

# Comparison of Artificial Intelligence Methods on the Example of Tea Classification Based on Signals from E-nose Sensors

Paweł Pławiak<sup>1,2,\*</sup>, Wojciech Maziarz<sup>3</sup>

<sup>1</sup>Department of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering  
AGH University of Science and Technology  
30 Mickiewicza Av., 30-059 Krakow, Poland

<sup>2</sup>Faculty of Physics, Mathematics and Computer Science, Cracow University of Technology,  
24 Warszawska st., F-5, 31-155 Krakow, Poland

<sup>3</sup>Department of Computer Science, Electronics and Telecommunications  
AGH University of Science and Technology  
30 Mickiewicza Av., 30-059 Krakow, Poland  
\*Corresponding Author: plawiak@agh.edu.pl

Copyright ©2013 Horizon Research Publishing All rights reserved.

**Abstract** The data collected from electronic nose systems are multidimensional and usually contain a lot of redundant information. In order to extract only the relevant data, different computational techniques are developed. The article presents and compares selected pattern recognition algorithms in application to qualitative determination of different brands of tea. The measured responses of an array of 18 semiconductor gas sensors formed input vectors used for further analysis. The initial data processing consisted on standardization, principal component analysis, data normalization and reduction. Soft computing one can divide into single method systems using neural networks, fuzzy systems, and hybrid systems like evolutionary-neural, neuro-fuzzy, evolutionary-fuzzy. All the presented systems were evaluated based on accuracy (generated error) and complexity (number of parameters and training time) criteria. A novel method of forming input data vector by aggregation of the first three principal components is also presented.

**Keywords** Artificial intelligence methods, Soft computing, Computational intelligence, Neural networks, Genetic algorithms, Fuzzy systems, Pattern Recognition, Signal processing, PCA, Tea, E-nose, Chemometrics

---

## 1 Introduction

A semiconductor gas sensors based on metal oxides are very popular due to their advantages: high sensitivity, low cost, small dimensions, low power consumption and easy integration with measurement circuit. Their

main drawbacks are small selectivity, drift, lack of long term stability and sensitivity to other parameters, i.e. temperature changes, water vapour presence in gas atmosphere, etc. [8]. The reactions between the sensitive layer and gas atmosphere are thermally activated and sensitivity to specific gas depends on the temperature. In the simplest case the only signal obtained from the resistive type sensor is usually its resistance measured at constant temperature. Due to this fact and because of sensors cross-sensitivity and small selectivity a strict determination of gas type and its concentration in unknown mixture is usually impossible.

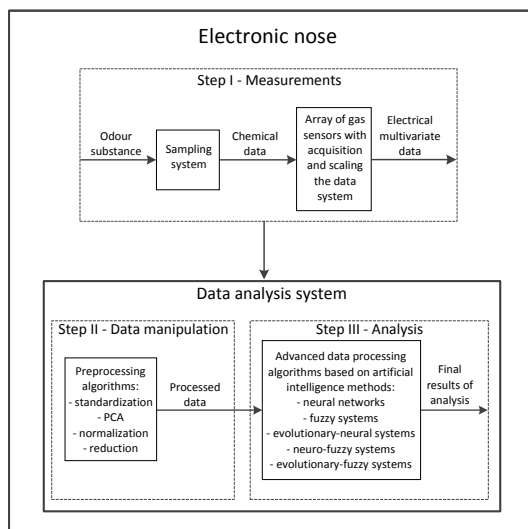
In order to solve this problem one can use different approaches. One of them is based on the gathering and processing of signals obtained from arrays of partially selective sensors [4, 11]. The sensitivity and selectivity of a sensor array can be highly enhanced by developing various pattern recognition algorithms [3, 14].

A promising and quite novel approach is the analysis of the dynamic response of a single sensor working at modulated temperature. In such a case, one sensor is equivalent to an array of sensors working at different temperatures. For sensors powered with a pulse voltage the average power consumption decreases and also the long-term stability of the sensor is improved [9, 13, 20]. The intentional temperature change according to a programmed profile can provide additional information consisted in time dependent non-linear sensor response. Such response is related to the adsorption and desorption processes at the semiconductor surface and is influenced by the concentration of the gas species and their chemical structure. These properties are used in gas sensor arrays working in electronic noses, where a multidimensional and non-linear responses are measured and analysed.

The collected information is usually too complex and impossible to analyse without any data processing system. This is a problem similar to the image analysis [25]. The procedure of feature extraction is typically performed by the standard pattern recognition methods used in chemometrics (i.e. PCA - Principal Component Analysis, CLA - Cluster Analysis, TM - Template Matching, DFA - Discriminant Function Analysis, TMLR - Transformed Multiple Linear Regression) [12] or signal processing (i.e. Fourier or wavelet analyses, ANN - Artificial Neural Networks, GA - Genetic Algorithms, FL - Fuzzy Logic, etc.) [6, 27].

As electronic nose (e-nose) is the technical system consisting of both the hardware and software [7]. A hardware part is an array of sensors responsible for detecting odours present in the measured gas atmosphere. The software is created by advanced algorithms that process information about volatile organic compounds (VOCs) and make the final analysis. Just as a biological analog, the e-nose is used for identification and classification of gas mixtures and rarely their concentrations [16]. The e-nose systems based on sensor arrays are frequently used in qualitative analyses of different species, i.e. of milk [2, 29], alcohols [1, 30] and tea [5, 10, 21, 28].

Figure 1 shows three stages of signal processing from measurements of odours to the final analysis using AI methods.



**Figure 1.** An example of electronic nose structure.

An essential part of the commercial e-nose devices is data processing unit data analysis system that performs data preprocessing, reduction and analysis. Algorithms employed should be relatively simple and not computationally demanding. In order to extract the relevant data only, minimize the computation time and prepare algorithms the best suited to be embedded in a hardware, a lot of different approaches are developed.

In this paper, the authors focused on a comparison of selected intelligent systems [26] based on computational intelligence, used to analyse the signals from the sensors array in order to classify brands of tea. The authors have developed several data analysis systems. To collect the data a commercial e-nose was used. The software was written in MATLAB<sup>1</sup>, as it contains very extensive

<sup>1</sup>MATLAB version R2010a

libraries of functions using AI methods.

## 2 Measurements

The measurements were performed on commercially available e-nose from Alpha MOS company - model FOX 4000 [15]. It consists of an array of 18 semiconductor gas sensors.

E-nose array consists of three types of metal oxide semiconductor sensors [17–19]:

- type T – SnO<sub>2</sub> active layer;
- type P – also based on SnO<sub>2</sub>; higher sensitivity and faster response time than T-type,
- type LY – based on Cr<sub>2-x</sub>Ti<sub>x</sub>O<sub>3+y</sub> (p-type) and WO<sub>3</sub> (n-type); LY are low power sensors.

Tea specimen means the vial broadcast with tea directly from the package. All tea specimens were heated before measurement for 5 min at 60°C. The volume injected into the measuring chamber of the gas mixture was 0.5 ml. Measurement of each specimen of tea lasted 121 seconds. Sampling was performed every second, and single sample contains information about a resistance value of 18 sensors. A set of 121 samples was performed. In effect, for a single specimen of tea we obtain a matrix, consisting of 121 rows corresponding to subsequent samples, and 18 columns corresponding to the responses of the sensors. Table 1 shows the matrices in which the data was collected.

The object of this study were data obtained for different brands of tea. For every kind of tea a number of data sets (from 5 to 25, see Table 3) was recorded and as a result we obtained 85 input data sets. After data selection process, described in next paragraphs, 10 brands of tea were taken into consideration: 1 - melissa; 2 - melon-apple; 3,4,5,6 - mint of 4 kinds; 7 - chamomile; 8 - vanilla; 9,10 - ginger of 2 kinds. As a result of the selection, some data was rejected and two brands of tea (mint and ginger) were divided into subtypes because they formed groups with tea specimens that have similar responses.

## 3 Data manipulation

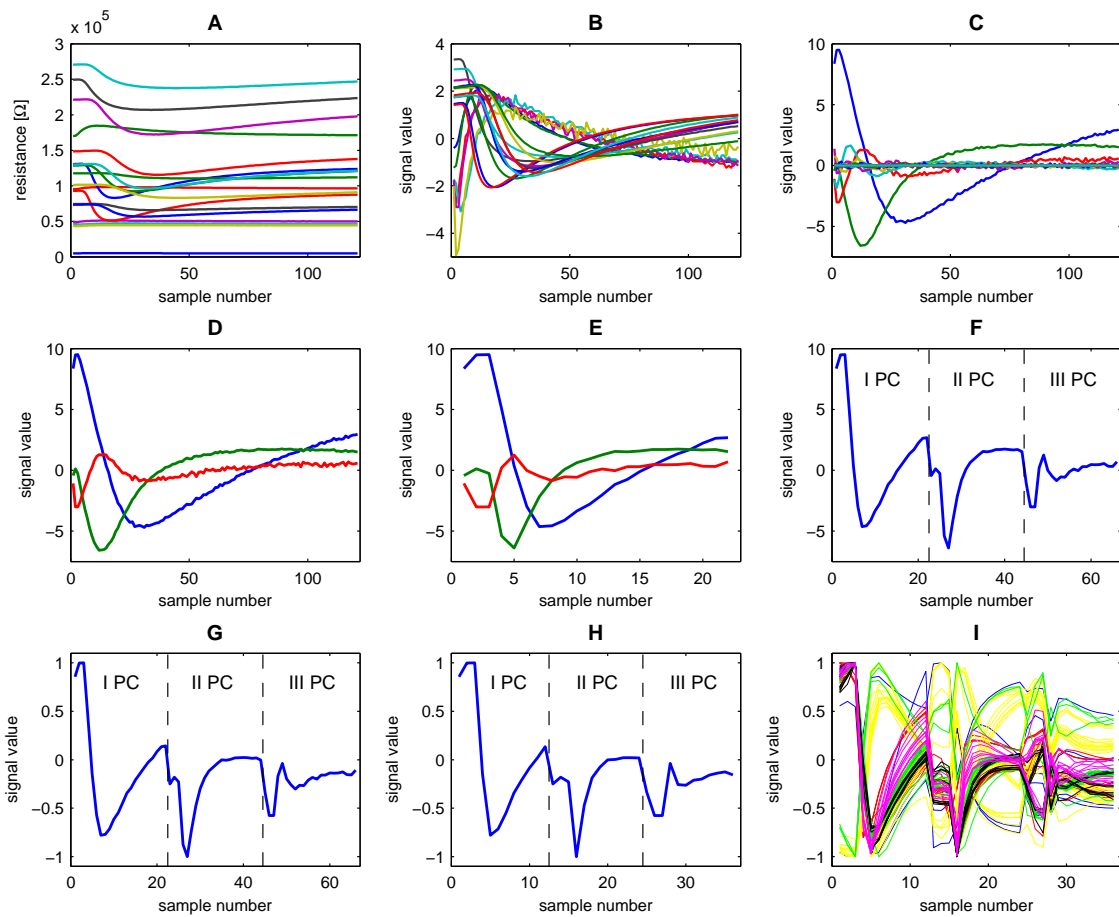
The first stage of data manipulation was preprocessing of measurement data: standardization, principal component analysis (PCA), normalization and data reduction.

### 3.1 Preprocessing

All of the preprocessing stages performed on a single specimen of tea are described in Table 2 and shown in Figure 2. As a result 85 vectors build of 36-elements were obtained. A part of them – after data selection - were used as input data for all the developed and evaluated AI systems.

### 3.2 Data selection

Due to some errors and inaccuracies observed in sensor responses, caused by i.e. improper dosing of the tea specimen, the data selection is necessary. The incorrect



**Figure 2.** The visualization of preprocessing stages for single specimen of tea. **A** – measured sensors response vectors, each color represents separate response; **B** – signals after the standardization; **C** – after PCA, each PC is presented with a different color; **D** – first three PCs; **E** – first three PCs - every six sample left; **F** – after aggregation of PC1 to PC3; **G** – signal after normalization; **H** – signal after reduction of samples - every second sample left; **I** – data for all specimens of tea after preprocessing, the specimens of tea corresponding to the same species of tea have the same color in graph (6 colors, 6 species of tea).

**Table 1.** Data matrices.

Raw data						Preprocessed data				
	Sensor 1	Sensor 2	Sensor 3	...	Sensor 18	Specimen 1	Specimen 2	Specimen 3	...	Specimen 48/18
Sample 1	$r_{1 1}$	$r_{1 2}$	$r_{1 3}$	...	$r_{1 18}$	$p_{1 1}$	$p_{1 2}$	$p_{1 3}$	...	$p_{1 48/18}$
Sample 2	$r_{2 1}$	$r_{2 2}$	$r_{2 3}$	...	$r_{2 18}$	$p_{2 1}$	$p_{2 2}$	$p_{2 3}$	...	$p_{2 48/18}$
Sample 3	$r_{3 1}$	$r_{3 2}$	$r_{3 3}$	...	$r_{3 18}$	$p_{3 1}$	$p_{3 2}$	$p_{3 3}$	...	$p_{3 48/18}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Sample 121	$r_{121 1}$	$r_{121 2}$	$r_{121 3}$	...	$r_{121 18}$	$p_{36 1}$	$p_{36 2}$	$p_{36 3}$	...	$p_{36 48/18}$
	$r \in [4.8 \cdot 10^3 ; 2.7 \cdot 10^5]$					$p \in [-1 ; 1]$				
	85 matrices, each matrix corresponds to one specimen of tea					48 columns for training set, 18 columns for test set each column of the matrix corresponds to one specimen of tea				
Reference matrix - systems with multiple outputs						Reference matrix - system with single output				
	Specimen 1	Specimen 2	Specimen 3	...	Specimen 48/18	Specimen 1	Specimen 2	Specimen 3	...	Specimen 48/18
Tea 1	$m_{1 1}$	$m_{1 2}$	$m_{1 3}$	...	$m_{1 48/18}$	$m_{1 1}$	$m_{1 2}$	$m_{1 3}$	...	$m_{1 48/18}$
Tea 2	$m_{2 1}$	$m_{2 2}$	$m_{2 3}$	...	$m_{2 48/18}$				...	
Tea 3	$m_{3 1}$	$m_{3 2}$	$m_{3 3}$	...	$m_{3 48/18}$				...	
⋮	⋮	⋮	⋮	⋮	⋮				⋮	
Tea 10	$r_{10 1}$	$r_{10 2}$	$r_{10 3}$	...	$r_{10 48/18}$				...	
	$m \in \{-1 ; 1\}$ or $m \in \{0 ; 1\}$					$m \in \{1 ; 2 ; 3 ; 4 ; 5 ; 6 ; 7 ; 8 ; 9 ; 10\}$				
	1 for the correct species of tea, and -1 or 0 for other 48 columns for training set, 18 columns for test set					appropriate integer indicating the correct species of tea 48 columns for training set, 18 columns for test set				

**Table 2.** Successive preprocessing stages of measured data, for single specimen of tea. PC is principal component.

Stage:	Name:	Description:	Number of dimensions:	Range:	Number of data:	Figure:
<b>I</b>	Raw data	121 samples 18 sensors responses	18	$[4.8 \cdot 10^3 ; 2.7 \cdot 10^5]$	$121 \cdot 18 = 2178$	<b>2A</b>
<b>II</b>	Standardization	standard deviation = 1 average value = 0	18	$[-4.9 ; 3.4]$	$121 \cdot 18 = 2178$	<b>2B</b>
<b>III</b>	Principal Component Analysis	maximizing the variance of the first coordinates	18	$[-6.6 ; 9.5]$	$121 \cdot 18 = 2178$	<b>2C</b>
<b>IV</b>	Reduction	dimensional reduction leaving first 3 PC	3	$[-6.6 ; 9.5]$	$121 \cdot 3 = 363$	<b>2D</b>
<b>V</b>	Reduction	downsampling: every six sample left	3	$[-6.4 ; 9.5]$	$22 \cdot 3 = 66$	<b>2E</b>
<b>VI</b>	Aggregation	aggregation of first 3 PC	1	$[-6.4 ; 9.5]$	$66 \cdot 1 = 66$	<b>2F</b>
<b>VII</b>	Normalization	to interval: $[-1 ; 1]$	1	$[-1 ; 1]$	$66 \cdot 1 = 66$	<b>2G</b>
<b>VIII</b>	Reduction	downsampling: one of two samples left	1	$[-1 ; 1]$	$36 \cdot 1 = 36$	<b>2H</b>
<b>IX</b>	Preprocessed data	all 85 specimens of tea preprocessed	1	$[-1 ; 1]$	$36 \cdot 1 = 36$	<b>2I</b>

measurements were rejected and two kinds of tea were divided into smaller subgroups. Results after selection are shown in Figure 3, where visual grouping of different kinds of tea is observed.

Figure 3 presents the separation of preprocessed data, but as it can be seen, it is rather difficult to interpret. Therefore, other methods showing the grouping of data were used. One of them is PCA that converts a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (PC). On the graph of PC1 vs. PC2 one can observe grouping of similar variables. Figure 4 which is equivalent to Figure 3 shows the grouping for all kinds of tea before and after selection.

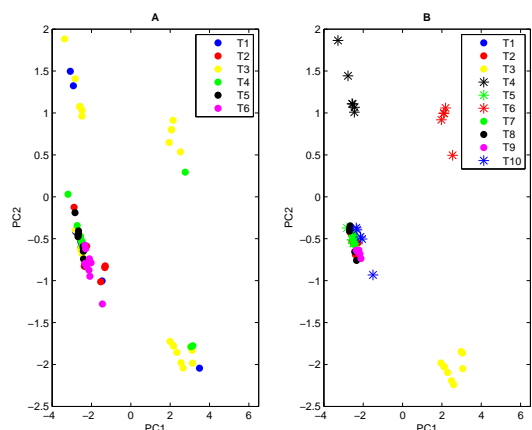
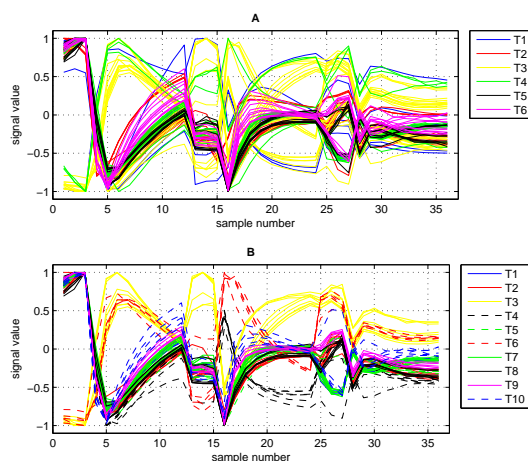
Table 3 provides a detailed description of the data, covering raw data, preprocessed data, reference matrices (matrices with model answers) and data selection. For the reference matrices the expected response should be understood as the correct brand of tea.

### 3.3 Training and test data sets

Intelligent systems learn on examples how to analyse input signals. Due to this fact one has to divide all of the elements into training data set used for learning and test data set used for evaluation of the system. In order to make properly this assignment, we draw the data and judge visually whether the elements are grouped or

**Table 3.** Description of data.

<b>Raw data</b>	
Rows	121 rows: correspond to the subsequent 121 samples
Columns	18 columns: correspond to the response of the 18 sensors
<b>Preprocessed data - input information</b>	
Rows	36 rows corresponding to the selected 36 samples
Columns	Columns corresponding to the subsequent specimens of tea: - for training set: 48 columns (elements of the set) - for test set: 18 columns (elements of the set)
<b>Reference matrices - output information</b>	
Systems with single output	- 1 row corresponding to 1 output, expected response represented by the appropriate integer from 1 to 10 - 48 columns for training set or 18 columns for test set
Systems with multiple outputs from the set {0,1}	- 10 row corresponding to 10 outputs, expected response: 1 for the correct species of tea, and 0 for other - 48 columns for training set or 18 columns for test set
Systems with multiple outputs from the set {-1,1}	- 10 row corresponding to 10 outputs, expected response: 1 for the correct species of tea, and -1 for other - 48 columns for training set or 18 columns for test set
<b>Data selection</b>	
Before selection	85 specimens of tea, 6 brands of tea: melissa (T1 - 10), melon-apple (T2 - 15), mint (T3 - 25), chamomile (T4 - 15), vanilla (T5 - 10) and ginger (T6 - 10)
After selection	66 specimens of tea, 10 brands of tea: melissa (T1 - 5), melon-apple (T2 - 10), mint - 4 kinds (T3 - 9, T4 - 6, T5 - 5, T6 - 5), chamomile (T7 - 10), vanilla (T8 - 6) and ginger - 2 kinds (T9 - 5, T10 - 5) - training set: 48 specimens of tea (T1 - 4, T2 - 7, T3 - 6, T4 - 4, T5 - 4, T6 - 4, T7 - 7, T8 - 4, T9 - 4, T10 - 4) - test set: 18 specimens of tea (T1 - 1, T2 - 3, T3 - 3, T4 - 2, T5 - 1, T6 - 1, T7 - 3, T8 - 2, T9 - 1, T10 - 1)



**Figure 3.** The visual grouping of data before and after selection. The solid lines show the basic teas, while the dashed line subtypes of teas obtained as a result of selection. Different brands of tea are marked with symbols T1 to T10. **Legend:** **A** – before selection - 6 species; **B** – after selection - 10 species;

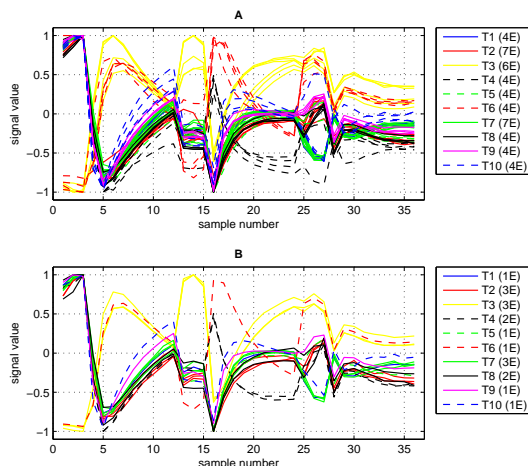
**Figure 4.** The result of PCA: the visual grouping of data before and after selection. Every kind of tea is presented with a different color. **Legend:** **A** – before selection - 6 species; **B** – after selection - 10 species.

separated into classes (tea brands).

Figure 5 show signals obtained for specific brands of tea in comparison to one chosen kind of tea - mint (T3). As we can see there is an example of good separation of T3 tea from others. Data of this kind are relatively easy to classify for data analysis systems.

Similarly, figure 6 show signals obtained for specific brands of tea in comparison to mint tea (T5). As we can see there is an example of almost no separation of T5 tea from others. The signal for T5 tea is very similar to signals form T2, T8 and T9. Data of this kind are very difficult for classification systems.

After selection, the data have been carefully divided for the training set and test set. Figure 7 shows the data for the a.m. sets. Note that the elements (data vectors related to specimens of tea) of the same kind of tea are grouped close together and are marked with the same color. Some groups of elements are clearly separated, while others overlap. The elements used to test the systems are within a set of elements used for learning systems. Shapes of signals in training and test data sets (for a given species of tea) are similar.



**Figure 7.** Separation of data (related to specimens of tea) into training and test sets. Numbers in brackets (i.e. 4E) mean the number of elements (i.e. 4) of a given set assigned to specific brand of tea. **Legend:** **A** – training set; **B** – test set.

## 4 Data analysis systems

The data contained in the preprocessed response vectors are fed into the next stage of recognition of odour – data analysis. It involves the assignment of the input data to appropriate classes. Such classification is performed using data analysis systems based on the methods of AI.

The advantage of techniques based on artificial intelligence (AI) methods lies in the properties inherited from their biological equivalent, such as learning and generalization of knowledge (ANN [23]), a global optimization (evolutionary algorithms) and the use of imprecise concepts (FL) [26]. Very high popularity of ANN in recent years has contributed to a number of different types of structures, data flow, type of transfer function of neurons and the method of training. When applying ANN to obtain maximum efficiency, the designer must

demonstrate relevant experience realized in the selection of relevant parameters.

The authors designed the data analysis systems that use both the single methods of AI and/or their combinations [22, 24]. Below are briefly characterized. A several types of ANN were evaluated and the results of their calculations were compared.

- **Single AI methods**

- **Artificial neural networks (ANN)**

- \* Feed-forward neural network (FF)
- \* Probabilistic neural network (PNN)
- \* Learning vector quantization network (LVQ)
- \* Recurrent neural network (RNN)
- \* Radial basis function neural network (RBF)

- **Sugeno fuzzy system (FUZZY)**

- **Hybrid systems**

- **Evolutionary-neural systems (ANN + GA)** - artificial neural networks, where the weights and biases are determined as a result of training by genetic algorithm

- \* Feed-forward neural network trained by genetic algorithm (FF + GA)
- \* Recurrent neural network trained by genetic algorithm (RNN + GA)

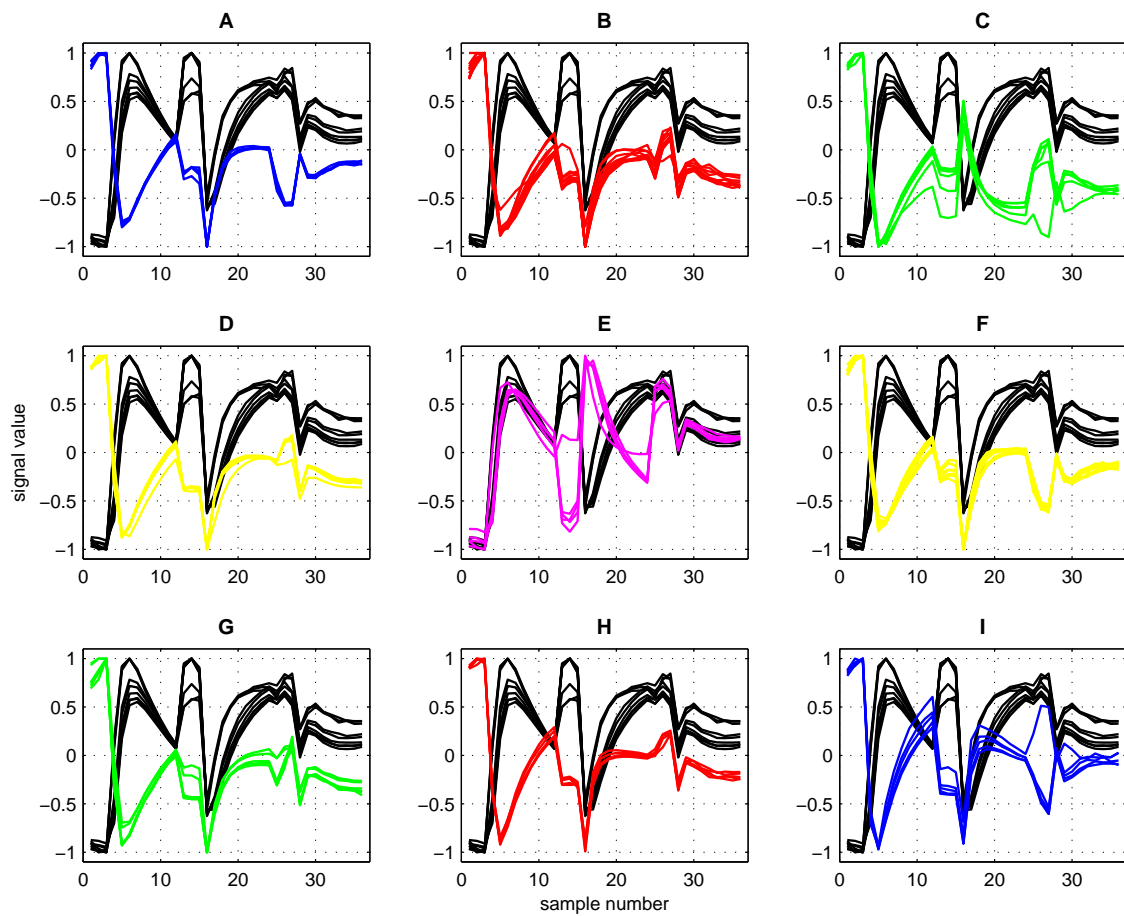
- **Neuro-fuzzy system (ANFIS)** - Sugeno-type fuzzy system transformed into an equivalent neural network, whose parameters are optimized using the method of backpropagation

- **Evolutionary-fuzzy system (FUZZY + GA)** - Mamdani-type fuzzy system with parameters optimized using a genetic algorithm

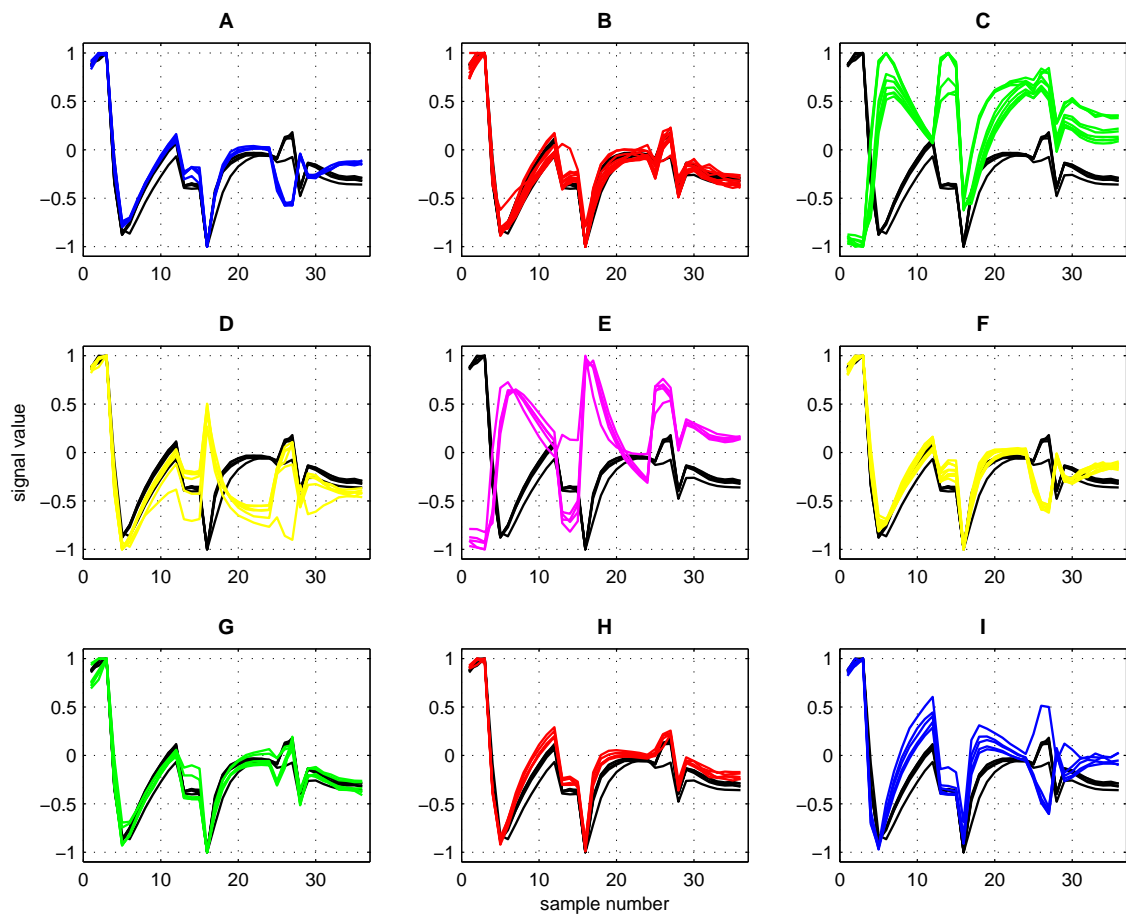
The most important parameters of these systems are presented in Table 4. For the ANN are: topology, the transfer function of neurons in successive layers and training algorithm. For the fuzzy system are: the type with the number of inputs - outputs in brackets and the number of inference rules. For the GA are: the number of individuals, the probability of crossover and mutation, the number of generation and the method of scaling and selection. For overall parameters are: the number of elements of the training and test sets, the number of variables (parameters determined during the training - such as weights and biases or parameters of membership functions in the premises and conclusions), and the number of correctly classified brands of tea.

### 4.1 Probabilistic neural network

Among all the tested systems the probabilistic neural network (PNN), received the best score in the evaluation. Its structure is shown in Figure 8. The results of calculations of the PNN system are presented in Figures 9 and 10. Table 5 presents its parameters.



**Figure 5.** The comparison of signals obtained for T3 mint tea on background of others species of tea. Mint tea T3 is in black, other brands of tea have different colors. Figure 5 was made from Figure 2H. **Legend:** T3 vs. **A** – T1; **B** – T2; **C** – T4; **D** – T5; **E** – T6; **F** – T7; **G** – T8; **H** – T9; **I** – T10.



**Figure 6.** The comparison of signals obtained for T5 mint tea on background of others species of tea. Mint tea T5 is in black, other brands of tea have different colors. Figure 5 was made from Figure 2H. **Legend:** T5 vs. **A** – T1; **B** – T2; **C** – T3; **D** – T4; **E** – T6; **F** – T7; **G** – T8; **H** – T9; **I** – T10.

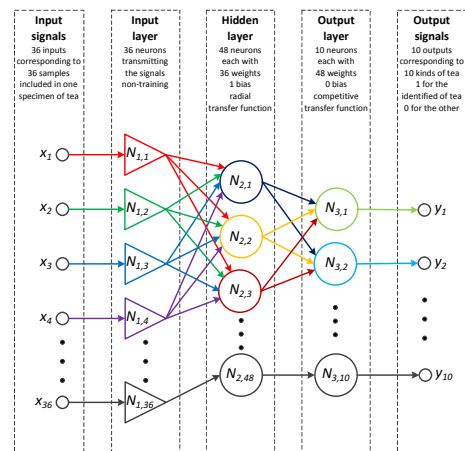


**Table 4.** The parameters of the probabilistic neural network (PNN).

System	Parameters					
	Neural network		Fuzzy system	Genetic algorithm		Overall
	topology/ transfer functions	training algorithm	type (I-O)/ inference rules	individuals/ crossover/ mutation	generations/ scaling/ selection	training/ test set variables/ efficiency
<b>Artificial neural networks (ANN)</b>						
<b>FF</b>	36 – 16 – 10 tangent sigmoid tangent sigmoid	LM	–	–	–	48 18 <b>762</b> <b>10:10</b>
<b>PNN</b>	36 – 48 – 10 radial competitive	mapping	–	–	–	48 18 <b>2256</b> <b>10:10</b>
<b>LVQ</b>	36 – 7 – 6 competitive linear	modification	–	–	–	32 13 <b>294</b> <b>6:10</b>
<b>RNN</b>	36 – 16 – 10 tangent sigmoid tangent sigmoid	LM	–	–	–	48 18 <b>1018</b> <b>10:10</b>
<b>RBF</b>	36 – 48 – 10 radial – linear	LM	–	–	–	48 18 <b>2266</b> <b>10:10</b>
<b>Fuzzy system (FUZZY)</b>						
<b>FUZZY</b>	–	–	Sugeno (36 - 10) 14	–	–	48 18 <b>6188</b> <b>10:10</b>
<b>Evolutionary-neural systems (ANN + GA)</b>						
<b>FF GA</b>	36 – 10 – 6 tangent sigmoid tangent sigmoid	GA	–	50 heuristic (0.7) uniform (0.01)	300 ranking stochastic	29 11 <b>436</b> <b>6:10</b>
<b>RNN GA</b>	36 – 10 – 6 tangent sigmoid tangent sigmoid	GA	–	50 heuristic (0.7) uniform (0.01)	300 ranking stochastic	29 11 <b>536</b> <b>6:10</b>
<b>Neuro-fuzzy system (ANFIS)</b>						
<b>ANFIS</b>	36 – 3554 – 1 radial – linear	BP	Sugeno (36 - 1) 48	–	–	48 18 <b>5232</b> <b>10:10</b>
<b>Evolutionary-fuzzy system (FUZZY + GA)</b>						
<b>FUZZY GA</b>	–	–	Mamdani (36 - 8) 11	50 heuristic (0.7) uniform (0.01)	300 ranking stochastic	37 14 <b>968</b> <b>8:10</b>

PNN consists of input, hidden and output layers. Each neuron of hidden layer (with radial transfer function) corresponds to one element from the training set. In the output layer each recognized class corresponds to one neuron (with competition transfer function). The only parameter influencing the learning process of PNN is smoothing coefficient. It represents the radial deviation of appropriate Gaussian functions. Its modification affects the range of influence (scope of impact) of “knowledge” contained in the training set on space of input signals. The PNN system learns quickly, its disadvantage is the size. This is due to the fact that this type of network consists of neurons corresponding to all the elements of the training set. This results in mapping of the entire training set in PNN structure.

The example of matching the PNN response to the model response obtained for a test set are presented in Figure 9. The accuracy of the system is related to the slope of the fitting line and to the shift of the response points. The ideal line has a slope of 45 degrees and the points of test and model response match. The example of classification results obtained for PNN for the test set are shown in Figure 10. The proper classifications are marked with green color and are placed on diagonal, improper are in red. The test dataset consisted of 18 elements (1 test specimen of T1, T5, T6, T9, T10, 2 specimens of T4 and T8, and 3 specimens of T2, T3, T7). The data inside the boxes are the results obtained for the system. Upper value is the number of classified test



**Figure 8.** The structure of the best system - PNN.

**Table 5.** The most important parameters of tested data analysis systems.

<b>Probabilistic neural network - PNN</b>	
<b>Structure</b>	
Data flow:	unidirectional
Topology:	multilayer
Number of input – output:	36 – 10
Range of input signals:	[-1, 1]
The set of model answers:	$(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}); a_i \in \{0, 1\}$
Number of layers under learning / hidden:	2 / 1
The number of neurons in subsequent layers:	36 – 48 – 10
Type of transfer function in the subsequent layers:	radial – competition
Value of the span (smoothing) coefficient of radial function:	0.5
<b>Training</b>	
Method:	supervised
Training algorithm:	mapping of the training set
Modification method of weights (biases):	–
Error function (adaptation):	–
The target value of the error:	–
Number of epochs:	–
The number of elements the training set:	48
The number of elements the test set:	18
<b>Variables</b>	
Number of weights per single neuron, in the subsequent layers:	36 – 48
Number of biases per single neuron, in the subsequent layers:	1 – 0
The total number of variables (weights + biases network):	$1728 + 480 + 48 + 0 = 2256$
<b>Effectiveness</b>	
Number of classified brands of tea:	10

specimens, below is placed corresponding percent value, i.e. the test set contained 3 specimens of T3 (16.7% of all test specimens). All tea specimens were properly classified (100% in the blue box in bottom right corner, all tea specimens are in green boxes on diagonal).

## 5 Evaluation criteria

A comparison of data analysis systems was performed on the basis of accuracy (AC) and complexity criterions (CC). The AC covers 60% of the total evaluation and CC covers 40%.

The results were collected from 30 completed analyses of each of the systems, except from systems trained with GA, which require long calculation time or ANFIS system for which the subsequent results were identical. All results were normalized in such a way that the worst score in a group of systems is assigned to 0% of the grade, and the best 100%. Assessment values were normalized by the formulas 1 and 2:

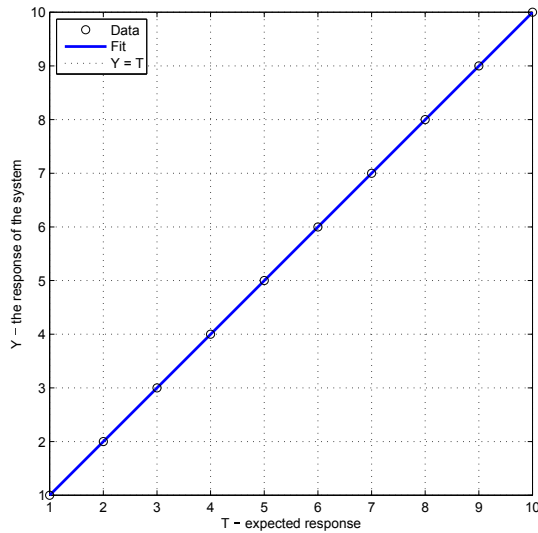
$$y_+ = \left( \frac{x - x_{min}}{x_{max} - x_{min}} \right) \cdot 100\% \quad (1)$$

$$y_- = \left( \frac{x_{max} - x}{x_{max} - x_{min}} \right) \cdot 100\% \quad (2)$$

where:  $y_+$  – normalized rating of the system for a given parameter  $x$ ; used if the rating increases with the parameter  $x$ ,  
 $y_-$  – normalized rating of the system for a given parameter  $x$ , used if the rating decreases with the parameter  $x$ ,  
 $x$  – a value of a parameter obtained by the current system evaluated,  
 $x_{max}$  – a maximum value of a parameter, obtained from all the analyses,  
 $x_{min}$  – a minimum value of a parameter, obtained from all the analyses

During the process of designing the systems the emphasis was placed on the criterion of accuracy, so the main goal was to achieve the lowest error. When the minimal error was reached, the overall number of system parameters was decreased.

The analyses concerned 10 systems evaluated on the basis of three parameters for accuracy criterion, and three parameters for complexity criterion. The evaluation was made on the basis of systems efficiency (it tells how well different brands of tea were classified). It was calculated in Matlab as a number of correct classifications of the pattern in relation to all classifications calculated for the entire test set, according to the



**Figure 9.** Perfect matching of the system’s response to the response model for a test set.

formula 3:

$$\delta = \left( \frac{1}{N} \cdot \sum_{i=1}^N X(i) = \begin{cases} 1 & \text{for correct classification} \\ 0 & \text{for incorrect classification} \end{cases} \right) \cdot 100\% \quad (3)$$

where:  $\delta$  – percentage of correctly classified patterns for the entire test data set,  
 $N$  – number of tea specimens in the test data set,  
 $X(i)$  –  $i$ -th tea specimen

The winning pattern (kind of tea) was based on the intermediate results indicated in accordance with the simple rule “winner takes all” (i.e. out of: 0,7; 0,2 and 0,1 it is 0,7 that wins and is transformed to 1,0). After normalization, the values of all a.m. parameters were multiplied by their respective weights. The individual components of evaluational criteria, together with their corresponding weights are described below.

• **Accuracy criterion (60%) - evaluated system error**

- Maximum number of correctly classified patterns. Calculations included three groups of systems: the best - capable to identify all brands of tea (T1–T10), the worst - six brands, and intermediate - eight. Weight – 35% of the assessment. (AC1)
- Average of results for analyses resulted in correctly classified patterns ( $\delta$ ) for the maximum number of classified teas, given in percentages, for a given system; weight equals to 20%. (AC2)
- Reliability - the number of error-free classifications for all performed calculations, given in percentages, calculated usually for 30 analyses; weight – 5%. (AC3)

1	1 5.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	3 16.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	3 16.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	2 11.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 5.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 5.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
7	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	3 16.7%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
8	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 11.1%	0 0.0%	0 0.0%	100% 0.0%
9	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 5.6%	0 0.0%	100% 0.0%
10	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 5.6%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%
	1	2	3	4	5	6	7	8	9	10	

**Figure 10.** PNN: the results of classification of T1–T10 tea brands for test set. The classifications are marked with the appropriate colors – green (proper, on diagonal) or red (improper).

• **Complexity criterion (40%) - number of system parameters (variables)**

- Number of variables (parameters) in an analysed system, weight 25% of the assessment. (CC1)
- Average computation time (in seconds) needed to train the system, weight 10%. (CC2)
- Average number of epochs needed to train the system, weight 5%. (CC3)

## 6 Discussion

An important issue not taken into account when evaluating the systems (in the tab. 6), is the accessibility of knowledge contained in the systems, used to transform information. In this respect, by far the best are systems based on FL, because we can easily interpret the knowledge contained in the base of inference rules. In contrast, the systems based on ANN have the knowledge contained in the weights and biases of neurons, which unfortunately is not understandable to us.

Table 6 and Figure 11 show the results of the evaluation of developed systems.

As one can see from table 6 systems based on radial neural networks exhibit the best classification of tea specimen patterns: PNN (WA3 parameter value - 91.68%) and RBF (WA3 - 91.64%). Although both systems have WA3 of similar value, the system based on PNN is marginally better than that based on RBF due to the smaller number of variables - biases, occurring in the last layer of the network (parameter CC1: 2256 vs. 2266). The winning systems have high efficiency in classifying the kinds of tea (over 90% of score). The systems based on radial neural networks (PNN and RBF) and fuzzy logic (ANFIS and FUZZY) obtained the best results in accuracy criteria, as they always classified all kinds of tea perfectly. The corresponding parameters AC1, AC2, AC3 and WA1 are equal to 100%. All a.m. systems have also the shortest training time (CC2 and

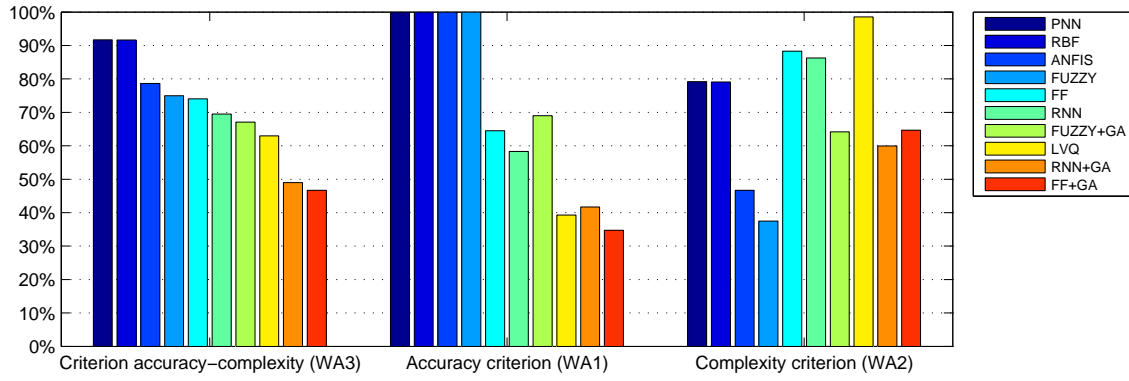


Figure 11. Results of qualitative analysis of systems from Table 6.

Table 6. The comparison of results obtained for the developed systems. Normalized values ( $y_+$  and  $y_-$ ) are given in parentheses. The parameters WA1, WA2 and WA3 are calculated using normalized values of  $y_+$  and  $y_-$ . The best results are marked green and the worst red.

Legend										
Accuracy criterion				Complexity criterion			Weighted average			
AC1	—	maximum number of correctly classified the brands of teas		CC1	—	number of variables (parameters)	WA1	—	accuracy criterion	
AC2	—	average number of correctly classified specimens of teas		CC2	—	average computation time	WA2	—	complexity criterion	
AC3	—	reliability - the number of error-free classification		CC3	—	average number of epochs	WA3	—	criterion accuracy-complexity	
The results of qualitative analysis of systems										
Pos.	Name	Accuracy criterion			Complexity criterion			Weighted average		
		AC1	AC2	AC3	CC1	CC2	CC3	WA1	WA2	WA3
	Unit	[ ]	[%]	[%]	[ ]	[s]	[ ]	[%]	[%]	[%]
	Weight	35%	20%	5%	25%	10%	5%	60%	40%	100%
1	PNN	10 (100%)	100% (100%)	100% (100%)	2256 (66.71%)	1.00 (100%)	1.00 (100%)	100%	79.19%	<b>91.68%</b>
2	RBF	10 (100%)	100% (100%)	100% (100%)	2266 (66.54%)	1.00 (100%)	1.00 (100%)	100%	79.09%	<b>91.64%</b>
3	ANFIS	10 (100%)	100% (100%)	100% (100%)	5232 (16.22%)	5.00 (99.32%)	20.00 (93.65%)	100%	46.67%	<b>78.67%</b>
4	FUZZY	10 (100%)	100% (100%)	100% (100%)	6188 (0%)	1.00 (100%)	1.00 (100%)	100%	37.50%	<b>75.00%</b>
5	FF	10 (100%)	72.22% (15.27%)	33.33% (13.04%)	762 (92.06%)	20.93 (96.60%)	141.73 (52.93%)	64.51%	88.30%	<b>74.03%</b>
6	RNN	10 (100%)	67.21% (0%)	23.33% (0%)	1018 (87.72%)	59.73 (89.98%)	86.17 (71.52%)	58.33%	86.26%	<b>69.50%</b>
7	FUZZY GA	8 (50%)	99.29% (97.83%)	90.00% (86.96%)	968 (88.56%)	380.00 (35.32%)	300.00 (0%)	69.02%	64.18%	<b>67.09%</b>
8	LVQ	6 (0%)	98.72% (96.09%)	90.00% (86.96%)	294 (100%)	4.33 (99.43%)	32.63 (89.42%)	39.27%	98.54%	<b>62.98%</b>
9	RNN GA	6 (0%)	100% (100%)	100% (100%)	536 (95.89%)	587.00 (0%)	300.00 (0%)	41.67%	59.93%	<b>48.97%</b>
10	FF GA	6 (0%)	96.36% (88.90%)	70.00% (60.87%)	436 (97.59%)	508.50 (13.40%)	292.00 (2.68%)	34.71%	64.68%	<b>46.69%</b>

CC3 values). The systems based on FL yielded worse marks resulting from the imposed priority on the accuracy criterion - they have an enormous complexity (CC1: 5232 and 6188). The best system according to complexity criteria was LVQ based on neural networks. The high value of parameter WA2 = 98.54% is connected with small number of variables (CC1 - 294). LVQ classified properly only 6 of 10 kinds of tea. FF network exhibits greater stability during learning and greater accuracy than RNN (AC2: 72.22% vs. 67.21% and AC3: 33.33% vs. 23.33%), with lower complexity due to a decreased number of variables and lower computation time (CC1: 762 vs. 1018 and CC2: 20.93 vs. 59.73).

The worst results exhibit systems based on ANN, both FF (WA3 - 46.69%) and RNN (WA3 - 48.97%), trained only with GA, probably because of too small training set. Both systems (FF and RNN + GA) characterize also the longest training time (CC2: 508.5 and 587, CC3: 292 and 300). The system using RNN + GA exhibits greater stability during learning and greater accuracy than FF + GA (AC2: 100% vs. 96.36% and AC3: 100% vs. 70%), at the expense of increased complexity due to increased number of variables and longer computation time (CC1: 536 vs. 436 and CC2: 587 vs. 508.5). System FUZZY based on fuzzy logic trained with GA was some better and it classified properly only 8 kinds of tea (AC1 - 8). It has the highest number of training epochs (CC3 - 300) and long training time (CC2 - 380).

As we can see, the systems based on the radial neural networks worked the best at solving problems requiring high accuracy, even if we have a small training set. Fuzzy systems are best suited to problems where the complexity does not play a significant role, but we require high accuracy and have access to collected knowledge.

## 7 Conclusions

The authors proposed the stages of data pre-processing needed to prepare relevant information for the developed AI systems. A novel method of forming input data vector by aggregation of the first three principal components was proposed. A dozen of data analysis systems based on ANN, fuzzy system and hybrid systems (evolutionary-neural, neuro-fuzzy and evolutionary-fuzzy) were developed and evaluated. The systems were used to classify ten kinds of tea. All the methods were optimized according to assumed criteria. The best results in complexity criterion was obtained for systems based on radial neural networks (both PNN and RBF) and fuzzy logic (FUZZY and ANFIS). These systems accurately classified all kinds of tea.

In future the authors plan to study systems based on ANN (FF and RNN) trained GA and then optimized by the LM algorithm. This solution should eliminate the main disadvantage of ANN, i.e a random choose of initial values of network weights and biases, what frequently reaching of local minimum, not global. Such combined system should have improved efficiency.

## Acknowledgements

The authors thank Professor Zakrzewska for valuable advice and guidance on the preparation of this paper. We are grateful also Dr. Lubecka and Dr. Czapl, who performed the measurements and supplied us with the data from the e-nose. This work has been funded in part by project "Micro- and nano- systems for chemistry and biomedical diagnostics - MNS-DIAG" co-financed by the European Union via the European Regional Development Fund and the Polish state budget in the framework of the Operational Programme Innovative Economy 2007-2013, contract No. UDA-POIG.01.03.01-00-014/08 and in part by Polish Ministry of Higher Education, Grant no. 11.11.120.614.

## REFERENCES

- [1] M. Aleixandre, J. Lozano, J. Gutierrez, I. Sayago, M.J. Fernandez, and M.C. Horrillo. Portable e-nose to classify different kinds of wine. *Sensors and Actuators B: Chemical*, 131:71–76, 2008.
- [2] K. Brudzewski, S. Osowski, and T. Markiewicz. Classification of milk by means of an electronic nose and svm neural network. *Sensors and Actuators B: Chemical*, 98:291–298, 2004.
- [3] S. Capone, M. Epifani, F. Quaranta, P. Siciliano, and L. Vasanelli. Application of semiconductor solgel sensor array to the discrimination of pollutants in air. *Thin Solid Films*, 391:314–319, 2001.
- [4] S. Capone, P. Siciliano, N. Barsan, U. Weimar, and L. Vasanelli. Analysis of CO and CH<sub>4</sub> gas mixtures by using a micromachined sensor array. *Sensors and Actuators B: Chemical*, 78:40–48, 2001.
- [5] C. Cimpoiu, V.-M. Cristea, A. Hosu, M. Sandru, and L. Seserman. Antioxidant activity prediction and classification of some teas using artificial neural networks. *Food Chemistry*, 127:1323–1328, 2011.
- [6] P. Corcoran and H. Shurmer. An intelligent gas sensor. *Sensors and Actuators A: Physical*, 41-42:192–197, 1994.
- [7] J.W. Gardner and P.N. Bartlett. A brief history of electronic noses. *Sensors and Actuators B: Chemical*, 18-19:211–220, 1994.
- [8] T. Grandke and W.H. Ko. *Sensors: A Comprehensive Survey, Vol. 1, Fundamentals and General Aspects, Update*. John Wiley & Sons, 1996.
- [9] R. Ionescu, E. Llobet, S. Al-Khalifa, J. Gardner, X. Vilanova, J. Brezmes, and X. Correig. Response model for thermally modulated tin oxide-based microhotplate gas sensors. *Sensors and Actuators B: Chemical*, 95:203–211, 2003.
- [10] R. Kaur, R. Kumar, A. Gulati, C. Ghanshyam, P. Kapur, and A.P. Bhondekar. Enhancing electronic nose performance: A novel feature selection approach using dynamic social impact theory and moving window time slicing for classification of Kangra orthodox black tea (*Camellia sinensis* (L.) O. Kuntze). *Sensors and Actuators B: Chemical*, 166-167:309–319, 2012.
- [11] D.S. Lee, J.S. Huh, and D.D.Lee. Classifying combustible gases using micro-gas sensor array. *Sensors and Actuators B: Chemical*, 93:1–6, 2003.
- [12] D. Massart, B. Vandeginste, S. Deming, Y. Michotte, and L. Kaufman. *Chemometrics: a textbook*. Elsevier, 1990.
- [13] W. Maziarz and T. Pisarkiewicz. Gas sensors in a dy-

- dynamic operation mode. *Measurement Science and Technology*, 19(055205), 2008.
- [14] W. Maziarz, P. Potempa, A. Sutor, and T. Pisarkiewicz. Dynamic response of a semiconductor gas sensor analysed with the help of fuzzy logic. *Thin Solid Films*, 436:127–131, 2003.
- [15] Alpha M.O.S. *Technical note*. Av Didier Daurat - Le St Exupery - 31400 Toulouse - France, October 2002. [www.alpha-mos.com](http://www.alpha-mos.com).
- [16] Alpha M.O.S. *Technical note N-P-01*. Av Didier Daurat - Le St Exupery - 31400 Toulouse - France, October 2002. [www.alpha-mos.com](http://www.alpha-mos.com).
- [17] Alpha M.O.S. *Technical note N-SAS-02*. Av Didier Daurat - Le St Exupery - 31400 Toulouse - France, October 2002. [www.alpha-mos.com](http://www.alpha-mos.com).
- [18] Alpha M.O.S. *Technical note N-SAS-03*. Av Didier Daurat - Le St Exupery - 31400 Toulouse - France, October 2002. [www.alpha-mos.com](http://www.alpha-mos.com).
- [19] Alpha M.O.S. *Technical note N-SAS-05*. Av Didier Daurat - Le St Exupery - 31400 Toulouse - France, October 2002. [www.alpha-mos.com](http://www.alpha-mos.com).
- [20] S. Nakata, K. Neya, and K.K. Takemura. Non-linear dynamic responses of a semiconductor gas sensor - Competition effect on the sensor responses to gaseous mixtures. *Thin Solid Films*, 391:293–298, 2001.
- [21] M. Palit, B. Tudu, N. Bhattacharyya, A. Dutta, P.K. Dutta, A. Jana, R. Bandyopadhyay, and A. Chatterjee. Comparison of multivariate preprocessing techniques as applied to electronic tongue based pattern classification for black tea. *Analytica Chimica Acta*, 675:8–15, 2010.
- [22] L. Rutkowski. *Computational intelligence: methods and techniques*. Springer, 2008.
- [23] R. Tadeusiewicz. *New Trends in Neurocybernetics. Computer Methods in Materials Science*, 10:1–7, 2010.
- [24] R. Tadeusiewicz. Place and Role of Intelligent Systems in Computer Science. *Computer Methods in Materials Science*, 10:193–206, 2010.
- [25] R. Tadeusiewicz. *How Intelligent Should Be System for Image Analysis?*, volume 339. Springer Verlag, Berlin, Heidelberg, New York, 2011. Preface to book: H. Kwasnicka, L.C. Jain (Eds.): *Innovations in Intelligent Image Analysis*. *Studies in Computational Intelligence*.
- [26] R. Tadeusiewicz. *Introduction to Intelligent Systems*. CRC Press, Boca Raton, 2011. In book: B.M. Wilamowski, J.D. Irvin (Eds.): *The Industrial Electronics Handbook Intelligent Systems*.
- [27] R. Tadeusiewicz and J. Morajda. *Artificial Intelligence Methods*. Cracov University of Economics Publishing House, Krakow, 2012. Chapter in book: Lula P., Paliwoda-Pekosz G. (eds.): *Analysis and Data Processing Computer Methods*.
- [28] B. Tudu, A. Jana, A. Metla, D. Ghosh, N. Bhattacharyya, and R. Bandyopadhyay. Electronic nose for black tea quality evaluation by an incremental RBF network. *Sensors and Actuators B: Chemical*, 138:90–95, 2009.
- [29] B. Wang, S. Xu, and D.-W. Sun. Application of the electronic nose to the identification of different milk flavorings. *Food Research International*, 43:255–262, 2010.
- [30] C. Wongchoosuk, A. Wisitsoraat, A. Tuantranont, and T. Kerdcharoen. Portable electronic nose based on carbon nanotube-SnO<sub>2</sub> gas sensors and its application for detection of methanol contamination in whiskeys. *Sensors and Actuators B: Chemical*, 147:392–399, 2010.