World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:9, No:5, 2015

Application and Assessment of Artificial Neural Networks for Biodiesel Iodine Value Prediction

Raquel M. de Sousa, Sofiane Labidi, Allan Kardec D. Barros, Alex O. Barradas Filho, Aldalea L. B. Marques

***Abstract*—**Several parameters are established in order to measure biodiesel quality. One of them is the iodine value, which is an important parameter that measures the total unsaturation within a mixture of fatty acids. Limitation of unsaturated fatty acids is necessary since warming of higher quantity of these ones ends in either formation of deposits inside the motor or damage of lubricant. Determination of iodine value by official procedure tends to be very laborious, with high costs and toxicity of the reagents, this study uses artificial neural network (ANN) in order to predict the iodine value property as an alternative to these problems. The methodology of development of networks used 13 esters of fatty acids in the input with convergence algorithms of back propagation of back propagation type were optimized in order to get an architecture of prediction of iodine value. This study allowed us to demonstrate the neural networks' ability to learn the correlation between biodiesel quality properties, in this caseiodine value, and the molecular structures that make it up. The model developed in the study reached a correlation coefficient (R) of 0.99 for both network validation and network simulation, with Levenberg-Maquardt algorithm.

***Keywords*—**Artificial Neural Networks, Biodiesel, Iodine Value, Prediction.

I. Introduction

THE global search for clean energy alternatives, i.e., those generated from renewable resources that do not exacerbate the problems of greenhouse gases in the atmosphere, has made several countries initiate the search for alternatives and to conduct research directed to resource consumption sustainable [1]-[5]. Biodiesel, as a source of biomass derived energy, is able to contribute with sustainability to this world energy demand. Since it is an energy source biodegradable, renewable, non-toxic and environmentally friendly [3].

There are many properties used to identify if the biodiesel is a quality fuel. Quality control of biodiesel is important since it can be influenced by many factors, including the quality of the raw material, the composition of the vegetable oil or animal fat fatty acids, production process, the use of other materials and parameters subsequent production [6]. Several international norms establish an appropriate standart, as European Stardard (EN) 14214 in Europe, American Society for Testing and Materials (ASTM) D6751 in the USA [3]. In Brazil, resolution nº45 of the National Agency of Petroleum, Natural Gas and Biofuels (ANP) obligatorily controls the quality of biodiesel by employing the rules of the Brazilian Association of Technical Standards (ABNT), American Society for Testing and Materials (ASTM), the International Organization for Stardardization (ISO) and the Comité Européen Normalisation (CEN).

Biodiesel has several properties that can be related to the raw material pollutants, the storage conditions and, mainly, the molecular structures that make it up. The inherent properties to the composition, such as cetane number, density, viscosity and iodine value, are directly related to percentage of each alkyl ester that makes up the raw material (vegetable oil or animal fat) [2], [5], [7], [8].

The iodine value (IV) is a measure of total unsaturation within a mixture of fatty acids, based on the formal addition of iodine in the double unions and determined by g of iodine/100g of sample. Diesel's IV is virtually the same as the value of the oil used in its production [2], [6]. This value influences on several important properties of biodiesel, such as cold filter plugging point and oxidation stability. Oils and biodiesel with higher value of iodine tend to be less stable and more sensitive to oxidation due to the high degree of unsaturation [4], [7]. Use of IV is justified since this parameter indicates a tendency for both oil and fat to oxidation, in addition to indicating a tendency for both oil and fat to polymerization and formation of deposits inside the motor [6].

Determination of that property is carried out through official methods developed by the American Oil Chemists' Society (AOCS) and the Association of Analytical Communities, although these methods tend to be laborious, long and involve several steps, using various solvents and toxic chemical products [7], [9]. Several free and non-invasive methods were developed in order to determine or predict the iodine value [4], [7], [9], [10]. Adewale et al. [7] and Baptista et al. [4] determined the iodine value by means of near-infrared spectroscopy techniques (FT-NIR). Nevertheless, these techniques can take a long time and be uncomfortable for the company. ANNs are presented here as a non-linear alternative for predicting quality parameters, when it comes to lower cost and less time.

There are several researches in relation to quality control measures of fuel and biofuel, by using ANNs as non-linear tool to predict. The study [5] used the neural network to determine the cetane number, which is one of the indicators of biodiesel quality related to the property of the esters. Çay et al. [11] used neural networks to predict the performance of motor and the exhaust gas emission for methanol and gasoline. Saldana et al. [12] used learning methods of machine that include support vector machines and artificial neural network in order to predict the density and viscosity of biofuel compounds. Jahirul et al. [1] and X. Meng et al. [3] used

Raquel M. de Sousa is with the Federal University of Maranhão, Av. dos Portugueses, 1996, Campus do Bacanga, São Luis, Maranhão. 65080-805, Brazil (e-mail: rachel.msousanet@gmail.com).

ANNs to predict the kinematic viscosity over influence of temperature. Besides there are studies such as [13] that use ANNs to predict the emission and performance of biodiesel.

This work presents a method based an ANNs to predict the iodine value through training and testing of several algorithms of weight correction derived from backpropagation. The studyis organized as follows: In Section II, we show the experimental data that was used and the development methodology of ANNs in the work. In Section III, the results and discussions and, finally, in section IV we show the final remarks.

II. Method

A. Experimental Data

The data used for training, validation and testing of the network were collected from the literature [14]-[22] and the other one was ceded by the Laboratory of Analysis and Research on Petroleum Analytical Chemistry (LAPQAP/UFMA) that have related to the composition of samples and the physicochemical analysis. These data have 98 samples with 13 fatty acids esters of input, and in output they have the quality parameter iodine value. The distribution of each type of fatty acids ester in the biodiesel samples collected and the predominance of esters, such as oleic acid (C18:1), linoleic acid (C18:2), lauric acid (C12:0) and palmitic acid (C16:0) can be verified from Fig. 1. The description for the FAMEs in this study (XX: X) is about the information of the number of carbons (XX) and the number to the right represents the number of unsaturations of the molecule. Biodiesel samples were derived from various types of oils and oleogenosas, including, methyl and ethyl biodiesel from soybean and castor oil, palm oil, flaxseed, rapeseed, jatropha, moringa, babassu oil, poultry fat and algae.

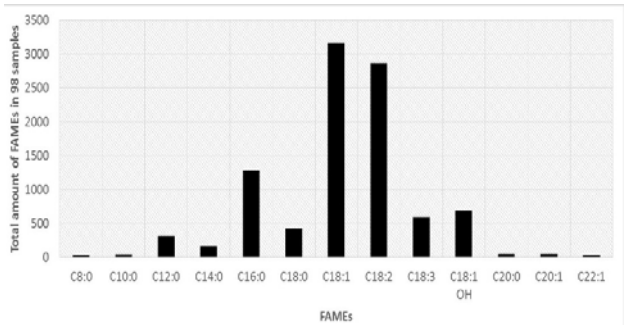

Fig. 1 Amount of predominant esters in the biodiesel samples

B. Methodology of the Development of Artificial Neural Networks (ANNs)

The ANNs used in this study are of the type of feedforward backpropagation, which is one of the most popular and used in various fields and it uses the supervised learning paradigm [3]. The networks were configured with two hidden layers with the neurons varied in each training from two to 25 neurons in each layer. Several algorithms can be used to optimize the ANNs' architectures, which enable the convergence of the results quickly and efficiently. Thus, seven algorithms for training, validation and simulation of the network were selected which are gradient descent pattern, gradient descent with momentum and adaptive learning rate, BFGS quasi-newton, scaled conjugate gradient, resilient propagation, one step secant and Levenberg Marquardt. We can observe from Fig. 2, a neural network feed forward back propagation, showed through the 13 inputs of fatty acids esters, two hidden layers: the first layer with seven neurons and the second hidden layer with five neurons. In addition, the quality property of iodine value as output. This was one of the combinations tested during networks' development process.

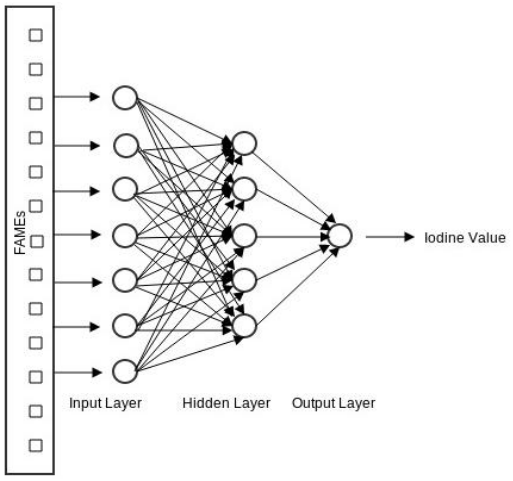

Fig. 2 Model of feed forward back propagation neural network with 13 inputs; two hidden layers containing 7 neurons in the first layer, 5 neurons in the second and a single output in the network

The selected algorithms are divided into two categories of training strategies, methods of first and second order. The methods of the first order are those where only the local gradient determines the direction of minimization, which are known as direction methods of greater decrease (steepest descent or gradient descent). The second-order methods are considered the most efficient way to do the training of neural networks of MLP type. These algorithms resort to a mathematical rigor based on well-defined models of unconstrained nonlinear optimization, thus not presenting a natural bond with the biological inspiration initially proposed for ANNs. We can quote Newton's method, which can be considered as the basic local method that uses second-order information, your application requires the calculation of the Hessian matrix [23]. By following the activation functions used in the present study were the logistic function and hyperbolic tangent function. Table I shows a brief summary of the scope of the study, the tested algorithms, the activation functions used, the stopping criteria and generalization of the network.

Altogether, 576 combinations of networks for each training algorithm were tested. The top ten network models of each training algorithm were saved through the k-fold cross-validation method [24]. In addition to the k-fold cross validation criterion with k = 11 as choice of the best network

among the candidates, the training technique with early stopping was in order to detect when over fitting start during the training of a network [25]. Over fitting occurs while the error in the training set decreases. However, the validation error begins to increase sometime. If the training does not stop at this minimum point, the network will essentially learn noise contained in the training data after this point [3], [26], [25]. The error calculation used to measure the network performance was the Mean Squared Error (MSE), which it interprets the difference between the value produced, indicated by $\hat{y}_i$, by the network and the desired value, y_i [3], given by:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 \quad (1)$$

Thus, k-10 parts were used for training and validation; and k = 11 used for testing. The inputs and outputs of data for training, validation and testing of the network were pre-processed between -1 and 1. The linear regression (R) and root mean square error (RMSE) are used in the network performance analysis. The R and RMSE value are an indication of the correlation between the outputs and targets of the network is given by:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (2)$$

$$RMSE = \frac{\sqrt{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}}{n} \quad (3)$$

TABLE I
THE SETTINGS OF THE TRAINED NETWORK, EACH VARIATION OF THE BACKPROPAGATION ALGORITHM WAS TESTED WITH ACTIVATION FUNCTIONS DISPLAYED, AND NETWORK ARCHITECTURE WITH TWO HIDDEN LAYERS OF NEURONS RANGING BETWEEN 2 AND 25

Training parameters	Value
Backpropagation algorithms	Gradient descent
	Gradient descent with momentum
	BFGS quasi-newton
	Scaled conjugate gradien
	Resilient Propagation
	One-Step-Secant
	Leverberg-Marquardt
Activation function	tansig function
	logsig function
	purelin function
Number hidden layers	2
Performance function	MSE
Criteria generalization	Cross validation K-fold
Maximum validation error	and early stopping
iterations	20
Number of neurons by layer	2 – 25

III. RESULTS AND DISCUSSIONS

In this study, the Matlab Version 2013b neural Netwoks Toolbox was used for the optimization process of ANNs. The experimental data of 98 samples with 13 esters of fatty acids in the input and the quality parameter of iodine value in the output were collected for training, validate and test of ANN method. The investigation the iodine value is very important to oxidation stability of biodiesel due to the number of unsaturation and because of warming of these unsaturated acids that can lead to formation of deposits or damage of lubricating oil [2], [6].

The ANNs were configured with two hidden layers and was set at two to 25 for each neuron layer of the neural network. The training and testing since studies were initially performed with ANN of only one hidden layer and low correlations were found, the predicted results for a layer were 0.76, 0.64, 0.91, for example, for the tested algorithms, while for networks with two hidden layers the coefficient reached 0.99. The decision taken to use two hidden layers in this work was justified by these results.

The coefficient of correlation results for choosing the early stopping criterion is reported in Table II. The use of cross validation by means of early stopping is an important step to avoid overfitting an underfitting. However, the choice of stopping point is an essential and critical criterion for network training. In this study, the definition of number of errors for early stopping was obtained through testing, with values of 6, 10, 15, 20 and 25. The Levenberg Marquardt was the convergence algorithm used for testing the values of errors, together with hyperbolic tangent activation function. Very small or very high values for validation with earlier stopping not allow the network to learn enough to generalize well so classifying incorrectly new data. As shown in Table II the results presented for error values equal to 6 in architecture network 8-15 with correlation coefficients of R = 0.98 for the training, R = 0.99 for validation and R = 0.35 for the testing. Already for error values equal to 25, the prediction results were also incorrect in the unknown patterns classification, we can mention the result of the network architecture 10-15 with R = 0.98 for training, R = 0.99 for validation and R = 0.54 for testing. By looking at the results of Table II, we can see that networks which generalized and obtained higher correlation coefficient for prediction were those having the value of 20 errors for early stopping, such as the network of architecture 12-5 with R = 0.97 for the training, R = 0.92 for the validation and R = 0.86 for the testing. Base on the results obtained with value of error equal to 20, this value was used with other algorithms in the following trainings.

From simulation of networks, correlation coefficient of prediction of higher value were showed by the algorithm Levenberg Marquardt, Resilient Propagation and One Step Secant, both of them with correlation of more than 0.95 for both validation and testing. The network architecture obtained with correlation coefficient of 0.99 for training, validation and testing was a 9-7 network with algorithm Levenberg Marquardt and function of logsig activation. Table III shows the results obtained of each algorithm that was trained with their respective correlation coefficients and RMSEs. Related works [5], [1], [3] already demonstrated the ability of the Levenberg Marquardt in obtaining satisfactory results in predicting other properties of biodiesel quality as viscosity and cetane number. In this study, we confirm that ability also for the iodine value.

TABLE II
THE RESULTS OF TESTS WITH EARLY STOPPING OF VALUE 6, 10, 15, 20 AND 25

Network	Value errors Early stopping	Training-R	Validation-R	Test-R
8–15	6	0.98	0.99	0.35
7–7	6	0.96	0.95	0.72
6–5	10	0.72	0.97	0.82
6–21	10	0.98	0.98	0.78
7–10	15	0.97	0.99	0.88
11–5	15	0.94	0.99	0.70
8–4	20	0.96	0.97	0.93
12–5	20	0.97	0.92	0.86
5–9	25	0.56	0.88	0.95
10–15	25	0.98	0.99	0.54

Results show the topology encountered and their corresponding results of correlation coefficient of training, validation and testing.

As expected, the networks were tested with only Gradient Descent had inferior results compared to other algorithms tested due to they became trapped in minimum locals, with R = 0.82, R = 0.93 and 0.94 respectively for training, validation and testing into the network architecture 9-4. However, the results obtained by the algorithms Resilient Propagation and One Step Secant remained equivalent. Moreover, although the time of the training execution was not measured in this study, it is necessary to highlight that some algorithms have more time of training due calculate the hessian matrix such as the algorithms of BFGS and Levenberg Marquardt [23].

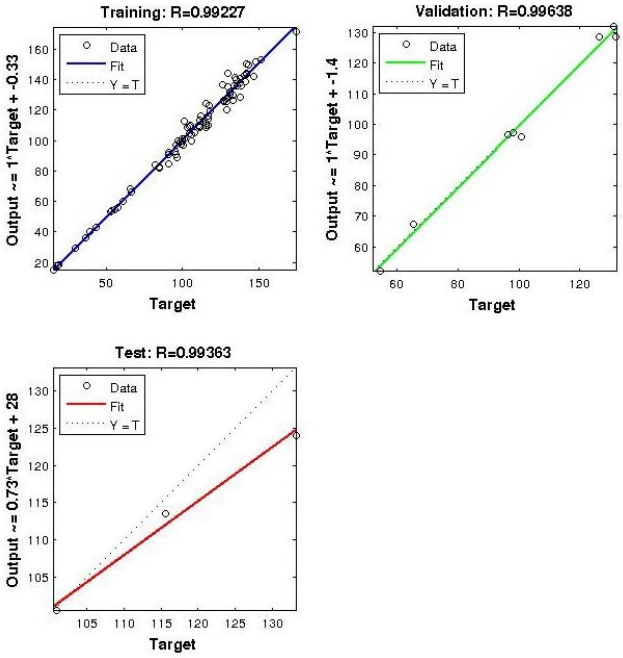

Fig. 3 Result of correlation coefficient of value 0.99, with Levenberg Marquardt and logistic function in the prediction of iodine value, in the phase of ANN's training, validation and testing

TABLE III
RESULTS OF THE CONVERGENCE ALGORITHMS

Network	Algorithms	Activation Function	Training - R	Training - RMSE	Validation - R	Validation - RMSE	Test - R	Test - RMSE
9–4	Gradient descent	tansig	0.82	20.75	0.93	25.07	0.94	28.36
17–6	Gradient descent with momentum	tansig	0.94	10.3	0.99	7.337	0.92	22.61
9–9	BFGS quasi-newton	logsig	0.95	10.08	0.99	10.1	0.90	12.67
24–2	Scaled conjugate gradient	tansig	0.96	11,54	0.98	10,65	0.92	17,46
23–19	Resilient Propagation	tansig	0.98	8.008	0.99	4.445	0.95	12.5
25–15	One Step Secant	tansig	0.94	9.966	0.98	15.43	0.96	24.98
9–7	Leverberg Marquardt	logsig	0.99	4.3035	0.99	2.677	0.99	2,5832

Table shows architectures with higher prediction index for each algorithm together with the activation function and the indices of prediction results for training, validation and testing.

In relation to other methods for predicting the iodine value, the ANNs showed a superior results demonstrated by correlation coefficients compared to the work of [7], where the correlation results reach 0.97 and 0.98 in the validation tests, while the method of neural networks reach R = 0.99 for both validation or test. Although other methods present similar results in the determination of iodine value [4], [9], as [4] which has prediction results also R = 0.99, using ANNs is much simpler and faster than these methods.

Fig. 3 shows the correlation coefficient of the value obtained and the target value of prediction, obtained by Levenberg Marquardt, both of them have a value of 0.99, demonstrating the network capacity to generalize and simulate new ester data of fatty acid and predict the iodine value.

IV. CONCLUSION

This study used 98 samples of ester fatty acids to predict the iodine value. Several networks were developed and tested, with several variations of backpropagation algorithm in order to get a better network architecture to predict the biodiesel quality property. This study researched the use of artificial neural networks to predict the quality parameter of iodine value in biodiesel. This is important since through the prediction of biodiesel's property and from the composition of esters of the raw material is possible to examine the viability of the use of such raw material for the synthesis of biodiesel within the quality parameters. In addition, some methods of analysis of biodiesel's properties can be long and/or complex. In this case, we achieved a better network architecture with the training algorithm Levenberg Marquardt with correlation index of 0.99 for both validation and testing. In this way, we can say that ANNs are a technology that can be used to predict the iodine value, as well as other properties related to the composition of esters of fatty acids.

REFERENCES

[1] M. I. Jahirul, W. Senadeera, P. Brooks, R. J. Brown, R. Situ, P. X.Pham, and A. R. Masri, "An Artificial Neutral Network (ANN) Model for Predicting Biodiesel Kinetic Viscosity as a Function of Temperature and Chemical Compositions," no. December, pp. 1–6, 2013.

[2] M. J. Ramos, C. M. Fernandez, A. Casas, L. Rodr´ıguez, and A. Pe´rez, "Influence of fatty acid composition of raw materials on biodiesel properties." Bioresource technology, vol. 100, no. 1, pp. 261–8, Jan.2009.

[3] X. Meng, M. Jia, and T. Wang, "Neural network prediction of biodiesel kinematic viscosity at 313K," Fuel, vol. 121, pp. 133–140, Apr. 2014.

[4] P. Baptista, P. Felizardo, J. C. Menezes, and M. J. Neiva Correia, "Multivariate near infrared spectroscopy models for predicting the iodine value, CFPP, kinematic viscosity at 40 degrees C and density at 15 degrees C of biodiesel." Talanta, vol. 77, no. 1, pp. 144–51, Oct. 2008.

[5] R. Piloto-Rodr´ıguez, Y. Sa´nchez-Borroto, M. Lapuerta, L. Goyos-Pe´rez, and S. Verhelst, "Prediction of the cetane number of biodiesel using artificial neural networks and multiple linear regression," Energy Con- version and Management, vol. 65, pp. 255–261, jan 2013.

[6] G. Knothe, J. Krahl, and J. V. Gerpen, Biodiesel Handbook (2nd Edition). AOCS Press, nov 2010.

[7] P. Adewale, O. Mba, M.-J. Dumont, M. Ngadi, and R. Cocciardi, "Determination of the iodine value and the free fatty acid content of waste animal fat blends using FT-NIR," Vibrational Spectroscopy, vol. 72, no. 0, pp. 72–78, 2014.

[8] I. P. Lobo, S. L. C. Ferreira, and R. S. d. Cruz, "Biodiesel: qualityparameters and analytical methods," Qu´ımica Nova, vol. 32, pp. 1596–1608, 2009.

[9] M. Orom´ı-Farru´s, G. Villorbina, J. Eras, F. Gatius, M. Torres, and R. Canela, "Determination of the iodine value of biodiesel using 1H NMR with 1,4-dioxane as an internal standard," Fuel, vol. 89, no. 11, pp. 3489–3492, nov 2010.

[10] M. Insausti, C. Romano, M. F. Pistonesi, and B. S. F. Band, "Simultaneous determination of quality parameters in biodiesel/diesel blends using synchronous fluorescence and multivariate analysis," Microchemical Journal, vol. 108, pp. 32–37, may 2013.

[11] Y. Cay, I. Korkmaz, A. C¸ ic¸ek, and F. Kara, "Prediction of engine performance and exhaust emissions for gasoline and methanol using artificial neural network," Energy, vol. 50, pp. 177–186, Feb. 2013.

[12] D. A. Saldana, L. Starck, P. Mougin, B. Rousseau, N. Ferrando, and B. Creton, "Prediction of Density and Viscosity of Biofuel Compounds Using Machine Learning Methods," Energy & Fuels, vol. 26, no. 4, pp.2416–2426, Apr. 2012.

[13] H. Sharon, R. Jayaprakash, M. Karthigai selvan, D. Soban kumar, a. Sun- daresan, and K. Karuppasamy, "Biodiesel production and prediction of engine performance using SIMULINK model of trained neural network," Fuel, vol. 99, pp. 197–203, sep 2012.

[14] J. d. J. Santos, "Biodiesel of Babassu: Thermal Verification, Oxidative and Binary Blends." Ph.D. dissertation, D. Sc. Thesis. Federal University of Paraíba, Brazil, 2008.

[15] M. B. Dantas, "Biodiesel blends: flow properties, thermal and oxidative stability during storage and monitoring," Ph.D. dissertation, Doctoral dissertation, D. Sc. Thesis. Federal University of Para´ıba. Brazil. Pag.138, 2006.

[16] R. A. Candeia, "Soybean Biodiesel: Synthesis, Degradation and Binary Mixtures," Ph.D. dissertation, Doctoral dissertation, D. Sc. Thesis. Federal University of Para´ıba. Brazil. Pag. 138, 2008.

[17] S. K. Padhi, "Preparation and Characterization of Biodiesel from Non-Edible Oils," Ph.D. dissertation, Doctoral dissertation, PhD thesis. National Institute of Technology, Rourkela, 2010.

[18] S. D. Sanford, J. M. White, P. S. Shah, C. Wee, M. A. Valverde, G. R. Meier, R. E. Group, and C. Report, "Feedstock and Biodiesel Characteristics Report," pp. 1–136, 2009.

[19] B. R. Moser, "Biodiesel production, properties, and feedstocks," In Vitro Cellular & Developmental Biology - Plant, vol. 45, no. 3, pp. 229–266, Mar. 2009.

[20] S. K. Hoekman, A. Broch, C. Robbins, E. Ceniceros, and M. Natarajan, "Review of biodiesel composition, properties, and specifications," Renewable and Sustainable Energy Reviews, vol. 16, no. 1, pp. 143–169, Jan. 2012.

[21] T. T. Kivevele, M. M. Mbarawa, a. Bereczky, T. Laza, and J. Madarasz, "Impact of antioxidant additives on the oxidation stability of biodiesel produced from Croton Megalocarpus oil," Fuel Processing Technology, vol. 92, no. 6, pp. 1244–1248, 2011.

[22] C. A. Nogueira, "Densities and Viscosities of Binary Mixtures of Babassu Biodiesel Cotton Seed or Soybean Biodiesel at Different Temperatures," vol. 55, no. 11, pp. 5305–5310, 2010.

[23] L. N. d. C. Silva, "Analysis and synthesis of learning strategies for artificial neural networks," Ph.D. dissertation, Doctoral dissertation, PhD thesis. State University of Campinas. Brazil. Pag. 210, 1998.

[24] R. Kohavi et al., "A study of cross-validation and bootstrap for accuracy estimation and model selection," in IJCAI, vol. 14, no. 2, 1995, pp.1137–1145.

[25] L. Prechelt, "Automatic early stopping using cross validation: quantifying the criteria," Neural Networks, vol. 11, no. 4, pp. 761–767, 1998.

[26] S. Haykin, Neural Networks: A Comprehensive Foundation, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.