https://dx.doi.org/10.21577/0103-5053.20220139

J. Braz. Chem. Soc., Vol. 34, No. 5, 684-693, 2023

Article

Mathematical Mapping Study of Synthetic Antioxidants Behavior in Biodiesel: Application of the Self-Organized Feature Map (SOFM) and Multilayer Perceptron (MLP)

Marco A. J. Clemente,[a] Heloisa H. P. Silva,[a] Julia W. Campos,[a] Nathan F. Silva,[a] Eduardo G. Sousa,[a] Ana C. G. Mantovani,[b] Karina B. Angilelli[a] and Dionisio Borsato[,a]*

[a]*Departamento de Química, Universidade Estadual de Londrina (UEL), 86057-970 Londrina-PR, Brazil*

[b]*Centro Universitário Ingá (UNINGÁ), 87035-510 Maringá-PR, Brazil*

The objective of the present work was to apply self-organizing feature maps (SOFM) and multilayer perceptron (MLP) to evaluate the protective capacity of *tert*-butylhydroquinone (TBHQ), butylhydroxyanisole (BHA), and butylhydroxytoluene (BHT) antioxidants against the biodiesel oxidation reaction. For this, the antioxidant concentration and the biodiesel compliance parameters were used as continuous input variables and the type of antioxidant as a categorical. The SOFM proved to be an adequate tool for the classification of biodiesel samples containing antioxidants. The performance of the 5 most active networks of the model ranged from 89 to 99% for training, testing, and validation with an error below 1.42×10^{-4}. Statistical tests applied to validate the model showed no significant difference between predicted and experimental values. The global sensitivity analysis showed that the relative protection factor (RPF) contributed with 34.89%, the antioxidant with 31.49%, the induction period with 10.69%, the water content with 6.00%, and the others all together with 16.93% in the construction of the regression models by MLP.

Keywords: topological map, artificial neural networks, TBHQ, BHA, BHT

Introduction

Biodiesel is a generic denomination for fuels produced from sustainable and renewable energy sources, such as vegetable oils, animal fat, and frying residues, to be used in compression ignition engines since its properties are generally similar to those of diesel.[1-8] However, these raw materials have chemical substances in their composition containing unsaturated bonds, which are transferred to biodiesel making the obtained product susceptible to oxidation.[9,10]

Oxidation reactions occur in stages, forming products such as acids, aldehydes, ketones, and polymers, among others, that can damage the engine and modify the biodiesel physicochemical properties, such as density, flash point, iodine index and cetane, cloud and pour point, acidity index, among others. These changes lead to a loss of quality and non-compliance with the standards required by legislation.[11-14] Therefore, to be commercialized, biodiesel must keep its essential characteristics little changed during the storage period. Thus, the maintenance of its quality for a long time has been a major concern for biodiesel producers and suppliers.[15,16]

The addition of antioxidants is an efficient and low-cost way to delay the oxidation reaction of biodiesel. Synthetic antioxidants are industrially used for this purpose, ensuring the properties of the biodiesel for a longer period.[17,18] The most used antioxidants are phenolic compounds that promote the removal or inactivation of free radicals produced in the biodiesel oxidation reaction, that is, interrupting the propagation of the radical oxidation. The most used are *tert*-butylhydroquinone (TBHQ), butylhydroxytoluene (BHT), and butylhydroxyanisole (BHA).[17,19]

To evaluate the concentration effect of different antioxidants added in biodiesel samples, computational tools, which work on data analysis such as artificial neural networks (ANNs) can be used. Several data processing techniques stand out: multilayers perceptron (MLP), radial basis and self-organizing feature maps (SOFM), and others.[20,21] ANNs try to mathematically model the logical operations that the brain performs when executing tasks. These tools have gained importance in pattern recognition, images and objects, construction engineering, food science, and financial forecasting.[20-26]

*e-mail: dborsato@uel.br
Editor handled this article: José Walkimar M. Carneiro

This work aimed to apply and adapt SOFM and MLP neural networks to study the efficiency of synthetic antioxidants in a mixture with commercial biodiesel.

Experimental

Biodiesel

Twelve samples of commercial B100 biodiesel were used, with no antioxidant additives, supplied by the Laboratory of Research and Analysis of Fuels of the Department of Chemistry of the State University of Londrina. Samples with induction period values below those allowed by the EN14214[27] standard were chosen to justify the use of synthetic antioxidants.

Biodiesel physical-chemical characterization

The analyzed parameters were: density (ASTM D4052),[28] flash point (ASTM D93),[29] water content (ASTM D6304),[30] acid number (ASTM D664),[31] viscosity (ASTM D445-06),[32] cloud point and pour point (ASTM D2500),[33] and oxidative stability (EN 14112).[34]

Antioxidants

The antioxidants used were butylhydroxyanisole (BHA) (Diadema, Brazil, Synth, 98.5% purity, P.A.), butylhydroxytoluene (BHT) (Guarulhos, Brazil, Biotec, 99.0% purity, P.A.) and *tert*-butylhydroquinone (TBHQ) (São Paulo, Brazil, Sigma-Aldrich, 99.0% purity, P.A.). The antioxidant masses were obtained with an analytical balance. The additions antioxidants in biodiesel were made in 1 g:100 g of biodiesel.

Relative protection factor

The relative protection factor (RPF) was determined by the ratio between the B100 biodiesel with antioxidant additives and the induction period of the control sample, multiplied by the antioxidant concentration.[35]

Kinetic parameter

To calculate the rate constant (k) at 110 ºC, it was determined the slope (equation 1) of the linear fit of the time (t / h) and the natural logarithm of the electrical conductivity (Λ), where Λ_0 is the initial conductivity.[9]

$$\ln \Lambda = \ln \Lambda_0 - kt \quad (1)$$

Artificial neural networks (ANN)

The physicochemical parameter values of the biodiesel samples, the concentration of antioxidants, the relative protection factor and rate constant at 110 ºC were processed by Statistica 13.4 software[36] applying the SOFM and MLP type artificial neural network modules. For the SOFM network, 100% of the samples were used for training, 12 × 12 topology, 7000 training cycles, learning rate ranging from 0.10 to 0.02, neighborhoods ranging from 3 to 0, and normal randomization of the network with a minimum mean equal to 0 and maximum variance equal to 1.

For the MLP network, the variables selected were antioxidant concentration in the proportion of 1 g:100 g of biodiesel, as the target continuous variable, the physicochemical parameters as the input continuous variables, and the antioxidants as the input categorical variable, where A = 1 for TBHQ, A = 2 for BHA and A = 3 for BHT. The random sampling selected 70% of data for training, 15% for testing, and 15% for validation.

Computational processing

All spreadsheets were processed on a computer with an Intel® Core TM i7-4790 CPU 3.60 GHz©, 32 GB of RAM, and 250 GB HDD.

Results and Discussion

The 12 samples of commercial biodiesel, without antioxidants, were submitted to density tests (D) in kg m^{-3}, flash point (FP) in °C, water content (W) in mg kg^{-1}, acid number (AN) in mg KOH g^{-1}, cloud point (CP) in °C, pour point (PP) in °C, kinematic viscosity at 40 °C (V) in mm^2 s^{-1} and oxidative stability with the determination of the control induction period (IPc) in h, and rate constant (k_c). For each 100 g of biodiesel, 0.0087 to 0.1028 g of TBHQ, 0.0650 to 0.1576 g of BHA, and 0.0433 to 0.1576 g of BHT were added separately. After the addition of the synthetic antioxidants to biodiesel, induction periods (IP), relative protection factor (RPF) and rate constant (k) were determined. The values obtained and the concentration of the antioxidants added (C) to perform each test are shown in Tables 1, 2, and 3.

The compliance parameters established by ANP standard No. 45[37] and 798,[38] are a minimum induction period of 12 h, density ranging from 850 to 900 kg m^{-3}, a minimum flash point of 100 °C, a maximum acid number of 0.5 mg KOH g^{-1}, a maximum water content of 350 mg kg^{-1}, viscosity ranging from 3.0 to 6.0 mm^2 s^{-1}, and cloud point and pour point in °C (minimum values for these parameters

are not defined in the Brazilian legislation). According to the data presented in Tables 1, 2, and 3, all parameters, except the induction period and water content, are according to the Brazilian legislation.

In Tables 1, 2, and 3, it is possible to observe that all antioxidants increased the biodiesel induction period, with TBHQ presenting greater influence, as it has one more hydroxyl group than BHA and BHT in its chemical structure. In addition, it has a higher relative protection factor in most experiments.

To evaluate the antioxidants applied, the parameters and values in Tables 1, 2, and 3, in that order, were grouped into a single spreadsheet and processed by cluster analysis and neural networks regression module of the Statistica 13.4 software.[36]

For both the cluster and regression modules, the seed

Table 1. TBHQ concentration and physicochemical parameters values of the biodiesel samples

C / (g 100 g^{-1})	RPF	IP / h	k / h^{-1}	D / (kg m^{-3})	FP / °C	W / (mg kg^{-1})	AN / (mg KOH g^{-1})	CP / °C	PP / °C	V / (mm^2 s^{-1})	k_c / h^{-1}	IPc / h
0.0599	70.99	15.32	0.15	887.20	118.60	305.40	0.34	6.00	2.00	4.79	0.91	3.60
0.0330	36.98	2.72	0.81	878.90	131.50	304.60	0.45	4.00	0.00	4.33	0.97	2.23
0.0340	35.66	2.70	0.89	878.90	131.50	304.60	0.45	4.00	0.00	4.33	0.97	2.23
0.0385	39.67	3.94	0.41	879.30	86.60	292.10	0.40	3.00	0.00	4.45	0.82	2.58
0.0340	39.46	4.34	0.63	878.50	141.50	260.90	0.45	3.00	0.00	4.50	0.78	3.24
0.0580	101.33	6.70	0.14	877.90	131.30	416.00	0.45	6.00	2.00	4.52	1.41	1.14
0.0369	27.25	2.20	0.89	877.20	129.50	353.60	0.23	7.00	3.00	4.55	1.38	2.19
0.1028	37.23	8.88	0.23	877.20	129.50	353.60	0.23	7.00	3.00	4.55	1.38	2.32
0.0534	23.67	2.93	1.50	877.20	129.50	353.60	0.23	7.00	3.00	4.55	1.38	2.32
0.0484	52.93	7.84	0.29	875.50	98.70	401.80	0.28	10.00	3.00	4.46	1.00	3.06
0.0503	35.91	4.01	0.67	876.30	109.50	390.53	0.30	9.00	4.00	4.57	1.38	2.22
0.0658	24.64	3.60	0.90	876.30	109.50	390.53	0.30	9.00	4.00	4.57	1.38	2.22
0.0222	60.97	8.04	0.34	876.20	110.00	310.00	0.29	6.00	2.00	4.44	0.44	5.94
0.0223	57.29	3.13	0.88	879.00	117.70	316.60	0.26	1.00	–1.00	4.22	1.06	2.45
0.0087	115.41	2.46	1.10	879.00	117.70	316.60	0.26	1.00	–1.00	4.22	1.06	2.45
0.0269	45.75	4.75	0.40	877.30	135.00	409.50	0.17	7.00	4.00	4.43	0.48	3.86
0.0455	30.58	5.37	0.45	877.30	135.00	409.50	0.17	7.00	4.00	4.43	0.48	3.86
0.0194	63.73	3.67	0.81	876.60	136.40	348.30	0.22	7.00	2.00	4.70	1.07	2.97
0.0427	47.81	6.06	0.43	876.60	136.40	348.30	0.22	7.00	2.00	4.70	1.07	2.97

C: concentration of the antioxidants added; RPF: relative protection factor; IP: induction period; k: rate constant; D: density; FP: flash point; W: water content; AN: acid number; CP: cloud point; PP: pour point; V: kinematic viscosity; k_c: rate constant; IPc: control induction period.

Table 2. BHA concentration and physicochemical parameters values of the biodiesel samples

C / (g 100 g^{-1})	RPF	IP / h	k / h^{-1}	D / (kg m^{-3})	FP / °C	W / (mg kg^{-1})	AN / (mg KOH g^{-1})	CP / °C	PP / °C	V / (mm^2 s^{-1})	k_c / h^{-1}	IPc / h
0.0650	30.59	7.16	0.39	887.20	118.60	305.40	0.34	6.00	2.00	4.79	0.91	3.60
0.0690	24.84	3.82	0.52	878.90	131.50	304.60	0.45	4.00	0.00	4.33	0.97	2.23
0.0990	21.27	5.43	0.41	879.30	86.60	292.10	0.40	3.00	0.00	4.45	0.82	2.58
0.0670	23.73	5.15	0.50	878.50	141.50	260.90	0.45	3.00	0.00	4.50	0.78	3.24
0.1576	18.76	3.37	0.73	877.90	131.30	416.00	0.45	6.00	2.00	4.52	1.41	1.14
0.1044	18.58	4.50	0.62	877.20	129.50	353.60	0.23	7.00	3.00	4.55	1.38	2.32
0.1043	20.96	6.69	0.39	875.50	98.70	401.80	0.28	10.00	3.00	4.46	1.00	3.06
0.1477	16.06	7.26	0.37	875.50	98.70	401.80	0.28	10.00	3.00	4.46	1.00	3.06
0.1492	18.36	6.08	0.41	876.30	109.50	390.53	0.30	9.00	4.00	4.57	1.38	2.22
0.1012	17.00	3.82	0.80	876.30	109.50	390.53	0.30	9.00	4.00	4.57	1.38	2.22
0.0475	25.41	7.17	0.38	876.20	110.00	310.00	0.29	6.00	2.00	4.44	0.44	5.94
0.1515	22.93	8.51	0.37	879.00	117.70	316.60	0.26	1.00	–1.00	4.22	1.06	2.45
0.0981	29.04	6.98	0.46	879.00	117.70	316.60	0.26	1.00	–1.00	4.22	1.06	2.45
0.0700	21.92	5.92	0.44	877.30	135.00	409.50	0.17	7.00	4.00	4.43	0.48	3.86
0.1067	21.33	6.76	0.41	876.60	136.40	348.30	0.22	7.00	2.00	4.70	1.07	2.97
0.0789	24.35	5.71	0.47	876.60	136.40	348.30	0.22	7.00	2.00	4.70	1.07	2.97
0.0651	28.12	5.44	0.44	876.60	136.40	348.30	0.22	7.00	2.00	4.70	1.07	2.97

C: concentration of the antioxidants added; RPF: relative protection factor; IP: induction period; k: rate constant; D: density; FP: flash point; W: water content; AN: acid number; CP: cloud point; PP: pour point; V: kinematic viscosity; k_c: rate constant; IPc: control induction period.

Table 3. BHT concentration and physicochemical parameters values of the biodiesel samples

C / (g 100 g^{-1})	RPF	IP / h	k / h^{-1}	D / (kg m^{-3})	FP / °C	W / (mg kg^{-1})	AN / (mg KOH g^{-1})	CP / °C	PP / °C	V / (mm^2 s^{-1})	k_c / h^{-1}	IPc / h
0.0796	25.55	7.32	0.37	887.20	118.60	305.40	0.34	6.00	2.00	4.79	0.91	3.60
0.0812	22.54	4.08	0.57	878.90	131.50	304.60	0.45	4.00	0.00	4.33	0.97	2.23
0.1189	17.70	5.43	0.36	879.30	86.60	292.10	0.40	3.00	0.00	4.45	0.82	2.58
0.0990	18.17	4.64	0.40	879.30	86.60	292.10	0.40	3.00	0.00	4.45	0.82	2.58
0.0772	18.79	4.70	0.59	878.50	141.50	260.90	0.45	3.00	0.00	4.50	0.78	3.24
0.1521	15.69	2.72	0.97	877.90	131.30	416.00	0.45	6.00	2.00	4.52	1.41	1.14
0.1020	14.41	3.41	0.85	877.20	129.50	353.60	0.23	7.00	3.00	4.55	1.38	2.32
0.1007	19.15	5.90	0.43	875.50	98.70	401.80	0.28	10.00	3.00	4.46	1.00	3.06
0.1540	15.51	7.31	0.30	875.50	98.70	401.80	0.28	10.00	3.00	4.46	1.00	3.06
0.1475	15.82	5.18	0.54	876.30	109.50	390.53	0.30	9.00	4.00	4.57	1.38	2.22
0.1038	21.22	4.89	0.57	876.30	109.50	390.53	0.30	9.00	4.00	4.57	1.38	2.22
0.0433	27.33	7.03	0.36	876.20	110.00	310.00	0.29	6.00	2.00	4.44	0.44	5.94
0.1576	21.55	8.32	0.35	879.00	117.70	316.60	0.26	1.00	–1.00	4.22	1.06	2.45
0.1310	22.00	7.06	0.35	879.00	117.70	316.60	0.26	1.00	–1.00	4.22	1.06	2.45
0.1076	25.87	6.82	0.43	879.00	117.70	316.60	0.26	1.00	–1.00	4.22	1.06	2.45
0.0697	18.58	5.00	0.44	877.30	135.00	409.50	0.17	7.00	4.00	4.43	0.48	3.86
0.1035	14.89	5.95	0.40	877.30	135.00	409.50	0.17	7.00	4.00	4.43	0.48	3.86
0.1056	17.66	5.54	0.45	876.60	136.40	348.30	0.22	7.00	2.00	4.70	1.07	2.97
0.1247	15.52	5.75	0.37	876.60	136.40	348.30	0.22	7.00	2.00	4.70	1.07	2.97

C: concentration of the antioxidants added; RPF: relative protection factor; IP: induction period; k: rate constant; D: density; FP: flash point; W: water content; AN: acid number; CP: cloud point; PP: pour point; V: kinematic viscosity; k_c: rate constant; IPc: control induction period.

for sampling equal to 1000 was selected. The choice of the random number seed generator is important in data analysis since it is related to the sequence of numbers provided by the program. When the same value is defined, the same sequence of random numbers will always be generated, which allows the reproduction of data whenever the program is initialized. This value should be changed if it is desirable to obtain a sampling of different data.[36]

In the cluster module, the concentration of antioxidants, in 1 g:100 g of biodiesel, and the other parameters were considered as continuous input variables, and the synthetic antioxidants used were defined as the categorical input variable, with A = 1 for TBHQ, A = 2 for the BHA and A = 3 for the BHT.

For data analysis by SOFM, the neighborhood parameter started at 3 and decreased to 0, the learning rate started at 1 and decreased exponentially to 0.02, with an increasing number of training cycles.

The number of cycles is an important parameter for the good performance of an ANN, as it refers to the number of times the network analyzes the data provided. This number must be determined so that at the end of the training step the error will be stabilized. However, it must not be too large because it would imply a longer processing time.

The choice of the network topology is fundamental for its good performance because if it is too small for the database, the neighborhood relationship can be so close between all neurons that all samples will be classified in just one group, suggesting no difference between them. If the network topology is too large, many groups of specialized neurons will be formed, due to the many possibilities of the neighborhood relationships. In this case, there is a large data dispersion, which may impair the grouping of the samples.[39] The topology used was chosen according to the number of samples analyzed, and the one with the best result was 12 × 12.

The SOFM network used is part of an ANN group called networks based on competitive learning models in which the output neurons of the grid compete with each other to be activated, and are mainly used to visualize clusters.[40]

According to Link *et al.*,[24] once the network has been initialized, there are three essential processes involved in its formation: competition, cooperation, and synaptic adaptation. The basic principle of the SOFM map is that it represents high-dimensional input vectors in a lower-dimensional 2D topology in an unsupervised way. Network neurons calculate their respective Euclidean distance values, and these values provide the basis for competition between them.[24,40,41] The neuron with the shortest Euclidean distance is declared the winner, that is, the one that best meets the stimulus provided by the input vector.[42]

In a cluster-type neural network, neurons are connected by directed and weighted edges. During training, the input and output signals are adjusted by the generated weights. The adjustments are based on previous errors and learning rates, and the connection pattern between neurons is called the network topology.[43]

This type of network uses unsupervised training, so the network seeks to find similarities based only on the input patterns. The main objective of Kohonen's self-organizing maps is to group input data that are similar to each other, forming classes or groupings called clusters.[44]

For training the network (Figure 1), 7000 cycles were used, chosen by previous experiments, which proved to be an adequate amount since the error stabilized at 0.02 after 6000 cycles.

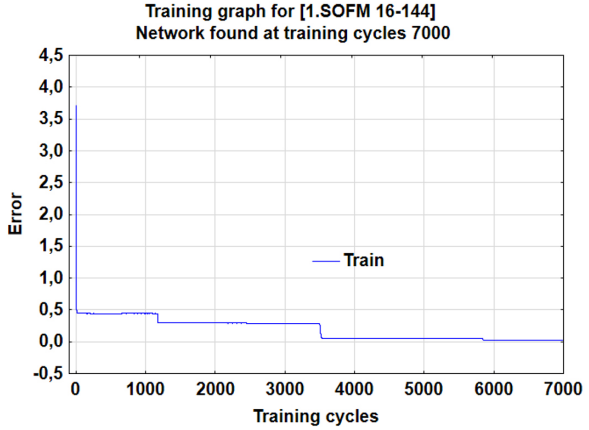

Figure 1. Error stabilization and the number of cycles used by the network SOFM.

The principle of training a network involves changing the weights of the winning neuron and its neighbors to make the weight vectors more similar to the input. This is done by gradually decreasing the learning rate and the neighborhood parameter at each iteration.[45]

Figure 2 shows the topological map provided by the network, the position of the antioxidants and frequency in the neurons, with values in parentheses and indicated by different colors. In the topological map, the definition of the groups is characterized by the presence of empty neurons between the clusters. If the topology is too small for the database, the neighborhood relationship between all neurons can be so close that the network can classify all the samples as a single cluster. If the network topology is too large, there is a large data dispersion, which can affect the clustering of the samples.[24,41,42]

In the map, it is possible to observe the formation of two well-defined groups of samples containing the antioxidants TBHQ and BHA and another containing BHT. On the right of the map is the TBHQ, on the upper left the BHA, and on the lower left the BHT. In the central part, we have neurons with BHT close to the TBHQ and also close to the BHA, therefore, they are more dispersed. However, we have more neurons containing BHT close to those with BHA, suggesting a similar behavior between these two antioxidants. Because we have empty neurons between the clusters, it can be concluded that the network was able to identify the differences in the behavior of the antioxidants used.

The BHA antioxidant located in the neuron (1,7), where the first number refers to the row and the second to the column in the map, presents only the wine color of the topological map with frequency 3. The three assays located in this neuron were performed with the same biodiesel sample and with similar BHA antioxidant concentrations.

In locations (9,5) and (7,2) the BHT samples presented frequency 2. They used different biodiesel for each neuron, but the same biodiesel for each test in the indicated positions. The same happens in positions (2,10), (4,10),

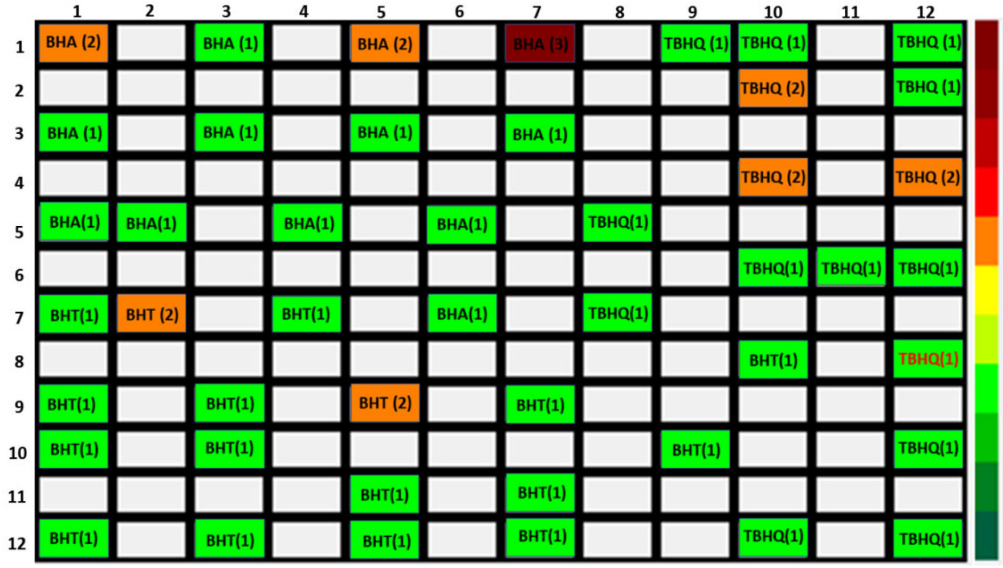

Figure 2. Distribution of samples containing antioxidant position, frequency, in parentheses, and the position of the winning neuron in red. The colors green, orange and wine indicate the frequencies in the neurons.

and (4,12), all containing TBHQ, with frequency 2, being close on the map because the IP of the samples is similar. In positions (1,1) and (1,5), with frequency 2, containing different BHA and biodiesel in both, even between the tests, they showed similar antioxidant concentrations.

In the central region of the map, positions (7,6), (9,7), and (7,8) are close, because they present similar antioxidant action, despite having different concentrations. The molecular structure of these compounds should be taken into account, as the antioxidant BHA was used in higher concentration, with 0.157 g:100 g, when compared to BHT with 0.101 g:100 g and TBHQ with 0.058 g:100 g.

The effectiveness of an antioxidant is related to its molecular structure, steric hindrance, and electronic behavior of the substituent. The phenolic antioxidants used have hydroxyl groups (–OH) that provide hydrogens that react with the free radicals formed in the biodiesel samples, delaying or decreasing the oxidation rate. TBHQ molecules have two –OH groups connected to the aromatic ring structure, while BHT and BHA have only one.[46,47]

To model the behavior of the antioxidants in biodiesel, the data from Tables 1, 2, and 3 were also grouped into a single spreadsheet and used in the regression module of the multilayer perceptron (MLP) artificial neural networks. In the worksheet, cases from 1 to 19 correspond to samples containing TBHQ, from 20 to 36 correspond to samples with BHA, and from 37 to 55 to samples with BHT. In the regression module, the concentration of antioxidants, in 1 g:100 g of biodiesel, was considered as the target continuous variable, the other parameters were considered as continuous input variables and the synthetic antioxidants were used as categorical input variables, with A = 1 for TBHQ, A = 2 for BHA and A = 3 for BHT.

5 to 20 hidden units were selected, and 200 networks were trained to solve the problem. The 5 best networks represent the relationship between the input and output variables, that is, those networks that reach the maximum correlation between the targets and the neural network outputs were chosen by the Statistica software.[36]

The Statistica MLP regression module specifies the number of hidden layers and the number of neuron units in each layer. For neurons of the hidden and output layers, the activation functions used were identity, logistic, tanh, and exponential. From the total number of samples chosen randomly, 70% were trained by the network, 15% for testing, and 15% for validation.[36]

The decay weight in the hidden layers and in the output layer varied between 10^{-4} and 10^{-3}, reducing the overfitting and improving the generalization efficiency of the network. Larger values can affect network performance in an unacceptable way.[36] Before the network initialization, the sum of the squares error (SOS) function was selected and the training algorithm used was the BFGS, individually proposed by Broyden-Fletcher-Goldfarb-Shanno.[36,41]

Choosing the number of hidden layers is important, as the greater the number the greater the non-linear mapping capacity. However, the excess of neurons in these layers leads to a reduction in the learning rate and over-fitting. The choice of 70% for training samples proved to be adequate because a larger number can cause an increase in processing time, reducing the efficiency of the network, and a small number can affect the fit and generalization ability of the network.[41,48]

The activation functions in a network transform the input signals of previous layer neurons into an output signal by a mathematical function, which is chosen before starting the regression module, and can influence the performance of the model obtained.[36,41,49]

To assess whether the number of cycles chosen was adequate, error graphs were obtained for the first and fourth networks of the five ones with the best performance. Figure 3 shows the number of cycles used to train and test the best performing 1.MLP 15-15-1 and the fourth best performing 4.MLP 15-20-1 networks. The first required 22 training cycles to reach stability and the latter 79. The error

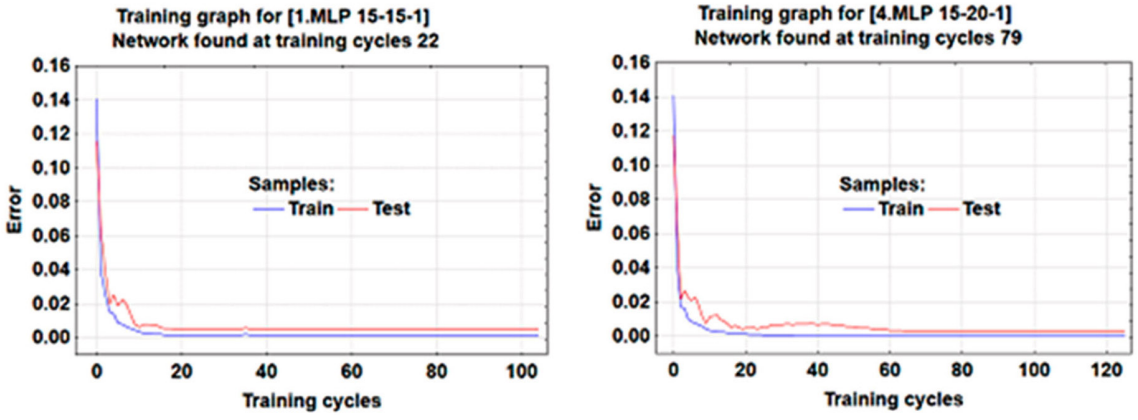

Figure 3. Error stabilization and the number of cycles used by the MLP network.

reduction, which represents the sum of squared differences between the target and output values (SOS) was fast, and small oscillations were observed only in the first training and testing cycles.

The global sensitive analysis, which provides information about the relative importance and evaluates the contribution of the input variables in the model construction, showed that the RPF contributed with 34.89%, the type of antioxidant with 31.49%, the induction period with 10.69%, the water content with 6.00% and the others all together with 16.93%. The variables order of importance was RPF > A > IP > W > k_c > CP > AN > D > PP > k > IPc > V > FP.

Tables 4, 5, and 6 show the cases used for training, testing, and validation, respectively, values of concentrations of synthetic antioxidants added to biodiesel (target) and estimated by the 5 best networks selected. In the networks column, the first number represents the 15 input variables, the second the number of hidden units and the last one is the number of outputs. In the last two columns are the mean values and standard deviations (StdD). The performance of the 5 most active networks ranged from 89 to 99% for training, from 90 to 99% for testing, and from 92 to 99% for validation. Furthermore, the error ranged from 1.00×10^{-6} to 2.25×10^{-4} for the training, from 6.00×10^{-6} to 4.47×10^{-4}

Table 4. Samples used for training, targeting, and predicted concentration of antioxidants (TBHQ, BHA and BHT), and statistical test employed

Sample	C (Target) / (g 100 g^{-1})	C (15-15-1) / (g 100 g^{-1})	C (15-5-1) / (g 100 g^{-1})	C (15-13-1) / (g 100 g^{-1})	C (15-20-1) / (g 100 g^{-1})	C (15-16-1) / (g 100 g^{-1})	Mean	StdD
1	0.060	0.059	0.032	0.061	0.060	0.049	0.052	0.012
2	0.033	0.033	0.037	0.037	0.033	0.052	0.038	0.008
3	0.034	0.034	0.037	0.038	0.034	0.052	0.039	0.007
4	0.039	0.038	0.037	0.037	0.038	0.052	0.041	0.007
5	0.034	0.035	0.032	0.033	0.034	0.048	0.036	0.007
6	0.058	0.057	0.043	0.051	0.058	0.061	0.054	0.007
7	0.037	0.037	0.040	0.040	0.037	0.058	0.043	0.009
8	0.103	0.103	0.043	0.112	0.103	0.061	0.084	0.030
9	0.053	0.054	0.039	0.048	0.053	0.057	0.050	0.007
13	0.022	0.022	0.029	0.024	0.022	0.046	0.029	0.010
14	0.022	0.021	0.038	0.026	0.022	0.053	0.032	0.013
16	0.027	0.026	0.032	0.027	0.027	0.049	0.032	0.010
17	0.046	0.046	0.033	0.040	0.046	0.050	0.043	0.007
19	0.043	0.044	0.036	0.036	0.043	0.053	0.042	0.007
20	0.065	0.064	0.071	0.062	0.065	0.081	0.069	0.008
21	0.069	0.068	0.097	0.074	0.069	0.093	0.080[a]	0.014
24	0.158	0.158	0.153	0.155	0.158	0.125	0.150	0.014
25	0.104	0.103	0.115	0.095	0.104	0.113	0.106	0.008
26	0.104	0.105	0.114	0.117	0.104	0.115	0.111[a]	0.006
27	0.148	0.148	0.118	0.132	0.148	0.117	0.133	0.015
30	0.048	0.048	0.059	0.042	0.047	0.072	0.054	0.012
31	0.152	0.148	0.127	0.138	0.152	0.105	0.134	0.019
32	0.098	0.100	0.119	0.106	0.098	0.102	0.105	0.008
33	0.070	0.071	0.073	0.073	0.070	0.088	0.075	0.008
34	0.107	0.109	0.096	0.108	0.107	0.102	0.104	0.005
35	0.079	0.078	0.092	0.086	0.078	0.100	0.087	0.009
37	0.080	0.080	0.077	0.080	0.080	0.090	0.081	0.005
41	0.077	0.078	0.078	0.075	0.077	0.084	0.078	0.003
42	0.152	0.151	0.161	0.155	0.152	0.128	0.150	0.012
43	0.102	0.103	0.120	0.093	0.102	0.118	0.107	0.011
45	0.154	0.154	0.128	0.142	0.154	0.122	0.140	0.015
46	0.148	0.149	0.141	0.139	0.147	0.127	0.141	0.009
47	0.104	0.103	0.136	0.128	0.104	0.126	0.119	0.015
48	0.043	0.043	0.062	0.042	0.043	0.078	0.054	0.016
50	0.131	0.133	0.134	0.135	0.131	0.112	0.129	0.010
52	0.070	0.070	0.077	0.077	0.070	0.095	0.078	0.011
53	0.104	0.104	0.080	0.100	0.103	0.097	0.097	0.010
54	0.106	0.108	0.102	0.109	0.106	0.108	0.107	0.003
55	0.125	0.122	0.104	0.117	0.124	0.109	0.115	0.009

[a]Hypothesis of homogeneity of variance not accepted. C: concentration of the antioxidants; StdD: standard deviations.

Table 5. Cases used for testing, targeting and predicted concentration of antioxidants (TBHQ, BHA and BHT) and statistical tests employed

Cases	C (Target) / (g 100 g^{-1})	C (15-15-1) / (g 100 g^{-1})	C (15-5-1) / (g 100 g^{-1})	C (15-13-1) / (g 100 g^{-1})	C (15-20-1) / (g 100 g^{-1})	C (15-16-1) / (g 100 g^{-1})	Mean	StdD
15	0.009	–0.008	0.033	0.022	0.008	0.050	0.021	0.022
18	0.019	0.011	0.033	0.024	0.017	0.051	0.027	0.016
28	0.149	0.165	0.133	0.139	0.151	0.123	0.142	0.016
36	0.065	0.061	0.091	0.076	0.061	0.099	0.078	0.017
39	0.119	0.101	0.110	0.100	0.116	0.104	0.106	0.007
40	0.099	0.083	0.108	0.088	0.103	0.103	0.097	0.011
44	0.101	0.083	0.122	0.119	0.098	0.120	0.108	0.017
49	0.158	0.162	0.137	0.146	0.151	0.113	0.142	0.019

C: concentration of the antioxidants; StdD: standard deviations.

Table 6. Cases used for validation, targeting, and predicted concentration of antioxidants (TBHQ, BHA and BHT), and statistical tests employed

Cases	C (Target) / (g 100 g^{-1})	C (15-15-1) / (g 100 g^{-1})	C (15-5-1) / (g 100 g^{-1})	C (15-13-1) / (g 100 g^{-1})	C (15-20-1) / (g 100 g^{-1})	C (15-16-1) / (g 100 g^{-1})	Mean	StdD
10	0.048	0.045	0.040	0.049	0.049	0.059	0.048	0.007
11	0.050	0.047	0.043	0.059	0.066	0.064	0.056	0.010
12	0.066	0.061	0.043	0.074	0.078	0.064	0.064	0.014
22	0.099	0.093	0.101	0.084	0.101	0.095	0.095	0.007
23	0.067	0.068	0.073	0.063	0.067	0.076	0.070	0.005
29	0.101	0.099	0.124	0.109	0.111	0.120	0.113	0.010
38	0.081	0.079	0.105	0.093	0.077	0.101	0.091	0.013
51	0.108	0.114	0.130	0.124	0.102	0.110	0.116	0.011

C: concentration of the antioxidants; StdD: standard deviations.

for the test and from 8.00×10^{-6} to 1.42×10^{-4} for validation, indicating the quality of the models represented by these networks.

The application of testing and validation (Tables 5 and 6) is recommended because the performance of a neural network is measured by how much it generalizes unseen data, that is, how well it predicts data that were not used during training. Like the test sample, a validation sample is not used to train the neural network but for additional verification of the models' performance. If the network performs consistently and well across testing and validation samples, then it is reasonable to assume that the network generalizes well unseen data.[36]

Since the Tukey's method is only valid if the variance is homogeneous, Levene's test was applied, which is a variance analysis of the absolute deviation values of the respective means of the group. The hypothesis of homogeneity was accepted because the *p*-values obtained were greater than 0.05, except in two cases that are presented in the table with an asterisk. The Tukey's test applied to the means did not show a significant difference at the 5% level in any of the cases, considering the values of the TBHQ, BHA, and BHT antioxidants concentrations added to the biodiesel and the average of the values obtained by the 5 best networks.[20,36]

Figure 4 shows the dispersion between the target and the output concentration values during the training of the two MLP-type networks that presented the best performance, which is a quality indication of the regression model used.

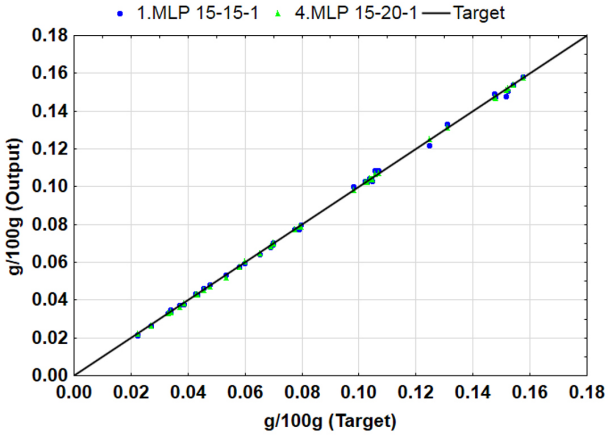

Figure 4. Dispersion graph between predicted values and target values obtained by the 1.MLP 15-15-1 and 4.MLP 15-20-1 networks.

Conclusions

In this work, the relationship between the biodiesel physicochemical parameters and the concentration of the synthetic antioxidants TBHQ, BHA and BHT added were studied, by applying SOFM (self-organizing feature maps) and MLP (multilayer perceptron) artificial neural networks type.

The SOFM neural network proved to be an adequate tool for the classification and recognition patterns of biodiesel samples containing TBHQ, BHA, and BHT.

The sensitive analysis showed that the relative protection factor and the type of antioxidant used were the variables that most influenced the construction of regression models using multilayer perceptron networks.

Statistical tests applied to validate the predictive model showed no significant difference between predicted and experimental values.

The application of mathematical modeling using SOFM and MLP neural networks to estimate the amount of antioxidant to be added to biodiesel can be interesting to improve manufacturing techniques, choice of antioxidant and help industries to produce biodiesel with induction period values within specifications.

Acknowledgments

The authors thank State University of Londrina (UEL), National Council for Scientific and Technological Development (CNPq), Coordination for the Improvement of Higher Education Personnel (CAPES), and Araucaria Foundation to Support Scientific and Technological Development of the State of Parana.

Author Contributions

Marco A. J. Clemente was responsible for conceptualization, data curation, formal analysis funding acquisition, validation, writing original draft; Heloisa H. P. Silva for visualization, data curation, formal analysis funding acquisition, investigation; Julia W. Campos for visualization, data curation, formal analysis funding acquisition, investigation; Nathan F. Silva for visualization, data curation, formal analysis funding acquisition, investigation; Eduardo G. Sousa for visualization, data curation, formal analysis funding acquisition, investigation; Ana C. G. Mantovani for validation, visualization, writing original draft, writing-review and editing; Karina B. Angilelli for project administration, resources, software, validation, visualization, writing original draft; Dionisio Borsato for investigation, project administration, resources, software, validation, visualization, writing original draft, writing-review and editing.

References

1. Cremonez, P. A.; Feroldi, M.; de Oliveira, C. J.; Teleken, J. G.; Meier, T. W.; Dieter, J.; Sampaio, S. C.; Borsatto, D.; *Ind. Crops Prod.* **2016**, *89*, 135. [Crossref]
2. Chendynski, L. T.; Angilelli, K. G.; Ferreira, B. A. D.; Borsato, D.; *Rev. Virtual Quim.* **2016**, *8*, 823. [Crossref]
3. Duda, K.; Wierzbicki, S.; Śmieja, M.; Mikulski, M.; *Fuel* **2018**, *212*, 202. [Crossref]
4. Simsek, S.; *Fuel* **2020**, *265*, 117026. [Crossref]
5. Viswanathan, K.; Taipabu, M. I.; Wu, W.; *Fuel* **2022**, *319*, 123660. [Crossref]
6. Ganesan, K.; Viswanathan, K.; Karthic, S. V.; Ekambaram, P.; Wu, W.; Vo, D. N.; *Fuel* **2022**, *319*, 123710. [Crossref]
7. Bhatt, A. N.; Shrivastava, N.; *Arch. Comput. Methods Eng.* **2022**, *29*, 897. [Crossref]
8. Shrivastava, N.; Khan, Z. M.; *Arch. Comput. Methods Eng*. **2018**, *25*, 707. [Crossref]
9. Borsato, D.; Dall'Antonia, L. H.; Guedes, C. L. B.; Maia, E. C. R.; de Freitas, H. R.; Moreira, I.; Spacino, K. R.; *Quim. Nova* **2010**, *33*, 1726. [Crossref]
10. Taipabu, M. I.; Viswanathan, K.; Wu, W.; Nagy, Z. K.; *Chem. Eng. J*. **2021**, *424*, 130199. [Crossref]
11. Figueredo, I. M.; Rios, M. A. S.; Cavalcante Jr., C. L.; Luna, F. M. T.; *Ind. Eng. Chem. Res.* **2020**, *59*, 18. [Crossref]
12. Nogales-Delgado, S.; Encinar, J. M.; González, J. F.; *Energies* **2019**, *12*, 1940. [Crossref]
13. Mantovani, A. C. G.; Chendynski, L. T.; Santana, V. T.; Borsato, D.; Di Mauro, E.; *Fuel* **2021**, *287*, 119531. [Crossref]
14. Longanesi, L.; Pereira, A. P.; Johnston, N.; Chuck, C. J.; *Biofuels, Bioprod. Biorefin.* **2022**, *16*, 265. [Crossref]
15. de Almeida, V. F.; García-Moreno, P. J.; Guadix, A.; Guadix, E. M.; *Fuel Process. Technol.* **2015**, *133*, 152. [Crossref]
16. Jeyakumar, N.; Narayanasamy, B.; Balasubramanian, D.; Viswanathan, K.; *Fuel* **2020**, *281*, 18614. [Crossref]
17. Silva, E. T.; Spacino, K. R.; Silva, L. R. C.; Romagnoli, É. S.; Angilelli, K. G.; Borsato, D.; *Acta Sci. Technol.* **2018**, *40*, 35108. [Crossref]
18. Chendynski, L. T.; Cordeiro, T.; Messias, G. B.; Mantovani, A. C. G.; Spacino, K. R.; Zeraik, M. L.; Borsato, D.; *Fuel* **2020**, *261*, 116379. [Crossref]
19. Rodrigues, J. S.; do Valle, C. P.; Uchoa, A. F. J.; Ramos, D. M.; da Ponte, F. A. F.; Rios, M. A. S.; Malveira, J. Q.; Ricardo, N. M. P. S.; *Renewable Energy* **2020**, *156*, 1100. [Crossref]
20. Clemente, M. A. J.; Silva, H. H. P.; Campos, J. W.; Mantovani, A. C. G.; Borsato, D.; *Quim. Nova* **2022**, *45*, 935. [Crossref]
21. Ramalingam, K.; Kandasamy, A.; Balasubramanian, D.; Palani, M.; Subramanian, T.; Varuvel, E. G.; Viswanathan, K.; *Environ. Sci. Pollut. Res*. **2020**, *27*, 24702. [Crossref]

22. Borsato, D.; Pina, M. V. R.; Spacino, K. R.; dos Santos Scholz, M. B.; Androcioli Filho, A.; *Eur. Food Res. Technol.* **2011**, *233*, 533. [Crossref]
23. Nobrega, M. M.; Bona, E.; Yamashita, F.; *Mater. Sci. Eng., C* **2013**, *33*, 4331. [Crossref]
24. Link, J. V.; Guimarães Lemes, A. L.; Marquetti, I.; dos Santos Scholz, M. B.; Bona, E.; *Food Res. Int.* **2014**, *59*, 1. [Crossref]
25. Mahmoudi, J.; Arjomand, M. A.; Rezaei, M.; Mohammadi, M. H.; *Civ. Eng. J.* **2016**, *2*, 1. [Crossref]
26. Sheela, K. G.; Deepa, S. N.; *Math. Probl. Eng.* **2013**, *2013*, ID 425740. [Crossref]
27. EN 14214: *Liquid Petroleum Products, Fatty Acid Methyl Esters (FAME) for Use in Diesel Engines and Heating Applications, Requirements and Test Methods*, European Committee for Standardization, 2012.
28. ASTM D4052-18a: *Standard Test Method for Density, Relative Density, and API Gravity of Liquids by Digital Density Meter*, ASTM International, West Conshohocken, PA, 2018. [Link] accessed in November 2022
29. ASTM D93-20: *Standard Test Methods for Flash Point by Pensky-Martens Closed Cup Tester*, ASTM International, West Conshohocken, PA, 2020. [Link] accessed in November 2022
30. ASTM D6304-20: *Standard Test Method for Determination of Water in Petroleum Products, Lubricating Oils, and Additives by Coulometric Karl Fischer Titration*, ASTM International, West Conshohocken, PA, 2020. [Link] accessed in November 2022
31. ASTM D664-11a: *Standard Test Method for Acid Number of Petroleum Products by Potentiometric Titration*, ASTM International, West Conshohocken, PA, 2018. [Link] accessed in November 2022
32. ASTM D445-06: *Standard Test Method for Kinematic Viscosity of Transparent and Opaque Liquids (and Calculation of Dynamic Viscosity)*, ASTM International, West Conshohocken, PA, 2006. [Link] accessed in November 2022
33. ASTM D2500-17a: *Standard Test Method for Cloud Point of Petroleum Products and Liquid Fuels*, ASTM International, West Conshohocken, PA, 2017. [Link] accessed in November 2022
34. EN 14112: *Fat and Oil Derivatives, Fatty Acid Methyl Esters (FAME), Determination of Oxidation Stability (Accelerated Oxidation Test)*; European Committee for Standardization: Brussels, 2003.
35. Spacino, K. R.; da Silva, E. T.; Angilelli, K. G.; Moreira, I.; Galão, O. F.; Borsato, D.; *Ind. Crops Prod.* **2016**, *80*, 109. [Crossref]
36. *Statistica Software for windows*, version 13.4; Tulsa, OK, USA, 2018.
37. Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP); Resolução ANP No. 45, de 25 Agosto de 2014, *Especificação do Biodiesel*; Diário Oficial da União: Brasília, 2014. [Link] accessed in November 2022
38. Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP); Resolução No. 798, de 1º de agosto de 2019; Altera a Resolução ANP No. 45, de 25 de agosto de 2014, que *Estabelece as Especificações deQualidade de Biodiesel, para Determinar a Obrigatoriedade da Aditivação do Biodiesel com Antioxidante e Estabelecer Novo Limite de Especificação da Característica Cstabilidade à Oxidação*; Diário Oficial da União (DOU), Brasília, No. 148, de 02/08/2019, p. 49. [Link] accessed in November 2022
39. Bona, E.; da Silva, R. S. S. F.; Borsato, D.; Bassoli, D. G.; *Acta Sci., Technol.* **2012**, *34*, 111. [Crossref]
40. Jagtap, S. S.; Shankar Sriram, V. S.; Subramaniyaswamy, V.; *Future Gener. Comput. Syst.* **2021**, *119*, 84. [Crossref]
41. Haykin, S.; *Redes Neurais: Princípios e Práticas*, 2nd ed.; Bookman: Porto Alegre, Brazil, 2001.
42. Kohonen, T.; *Neural Networks* **2013**, *37*, 52. [Crossref]
43. Köhnen, C. S.; Priesmann, J.; Nolting, L.; Kotzur, L.; Robinius, M.; Praktiknjo, A.; *Int. J. Energy Res.* **2022**, *46*, 4550. [Crossref]
44. Mawane, J.; Naji, A.; Ramdani, M.; *Indones. J. Electr. Eng. Comput. Sci.* **2022**, *26*, 482. [Crossref]
45. Draghici, S.; Potter, R. B.; *Bioinformatics* **2003**, *19*, 98. [Crossref]
46. Ryu, K.; *Bioresour. Technol* **2010**, *101*, S78. [Crossref]
47. Hossaina, M.; Sujan, S. M. A.; Jamal, M. S.; Chowdhury, M. S.; Techato, K.; Fakhruddin, A. N. M.; *BioResources* **2021**, *16*, 7021. [Crossref]
48. Wu, Q.; Ma, Z.; Xu, G.; Li, S.; Chen, D.; *IEEE Access* **2019**, *7*, 64686. [Crossref]
49. Bishop, C. M.; *Neural Networks for Pattern Recognition*; Oxford University Press: New York, 2007.

Submitted: July 23, 2022
Published online: November 8, 2022