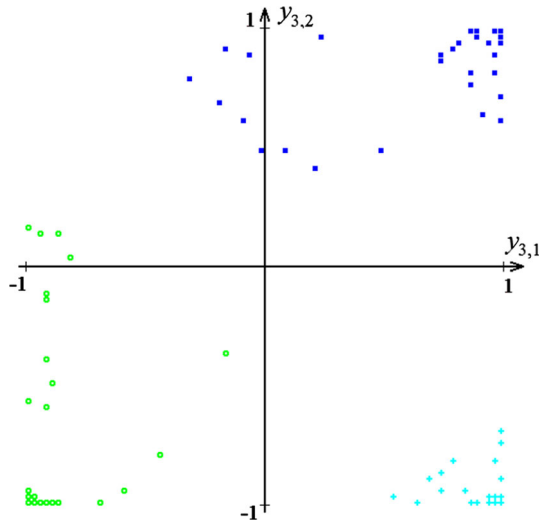
Figure 3 presents the result obtained using the learning criterion in which the network has three outputs representing separate coal classes. It can be observed in the figure that signals being the response to data representing samples of coal of a given class accumulate in aggregations. Furthermore, it can be seen that these aggregations can be separated from each other that makes it significantly more readable than with the use of earlier criteria. Each of the three classes formed one aggregation. As a result, by using the analyzed criterion, we can indicate the possibility to divide the space of features into subareas occupied by different classes in a way more readable than with the use of the previous criteria. In the case of the used criterion, information on belonging of data vectors to specific classes is used during learning. We may thus indicate whether the network correctly learned to recognize classes or not. Then, why do we need to conduct the additional analysis concerning the possibility to divide the space of features into subareas occupied by different classes? However, it must be noted that in the case of such networks, conclusions drawn from the visualization obtained by



**Fig. 2** The view of the analyzed coal data obtained using the criterion for the autoassociative neural network's learning with parameter ITER = 21004



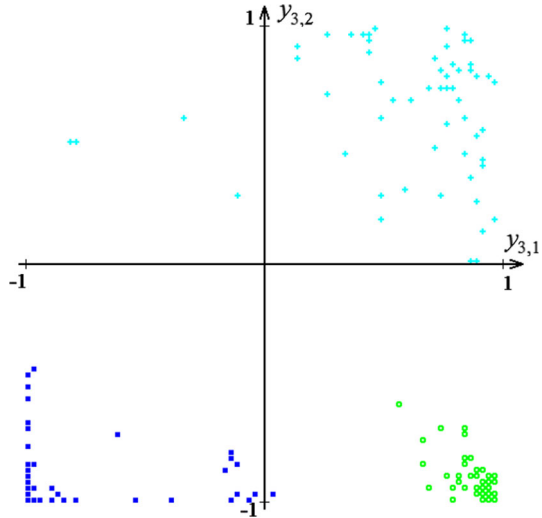
**Fig. 3** The view of the analyzed coal data with the learning criterion in which the network has three outputs the values at which are to represent separate classes of coal with parameter ITER = 1000



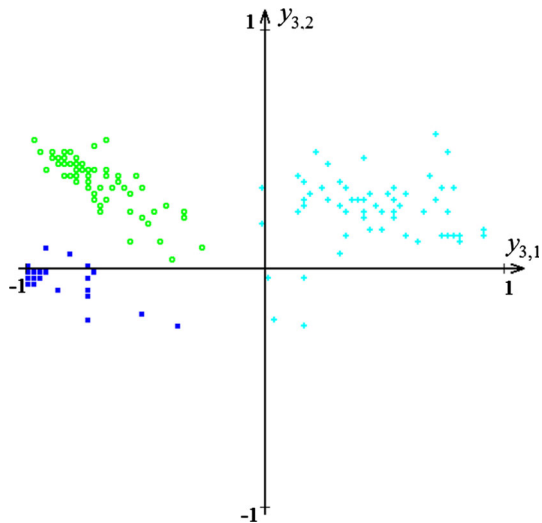
these networks can be independent of the fact whether learning in the case of such a network was successful or not. It may turn out that despite the lack of success in the network's learning, the view from the interlayer will allow to indicate the possibility to easily separate areas containing different classes. On the other hand, even when the neural network learns to correctly recognize a learning sequence, this does not mean that easy separation of areas containing different classes is possible. Even then they can overlap. As a result, the learning fact itself does not allow to state that the selected features are sufficient for the possibility to correctly and easily distinguish areas occupied by different classes.

Figure 4 presents the result obtained using the learning criterion in which the network has one output at which the appropriate values are to represent separate coal classes. It can be observed in the figure that signals being the response to data representing samples of coal of a given class accumulate in aggregations. It can be seen that these aggregations can be

**Fig. 4** The view of the analyzed coal data with the learning criterion in which the network has one output at which the appropriate values are to represent separate classes of coal with parameter ITER = 4000



**Fig. 5** The view of the analyzed coal data with the learning criterion and number of learning repetitions the same as in Fig. 4, but with different initial values of the drawn network weights



separated from each other to make it more readable as compared to the earlier criterion. Each of the three classes formed one aggregation. As a result, by using the analyzed criterion, we can indicate the possibility to divide the space of features into subareas occupied by different classes in a way equally readable as with the use of the previous criterion. The described criterion differs from the previous one in the number of network outputs with the network learning to reply to the question: what class we are to deal with. As it can be seen, this does not cause any change in the readability of results.

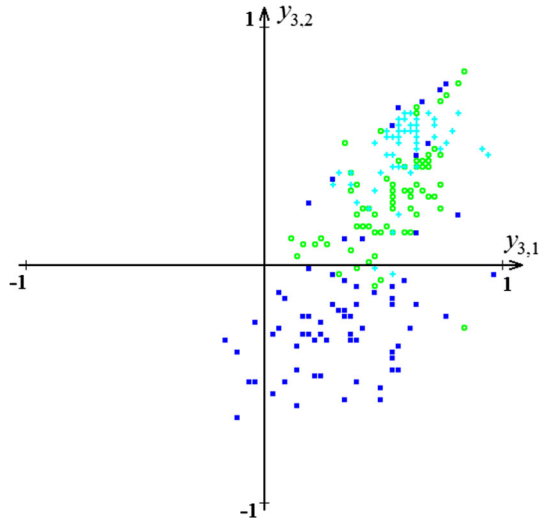
Figure 5 presents the result of the operation of the same network as in Fig. 4, but with different initial values of the drawn network weights. Despite a different distribution of point aggregations, the obtained view allows to indicate the possibility to divide the space of features into subareas occupied by different classes in an equally readable way.

The idea for the last criterion appeared as a result of experiments related to the visualization of multidimensional data. There were situations in which the obtained view became more readable and then less readable as a result of the network's learning. As a result of the further learning, the view became more readable again. Thus, a question, what effect the degree of the network's learning has on the readability of the visualization results, appeared. It turned out that we do not have to aim at the best network's training to obtain correct conclusions resulting from the qualitative analysis—it does not make any difference. The above conclusion is inconsistent with the general approach to using neural networks, where the indicator of the correct network operation is usually the degree of its learning. However, in such a general case we want the network to correctly recognize objects belonging to the input space. Therefore, in such a general situation, the degree of the network's learning (indicator specifying the correct responses) is significant. However, the qualitative analysis using the visualization of multidimensional data is a completely different case. Here, the degree of the network's learning does not matter at all. However, the indication of the possibility to divide the input space in terms of belonging to classes is of significance. If in a given moment we can indicate this based on the obtained view, this means that there is a representation allowing to divide such a space. The representation realized by the network part used for the visualization is such a representation. The existence of this representation is thus irrespective of the degree of the network's learning. It is possible to imagine a situation in which the network without learning allows to indicate the possibility to divide the space of features—an example of such a situation will be presented in the further part. The criterion in this situation can only be a “pretext” enabling to change the network into a specified direction, not necessarily consistent with the improvement in the readability of the visualization results. For the qualitative analysis purposes, it is sufficient that there will be a moment somewhere on the way in which the readability of the visualization results will be satisfying. For the same reason, the control of over-learning is of no significance. It also turns out that the network outputs are over-learned and yet the visualization layer shows readable views. This may be caused by the fact that layers located behind the visualization layer were subjected to over-learning, and thus, their over-learning does not affect the readability of results of the visualization layer. Therefore, in this type of application of the network, over-learning should be redefined as a situation in which the readability of results of the visualization layer is continuing to deteriorate.

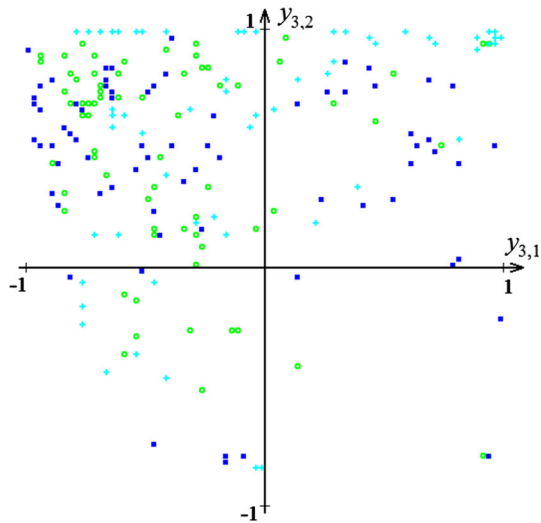
Figure 6 presents the view obtained using the learning criterion in which the network has one output at which the value is to be equal to a random value attributed to a given sample. Random values are attributed to each sample before learning starts. It can be observed in the figure that signals being the response to data representing samples of coal of a given class accumulate in aggregations. These aggregations partially overlap, thus based on the obtained view we cannot indicate the possibility to separate them from each other. However, the mere fact that with the criterion coercing random values at the network output, there are clear aggregations of points representing the same classes formed in the picture is very interesting. How is it possible that with the same criterion, we obtain any ordering? Let us analyze the results obtained in the next figures to explain this.

Figure 7 presents the view obtained as a result of the further network's learning. It can be seen that the readability of results deteriorated completely—signals being the response to data representing samples of coal of all classes is mixed up with each other. Based on the obtained view, it is not possible to observe grouping of signals being the response to data representing samples of coal of the same class. Better results were not obtained as a result of further learning. As a result, the network with the assumed criterion aims to randomly distribute points in the space of signals. It is consistent with intuition—if the criterion is based on random values, then the space of signals will also aim at random values. But why during

**Fig. 6** The view of the analyzed coal data with the learning criterion in which the network has one output at which the value is to be equal to a random value attributed to a given sample, with parameter ITER = 3990



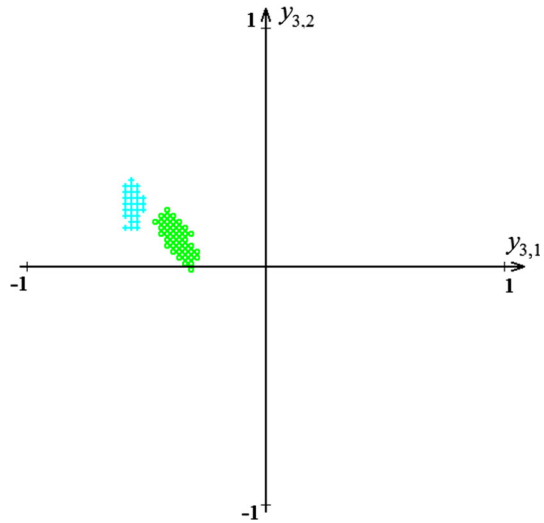
**Fig. 7** The view of the analyzed coal data with the same learning criterion as in Fig. 6 obtained as a result of the further network's learning with parameter ITER = 7000



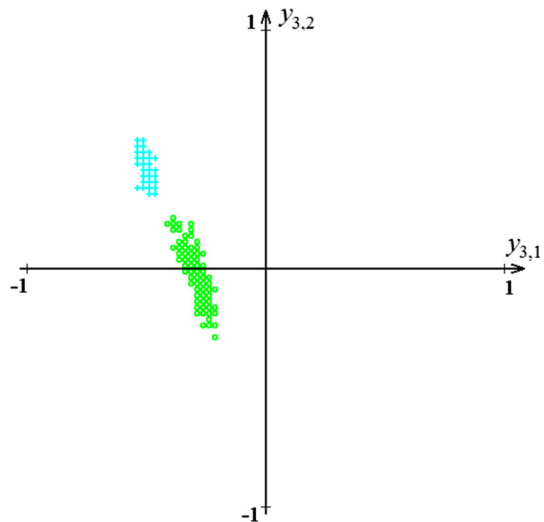
learning, before more accurate network's learning, did we obtain views in which clearly ordered aggregations were formed?

Artificially generated seven-dimensional data were used to explain the raised questions. These data were prepared in such a way that points belonging to two classes were placed in separate subareas of a seven-dimensional space. Each of these subareas contained a shape of a seven-dimensional cube. The criterion for network's learning was the same as in Figs. 6 and 7, that is for each data vector, a value is to appear at the network output as a result of learning was drawn. Figure 8 presents the view obtained for artificially generated data before the network's learning starts. As a result, values of weights at the moment of obtaining this view were equal to a random value. It can be seen that the view obtained using the network without its learning allows to indicate the possibility to divide the space into subareas occupied by different classes. Figure 9 presents that the view becomes even more readable as a result of

**Fig. 8** The view of artificially generated seven-dimensional data with the same learning criterion as in Figs. 6 and 7 obtained before the network's learning started, that is with parameter ITER = 0



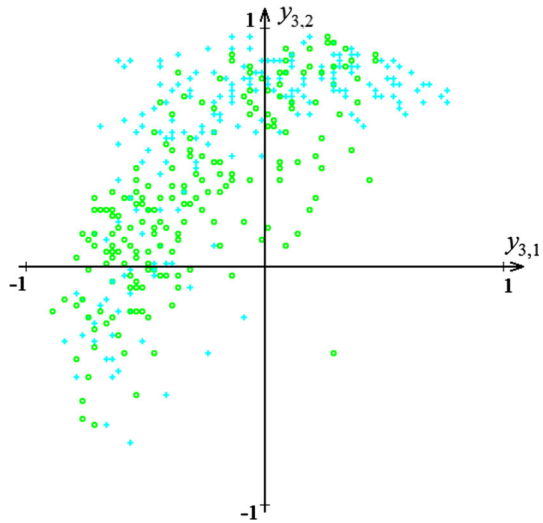
**Fig. 9** The view of the same artificially generated seven-dimensional data as in Fig. 8 with the same learning criterion, obtained as a result of the network's learning, with parameter ITER = 100



the network's learning—subareas occupied by different classes move away from each other. However, as a result of further learning, the situation deteriorates and stays so.

Figure 10 presents the view obtained as a result of the further network's learning. It can be observed that signals being the response to data representing different classes are mixed up with each other. Based on the obtained view, it is not possible to indicate the possibility to divide the space into subareas occupied by different classes anymore. Better results were not obtained as a result of further learning. Let us repeat the already raised question: why during learning the network and even for the unlearned network, did we obtain views in which clearly ordered aggregations were formed? The response results directly from the network operation principle presented in Sect. 3. Let us take the initial situation before the network's learning as an example. Then, weights assume random values. It is sufficient that the sum of values of vectors' coordinates of a given class significantly differs from the sum of the value of vectors' coordinates of the remaining classes. Then, these coordinates after

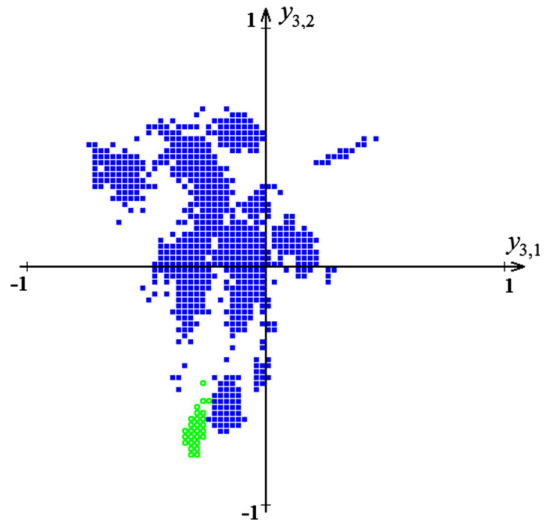
**Fig. 10** The view of the same artificially generated seven-dimensional data as in Figs. 8 and 9 with the same learning criterion, obtained as a result of the network's learning, with parameter ITER = 2000



their introduction into inputs and multiplication by the initial random values of weights, after summing up statistically, will also differ from the remaining ones. Thus, values which for vectors of a given class statistically differ from other classes will appear at outputs of neurons of the first layer. These values will be statistically maintained on their way through the next layers of neurons with random weights. Therefore, before the network's learning starts, in exceptional cases, it is possible to obtain readable results of the qualitative analysis of multidimensional data. The described criterion is also the proof that more accurate network's learning, omitting even the phenomenon of the network over-learning, does not have to cause the increase in the readability of the qualitative analysis results.

By summing up the results of analysis of seven-dimensional data describing different classes of coal, it can be stated that the readability of obtained results strongly depends on the applied criterion for neural network's learning. The readability of obtained results until this moment in the paper was determined in an intuitive way. It can be observed that separation of areas occupied by signals belonging to different classes, e.g., in Fig. 3 is significantly easier than, e.g., in Fig. 2. It must be noted that visualization of multidimensional data in this work is used solely to determine a specific property of this data, that is, to determine whether given features allow to separate the space of features into subareas occupied by different classes. The possibility to reply to the question asked above is thus important from the point of view of the analysis conducted in the paper. For example, it is not important how much of information contained in multidimensional data will be retained in the obtained view. Moreover, it can be assumed that the more general information will be retained, the less readable the result of the specific targeted analysis can be. The example of the autoassociative neural network can confirm such an assumption. After the successful training of network, the complete information from the network inputs appear at its outputs. In order to make it happen, this complete information must go through every layer of the network, and therefore through the layer used for visualization (consisting of two neurons). However, as can be observed in Fig. 2, the view obtained using such a network is not at all more readable among the obtained views. Therefore, to evaluate the effectiveness of qualitative analysis it is best to use the criterion allowing to evaluate solely the readability of this analysis. The example of this is a criterion which consists in drawing a curve separating images of points belonging

**Fig. 11** The view of five-dimensional data obtained as a result of the printed text recognition, obtained using the criterion for the autoassociative neural network learning, with parameter ITER = 49



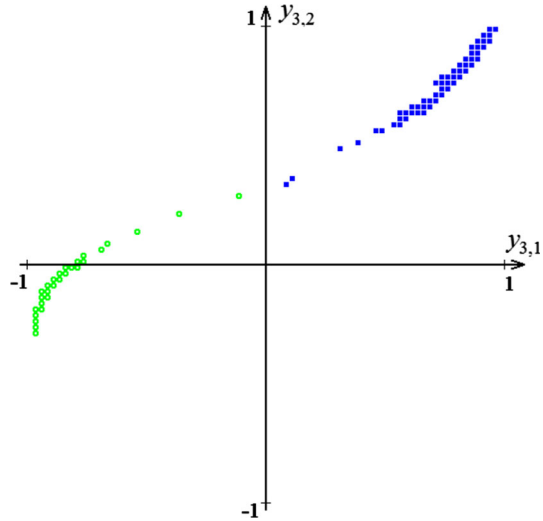
to different classes is shown in figure [42]. The more complicated this curve is, the less readable the view allowing to indicate the possibility to separate subsets of points from each other is. It was assumed that the curve consists of arcs and the more complicated it is, the more inflection points it has. Inflection points are points joining arcs turning into different directions. This criterion was formulated in order to create a ranking of different methods of multidimensional visualization used for the qualitative analysis [42]. Assuming such a criterion, it can be easily noticed that in Figs. 3, 4, 5, 8 and 9 each class can be separated from the remaining ones by means of a curve without inflection points. Such views constitute the most readable result of the qualitative analysis possible to obtain. This means that the use of any other method will not bring more readable views; thus, every other method can be at most as good as the above-mentioned one in the sense of the assumed criterion. It can be easily noticed that in Fig. 2, curve separating a given class from the remaining ones will have inflection points; thus, it is less readable from the point of view of the assumed criterion. In the remaining figures, classes cannot be separated; thus, they are beyond the criterion.

With the use of the previously described criteria for neural networks' learning, also five-dimensional real data obtained as a result of the printed text recognition were presented. The way of obtaining these data is already described [44]. Each sample was created as a result of separating five features; therefore, these data can be interpreted as a set placed in a five-dimensional space of features. This set consisted of 4810 samples representing all alphabet characters. An exemplary goal of the conducted analysis was to determine whether samples representing character "m" occupy separate subareas of a multidimensional space of features. This allows to state whether five analyzed features are sufficient for the correct recognition of this character.

Figures 11, 12, 13 and 14 present the obtained views of the analyzed five-dimensional data using different criteria for the neural network's learning. In each of these figures, signals representing character "m" are marked with a symbol of a circle (o) and signals representing the remaining characters with a square ( $\square$ ). In Figs. 12 and 13, signals representing character "m" can be separated from signals representing the remaining characters by a curve without inflection points. In Fig. 11, signals representing character "m" can be separated from signals



**Fig. 12** The view of five-dimensional data obtained as a result of the printed text recognition with the learning criterion in which the network has two outputs. The value of the first output represents the occurrence of character “m” and value of the second output the remaining characters, with parameter ITER = 190



representing the remaining characters by a curve with one inflection point. In Fig. 14, signals representing character “m” cannot be separated from the remaining characters.

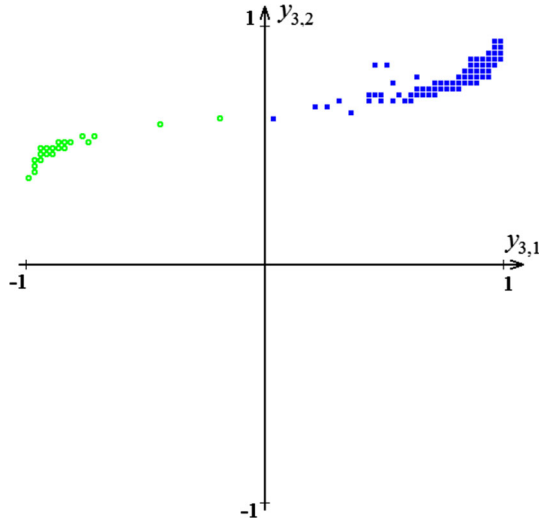
Figure 11 presents the result obtained using the criterion for the autoassociative neural network’s learning. It can be observed that signals of the interlayer used to visualize which are the response to samples representing character “m” accumulated in an aggregation. It can be noticed that this aggregation can easily be separated from signals representing the remaining characters. As a result, by using the analyzed criterion, we can state the possibility to separate a subarea of the space of features occupied solely by samples of character “m.”

Similarly, Fig. 12 presents the result obtained using a learning criterion in which the network has two outputs. Value one obtained at the first output represents the occurrence of character “m,” and value one obtained at the second output represents the occurrence of the remaining characters. It can be observed in this figure in a more readable way than in the previous one that signals to be the response to samples representing character “m” can easily be separated from signals representing the remaining characters. As a result, by using the analyzed criterion, we can state the possibility to separate a subarea of the space of features occupied solely by samples of character “m” in a more readable way than by using the previous criterion.

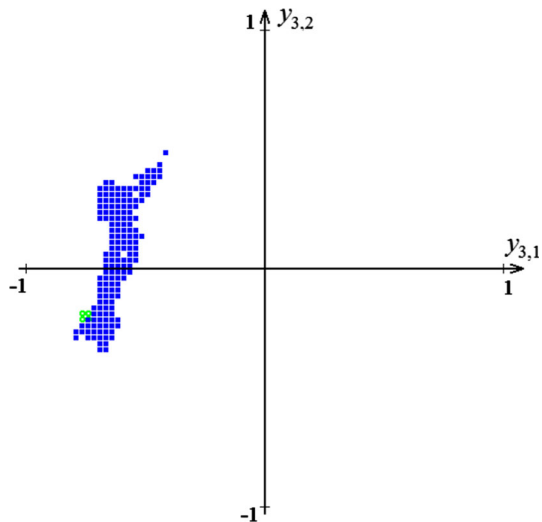
Figure 13 presents the result obtained using the learning criterion in which the network has one output at which the value equal to 1 represents the occurrence of character “m” and value equal to 0 represents the occurrence of the remaining characters. It can be observed in this figure in an equally readable way as in the previous one that signals to be the response to samples representing character “m” can easily be separated from signals representing the remaining characters. As a result, by using the analyzed criterion, we can state the possibility to separate a subarea of the space of features occupied solely by samples of character “m” in an equally readable way as by using the previous criterion.

Figure 14 presents the view obtained using the learning criterion in which the network has one output at which the value is to be equal to a random value attributed to each sample. It can be observed that signals being the response to samples representing character “m” accumulated in an aggregation. This aggregation partially overlaps the area occupied by signals representing the remaining characters. As a result, by using the analyzed criterion,

**Fig. 13** The view of five-dimensional data obtained as a result of the printed text recognition with the learning criterion in which the network has one output at which value 1 represents character “m” and value 0 represents the remaining characters, with parameter ITER = 500



**Fig. 14** The view of five-dimensional data obtained as a result of the printed text recognition with the learning criterion in which the network has one output at which the value is to be equal to a random value attributed to a given sample, with parameter ITER = 300

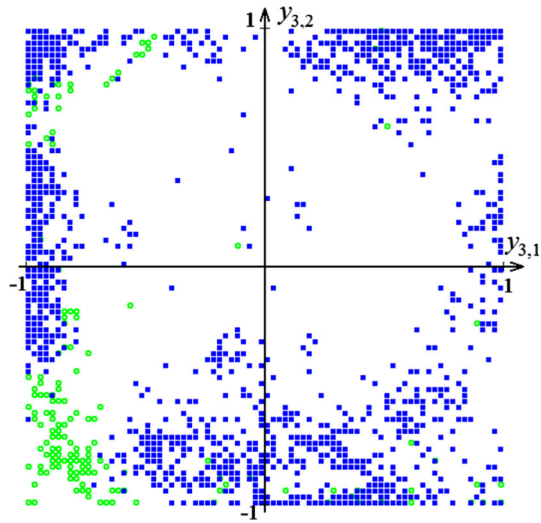


we cannot state the possibility to separate a subarea of the space of features occupied solely by samples of character “m.”

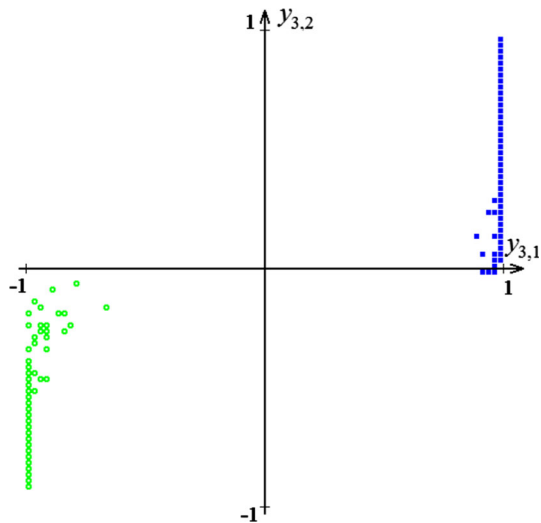
The additional set which is presented by using the criteria for neural networks’ learning described earlier constituted 216-dimensional data. These data were created as a result of separating 216 features (profile correlations) from images presenting hand-written digits that is publicly available [45] as part of the data set called “Multiple Features”. This set consists of 2000 samples representing all digits, 200 samples representing each digit. An exemplary goal of the conducted analysis was to determine whether samples representing digit 1 occupy separate subareas of a multidimensional space of features. This in turn allows to state whether 216 analyzed features are sufficient for the correct recognition of this digit.

Figures 15, 16, 17 and 18 present the obtained views of the analyzed 216-dimensional data using different criteria for the neural network’s learning. Signals representing digit 1 are marked with a symbol of a circle (o) and signals representing the remaining digits with

**Fig. 15** The view of 216-dimensional data representing hand-written digits, obtained using the criterion for the autoassociative neural network learning, with parameter  $ITER = 1000$



**Fig. 16** The view of 216-dimensional data representing hand-written digits, obtained with the learning criterion in which the network has two outputs. The value of the first output represents the occurrence of digit 1 and value of the second output the remaining digits, with parameter  $ITER = 45$

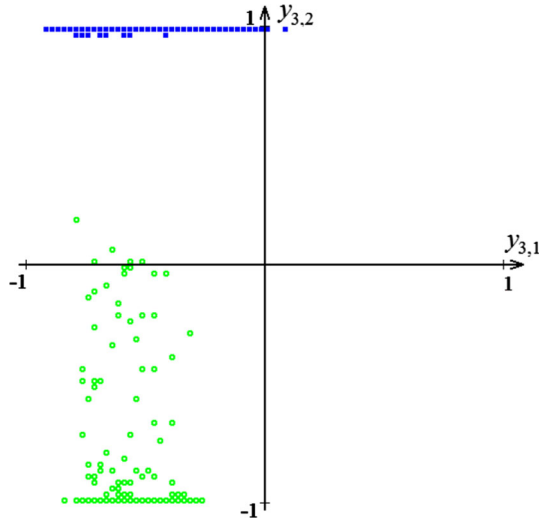


a square ( $\square$ ) as shown in the figures. In Figures 16 and 17, signals representing digit 1 can be separated from signals representing the remaining digits by a curve without inflection points. Similarly, in Figs. 15 and 18, signals representing digit 1 cannot be separated from the remaining digits.

Figure 15 presents the result obtained using the criterion for the autoassociative neural network's learning. It can be observed in it that signals being the response to samples representing digit 1 partly accumulate in aggregations. In many places, these aggregations overlap signals representing the remaining digits. As a result, the use of the analyzed criterion does not allow to state the possibility to separate a subarea of the space of features occupied solely by samples of digit 1.

Figure 16 presents the result obtained using a learning criterion in which the network has two outputs. Value one obtained at the first output represents the occurrence of digit 1, and value one obtained at the second output represents the occurrence of the remaining digits. It

**Fig. 17** The view of 216-dimensional data representing hand-written digits, obtained with the learning criterion in which the network has one output at which value 1 represents digit 1 and value 0 represents the remaining digits, with parameter ITER = 100



can be observed in the figure that signals being the response to samples representing digit 1 can be easily separated from signals representing the remaining digits. As a result, by using the analyzed criterion, we can state the possibility to separate a subarea of the space of features occupied solely by samples of digit 1.

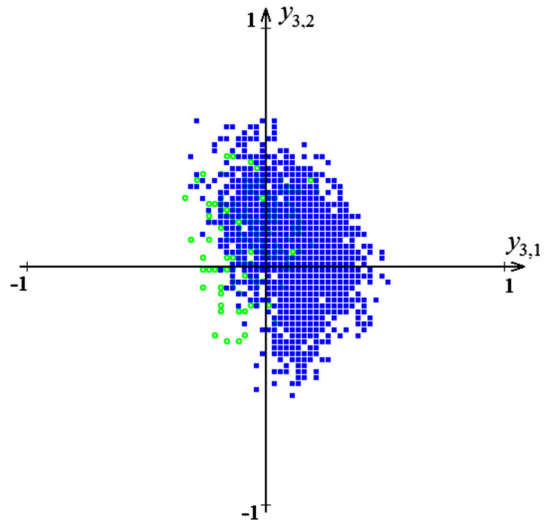
Figure 17 presents the result obtained using the learning criterion in which the network has one output at which the value equal to 1 represents the occurrence of digit 1 and value equal to 0 represents the occurrence of the remaining digits. It can be observed in this figure in an equally readable way as in the previous one that signals being the response to samples representing digit 1 can easily be separated from signals representing the remaining digits. As a result, by using the analyzed criterion, we can state the possibility to separate a subarea of the space of features occupied solely by samples of digit 1 in an equally readable way as by using the previous criterion.

Figure 18 presents the view obtained using the learning criterion in which the network has one output at which the value is to be equal to a random value attributed to each sample. It can be observed in it that signals being the response to samples representing digit 1 in majority overlap the area occupied by signals representing the remaining digits. As a result, by using the analyzed criterion, we cannot state the possibility to separate a subarea of the space of features occupied solely by samples of digit 1.

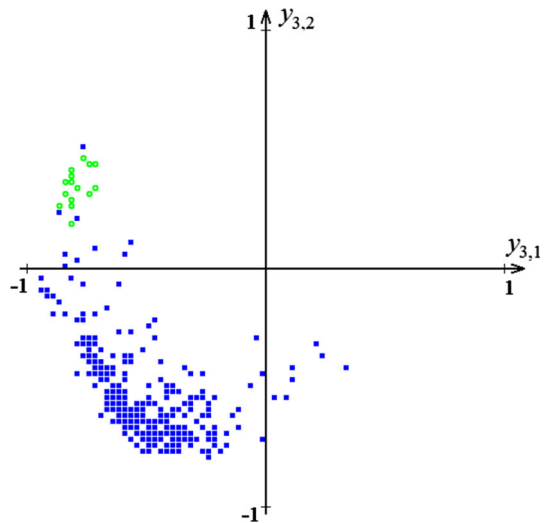
It must be noted that the methods compared in the paper can be used on any data set. The number of network inputs assumes the value equal to the number of dimensions of the analyzed set. But the length of the learning sequence is equal to the number of samples of the analyzed data set. With the change in the analyzed data set, the number of the network inputs and the length of the learning sequence can change.

The greatest advantage during the qualitative analysis of multidimensional data through their visualization is the fact that some data properties without any additional analysis can plainly be seen. However, these are not always the properties we would like to find out about. Therefore, it was decided to verify the effectiveness of criteria for searching the assumed information which is described in the paper. For this purpose, 20 sets of multidimensional data obtained from one of the publicly available repositories were randomly selected. By applying the same analysis of each of these data sets as those conducted in the paper and

**Fig. 18** The view of 216-dimensional data representing hand-written digits, obtained with the learning criterion in which the network has one output at which the value is to be equal to a random value attributed to a given sample, with parameter ITER = 10



**Fig. 19** The view of 15-dimensional Leaf data set, obtained with the learning criterion in which the network has one output at which the value is to be equal to a random value attributed to a given sample, with parameter ITER = 97



described above, it was decided to verify the possibility to separate a subarea of the space of features occupied solely by samples of a randomly selected class. Of course, if data contained information on only two classes, then there was no need to select a class. If one of dimensions in the data was an ordinal number, then such a dimension was removed, because frequently it itself allowed to recognize the class.

Obtained results are presented in Table 1. It can be analyzed that data sets varied greatly. If there is 'yes' in the column related to the effectiveness of a given learning criterion of the network, this means that with its use, the visualization allows to state the possibility to separate a subarea of the space of features occupied solely by samples of a randomly selected class. If there appears 'no' in a given place, this means that obtaining the view allowing to draw such a conclusion was not possible.

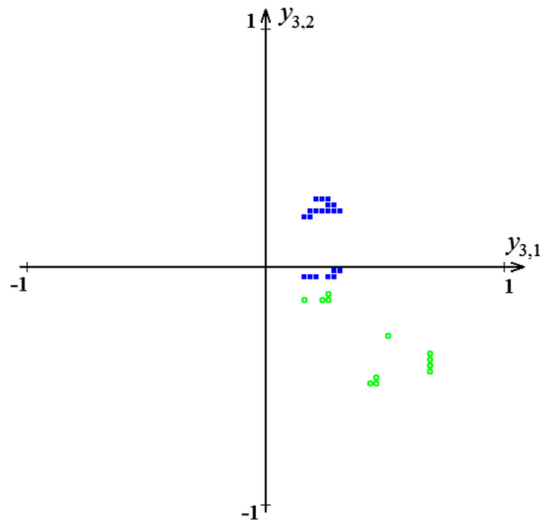
As can be observed, the use of the learning criterion of the autoassociative neural network allowed to obtain information on the possibility to divide the space of features in the case of

**Table 1** Results of the qualitative analysis of randomly selected 20 sets of multidimensional data obtained from one of publicly available repositories using different learning criteria

Data set	Analyzed class	Dimensions	Instances	Learning criteria			
				1	2	3	4
Car evaluation [45]	Car acceptability: v-good	6	1728	No	Yes	Yes	No
Cardiography [45]	Class code: pathologic	21	2126	No	No	No	No
Climate model simulation crashes [45,46]	Simulation outcome: success	18	540	No	Yes	Yes	No
Forest type mapping [45,47]	'Mixed deciduous' forest	27	325	No	Yes	Yes	No
Leaf [45,48]	Acer palmatum	15	340	Yes	Yes	Yes	Yes
Steel plates faults [45,49]	Types: pastry	27	1941	No	No	No	No
Thoracic surgery data [45,50]	Death within 1 year	16	470	No	No	No	No
Wine [45]	recognition class: 1	13	178	Yes	Yes	Yes	No
Acute inflammations [45,51]	Inflammation of urinary bladder: yes	6	120	Yes	Yes	Yes	Yes
Banknote authentication [45]	Authentication: yes	4	1372	No	Yes	Yes	No
Breast tissue [45]	Class: carcinoma	9	106	Yes	Yes	Yes	No
Connectionist bench (sonar, mines vs. rocks) [45]	Sonar: Mines	60	208	No	Yes	Yes	No
Contraceptive method choice [45]	Contraceptive method used: no	9	1473	No	No	No	No
Wilt [45,52]	Class: diseased trees	5	4339	No	No	No	No
Quality assessment of digital colposcopies [45,53]	Subjective assessment of the consensus	62	287	No	Yes	Yes	No
QSAR biodegradation [45,54]	Ready biodegradable	41	1055	No	No	No	No
Statlog (image segmentation) [45]	Class: sky	19	2310	Yes	Yes	Yes	No
Wall-following robot navigation data [45]	Class: move-forward	24	5456	No	No	No	No
Glass identification [45]	Type of glass: headlamps	9	214	Yes	Yes	Yes	Yes
Wholesale customers [45,55]	Customer sale channel: hotel/restaurant/cafe	7	440	No	No	No	No
Method effectiveness				30%	60%	60%	15%

The designation of learning criteria: 1—criterion of the autoassociative neural network, 2—criterion in which the network has two outputs representing separate classes, 3—criterion in which the network has one output whose values represent separate classes and 4—criterion in which the network has one output representing a random value attributed to a given sample

**Fig. 20** The view of six-dimensional Acute Inflammations data set, obtained with the learning criterion in which the network has one output at which the value is to be equal to a random value attributed to a given sample, with parameter ITER = 181



six data sets. This means that using this criterion allowed to obtain the searched information in the case of 30% of sets. The best results were obtained in the case of two learning criteria, in which the information on belonging of a sample to a given class should appear at one output or at two outputs as a result of learning. In the case of using both these criteria, information on the possibility to divide the space of features was obtained in the case of 12 data sets, which constitutes 60% of analyzed sets. This means that each of these two criteria allowed to obtain the searched information twice as often as in the case of the autoassociative neural network. This constitutes a perfect result and is quite surprising. The result obtained using a criterion in which the network has one output at which the value is to be equal to the random value attributed to each sample is even more surprising. In the case of using this criterion, information on the possibility to divide the space of features was obtained in the case of three data sets, which constitutes 15% of analyzed sets. This means that using such a seemingly absurd criterion allowed to obtain the searched information only twice less frequently than in the case of the autoassociative neural network. It is a very interesting result; therefore, Figs. 19, 20 and 21 present views obtained using this criterion on these sets, on which the information on the possibility to divide the space of features was obtained.

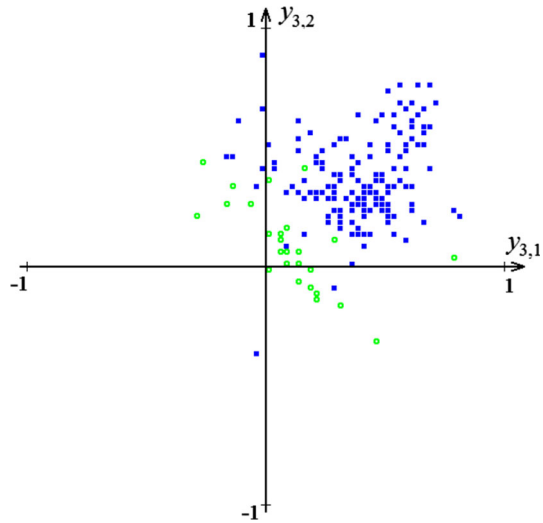
The best results were obtained with the use of criteria with which as a result of learning, information on belonging of a sample to a given class should appear at one output or several outputs. These views turned out to be more readable than with the use of the autoassociative neural network. However, in the previous paper, in which the ranking of selected methods of data qualitative analysis through its visualization was developed, autoassociative neural networks occupied the first position [42]. Combining the above facts, it can be stated that the method using the criterion which turned out to be the best in this paper, simultaneously becomes the best among all methods compared in this earlier ranking [42].

## 5 Conclusions

With the use of the learning criterion of the autoassociative neural network, signals being the response to data representing samples of coal of a given class accumulated in aggregations.



**Fig. 21** The view of nine-dimensional Glass Identification data set, obtained with the learning criterion in which the network has one output at which the value is to be equal to a random value attributed to a given sample, with parameter  $ITER = 1196$



These aggregations could easily be separated from each other. Classes 34.2 and 35 formed two aggregations of points each, while class 31 formed two aggregations of points in the figure and the third subarea of the figure occupied by two points. As a result, by using the analyzed criterion, we can indicate the possibility to divide the space of features into subareas occupied by different classes. In the case of using the same criterion on five-dimensional real data obtained as a result of the printed text recognition, the possibility to divide the space of features into subareas occupied by different classes could be stated. While in the case of using the same criterion on 216-dimensional real data presenting hand-written digits, this possibility could not be stated.

The most readable views were obtained using the criteria with which, as a result of learning, information on belonging of a given sample to a given class should appear at one output or several outputs. In the case of the use of these criteria, views in which each of classes formed one aggregation were obtained. These aggregations could be separated from each other in a way that is significantly more readable than with the use of the remaining criteria. As a result, by using these criteria, we can indicate the possibility to divide the space of features into subareas occupied by different classes in a way more readable than with the use of the remaining criteria.

Even such a seemingly nonsensical learning criterion, in which the output value is to be equal to a random value attributed to a given sample, can lead to obtaining information on data.

The more accurate network's learning, omitting even the phenomenon of the network over-learning, does not have to cause an increase in the readability of the qualitative analysis results. It may also cause deterioration in the readability of the qualitative analysis results. Therefore, with this kind of analysis, the attention should not be paid to the degree of the network's learning.

In exceptional cases, even the network without learning (before the learning process starts) allows obtaining readable results of the qualitative analysis of multidimensional data through its visualization.

From the point of view of the analysis conducted in the paper, it is not important how much of information contained in multidimensional data will be retained in the obtained

view. Moreover, it can be assumed that the more general information will be retained, the less readable can be the result of the specific targeted analysis. Therefore, to evaluate the effectiveness of qualitative analysis it is best to use the criterion allow evaluating solely the readability of this analysis.

During the analysis of randomly selected 20 sets of multidimensional data obtained from one of publicly available repositories, interesting results were obtained. Use of the learning criterion of the autoassociative neural network allowed to obtain information on the possibility to divide the space of features in the case of six data sets (30%). The same information in the case of two learning criteria, in which the information on belonging of a sample to a given class should appears at one output or at two outputs as a result of learning, was obtained in the case of 12 data sets (60%). Using a criterion in which the network has one output at which the value is to be equal to the random value attributed to each sample depicts the most surprising result. In the case of using this criterion, information on the possibility to divide the space of features was obtained in the case of three data sets (15%). This means that using such a seemingly absurd criterion allows obtaining the searched information only twice less frequently than in the case of the autoassociative neural network.

## Compliance with ethical standards

**Conflict of interest** The author declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Brandoli B, Eler D, Paulovich F, Minghim R, Batista J (2010) Visual data exploration to feature space definition. In: Proceedings of the 23rd SIBGRAPI conference on graphics, patterns and images, SIBGRAPI 2010, pp 32–39. <https://doi.org/10.1109/SIBGRAPI.2010.13>
2. Poco J, Etemadpour R, Paulovich FV, Long TV, Rosenthal P, Oliveira MCF, Linsen L, Minghim R (2011) A framework for exploring multidimensional data with 3D projections. *Comput Graph Forum* 30(3):1111–1120. <https://doi.org/10.1111/j.1467-8659.2011.01960.x>
3. Krause J, Perer A, Bertini E (2014) Infuse: interactive feature selection for predictive modeling of high dimensional data. *IEEE Trans Vis Comput Graph* 20(12):1614–1623. <https://doi.org/10.1109/TVCG.2014.2346482>
4. Jamroz D (2017) The perspective-based observational tunnels method: a new method of multidimensional data visualization. *Inf Vis* 16(4):346–360. <https://doi.org/10.1177/1473871616686634>
5. Tharwat A (2016) Principal component analysis—a tutorial. *Int J Appl Pattern Recognit* 3(3):197–240. <https://doi.org/10.1504/IJAPR.2016.079733>
6. Jolliffe I, Cadima J (2016) Principal component analysis: a review and recent developments. *Philos Trans R Soc A Math Phys Eng Sci.* <https://doi.org/10.1098/rsta.2015.0202>
7. Adachi K (2016) Multidimensional scaling. In: *Matrix-based introduction to multivariate data analysis*. Springer, Singapore, pp 243–253. [https://doi.org/10.1007/978-981-10-2341-5\\_16](https://doi.org/10.1007/978-981-10-2341-5_16)
8. Saeed N, Nam H, Imtiaz UHM, Bhatti DM (2018) A survey on multidimensional scaling. *ACM Comput Surv* 51(3):47.1–47.25. <https://doi.org/10.1145/3178155>
9. Assa J, Cohen-Or D, Milo T (1999) RMAP: a system for visualizing data in multidimensional relevance space. *Visual Comput* 15(5):217–234. <https://doi.org/10.1007/s003710050174>

10. Niedoba T (2015) Application of relevance maps in multidimensional classification of coal types. *Arch Min Sci* 60(1):93–106. <https://doi.org/10.1515/amsc-2015-0007>
11. Johansson J, Forsell C (2016) Evaluation of parallel coordinates: overview, categorization and guidelines for future research. *IEEE Trans Vis Comput Graph* 22(1):579–588. <https://doi.org/10.1109/TVCG.2015.2466992>
12. Watanabe A, Itoh T, Chiba K, Kanazaki M (2017) A scatterplots selection technique for multi-dimensional data visualization combining with parallel coordinate plots. In: 21st international conference information visualisation (IV), London, pp 78–83. <https://doi.org/10.1109/IV.2017.63>
13. Swee CT, Jeksen T (2018) Blind spots in star coordinate visualization: analysis and correction. *Pattern Recognit Lett* 106:7–12. <https://doi.org/10.1016/j.patrec.2018.02.005>
14. Swee CT, Jeksen T (2017) Lost in translation: the fundamental flaws in star coordinate visualizations. *Procedia Comput Sci* 108:2308–2312. <https://doi.org/10.1016/j.procs.2017.05.087>
15. Maaten LJPVD, Hinton GE (2008) Visualizing high-dimensional data using t-SNE. *J Mach Learn Res* 9:2579–2605
16. Rice I (2018)  $\Gamma$ -stochastic neighbour embedding for feed-forward data visualization. *Inf Vis* 17(4):306–315. <https://doi.org/10.1177/1473871617715212>
17. Prechelt L (1998) Early stopping-but when? In: Orr GB, Müller KR (eds) *Neural networks: tricks of the trade*. Springer, Berlin, pp 55–69
18. Orr GB, Müller KR (2003) *Neural networks: tricks of the trade*. Springer, Berlin. <https://doi.org/10.1007/978-3-642-35289-8>
19. Bengio Y (2012) Practical recommendations for gradient-based training of deep architectures. In: Montavon G, Orr GB, Müller KR (eds) *Neural networks: tricks of the trade*. Springer, Berlin, pp 437–478
20. Snoek J, Larochelle H, Adams RP (2012) Practical Bayesian optimization of machine learning algorithms. In: *Advances in neural information processing systems*, pp 2951–2959
21. Mahendran A, Vedaldi A (2015) Understanding deep image representations by inverting them. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), pp 5188–5196
22. Pezzotti N, Holtt T, Van Gemert J, Lelieveldt BPF, Eisemann E, Vilanova A (2018) DeepEyes: progressive visual analytics for designing deep neural networks. *IEEE Trans Vis Comput Graph* 24(1):98–108. <https://doi.org/10.1109/TVCG.2017.2744358>
23. Yosinski J, Clune J, Nguyen A, Fuchs T, Lipson H (2015) Understanding neural networks through deep visualization. In: *Deep learning workshop, international conference on machine learning (ICML)*
24. Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. In: *Proc. European conference on computer vision*. Springer, Berlin, pp 818–833
25. Rauber PE, Fadel SG, Falcao AX, Telea AC (2017) Visualizing the hidden activity of artificial neural networks. *IEEE Trans Vis Comput Graph* 23(1):101–110. <https://doi.org/10.1109/TVCG.2016.2598838>
26. Aubry M, Russell B (2015) Understanding deep features with computer generated imagery. In: *International conference on computer vision (ICCV)*, pp 2875–2883
27. Kohonen T (1989) *Self-organization and associative memory*. Springer, Berlin. <https://doi.org/10.1007/978-3-642-88163-3>
28. Dzemyda G, Kurasova O, Zilinskas J (2015) *Multidimensional data visualization. Methods and applications*. Springer, New York. <https://doi.org/10.1007/978-1-4419-0236-8>
29. Dzemyda G, Medvedev V, Lupeikiene A, Kurasova O, Caplinskas A (2016) Big multidimensional datasets visualization using neural networks—efficient decision support. *Complex Syst Inf Model Q*. <https://doi.org/10.7250/csinq.2016-6.01>
30. Aldrich C (1998) Visualization of transformed multivariate data sets with autoassociative neural networks. *Pattern Recognit Lett* 19(8):749–764. [https://doi.org/10.1016/S0167-8655\(98\)00054-3](https://doi.org/10.1016/S0167-8655(98)00054-3)
31. Kaklauskas A, Dzemyda G, Tupenaite L, Voitau I, Kurasova O, Naimaviciene J, Rassokha Y, Kanapeckiene L (2018) Artificial neural network-based decision support system for development of an energy-efficient built environment. *Energies*. <https://doi.org/10.3390/en11081994>
32. Hu R et al (2018) ELM-SOM: a continuous self-organizing map for visualization. In: 2018 International joint conference on neural networks (IJCNN), Rio de Janeiro, pp 1–8. <https://doi.org/10.1109/IJCNN.2018.8489268>
33. Mohebi E, Bagirov A (2016) Constrained self organizing maps for data clusters visualization. *Neural Process Lett* 43:849. <https://doi.org/10.1007/s11063-015-9454-1>
34. Brito da Silva LE, Wunsch DC (2018) An information-theoretic-cluster visualization for self-organizing maps. *IEEE Trans Neural Netw Learn Syst* 29(6):2595–2613. <https://doi.org/10.1109/TNNLS.2017.2699674>
35. Cordel MII, Azcarraga A (2018) A new method for emulating self-organizing maps for visualization of datasets. *Int J Comput Intell Appl* 17(3):1850014. <https://doi.org/10.1142/S1469026818500141>

36. Gonen M (2014) Coupled dimensionality reduction and classification for supervised and semi-supervised multilabel learning. *Pattern Recognit Lett* 38:132–141. <https://doi.org/10.1016/j.patrec.2013.11.021>
37. Hartono P (2018) Classification and dimensional reduction using restricted radial basis function networks. *Neural Comput Appl* 30(3):905–915. <https://doi.org/10.1007/s00521-016-2726-5>
38. Wei W, Song H, Li W, Shen P, Vasilakos A (2017) Gradient-driven parking navigation using a continuous information potential field based on wireless sensor network. *Inf Sci* 408:100–114. <https://doi.org/10.1016/j.ins.2017.04.042>
39. Lu X, Zhao Y, Liu M (2018) Self-learning interval type-2 fuzzy neural network controllers for trajectory control of a Delta parallel robot. *Neurocomputing* 283:107–119. <https://doi.org/10.1016/j.neucom.2017.12.043>
40. Bryson AE, Ho YC (1969) *Applied optimal control*. Blaisdell, New York
41. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323:533–536
42. Jamroz D, Niedoba T (2015) Comparison of selected methods of multi-parameter data visualization used for classification of coals. *Physicochem Probl Mineral Process* 51(2):769–784. <https://doi.org/10.5277/ppmp150233>
43. Niedoba T (2013) Multidimensional characteristics of random variables in description of grained materials and their separation processes [in Polish]. IGSMiE PAN, Krakow
44. Jamroz D (2014) Application of multidimensional data visualization in creation of pattern recognition systems. In: *Man-machine interactions 3, ICMMI 2013, advances in intelligent systems and computing*, vol 242. Springer, pp 443–450. [https://doi.org/10.1007/978-3-319-02309-0\\_48](https://doi.org/10.1007/978-3-319-02309-0_48)
45. Dua D, Karra Taniskidou E (2017) UCI machine learning repository [<http://archive.ics.uci.edu/ml>]. University of California, School of Information and Computer Science, Irvine. <http://archive.ics.uci.edu/ml>. Accessed 6 Feb 2019
46. Lucas DD, Klein R, Tannahill J, Ivanova D, Brandon S, Domyancic D, Zhang Y (2013) Failure analysis of parameter-induced simulation crashes in climate models. *Geosci Model Dev Discuss* 6:585–623. <https://doi.org/10.5194/gmd-6-1157-2013>
47. Johnson B, Tateishi R, Xie Z (2012) Using geographically-weighted variables for image classification. *Remote Sens Lett* 3(6):491–499
48. Silva PFB, Marcal ARS, Almeida da Silva RM (2013) Evaluation of features for leaf discrimination. In: *Springer lecture notes in computer science*, vol 7950, pp 197–204
49. Dataset provided by Semeion, Research Center of Sciences of Communication, Via Sersale 117, 00128, Rome, Italy. [www.semeion.it](http://www.semeion.it). Accessed 6 Feb 2019
50. Zieba M, Tomczak JM, Lubicz M, Swiatek J (2013) Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients. *Appl Soft Comput* 14:99–108. <https://doi.org/10.1016/j.asoc.2013.07.016>
51. Czerniak J, Zarzycki H (2003) Application of rough sets in the presumptive diagnosis of urinary system diseases. In: *Artificial intelligence and security in computing systems, ACS'2002 9th international conference proceedings*. Kluwer Academic Publishers, pp 41–51
52. Johnson B, Tateishi R, Hoan N (2013) A hybrid pansharpening approach and multiscale object-based image analysis for mapping diseased pine and oak trees. *Int J Remote Sens* 34(20):6969–6982
53. Fernandes K, Cardoso JS, Fernandes J (2017) Transfer learning with partial observability applied to cervical cancer screening. In: *Iberian conference on pattern recognition and image analysis*. Springer
54. Mansouri K, Ringsted T, Ballabio D, Todeschini R, Consonni V (2013) Quantitative structure—activity relationship models for ready biodegradability of chemicals. *J Chem Inf Model* 53:867–878
55. Abreu N (2011) *Análise do perfil do cliente Recheio e desenvolvimento de um sistema promocional*. Mestrado em Marketing, ISCTE-IUL, Lisbon



**Dariusz Jamróz** is adjunct professor at the Department of Applied Computer Science, AGH University of Science and Technology. He holds a Ph.D. in Computer Science, and his research interests include artificial intelligence, data analysis, data mining, multidimensional visualization, information visualization and scientific visualization.