Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

Computational Intelligence Applied to the Mathematical Modeling of the Esterification of Fatty Acids with Sugars

Lorenzo G. Tonetti[a], Ruy de Sousa Jr.[a*]

[a] Federal University of São Carlos, Department of Chemical Engineering, São Carlos-SP, Brazil
* Corresponding Author: ruy@ufscar.br.

ABSTRACT

The mathematical modeling of enzymatic reactors for esterification of fatty acids with sugars in the production of biosurfactants has been a useful tool for studying and optimizing the process. In particular, artificial neural networks and fuzzy systems emerge as promising methods for developing models for those processes. In this work, regarding artificial neural networks application, coupling of networks to reactor mass balances was considered in hybrid models to infer reactant concentrations over time. Computationally, an algorithm was constructed incorporating material balances, neural reaction rates, and step-by-step numerical integration (employing the classical Runge-Kutta method). Besides, based on an available set of experimental data, fuzzy logic was applied for modeling and optimization of the conversion of esterification as a function of operational process parameters (such as time, temperature and molar ratio of substrates). All computational development was carried out using the Matlab software. The neural networks were able to predict the kinetic behavior of xylose esterification process by applying them to reactor mass balances, obtaining R^2 values above 0.94, indicating a good predictive capacity. The trained fuzzy models were able to simulate the relationships between input variables and the output variable, enabling the construction of surfaces and estimating the optimal operational condition at 60 h of reaction, 55°C, molar ratio of substrates of 5:1 and enzyme loading of 37.5 U/g. The same condition was obtained when applying the particle swarm optimization algorithm. Thus, this work presented a complete tool based on computational intelligence for modeling, simulation, and optimization of biosurfactant synthesis.

Keywords: Biosurfactants, Artificial neural networks, Fuzzy modeling.

INTRODUCTION

Given the need for more sustainable alternatives to products available in the market, biosurfactants emerge as a promising option to synthetic surfactants [1]. Like as to synthetic surfactants, biosurfactants can form micelles and reduce surface and interfacial tensions, making them suitable for commercial applications [2]. Biosurfactants stand out to synthetic surfactants due to their biodegradability, low toxicity, use of substrates from renewable sources and waste from various production processes, which can reduce manufacturing costs, making them more attractive to industries [3,4]. Furthermore, biosurfactants are more efficient as they further reduce surface tension compared to chemical surfactants, requiring a lower concentration for maximum surface tension reduction, in a way that can be used in commercial products of the food, beverage, and cosmetics industries [5,6].

The esterification of fatty acids with carbohydrates has a significant commercial interest [7], because it yields sugar fatty acid esters, which find extensive applications in the food, pharmaceutical, cosmetic, and petrochemical industries [8]. This reaction can be carried out through either chemical or enzymatic routes, with the chemical route requiring high temperatures and an alkaline catalyst, potentially leading to the formation of toxic by-products [9]. In contrast, the enzymatic route offers mild reaction conditions, along with high selectivity, low toxicity, and high product quality and purity [10]. The mathematical modeling of enzymatic reactors for the esterification of fatty acids with sugars in the production of

biosurfactants can be a useful tool for studying and optimizing the process. Particularly, artificial neural networks and fuzzy systems emerge as promising methods for developing models for the mentioned process.

Artificial Neural Networks (ANNs) are a powerful technique for problem-solving in chemical engineering, capable of modeling nonlinear systems with a relatively simple neural network structure [11,12]. The use of ANNs in the mathematical modeling of enzymatic esterification reactors has been extensively studied by various authors, aiming to optimize the production of biodiesel, esters of aromatic carboxylic acids, and biosurfactants [13-15]. ANNs also may be used to enhance productivity, quality, and consistency, having applications in medical chemistry, hydrology, engineering, and pharmaceutical research [16,17].

The Adaptive Neuro-Fuzzy Inference System (ANFIS) is a technique that combines the power of neural networks and fuzzy inference systems, providing a method for the fuzzy modeling procedure to learn information on a data set [18,19]. ANFIS has been presented as a powerful tool for optimizing the reaction parameters of enzymatic reactors to produce biosurfactants and biodiesel, and other applications [20-22].

The production of biosurfactants through the esterification of fatty acids (oleic and lauric) and carbohydrates (fructose, lactose and xylose) in the presence of solvents such as tert-butanol and 2-methyl-2-butanol, using the commercial immobilized lipase B from *Candida antarctica* and with silica magnetic microparticles (SMMPs) as biocatalysts, has been experimentally evaluated by Lima et al. [23,24]. Torres [14], in turn, performed the kinetic modeling of this process through three distinct approaches: semi-mechanistic phenomenological modeling, the use of a model based on the Ping Pong Bi Bi mechanism; by using artificial neural networks in an empirical kinetic modeling; and the initial development of a fuzzy model.

In this context, the present work aims to contribute to the mathematical modeling of biosurfactant production by developing a hybrid-neural model capable of calculating xylose concentration values over time. This involves using neural networks to directly infer reaction rates. More specifically, hybrid-neural models for inferring the concentrations of reactants over time were developed, employing neural networks to directly provide reaction rates, being the kinetics coupled with reactor mass balances. Additionally, to complement the fuzzy model developed by Torres et al. [25] (also described in Torres [14]), data was incorporated from the factorial design by Lima et al. [23], allowing a more comprehensive understanding of the effect of input variables on the system output. This way, fuzzy logic was employed to create a model aimed at predicting optimal operational process conditions (such as time, temperature, molar ratio of substrates, etc.), to maximize product conversion.

MATERIAL AND METHODS

Application of Neural Networks to Infer Reaction Rates

For the development of artificial neural networks, data provided by Lima et al. [23] were employed, pertaining to the kinetics of xylose ester synthesis obtained through the esterification of oleic or lauric acid in tert-butyl alcohol medium. The experimental conditions included 300 rpm agitation, 1g of molecular sieve, a fatty acid/xylose molar ratio (SMR) of 5:1 (or 1:0.2), an enzyme activity load of 37.5 U/g of acid, and an initial xylose concentration of 20mM. Kinetic data were collected from the beginning of the reaction up till 72 hours, conducting experiments with both fatty acids at temperatures of 46ºC and 55ºC. The biocatalysts employed were commercial lipase B from *Candida antarctica* immobilized (CALB-IM-T2-350) and derivatives of CALB immobilized on silica magnetic microparticles (SMMPs) with octyl groups (CALB-SMMP-octyl) or with octyl and glutaraldehyde groups (CALB-SMMP-octyl-GLU).

To construct the neural network for the hybrid-neural model application, concentrations of xylose (mM) and reaction temperature were used. The neural networks were grouped based on the type of fatty acid and biocatalyst. The target response variable used in training the networks was the sugar consumption rate ($mmol.L^{-1}.h^{-1}$), calculated using the numerical derivative method from experimental data on concentration and time. Table 1 presents the sets of neural networks trained with experimental data for both temperatures, 46ºC and 55ºC.

Table 1: Specification of each trained ANN set.

Network	Biocatalyst	Fatty Acid
ANN1-3	CALB-IM-T2-350	Oleic
ANN2-1	CALB-IM-T2-350	Lauric
ANN3-1	CALB-SMMP-octyl	Lauric
ANN4-2	CALB-SMMP-octyl-GLU	Oleic
ANN5-2	CALB-SMMP-octyl-GLU	Lauric

To develop the hybrid-neural model, the coupling of neural networks to reactor mass balances was performed, aiming to infer concentrations of reactants over time. As the experiments were conducted in a batch reactor, Equation 1 represents the mass balance for this reactor type.

$$\frac{dX}{dt} = -\Gamma_X \quad (1)$$

Where Γ_X is the xylose consumption rate, X is the xylose concentration, and t is the reaction time. In the implementation of the hybrid-neural model, Γ_X is the output

of the neural network; therefore, is possible to obtain xylose concentration values over time by solving the differential equation. For this purpose, the classical Runge-Kutta method can be used for integration. The integration method was applied using the "ode15s" subroutine in the software Matlab.

The construction and training of neural networks were performed using the "feedforwardnet" function in the software Matlab. This function also allows data splitting for training, validation, and testing, with a distribution of 50%, 25%, and 25%, respectively, ensuring at least three data points for each training stage.

In this work, feed-forward neural networks with an input layer, a hidden layer, and one neuron in the output layer were employed. The training was conducted using the "trainlm" function, which implements the Levenberg-Marquardt method. The model's performance criteria included evaluating the determination coefficients for both the network training stage and the final response of the hybrid-neural model.

Development of Fuzzy Systems Modeling

For the development of the ANFIS modeling, kinetic data provided by Lima et al. [23] were employed for the xylose esterification with fatty acids (oleic or lauric) in tert-butyl alcohol medium. The experimental conditions included 300 rpm agitation, 1g of molecular sieve, a fatty acid/sugar molar ratio of 1:0.2, 20mM of xylose, and an enzyme load of 37.5 U/g of acid. Data collection spanned from the beginning of the reaction until 72 hours, conducting the experiment at temperatures of 46ºC and 55ºC for each type of biocatalyst used, namely CALB-IM-T2-350, CALB-SMMP-octyl, and CALB-SMMP-octyl-GLU.

In addition to the kinetic data provided by Lima et al. [23], data from an experimental design procedure were also utilized in this stage. The operational conditions are outlined in Table 2, with a reaction time of 48 hours. These data were used in conjunction with the kinetic data for constructing the fuzzy models.

Table 2: Operational Conditions of the factorial design [23].

Fatty acid	Oleic	Lauric	
Temperature(ºC)	40.6	46.0	51.30
Acid/sugar ratio (SMR)	0.60	1.00	1.39
Enzymatic Load (U/g)	30.10	37.50	44.90

The Neuro Fuzzy Designer tool in the software Matlab was employed for model construction, enabling the creation of the ANFIS model through graphical user interfaces. The datasets were separated based on the catalyst type used, resulting in three distinct models: ANFIS1 for CALB-IM-T2-350, ANFIS2 for CALB-SMMP-octyl, and ANFIS3 for CALB-SMMP-octyl-GLU. To enhance the fitting quality, additional data points were incorporated through interpolation in the kinetic data at the 36-hour mark.

The input data for ANFIS were organized in a file, with columns representing the values for fatty acid, temperature, SMR (substrate molar ratio), enzyme load, time, and xylose conversion, respectively. Each row represented one of the experimental data points. These data were loaded into the software, and in conjunction with the Neuro Fuzzy Designer tool, linguistic values of the model and the membership function were defined.

Optimization of reaction parameters through PSO

From the trained ANFIS networks, the particle swarm optimization (PSO) algorithm was employed to assess the optimized operational condition for the xylose esterification reaction in each network and for each fatty acid. The software Matlab, utilizing the PSO routine, can be utilized for this purpose, enabling the application of the PSO algorithm in the fuzzy network. Due to the routine's nature of seeking the global minimum, the ANFIS network output was defined as negative to pursue the global maximum of the network [26]. Import to say that other optimization methods could also be considered, such as genetic algorithms or simulated annealing. All are classical robust algorithms often used to solve global optimization problems.

RESULTS

Development of Modeling by Artificial Neural Networks

The neural networks were trained with the aim to inferring xylose consumption rate in the esterification reaction during biosurfactant synthesis, using concentration and temperature information as input for the network. During training, the network utilized experimental data to learn data behaviors, while validation was conducted to assess the network's generalization ability using some data, and testing was performed to evaluate the neural model's performance. The adopted criterion was the coefficient of determination (R^2), obtained by comparing the network output value with the experimental value of the output variable. For the network to be considered promising, the final R^2 values need to exceed 0.95.

The number of neurons in the hidden layer was determined following the same training criterion, evaluating the final behavior of the network in the hybrid model. There is a possibility that the network is promising with two different quantities of neurons, or they differ only due to the randomness of the training initialization. For these reasons, it is possible to obtain more than one network for each dataset. The neural networks ANN1-3, ANN3-1, ANN4-2 and ANN5-2 have three neurons in the

hidden layer, while the neural network ANN2-1 has two neurons. Overall, all networks have R^2 values above 0.98, due to the criterion adopted during training, with some exceeding 0.99.

The developed neural networks were coupled to reactor mass balances in a hybrid model to infer the concentrations of reactants over time, using the Runge-Kutta method for numerical integration. To achieve this, Equation 1 is applied in conjunction with these networks, and the differential equations are solved using the "ode15s" subroutine. Figure 1 illustrates the resulting curves of the hybrid modeling for the ANN2-1 neural network, where the determination coefficients of the model with the experimental data at temperatures of 46°C and 55°C are, respectively, 0.9741 and 0.9935. Overall, the results obtained from the hybrid modeling with direct neural kinetics fully satisfy the conditions imposed as performance criterion, where the value of R^2 should be greater than 0.95 (except for the RNA4-2, which presented a slightly lower but also good value of 0.9249 at 46°C). Thus, the application of mass balance together with neural networks has successfully constructed kinetic curves capable of inferring the concentration of xylose throughout the esterification reaction.

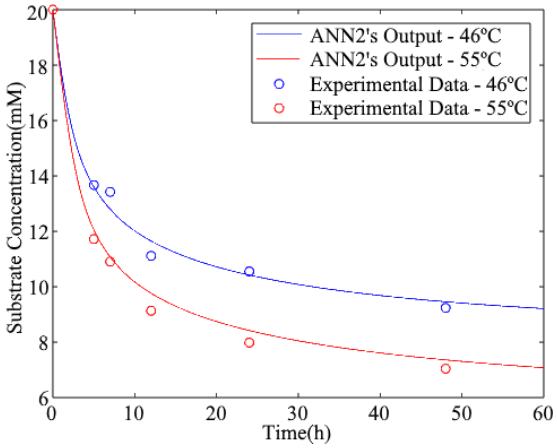

Figure 1: Response curves for the hybrid-neural model at temperatures of 46°C and 55°C, using the neural network ANN2-1.

Development of Fuzzy System Modeling

The application of fuzzy logic in the mathematical modeling of biosurfactant production through esterification was carried out using a dataset derived from a factorial design and another set of kinetic data. These datasets were classified based on the catalyst used, generating fuzzy networks for each distinct catalyst. With the trained ANFIS models, the ability to simulate experimental data was tested. For this purpose, a regression of the model-derived data with experimental data was performed. Coefficient of determination values close to or greater than 0.98 were obtained, indicating that the training was effective in preparing the network to accurately infer the experimental data. Figure 2 illustrates the kinetics resulting from the fuzzy model prediction at temperatures of 46°C and 55°C, with the lauric acid experimental data.

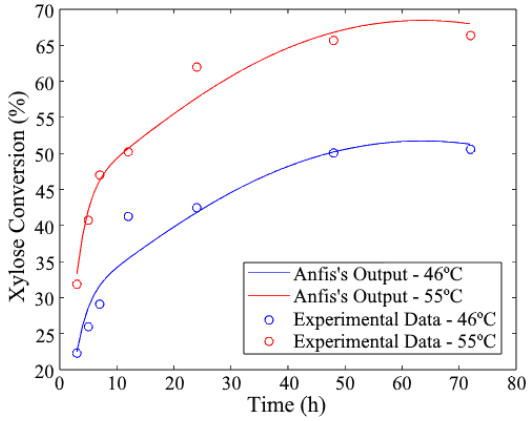

Figure 2: ANFIS modeling at temperatures of 46°C and 55°C from ANFIS3-1 network, for lauric acid.

Fuzzy modeling has a high ability to infer the output variable based on an input value within its domain. In this context, surface plots were developed from combinations of input variables, evaluating batch temperature (constant during each batch) combined with time (to terminate the batch), fatty acid/sugar ratio combined with enzyme load, and temperature combined with fatty acid/sugar ratio. Figure 3 illustrates these surfaces, allowing a comprehensive visualization of the fuzzy modeling. The problem had a pattern input of a fatty acid/sugar ratio of 5.0, temperature of 55°C, enzyme load of 37.50 UE/g, and a reaction time of 48 hours, when kept invariable (i.e., when some not being one of the variables whose effects are being analysed). The obtained results show the region with global maximum conversion under the following conditions: above 48 hours, temperature of 55°C, acid/sugar ratio of 1:0.2, and enzymatic load of 37.5 U/g, even with variations in linguistic values and the biocatalyst used. This is consistent with the results obtained by Torres [14]. Additionally, the local maximum appearing in regions of SMR values below 1:0.2 represent the optimal condition for the data obtained in the factorial design by Lima et al. [23]. One impact of including data from the experimental design is the possibility of also identifying this local maximum.

Application of particle swarm optimization in the ANFIS model

The particle swarm optimization algorithm was applied to the constructed fuzzy models with the aim of optimizing operational conditions for biosurfactant synthe-

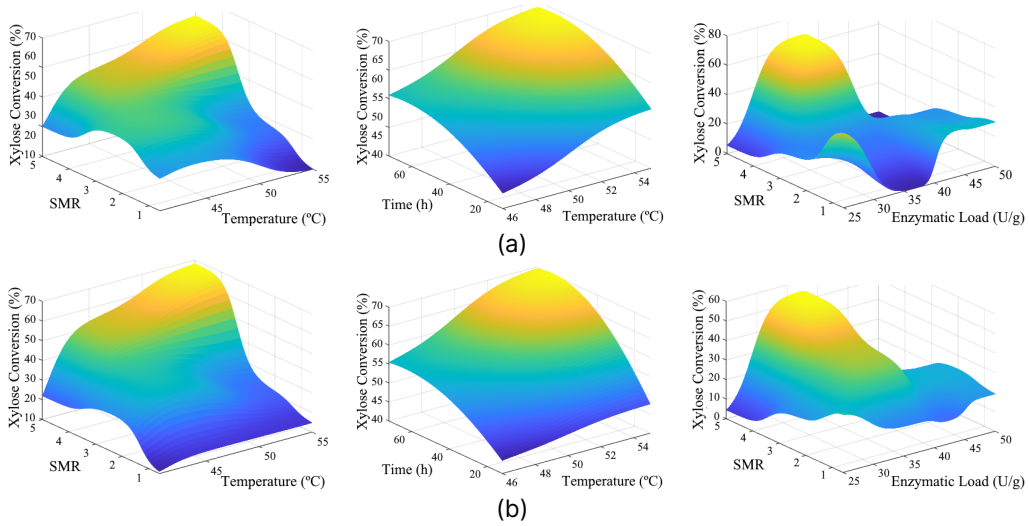

Figure 3: Surfaces developed by the ANFIS model for lauric acid (a) and oleic acid (b) from the ANFIS3-1.

sis. The optimization was carried out using the "particleswarm" command in software Matlab. Table 3 presents the optimized conditions obtained by the optimization algorithm for each of the developed models (the respective optimal temperature and SMR values for all networks are 55ºC and 5).

Table 3: Optimized conditions obtained by PSO.

Network	FA	Load (U/g)	t (h)	(%)
ANFIS1-2	Lauric	37.48785	64.931	69.19
ANFIS1-2	Oleic	37.48769	63.714	70.16
ANFIS2-1	Lauric	40.41127	65.509	74.12
ANFIS2-1	Oleic	40.41144	65.508	68.65
ANFIS3-1	Lauric	36.90347	63.572	68.58
ANFIS3-1	Oleic	36.90348	63.573	70.33

The global optimal conditions found by the particle swarm optimization algorithm show values close to the point with the highest conversion (temperature of 55ºC, SMR of 5, enzyme load of 37.5 U/g, and 72 hours of reaction).

CONCLUSION

The training of neural networks was effective in producing neural networks capable of inferring experimental data. When applying the hybrid-neural model, the neural networks were able to predict the kinetic behavior of the xylose esterification process in biosurfactant synthesis, applied to reactor mass balances.

The fuzzy models were able to simulate the relationships between input variables (temperature, time, SMR, and enzyme load) and the output variable (xylose conversion), allowing the construction of various response combinations. The application of the particle swarm optimization to fuzzy models demonstrated the ability of this optimization technique to estimate the optimized operational condition.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. Besides, this work was carried out with the support of the National Council for Scientific and Technological Development, CNPq.

REFERENCES

1. Nitschke M, Pastore GM. Biosurfactants: properties and applications. *Quím Nova*. 25:772–6 (2002)
2. Drakontis, CE. Amin, S. Biosurfactants: Formulations, properties, and applications. *COCIS* 48:77-90 (2020)
3. Kreling NE, Zaparoli M, Margarites AC, Zampieri D, Colla LM. Biosurfactant production: intracellular manoproteins and extracellular sophorolipids by *Saccharomyces cerevisiae*. *ABES* 24:1209-19 (2019)
4. Santos ECL, Miranda DAR, Silva ALS, López AMQ. Biosurfactant Production by Bacillus strains isolated from sugar cane mill wastewaters. *Braz arch biol technol* 67 (2024). https://doi.org/10.1590/1678-4324-2019170630
5. Cameotra SS, Makkar RS. Recent applications of biosurfactants as biological and immunological molecules. *COMICR 7*:262–6 (2004)
6. Oliveira EM, Sales VHG, Dias EDC, Andrade MS, Araújo RS, Borges WL, et al. . Biosurfactant by *Serratia* sp. BR13816: Fermentation Optimization and Nanoemulsion Formation. *Braz arch biol technol* 67 (2024). https://doi.org/10.1590/1678-4324-2024220225.

7. Vescovi V, Santos JBC, Tardioli, PW. Porcine pancreatic lipase hydrophobically adsorbed on octyl-silica: A robust biocatalyst for syntheses of xylose fatty acid esters. *Biocatal Biotrans* 35:298–305 (2016)
8. Torres ACL, Lima LN, Tardioli PW, Júnior RS. Mathematical modeling of enzymatic syntheses of biosurfactants catalyzed by immobilized lipases. *Reac Kinet Mech Cat* 130:699–712 (2020)
9. Yan Y, Bornscheuer, UT, Stadler, G. Lutz-Wahl S, Reuss M, Schmid, RD. Production of sugar fatty acid esters by enzymatic esterification in a stirred-tank membrane reactor: Optimization of parameters by response surface methodology. *J Amer Oil Chem Soc* 78:147–53 (2001)
10. An D, Zhang X, Liang F, Xian M, Feng D, Ye Z. Synthesis, surface properties of glucosyl esters from renewable materials for use as biosurfactants. *Colloids Surfaces A* 577:257–64 (2019)
11. Beigi M, Torki-Harchegani M, Mahmoodi-Eshkaftaki M. Prediction of paddy drying kinetics: A comparative study between mathematical and artificial neural network modelling. *Chem Ind & Chem Eng Q* 23:251–8 (2017). https://doi.org/10.2298/CICEQ160524039B
12. Eyng E, Silva FV, Palú F, Fileti AMF. Neural network based control of an absorption column in the process of bioethanol production. *Braz arch biol technol* 52:961–72 (2009). https://doi.org/10.1590/S1516-91320090004000020
13. Manohar B, Divakar S. An artificial neural network analysis of porcine pancreas lipase catalysed esterification of anthranilic acid with methanol. *Process Biochem* 403372–6 (2005)
14. Torres ACL. Modelagem matemática para sínteses enzimáticas de biossurfactantes catalisadas por lipases imobilizadas. São Carlos: Universidade Federal de São Carlos (2021)
15. Soltani S, Shojaei TR, Khanian N, Choong TSY, Asim N, Zhao Y. Artificial neural network method modeling of microwave-assisted esterification of PFAD over mesoporous TiO2—ZnO catalyst. *J Renew Energy* 87:760–73 (2022)
16. Singh I, Kaur J, Kaur S, Barik BR, Pahwa R. Artificial Neural Networks and Neuro-Fuzzy Models: Applications in Pharmaceutical Product Development. *Braz arch biol technol* 66 (2023). https://doi.org/10.1590/1678-4324-2023210769
17. Souza GR, Bello IP, Corrêa FV, Oliveira LFC. Artificial Neural Networks for Filling Missing Streamflow Data in Rio do Carmo Basin, Minas Gerais, Brazil. *Braz arch biol technol* 63 (2020). https://doi.org/10.1590/1678-4324-2020180522
18. Betiku E, Odude VO, Ishola NB, Bamimore A, Osunleke AS, Okeleye AA. Predictive capability evaluation of RSM, ANFIS and ANN: A case of reduction of high free fatty acid of palm kernel oil via esterification process. *Energy Convers Manag* 124:219–30 (2016)
19. Bressan GM, Azevedo BCF de, Souza RM de. A Fuzzy Approach for Diabetes Mellitus Type 2 Classification. *Braz arch biol technol* 63 (2020). https://doi.org/10.1590/1678-4324-2020180742
20. Santos B, Ponezi A, Fileti A. Development of Artificial Intelligence Models to Monitor Biosurfactant Concentration in Real-Time using Waste as Substrate in Bioreactor Through Fermentation by *Bacillus subtilis*. *Chem Eng Trans* 57:1009-14 (2017)
21. Ighose BO, Adeleke IA, Damos M, Junaid HA, Okpalaeke KE, Betiku E. Optimization of biodiesel production from Thevetia peruviana seed oil by adaptive neuro-fuzzy inference system coupled with genetic algorithm and response surface methodology. *Energy Convers Manag* 132:231–40 (2017)
22. Sadi M, Ahari SJ, Zarrinpashne S. Application of adaptive neuro fuzzy inference system to modeling oxidative coupling of methane reaction at elevated pressure. *J Pet Technol* 4: 43-55 (2014)
23. Lima LN, Vieira GNA, Kopp W, Tardioli PW, Giordano RLC. Mono- and heterofunctionalized silica magnetic microparticles (SMMPs) as new carriers for immobilization of lipases. *J Mol Catal B Enzym* 133:S491–99 (2016)
24. Lima LN, Mendes AA, Fernandez-Lafuente R, Tardioli PW, Giordano RLC. Performance of Different Immobilized Lipases in the Syntheses of Short- and Long-Chain Carboxylic Acid Esters by Esterification Reactions in Organic Media. *Molecules* 23:766 (2018)
25. Torres ACL, Akisue RA, Lima LN, Tardioli PW, Júnior RS. Computational intelligence applied to the mathematical modeling of enzymatic syntheses of biosurfactants. *Comput Aided Chem* Eng 51;139-44 (2022)
26. Kennedy J, Eberhart R. Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks* 4:1942–48 (1995)