Bulgarian Journal of Agricultural Science, 31 (No 6) 2025, 1234–1244

Application of intelligent classifier for multi-dimensional identification of food and beverages

Mariyana Sestrimska, Tanya Titova* and Veselin Nachev

University of Food Technologies, Department of Electrical Engineering, Electronics and Automation, Plovdiv, Bulgaria
**Corresponding author:* t_titova@abv.bg

Abstract

Sestrimska, M., Titova, T. & Nachev, V. (2025). Application of intelligent classifier for multi-dimensional identification of food and beverages. *Bulg. J. Agric. Sci., 31*(6), 1234–1244

This publication presents models of intelligent recognition systems aimed at the non-destructive quality of food and beverages, in particular, fruit yogurts and tea blends with herbal and fruit ingredients. The experimental study is based on obtaining diffuse reflectance spectral curves for the tea samples and measuring basic physicochemical parameters, such as pH, total soluble solids, and color, for the sour milks.

Naive Bayes Classifier, Back-propagation Artificial Neural Network, and Self-Organizing Map models were synthesized, and the highest classification accuracy was achieved with PC-BP-ANN.

Keywords: Naive Bayes classification; Back-propagation Artificial Neural Network; Self-Organizing Map; PCA; yogurt; tea

1. Introduction

In scientific research, it is becoming increasingly important to solve problems of a diverse nature, both theoretical and applied, to make correct conclusions and predictions. The assumption that the world around us and its issues are multiparametric systems leads to the need to use multivariate research methods, based on which reliable classification, modeling, and prediction of specific properties inherent in the research objects can be carried out.

On this basis, it is possible to create mathematical models for predicting the state of objects in various applications, such as quality assessment, authenticity, and technological control.

The processing of multivariate data, obtained directly from simple sensors or through a sophisticated measurement process, has several advantages. It enables the exploitation of both unique and redundant information from each source, as well as the synergistic information gained from their combined use (Nikovski, 2022). Multivariate data tasks require the use of specific methods for evaluating the parameters of the object under study, optimizing analytical signals, and studying their correlation with various experimental factors. Interpretation of analytical data is successfully achieved by using the chemometric methods for "*intelligent analysis*".

Wold et al. (2001) define chemometrics as a part of chemical disciplines in which, based on mathematical, statistical, and other methods of formal logic, models are created for a specific product (object). Optimum or modified procedures are sought to experiment as well as to extract maximum useful information from the acquired experimental data.

A large part of scientific research in the food industry in recent years has been devoted to the idea of creating a conceptual model for the evaluation of quality indicators of food and beverages.

There is a growing interest in creating adaptive intelligent recognition techniques that seek an analogue to the human sensory system. Several studies have focused on the im-

plementation of non-destructive recognition (classification) systems, including UV/VIS/NIR spectral analysis, Mass spectrometry, "electronic nose", and "electronic tongue" techniques.

The authors Huang et al. (2015), used an electronic nose to classify honey according to its botanical and geographical origin. They analyzed a total of 84 samples, divided into two classes based on their geographical origin: China and Australia.

The research employed elements of pattern recognition methods, including Principal Component Analysis (PCA), Partial Least Squares (PLS), and Support Vector Machine (SVM), to create predictive models for key physicochemical parameters of honey, such as glucose, fructose, hydromethylfurfural, amylase activity, and acidity. PLS-SVM model achieved the highest prediction accuracy of.

The authors demonstrate that the use of an electronic nose, combined with chemometric approaches for data analysis, can be successfully applied to identify honey based on its botanical and geographical origin of its constituents (Huang et al., 2015).

Liu et al. (2013) used high-performance gas chromatography and an electronic nose to characterize a modified variety of white pepper (*Piper nigrum L.*). The total protein, lipid, starch, piperine, and essential oil contents were analyzed. The electronic nose system was used to create a general gas profile of the investigated samples.

Parra et al. (2006) propose an electronic tongue system for the identification of red wine samples, based on the aging time and the type of barrels in which they were stored. The experimental data were analyzed using PCA. The method divides the studied samples into two classes based on aging time (3 and 6 months). Each of the classes is divided into nine subclasses (cluster groups) according to the type of material from which the barrel is made. Based on the analysis, a predictive model was developed to predict the aging time of the wine. The predictive model was developed using the Soft Independent Model of Class Analogy (SIMCA) algorithm.

Gang and Yang (2013) studied the classification possibilities of four different coffee brands based on VIS/NIR spectral analysis. For this purpose, one qualitative and three quantitative discrimination algorithms were considered – PCA, Linear Discriminant Analysis (LDA), artificial neural network, with „backpropagation " training algorithm (BP-ANN), and SVM classifier.

Khashman (2011) describes an automated system for classifying individual portions of chicken meat. The research aims to build and train an artificial neural network to recognize six different portions of chicken: bone-in breast, leg, boneless fillet, whole leg, and shoulder wings. The development involved taking high-resolution RGB color images of 288 portions of meat. To ensure efficient training of the neural network, the pictures were taken at eight different angles.

A three-layer artificial neural network with 400 neurons in the input layer, 30 neurons in the hidden layer, and six output neurons was trained to recognize the different portions of a chicken, achieving an accuracy of 97.57% in 206 seconds.

According to the same author, the good results achieved in the study could be used to implement an automated sorting system that does not have physical contact with the raw poultry meat. Such a system would limit the risk of biological contamination of the food product and reduce labor costs.

The identification of foods with varying compositions is crucial for accurately determining their quality and safety.

In the context of the presented works, the research aims to synthesize intelligent classifiers for analyzing types of fruit yogurt and multi-component herbal tea mixtures with fruit ingredients.

2. Methodology section

Bulgarian yogurt is among the most popular food products produced and consumed worldwide. It is considered part of food products with a high energy value and a positive bioactive effect, usually enhanced by the addition of prebiotic ingredients and probiotic bacteria (Aryana and . McGrew, 2007; Staffolo et al., 2004). To increase the functionality and antioxidant capacity of dairy products, various types of fruit are often added.

2.1. *Preparation of dairy fruit samples*

For the study, 360 samples of fruit-sour milk were prepared and divided into three groups. The first group consists of 120 yogurt samples from three different brands, purchased from the Bulgarian retail network. A fruit puree of strawberries, raspberries, and blueberries is added to the milk base in a ratio of 80% to 20%.

The second group of samples consists of 120 samples of three types of fruit yogurts from a single brand.

The third set of 120 samples consisted of three different brands of commercial yogurt colored with food dye E122 (azorubin, carmoisine). The dye was added to ensure that the color of the dyed samples matched the color of the samples from the other two groups as closely as possible.

The experimental study of the prepared dairy-fruit samples involves measuring the hydrogen index (*pH*), the total content of soluble dry substances expressed as sucrose equivalent (*Brix*), and determining color indicators in two color spaces – *CIELab* and *XYZ*.

A digital pH-meter (PH-201, Micronix) was used to determine the pH of the milk and fruit samples, and all measurements were performed at a relatively constant temperature of 20–22°C.

Total soluble solids were measured using a refractometer (HI 9680, HANNA).

The color descriptors of the studied samples were determined using a fiber optic spectrometer (USB4000, Ocean Optics) in combination with specialized software, SpectraSuite, for obtaining experimental data, which are recorded and stored in digital and graphic form.

The *CIELab* model defines color evaluations through the following components: L^* – illuminance, brightness (ratio between white and black color), a^* – coordinates on the axis corresponding to red-green color, b^* – coordinates on the axis corresponding to yellow-blue color, c^* – color saturation, h^* – color shade.

The XYZ space is a modification of the RGB color model. Based on the conducted experiment, a 10-dimensional vector of the form: $\vec{x}$ (*pH, TSS, X, Y, Z, L*, a*, b*, c*, h**). As a result, a sample set was obtained, including primary information for each object in the form of a vector of numerical values (with a dimension of 360 elements).

2.2. *Preparation of tea samples*

The study included 350 samples of five different types of tea. The first class (group) consists of 100 samples of herbal tea with the following ingredients: linden, chamomile, and mint. Class 2 is composed of 49 samples, a blend of black tea with additions of lemon and lemon peel. In Class 3, 100 samples of the tea type "Ceylon Melange" are grouped, which is a type of black tea grown in Sri Lanka's tea plantations. The tea boasts a rich, aromatic profile and a balanced, sweet, and astringent taste.

Classes 4 and 5 include 100 and 40 tea samples, respectively, which are multi-component mixtures with fruit ingredients, such as papaya, passion fruit, melon, nuts, rose leaves, strawberries, raspberries, blackberries, and blueberries. Multi-sensorial blend with a pronounced fruity taste and aroma.

Tea bags (2–3 g) of each type are steamed with distilled water heated to 100°C for 3–5 minutes. The samples thus prepared are cooled to room temperature (22–23°C), poured into 20 mL glass closed tubes, and prepared for further processing. All tea samples were prepared and stored under the same conditions.

For the prepared tea blends, the diffuse reflectance spectral characteristics in the VIS/NIR range of the electromagnetic spectrum from 470 nm to 1010 nm are measured. Each sample is placed in a 9 cm diameter glass petri dish, with the liquid filling approximately two-thirds of the petri dish's volume. The spectrometer probe is positioned 15 cm from the petri dish, at an angle of 45°. Five spectral features were taken for each sample, and the results were averaged during post-processing.

2.3. *Principal component analysis (PCA)*

One of the most common methods for reducing the dimensionality of a dataset is the statistical technique of *Principal Component Analysis* (PCA). The method diagonalizes the covariance matrix of the data, decorrelates the signal in the transformed region, thereby minimizing the mean *squared error* of the data compression, and merges the elements with the most significant variance (Cabrita et al., 2012). In the space of variables, new directions are formed, where the distance between the points for the corresponding objects (samples) is the largest. These directions are referred to as principal components or eigenvectors. In the calculation process, the principal components are arranged in ascending order of importance. The first eigenvector realizes the most significant part of the variation of the experimental data, the second principal component contains the information not explained by the first, and so on.

The principal components method is an effective way to visualize the information contained in a dataset. It allows for grouping the studied objects according to a specific feature.

2.4. *Naive Bayes classification*

The Bayes Classifier is a classic approach within probabilistic-statistical methods for pattern recognition. The strategy is based on Bayes' Theorem, which allows for predicting class membership by estimating the a priori probability of a given event's realization.

The mathematical expression of Bayes' theorem is as follows (Jadhav and Channe, 2016; Taheri and Mammadov, 2013):

$$P(D_j|s) = \frac{P(s|D_j).P(D_j)}{\sum_{i=1}^{m} P(D_i).P(s|D_i)} \quad (1)$$

where: D_i are the defined classes D = $\{D_1, D_2, ..., D_m\}$;

$P(D_j|s)$ is the posterior probabilities for a feature vector object $s_j = [s_{1j}, s_{2j}, ..., s_{nj}]$;

$P(D_j)$ – prior probabilities for the classes;

$P(s|D_j)$ – a priori probabilities for belonging to a feature vector to a corresponding class D_j;

$\sum_{i=1}^{m} P(D_i.P(s|D_j)$ – normalizing component,

The final classification decision is arrived at by maximizing the posterior probability where the vector s is most strongly associated with the class D_j:

$$s \in D_j \Leftrightarrow \underset{D}{\text{argmax}}[P(D_j|s)]. \quad (2)$$

The Bayes Classifier yields relatively good results when implemented with physically distinct quality features. The main drawback of this approach could be the difficulty the algorithm encounters when training with large datasets (Dresp-Langley and Wandeto, 2021; Blanquero et al., 2021).

2.5. *Self-Organizing Map (Kohonen Network)*

Self-organizing maps (SOMs) are unsupervised learning systems that have a two-layer structure consisting of an input layer and an output competitive layer (Pedró et al., 2019). The input layer contains object features. The output layer is a 2-D self-organizing map in which the neurons of the network "compete" to achieve the best proximity between the model and the object under study. Each neuron of the output layer is functionally connected to the neurons of the input layer. In Figure 1, the structure of Kohonen's self-organizing neural network is graphically presented.

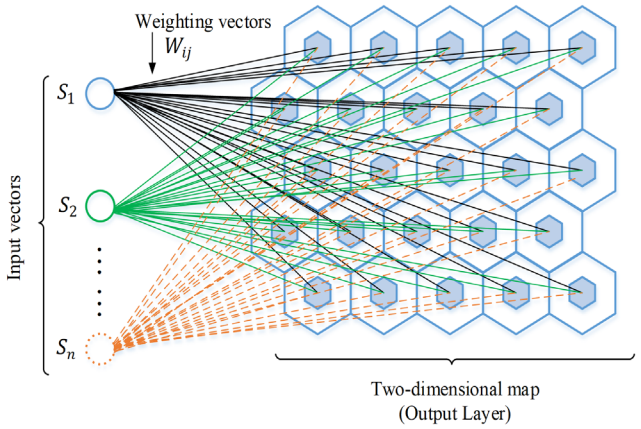

Fig. 1. Schematic diagram of the *Self-Organizing* Network

Source: Authors' own elaboration

The neurons of the network compete with each other, according to classical regression dependence:

$$modelNet = \Sigma_{i=1}^{n} S_{ij}.W_{ij} \quad (3)$$

where: S_{ij} – input vectors for the artificial neural network;

W_{ij} – weight vector connecting the i^{th} neuron of the input layer to the element of the output map.

The values of the weight vectors are modified adaptively for each iteration (repetition) of the learning process.

The learning algorithm for self-organizing neural networks is called *Winner Takes All* or *Winner Takes Most*. In the initial stage of training, the weight vectors have randomly and uniformly distributed values in the network architecture. When a data vector enters the input layer, each neuron in the output layer strives to achieve the most significant similarity to the elements of the input vector. The "winning" neuron is the one for which the value of the product $(S_{ij}.W_{ij})$ is the largest, resulting in the winning neuron generating a "+1" at the output of the network (Musci et. al, 2020; Pedró et. al, 2019).

The winning neurons and their neighbors are determined in the training process by sequentially adjusting the weights according to the formula:

$$W_{ij}(\mathrm{k}+1) = W_{ij}(\mathrm{k}) + \alpha(t).[S_{ij}(\mathrm{k}) - W_{ij}(\mathrm{k})] \quad (4)$$

where: $\alpha(t)$ is a monotonic decreasing function defining the learning rate (Dresp-Langley and Wandeto, 2021; Damyanov, 2006; Mici et al., 2018).

To determine the neurons adjacent to the winner, a similarity metric function is defined, which aims to find the minimum distance between the weight vector for the corresponding neuron and the input vector. The most commonly used measure of proximity is Euclidean distance:

$$D(S_j, W_{ij}) = \sqrt{\Sigma_{j=1}^{n}(S_j - W_{ij})^2} \quad (5)$$

Kohonen's artificial self-organizing networks find application in practice, with their main goal being to compress high-dimensional input data into a 2D hexagonal map that represents the studied objects using a smaller number of neural weights (Zhang et al., 2022).

2.6. *Backpropagation Artificial Neural Network (BP-ANN)*

The *Backpropagation algorithm* refers to the "supervisory" methods of machine learning (machine learning). *Backpropagation Artificial Neural Network* has the architecture of a multilayer perceptron, made up of one input layer, one (or several) hidden layers, and one output layer. Each input vector corresponds to a target output vector, which is composed of the desired values for the elements in the output layer of the neural network – the weight of the net elements is recalculated (adjusted) using a rule that aims to minimize the error in the output layer (Skowron et al., 2019; Amin, 2020).

The training stage proceeds in two conditional directions: forward and reverse. In the straight-through run, the training vectors are fed to the network's input, and successively, in the following layers, the values of the weight vectors are calculated. The element values of the output layer are compared to the desired output for a corresponding input vector. In the reverse pass, the error between the obtained and desired values for the output vector is re-propagated through

the intermediate (hidden) layers of the network, aiming to minimize the sum of squares error of the output (Saritas and Yasar, 2019).

3. Discussion

3.1. *Identification of dairy fruit samples*

In the study of fruit yogurts, three groups of 120 samples were selected, comprising fruit, commercial brands, and artificially colored samples. The qualitative parameters, acidity, total soluble solids, and color, were measured for them.

The diagrams in Figure 2 present the distribution of a part of the signs selected from the complete set.

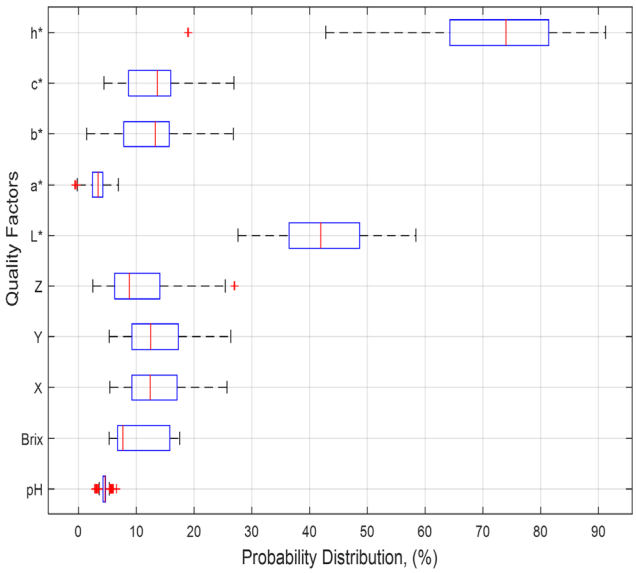

Fig. 2. Box plots for the yogurt dataset
Source: Authors' own elaboration

For the dairy and fruit samples from the three analyzed groups, a Naive Bayes Classifier was synthesized based on two defining characteristics: pH and total soluble solids (TSS). Figure 3 shows the Gaussian distribution of the two factors.

The Naive Bayes Classifier model predicts the class (group) belonging of the studied samples with a total of 8 errors between the first and third groups. Table 1 presents the error matrix for the synthesized classifier.

To utilize the entire available experimental data set for dairy-fruit products, the principal components analysis method was employed. In Figure 4 the location of the sour milk samples is visualized in the space formed by the first three eigenvectors, the result of the PCA analysis. The formation of three relatively compact cluster groups was observed, corresponding to the a priori grouping of the studied samples.

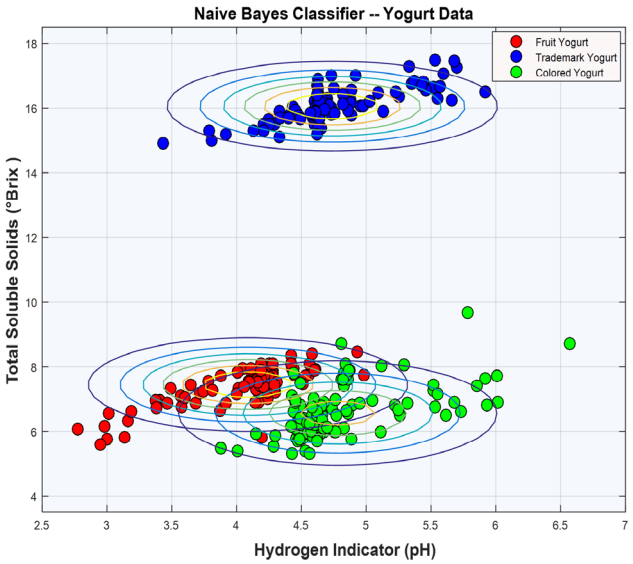

Fig. 3. Gaussian contours for two factors – pH and TSS
Source: Authors' own elaboration

Table 1. Confusion matrix

Naive Bayes Classifier		Actual Class			e_i, (%)
		FY	TY	CY	
Predicted Class	FY	116	0	4	3.33
	TY	0	120	0	0.00
	CY	4	0	116	3.33
g_i, (%)		0.33	0.00	0.33	2.22

*FY – Fruit Yogurt, TY – Trademark Yogurt, CY – Colored Yogurt,
g_i – false omission rate; e_i – false discovery rate
Source: Authors' own elaboration

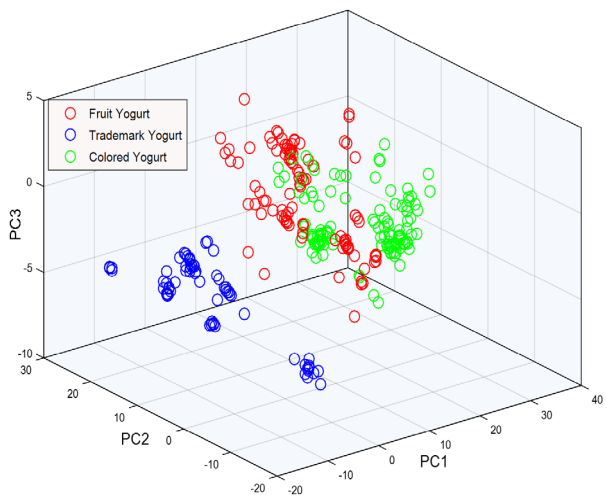

Fig. 4. Projection of the analyzed samples in the space formed by the first three principal components
Source: Authors' own elaboration

The PCA algorithm calculates ten new determinants. In Figure 5, the relationship between the 10 principal components calculated using the statistical approach and the corresponding eigenvector values is presented. The first five principal components collectively account for 99.71% of the variation in the original experimental data; i.e., the five eigenvectors sufficiently accurately characterize the studied objects (samples), and it would be appropriate to use them as input vectors in the synthesis of a recognition classifier.

Table 2 presents the error matrix resulting from the PC-NB prediction model. A relatively large resulting error of 18.33% is obtained, which can be attributed to the loss of information in the reduction of the feature space by PCA.

At the next stage of the study, a model of Kohonen's self-organizing neural network was synthesized. The neurons in the input layer of the network correspond to the first five principal components resulting from the PCA analysis. Figure 6 presents a projection of the factor space defined by the self-organizing network. In this case, the green asterisks represent a projection of the experimental points organized according to the grid algorithm. The red points are the weighted positions of the neurons, and the gray lines express the metric measure of proximity between neighboring neurons – the squared Euclidean distance. Each grid element represents a model, a point projected within the feature space defined by the experimental data. Neurons compete to interpret the experimental values as accurately as possible, with the "winning" neuron being the one for which the model most closely approximates the values of the input vector.

Figure 7 is a projection of the output layer of the Kohonen network – two-dimensional hexagonal map with dimensions of 10×10 neurons. The elements (neurons) in the output layer are determined according to the type of study.

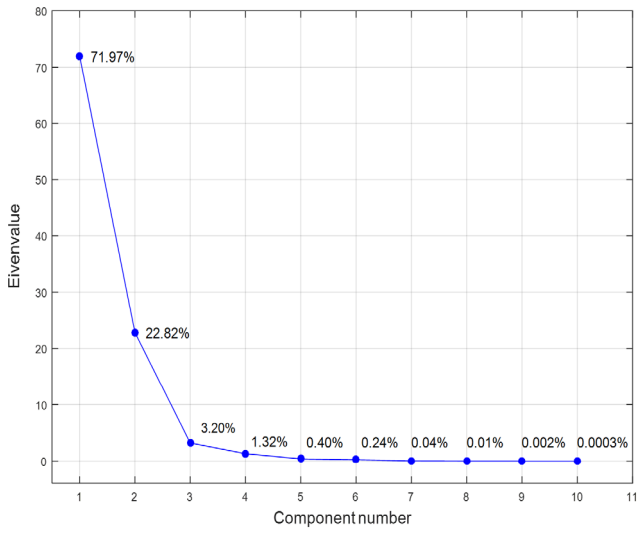

Fig. 5. Relationship between eigenvectors and their eigenvalues
Source: Authors' own elaboration

Table 2. Confusion matrix

PC – NB model		Actual Class			e_i, (%)
		FY	TY	CY	
Predicted Class	FY	107	1	12	10.83
	TY	10	110	0	8.33
	CY	40	3	77	35.83
g_i, (%)		31.85	3.51	13.48	18.33

Source: Authors' own elaboration

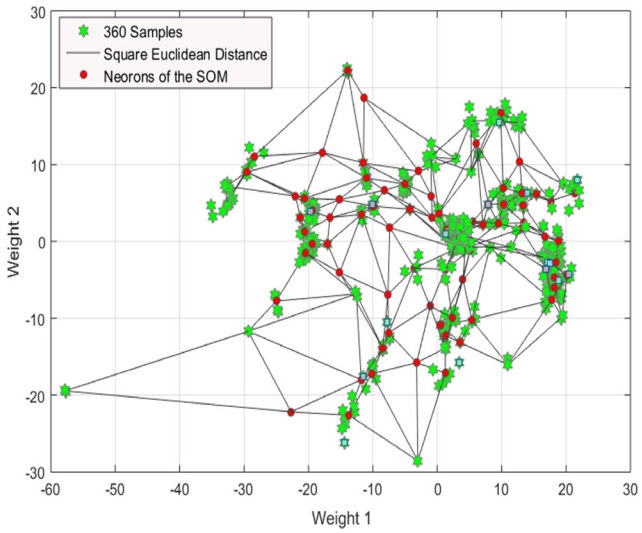

Fig. 6. Self-organizing map weight positions
Source: Authors' own elaboration

The artificial neural network clusters the examined milk-fruit samples into three distinct groups. Dark hexagons represent the "winning" neurons. Their area is proportional to the number of samples strongly associated with that element. Low-density units are assumed to be cluster limits.

Table 3 presents the confusion matrix, which reflects the overall classification error of the PC-SOM model. A total of 8 samples were incorrectly predicted to belong to a group. One sample from group 2 was misidentified as part of the fruit-dairy products (group 1) and seven samples from the colored products were misclassified as part of group 1. In the specific case, Kohonen's self-organizing network achieves a classification accuracy of 97.92% (with an error rate of 2.08%).

Kohonen's artificial neural networks are a good option for analyzing high-dimensional experimental data. The out-

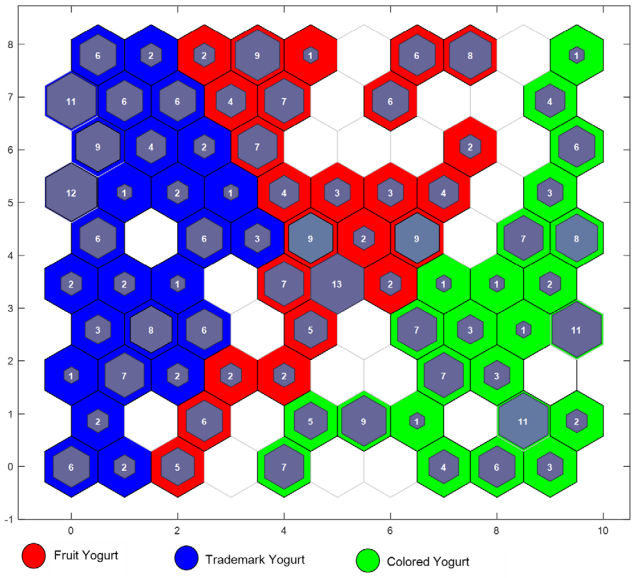

Fig. 7. Self-organizing map for dairy fruit samples
Source: Authors' own elaboration

Table 3. Confusion matrix

PC – SOM model		Actual Class			e_i, (%)
		FY	TY	CY	
Predicted Class	FY	120	1	7	6.25
	TY	0	119	0	0
	CY	0	0	113	0
g_i, (%)		0.00	0.83	5.83	2.08

Source: Authors' own elaboration

put layer is a 2-D hexagonal map model, on which the analyzed objects are grouped according to the grid algorithm.

In the present study of fruit yogurt, high classification accuracy was obtained using PC-SOM. The proposed methodology can be applied to identify food products with varying ingredients, component ratios, and quality levels.

To perform a comparative analysis and select the best classifier, a neural network was synthesized. It has a three-layer structure with *Tangent-sigmoidal* and *Softmax* in the hidden and output layers, respectively. The number of input training vectors in the network is equal to the first five principal components, accounting for approximately 99% of the variation in the experimental data. Neurons in the output layer are coded in binary code combinations.

In percentage terms, the experimental database is divided as follows: 70% for training, 15% for validation, and 15% for testing.

Figures 8 and 9 present the error matrices for the synthe-

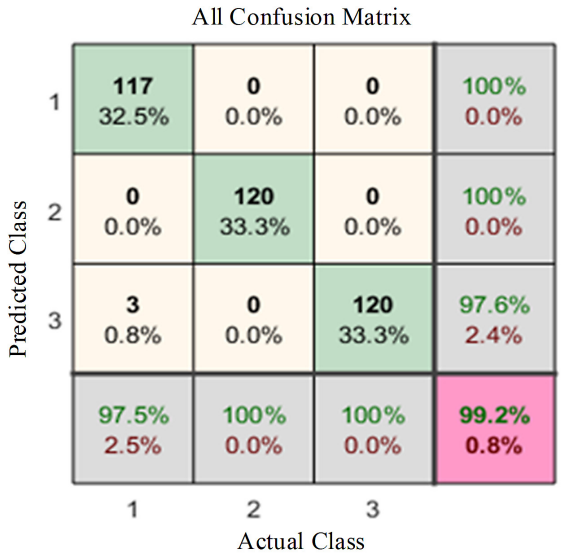

Fig. 8. Summary confusion matrix for dairy fruit samples
Source: Authors' own elaboration

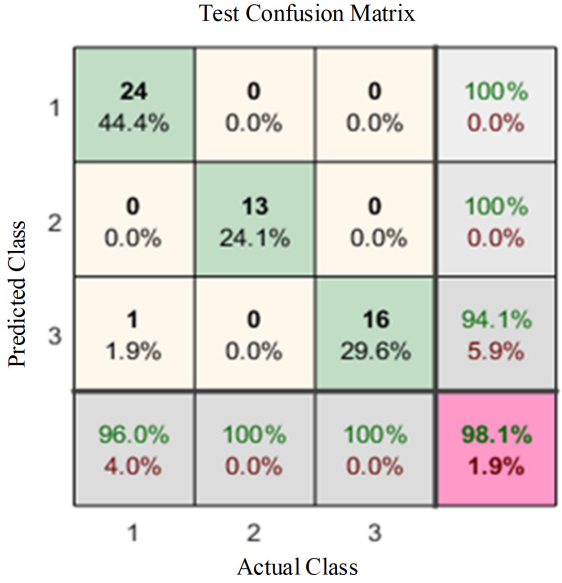

Fig. 9. Confusion matrix for the test set of the back-propagation artificial neural network
Source: Authors' own elaboration

sized neural network during training and testing, respectively. In the training process, three samples were misclassified as part of group 3 instead of group 1. When running the test set, one sample was misidentified, with the error again between fruit and unnatural colored products.

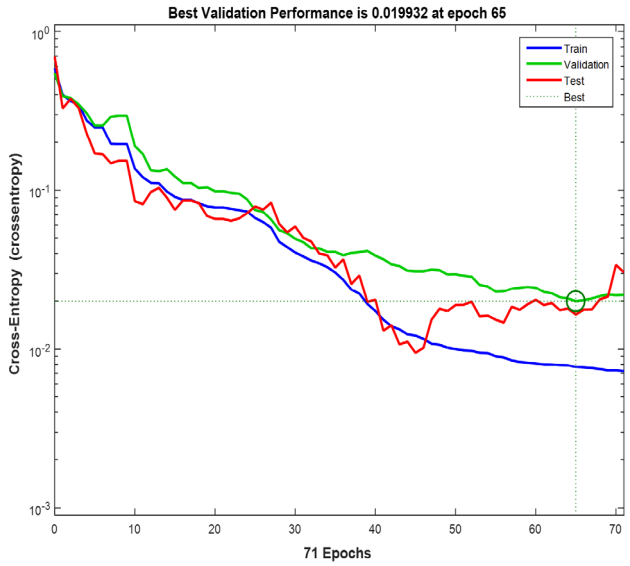

Fig. 10. Plot of mean squared error (MSE) for the learning, test, and validation set of BP-ANN
Source: Authors' own elaboration

Figure 10 presents the variation of the MSE error during training, testing, and validation of the network. The test and validation curves exhibit a similar appearance, characterized by a decreasing pattern of variation, with no overtraining of the network observed.

The overall training and test classification accuracies of the BP-ANN model are 99.2% and 98.1%, respectively.

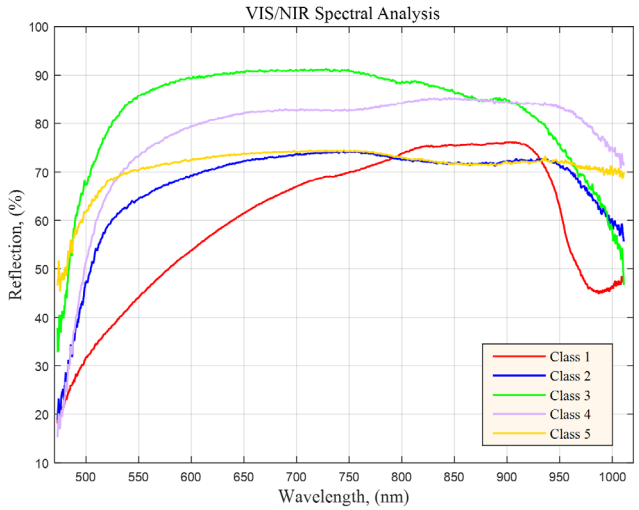

Fig. 11. VIS/NIR spectral characteristics of five types of tea mixtures
Source: Authors' own elaboration

3.2. *Classification of multicomponent tea blends*

A total of 350 tea samples from five types of multicomponent mixtures were included in the study. Diffuse reflectance spectral characteristics were taken for them in the range from 470 nm to 1010 nm of the electromagnetic spectrum.

Figure 11 shows the averaged spectral curves for the five types of tea blends. It can be observed that each of the five classes of analyzed samples exhibits a distinct spectral characteristic.

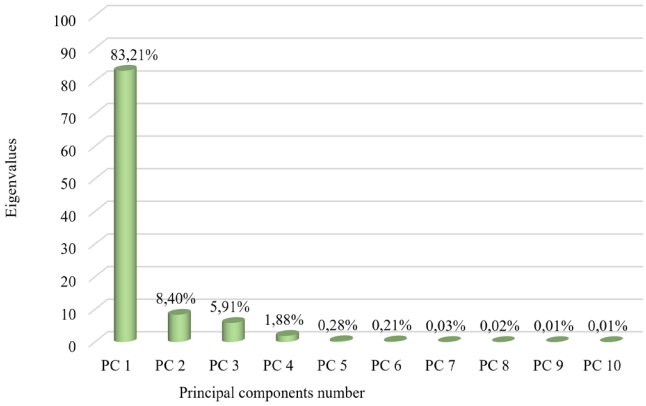

Fig. 12. Eigenvectors, eigenvalues, and variance from principal component analysis
Source: Authors' own elaboration

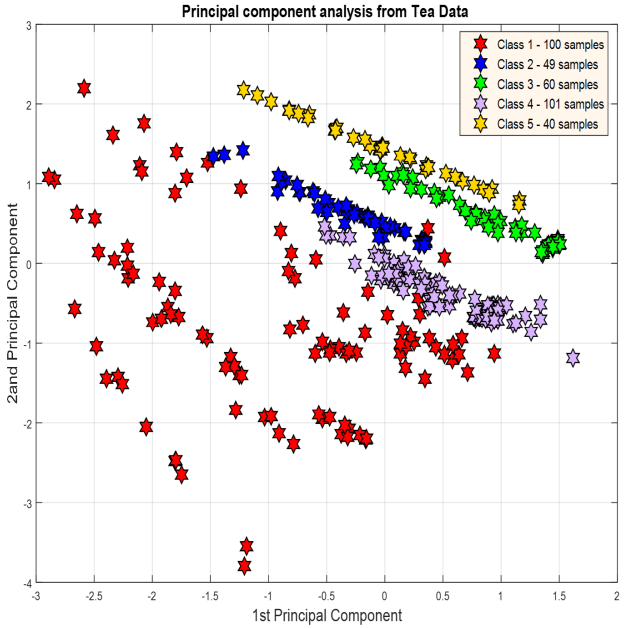

Fig. 13. Score cluster plot in the plane formed by the first and second principal components
Source: Authors' own elaboration

As a result of the experiment, a feature space of the type was formed, i.e., 2830 points were obtained from the spectral curves for each of the 350 analyzed samples.

For the formed database, the Principal Components Analysis method was applied to recalculate and reduce the number of qualitative characteristics accordingly.

The PCA algorithm yields 10 new determinants, whose informative value is represented in Figure 12. The first five principal components collectively account for 99.68% of the variation in the experimental data.

Figures 13 and 14 present projections of the analyzed tea samples in the planes formed by the first and second principal components, respectively.

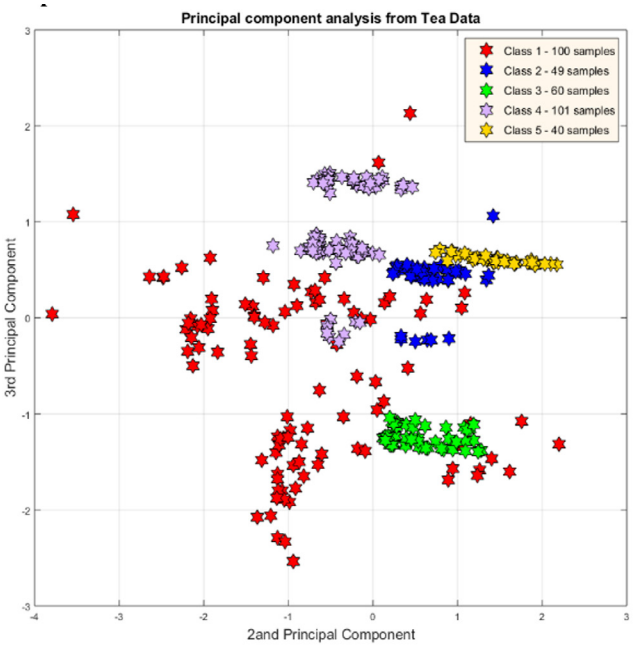

Fig. 14. Score cluster plot in the plane formed by the second and third principal components
Source: Authors' own elaboration

Table 4. Confusion matrix

PC – NB model		Actual Class					e_i, (%)
		CL 1	CL 2	CL 3	CL 4	CL 5	
Predicted Class	CL 1	70	8	0	22	0	30.0
	CL 2	6	30	0	13	0	38.8
	CL 3	0	1	45	14	0	25.0
	CL 4	5	1	34	53	8	47.5
	CL 5	4	11	14	0	11	72.5
g_i, (%)		19.7	41.2	51.6	48.0	42.1	40.3

*" CL" – Class; g_i – false omission rate; e_i – false discovery rate
Source: Authors' own elaboration

A Bayesian classifier was synthesized, using the first five principal components as training vectors. From the error matrix presented in Table 4, it is evident that the PC-NB model accurately identifies the studied tea samples with an overall error of approximately 40%.

At the next stage of the study, a predictive algorithm is sought to recognize the tested samples with higher accuracy. For this purpose, a PC model, specifically a *Self-Organizing Map,* has been implemented. The five eigenvectors calculated by PCA are selected as input vectors of the network. The output layer, with 15 × 15 neurons, is presented in Figure 15. Five distinct cluster groups formed, with 17 samples from class 1 (herbal blends) misidentified as part of class 4 (tea with fruit ingredients). Also, five samples from class 5 were misclassified. Two of these were predicted as part of class 2 (black tea with lemon admixture) and three were misrecognized as part of class 4 (fruit tea). The total classification error of the model is 6.3%. The predicted class affiliation of the analyzed tea samples is presented in Table 5.

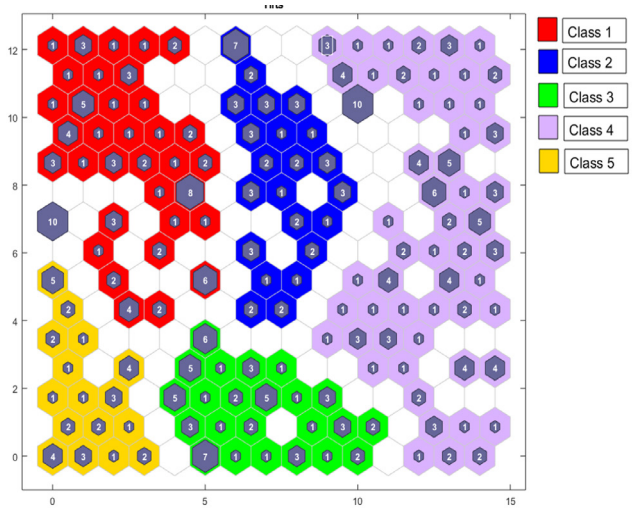

Fig. 15. Clustering through tea data
Source: Authors' own elaboration

Table 5. Confusion matrix

PC – SOM model		Actual Class					e_i, (%)
		CL 1	CL 2	CL 3	CL 4	CL 5	
Predicted Class	CL 1	83	0	0	0	0	0.0
	CL 2	0	49	0	0	2	3.9
	CL 3	0	0	60	0	0	0.0
	CL 4	17	0	0	101	3	16.5
	CL 5	0	0	0	0	35	0.0
g_i, (%)		20.5	0.0	0.0	0.0	12.5	6.3

Source: Authors' own elaboration

Figure 16 presents the so-called unified distance matrix. It helps to define the limits of the cluster groups on the source layer of the Kohonen network. Light colors represent highly interconnected objects, while dark colors signify more strongly differentiated elements, i.e., smaller and larger values for the squared Euclidean distance, respectively.

Figure 17 presents the error matrix for a synthesized PC-Backpropagation Artificial Neural Network. The network identified tea samples with a total of three errors, with 3 class 1 samples misclassified as class 4. The achieved classification accuracy of PC-BP-ANN is within 99.1%.

Summary information on the accuracy of the synthesized PC-NB, PC-SOM, and PC-BP-ANN for the two types of food products analyzed is presented in Figure 18.

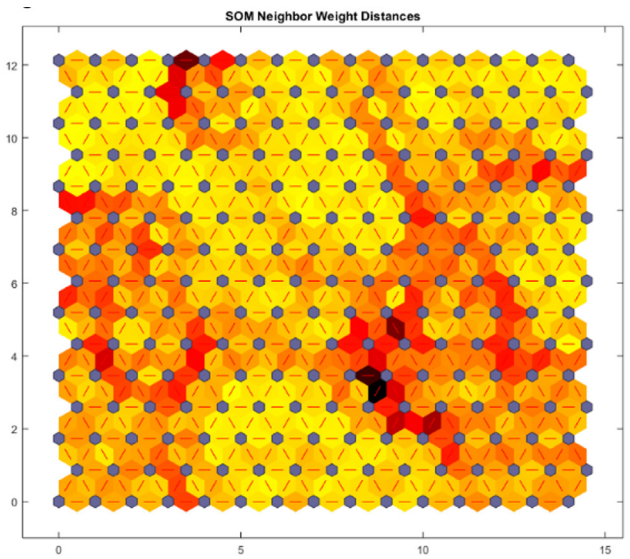

Fig. 16. Unified matrix of distances
Source: Authors' own elaboration

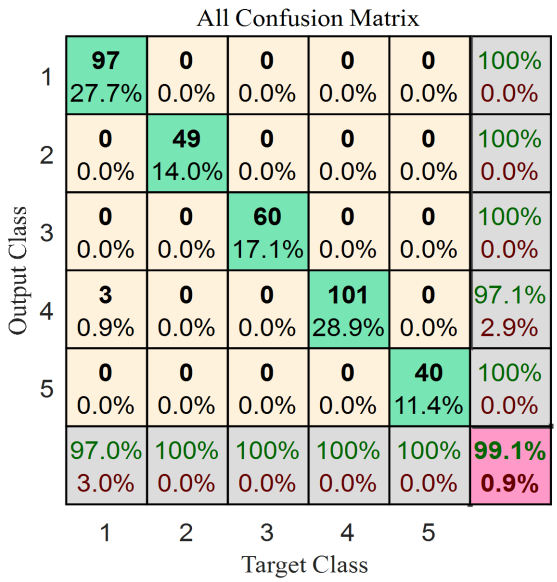

Fig. 17. Confusion matrix for PC-BP-ANN
Source: Authors' own elaboration

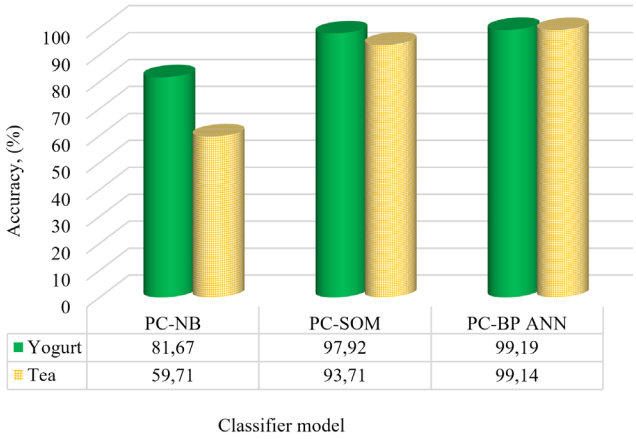

Fig. 18. Generalized classification accuracy of the synthesized intelligent classifiers
Source: Authors' own elaboration

Intelligent classifiers, such as artificial neural networks, perform better than Bayesian classifiers. The highest accuracy in predicting class membership for both sample types was achieved using an artificial neural network with backpropagation.

4. Conclusion

Traditionally used instrumental techniques, compatible with computer technologies and modern data processing methods, lead to the creation of intelligent, adaptive sensor systems that can determine authenticity, degree of freshness, presence of impurities, and a number of other indicators that determine the quality of food products.

This publication presents a strategy for synthesizing a multisensory quality system that combines experimental data sets, mathematical approaches for extracting new, informatively valuable features, and the development of intelligent classifiers tailored to identify foods and beverages with different quality indicators.

References

Amin, S. (2020). Backpropagation – Artificial Neural Network (BP-ANN): Understanding gender characteristics of older driv-

er accidents in West Midlands of United Kingdom. *Safety Science, 122*, 371 – 382.

Aryana, K. J. & McGrew, P. (2007). Quality attributes of yogurt with Lactobacillus casei and various prebiotics. *Lebensmittel Wissenschast and Technologie, 40*, 1808 – 1814.

Blanquero, R., Carrizosa, E., Ramírez-Cobo, P. & Sillero-Denamiel, M. R. (2021). Variable selection for Naïve Bayes classification. *Computers and Operations Research, 135*, 292 – 303.

Cabrita, M., Aires-De-Sousa, J., Da Silva, M. G., Rei, F. & Freitas, A. C. (2012). Multivariate statistical approaches for wine classification based on low molecular weight phenolic compounds. *Australian Journal of Grape and Wine Research, 18*, 138 – 146.

Damyanov, C. (2006). Non-destructive quality recognition in automatic sorting systems of food products -2nd Ed. UFT *Academic Publishing House, Plovdiv*, 349 (Bg).

Dresp-Langley, B. & Wandeto, J. M. (2021). Human symmetry uncertainty detected by a self-organizing neural network map. *Symmetry, 13*, 299 – 314.

Gang, L. & Yang, H. (2011). Discrimination of different brands of Nescafé coffee using VIS-NIR spectroscopy and comparative study. *International Conference on Agricultural and Biosystems Engineering, Advances in Biochemical Engineering, 1*(2), 163 – 166.

Huang, L., Liu, H., Zhang, B. & Wu, D. (2015). Application of electronic nose with multivariate analysis and sensor selection for botanical origin identification and quality determination of honey. *Food and Bioprocess Technology, 8*, 359 – 370.

Jadhav, S. D. & Channe, H. P. (2016). Comparative study of k-NN, Naive Bayes and Decision tree classification techniques. *International Journal of Science and Research, 5,* 1842 – 1845.

Khashman. (2011). Automatic identification system for raw poultry portions. *Journal of Food Process Engineering, 35*, 727 – 734. Wiley Periodicals, Inc.

Liu, H., Zeng, F., Wang, Q., Wu, H. & Tan, L. (2013). Studies on the chemical and flavor qualities of white pepper (Piper nigrum L.) derived from five new genotypes. *Journal European Food Research and Technology, 237*, 245 – 251.

Mici, L., Parisi, G. I. & Wermter, S. (2018). A self-organizing neural network architecture for learning human-object interactions. *Neurocomputing, 307*, 14 – 24.

Musci, M., Parigi, G., Cantoni, V. & Piastra, M. (2020). A scalable multi-signal approach for the parallelization of self-organizing neural networks. *Neural Networks, 123*, 108 – 117.

Nikovski, P. (2022). Multi-sensor configurations in data fusion systems. In AIP Conference Proceedings. *AIP Publishing, 2570*(1), 020018. https://doi.org/10.1063/5.0099614.

Parra, V., Arrieta, A., Ferńandez-Escudero, J., Íniguez, M., de Saja, J. A. & Rodrìguez-Méndez, M. (2006). Monitoring of the ageing of red wines in oak barrels by means of an hybrid electronic tongue. *Analytica Chimica Acta, 563,* 229 – 237.

Pedró, M., Martín-Martínez, J., Maestro-Izquierdo, M., Rodríguez, R. & Nafría, M. (2019), Self-organizing neural networks based on oxRam devices under a fully unsupervised training scheme. *Materials, 12*(21), 3482. https://doi.org/10.3390/ma12213482.

Saritas, M. & Yasar, A. (2019). Performance analysis of ANN and Naive Bayes classification algorithm for data classification. *International Journal of Intelligent Systems and Applications in Engineering, 7*, 88 – 91.

Skowron, M., Wolkiewicz, M., Orlowska-Kowalska, T. & Kowalski, C. T. (2019). Application of self-organizing neural networks to electrical fault classification in induction motors. *Journal: Applied Sciences, 9*, 616 – 638.

Staffolo, M. D., Bertola, N. & Martino, M. A. (2004). Bevilacqua, Influence of dietary fiber addition on sensory and rheological properties of yogurt. *International Dairy Journal, 14*, 783 – 789.

Taheri, S. & Mammadov, M. (2013). Learning the Naive Bayes classifier with optimization models. *International Journal of Applied Mathematics and Computer Science, 23*, 787 – 795.

Wold, S., Sjöström, M. & Eriksson, L. (2001). PLS – regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58*, 109 – 130.

Zhang, K., Lin, N., Tian, G., Yang, J., Wang, D. & Jin, Z. (2022). Unsupervised-learning based self-organizing neural network using multi-component seismic data: Application to Xujiahe tight-sand gas reservoir in China. *Journal of Petroleum Science and Engineering, 209,* 109964, https://doi.org/10.1016/j.petrol.2021.109964.

Received: **April 10, 2024**; *Approved*: **July, 04, 2024**; *Published*: **December, 2025**