

Article

Transesterification of Castor Oil into Biodiesel: Predictive Modeling with Machine Learning and Genetic Algorithm

Vivian Lima dos Santos , Luiz Carlos Lobato dos Santos *  and George Simonelli 

Oil, Gas, and Biofuels Research Laboratory (PGBio), Postgraduate Program in Chemical Engineering (PPEQ), Federal University of Bahia (UFBA), Rua Prof. Aristides Novis 02, Federação, Salvador 40210-630, BA, Brazil; vivianlima@ufba.br (V.L.d.S.); ggsimonelli@gmail.com (G.S.)

* Correspondence: lclsantos@ufba.br

Abstract

The growing demand for energy and the environmental impacts of fossil fuels have driven the search for sustainable alternatives such as biodiesel. Castor oil stands out as a promising non-edible feedstock but requires optimization strategies to overcome challenges in its conversion to biodiesel. This study developed a predictive model to determine the optimal parameters for homogeneous alkaline or acid transesterification of castor oil, aiming to maximize fatty acid methyl ester (FAME) yield. A dataset of 406 operating conditions from the literature was used to train and evaluate six models: Multilayer Perceptron with logistic sigmoid activation (MLP-logsig), hyperbolic tangent activation (MLP-tansig), Radial Basis Function network (RBF), hybrid RBF + MLP, Random Forest (RF), and Adaptive Neuro-Fuzzy Inference System (ANFIS). The MLP-tansig achieved the best performance in training, validation, and testing ($R > 0.98$). However, when combined with a Genetic Algorithm (GA), it generated infeasible parameters. Conversely, the RBF + GA combination yielded results consistent with the literature: molar ratio 19.35:1, alkaline catalyst 1.13% w/w , temperature 50 °C, reaction time 70 min, and stirring speed 548.32 rpm, achieving 100% FAME yield. This approach reduces the need for extensive experimental testing, offering a cost- and time-efficient solution for optimizing biodiesel production.

Keywords: biodiesel; castor oil; predictive modeling; genetic algorithm; artificial neural networks



Academic Editor: Jaehoon Kim

Received: 15 September 2025

Revised: 18 October 2025

Accepted: 30 October 2025

Published: 4 November 2025

Citation: Santos, V.L.d.; Santos, L.C.L.d.; Simonelli, G.

Transesterification of Castor Oil into Biodiesel: Predictive Modeling with Machine Learning and Genetic Algorithm. *Biomass* **2025**, *5*, 71. <https://doi.org/10.3390/biomass5040071>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Humanity is currently undergoing a period of increasing industrialization and technological advancement, resulting in a steadily growing energy demand. At present, this demand is still predominantly supplied by non-renewable sources such as fossil fuels [1]. However, this dependence poses significant risks. Petroleum exploration has become progressively more challenging. Onshore oil fields are declining at a rate of 3.1% per year, requiring investments in additional capacity and advanced technologies to access remote or complex reserves, such as offshore and mature fields [2]. This situation can negatively affect the market value of the product. Moreover, the combustion of fossil fuels releases pollutant gases such as nitrogen oxides and sulfur oxides, which are responsible for environmental issues including the greenhouse effect and acid rain. Human health is also severely affected by air pollution. Miller et al. (2024) published a concerning review reporting that air pollution accounts for 8.8 million premature deaths annually. Respiratory diseases are the most prevalent, including asthma, bronchitis, and bronchoconstriction, but

air pollution can also lead to cardiovascular problems, kidney damage, and neurological disorders, among others [3].

Given the problems associated with the combustion of fossil fuels, environmental legislation around the world has become increasingly strict. Following the publication of the new Global Air Quality Guidelines by the World Health Organization (WHO), the European Commission revised its Ambient Air Quality Directive, aiming to align European Union standards with WHO recommendations [4]. Canada follows the same trend, establishing in the Canadian Ambient Air Quality Standards Handbook (2025) the reduction in NO₂ (per hour) from 60 ppb in 2020 to 42 ppb in 2025, and in SO₂ from 70 ppb to 65 ppb [5]. In Latin America, Brazil stands out in the environmental agenda. In a 2018 resolution of the National Environmental Council, the country established a plan that foresees a progressive reduction in the permitted pollutant levels until 2044 [6]. It also presents programs to encourage energy transition, such as RenovaBio, which sets decarbonization targets and certification of biofuel production; and Resolution No. 16 of the National Energy Policy Council, which mandates the addition of 14% biodiesel to fossil diesel [7,8].

Initiatives that promote the transition to cleaner fuels, i.e., with lower emissions of toxic gases, contribute to mitigating impacts on human health and the environment. Biodiesel is a widely studied energy source for this purpose, standing out for its biodegradable and environmentally friendly nature [9]. A variety of feedstocks can be used in its production, such as vegetable oils, animal fats, or microalgae oils. Although several vegetable oils can be employed in biodiesel production, competition with the food industry—such as in the case of soybean and corn—reinforces the need to invest in non-edible feedstocks. In this regard, castor oil (*Ricinus communis* L.) represents an interesting alternative. Castor seeds provide a high oil yield, ranging from 40 to 50%, and their cultivation requires few inputs, making it a low-cost feedstock [10].

Biodiesel production methods may vary, but homogeneous transesterification is the most widely used industrial technique, which can employ either acidic or basic catalysts [11]. It is worth emphasizing that the characteristics of the oil, the reagents, and the process parameters directly affect the yield and quality of biodiesel [12]. In homogeneous basic transesterification, low catalyst concentrations reduce efficiency due to limited active sites and alcohol/oil emulsion formation, while excessive amounts increase viscosity and saponification, hampering production [13]. Regarding temperature, for example, it affects reaction time and conversion rate, so that an optimal temperature results in reduced oil viscosity and improved yield [14]. Therefore, it is necessary to understand process conditions and determine the best operating parameters by combining technology with production. In this context, machine learning (ML) emerges as a promising approach to model these relationships.

Machine Learning (ML) is a subset of artificial intelligence that focuses on the development of algorithms and techniques capable of learning patterns from data without explicit programming [15]. It is a powerful tool for identifying relationships in nonlinear systems that would be difficult to approach using conventional methods [16]. Moreover, these techniques can be applied to diverse problems such as regression, classification, and prediction. ML encompasses multiple approaches, including deep learning (a subfield that incorporates artificial neural networks—ANNs), ensemble learning, and other methods.

Imai et al. (2020) analyzed ANNs as a tool to predict adverse drug reactions, concluding that they have the potential to support clinical decision-making and reduce risks to patients [17]. In another study, Çolak (2021) applied ANNs to predict the thermophysical properties of nanofluids, demonstrating their superiority compared to traditional methods [18]. Santana-Santos et al. (2022) employed Random Forest (RF) as the basis of a classifier for DNA methylation analysis in the diagnosis of central nervous system tumors,

highlighting its essential role in achieving high accuracy for complex methylation patterns. Validation of the method showed performance comparable to the gold standard, with 92% agreement [19]. Currently, various ML architectures are available for different purposes, such as Random Forest (RF), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), Multilayer Perceptrons (MLPs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and Radial Basis Function networks (RBFs).

The application of ML associated with castor oil has already been described in the literature under different approaches. In biolubricant production, MLP models have shown significant effectiveness, as in the work of Ahmad et al. (2022) with castor oil and iron oxide nanoparticles, which achieved a 94% yield [20]. Shojaeefard et al. (2012) experimentally investigated the performance and emissions of a direct injection engine using blends of castor oil biodiesel with conventional diesel and predicted these parameters using two artificial neural network approaches: Feedforward and Group Method of Data Handling (GMDH) [21]. Both studies indicate a positive correlation with the use of neural networks, promoting optimization, saving time, and reducing operational costs, as they minimize the need for extensive experimental testing.

The study by Yue et al. (2018) presents an approach closely aligned with this research, modeling biodiesel production from castor oil using the ANFIS (Adaptive Neuro-Fuzzy Inference System), with operational parameters as inputs and methyl ester yield as the output [22]. However, although they demonstrate the effectiveness of ANFIS, the study provides only operational ranges rather than specific values and employs a smaller dataset (156 entries) compared to the 406 datasets used in the present study.

Jana et al. (2022) address the optimization of waste cooking oil transesterification using CaO catalyst and various machine learning techniques, including type-1 fuzzy logic system (T1FLS), response surface methodology (RSM), ANFIS, and type-2 fuzzy logic system (T2FLS) [23]. Unlike our study, however, they do not propose a mechanism for optimizing operational parameters and utilize a different feedstock.

Despite advances in the application of ML to biodiesel production, gaps remain in the literature, particularly regarding comparative studies of different neural network architectures—such as MLP, RBF, RF, and ANFIS—and the optimization of performance metrics such as MSE (Mean Squared Error) and the correlation coefficient (R). There is a scarcity of comprehensive analyses evaluating multiple architectures, and most research is limited to direct yield prediction, without exploring inverse optimization approaches—which use the desired yield as an input to determine optimal combinations of temperature, molar ratio, reaction time, and other critical parameters.

Therefore, this work addresses this gap by proposing a hybrid model that not only predicts biodiesel yield with high accuracy but also, innovatively, integrates a genetic algorithm to perform inverse optimization. This approach automatically determines the ideal operational parameters for a target yield, offering a robust solution with greater industrial applicability.

2. Materials and Methods

2.1. Tools

This study employed MATLAB R2025a (Matrix Laboratory) software to develop and implement the codes for the models—Multilayer Perceptron (MLP), Radial Basis Function Networks (RBF), Random Forest, Hybrid (MLP + RBF), and Adaptive Neuro-Fuzzy Inference System (ANFIS). MATLAB is an interactive system with a matrix-based programming language, supporting fuzzy logic, machine learning algorithms, and other advanced computational techniques.

2.2. Kinetic Fundamentals of Transesterification

The transesterification of castor oil is a multi-step reaction involving the sequential conversion of triglycerides into diglycerides, monoglycerides, and finally into fatty acid methyl esters (biodiesel) and glycerol, as illustrated in Figure 1. The transesterification reaction is characterized as non-elementary and reversible in nature.

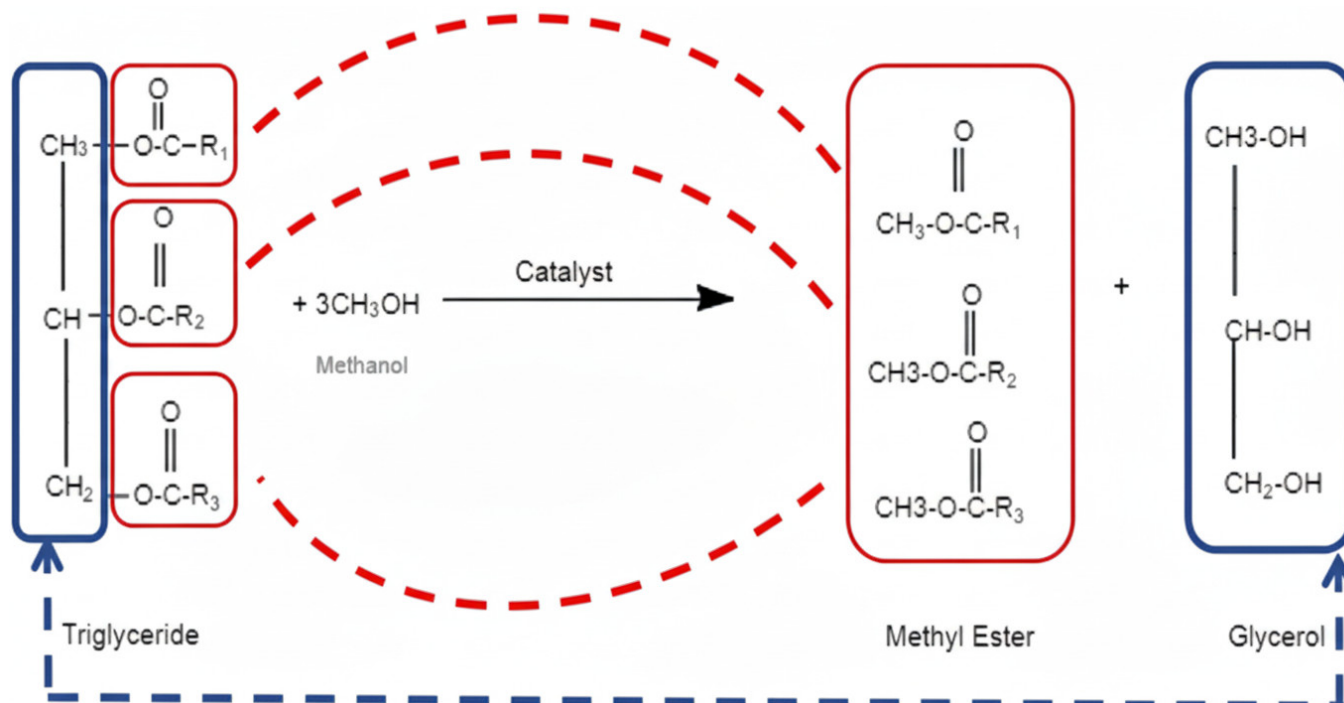


Figure 1. Biodiesel production via transesterification [24].

Ramezani et al. (2010) proposed a pseudo-first-order reaction model in their study, adopting the following simplifications: an irreversible reaction and a constant molar concentration of methanol during the reaction, to predict a simple kinetic equation for the initial stage of the reaction [25]. The integrated kinetic Equation (1) is given by:

$$\ln(3 - X_{ME}) = -kt + \ln 3 \quad (1)$$

where X_{ME} is the methyl ester yield, k is the apparent rate constant, and t is the reaction time.

Although this work does not aim to establish an explicit kinetic model, the use of ML algorithms serves as a computational alternative to conventional kinetic models, as they are capable of capturing nonlinear relationships between operational variables and the reaction yield.

2.3. Data and Preprocessing

The data used for training, validation, and testing of the neural networks, the fuzzy inference system, and the ensemble learning model were collected from the literature. The search was carried out on the platforms ScienceDirect, Taylor & Francis Group Online, SpringerLink, and Google Scholar, using the phrases: “Homogeneous transesterification of castor oil,” “Homogeneous transesterification of castor oil to biodiesel,” and “Production of biodiesel from castor oil.”

For the construction of the database, 406 labeled datasets were collected, containing information on the operating conditions for homogeneous alkaline or acid transesterification of castor oil. The input variables are: type of catalyst (basic/acid), alcohol-to-oil molar ratio, catalyst concentration (% w/w), transesterification temperature (°C), reaction time (min),

and stirring speed (rpm). The output variable is biodiesel yield (FAME). The minimum and maximum values of each variable are presented in Table 1. The detailed database composition and its complete operating ranges are presented in Table S7 (Supplement 2). The qualitative variable—type of catalyst—was handled using one-hot encoding. Information on stirring speed is more limited in the literature, due to the tendency of authors not to report numerical values for this variable. However, this limitation does not compromise the models, since artificial neural networks are capable of learning from incomplete and noisy data [26].

Table 1. Lower and Upper Bounds of the Process Variables.

Input Variable	Range of Variation
Molar ratio (alcohol:oil)	3:1–31.48:1
Catalyst concentration	0.50–4 wt%
Temperature	25–70 °C
Reaction time	10–600 min
Stirring speed	300–1200 rpm

In the pre-processing stage, numerical data were normalized to optimize convergence during model training. Given the nature of the data analyzed in this study, the range between 0 and 1 was deemed most appropriate, as the input data contained no negative values.

To optimize the models' hyperparameters (such as the number of neurons in MLPs or the spread in RBFs), auxiliary routines were implemented to automatically vary these parameters. Their performance was evaluated using balanced metrics (mean and standard deviation of R across the training/validation/test sets). The configurations with the best metrics (MSE and R) were replicated in the final models. The specific simulation parameters for each trained machine learning architecture (MLP-logsig, MLP-tansig, RBF, Hybrid RBF + MLP, Random Forest, and ANFIS) are de-tailed in Tables S1–S6 (Supplement 1). The trained weights were then saved for later integration with the Genetic Algorithm (GA), enabling the GA to utilize the relationships learned by the network to optimize the process variables.

2.4. Network Architectures

2.4.1. Multilayer Perceptron (MLP)

The Multilayer Perceptron (MLP) is a type of feedforward artificial neural network composed of layers of neurons. These networks contain a hidden layer situated between the input and output, enabling the extraction of more sophisticated features. They are used for classification, regression, and pattern recognition problems [27]. This study utilizes a two-layer Feedforward Neural Network with either logsig (log-sigmoid) or tansig (tan-sigmoid) hidden neurons and linear output neurons. The nftool function in MATLAB was used for its implementation. For network training, the data were randomly divided into three groups: training (70%), testing (15%), and validation (15%). This MLP was trained using the Levenberg–Marquardt algorithm to accelerate model convergence.

2.4.2. Radial Basis Function Networks (RBF)

Radial Basis Function Networks (RBFs) are a type of compact neural network with a high convergence rate, functioning as a system that compares new information to previously known examples [28]. This approach allows the model to be applied to different tasks, such as classification, regression, and linear interpolation, depending on the chosen loss

function [29]. In this study, the RBF architecture consists of two layers: hidden neurons with a Gaussian activation function (Equation (2)) and linear output neurons.

$$z_j(x) = \exp\left(-\frac{\|x - c_j\|^2}{2\sigma_j^2}\right) \quad (2)$$

where:

x : Input vector of the neural network.

c_j : Center of the Gaussian function for the j -th hidden layer neuron.

$\|x - c_j\|$: Euclidean distance between the input vector and the RBF center.

σ_j : Smoothness adjustment factor (“spread”) of the Gaussian function.

The data were divided into two groups: 70% for training and 30% for testing. In MATLAB, the network is defined using the newrb function. This function trains the RBF by progressively adding hidden neurons; at each iteration, it selects the sample with the highest prediction error and defines a new RBF center at that sample, adjusting the output layer weights based on the neurons already added. The stopping criterion is met when the network error falls below the tolerance value (goal = 0.001) or when the maximum number of neurons is reached (maxNeurons = 40).

2.4.3. Hybrid Network

The hybrid model integrates a Radial Basis Function (RBF) layer (Gaussian function) with a Multilayer Perceptron (MLP) layer (log-sigmoid function) in series. This architecture combines the local sensitivity of the RBF with the generalization capability of the MLP to enhance predictive performance. The network’s operation is illustrated in the flowchart of Figure 2.

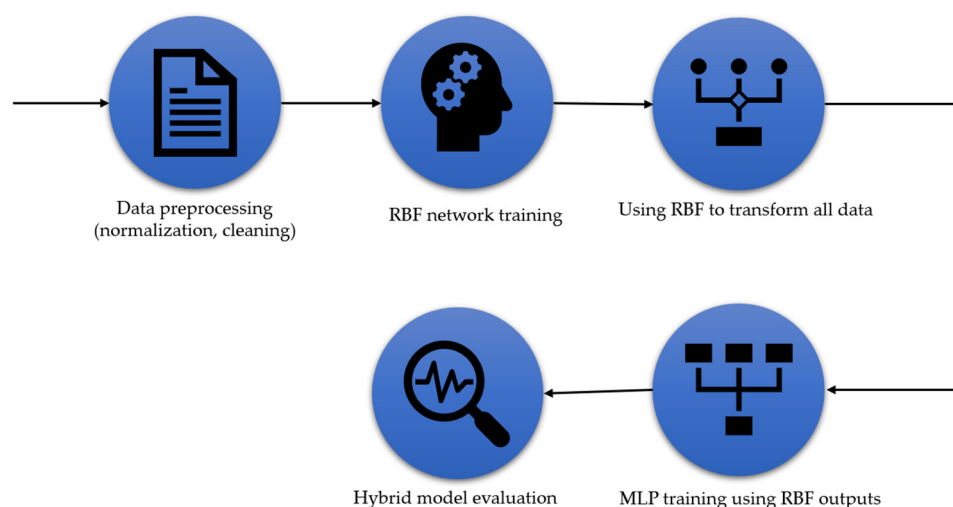


Figure 2. Hybrid Network diagram.

2.4.4. Random Forest

Random Forest is not a neural network; this ensemble learning algorithm combines multiple random decision trees to improve model accuracy and robustness. The algorithm has been primarily used for classification, prediction, and regression problems [30]. Each tree is trained on a random subset of the data and a random subset of input variables. The final prediction is the average (for regression) or the majority vote (for classification) of the results from the individual trees [31]. The model training in this article is based on a combined process of random sampling and iterative evaluation, utilizing cross-validation (K-Fold with $K = 5$) to ensure robustness. Key parameter combinations were tested to find the optimal indicators:

- Number of trees in the forest (TreeBagger parameter): 50 to 500;
- Minimum number of observations per leaf (MinLeafSize parameter): 1 to 10;
- Maximum number of splits per tree (MaxNumSplits parameter): 10 to 100;
- Number of features evaluated at each split (NumPredictorsToSample parameter): 1 to 10.

The use of MATLAB's TreeBagger function allows for control over details such as the randomization of input variables (NumPredictorsToSample) and tree complexity (MinLeafSize and MaxNumSplits).

2.4.5. Adaptive Neuro-Fuzzy Inference System (ANFIS)

The Adaptive Neuro-Fuzzy Inference System (ANFIS) is a hybrid system that combines artificial neural networks and fuzzy logic. This model thus integrates the learning capability of neural networks with the easily interpretable reasoning of fuzzy logic. It is frequently used for regression, prediction, and optimization problems [32]. In this study, the network is generated by genfis using Subtractive Clustering, a MATLAB tool that automatically defines the number of fuzzy rules and membership functions by analyzing data density. The ClusterInfluenceRange parameter controls the sensitivity to cluster formation [33]. For algorithm optimization, Gradient Descent and Least Squares Estimation are used. To ensure the robustness of the standard ANFIS model, stratified validation (70% training/30% validation) was implemented. This approach enables a more comprehensive evaluation of the model's generalization behavior by comparing the training and validation R coefficients.

2.5. Performance Indicators

Performance indicators are used to evaluate the performance of the algorithms, aiming to assess the model's generalization capability. These are essential for identifying the presence of overfitting and underfitting. The present work utilized: Mean Absolute Error (MAE), Mean Squared Error (MSE), Correlation Coefficient (R), Root Mean Squared Error (RMSE), Regression Plot, and Performance Plot. Equation (3) represents MAE, Equation (4) represents MSE, and Equation (5) represents RMSE [34].

$$\text{MAE} = \frac{(y_i - y_p)}{n} \quad (3)$$

$$\text{MSE} = \frac{\Sigma(y_i - y_p)^2}{n} \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{\Sigma(y_i - y_p)^2}{n}} \quad (5)$$

where:

n : number of samples,

y_i : actual value,

y_p : predicted value.

The Regression Plot identifies the relationship between the actual values and the values predicted by the models, showing how closely the model approximated the actual results. The Performance Plot, on the other hand, indicates the evolution of the error during the training process, allowing us to identify whether the model is learning correctly, has stopped improving, or is suffering from overfitting.

2.6. Optimization Model: Genetic Algorithm

The genetic algorithm was implemented to promote process optimization, being responsible for predicting the ideal conditions for alkaline or acid homogeneous transesterification parameters aiming for a yield close to 100%. This is an algorithm inspired by

biological evolution and the theory of natural selection. Lambora et al. (2019) indicate this as an important optimization tool, capable of solving complex problems with or without constraints [35]. In this work, the GA is employed in a reverse optimization strategy. While conventional modeling predicts the yield from given inputs, our approach inverts this logic: we define the target yield and use the GA to find the input parameters that achieve it. The trained ml models (MLP-tansig, RBF, hybrid MLP + RBF, Random Forest, and ANFIS) were saved and integrated with a customized and optimized genetic algorithm (GA) to maximize biodiesel yield, as represented in the process flowchart (Figure 3).

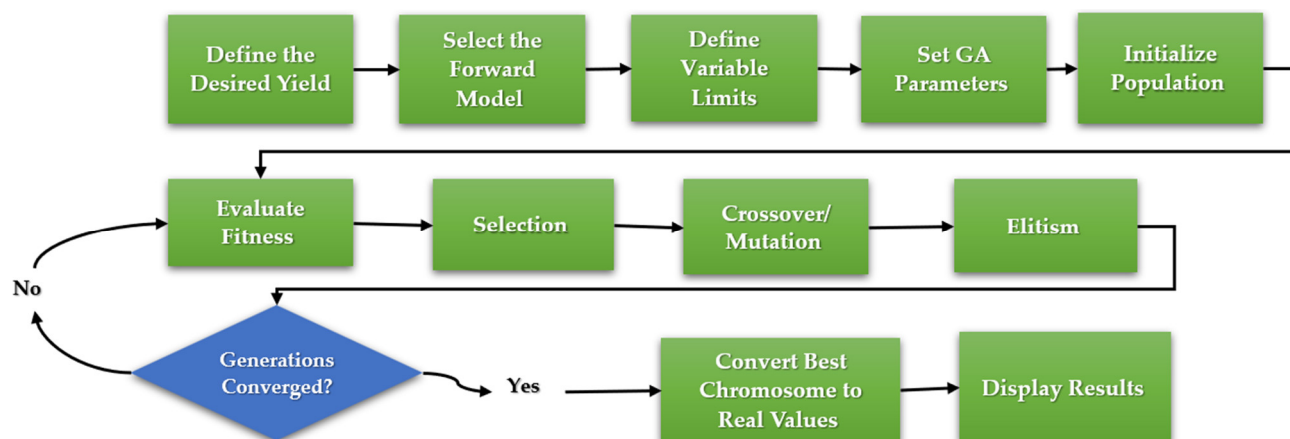


Figure 3. Genetic Algorithm Flowchart.

The GA operates with each chromosome encoding variables such as molar ratio, temperature, reaction time, and catalyst type (one-hot encoding for categories). The fitness function evaluates the yield predicted by the model and applies penalties for values outside the desired range (90–100%). The algorithm employs roulette wheel selection, crossover with one-hot correction, and adaptive mutation (Gaussian for continuous variables and binary swap for categories), in addition to elitism to preserve the best solution. Convergence is monitored across generations, with final results converted back to the original data scales. Population size, number of generations, crossover rate, and mutation rate were varied to seek the most coherent results with the literature and closest to the desired outcome. This approach was replicated for all models, ensuring a fair comparison between architectures.

3. Results

3.1. Model Comparison

The MLP-logsig, MLP-tansig, RBF, RBF + MLP, Random Forest, and ANFIS models were individually trained, testing various basic parameters of each model. The variation of parameters allowed finding the best performance indicators, namely, correlation coefficient (R) closest to 1 (for a positive correlation), mean squared error (MSE) closest to 0 (to minimize absolute errors), RMSE with minimized values (to reduce deviations in FAME yield) and a Mean Absolute Error (MAE) also close to 0, ensuring a robust assessment of the models' average accuracy.

In the MLP neural network, both with the logistic sigmoid and hyperbolic tangent activation functions, the number of neurons in the hidden layer was adjusted. In turn, in the RBF network, the spread parameter was optimized, a factor that defines the width of the radial basis function. The hybrid model alternates between different neurons in the Multilayer Perceptron hidden layer and spreads in the Radial Basis Function Networks. The RF model has several parameters that can be varied: number of trees in the forest; minimum observations per leaf; maximum splits per tree; and features evaluated at each split. In ANFIS, since Cluster Influence Range was applied, only this parameter and the number of

epochs were optimized. To ensure reproducibility, the Random Number Generator (RNG) was defined for all codes and was also varied, generating diverse results.

After performing several simulations, the results of the performance metrics will be analyzed. It is important to emphasize that RBF and RF do not present validation data due to the characteristics of their algorithms. In contrast, ANFIS inherently requires only training in its base characteristics, with validation being employed for better evaluation.

Figure 4 presents a comparative analysis of the RMSE values in absolute units (%), allowing a direct quantitative assessment of the model. The RMSE on a real scale indicates the average percentage error in the model. For example, for an expected yield of 80%, the RF model would demonstrate an error of $\pm 11.55\%$, generating predictions in the range of 68.5% to 91.5%, making it unviable for process control.

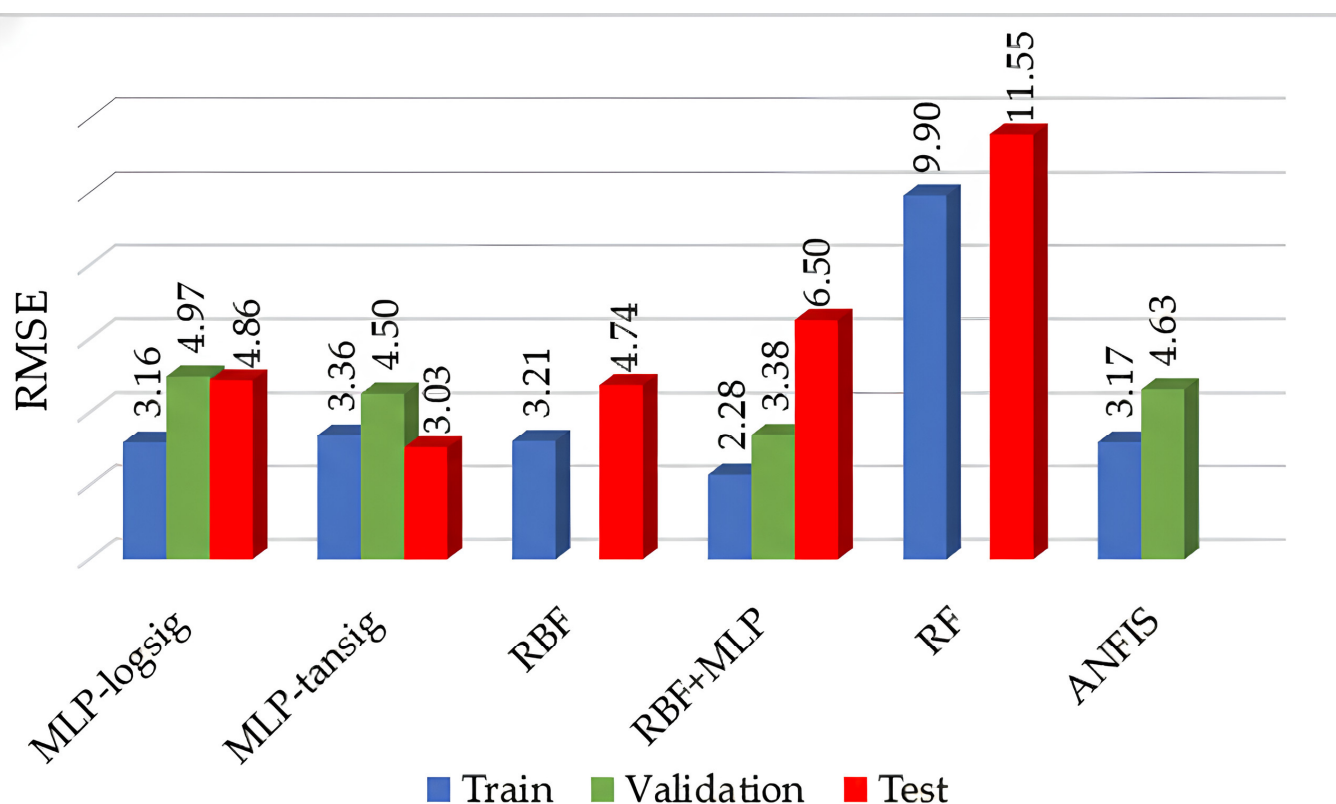


Figure 4. Model Performance in Terms of RMSE (%).

The MLP-logsig and ANFIS models exhibit validation/test error approximately 50% higher than training, highlighting the need for adjustments and adaptations for these models. The hybrid model (RBF + MLP) displays excellent training performance but the second-worst test value ($\pm 6.50\%$) compared to the others. The model that showed the best test results was MLP-tansig, indicating that its predictions are, on average, only $\pm 3.03\%$ away from the actual FAME yield value.

The other performance metrics (R and MSE) are analyzed in Figures 5 and 6, respectively, highlighting the most significant values obtained for each model configuration and parameters evaluated.

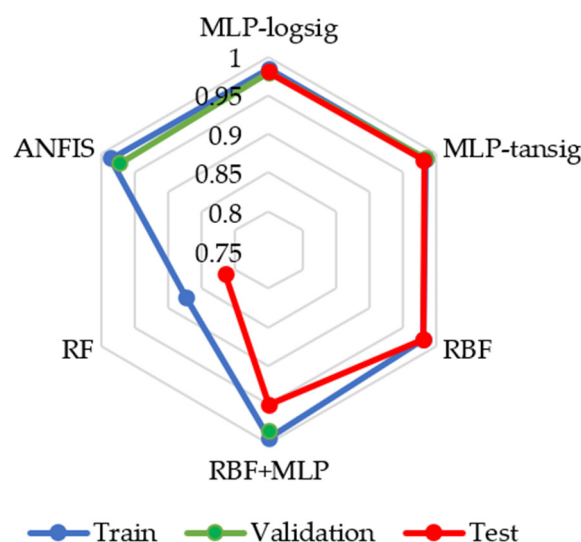


Figure 5. Model Performance in Terms of R Coefficient.

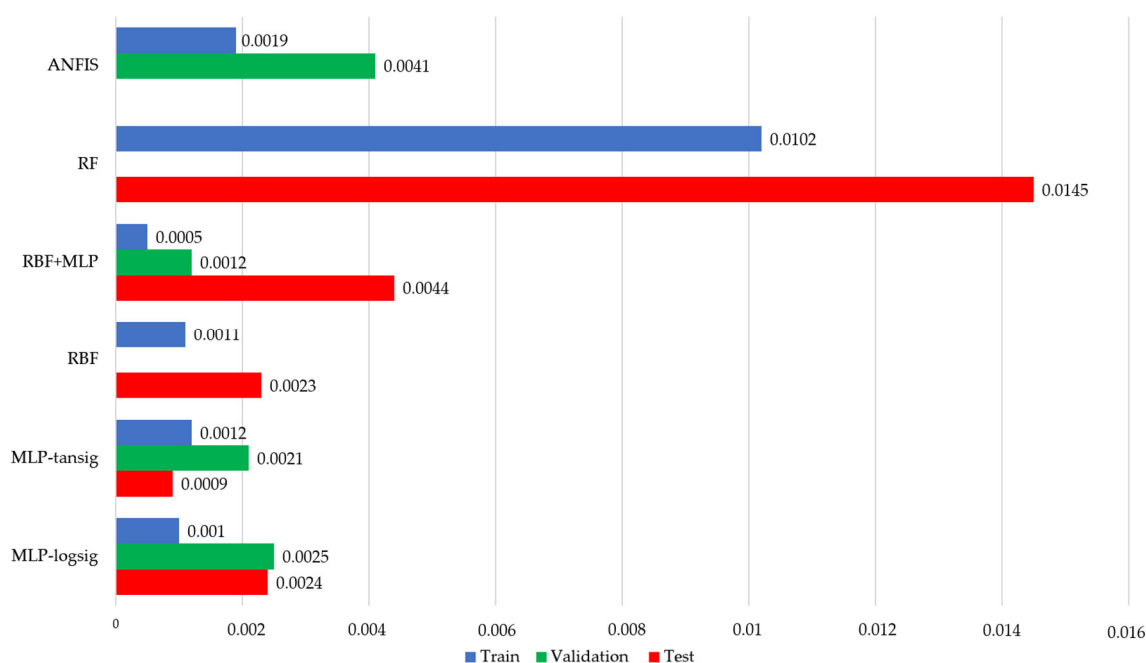


Figure 6. Model Performance in Terms of MSE.

It can be observed that the Random Forest recorded the lowest training R value, indicating a lower ability to interpret the relationship between generated and actual values, being inferior to the other models. In testing, RF showed a lower test R value than training, suggesting a decline in generalization capability. The MSE values are also slightly higher than the others, indicating that this model lacks good precision for predicting the data. This is a clear sign of underfitting.

The hybrid network (RBF + MLP) exhibits excellent training performance with $R_{\text{train}} = 0.9932$, superior to all other models. However, the discrepancy with the test R value of 0.9504 indicates that the network cannot transfer its learning efficiently, indicative of overfitting.

The other algorithms demonstrate excellent R indices for training, testing, and validation, with values close to 1, revealing strong correlation and good generalization. The mean squared errors were consistently low.

Given the excellent indicators, identifying the network with the best performance—capable of combining consistency, generalization, and resistance to overfitting and underfitting—poses a challenge. The MLP-logsig, MLP-tansig, RBF, and ANFIS algorithms exhibit strong positive correlation during training. The MLP-tansig recorded R values above 0.98 in both training and validation. In contrast, MLP-logsig and ANFIS, despite excellent training performance ($R > 0.98$), showed slightly lower R values in validation (< 0.98). As for RBF, due to its architecture, it does not include validation, making it necessary to compare its test R value with that of MLP-tansig. Although they demonstrated similar test performance, MLP-tansig was chosen due to its: slightly superior training performance (training R 0.0030 higher), better stability across datasets, and lower test MSE.

The MAE results, presented in Figure 7, indicate the superiority of the MLP-tansig model, which exhibited the lowest mean absolute errors across all datasets: training (0.019), validation (0.031), and test (0.020). This signifies that, on average, the model's predictions deviate by only approximately 1.9% to 3.1% from the actual FAME value, confirming its precision and robustness. The RBF and ANFIS models demonstrated intermediate performance, consistent with previous analyses. Finally, the Random Forest model recorded the highest MAE values in both training (0.064) and testing (0.075), a strong indicator of underfitting.

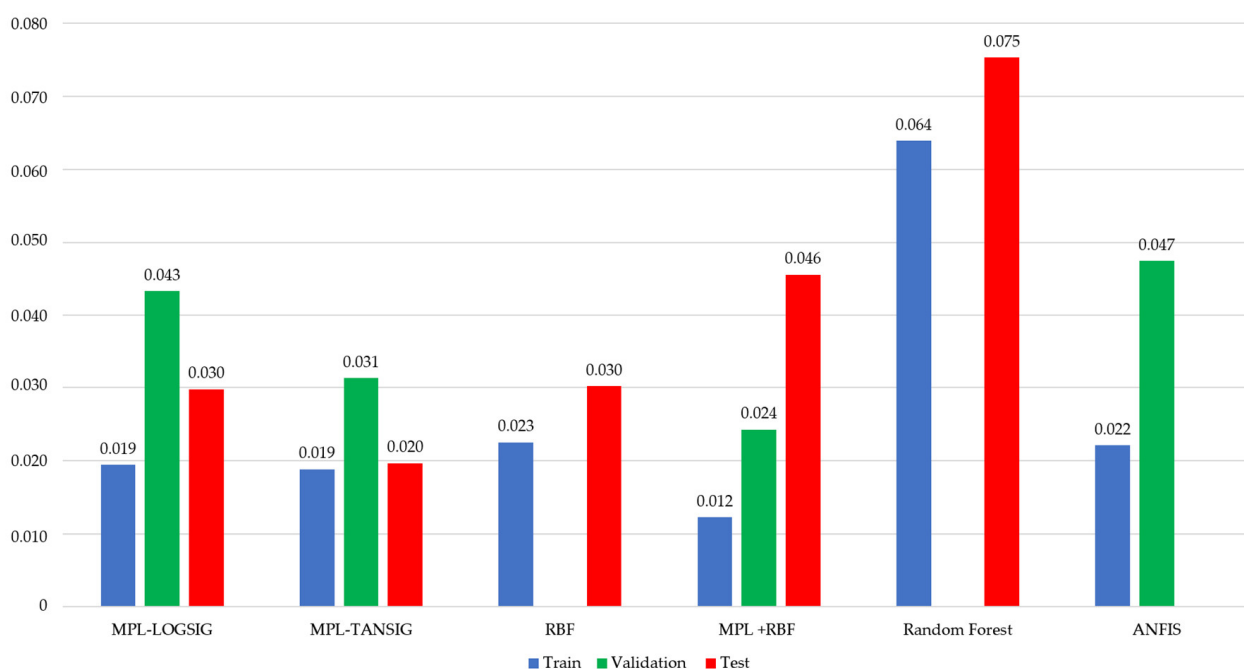


Figure 7. Model Performance in Terms of MAE.

In summary, the integrated analysis of the R, RMSE, MSE, and MAE metrics identified the MLP-tansig model as the most efficient, exhibiting a high correlation ($R > 0.98$) and the lowest absolute errors ($MAE \approx 2\text{--}3\%$, $RMSE = 3.03\%$), demonstrating superior accuracy and robustness. The RBF network established itself as the second-best option, with performance very close to MLP-tansig on the test data, albeit with slight inferiority in stability across datasets and in the MSE. The remaining models, despite showing good indicators in some aspects, were discarded due to issues of overfitting (RBF + MLP) or underfitting (Random Forest).

3.2. Best-Performing Network in the Training Phase

As discussed in the previous section, the Multilayer Perceptron network with a hyperbolic tangent activation function and linear output demonstrated the best performance indicators. Figure 8 shows a representation of this network's architecture with 19 neurons in the hidden layer.

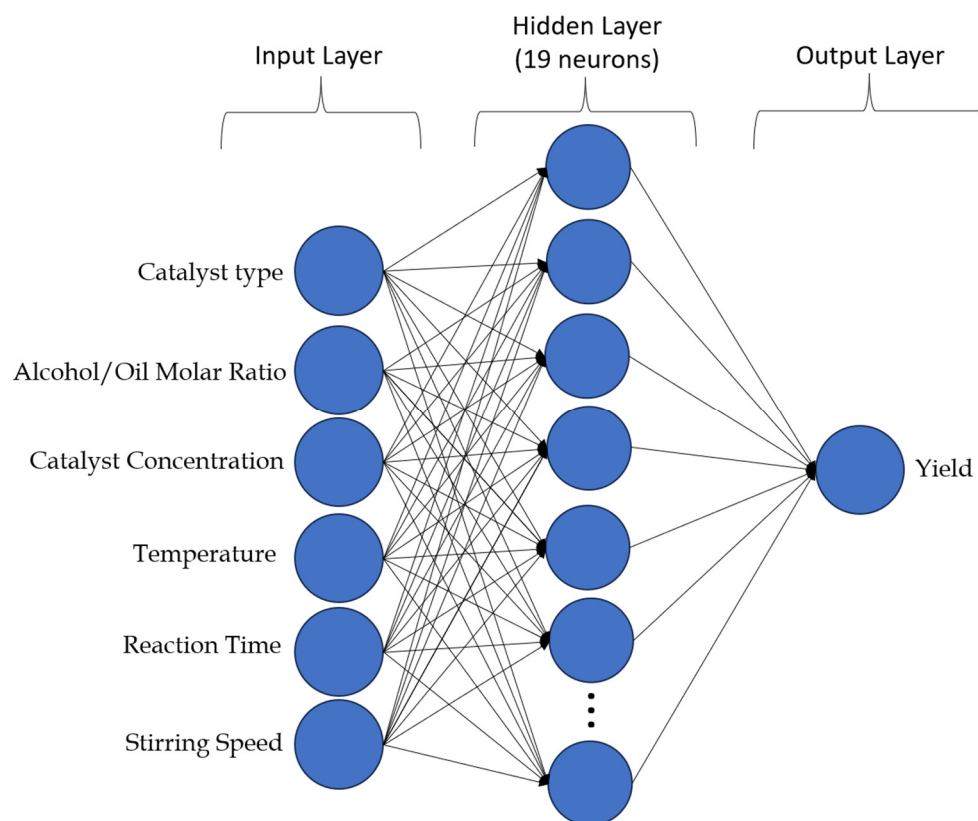


Figure 8. Multilayer Perceptron tansig Network Architecture.

The performance plot, Figure 9, shows the best validation performance at epoch 15, with an MSE of 0.0020936. It can be stated that the training was successful, stopping at the appropriate time, indicating good generalization capability of the model.

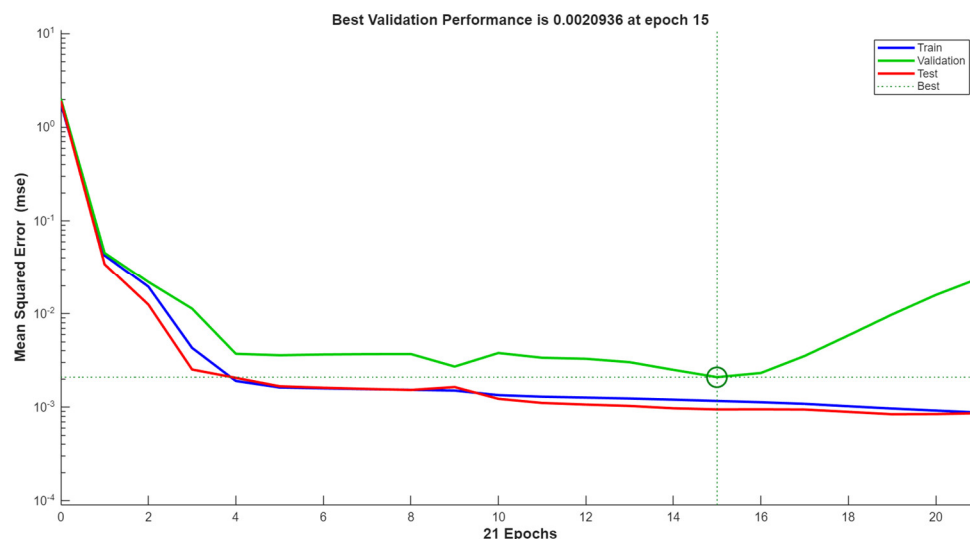


Figure 9. Performance Plot for Multilayer Perceptron tansig.

Figure 10 presents the regression plots between the actual and predicted values for the training, validation, test, and complete sets. The plot shows nearly perfect linear correlations across all sets, along with excellent consistency between training data and unseen data. The high correlation values ($R > 0.98$) in all cases indicate the robustness and generalization capability of the proposed model, with no signs of overfitting.

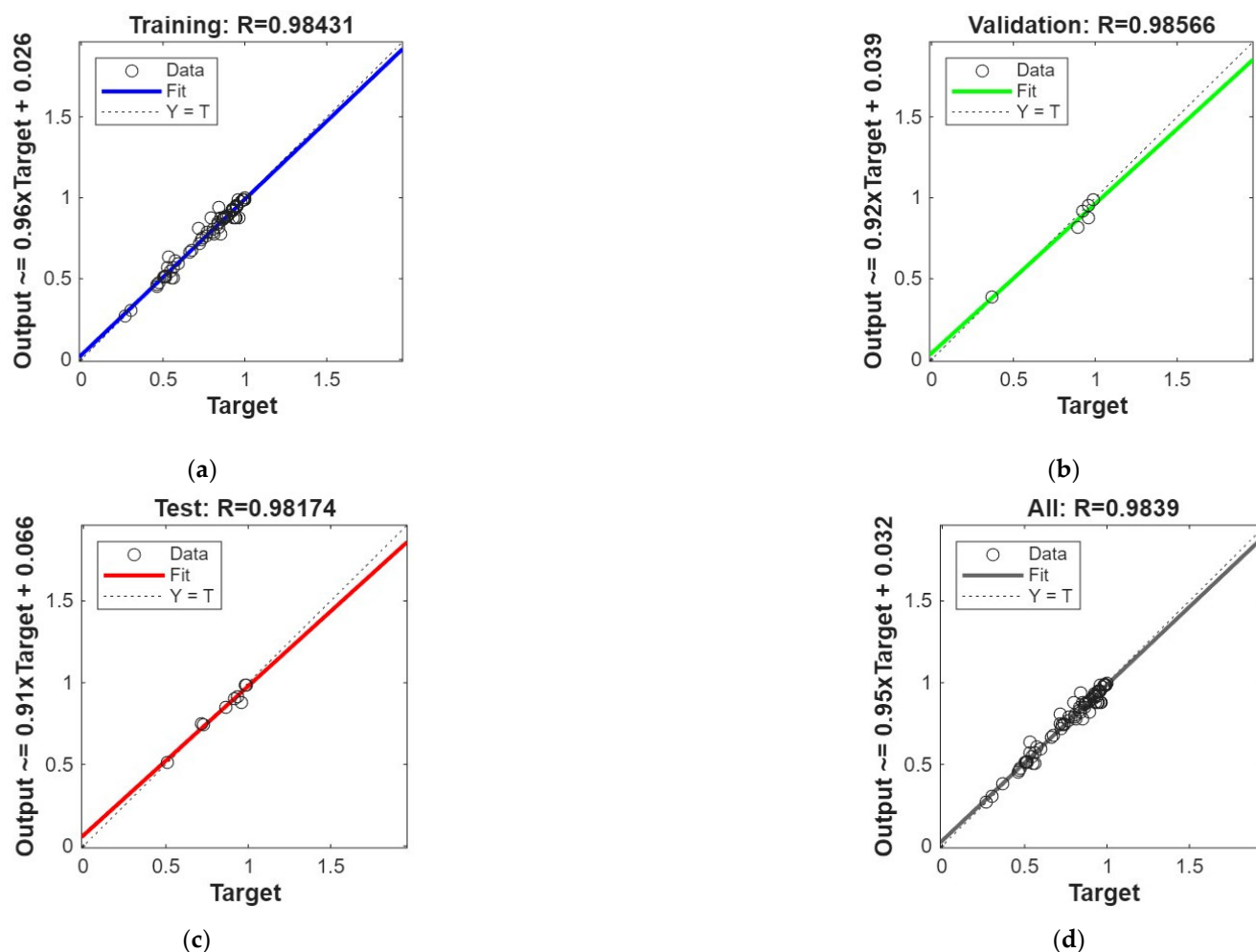


Figure 10. Regression plots for MLP-tansig: (a) Training; (b) Validation; (c) Test; (d) All.

3.3. Applying the GA

After identifying the best configurations for each model and determining which one performs the best, the genetic algorithm is applied in conjunction with the already trained model. In this stage, the desired yield has already been defined as 90–100% (FAME).

Next, the boundaries of the input variables (the same as described in Table 1) are defined, always considering the evaluation of both acid and basic homogeneous transesterification.

To enable a comparison between the results generated by the different GA parameters, the following parameters were varied: population size, number of generations, crossover rate, and mutation rate. The output is expected to indicate the catalyst type with the best performance and the most satisfactory operational variables.

Although the MLP-tansig network demonstrated excellent performance during training, its combination with the GA did not yield adequate results. Specifically, for the same genetic parameters presented in Figure 11, the MLP-tansig model suggested a reaction time of 281.93 min and a catalyst concentration of 2.35%—values significantly higher than those typically reported in the literature for this type of reaction. In contrast, the RBF generated more coherent and feasible values, aligned with the operational conditions usually employed. As discussed in Section 3.1, the RBF also achieved excellent results, ranking as the second-best during training.

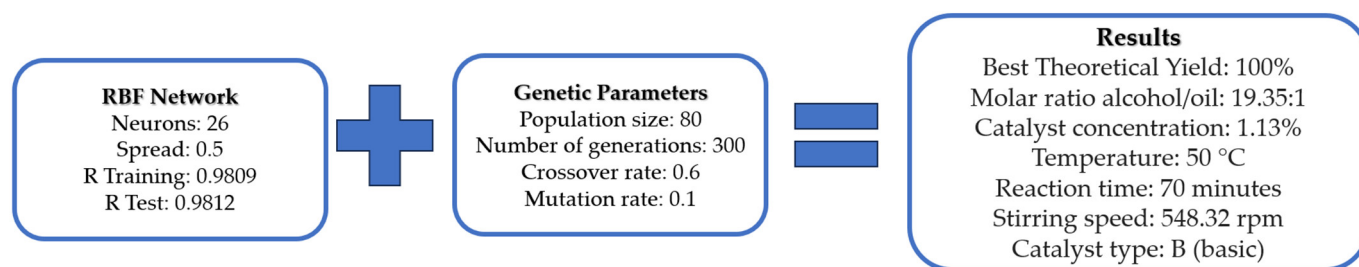


Figure 11. Diagram: Result of the RBF + GA Analysis.

In Figure 11, presents the RBF + GA analysis results diagram, highlighting the RBF network parameters along with its correlation coefficients (R) for training and testing, and the Genetic Algorithm configuration. The genetic parameters were defined to balance computational efficiency and solution search comprehensiveness: a crossover rate of 60% for efficient recombination of promising solutions, and a mutation rate of 10% to preserve population diversity. This configuration aims to ensure stable and rapid convergence during optimization. The application of this strategy yielded the following optimal operating conditions.

4. Discussion

In order to better analyze the model's consistency with data from other articles, scatter plots (Figures 12 and 13) were used to compare the optimal operational conditions found by the model (red point) and the process variables that achieved the highest yield in each scientific article (blue points).

The genetic algorithm indicates homogeneous basic catalyst as the best alternative for biodiesel production from castor oil. This prediction aligns with the literature, which recognizes that it is generally more efficient and cheaper than acid catalysts [36]. Basic catalysts accelerate the reaction approximately 4000 times more than acid catalysts, such as hydrochloric acid (HCl) [37]. This significant kinetic difference explains the rare adoption of acid catalysts in industrial biodiesel production processes.

In Figure 12, it can be observed that (a) temperature, (b) catalyst concentration, and (c) time are all within the range of the literature data. The model's temperature was 50 °C, while the average across the 26 articles in the graph was 54.7 °C. However, Akhabue and Okwundu (2017) observed that temperatures above 50 °C reduce FAME yield due to methanol evaporation [38]. Furthermore, excessively high temperatures can promote parallel saponification reactions, which actively consume both triglycerides and catalyst, consequently leading to a decrease in biodiesel conversion yield [14]. The optimal catalyst concentration assigned by the model is 1.13%, while the average across the 26 articles is 1.19%, a difference of 0.06%. Elevated concentrations of basic catalysts adversely affect the main transesterification reaction, promoting the formation of byproducts such as soap and water [39]. Thus, the model's value falls within acceptable limits to promote reaction acceleration without compromising time and yield through parallel reactions.

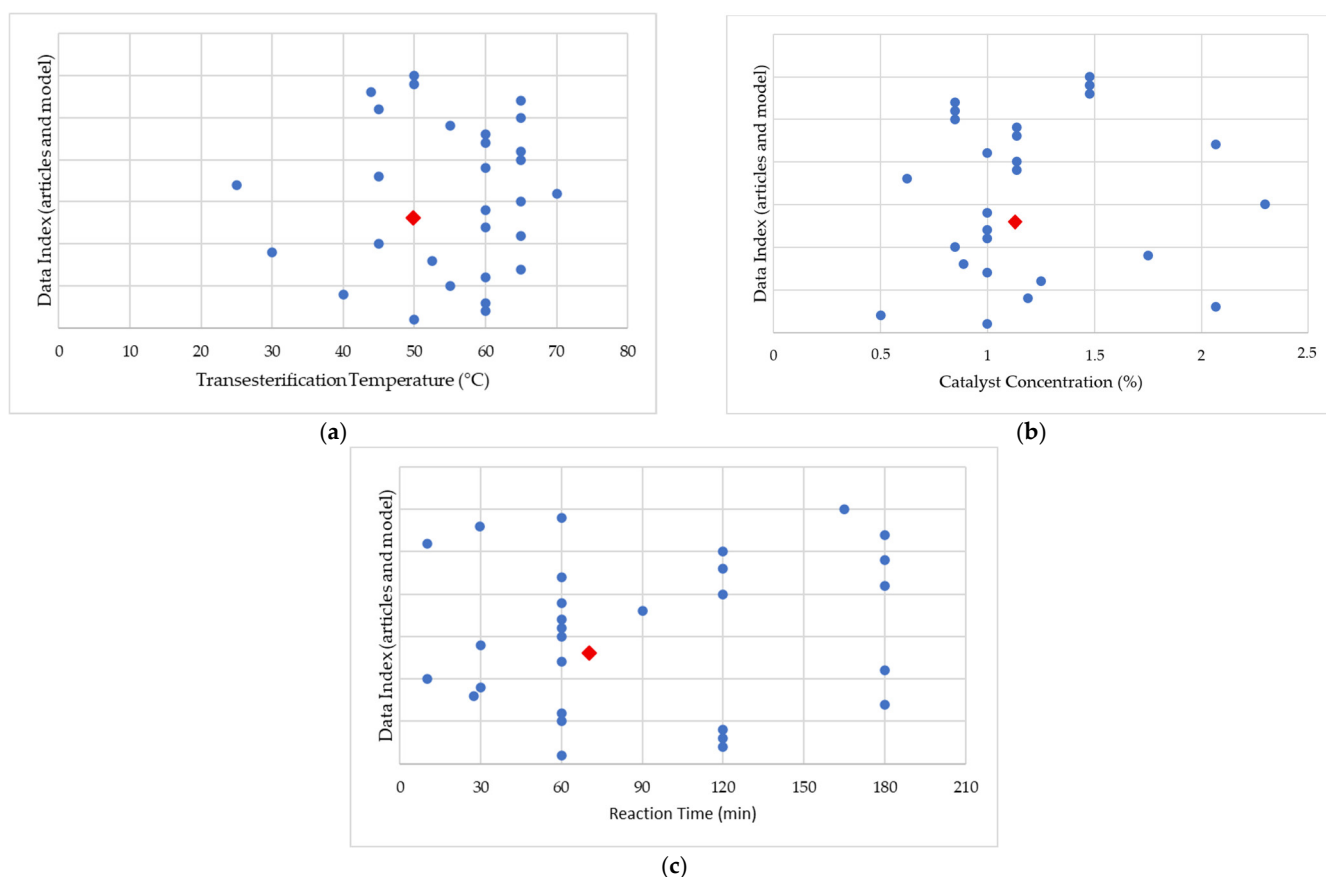


Figure 12. Comparison between the model (RBF + GA) and data reported in the literature: (a) Transesterification Temperature (°C) [25,39–65]; (b) Catalyst Concentration (%) [25,38–43,45–52,55–65]; and (c) Reaction Time (min) [25,38–65]. Caption: ♦ Optimization generated by the model; • Values found in the literature.

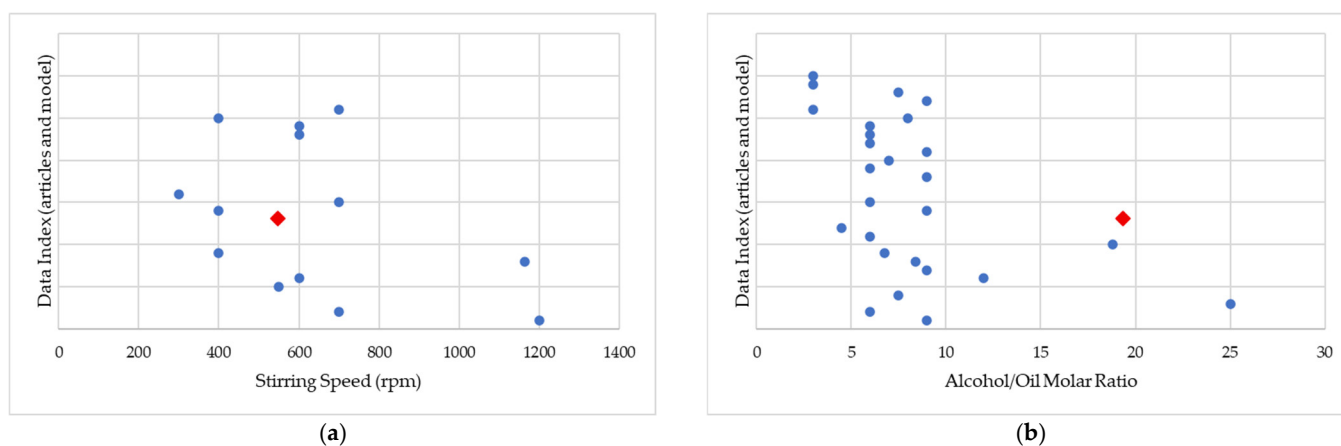


Figure 13. Comparison between the model (RBF + GA) and data reported in the literature: (a) Stirring Speed (rpm) [25,40,41,44,46,47,51–53,59,60] and (b) Alcohol/Oil Molar Ratio [25,38–43,45–52,55–65]. Caption: ♦ Optimization generated by the model; • Values found in the literature.

Regarding reaction time, data from 29 studies were collected. The reaction time of 70 min, considered ideal by the model, falls within the dispersion range of the literature data, validating its coherence. The wide dispersion of the blue points suggests there is no clear consensus regarding the optimal reaction time. When analyzing the studies by Masango et al. (2024) and Elango et al. (2019), both focusing on the optimization of homogeneous transesterification of castor oil, a divergence is observed [39,40]. Masango et al. (2024)

state that the reaction reached its steady state after 90 min [39]. In contrast, Elango et al. (2019) report the optimal FAME yield at a reaction time of 60 min. The divergence between the studies may reflect distinct experimental conditions [40]. Mohiddin et al. (2021) emphasize that a short reaction time may result in inadequate dispersion of the feedstock and alcohol. Conversely, an excessively long reaction time maintains the system at equilibrium, favoring the occurrence of reverse reactions, such as ester hydrolysis and saponification, due to the reversible nature of the transesterification reaction. At the optimal reaction time, triglycerides are sequentially converted into diglycerides and monoglycerides—key intermediates that, when present at adequate concentrations, accelerate the formation of methyl esters (biodiesel). Therefore, this progressive accumulation of reactive intermediates shifts the chemical equilibrium towards the desired product [13].

Stirring speed is an important parameter for maximizing FAME yield, as it promotes efficient particle collision and diffusion within the reaction mixture. This agitation effect contributes to reducing the reaction time and increasing biodiesel conversion [12]. However, beyond a certain agitation speed, no significant yield improvement is observed, as the optimal speed varies depending on the specific feedstock. In the present work, the stirring speed (Figure 13a) was analyzed by comparing 13 articles, a smaller number than the other variables, as it is a parameter less frequently cited numerically, yet important for the process. The red point (model) falls within the dispersion of the data, indicating conformity with the literature, for an optimal rotational speed of 548.32 rpm.

It is well-established that the alcohol-to-oil molar ratio also strongly influences biodiesel yield. To promote the transesterification reaction, the stoichiometric ratio requires an excess of alcohol in a minimum proportion of 3:1 [14]. According to Le Chatelier's principle, the rate of product formation increases with higher reactant concentrations; that is, biodiesel formation tends to be favored by a higher alcohol concentration [12]. In the scatter plot of the alcohol-to-oil molar ratio (Figure 13b), it can be observed that, among all the variables analyzed, the optimal condition identified by the model for this parameter was the only one divergent from the central tendency region of the points. However, Yue et al. (2018), a study evaluating 156 secondary experimental datasets on castor oil biodiesel, suggests a prediction range of 6:1 to 25:1 [22]. Thus, the model's prediction of 19.35:1 falls within this range.

In summary, the present work aligns with contemporary biodiesel research, which has demonstrated a growing and diverse application of computational methodologies for process optimization. For instance, Jin et al. (2023) investigated the performance of four algorithms (kNN, SVM, AdaBoost, and Random Forest Regression) in predicting biodiesel yield via transesterification. Their results demonstrated that the Random Forest Regression model was the most suitable, exhibiting RMSE values of 2.778 for training and 5.178 for validation, and showing superior predictive accuracy among the compared models [66]. These performance metrics are comparable in robustness to those obtained and employed in our study, thereby reinforcing the coherence and reliability of predictive approaches in this field.

This study distinguishes itself by conducting a comprehensive comparative analysis of six machine learning architectures—MLP-logsig, MLP-tansig, RBF, hybrid MLP + RBF, Random Forest, and ANFIS—specifically for modeling the transesterification of castor oil. Furthermore, combining machine learning models with GA for reverse optimization of operational conditions represents a significant contribution. While numerous previous studies focused solely on yield prediction, this work advances the field by inverting the traditional logic, using the desired yield as input to determine the optimal process parameters [23,66,67].

Other literature reviews, such as the one by Awogbemi and Von Kallon (2023), highlight the efficacy of models like Artificial Neural Networks (ANN), Response Surface Methodology (RSM), and ANFIS, confirming that the use of machine learning can increase biodiesel yield to levels between 84% and 98% for different feedstocks [68]. The model proposed in our study aligns with this trend, identifying a theoretical optimum that reaches a 100% yield. Furthermore, Buasri et al. (2023) demonstrated the application of machine learning to optimize biodiesel production from waste cooking oil in microwave reactors [69]. Their work reported excellent performance metrics, with a coefficient of determination (R^2) of 0.9988 for the Box–Behnken design (BBD) model and a correlation coefficient (R) of 0.9994 for the ANN model. These metrics are comparable in magnitude to the high predictive accuracy achieved in the present study, reinforcing the robustness of advanced computational approaches in this field.

The performance metrics of average $R > 0.98$, average MSE of 0.0014, and average RMSE of 3.63% demonstrate the high accuracy of the MLP-tansig model, validating its generalization capability without signs of overfitting or underfitting. The agreement between the optimal conditions predicted by the model and the literature data reinforces the model's reliability.

The modeling approach can be extended to other vegetable oils, provided the input parameters are adjusted accordingly. This flexibility allows for the analysis of biodiesel production with varying raw material availabilities. In biorefineries, the model could prove highly valuable by integrating monitoring systems and dynamically adjusting process conditions to maintain maximum yield.

5. Conclusions

In the present study, approximately 406 labeled datasets of biodiesel production from alkaline and acid homogeneous transesterification of castor oil under different operational conditions were extracted from the literature. The MLP-tansig neural network with linear output and 19 neurons in the hidden layer demonstrated the best training performance, with average indicators of $R > 0.98$. The GA optimized the operational conditions, targeting a theoretical optimum of 100% ester content. The results of the GA combined with the RBF neural network—yielding the most consistent outcomes—indicated optimal conditions: alcohol/oil molar ratio of 19.35:1, catalyst concentration of 1.13% (w/w), temperature of 50 °C, reaction time of 70 min, stirring speed of 548.32 rpm, and alkaline catalyst. Validation with experimental data confirmed the model's efficacy. Thus, the developed model demonstrates proven effectiveness in predicting FAME yields from operational conditions of homogeneous transesterification, establishing reliable quantitative relationships between process parameters and reaction efficiency. Nevertheless, it is important to acknowledge that, despite its high predictive performance, the study uses data compiled from multiple sources, which introduces uncontrolled variability between experiments. Differences in oil quality, catalyst purity, and operational conditions may generate noise in the data. Future research should validate the model with standardized experimental datasets to enhance its accuracy and industrial applicability. Furthermore, we intend to extend this machine learning approach to model the effects of surfactants and cosolvents as additives in biodiesel transesterification reactions. Previous studies by Simonelli et al. (2019) and Mascarenhas et al. (2024) have demonstrated the significant potential of these additives in optimizing biodiesel production, and their integration with ML-based modeling represents a promising avenue for improving process efficiency and yield under more complex reaction systems [70,71].

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biomass5040071/s1>, Supplement S1: Table S1: MATLAB simulation parameters for the Multilayer Perceptron (MLP)—Logistic Sigmoid model; Table S2: MATLAB simulation parameters for the Multilayer Perceptron (MLP)—Hyperbolic Tangent model; Table S3: MATLAB simulation parameters for the Radial Basis Function (RBF) network model; Table S4: MATLAB simulation parameters for the Hybrid RBF + MLP model; Table S5: MATLAB simulation parameters for the Random Forest model; Table S6: MATLAB simulation parameters for the Adaptive Neuro-Fuzzy System model; and Supplement S2: Table S7: Database composition and operating ranges for ML modeling.

Author Contributions: Writing—review and editing, V.L.d.S., G.S. and L.C.L.d.S.; writing—original draft, V.L.d.S., G.S. and L.C.L.d.S.; validation, V.L.d.S., G.S. and L.C.L.d.S.; methodology, V.L.d.S.; investigation, V.L.d.S.; data curation, V.L.d.S., G.S. and L.C.L.d.S.; conceptualization, V.L.d.S., G.S. and L.C.L.d.S.; supervision, G.S. and L.C.L.d.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001. The authors would like to acknowledge the financial support received from the National Council for Scientific and Technological Development—CNPq.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: All relevant data supporting the results of this article are available in the Supplementary Materials section and within the manuscript body.

Acknowledgments: During the preparation of this study, the authors used Generative Artificial Intelligence (ChatGPT-4, OpenAI) for the purposes of code structuring, debugging and syntax optimization. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

ML	Machine Learning
MLP	Multilayer Perceptron
RBF	Radial Basis Function
RF	Random Forest
ANFIS	Adaptive Neuro-Fuzzy Inference System
GA	Genetic Algorithm
FAME	Fatty Acid Methyl Esters
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
R	Coefficient of Correlation

References

1. Zhu, D. New Advances in Oil, Gas, and Geothermal Reservoirs. *Energies* **2023**, *16*, 477. [CrossRef]
2. Al-Fattah, S.M. Non-OPEC conventional oil: Production decline, supply outlook and key implications. *J. Pet. Sci. Eng.* **2020**, *189*, 107049. [CrossRef]
3. Miller, M.R.; Landrigan, P.J.; Arora, M.; Newby, D.E.; Münzel, T.; Kovacic, J.C. Environmentally Not So Friendly: Global Warming, Air Pollution, and Wildfires. *J. Am. Coll. Cardiol.* **2024**, *83*, 2291–2307. [CrossRef]
4. European Union. DIRECTIVE (EU) 2024/2881 of 23 October 2024 on Ambient Air Quality and Cleaner Air for Europe (Recast). *Official Journal of the European Union*. 2024. L 2024/2881. Available online: <http://data.europa.eu/eli/dir/2024/2881/oj> (accessed on 16 July 2025).
5. Canadian Council of Ministers of the Environment. *Canadian Ambient Air Quality Standards Handbook*; Canadian Council of Ministers of the Environment: Winnipeg, MB, Canada, 2025.

6. Conselho Nacional do Meio Ambiente (CONAMA). Resolution Conama No. 491, 19 November 2018. Establishes National Air Quality Standards and Guidelines. *Diário Oficial da União*. 19 November 2018; p. 155, Revised in June 2024. Available online: <https://www.siam.mg.gov.br/sla/download.pdf?idNorma=51160> (accessed on 16 July 2025).
7. Conselho Nacional de Política Energética (CNPE). Resolução N° 3, de 20 de Março de 2023. Altera a Resolução CNPE n° 16, de 29 de Outubro de 2018, Que Dispõe Sobre a Evolução da Adição Obrigatória de Biodiesel ao Óleo Diesel Vendido ao Consumidor Final. *Diário Oficial da União*. 20 March 2023, p. 2. Available online: <https://www.legisweb.com.br/legislacao/?id=443705> (accessed on 16 July 2025).
8. Presidência da República. LEI N° 13.576, de 26 de dezembro de 2017. Dispõe Sobre a Política Nacional de Bio-combustíveis (RenovaBio) e dá Outras Providências. *Diário Oficial da União*. 26 December 2017. Available online: <https://www2.camara.leg.br/legin/fed/lei/2017/lei-13576-26-dezembro-2017-786013-publicacaooriginal-154631-pl.html> (accessed on 16 July 2025).
9. Takase, M.; Zhao, T.; Zhang, M.; Chen, Y.; Liu, H.; Yang, L.; Wu, X. An expatiated review of neem, jatropha, rubber and karanja as multipurpose non-edible biodiesel resources and comparison of their fuel, engine and emission properties. *Renew. Sustain. Energy Rev.* **2015**, *43*, 495–520. [CrossRef]
10. Sáez-Bastante, J.; Pinzi, S.; Jiménez-Romero, F.J.; Luque de Castro, M.D.; Priego-Capote, F.; Dorado, M.P. Synthesis of biodiesel from castor oil: Silent versus sonicated methylation and energy studies. *Energy Convers. Manag.* **2015**, *96*, 561–567. [CrossRef]
11. Vilas Bôas, R.N.; Mendes, M.F. A review of biodiesel production from non-edible raw materials using the transesterification process with a focus on influence of feedstock composition and free fatty acids. *J. Chil. Chem. Soc.* **2022**, *67*, 5433–5444. [CrossRef]
12. Maheshwari, P.; Haider, M.B.; Yusuf, M.; Klemeš, J.J.; Bokhari, A.; Beg, M.; Al-Othman, A.; Kumar, R.; Jaiswal, A.K. A review on latest trends in cleaner biodiesel production: Role of feedstock, production methods, and catalysts. *J. Clean. Prod.* **2022**, *355*, 131588. [CrossRef]
13. Mohiddin, M.N.B.; Tan, Y.H.; Seow, Y.X.; Kandedo, J.; Mubarak, N.M.; Abdullah, M.O.; Chan, Y.S.; Khalid, M. Evaluation on feedstock, technologies, catalyst and reactor for sustainable biodiesel production: A review. *J. Ind. Eng. Chem.* **2021**, *98*, 60–81. [CrossRef]
14. Stanescu, R.-C.; Leahu, C.-I.; Soica, A. Aspects regarding the modelling and optimization of the transesterification process through temperature control of the chemical reactor. *Energies* **2023**, *16*, 2883. [CrossRef]
15. Shinde, P.P.; Shah, S. A Review of Machine Learning and Deep Learning Applications. In Proceedings of the Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, Pune, India, 16–18 August 2018.
16. Walsh, J.; Neupane, A.; Koirala, A.; Li, M.; Anderson, N. Review: The evolution of chemometrics coupled with near infrared spectroscopy for fruit quality evaluation. II. The rise of convolutional neural networks. *J. Near Infrared Spectrosc.* **2023**, *31*, 109–125. [CrossRef]
17. Imai, S.; Takekuma, Y.; Kashiwagi, H.; Miyai, T.; Kobayashi, M.; Iseki, K.; Sugawara, M. Validation of the usefulness of artificial neural networks for risk prediction of adverse drug reactions used for individual patients in clinical practice. *PLoS ONE* **2020**, *15*, e0236789. [CrossRef]
18. Çolak, A.B. Experimental Analysis with Specific Heat of Water-Based Zirconium Oxide Nanofluid on the Effect of Training Algorithm on Predictive Performance of Artificial Neural Network. *Heat Transf. Res.* **2021**, *52*, 67–93. [CrossRef]
19. Santana-Santos, L.; Kam, K.L.; Dittmann, D.; De Vito, S.; McCord, M.; Jamshidi, P.; Fowler, H.; Wang, X.; Aalsburg, A.M.; Brat, D.J.; et al. Validation of Whole Genome Methylation Profiling Classifier for Central Nervous System Tumors. *J. Mol. Diagn.* **2022**, *24*, 924–934. [CrossRef] [PubMed]
20. Ahmad, U.; Naqvi, S.R.; Ali, I.; Saleem, F.; Mehran, M.T.; Sikandar, U.; Juchelková, D. Biolubricant production from castor oil using iron oxide nanoparticles as an additive: Experimental, modelling and tribological assessment. *Fuel* **2022**, *324*, 124565. [CrossRef]
21. Shojaeefard, M.H.; Etghani, M.M.; Akbari, M.; Khalkhali, A.; Ghobadian, B. Artificial neural networks based prediction of performance and exhaust emissions in direct injection engine using castor oil biodiesel-diesel blends. *J. Renew. Sustain. Energy* **2012**, *4*, 063130. [CrossRef]
22. Yue, X.; Chen, Y.; Chang, G. Accurate modeling of biodiesel production from castor oil using ANFIS. *Energy Sources Part A Recovery Util. Environ. Eff.* **2018**, *40*, 432–438. [CrossRef]
23. Jana, D.K.; Bhattacharjee, S.; Roy, S.; Dostál, P.; Bej, B. The Optimization of Biodiesel Production from Waste Cooking Oil Catalyzed by Ostrich-Eggshell Derived CaO through Various Machine Learning Approaches. *Clean. Energy Syst.* **2022**, *3*, 100033. [CrossRef]
24. Kibar, M.E.; Hilal, L.; Çapa, B.T.; Bahçivanlar, B.; Abdeljelil, B.B. Assessment of homogeneous and heterogeneous catalysts in transesterification reaction: A mini review. *ChemBioEng Rev.* **2023**, *10*, 412–422. [CrossRef]
25. Ramezani, K.; Rowshanzamir, S.; Eikani, M.H. Castor Oil Transesterification Reaction: A Kinetic Study and Optimization of Parameters. *Energy* **2010**, *35*, 4142–4148. [CrossRef]
26. Haykin, S. *Redes Neurais—Princípios e Práticas*, 2nd ed.; Bookman: Porto Alegre, Brazil, 2007; pp. 27–59.

27. Riahi-Madvar, H.; Dehghani, M.; Seifi, A.; Salwana, E.; Shamshirband, S.; Mosavi, A.; Chau, K.W. Comparative analysis of soft computing techniques RBF, MLP, and ANFIS with MLR and MNLR for predicting grade-control scour hole geometry. *Eng. Appl. Comput. Fluid Mech.* **2019**, *13*, 529–550. [\[CrossRef\]](#)
28. Tao, J.; Yu, Z.; Zhang, R.; Gao, F. RBF neural network modeling approach using PCA based LM–GA optimization for coke furnace system. *Appl. Soft Comput.* **2021**, *111*, 107691. [\[CrossRef\]](#)
29. Aggarwal, C.C. Radial Basis Function Networks. In *Neural Networks and Deep Learning*; Springer International Publishing AG: Cham, Switzerland, 2018; pp. 217–233.
30. Liu, Y.; Wang, Y.; Zhang, J. New Machine Learning Algorithm: Random Forest. In Proceedings of the ICICA 2012, Chengde, China, 14–16 September 2012; Liu, B., Ma, M., Chang, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7473, pp. 246–252.
31. Li, H.; Lin, J.; Lei, X.; Wei, T. Compressive strength prediction of basalt fiber reinforced concrete via random forest algorithm. *Mater. Today Commun.* **2022**, *30*, 103117. [\[CrossRef\]](#)
32. Janardhana, K.; Sridhar, S.; Dixit, C.K.; Deivakani, M.; Tamilselvi, S.; Kaladgi, A.R.; Afzal, A.; Baig, M.A.A. ANFIS modeling of biodiesels' physical and engine characteristics: A review. *Heat Transf.* **2021**, *50*, 8052–8079. [\[CrossRef\]](#)
33. Karaboga, D.; Kaya, E. Adaptive network based fuzzy inference system (ANFIS) training approaches: A comprehensive survey. *Artif. Intell. Rev.* **2019**, *52*, 2263–2293. [\[CrossRef\]](#)
34. Ponkumar, G.; Jayaprakash, S.; Kanagarathinam, K. Advanced Machine Learning Techniques for Accurate Very-Short-Term Wind Power Forecasting in Wind Energy Systems Using Historical Data Analysis. *Energies* **2023**, *16*, 5459. [\[CrossRef\]](#)
35. Lambora, A.; Gupta, K.; Chopra, K. Genetic Algorithm- A Literature Review. In Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), Faridabad, India, 14–16 February 2019.
36. Osorio-González, C.S.; Gómez-Falcon, N.; Sandoval-Salas, F.; Saini, R.; Brar, S.K.; Ramírez, A.A. Production of Biodiesel from Castor Oil: A Review. *Energies* **2020**, *13*, 2467. [\[CrossRef\]](#)
37. Mamona. Available online: <https://www.embrapa.br/agencia-de-informacao-tecnologica/tematicas/agroenergia/biodiesel/materias-primas/mamona> (accessed on 2 August 2025).
38. Akhabue, C.E.; Okwundu, O.S. Monitoring the transesterification reaction of castor oil and methanol by ultraviolet visible spectroscopy. *Biofuels* **2017**, *10*, 729–736. [\[CrossRef\]](#)
39. Masango, S.B.; Ngema, P.T.; Olagunju, O.A.; Ramsuroop, S. The Effect of Reaction Temperature, Catalyst Concentration and Alcohol Ratio in the Production of Biodiesel from Raw and Purified Castor Oil. *Adv. Chem. Eng. Sci.* **2024**, *14*, 137–154. [\[CrossRef\]](#)
40. Elango, R.K.; Sathiasivan, K.; Muthukumaran, C.; Thangavelu, V.; Rajesh, M.; Tamilarasan, K. Transesterification of castor oil for biodiesel production: Process optimization and characterization. *Microchem. J.* **2019**, *145*, 1162–1168. [\[CrossRef\]](#)
41. Armendáriz, J.; Lapuerta, M.; Zavala, F.; García-Zambrano, E.; Ojeda, M.C. Evaluation of eleven genotypes of castor oil plant (*Ricinus communis* L.) for the production of biodiesel. *Ind. Crops Prod.* **2015**, *77*, 484–490. [\[CrossRef\]](#)
42. Banerjee, A.; Varshney, D.; Kumar, S.; Chaudhary, P.; Gupta, V.K. Biodiesel production from castor oil: ANN modeling and kinetic parameter estimation. *Int. J. Ind. Chem.* **2017**, *8*, 253–262. [\[CrossRef\]](#)
43. Canoira, L.; García Galeán, J.; Alcántara, R.; Lapuerta, M.; García-Contreras, R. Fatty acid methyl esters (FAMEs) from castor oil: Production process assessment and synergistic effects in its properties. *Renew. Energy* **2010**, *35*, 208–217. [\[CrossRef\]](#)
44. Das, M.; Sarkar, M.; Datta, A.; Santra, A. An experimental study on the combustion, performance and emission characteristics of a diesel engine fuelled with diesel-castor oil biodiesel blends. *Renew. Energy* **2018**, *119*, 174–184. [\[CrossRef\]](#)
45. Encinar, J.M.; González, J.F.; Martínez, G.; Sánchez, N.; González, C.G. Synthesis and characterization of biodiesel obtained from castor oil transesterification. *RE&PQJ* **2011**, *1*, 1078–1083. [\[CrossRef\]](#)
46. Soliman, A.; Ismail, A.R.; Khater, M.; Abu Amr, S.A.; El-Gendy, N.S.; Ezzat, A.A. Response surface optimization of a single-step castor oil-based biodiesel production process using a stator-rotor hydrodynamic cavitation reactor. *Environ. Sci. Pollut. Res.* **2024**, *31*, 60601–60618. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Hailegiorgis, S.M.; Hasraff, M.A.; Khan, S.N.; Ayoub, M. Methanolysis of Castor Oil and Parametric Optimization. *Procedia Eng.* **2016**, *148*, 546–552. [\[CrossRef\]](#)
48. Sánchez, N.; Sánchez, R.; Encinar, J.M.; González, J.F.; Martínez, G. Complete analysis of castor oil methanolysis to obtain biodiesel. *Fuel* **2015**, *147*, 95–99. [\[CrossRef\]](#)
49. Hiwot, T. Investigation of the Chemical Composition, Characterization and Determination of Energy Content for Renewable Energy Source (Biodiesel) Produced from Non-Edible Ethiopian Seeds' Particularly Castor Seed (*Ricinus communis*) Using Homogeneous Catalysis. *Int. Lett. Chem. Phys. Astron.* **2014**, *18*, 63–74. [\[CrossRef\]](#)
50. Hurtado, B.; Posadillo, A.; Luna, D.; Bautista, F.M.; Hidalgo, J.M.; Luna, C.; Calero, J.; Romero, A.A.; Estevez, R. Synthesis, Performance and Emission Quality Assessment of Ecodiesel from Castor Oil in Diesel/Biofuel/Alcohol Triple Blends in a Diesel Engine. *Catalysts* **2019**, *9*, 40. [\[CrossRef\]](#)
51. Keera, S.T.; El Sabagh, S.M.; Taman, A.R. Castor oil biodiesel production and optimization. *Egypt. J. Pet.* **2018**, *27*, 979–984. [\[CrossRef\]](#)

52. Khan, I.U.; Chen, H.; Yan, Z.; Chen, J. Extraction and Quality Evaluation of Biodiesel from Six Familiar Non-Edible Plants Seeds. *Processes* **2021**, *9*, 840. [\[CrossRef\]](#)
53. Lamiel, C.S.J.; Manocan, M.C.C.C.; Marasigan, G.P.M.; Dimaano, M.N.R. Optimization of Transesterification Parameters in *Ricinus communis* L. (Castor) Seed Oil for Biodiesel Production: Reaction Temperature Based at 70 °C. *Int. J. Eng. Res. Technol.* **2015**, *4*, 472–476.
54. Mahla, S.K.; Dhir, A.; Singla, V.; Rosha, P. Investigations on Environmental Emissions Characteristics of CI Engine Fuelled with Castor Biodiesel Blends. *J. Environ. Biol.* **2018**, *39*, 353–357. [\[CrossRef\]](#)
55. Meneghetti, S.M.P.; Meneghetti, M.R.; Wolf, C.R.; Silva, E.C.; Lima, G.E.S.; Silva, L.L.; Serra, T.M.; Cauduro, F.; Oliveira, L.G. Biodiesel from Castor Oil: A Comparison of Ethanolysis versus Methanolysis. *Energy Fuels* **2006**, *20*, 2262–2265. [\[CrossRef\]](#)
56. Najim, Y.H.; Al-Abdraba, W.M.S.; Ahmad, A.H. Effects of Temperature, Alkaline Catalysts and Molar Ratio of Alcohol to Oil on the Efficiency of Production Biodiesel from Castor Oil. *Kirkuk Univ. J.* **2016**, *11*, 56–69.
57. Nakarmi, A.; Joshi, S. A Study on Castor Oil and Its Conversion into Biodiesel by Transesterification Method. *Nepal J. Sci. Technol.* **2014**, *15*, 45–52. [\[CrossRef\]](#)
58. Pattnaik, S.; Mathur, B.; Desai, A.; Patel, A.; Chowdhury, P. Biodiesel Production from Non-Edible Castor and Sesame Oils via Homogeneous Transesterification: Comparative Physico-Chemical Evaluation. *Chem. Pap.* **2025**, *79*, 3951–3961. [\[CrossRef\]](#)
59. Peña, R.; Romero, R.; Martinez, S.L.; Ramos, M.J.; Martinez, A.; Natividad, R. Transesterification of Castor Oil: Effect of Catalyst and Co-Solvent. *Ind. Eng. Chem. Res.* **2009**, *48*, 1186–1189. [\[CrossRef\]](#)
60. Pradhan, S.; Saha, C. Transesterification and Reactive Extraction of Castor Oil for Synthesis of Biodiesel/Biolubricant. In Proceedings of the International Conference on Innovative Research on Renewable Energy Technologies, Malda, West Bengal, India, 25–27 February 2021; IOP Conf. Ser. Earth Environ. Sci. IOP Publishing: Philadelphia, PA, USA, 2021; Volume 785, p. 012005. [\[CrossRef\]](#)
61. Sánchez, N.; Encinar, J.M.; Nogales, S.; González, J.F. Biodiesel Production from Castor Oil by Two-Step Catalytic Transesterification: Optimization of the Process and Economic Assessment. *Catalysts* **2019**, *9*, 864. [\[CrossRef\]](#)
62. Setiadji, S.; Tanyela, T.; Sudiarti, T.; Prabowo, E.; Wahid, B. Alternatif Pembuatan Biodiesel Melalui Transesterifikasi Minyak Castor (*Ricinus communis*) Menggunakan Katalis Campuran Cangkang Telur Ayam dan Kaolin. *J. Kim. Val.* **2017**, *3*, 1–10.
63. Thakkar, K.; Kachhawaha, S.S.; Kodgire, P.; Keshav, M. Effectiveness of RSM Based Central Composite Design for Optimization of In-Situ Biodiesel Production Process from Castor Seeds. In Proceedings of the Advances in Thermal-Fluids Engineering (ATFE 2021), Gandhinagar, India, 25–26 March 2021; IOP Conf. Ser. Mater. Sci. Eng.. IOP Publishing: Philadelphia, PA, USA, 2021; Volume 1146, p. 012008. [\[CrossRef\]](#)
64. Thomas, T.P.; Birney, D.M.; Auld, D.L. Viscosity Reduction of Castor Oil Esters by the Addition of Diesel, Safflower Oil Esters and Additives. *Ind. Crops Prod.* **2012**, *36*, 267–270. [\[CrossRef\]](#)
65. Thomas, T.P.; Birney, D.M.; Auld, D.L. Optimizing Esterification of Safflower, Cottonseed, Castor and Used Cottonseed Oils. *Ind. Crops Prod.* **2013**, *41*, 102–106. [\[CrossRef\]](#)
66. Jin, X.; Li, S.; Ye, H.; Wang, J.; Wu, Y.; Zhang, D.; Ma, H.; Sun, F.; Pugazhendhi, A.; Xia, C. Investigation and Optimization of Biodiesel Production Based on Multiple Machine Learning Technologies. *Fuel* **2023**, *348*, 128546. [\[CrossRef\]](#)
67. Arunyanart, P.; Simasatitkul, L.; Juyploy, P.; Kotluklan, P.; Chanbunrung, J.; Seeyangnok, S. The Prediction of Biodiesel Production Yield from Transesterification of Vegetable Oils with Machine Learning. *Results Eng.* **2024**, *24*, 103236. [\[CrossRef\]](#)
68. Awogbemi, O.; Kallon, D.V.V. Application of machine learning technologies in biodiesel production process—A review. *Front. Energy Res.* **2023**, *11*, 1122638. [\[CrossRef\]](#)
69. Buasri, A.; Sirikoom, P.; Pattane, S.; Buachum, O.; Loryuenyong, V. Process optimization of biodiesel from used cooking oil in a microwave reactor: A case of machine learning and Box–Behnken design. *ChemEngineering* **2023**, *7*, 65. [\[CrossRef\]](#)
70. Simonelli, G.; Moraes, C.; Pires, C.A.M.; Santos, L.C.L. Multivariate study and optimization of biodiesel production using commercial surfactants. *Chem. Ind. Chem. Eng. Q.* **2019**, *25*, 183–192. [\[CrossRef\]](#)
71. Mascarenhas, N.O.; Pereira, M.A.; Pires, C.A.M.; Simonelli, G.; Santos, L.C.L. Production, optimization, and evaluation of thermal stability of palm oil biodiesel produced using a natural coconut oil-based surfactant. *Biomass Convers. Biorefinery* **2024**, *14*, 9455–9472. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.