





Review

Machine Learning Applications in Biofuels' Life Cycle: Soil, Feedstock, Production, Consumption, and Emissions

Iftikhar Ahmad ^{1,*}, Adil Sana ¹, Manabu Kano ², Izzat Iqbal Cheema ^{3,4}, Brenno C. Menezes ⁵, Junaid Shahzad ¹, Zahid Ullah ¹, Muzammil Khan ¹ and Asad Habib ⁶

¹ Department of Chemical and Materials Engineering, National University of Sciences and Technology, Islamabad 44000, Pakistan; asana.pse2scme@student.nust.edu.pk (A.S.); jshahzad.pse2scme@student.nust.edu.pk (J.S.); zahid.pse2scme@student.nust.edu.pk (Z.U.); mkhan.pse2scme@student.nust.edu.pk (M.K.)

² Department of Systems Science, Kyoto University, Kyoto 606-8501, Japan; manabu@human.sys.i.kyoto-u.ac.jp

³ Department of Chemical, Polymer and Composite Materials Engineering, University of Engineering and Technology, New Campus, Lahore 54890, Pakistan; izzatcheema@uet.edu.pk

⁴ Center for Energy Research and Development, University of Engineering and Technology, New Campus, Lahore 39021, Pakistan

⁵ Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha 34110, Qatar; BMenezes@hbku.edu.qa

⁶ Institute of Computing, Kohat University of Science and Technology, Kohat 26000, Pakistan; asadhabib@kust.edu.pk

* Correspondence: iftikhar.salarzai@scme.nust.edu.pk



Citation: Ahmad, I.; Sana, A.; Kano, M.; Cheema, I.I.; Menezes, B.C.; Shahzad, J.; Ullah, Z.; Khan, M.; Habib, A. Machine Learning Applications in Biofuels' Life Cycle: Soil, Feedstock, Production, Consumption, and Emissions. *Energies* **2021**, *14*, 5072. <https://doi.org/10.3390/en14165072>

Academic Editors: Lyes Bennamoun and Byong-Hun Jeon

Received: 3 May 2021

Accepted: 3 August 2021

Published: 18 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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/>).

Abstract: Machine Learning (ML) is one of the major driving forces behind the fourth industrial revolution. This study reviews the ML applications in the life cycle stages of biofuels, i.e., soil, feedstock, production, consumption, and emissions. ML applications in the soil stage were mostly used for satellite images of land to estimate the yield of biofuels or a suitability analysis of agricultural land. The existing literature have reported on the assessment of rheological properties of the feedstocks and their effect on the quality of biofuels. The ML applications in the production stage include estimation and optimization of quality, quantity, and process conditions. The fuel consumption and emissions stage include analysis of engine performance and estimation of emissions temperature and composition. This study identifies the following trends: the most dominant ML method, the stage of life cycle getting the most usage of ML, the type of data used for the development of the ML-based models, and the frequently used input and output variables for each stage. The findings of this article would be beneficial for academia and industry-related professionals involved in model development in different stages of biofuel's life cycle.

Keywords: bio-energy; artificial intelligence; industry 4.0; biodiesel; biogas; renewable energy; supply chain

1. Introduction

Machine Learning (ML) is one of the major forces driving the fourth industrial revolution, typically known as Industry 4.0. ML enables a computer system to solve complex research questions through implicit and automated “learning” that additionally self-improves without being explicitly preprogrammed to do so. From an algorithmic point of view, the term machine refers to an automated process that incrementally updates its problem-solving capability through successive iterations based on inputs from external variants. The concept of ML was introduced by Samuel [1], one of the pioneers of modern Artificial Intelligence (AI).

Machine Learning methods are broadly classified into supervised learning, unsupervised learning, and reinforcement learning. The supervised learning is performed on labeled output for a given set of input [2]. The trained algorithm is then used to predict

output for new data-sets. The supervised learning is applied both for classification and regression cases. It can be further categorized into decision tree learning, association rule learning, inductive logic learning and support vector machine. The dominantly used algorithms include linear regression, logistic regression, Neural Networks (NN), decision tree, Support Vector Machine (SVM), Random Forest (RF), naive Bayes, and k-nearest neighbor. The unsupervised learning does not require labeled output [3], and the method discovers patterns within the data. It is commonly applied for clustering instead of classification or regression. It can be further classified into clustering, similarity and metric learning, sparse dictionary learning, and Genetic Algorithms (GA). The reinforcement learning does not require a supervisor or network trainer [4]. It does not directly compare correct pattern output with the actual output. Unlike the supervised learning, the reinforcement algorithms rewards or punishes a network based on acceptability of its outcome. Further classes of the reinforcement learning include Bayesian networks, manifold learning and deep learning.

ML has been applied across almost all types of industrial processes for various purposes, such as data mining by Ge et al. [5], process automation by De et al. [6], process-aware attacks on industrial control systems by Keliris et al. [7], predictive maintenance by Kanawaday and San [8], optimization by [9], process monitoring and fault diagnosis by [10], and industrial tomography by Rymarczyk et al. [11]. Extensive reviews have been performed to analyze the state-of-the-art of ML applications in process industries, such as chemical and petrochemical by Min et al. [12], petroleum industry by Anifowose et al. [13], cement industry by Oey et al. [14], pharmaceutical industry by Guo et al. Unnikrishnan et al. [15], and iron and steel industry by [16]. The integration of ML in conventional industrial systems has brought about precision-based automation that subsequently resulted in the fabrication of high-quality products at a minimal cost in a shorter time. Within the process industry field, biofuel production in particular is getting prominence in the energy sector due to the renewable nature of its raw materials and environmental friendly products, i.e., biofuels.

ML is capable of playing a vital role in realizing the efficient and stable operation of biofuel production. The application of ML across all stages of the life cycle of biofuels, i.e., soil, feedstock, production, consumption, and emission, as shown in Figure 1, has been reported extensively in the literature. SewsynkerSukai et al. [17] and Ardabili et al. [18] presented an overview of ML and deep learning applications in biofuels. However, the literature lacks a review that discusses ML applications for all stages of the life cycle of biofuels. This article aims to fill this gap by presenting a systematic review that answers the following questions: (i) What are the dominant ML methods in the life cycle of biofuels? (ii) Which stage of the life cycle gets more usage of ML? (iii) What type of data is used for the development of the ML-based models? and (iv) What are the stage-wise trends for input and output variables? The findings of this article would be equally beneficial for academia and industry practitioners and theoreticians involved in the model development for different stages of the life cycle of biofuels.

This article is structured in four distinct sections. Section 2 presents the methodology adopted to perform the literature review and the approach used to obtain the citation data set. Section 3 documents the results and discussions of the review. Meanwhile, conclusions and future prospects are outlined in Section 4.

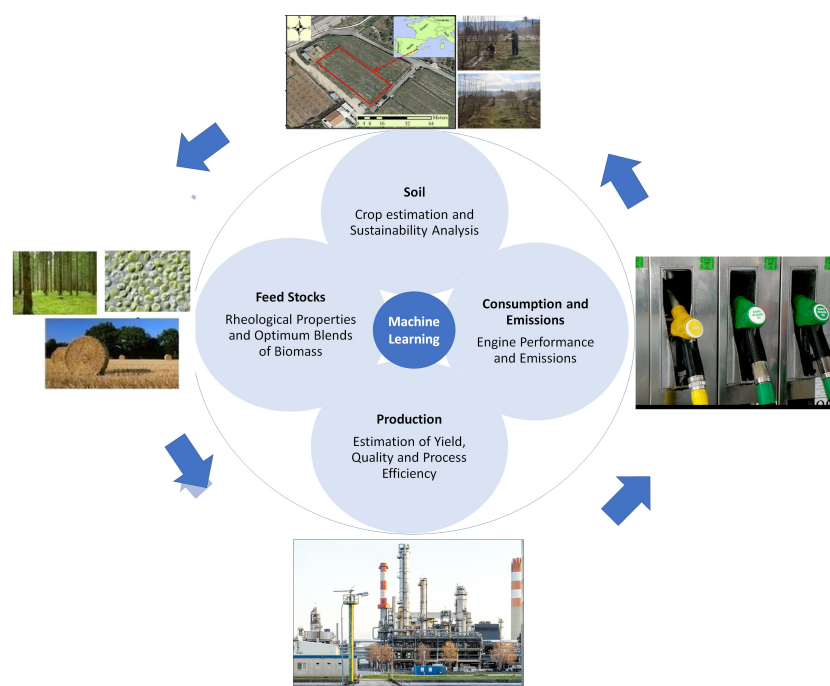


Figure 1. Machine Learning applications in the life cycle of biofuels.

2. Methodology

For literature retrieval, the Web of Science and Google Scholar databases were mainly used. To document the literature in a logical and succinct way accommodating in a review paper, only research articles were reviewed. The dominant part of the literature was collected from the Web of Science database; however, Google Scholar was also used for collecting some gray literature regarding ML applications in soil stage and feedstock stages. Keywords listed in Table 1 were used to search relevant literature in the Web of Science database.

Table 1. Keywords used in literature survey.

Categories	Keywords Used for Search in the Databases
Machine Learning	artificial neural networks boosting data-based data-driven decision tree deep learning dimensionality reduction algorithms discriminant analysis ensemble learning estimation extreme learning machine genetic algorithm inference kNN K-Means least-squares logistic regression linear regression machine learning moving average multi layered perception Naive Bayes neuro fuzzy partial least squares principal component analysis prediction random forest(s) soft sensor support vector machine virtual sensor
Biofuels	bioalcohol biodiesel bioethers biofuels biogas biohydrogen dimethylfuran green diesel
Soil	drone land land image satellite soil surveillance
Feedstock	algae and aquatic biomass biomass biosolid(s) corn energy cane feed feedstock forest thinning high biomass sorghum hybrid poplars industrial waste gases logging residues lignocellulosic crops lignocellulosic residues manure slurries micro algae miscanthus municipal waste oil-based residues oil crops organic residues plant plastics raw material(s) shrub willows sludge starch crops sugar crops sweet sorghum switch grass vegetable oil waste waste food waste gases
Production	catalytic synthesis distillation drying fermentation gas cleaning gasification operation process product production reactor refining unit water gas shift yield
Consumption & emissions	air pollution carbon emission emission energy potential engine environment exhaust gases fuel consumption fuel quality fuel use green house gases greenhouse mileage

The following rules were applied while deciding the relevance of the articles: (i) the title of the article should contain at least one word from each of the ML and biofuels categories of keywords as shown in Table 1, and (ii) the title should also contain at least one word from any of the other four categories, i.e., soil, feedstock, production, consumption,

and emissions. The collected literature was divided into four stages of the life cycle on the basis of keywords found in the second step of the above-mentioned procedure.

In the first round, the number of papers related to soil, feed stock, production, and consumption and emission stages were 10, 91, 90, and 75, respectively. The collected literature was initially studied to screen-out the irrelevant papers. Some papers were shifted across the stages. After the screening, the production stage contained the largest number of papers with 64 papers on ML applications followed by consumption and emission with 43 papers, feedstock with 12 papers, and soil with 5 papers.

3. Applications of ML Methods in the Life Cycle of Biofuels

In this section, the stage-wise application of Machine Learning (ML) methods is discussed. This section is divided into the following subsections: Section 3.1 Soil, Section 3.2 Feedstock, Section 3.3 Production and Section 3.4 Consumption, Engine Performance and Emission Stages. A summary of applications is presented in Section 3.5.

3.1. Soil

Several studies on ML application in the soil stage of the life cycle of biofuels have been reported both at the tree and plot levels. The ML applications in soil phase are summarized in the Table 2 and discussed below.

Table 2. Summary of ML applications in the soil phase.

ML Method	Input Variables	Output Variables	Error Range	References
Linear Mixed-Effects (LME) regression, Random Forest (RF), Support Vector Regression (SVR)	Tree Crowns	Biomass estimation		[19]
Extremely Randomized Trees (ERT), Random Forest (RF) model	Average precipitation, temperature, solar radiation, atmospheric CO ₂ , wind speed	Sorghum biomass yield		[20]
Partial Least Square Discriminant Analysis (PLS-DA), Principal Component Analysis Discriminant Analysis (PCA-DA), Neural Network (NN), Random Forest (RF), Support Vector Machine (SVM) with Linear Classifier (SVML), SVM with Nonlinear Kernel (SVML _G), SVM with Radial basis Kernel (SVM _R), SVM with Radial basis Kernel with Polynomial basis Kernel (SVM _P), eXtreme Gradient Boosting-xgbtree method (GBT), eXtreme Gradient Boosting-xgb DART method (GBD), eXtreme Gradient Boosting-xgb Linear method (GBL), simple linear model (LM)	Saponification value, iodine value and the poly unsaturated fatty acids content of feedstock	Sorghum biomass yield	R ² ≥ 0.50	[21]
Boosted Regression Tree (BRT)	Climate, soil characteristics, farming practices	Future life-cycle environmental impacts of corn production	0.7800 ≤ R ² ≤ 0.8200	[22]
Random Forest Model (RFM)	Average temperature, average precipitation, slope, soil characteristics, diurnal temperature	Crop yield	0.8300 ≤ R ² ≤ 0.9000	[23]

For example, Gleason and Im [19] compared the Linear Mixed-Effects (LME) regression, Support Vector Regression (SVR), Random Forest (RF), and Cubist for prediction of biomass in a 40% to 60% canopy closure forest. SVR outperformed the other methods. Huntington et al. [20] used RF to predict future trends in sorghum bicolor yield under two irrigation regimes and four Greenhouse Gas (GHG) emission scenarios. The RF model trained on uniquely identified data, identified by year and country, achieved reasonable prediction accuracy. In another study by Habyarimana [21], various ML methods were used for the prediction of sorghum biomass yield based on satellite images of sorghum

fields. The ML methods included PLS Discriminant Analysis (PLS-DA), PCA Discriminant Analysis (PCA-DA), ANN, RF, SVM with Nonlinear Kernel (SVM-G), SVM with Radial basis Kernel (SVM-R), SVM with Radial basis Kernel with Polynomial basis Kernel (SVM-P), SVM with Linear Classifier (SVM), eXtreme Gradient Boosting-xgbtree method (GBT), eXtreme Gradient Boosting-xgbLinear method (GBL), eXtreme Gradient Boosting-xgb DART method (GBD), and a simple linear model. The GBT method outperformed the other methods. Lee et al. [22] used the Boosted Regression Tree (BRT) model to assess environmental impacts of corn production for the years 2022 to 2100. The study was conducted in the context of four emissions scenarios where the BRT model achieved correlation coefficients of 0.82 and 0.78, in estimating eutrophication impacts and global warming, respectively. Yang et al. [23] used a two-step ML approach. Initially, the Gaussian Process Model (GPM) was used for crop yield down-scaling followed by yield estimation through the RF model. The GPM, a Bayesian inference method, helped in realizing accurate estimations.

3.2. Feedstock

ML applications in the feedstock stage of biofuels have received attention in the literature, mostly recently. The ML applications in feedstock phase are summarized in Table 3 and discussed below.

Mahanty et al. [24] used ANN and statistical regression models to predict specific methane yield in the production of biogas from industrial sludges. The ANN model performed better than the statistical regression model. It was revealed that sludges from the chemical industry have a relatively higher impact on methane in the produced biogas. Mairizal et al. [25] used multiple linear regressions to predict viscosity, Flash Point (FP), density, higher heating value, and oxidative stability of biodiesel produced from sunflower oil, peanut oil, hydrogenated coconut oil, hydrogenated copra oil, beef tallow, rapeseed oil, and walnut oil. It was inferred from the results that the addition of PU/MU as an independent parameter increase prediction performance. Giwa et al. [26] used ANN to estimate Cetane Number (CN), Kinematic Viscosity (KV), Flash Point (FP), and density of biodiesel produced from fatty acid. Accuracy of estimation and the average of absolute deviation of the model were as follows: CN (96.6%; 1.637%), FP (99.07%; 0.997%), KV (95.80%; 1.638%), and density (99.40%; 0.101%). Tchameni et al. [27] used Multiple Non-Linear Regression (MNL) and ANN for prediction of rheological properties of waste vegetable oil for production of biodiesel. The ANN model had superior performance over the MNL method. Dahunsi [28] used single and multiple linear regressions to estimate methane yield in biomass structural components. A fairly high correlation was found between the chemical composition and methane potentials of the biomass.

Reimann et al. [29] used ML methods such as Naive Bayes, RF and ANN for the classification of micro-algae cells, with RF outperforming the other ML techniques. It was inferred that pairing the RF based modeling framework with microscopic features of samples may result with high-resolution distinction and quantification of different species within a lesser time frame when compared with the conventional lab based approach. Tang et al. [30] applied MLR and RF for the prediction of the yield and hydrogen of bio-oil using information of biomass compositions and pyrolysis conditions. The results verified that RF has a better performance when compared with MLR. Shahbeig et al. [31] developed an SVR-based model to predict the thermal characteristics of biomasses. The predicted results were aligned with experimental findings with a correlation coefficient of 0.9999. Ighalo et al. [32] developed LRA and Stochastic Gradient Descent (SGD) based models to predict the Higher Heating Value (HHV) of biomass. The LRA model was observed to be more accurate. Kumar et al. [33] developed Artificial Neural Network (ANN) coupled with Genetic Algorithm (GA) to increase the lipid yield. The input parameters were glycerol, NH_4Cl , MgSO_4 and KH_2PO_4 , which were screened by Plackett Burman design. The obtained value of regression correlation coefficient value was 0.9918. Maximum biomass concentration of $15.16 \pm 0.69 \text{ g L}^{-1}$ was achieved with $0.49 \pm 0.02 \text{ g}$ lipid per gram of yeast. By examining the lipid composition, the main fatty acids revealed, in order of

their relative richness (% *w/w*), were oleic, tridecanoic, palmitic, stearic, palmitoleic and linolelagic acids. Cheng et al. [34] developed RF model for biochar production through slow pyrolysis process. Output of the model were yields and quality of biochar produced from slow pyrolysis from different feedstocks. The feedstock compositions, reaction temperature, resistance time, and heating rate were used as the model inputs. The model outputs were used with Life Cycle Assessment (LCA) and economic analysis to find net Global Warming Potential (GWP), Energy Return on Investment (EROI), and Minimum Product Selling Price (MPSP). Cheng et al. [35] developed MLR, regression tree (RT), and RF models to predict yields and characteristics of products (biocrude, hydrochar, gas, and aqueous co-product) from Hydrothermal Treatment (HTT) of various feedstocks. Feedstocks' characteristics together with reaction temperature, reaction time, and initial concentration were used as inputs of the models. The model outputs were used with LCA and economic analysis to find net GWP, and energy return on investment EROI.

Table 3. Summary of ML applications in the feedstock phase.

ML Method	Input Variables	Output Variables	Error Range	References
ANN, Statistical Regression Model (SRM)	Fractions of sluges of paper, chemical, petrochemical, automobile, food industries	Specific methane yield	$0.7300 \leq R^2 \leq 0.9900$	[24]
Multiple Linear Regressions	Saponification value, iodine value and the polyunsaturated fatty acids content of feedstock	Biodiesel's viscosity, density, Flash Point (FP), Higher Heating Value (HHV) and oxidative stability	$R^2 = 0.9900$	[25]
ANN	Palmitic, stearic, oleic, linoleic, linolenic acids, temperature	Cetane Number (CN), Flash Point (FP), Kinematic Viscosity (KV) and density of biodiesel	$R = 0.958$	[26]
ANN, MNLR	Biodiesel content, aging temperature	Apparent viscosity, plastic viscosity and yield	$0.9960 \leq R^2 \leq 0.9970$	[27]
Single and Multiple Regression Model	Chemical composition of the biomass	Methane yield	$R^2 = 0.6300$	[28]
Naive Bayes, RF, ANN	Microscopic features of samples of microalgae cells	Classification of microalgae cells	Correlation coefficient = 0.9950	[29]
MLR, RF	Biomass compositions, pyrolysis conditions	Yield, hydrogen of bio-oil	$0.1660 \leq R^2 \leq 0.9200$	[30]
SVR-based model	Bacterial biomass, dimensionality of differential thermogravimetric (DTG)	Thermal characteristics of biomasses: enthalpy change, Gibb's free energy, entropy change and high heating value	$R^2 = 0.9999$	[31]
LRA and stochastic gradient descent (SGD)	78 lines of combined proximate and ultimate analysis data	High heating value of biomass	$R^2 = 0.9999$	[32]
ANN, GA	Glycerol, NH_4Cl , MgSO_4 , KH_2PO_4	Lipid yield		[33]
RF	The feedstock compositions, reaction temperature, resistance time, and heating rate	yield and quality	$0.7800 \leq R^2 \leq 0.8700$	[34]
MLR, regression tree (RT), and RF	Feedstocks' characteristics, reaction temperature, reaction time, and initial concentration	Biocrude, hydrochar, gas, and aqueous co-product	$0.1600 \leq R^2 \leq 0.9000$	[35]

3.3. Production

Several studies have been reported for the prediction and optimization of quality, yield (quantity), and process conditions, i.e., pressure, temperature, flow rate, etc., in the production stage of the life cycle of biofuels. These studies are categorized based on the type of biofuel produced, such as biodiesel, biogas, biohydrogen, and other miscellaneous cases. Trends in ML applications in the production stage are summarized in Table 4 and discussed in the following subsections.

Table 4. Summary of ML applications in the production phase. (The numbers in round brackets indicate the number of papers for the preceding keywords).

Production Process	Purpose	ML Method Ranking	Input Variables Ranking	Error Range	References
Biodiesel	Quality estimation	ANN (2), Least Squares Boosting (LSBoost) integrated with polynomial chaos expansion (PCE) method, regression models, principle component analysis	Reaction temperature (4), reaction time (2), metal ratio, and calcination temperature, flow rate, pressure, reactor residence time, reflux rate, oil fraction, methanol-to-oil molar ratio, catalyst concentration	$0.5960 \leq R^2 \leq 0.9976$	[36–40]
	Yield estimation	ANN (11), ANFIS (2), Linear Regression (LR)	Temperature (11), methanol-to-oil molar ratio (10), catalyst concentration (9), reaction time (7), organic loading rate, influent–effluent pH, H ₂ SO ₄ concentration, total volatile fatty acid (VFA) of the effluent, xylose, influent–effluent alkalinity, initial pH, pressure, reactor diameter, liquid height and ultrasound intensity	$0.3500 \leq R^2 \leq 0.9978$	[41–51]
	Quality and yield optimization	ANN-GA (9), genetic algorithm-based support vector machines (3), ANFIS-GA (2), multi-objective genetic algorithm (2), multivariate regression analysis, ELM-RSM, GA-LSSVM, HGAPSO-LSSVM, multi-objective optimization with Orthogonal collocation on finite elements (OCFE) method	Methanol-to-oil molar ratio (15), reaction temperature (14), reaction time (13), stirring speed (6), catalyst concentration (6), catalyst weight (5), dosage of NaOH catalyst (3), humidity (3), impurities (3), mixing time (3), Free Fatty Acid (FFA) content, sulfuric acid-to-rice bran ratio, duty-cycle, methanol-to-rice bran ratio, FAME concentration, overall heat duty, initial acid value of vegetable oil, calcination temperature, reactor’s residence time, and pressure	$0.8690 \leq R^2 \leq 0.9999$	[52–67]
	Estimation and optimization of process conditions and efficiency	ANN (3), multi-objective optimization program with genetic algorithm (2), ANFIS, Artificial Linear Interdependent Fuzzy Multi-Objective Optimization (ALFIMO), Multivariate Curve Resolution Alternative Least Square (MCR-ALS), Group Interaction Parameters (GIP) with GA, and Extreme Learning Machine (ELM) with Lyapunov analysis	Reaction temperature (3), concentration (2), water content and reaction time (2), methanol-to-oil molar ratio (2), residence time, vapor pressure, heat capacity, liquid molar volume, and liquid viscosity, ethanol-to-oil molar ratio, heat of vaporization, initial CO ₂ pressure, pH, reaction time, reaction temperature, metal ratio, heat of formation, calcination temperature, and phase equilibrium	$0.6580 \leq R^2 \leq 0.9990$	[36,68–78]

Table 4. Cont.

Production Process	Purpose	ML Method Ranking	Input Variables Ranking	Error Range	References
Biogas	Quality estimation	ANN (2), ANFIS, and multiple regression model	total solids, fixed solids, volatile solids, Volatile fatty acids (VFAs), pH, inflow rate, COD, ammonia (NH ₄), pH, total dissolved solids (TDS), total Kjeldahl nitrogen (TKN), alkalinity (Alk), chloride (Cl), conductivity (Cond), and total phosphorus (TP)	$0.7500 \leq R^2 \leq 0.9200$	[79–81]
	Yield estimation	ANN (5), Multiple Non-Linear Regression (MNL) models, and Partial Least Squares Regression (PLS-R)	pH (3), temperature (2), Total Volatile Solids (TVS) (2), Volatile Fatty Acids (VFAs) (2), composition, time, Moisture Content (MC), and CH ₄ , total Kjeldahl nitrogen, total Chemical Oxygen Demand (COD), total phosphorus, hydraulic retention time, ammonium, alkalinity, instrument measurements (FFT acoustic spectra), and Total Solids (TS)	$0.8700 \leq R^2 \leq 0.9983$	[82–86]
	Estimation of process performance, and optimization of quality and yield	ANN-GA (5), ANFIS	Total Solids (TS) (2), pH (2), temperature (2) digestion time and C-to-N ratio (2), cow dung (2), retention time, Total Volatile Solids (TVS), organic loading rate, duty cycle, mass of poultry droppings, plantain peel, piggery waste, stirring intensity of substrates, paper waste, banana stem, saw dust, and rice bran	$0.8700 \leq R^2 \leq 0.9900$	[86–92]
Biohydrogen	Yield estimation	ANN (3), hybrid fuzzy clustering-ranking approach with ANN, gray model	Biomass concentrations (5), pH (4), substrate (2), temperature (2), time, agitation speed and flow rate	$0.7500 \leq R^2 \leq 0.9999$	[93–97]
Miscellaneous	Bioethanol yield estimation and optimization	Fuzzy Neural Network (FNN) and PSO	Temperature, glucose content, and fermentation time	$R^2 = 0.9900$	[98]
	Bisabolene yield estimation and optimization	Convolutional neural networks-based multi-objective optimization with hybrid stochastic search optimization algorithm: random search (RS), PSO and SA	Incident light intensity, recycling gas flow rate, number of holes, diameter of holes, and cardinal coordinates of sample	mean error < 1%	[99]

3.3.1. Biodiesel

The studies reported in the biodiesel production phase are further classified based on the output of the ML models, i.e., quality, yield, and process efficiency.

Quality Estimation

Soltani et al. [36] used the ANN model to determine optimum conditions to get the desired nanocrystalline size of mesoporous SO_3HZnO catalyst. Optimized conditions were calcine temperature of $700\text{ }^\circ\text{C}$, $160\text{ }^\circ\text{C}$ reaction temperature, 18 min reaction time, and 4 mmol of Zn concentration. Ahmad et al. [37] used Least Squares Boosting (LSBoost) integrated with polynomial chaos expansion method in the production of vegetable oil based biodiesel under uncertainty. The average Mean Absolute Deviation Percent (MADP) values in the predicted values of the target output were 0.84 in response to 1% uncertainty in each input variable of the models. Gulum et al. [38] used regression and ANN models to predict viscosity and density of ternary blends consisting of biodiesel, diesel, and vegetable oil. Exponential and rational models previously reported in the literature were compared with regression models and the ANN approach. Tomaz-zoni [39] used PCA to estimate viscosity, relative density, and percentage conversion of vegetable oil to methyl esters in the production of diesel from vegetable oil. Through the use of PCA, they were able to differentiate between pure samples waste oil, diesel, and biodiesel from their respective blends. Sarve et al. [40] used ANN and Central Composite Design (CCD) to predict Fatty Acid Methyl Ester (FAME) content in the production of biodiesel from sesame oil. The study revealed that catalyst concentration has the highest impact on the FAME contents in the final product. The ANN model showed better performance.

Yield Estimation

Several studies based on biodiesel yield prediction through ML methods have been reported in the literature where the biodiesel was formed from jatropha-algae, castor oil, and anaerobic sludge, e.g., [41–43]. Kumar et al. [41] developed an ANN model to predict biodiesel yield using various jatropha-algae oil blends as inputs. Banerjee et al. [42] used an ANN model for predicting the fractional formation of FAME. They also devised a kinetic model using the experimental and computed data. The experimental and the ANN-based predicted data were used to estimate the rate constants of the kinetic model. The ANN model was able to predict the % FAME yield within 8% deviation. Kanat and Saral [43] used ANN to estimate the production rate of biodiesel from anaerobic sludge in a thermophilic up-flow anaerobic sludge blanket reactor. Longer time periods for the moving average showed a higher correlation coefficient of 0.927.

In several studies, ANN was used together with Response Surface Methodology (RSM) for the prediction of the yield of biodiesel from: Jatropha-algae oil by Kumar et al. [44], goat tallow by Chakraborty and Sahu [45], and enterobacter species by Pandu et al. [46]. In a study by Kumar et al. [44], ANN outperformed the RSM. Chakraborty and Sahu et al. [45] used RSM and ANN for identifying optimal parametric values that result in maximum FA conversion while maintaining FAME yield that met the American Society for Testing and Materials (ASTM) biodiesel specifications. ANN and RSM had comparable predictability performance. Pandu et al. [46] compared performance of RSM model and ANN model. The ANN model outperformed the RSM model.

Kumar et al. [47] used ANN and Linear Regression (LR) to predict soybean-based biodiesel yield where the ANN outperformed the LR. Moradi et al. [48] used ANN and kinetic models to estimate the yield of biodiesel from Soybean oil. The ANN exhibited the capability of learning from experimental data and is simple to apply in comparison to the classic kinetic modeling method. Guo and Baghban [49] and Mostafaei et al. [50] used Adaptive Neuro-Fuzzy Interference System (ANFIS) to estimate biodiesel yield from algae oil blend, vegetable oil, and waste cooking oil. They achieved a low absolute deviation between experimental data and ANFIS model based predicted data. It confirmed suitability of ANFIS for prediction of the biodiesel yield from vegetable oil. Maran and Priya [51]

used ANN and RSM to predict biodiesel yield from muskmelon oil. The ANN model, again, outperformed the RSM model.

Quality and Yield Optimization

Several studies have focused on the optimization of the yield and quality of biodiesel. Bobadilla et al. [52] used a GA-based SVM to estimate and optimize biodiesel yield of specific properties such as higher heating value with decreased viscosity, density, and turbidity from waste cooking oil. To produce biodiesel of high quality, the optimum inputs were dosage of catalyst (NaOH) from 1.00 to 1.38 wt%, molar ratio from 6.0 wt% to 8.4, mixing speed from 500 to 999.99 rpm, time from 20.00 to 26.94 min, temperature from 28.75 to 37.5 °C, humidity from 0 to 2.31 wt%, and impurities from 0 to 2.99 wt%. Cheng et al. [53] used a GA-based evolutionary SVM to predict and optimize the final acid value of oil in the production of biodiesel from rice bran. They found GA-ESVM better than ANN-GA and SVM. Sivamani et al. [54] used RSM and ANN coupled with GA to predict the yield of Simarouba giauca biodiesel. For higher yield, the optimum values for oil-to-alcohol ratio, temperature, and duration were found to be 1:6.22, 677.25 °C, and 20 h, respectively. Ighose et al. [55] used an ANFIS coupled with RSM and GA to realize high yield of biodiesel from *Thevetia peruviana* seed oil via the transesterification process. The ANFIS outperformed the RSM model. In addition, the use of GA resulted in a higher yield than RSM in relatively less time and less catalyst loading.

Dhingra et al. [56] used ANN and GA to predict and optimize yield in the production of polanga oil based biodiesel. They combined the ANN, RSM, and GA for predicting the optimized reaction conditions which resulted in a biodiesel yield of 91.08% by weight significantly higher than 78.8% obtained through RSM alone. Ishola et al. [57] used ANN, ANFIS, and GA to estimate and optimize biodiesel yield (methyl esters) from sorrel oil. The ANFIS model outperformed the ANN model while RSM was the lowest performer in terms of prediction accuracy. In addition, GA also outperformed the RSM and obtained the highest biodiesel yield (methyl esters) of 99.71 wt% at the methanol-to-oil molar ratio, the catalyst weight and reaction time of 8:1, 1.23 wt%, and 43 min, respectively. Silitonga et al. [58] used ANN integrated with ant colony optimization to determine the minimum acid value and maximum biodiesel yield from cerberamanghas. For esterification, the optimum methanol-to-oil molar ratio was 10.5:1 while the best values for reaction time and reaction temperature were 71 min and 54.5 °C, respectively. The optimized catalyst concentration, reaction temperature, methanol-to-oil molar ratio, stirring speed, and reaction time for transesterification were 1.1 wt%, 55 °C, 10.9:1, 1100 rpm, and 72 min, respectively.

Chakraborty et al. [59] used multivariate regression analysis to predict optimum operating conditions for mustard oil (MO)-based biodiesel yield. Optimal values of methanol-to-MO molar ratio, calcination temperature, catalyst concentration, and stirrer speed, were 13.13:1, 950 °C, 3.44 wt%, and 890 rpm, respectively. Goharimanesh et al. [60] used multi-objective GA to obtain optimum reaction temperature for maximizing biodiesel production, amount of ester, and alcohol.

Oladipo et al. [61] used CCD and ANN to maximize FAME content in the production of biodiesel from crude neem, jatropha, and waste cooking oils. They also found that safe reuse of the mesoporous catalyst can be carried out up-to five cycles. Rajendra et al. [62] used ANN, GA, and central composite in rotatable design to optimize the final acid value of oil in the production of biodiesel from jatropha, simarouba, mahua, and rice bran oils. ANN-GA helped in determining optimum process conditions to obtain high yield.

Fahimi and Cremaschi [63] used ANN to predict virgin oil and methanol-based biodiesel yield. An optimization model was used to determine the minimized net present sink in synthesis of biodiesel. The models for unit operation, thermodynamics, and mixing were replaced by the surrogate models to reduce computational load. Betiku et al. [64] used CCD, ANN, and RSM to determine high biodiesel yield from neem seed oil. The acid value of NO was significantly reduced by one-step esterification. Optimization of transesterification of pretreated NO using KOH as catalyst resulted in 99% yield of biodiesel. Zhang and

Niu [65] used LS-SVM with GA to estimate and optimize biodiesel yield from castor oil. Based on high accuracy in prediction, the use of the LS-SVM model was recommended for efficient prediction in the process. Optimum values of catalyst weight was 13 g, MOR at 625, the temperature at 4060 °C, and time of 1240 min. Mujtaba et al. [66] used ELM and RSM together with cuckoo search algorithm to find best cold flow and lubricity characteristics of biodiesel produced from the palm-sesame oil blend. Bemani et al. [67] developed LSSVM model for estimation of the cetane number of biodiesel. The LSSVM was coupled with GA, PSO and hybrid of GA and PSO (HGAPSO) algorithms for the process optimization.

Estimation and Optimization of Process Conditions and Efficiency

Karimi et al. [68] used RSM and ANN to estimate FAME content and exergetic efficiency in waste cooking oil-based production of biodiesel. The method performed well in achieving high quality and exergetic efficiency of the process by optimizing the input variables. Reaction time, immobilized lipase, concentrations of water, and concentrations of methanol were the design variables. The catalyst concentration of 35%, the water content of 12%, methanol to WCO molar ratio of 6.7 and reaction time of 20 h achieved FAME content and exergy efficiency of 86% and 80.1%, respectively. Aghbashlo et al. [69] used ANFIS with GA and linear interdependent fuzzy multi-objective optimization to predict Functional Exergy Efficiency (FEE), Normalized Exergy Destruction (NED), Universal Exergy Efficiency (UEE), and Conversion Efficiency (CE) in production of biodiesel. Optimum values for transesterification temperature, residence time, and methanol-to-oil molar ratio were found to be 60 °C, 10 min, and 6.20, respectively. Patle et al. [70] used multi-objective GA optimization to estimate heat duty, profit, and organic waste in the palm waste cooking oil based biodiesel production. Waste cooking oil flow rate was the factor affecting heat duty, profit, and organic waste. Shukri et al. [71] used ANN to predict pressure in-cylinder in a bar, heat release in percentage, volume generated, and thermal efficiency in percentage in palm oil methyl ester blends-based biodiesel production. Biodiesel 10% blend (B10) was found to be more efficient due to the high heating value and cetane number. Sarve et al. [72] used ANN to estimate ethanol-to-oil molar ratio, temperature, reaction time, and initial CO₂ pressure in mahua oil based production of biodiesel. Sensitivity analysis of the ANN model was performed where the temperature was found to be the most effective variable followed by reaction time, ethanol-to-oil molar ratio, and initial CO₂ pressure. ANN outperformed the RSM model both in data fitting and prediction accuracy. Kuen et al. [73] applied an automatic tune control scheme consisting of Recursive Least Squares (RLS) and Internal Model Control (IMC) integration to get optimized values of the adaptive controller parameters for the biodiesel transesterification reactor. For introduced disturbance of 5% rise in the reactor temperature and concentration loops from nominal values, in comparison to conventional PID controllers, adaptive controllers' response time was much faster, i.e., 370 s and 380 s, respectively.

Rouchi et al. [74] used a Multivariate Curve Resolution Alternative Least Square (MCR-ALS) for interpretation and control of the reaction towards the desired route. For the said purpose, the number of components, concentration profiles, spectral, and reaction kinetics were evaluated for Soybean oil-based biodiesel with reagents consisting of methanol and NaOH. The correlation coefficient and standard deviation of residuals were 0.99992 and 0.00765, respectively, which showed an advantage of MCR-ALS. Lopez-Zapata et al. [75] used virtual sensors based on the extended Kalman filter to estimate concentrations of triglycerides, monoglycerides, methyl ester, diglycerides, glycerol and alcohol in jatropha oil-based production of biodiesel. The method had the potentials of real-time implementation because it needed only a few measured variables, such as temperature and pH. Nicola et al. [76] used multi-objective GA optimization to realize maximum purity of important compounds and minimum energy requirements in the production of vegetable oil-based biodiesel by two processes. The specific energy consumptions for process schemes were 2.7 MJ/kg and 1.5 MJ/kg that met the required standards.

Fahmi and Cremaschi [63] used ANN as a surrogate model to identify the superstructure and operating conditions which minimized the net present sink in the production of biodiesel. ANN was used as a surrogate model which resulted in a less complex model with an efficient representation of the process synthesis. Soltani et al. [36] used ANN to model nanocrystalline-sized mesoporous zinc oxide (SO₃H-ZnO) catalyst for the efficient production of biodiesel from palm fatty acid distillate-based production of biodiesel. The prediction error was within an acceptable range of 2.73%. Noriega and Narvaez [77] used Group Interaction Parameters (GIP) to predict Liquid-Liquid Equilibrium (LLE) in the vegetable oil based production of biodiesel. The most influential variable on LLE was the overall mass fraction followed by the length of alcohol chain. Wong and Wong [78] used Extreme Learning Machines (ELM) with Lyapunov analysis to predict the Air-to-Fuel Ratio (AFR) in the production of biodiesel from biofuel blends. The proposed approach resulted in effectively regulating Air-Fuel Ratio (AFR) to the desired level. The control strategy outperformed the engine built-in AFR controller and was highly recommended for dual-injection engines.

3.3.2. Biogas

Quality Estimation

Tufaner et al. [79] used ANN for simulating and optimizing operating conditions of Upflow Anaerobic Sludge Blanket (UASB) reactors for biogas generation. It was observed that ANN can efficiently predict the biogas yield from a laboratory-scale UASB reactors. Asadi et al. [80] used ANN and ANFIS with subtractive clustering, Fuzzy C-Means Clustering (FCMC), and grid partition for prediction of biogas production rate from an anaerobic digesters. Based on the results, the ANFIS-FCMC model outperformed the other sets of models. Akkaya et al. [81] used the multiple regression model in the production of biogas from landfill leachates. The proposed method demonstrated sufficient prediction accuracy.

Yield Estimation

Ghatak and Ghatak [82] used ANN to predict the yield of biogas from cattle dung, sugarcane bagasse, bamboo dust, and sawdust under mesophilic and thermophilic conditions. The capability of ANN modeling significantly reduced the processing time required for control of the process. Nair et al. [83] used ANN to evaluate biogas yield in an anaerobic bioreactor from the organic fraction of municipal solid consisting of vegetable waste, food waste, and yard trimming. It was inferred that an optimized CH₄ recovery can be reached at pH range between 6.6 and 7.1 with Total Volatile Solids (TVS) from 77 to 84%. Antwi et al. [84] used different training algorithms for ANN models along with multiple nonlinear regression (MnLR) to estimate biogas and methane yield from chemical, industrial sludges of paper, automobile, petrochemical, and food industries. It was concluded that conjugate gradient backpropagation and the Quasi-Newton method were the best among eleven training algorithms.

Ihunegbo et al. [85] used PLS to predict the yield of biogas from bioslurry. The results indicate that the acoustic chemometrics is a reliable Process Analytical Technologies (PAT) approach to monitor Total Solids (TS) in complex bioslurry and the same concept can be extended to other biomass processing industries as well. Qdais et al. [86] used ANN to predict biogas yield in the production of biogas from the waste digester. The ANN model was effective in capturing the important features of the variables involved in biogas digester operation for methane production.

Optimization of Quality and Yield

Qdais et al. [86] also integrated ANN model with the GA for optimizing operational parameters that resulted in 6.9% increase in yield. Dibaba et al. [87] determined that the best performance of Upflow Anaerobic Contactor (UAC) with 87% COD removal, and hydraulic retention time of 16.67 days where an increase of 7.4% in biogas production was realized. Barik and Murugan [88] used ANN and GA to estimate and optimize the

yield of biogas from cattle dung and seed cake of Karanja in co-digestion. The product quality using co-digestion of cake of Karanja and cattle dung mixture was higher than when using cattle dung samples for a mixing ratio of 1 cake of Karanja to 3 cattle dung. Oloko-Oba [89] used ANN integrated with GA to predict biogas production from cow dung, poultry droppings, and piggery waste. The optimal amount of poultry droppings, cow dung, plantain peel, and piggery waste were 0.7 kg, 0.0004 kg, 0.29 kg, and 0.61 kg, respectively. Zareei and Khodaei [90] used the ANFIS to estimate and optimize the production of anaerobic digestion-based biogas from cow manure and maize straw. The ANFIS model helped in optimizing the process conditions that resulted in 8% rise in production. Kana et al. [91] used ANN and GA to predict the optimum combination of rice bran, paper waste, banana stem, sawdust, and concentration of cow dung that enhanced the yield of biogas. Akbas et al. [92] used integrated ANN and Particle Swarm Optimization (PSO) model for robust control of production system of biogas. The integrated estimation and optimization framework increased the production and quality of biogas, and boosted the quantity of electricity production at the affiliated wastewater treatment facility.

3.3.3. Biohydrogen

Nasr et al. [93] used the ANN model for the prediction of hydrogen production from different substrates. The initial pH, temperature, initial substrate, biomass concentrations, and time were used as inputs of the model. The ANN exhibited high capability in capturing the correlation among parameters and the process output. Whiteman and Kana [94] predicted hydrogen yield using ANN and RSM. It was observed that ANN has greater accuracy than RSM. Ren et al. [95] used gray and ANN model for prediction of biohydrogen yield from feedstocks comprising of agricultural residues, paper wastes, and wood chips. The gray box model outperformed the ANN model in predicting the output in the context of uncertain data. Prakasham et al. [96] integrated ANN with GA for the prediction of biohydrogen yield from mixed anaerobic microbial consortia. The optimization strategy resulted in 16% increase in biohydrogen yield. Aghbashlo et al. [97] used a novel hybrid fuzzy clustering-ranking method with ANN to predict exergetic efficiencies in the production of hydrogen from photo-fermentation. Optimum values of flow rate of syngas and agitation speed were 13.68 mL/min and 348.62 rpm, respectively.

3.3.4. Miscellaneous (Bioethanol, Bisabolene)

Ezzatzadegan et al. [98] used Fuzzy Neural Network (FNN) and PSO to predict the yield of bioethanol from corn stover. The optimum fermentation time and required temperature were 69.39 h and 34.5 °C, respectively. Del Rio-Chanona et al. [99] used ANN-based multi-objective optimization with a hybrid stochastic search optimization in bisabolene production from microalgal biofuel.

3.4. Consumption, Engine Performance and Emissions

ML application in consumption, engine performance, and emissions are mostly performed simultaneously in the studies reported in the literature. Hence, all these aspects in this section were combined. Dominant ML methods in these studies were ANN, ANFIS, Extreme learning, SVM, and PLS. The studies are predominantly based on biodiesel as a fuel hence classification is carried based on the ML methods as shown in Table 5 and discussed in the Sections 3.4.1–3.4.4.

Table 5. Summary of ML applications in the consumption and emission phase.

Types of ML Methods	Input Variables	Error Range	References
ANN	Biofuel blend (22), engine speed (15), load (11), cetane number (4), output torque (3), density of fuels (3), compression ratios (3), intake air temperature (2), EPS content (2), lower heating value (2), CO ₂ , hydrogen flow rates, percent fuel for non-EGR engine, H ₂ , brake power, nano size, rpm, CNG flow rate, specific gravity, average molecular weight, net heat of combustion, Kinematic Viscosity (KV), time, fuel temperature, C-to-H ratio, engine crank angle, performance of a compression ignition engine, compression ratio, injection timing, CH ₄ ratio of the fuel, maximum cylinder pressure, pilot fuel and natural gas consumption, Air-to-Fuel Ratio (AFR) exhaust emissions, exhaust temperature values, smoke, HC, CO, fuel mass flow rate, injection pressure, and throttle position, biodiesel volume, and NO _x	$0.5420 \leq R^2 \leq 0.9990$	[100–126]
ANFIS	Double bonds (3), blend (3), load (3), injection timings, SV, IV, molar weight, speed, acceleration, rpm, VSP, passenger count, MAP, temperature, the average carbon numbers (2)	$0.9440 \leq R^2 \leq 0.9990$	[127–133]
Extreme learning method	Biodiesel ratio (5), engine speed (5), fuel consumption (5) engine torque (3), concentrations of the emissions (3), idle air valve normal position (2), fuel injection time (2), ignition advance, throttle position, carbon monoxide, nitrogen oxide, smoke opacity, brake thermal efficiency, combustion characteristics, exhaust emissions, performance, and the Air-to-Fuel Ratio (AFR)		[134–139]
Support vector machine (SVM) algorithm, PLS, and nonlinear regression	Composition, injection timing (I), power (P), and blend ratio (B)	$0.9500 \leq R^2 \leq 0.9970$	[140–142]

3.4.1. ANN

ANN application for estimation of compositions of the emissions and performance of biofuels-based engines are categorized based on the feedstock of the biofuels, such as vegetable oil, waste cooking oil, and non-vegetable oil.

Ismail et al. [100] used ANN to predict CO, CO₂, NO, unburned hydrocarbons, maximum heat release rate, the maximum pressure, location of maximum HRR, location of maximum pressure, and cumulative HRR (CuHRR) of an engine using the soybean and palm oil-based biodiesel. The ANN model demonstrated high prediction capability of engine combustion and emission behavior. Sharon et al. [101] used ANN to predict hydrocarbon, brake thermal efficiency, brake specific fuel consumption, NO_x, CO, and smoke density of biodiesel produced from vegetable fried oil and non-vegetable fried oil. The prediction accuracy for B15, B30, B60, and B90 was in an acceptable range. Javed et al. [102] used different training structures of ANN to predict BTE, BSFC, O₂, CO, NO_x, HC, CO₂, and EGT of biodiesel produced from jatropha methyl ester. Levenberge Marquardt as a training algorithm with 16 neurons gave the best prediction performance. Canakci et al. [103] used ANN to predict emissions, flow rates, engine load, maximum injection pressure, thermal efficiency, and maximum cylinder gas pressure. The ANN performed well in terms of prediction of the output except emissions, such as NO_x, CO, and UHC where mean error in prediction was higher. Oguz et al. [104] used ANN to estimate hourly fuel consumption, power, moment, and specific fuel consumption of biodiesel. The ANN model was found suitable for the prediction of engine performance. Barma et al. [105] used ANN to predict mechanical efficiency, mean effective pressure, Air-to-Fuel Ratio (AFR), fuel consumption,

and torque of engine consuming biodiesel engine. The BPANN gave adequate prediction accuracy for the different fuel blends. Celebi et al. [106] used ANN to estimate the noise and vibration level of biodiesel produced from a blend of sunflower, conventional diesel, and canola biodiesel. The ANN model outperformed the Linear Regression (LR) model.

Javed et al. [107] used ANN model for prediction of noise of engine operating on biodiesel with hydrogen dual-fueled zinc-oxide nanoparticle blend. The least noise was found for the H₂ flow rate of 1.5 L/min. Aydogan et al. [108] used ANN for estimation of engine performance, the engine torque, engine power, the Specific Fuel Consumption (SFC), and EGT of engine using cotton and rapeseed oils biodiesel. The ANN accurately predicted the engine performance, the engine torque, SFC, and EGT. Shojaeefard et al. [109] used ANN to estimate performance and exhaust emissions, BSFC, brake power, and exhaust emissions of DI engine working on biodiesel developed from castor oil. The ANN performance was compared with a group method of data handling. The ANN model was better in terms of prediction accuracy but the group method of data handling models was superior in terms of simplicity. Sharma et al. [110] used ANN for estimation of performance, BSFC, exhaust temperature, and exhaust composition of an engine using Polanga biodiesel. A very highly accurate prediction with a correlation coefficient closed to one was achieved. Omidvarborna et al. [111] used ANN to predict NO_x emissions and concentration of NO_x from EGR engines and non-EGR engines using soybean based biodiesel. The application of ANN was recommended for the estimation of NO_x emissions from both EGR engines and non-EGR engines. Karthickeyan et al. [112] used ANN model for estimation of performance and emissions characteristics from engine using orange oil-based biodiesel. Orange oil Methyl Ester (OME) with the Variable Compression Ratio (VCR) engine demonstrated higher efficiency and lesser fuel consumption. Menon and Krishnasamy [113] used ANN with GA to optimize emission characteristics and performance of a biodiesel engine. For realizing optimum biodiesel composition, the total saturated methyl ester contents were from 36 to 43 wt% and unsaturated contents were from 55 to 63 wt%, respectively.

Ghobadian et al. [114] used ANN to estimate fuel consumption, torque and emission of engine working on waste cooking oil-based biodiesel. The correlation coefficient and Mean Squared Error (MSE) for torque, SFC, CO, and HC were close to 1 and 0.0004, respectively. Pai et al. [115] used ANN to estimate emission characteristics and performance of a variable compression ratio CI-engine working on waste cooking oil based biodiesel. The mean error values of ANN were less than 8%, which is acceptable. Muralidharan and Vasudevan [116] used ANN to predict emission and performance of a four-stroke variable compression ratio engine and a single-cylinder using cooking oil-based biodiesel. A good agreement was found between predicted and experimental measurements. Najafi et al. [117] used ANN to predict energy and exergy efficiency, and exhaust temperature in the usage of waste cooking oil-based biodiesel. The ANN was more efficient compared to the first-principle models. Kannan et al. [118] used ANN for prediction of performance, torque, power, and specific fuel consumption of a biodiesel engine. The optimum values of injection timing and injection pressure were 25.5° Before Top Dead Center (BTDC) and 280 bar, respectively. Jaliliantabar et al. [119] used ANN for prediction of emissions, load, and speed of an engine working on blend of biodiesel fuel derived from waste cooking oil in diesel. An optimum operation was achieved with the reduction for CO, CO₂, HC, NO_x and smoke emissions approximately 47.25%, 48.23%, 52.7%, 94.55% and 44.29%, respectively. Kurtgoz et al. [120] used ANN to predict biogas engine performance, BSFC, thermal efficiency (TE), and volumetric efficiency (VE) of biogas produced from bovine manure. It was concluded that ANN can accurately estimate TE, BSFC, and VE values.

Aydogan [121] used ANN to predict NO_x, SFC, and maximum cylinder inner pressure caused by the usage of various blends of biodiesel, bioethanol, and diesel. ANN exhibited high prediction accuracy with a correlation coefficient of 0.98. Ilangkumaran et al. [122] used ANN for estimation of the engine performance, HC, CO, CO₂, NO_x, BTE, and smoke from engine working on biodiesel from fish oil. The ANN model exhibited high prediction accuracy. Tosun et al. [123] applied Linear Regression (LR) and ANN to predict torque and

exhaust emissions (CO, NO_x) of a naturally aspirated diesel engine running on biodiesel-alcohol mixtures. ANN had more accurate results than LR. Dharma et al. [124] used ANN to predict emission characteristics and performance of a single-cylinder DI-engine using mixed biodiesel-diesel fuel blends. The ANN model was able to accurately predict the outputs for different blends of the fuel. Najafi et al. [125] used ANN with GA to estimate exhaust emissions including NO_x, PM, CO, and UHC of biodiesel blend of glycerol triacetate. With the use of biodiesel and additive, a reduction of emission of NO_x and CO up to 63% and 42%, respectively, was realized. The PM was also substantially reduced by 27 times. Ozgur et al. [126] used ANN to predict CO, CO₂, NO_x and NO₂ emissions from engine using soybean oil-based biodiesel. Consequently, a close agreement was found between the predicted and experimental results.

3.4.2. Neuro Fuzzy Inference System

ZareNezhad and Aminian [127] used ANFIS to predict surface tension of biodiesel prepared from soybean, rapeseed, palm, and sunflower. The ANFIS-based framework outperformed the reported work and the surface tensions values for ten different biodiesels. A high correlation was found between the model estimated values and the experimental data. Gopalakrishnan et al. [128] used ANFIS and the Dynamic Evolving NFIS (DENFIS) to predict emission from transit bus using real-world data of NO_x, HC, CO, CO₂ and PM of biodiesel. The ANFIS outperformed the DENFIS in prediction of emissions. Mostafaei et al. [129] used ANFIS models to predict the cetane number of biodiesels from its FAME composition. The ANFIS models developed by Fuzzy C-Means (FCM) and grid partition FIS techniques have higher final desirability of 0.718, and 0.857, respectively. Sakthivel et al. [130] used Fuzzy logic and GA to predict emission, performance, and combustion parameters of CI-engine working on biodiesel from fish oil. For high engine performance and reduction in emissions, best blends were identified. The exact biodiesel proportion for no-load, 25, 50, 75, and 100% loads were found out using Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) as 17%, 17%, 18%, 17%, and 20%, respectively. Sakthivel [131] applied fuzzy logic to predict BTE, HC, EGT, NO_x, smoke, CO, CO₂, Combustion Delay (CD), Ignition Delay (ID), and Maximum Rate of Pressure Rise (MRPR) of biodiesel produced from fish oil. The fuzzy approach had an edge over theoretical and empirical methods in terms of prediction accuracy. Debnath et al. [132] used GA to predict BTE, CO, NO_x of engine butanol-based biodiesel. Less biodiesel and higher butanol percentage yields a good impact on performance and emission. Blend with 10% butanol, 10% biodiesel, and 80% diesel resulted in high heat release rate, cylinder pressure, BTE, and reduced NO_x. Ardabili et al. [133] used ANFIS model for the estimation of cetane number of biodiesel. It was also investigated that a rise in the carbon number of FAMEs increases the viscosity, cetane number, and HHV. However, a rise in the number of double bonds causes a decrease in viscosity, cetane number, and HHV.

3.4.3. Extreme Learning Method

Silitonga et al. [134] used ELM for prediction of BSFC and thermal efficiency of an engine running on biodiesel and bioethanol blends. The biodiesel-bioethanol-diesel blends had oxidation stability of more than 20 h that showed the potential of their commercialization. Silitonga et al. [135] used a Kernel-based ELM (K-ELM) to predict exhaust emissions, performance, and characteristics of combustion of biodiesel. K-ELM demonstrated the high potential of application in prediction and process optimization of biodiesel derived from a variety of feedstocks. Wong et al. [136] used K-ELM for prediction of the fuel consumption and emissions characteristic of engine working on biodiesel. With the K-ELM, Cuckoo Search (CS) was then used to determine the optimal biodiesel ratio. The CS optimization was compared with experimental results and PSO. In computational time, the CS and PSO were similar. However, in case of PSO, the time for tuning the parameters was more than CS because CS had a lower number of user-defined parameters than PSO. Wong and Wong [137] used Bayesian ELM (B-ELM) and metaheuristic optimization to predict fuel

consumption, the k-value, and emission characteristics from engine using gasoline and ethanol. The metaheuristic optimization methods was also applied for identifying optimal ECU setup.

Wong et al. [138] used ELM, LS-SVM, and RB-FNN to predict performance and the concentrations of NO_x, CO, HC, CO₂, and PM in emissions from engine working on biodiesel. The ELM performed better than LS-SVM and RB-FNN. Wong et al. [139] compared the prediction accuracy of SB-ELM with conventional ELM, B-ELM, and ANN to estimate performance parameters of engine, fuel consumption, and the concentrations of CO, HC, and CO₂ emissions. SB-ELM outperformed the other methods.

3.4.4. Support Vector Machine and Least Square Methods

Alves and Poppi [140] used an SVM and PLS to estimate the biodiesel content in fuel blend. A comparison of SVM and PLS models was made, where SVM outperformed the former in terms of accuracy in prediction. Maheshwari et al. [141] used nonlinear regression to estimate the performance and emission characteristics, smoke, BTE, HC, and NO_x emissions of biodiesel from Karanja. 3% biodiesel-diesel blend was found optimum for emissions and efficiency. Shamshirband et al. [142] used SVM Wavelet Transform (SVM-WT), SVM-RBF, SVM Firefly Algorithm (SVM-FFA), SVM based on quantum particle swarm optimization (SVM-QPSO) and ANN to predict exergetic parameter of a diesel engine and exhaust hot gas, exergy transfer rate to the cooling water, fuel exergy rate, and sustainability index of waste oil based biodiesel. The SVMWT approach was more efficient in prediction of exergetic efficiency and identification of best combustion properties, and fuel composition.

3.5. Application Summary

In the soil stage, ensemble and SVM were the most commonly used ML methods. The input variables for the ML application were soil characteristics, average precipitation, temperature, solar radiation, and wind speed. The output variables for the ML application were biomass yield and future life cycle environmental impacts. In the feedstock stage, ANN, statistical regression model, multiple linear regression, RF, SVR, and multiple nonlinear regression were used. The input variables for the ML application were blend composition, mixing speed, mixing time, heating values, temperature, size dimensions of microalgai, and characteristics of fluorescence signals. The dominant output variables were viscosity, density, flash point, higher heating values, oxidating stability, cetane number, yield and characteristics of biocrude and hydrochar, and fraction of methane.

The production phase was divided with regard to the type of biofuels produced, such as biodiesel, biogas, biohydrogen, and miscellaneous. The biodiesel category was further divided into four categories based on the nature of the application, such as (1) quality estimation, (2) yield estimation, (3) quality, and yield optimization, and (4) estimation and optimization of process conditions, and efficiency.

For the quality estimation, the dominant ML method was ANN followed by the regression model. The top five most commonly used input variables were reaction time, reaction temperature, calcination temperature, flow rate, and pressure. In the yield estimation, the dominant ML method was ANN followed by ANFIS. The top five most commonly used input variables were catalyst concentration, reaction time, temperature, methanol-to-oil molar ratio, and total volatile fatty acid (VFA) of the effluent.

In the quality and yield optimization section, GA-based ANN was the dominant ML method followed by GA-based SVM and ANFIS. The top five most commonly used input variables were methanol-to-oil molar ratio, reaction temperature, stirring speed, reaction time, and catalyst concentration. In the process efficiency estimation and optimization section, ANN was the dominant ML method followed by ANFIS in combination with various optimization methods. The top five most commonly used input variables were concentration, water content, reaction time, temperature, and methanol-to-oil molar ratio.

The biogas category was categorized based on the nature of applications, such as (1) quality estimation, (2) yield estimation, and (3) optimization of quality and yield. In the

quality estimation, ANN was the dominant ML method followed by ANFIS and multiple regression models. The topmost commonly used input variables were volatile fatty acids (VFAs), total solids, fixed solids, volatile solids, and pH. In the yield estimation, ANN was the dominant ML method followed by MNL models and PLS. The top five most commonly used input variables were temperature, pH, TVS, VFAs, and composition. In the quality and yield optimization section, ANN-GA was the dominant ML method followed by ANFIS. The top five most commonly used input variables were TS, TVS, pH, temperature, and carbon-to-nitrogen ratio.

For the biohydrogen, ANN and its integration with GA were used for yield prediction and optimization. The input variables were pH, substrate, biomass concentrations, temperature, and time. The output variables were biohydrogen yield, exergetic outputs, and COD removal. In the miscellaneous category, comprised of bisabolene and bioethanol, ANN, FNN and their integration with optimization techniques such as PSO were used. The input variables were incident light intensity, recycling gas flow rate, cardinal coordinates of sample, temperature, glucose content, and fermentation time.

The consumption, engine performance, and emission cases were studied simultaneously in most of the related papers, hence they were reviewed in a single subsection. Biodiesel was the dominant type of biofuels in this stage so the literature was rather classified based on the type of ML methods, such as ANN, ANFIS, ELM, and SVM.

In the ANN application, the top five input variables were biofuel blend, engine speed, load, cetane number, and output torque. The top five output variables were emission characteristics NO_x , CO, CO_2 , BSFC and temperature. In the ANFIS application, the top five input variables were double bonds, blend, load, average carbon numbers, and temperature. The top five output variables were BTE, NO_x , CO, smoke, and CO_2 , respectively. In the ELM application, the top five input variables were biodiesel ratio, engine speed, engine torque, fuel injection time, and idle air valve normal position. The top five output variables were fuel consumption, brake thermal efficiency, performance, exhaust emissions, and fuel concentrations. In the SVM and LS methods, the input variable was composition. Meanwhile, the output variable was the yield.

The overall trends observed in the ML application in the biofuels' life cycle are shown in Figure 2. The phase-wise application of ML is shown in Figure 2a. Meanwhile, the number of publications in the subject area has been consistently increasing (except for the years 2014 and 2018) as shown in Figure 2b. The number of publications related to the ML methods is shown in Figure 2c where the use of ANN was reported to be 72, followed by GA at 20, SVM at 15, and ELM at 14. The emergence of the ELM that belongs to the second generation of ML is shown in Figure 2d. Contribution in terms of studies conducted in the subject area is reported from across the globe as shown in Figure 3. The leading country was India with 25 publications followed by Iran, Turkey, Malaysia, and the USA with 21, 14, 13, and 12 publications, respectively.

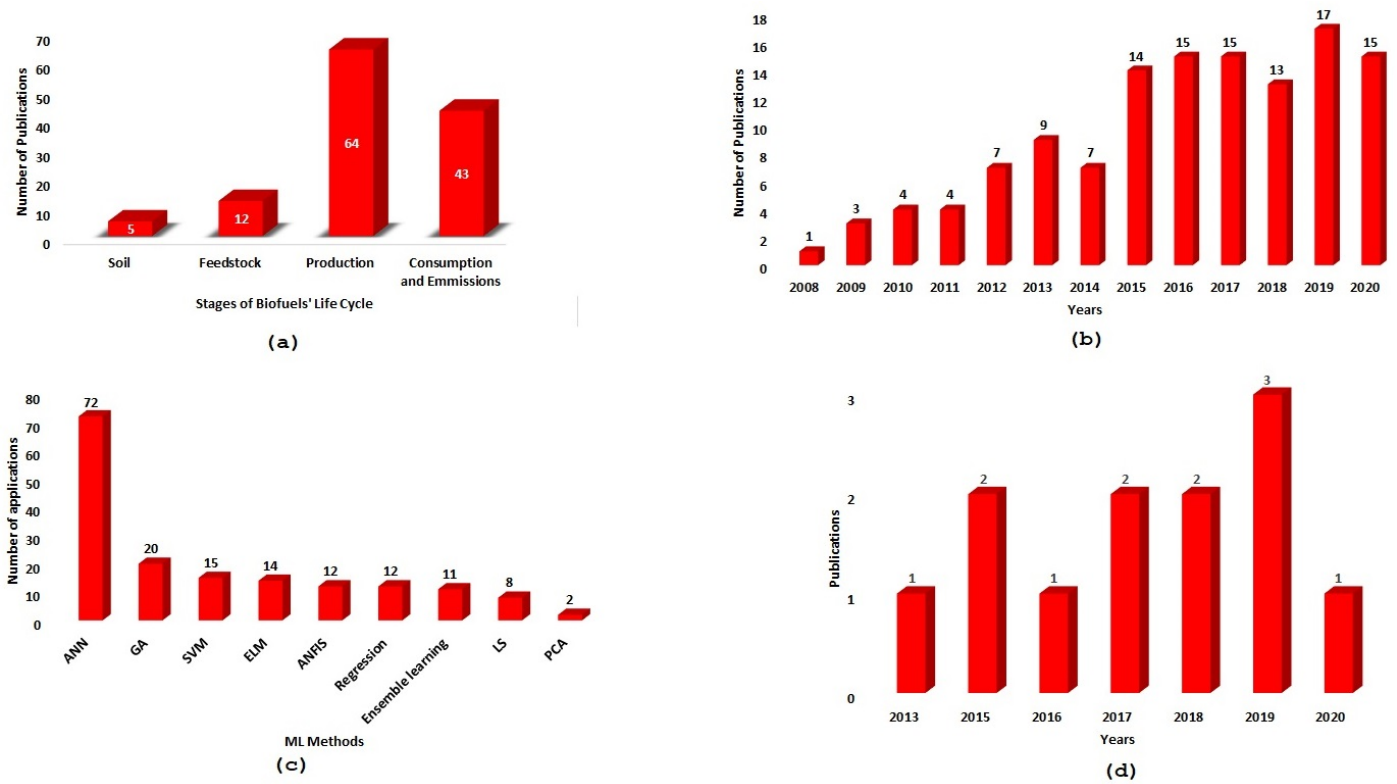


Figure 2. (a) Stage-wise share of publications, (b) Year-wise publications in the subject area, (c) ML method-wise publications, and (d) Year-wise publication from the second generation ML methods.

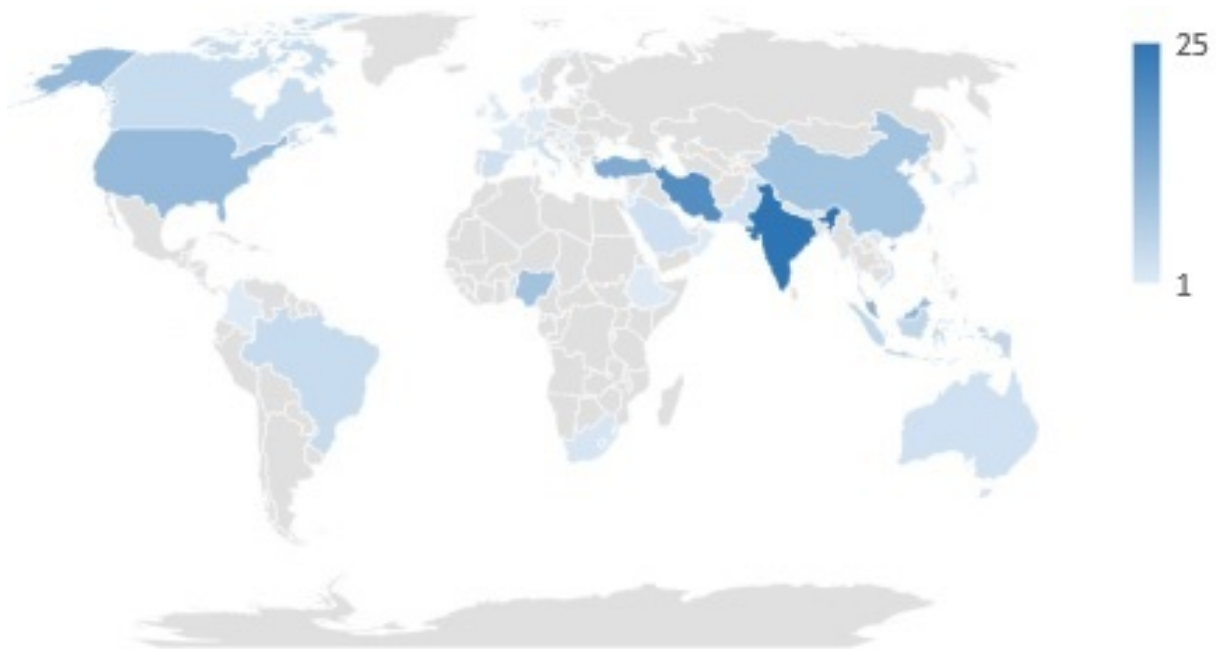


Figure 3. Region-wise publications on ML applications in biofuels' life cycle.

4. Conclusions and Future Work

In the life cycle of biofuels, out of the four stages, the production stage holds a 52% share of the reported ML applications followed by consumption and emission, soil, and feedstocks stages with 35%, 9%, and 4%, shares, respectively. ANN has been consistently dominant from the first generation of ML methods. Interestingly, GA based optimization was the second-highly reported work after ANN. GA outperformed the conventionally used RSM approach when compared in several studies in realizing optimum process conditions. The ML methods in descending order in terms of application are ANN, GA, SVM, ELM, ANFIS, Regression (linear/non-linear), ensemble learning, LS, and PCA. The application of the second generation was reported for the first time in 2013. ELM was the dominant out of the second generation ML methods with several variations reported every year after 2013. The contribution in terms of studies conducted in the subject area was reported from across the globe, however, India, Iran, Turkey, Malaysia, and the USA collectively form 54% of authors' affiliation from the reported articles.

Although applications of ML in biofuels are found in the whole life cycle, there is no research applying ML to cover integrated supply chains in biofuels involving agriculture production, feedstock management, quality control, bioprocessing to transform biomass into biofuels and the consumption and related emissions of the biofuels altogether. The challenge is to better determine the interplay over the decision-making among multiple players of multi-commodities of the value chain of biofuels.

Making decisions on complex design, operation, and control of today's industry may count on the novel capabilities of advanced analytics in engineering. A potential future application of advanced analytics is to model Supply Chain Resilience (SCR) of transactions, logistics, operations, etc., of such complex representation of supply chain elements in the industry. Subsequently, ML approaches can be used to determine optimizable surrogate models to correlate independent variables (e.g., resistance and recovery of the SCR) to the dependent variable SCR.

A ML methodology to quantify SCR based on continuous x and binary y variables of resistance (avoidance and containment) and recovery (stabilization and return) can consider ad hoc relationships of dependent and independent variables to be part of the SCR predictions in the ML method. Such SCR algebraic or analytical formulas obtained in constrained decision regression approach can be used in optimization and control problems to move from traditional independent networks in order to create more flexibility in fulfilling demand through the complementary behavior of heterogeneous resources. Such coupling of multi-layered networks paves the way for optimal resource exchange, efficient decision making, and knowledge discovery through developing ML, control, and optimization techniques for large-scale interdependent decision making.

Author Contributions: Conceptualization, I.A.; formal analysis, I.A., A.S., I.I.C., B.C.M., J.S., Z.U., M.K. (Muzammil Khan); investigation, I.A., A.S., I.I.C., B.C.M., J.S., Z.U., M.K. (Muzammil Khan); resources, I.A. and M.K.; writing—original draft preparation, I.A., A.S., I.I.C., B.C.M., A.H.; writing—review and editing, I.A., A.S., I.I.C., B.C.M., and M.K. (Manabu Kano); visualization, I.A. and M.K. (Manabu Kano); supervision, I.A. and M.K. (Manabu Kano). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript:

ALFIMO	Artificial Linear Interdependent Fuzzy Multi-Objective Optimization
AFR	Air-to-Fuel Ratio
AI	Artificial Intelligence
ANFIS	Adaptive Neuro-Fuzzy Interference System
ANN	Artificial Neural Networks
ALS	Alternative Least Square
ASTM	American Society for Testing and Materials
BRT	Boosted Regression Tree
CE	Conversion Efficiency
CN	Cetane Number
COD	Chemical Oxygen Demand
CS	Cuckoo Search
DA	Discriminant Analysis
ELM	Extreme Learning Machine
ERT	Extremely Randomized Trees
FAME	Fatty Acid Methyl Ester
FCM	Fuzzy C-Means
FCMC	Fuzzy C-Means Clustering
FEE	Functional Exergy Efficiency
FP	Flash Point
GA	Genetic Algorithm
GBD	eXtreme Gradient Boosting-xgbDART
GBL	eXtreme Gradient Boosting-xgbLinear
GBT	eXtreme Gradient Boosting-xgbtree
GIP	Group Interaction Parameters
GPM	Gaussian Process Model
K-ELM	Kernel-based Extreme Learning Machine
KV	Kinematic Viscosity
LLE	Liquid-Liquid Equilibrium
LME	Linear Mixed-Effects
LR	Linear Regression
LS	Least Square
MCR	Multivariate Curve Resolution
ML	Machine Learning
MNLR	Multiple Non-Linear Regression
MSE	Mean Squared Error
NED	Normalized Exergy Destruction
NN	Neural Network
OME	Orange oil Methyl Ester
PAT	Process Analytical Technologies
PCA	Principal Component Analysis
PLS	Partial least square
PU/MU	Mono and poly-unsaturated fatty acids balance
R ²	Correlation coefficient/Coefficient of determination
RB-FNN	Radial Basis Function Neural Network
RF	Random Forest
RFM	Random Forest Model
RLS	Recursive Least Squares
RSM	Response Surface Methodology
SFC	Specific Fuel Consumption
SVM	Support Vector Machines

SVM-FFA	Support Vector Machine Firefly Algorithm
SVM-QPSO	Support vector machine based on quantum particle swarm optimization
SVM-RBF	Support Vector Machine with the Radial Basis Function
SVM-WT	Support Vector Machine Wavelet Transform
SVR	Support Vector Regression
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
TS	Total Solids
TVS	Total Volatile Solids
UAC	Upflow Anaerobic Contactor
UASB	Upflow Anaerobic Sludge Blanket
UEE	Universal Exergy Efficiency
VCR	Variable Compression Ratio
VFA	Total volatile fatty acid

References

- Samuel, A.L. Some Studies in Machine Learning Using the Game of Checkers. *IBM J. Res. Dev.* **1959**, *3*, 210–229. [[CrossRef](#)]
- Shobha, G.; Rangaswamy, S. Computational analysis and understand of natural languages: Principles, methods and applications. In *Handbook of Statistics*; North-Holland: Amsterdam, The Netherlands, 2018.
- El Boucheffry, K.; de Souza, R.S. Learning in big data: Introduction to machine learning. In *Knowledge Discovery in Big Data from Astronomy and Earth Observation*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 225–249.
- Popovic, D. Intelligent Control with Neural Networks. In *Soft Computing and Intelligent Systems*; Elsevier: Amsterdam, The Netherlands, 2000; pp. 419–467.
- Ge, Z.; Song, Z.; Ding, S.X.; Huang, B. Data mining and analytics in the process industry: The role of machine learning. *IEEE Access* **2017**, *5*, 20590–20616. [[CrossRef](#)]
- De, R.; Rajan, A.; Govindaraj, K.; Kinage, A.; Ramamurthy, R.K.; Schreder, J.; Peters, C. System and Method for Industrial Process Automation Controller Farm with Flexible Redundancy Schema and Dynamic Resource Management through Machine Learning. U.S. Patent 10,416,630, 17 September 2019.
- Keliris, A.; Salehghaffari, H.; Cairl, B.; Krishnamurthy, P.; Maniatakos, M.; Khorrami, F. Machine learning-based defense against process-aware attacks on industrial control systems. In Proceedings of the 2016 IEEE International Test Conference (ITC), Fort Worth, TX, USA, 15–17 November 2016; pp. 1–10.
- Kanawaday, A.; Sane, A. Machine learning for predictive maintenance of industrial machines using IoT sensor data. In Proceedings of the 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 24–26 November 2017; pp. 87–90.
- Weichert, D.; Link, P.; Stoll, A.; Rüping, S.; Ihlenfeldt, S.; Wrobel, S. A review of machine learning for the optimization of production processes. *Int. J. Adv. Manuf. Technol.* **2019**, *104*, 1889–1902. [[CrossRef](#)]
- Lahdhiri, H.; Said, M.; Abdellafou, K.B.; Taouali, O.; Harkat, M.F. Supervised process monitoring and fault diagnosis based on machine learning methods. *Int. J. Adv. Manuf. Technol.* **2019**, *102*, 2321–2337. [[CrossRef](#)]
- Rymarczyk, T.; Kozłowski, E.; Kłosowski, G.; Niderla, K. Logistic Regression for Machine Learning in Process Tomography. *Sensors* **2019**, *19*, 3400. [[CrossRef](#)]
- Min, Q.; Lu, Y.; Liu, Z.; Su, C.; Wang, B. Machine learning based digital twin framework for production optimization in petrochemical industry. *Int. J. Inf. Manag.* **2019**, *49*, 502–519. [[CrossRef](#)]
- Anifowose, F.A.; Labadin, J.; Abdulraheem, A. Ensemble machine learning: An untapped modeling paradigm for petroleum reservoir characterization. *J. Pet. Sci. Eng.* **2017**, *151*, 480–487. [[CrossRef](#)]
- Oey, T.; Jones, S.; Bullard, J.W.; Sant, G. Machine learning can predict setting behavior and strength evolution of hydrating cement systems. *J. Am. Ceram. Soc.* **2020**, *103*, 480–490. [[CrossRef](#)]
- Unnikrishnan, S.; Donovan, J.; Macpherson, R.; Tormey, D. Machine Learning for Automated Quality Evaluation in Pharmaceutical Manufacturing of Emulsions. *J. Pharm. Innov.* **2020**, *15*, 392–403. [[CrossRef](#)]
- Guo, S.; Yu, J.; Liu, X.; Wang, C.; Jiang, Q. A predicting model for properties of steel using the industrial big data based on machine learning. *Comput. Mater. Sci.* **2019**, *160*, 95–104. [[CrossRef](#)]
- Sewsynker-Sukai, Y.; Faloye, F.; Kana, E.B.G. Artificial neural networks: An efficient tool for modelling and optimization of biofuel production (a mini review). *Biotechnol. Biotechnol. Equip.* **2017**, *31*, 221–235. [[CrossRef](#)]
- Ardabili, S.; Mosavi, A.; Várkonyi-Kóczy, A.R. Systematic review of deep learning and machine learning models in biofuels research. In *International Conference on Global Research and Education*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 19–32.
- Gleason, C.J.; Im, J. Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sens. Environ.* **2012**, *125*, 80–91. [[CrossRef](#)]
- Huntington, T.; Cui, X.; Mishra, U.; Scown, C.D. Machine learning to predict biomass sorghum yields under future climate scenarios. *Biofuels Bioprod. Biorefin.* **2020**, *14*, 566–577. [[CrossRef](#)]

21. Habyarimana, E.; Piccard, I.; Catellani, M.; De Franceschi, P.; Dall'Agata, M. Towards predictive modeling of sorghum biomass yields using fraction of absorbed photosynthetically active radiation derived from sentinel-2 satellite imagery and supervised machine learning techniques. *Agronomy* **2019**, *9*, 203. [[CrossRef](#)]
22. Lee, E.K.; Zhang, W.J.; Zhang, X.; Adler, P.R.; Lin, S.; Feingold, B.J.; Khwaja, H.A.; Romeiko, X.X. Projecting life-cycle environmental impacts of corn production in the US Midwest under future climate scenarios using a machine learning approach. *Sci. Total Environ.* **2020**, *714*, 136697. [[CrossRef](#)] [[PubMed](#)]
23. Yang, P.; Zhao, Q.; Cai, X. Machine learning based estimation of land productivity in the Contiguous US using biophysical predictors. *Environ. Res. Lett.* **2020**, *15*, 074013. [[CrossRef](#)]
24. Mahanty, B.; Zafar, M.; Park, H.S. Characterization of co-digestion of industrial sludges for biogas production by artificial neural network and statistical regression models. *Environ. Technol.* **2013**, *34*, 2145–2153. [[CrossRef](#)] [[PubMed](#)]
25. Mairizal, A.Q.; Awad, S.; Priadi, C.R.; Hartono, D.M.; Moersidik, S.S.; Tazerout, M.; Andres, Y. Experimental study on the effects of feedstock on the properties of biodiesel using multiple linear regressions. *Renew. Energy* **2020**, *145*, 375–381. [[CrossRef](#)]
26. Giwa, S.O.; Adekomaya, S.O.; Adama, K.O.; Mukaila, M.O. Prediction of selected biodiesel fuel properties using artificial neural network. *Front. Energy* **2015**, *9*, 433–445. [[CrossRef](#)]
27. Tchameni, A.P.; Zhao, L.; Ribeiro, J.X.; Li, T. Predicting the rheological properties of waste vegetable oil biodiesel-modified water-based mud using artificial neural network. *Geosyst. Eng.* **2019**, *22*, 101–111. [[CrossRef](#)]
28. Dahunsi, S. Mechanical pretreatment of lignocelluloses for enhanced biogas production: Methane yield prediction from biomass structural components. *Bioresour. Technol.* **2019**, *280*, 18–26. [[CrossRef](#)]
29. Reimann, R.; Zeng, B.; Jakopec, M.; Burdukiewicz, M.; Petrick, I.; Schierack, P.; Rödiger, S. Classification of dead and living microalgae *Chlorella vulgaris* by bioimage informatics and machine learning. *Algal Res.* **2020**, *48*, 101908. [[CrossRef](#)]
30. Tang, Q.; Chen, Y.; Yang, H.; Liu, M.; Xiao, H.; Wu, Z.; Chen, H.; Naqvi, S.R. Prediction of bio-oil yield and hydrogen contents based on machine learning method: Effect of biomass compositions and pyrolysis conditions. *Energy Fuels* **2020**, *34*, 11050–11060. [[CrossRef](#)]
31. Shahbeig, H.; Nosrati, M. Pyrolysis of biological wastes for bioenergy production: Thermo-kinetic studies with machine-learning method and Py-GC/MS analysis. *Fuel* **2020**, *269*, 117238. [[CrossRef](#)]
32. Ighalo, J.O.; Adeniyi, A.G.; Marques, G. Application of linear regression algorithm and stochastic gradient descent in a machine-learning environment for predicting biomass higher heating value. *Biofuels Bioprod. Biorefin.* **2020**, *14*, 1286–1295. [[CrossRef](#)]
33. Kumar, R.; Dhanarajan, G.; Sarkar, D.; Sen, R. Multi-fold enhancement in sustainable production of biomass, lipids and biodiesel from oleaginous yeast: An artificial neural network-genetic algorithm approach. *Sustain. Energy Fuels* **2020**, *4*, 6075–6084. [[CrossRef](#)]
34. Cheng, F.; Luo, H.; Colosi, L.M. Slow pyrolysis as a platform for negative emissions technology: An integration of machine learning models, life cycle assessment, and economic analysis. *Energy Convers. Manag.* **2020**, *223*, 113258. [[CrossRef](#)]
35. Cheng, F.; Porter, M.D.; Colosi, L.M. Is hydrothermal treatment coupled with carbon capture and storage an energy-producing negative emissions technology? *Energy Convers. Manag.* **2020**, *203*, 112252. [[CrossRef](#)]
36. Soltani, S.; Rashid, U.; Roodbar Shojaei, T.; Nehdi, I.A.; Ibrahim, M. Modeling of the nanocrystalline-sized mesoporous zinc oxide catalyst using an artificial neural network for efficient biodiesel production. *Chem. Eng. Commun.* **2019**, *206*, 33–47. [[CrossRef](#)]
37. Ahmad, I.; Ayub, A.; Ibrahim, U.; Khattak, M.K.; Kano, M. Data-Based Sensing and Stochastic Analysis of Biodiesel Production Process. *Energies* **2019**, *12*, 63. [[CrossRef](#)]
38. Gulum, M.; Onay, F.K.; Bilgin, A. Evaluation of predictive capabilities of regression models and artificial neural networks for density and viscosity measurements of different biodiesel-diesel-vegetable oil ternary blends. *Environ. Clim. Technol.* **2018**, *22*, 179–205. [[CrossRef](#)]
39. Tomazzoni, G.; Meira, M.; Quintella, C.M.; Zagonel, G.F.; Costa, B.J.; de Oliveira, P.R.; Pepe, I.M.; da Costa Neto, P.R. Identification of vegetable oil or biodiesel added to diesel using fluorescence spectroscopy and principal component analysis. *J. Am. Oil Chem. Soc.* **2014**, *91*, 215–227. [[CrossRef](#)]
40. Sarve, A.; Sonawane, S.S.; Varma, M.N. Ultrasound assisted biodiesel production from sesame (*Sesamum indicum* L.) oil using barium hydroxide as a heterogeneous catalyst: Comparative assessment of prediction abilities between response surface methodology (RSM) and artificial neural network (ANN). *Ultrason. Sonochem.* **2015**, *26*, 218–228. [[CrossRef](#)] [[PubMed](#)]
41. Kumar, S.; Jain, S.; Kumar, H. Prediction of jatropha-algae biodiesel blend oil yield with the application of artificial neural networks technique. *Energy Sources Part A Recover. Util. Environ. Eff.* **2019**, *41*, 1285–1295. [[CrossRef](#)]
42. Banerjee, A.; Varshney, D.; Kumar, S.; Chaudhary, P.; Gupta, V. Biodiesel production from castor oil: ANN modeling and kinetic parameter estimation. *Int. J. Ind. Chem.* **2017**, *8*, 253–262. [[CrossRef](#)]
43. Kanat, G.; Saral, A. Estimation of Biogas Production Rate in a Thermophilic UASB Reactor Using Artificial Neural Networks. *Environ. Model. Assess.* **2008**, *14*, 607–614. [[CrossRef](#)]
44. Kumar, S.; Jain, S.; Kumar, H. Process parameter assessment of biodiesel production from a Jatropha-algae oil blend by response surface methodology and artificial neural network. *Energy Sources Part A Recover. Util. Environ. Eff.* **2017**, *39*, 2119–2125. [[CrossRef](#)]
45. Chakraborty, R.; Sahu, H. Intensification of biodiesel production from waste goat tallow using infrared radiation: Process evaluation through response surface methodology and artificial neural network. *Appl. Energy* **2014**, *114*, 827–836. [[CrossRef](#)]

46. Pandu, K.; Joseph, S.; Arun, N.; Sundaramoorthy, K. Optimization of biohydrogen production by *Enterobacter* species using artificial neural network and response surface methodology. *J. Renew. Sustain. Energy* **2013**, *5*, 033104.
47. Kumar, S. Comparison of linear regression and artificial neural network technique for prediction of a soybean biodiesel yield. *Energy Sources Part A Recover. Util. Environ. Eff.* **2020**, *42*, 1425–1435. [[CrossRef](#)]
48. Moradi, G.; Dehghani, S.; Khosravian, F.; Arjmandzadeh, A. The optimized operational conditions for biodiesel production from soybean oil and application of artificial neural networks for estimation of the biodiesel yield. *Renew. Energy* **2013**, *50*, 915–920. [[CrossRef](#)]
49. Guo, J.; Baghban, A. Application of ANFIS strategy for prediction of biodiesel production using supercritical methanol. *Energy Sources Part A Recover. Util. Environ. Eff.* **2017**, *39*, 1862–1868. [[CrossRef](#)]
50. Mostafaei, M.; Javadikia, H.; Naderloo, L. Modeling the effects of ultrasound power and reactor dimension on the biodiesel production yield: Comparison of prediction abilities between response surface methodology (RSM) and adaptive neuro-fuzzy inference system (ANFIS). *Energy* **2016**, *115*, 626–636. [[CrossRef](#)]
51. Maran, J.P.; Priya, B. Comparison of response surface methodology and artificial neural network approach towards efficient ultrasound-assisted biodiesel production from muskmelon oil. *Ultrason. Sonochem.* **2015**, *23*, 192–200. [[CrossRef](#)] [[PubMed](#)]
52. Bobadilla, M.; Fernandez, R.; Lostado-Lorza, R.; Somovilla Gómez, F.; Vergara, E. Optimizing Biodiesel Production from Waste Cooking Oil Using Genetic Algorithm-Based Support Vector Machines. *Energies* **2018**, *11*, 2995. [[CrossRef](#)]
53. Cheng, M.Y.; Prayogo, D.; Ju, Y.H.; Wu, Y.W.; Sutanto, S. Optimizing mixture properties of biodiesel production using genetic algorithm-based evolutionary support vector machine. *Int. J. Green Energy* **2016**, *13*, 1599–1607. [[CrossRef](#)]
54. Sivamani, S.; Selvakumar, S.; Rajendran, K.; Muthusamy, S. Artificial neural network–genetic algorithm-based optimization of biodiesel production from *Simarouba glauca*. *Biofuels* **2019**, *10*, 393–401. [[CrossRef](#)]
55. Ighose, B.O.; Adeleke, I.A.; Damos, M.; Junaid, H.A.; Okpalaeke, K.E.; 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.* **2017**, *132*, 231–240. [[CrossRef](#)]
56. Dhingra, S.; Dubey, K.K.; Bhushan, G. A polymath approach for the prediction of optimized transesterification process variables of polanga biodiesel. *J. Am. Oil Chem. Soc.* **2014**, *91*, 641–653. [[CrossRef](#)]
57. Ishola, N.B.; Okeleye, A.A.; Osunleke, A.S.; Betiku, E. Process modeling and optimization of sorrel biodiesel synthesis using barium hydroxide as a base heterogeneous catalyst: Appraisal of response surface methodology, neural network and neuro-fuzzy system. *Neural Comput. Appl.* **2019**, *31*, 4929–4943. [[CrossRef](#)]
58. Silitonga, A.S.; Mahlia, T.M.I.; Shamsuddin, A.H.; Ong, H.C.; Milano, J.; Kusumo, F.; Sebayang, A.H.; Dharma, S.; Ibrahim, H.; Husin, H.; et al. Optimization of *Cerbera manghas* Biodiesel Production Using Artificial Neural Networks Integrated with Ant Colony Optimization. *Energies* **2019**, *12*, 3811. [[CrossRef](#)]
59. Chakraborty, R.; Das, S.; Pradhan, P.; Mukhopadhyay, P. Prediction of optimal conditions in the methanolysis of mustard oil for biodiesel production using cost-effective mg-solid catalysts. *Ind. Eng. Chem. Res.* **2014**, *53*, 19681–19689. [[CrossRef](#)]
60. Goharimanesh, M.; Lashkaripour, A.; Akbari, A.A. Optimization of biodiesel production using multi-objective genetic algorithm. *J. Appl. Sci. Eng.* **2016**, *19*, 117–124. [[CrossRef](#)]
61. Oladipo, A.S.; Ajayi, O.A.; Oladipo, A.A.; Azarmi, S.L.; Nurudeen, Y.; Atta, A.Y.; Ogunyemi, S.S. Magnetic recyclable eggshell-based mesoporous catalyst for biodiesel production from crude neem oil: Process optimization by central composite design and artificial neural network. *C. R. Chim.* **2018**, *21*, 684–695. [[CrossRef](#)]
62. Rajendra, M.; Jena, P.C.; Raheman, H. Prediction of optimized pretreatment process parameters for biodiesel production using ANN and GA. *Fuel* **2009**, *88*, 868–875. [[CrossRef](#)]
63. Fahmi, I.; Cremaschi, S. Process synthesis of biodiesel production plant using artificial neural networks as the surrogate models. *Comput. Chem. Eng.* **2012**, *46*, 105–123. [[CrossRef](#)]
64. Betiku, E.; Omilakin, O.R.; Ajala, S.O.; Okeleye, A.A.; Taiwo, A.E.; Solomon, B.O. Mathematical modeling and process parameters optimization studies by artificial neural network and response surface methodology: A case of non-edible neem (*Azadirachta indica*) seed oil biodiesel synthesis. *Energy* **2014**, *72*, 266–273. [[CrossRef](#)]
65. Zhang, Y.; Niu, C. Toward estimation of biodiesel production from castor oil using ANN. *Energy Sources Part A Recover. Util. Environ. Eff.* **2018**, *40*, 1469–1476. [[CrossRef](#)]
66. Mujtaba, M.; Masjuki, H.; Kalam, M.; Ong, H.C.; Gul, M.; Farooq, M.; Soudagar, M.E.M.; Ahmed, W.; Harith, M.; Yusoff, M. Ultrasound-assisted process optimization and tribological characteristics of biodiesel from palm-sesame oil via response surface methodology and extreme learning machine-Cuckoo search. *Renew. Energy* **2020**, *158*, 202–214. [[CrossRef](#)]
67. Bemani, A.; Xiong, Q.; Baghban, A.; Habibzadeh, S.; Mohammadi, A.H.; Doranehgard, M.H. Modeling of cetane number of biodiesel from fatty acid methyl ester (FAME) information using GA-, PSO-, and HGAPSO-LSSVM models. *Renew. Energy* **2020**, *150*, 924–934. [[CrossRef](#)]
68. Karimi, M.; Jenkins, B.; Stroeve, P. Multi-objective optimization of transesterification in biodiesel production catalyzed by immobilized lipase: Multi-objective optimization of biodiesel production. *Biofuels Bioprod. Biorefin.* **2016**, *10*. [[CrossRef](#)]
69. Aghbashlo, M.; Hosseinpour, S.; Tabatabaei, M.; Soufiyan, M.M. Multi-objective exergetic and technical optimization of a piezoelectric ultrasonic reactor applied to synthesize biodiesel from waste cooking oil (WCO) using soft computing techniques. *Fuel* **2019**, *235*, 100–112. [[CrossRef](#)]

70. Patle, D.S.; Sharma, S.; Ahmad, Z.; Rangaiah, G. Multi-objective optimization of two alkali catalyzed processes for biodiesel from waste cooking oil. *Energy Convers. Manag.* **2014**, *85*, 361–372. [[CrossRef](#)]
71. Shukri, M.R.; Rahman, M.; Ramasamy, D.; Kadirgama, K. Artificial neural network optimization modeling on engine performance of diesel engine using biodiesel fuel. *Int. J. Automot. Mech. Eng.* **2015**, *11*, 2332–2347. [[CrossRef](#)]
72. Sarve, A.N.; Varma, M.N.; Sonawane, S.S. Response surface optimization and artificial neural network modeling of biodiesel production from crude mahua (*Madhuca indica*) oil under supercritical ethanol conditions using CO₂ as co-solvent. *RSC Adv.* **2015**, *5*, 69702–69713. [[CrossRef](#)]
73. Kuen, H.Y.; Mjalli, F.S.; Koon, Y.H. Recursive Least Squares-Based Adaptive Control of a Biodiesel Transesterification Reactor. *Ind. Eng. Chem. Res.* **2010**, *49*, 11434–11442. [[CrossRef](#)]
74. Rouchi, M.B.; Khorrami, M.K.; Garmarudi, A.B.; de la Guardia, M. Application of infrared spectroscopy as Process Analytics Technology (PAT) approach in biodiesel production process utilizing Multivariate Curve Resolution Alternative Least Square (MCR-ALS). *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2019**, *213*, 347–353. [[CrossRef](#)]
75. López-Zapata, B.; Adam Medina, M.; Alvarez Gutierrez, P.; Castillo González, J.; Hernandez-De-Leon, H.; Valdés, L. Virtual Sensors for Biodiesel Production in a Batch Reactor. *Sustainability* **2017**, *9*, 455. [[CrossRef](#)]
76. Nicola, G.; Moglie, M.; Pacetti, M.; Santori, G. Bioenergy II: Modeling and Multi-Objective Optimization of Different Biodiesel Production Processes. *Int. J. Chem. React. Eng.* **2010**, *8*. [[CrossRef](#)]
77. Noriega, M.A.; Narváez, P.C. UNIFAC correlated parameters for liquid-liquid equilibrium prediction of ternary systems related to biodiesel production process. *Fuel* **2019**, *249*, 365–378. [[CrossRef](#)]
78. Wong, K.I.; Wong, P.K. Adaptive air-fuel ratio control of dual-injection engines under biofuel blends using extreme learning machine. *Energy Convers. Manag.* **2018**, *165*, 66–75. [[CrossRef](#)]
79. Tufaner, F.; Avşar, Y.; Gönüllü, M.T. Modeling of biogas production from cattle manure with co-digestion of different organic wastes using an artificial neural network. *Clean Technol. Environ. Policy* **2017**, *19*, 2255–2264. [[CrossRef](#)]
80. Asadi, M.; Guo, H.; McPhedran, K. Biogas production estimation using data-driven approaches for cold region municipal wastewater anaerobic digestion. *J. Environ. Manag.* **2020**, *253*, 109708. [[CrossRef](#)] [[PubMed](#)]
81. Akkaya, E.; Demir, A.; Varank, G. Estimation of biogas generation from a uasb reactor via multiple regression model. *Int. J. Green Energy* **2015**, *12*, 185–189. [[CrossRef](#)]
82. Ghatak, M.D.; Ghatak, A. Artificial neural network model to predict behavior of biogas production curve from mixed lignocellulosic co-substrates. *Fuel* **2018**, *232*, 178–189. [[CrossRef](#)]
83. Nair, V.V.; Dhar, H.; Kumar, S.; Thalla, A.K.; Mukherjee, S.; Wong, J.W. Artificial neural network based modeling to evaluate methane yield from biogas in a laboratory-scale anaerobic bioreactor. *Bioresour. Technol.* **2016**, *217*, 90–99. [[CrossRef](#)]
84. Antwi, P.; Li, J.; Boadi, P.O.; Meng, J.; Shi, E.; Deng, K.; Bondinuba, F.K. Estimation of biogas and methane yields in an UASB treating potato starch processing wastewater with backpropagation artificial neural network. *Bioresour. Technol.* **2017**, *228*, 106–115. [[CrossRef](#)] [[PubMed](#)]
85. Ihunegbo, F.N.; Madsen, M.; Esbensen, K.H.; Holm-Nielsen, J.B.; Halstensen, M. Acoustic chemometric prediction of total solids in bioslurry: A full-scale feasibility study for on-line biogas process monitoring. *Chemom. Intell. Lab. Syst.* **2012**, *110*, 135–143. [[CrossRef](#)]
86. Qdais, H.A.; Hani, K.B.; Shatnawi, N. Modeling and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm. *Resour. Conserv. Recycl.* **2010**, *54*, 359–363. [[CrossRef](#)]
87. Dibaba, O.; Lahiri, S.; T'Jonck, S.; Dutta, A. Experimental and Artificial Neural Network Modeling of a Upflow Anaerobic Contactor (UAC) for Biogas Production from Vinasse. *Int. J. Chem. React. Eng.* **2016**, *14*, 1241–1254. [[CrossRef](#)]
88. Barik, D.; Murugan, S. An Artificial Neural Network and Genetic Algorithm Optimized Model for Biogas Production from Co-digestion of Seed Cake of Karanja and Cattle Dung. *Waste Biomass Valoriz.* **2015**, *6*, 1015–1027. [[CrossRef](#)]
89. Oloko-Oba, M.I.; Taiwo, A.E.; Ajala, S.O.; Solomon, B.O.; Betiku, E. Performance evaluation of three different-shaped bio-digesters for biogas production and optimization by artificial neural network integrated with genetic algorithm. *Sustain. Energy Technol. Assess.* **2018**, *26*, 116–124. [[CrossRef](#)]
90. Zareei, S.; Khodaei, J. Modeling and optimization of biogas production from cow manure and maize straw using an adaptive neuro-fuzzy inference system. *Renew. Energy* **2017**, *114*, 423–427. [[CrossRef](#)]
91. Kana, E.G.; Oloke, J.; Lateef, A.; Adesiyun, M. Modeling and optimization of biogas production on saw dust and other co-substrates using Artificial Neural network and Genetic Algorithm. *Renew. Energy* **2012**, *46*, 276–281. [[CrossRef](#)]
92. Akbaş, H.; Bilgen, B.; Turhan, A.M. An integrated prediction and optimization model of biogas production system at a wastewater treatment facility. *Bioresour. Technol.* **2015**, *196*, 566–576. [[CrossRef](#)]
93. Nasr, N.; Hafez, H.; Naggat, M.H.E.; Nakhla, G. Application of artificial neural networks for modeling of biohydrogen production. *Int. J. Hydrogen Energy* **2013**, *38*, 3189–3195. [[CrossRef](#)]
94. Whiteman, J.; Kana, E. Comparative Assessment of the Artificial Neural Network and Response Surface Modelling Efficiencies for Biohydrogen Production on Sugar Cane Molasses. *BioEnergy Res.* **2014**, *7*, 295–305. [[CrossRef](#)]
95. Ren, J.; Gao, S.; Tan, S.; Dong, L. Prediction of the yield of biohydrogen under scanty data conditions based on GM (1, N). *Int. J. Hydrogen Energy* **2013**, *38*, 13198–13203. [[CrossRef](#)]
96. Prakasham, R.; Sathish, T.; Brahmaiah, P. Imperative role of neural networks coupled genetic algorithm on optimization of biohydrogen yield. *Int. J. Hydrogen Energy* **2011**, *36*, 4332–4339. [[CrossRef](#)]

97. Aghbashlo, M.; Shamshirband, S.; Tabatabaei, M.; Yee, P.L.; Larimi, Y.N. The use of ELM-WT (extreme learning machine with wavelet transform algorithm) to predict exergetic performance of a DI diesel engine running on diesel/biodiesel blends containing polymer waste. *Energy* **2016**, *94*, 443–456. [[CrossRef](#)]
98. Ezzatzadegan, L.; Morad, N.A.; Yusof, R. Prediction and optimization of ethanol concentration in biofuel production using fuzzy neural network. *J. Teknol.* **2016**, *78*. [[CrossRef](#)]
99. del Rio-Chanona, E.A.; Wagner, J.L.; Ali, H.; Fiorelli, F.; Zhang, D.; Hellgardt, K. Deep learning-based surrogate modeling and optimization for microalgal biofuel production and photobioreactor design. *AIChE J.* **2019**, *65*, 915–923. [[CrossRef](#)]
100. Ismail, H.M.; Ng, H.K.; Queck, C.W.; Gan, S. Artificial neural networks modelling of engine-out responses for a light-duty diesel engine fuelled with biodiesel blends. *Appl. Energy* **2012**, *92*, 769–777. [[CrossRef](#)]
101. Sharon, H.; Jayaprakash, R.; Sundaresan, A.; Karuppasamy, K. Biodiesel production and prediction of engine performance using SIMULINK model of trained neural network. *Fuel* **2012**, *99*, 197–203. [[CrossRef](#)]
102. Javed, S.; Murthy, Y.S.; Baig, R.U.; Rao, D.P. Development of ANN model for prediction of performance and emission characteristics of hydrogen dual fueled diesel engine with Jatropa Methyl Ester biodiesel blends. *J. Nat. Gas Sci. Eng.* **2015**, *26*, 549–557. [[CrossRef](#)]
103. Canakci, M.; Ozsezen, A.N.; Arcaklioglu, E.; Erdil, A. Prediction of performance and exhaust emissions of a diesel engine fueled with biodiesel produced from waste frying palm oil. *Expert Syst. Appl.* **2009**, *36*, 9268–9280. [[CrossRef](#)]
104. Oğuz, H.; Saritas, I.; Baydan, H.E. Prediction of diesel engine performance using biofuels with artificial neural network. *Expert Syst. Appl.* **2010**, *37*, 6579–6586. [[CrossRef](#)]
105. Barma, S.; Das, B.; Giri, A.; Majumder, S.; Bose, A. Back propagation artificial neural network (BPANN) based performance analysis of diesel engine using biodiesel. *J. Renew. Sustain. Energy* **2013**, *3*, 013101. [[CrossRef](#)]
106. Çelebi, K.; Uludamar, E.; Tosun, E.; Yıldızhan, Ş.; Aydın, K.; Özcanlı, M. Experimental and artificial neural network approach of noise and vibration characteristic of an unmodified diesel engine fuelled with conventional diesel, and biodiesel blends with natural gas addition. *Fuel* **2017**, *197*, 159–173. [[CrossRef](#)]
107. Javed, S.; Baig, R.U.; Murthy, Y.S. Study on noise in a hydrogen dual-fuelled zinc-oxide nanoparticle blended biodiesel engine and the development of an artificial neural network model. *Energy* **2018**, *160*, 774–782. [[CrossRef](#)]
108. Aydogan, H.; Altun, A.A.; Ozcelik, A.E. Performance analysis of a turbocharged diesel engine using biodiesel with back propagation artificial neural network. *Energy Educ. Sci. Technol. Part A* **2011**, *28*, 459–468.
109. Shojaeefard, M.; Etghani, 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](#)]
110. Sharma, A.; Sahoo, P.K.; Tripathi, R.; Meher, L.C. Artificial neural network-based prediction of performance and emission characteristics of CI engine using polanga as a biodiesel. *Int. J. Ambient Energy* **2016**, *37*, 559–570. [[CrossRef](#)]
111. Omidvarborna, H.; Kumar, A.; Kim, D. Artificial neural network prediction of NOx emissions from E GR and non-EGR engines running on soybean biodiesel fuel (B5) during cold idle mode. *Environ. Prog. Sustain. Energy* **2016**, *35*, 1537–1544. [[CrossRef](#)]
112. Karthickeyan, V.; Balamurugan, P.; Rohith, G.; Senthil, R. Developing of ANN model for prediction of performance and emission characteristics of VCR engine with orange oil biodiesel blends. *J. Braz. Soc. Mech. Sci. Eng.* **2017**, *39*, 2877–2888. [[CrossRef](#)]
113. Menon, P.R.; Krishnasamy, A. A composition-based model to predict and optimize biodiesel-fuelled engine characteristics using artificial neural networks and genetic algorithms. *Energy Fuels* **2018**, *32*, 11607–11618. [[CrossRef](#)]
114. Ghobadian, B.; Rahimi, H.; Nikbakht, A.; Najafi, G.; Yusaf, T. Diesel engine performance and exhaust emission analysis using waste cooking biodiesel fuel with an artificial neural network. *Renew. Energy* **2009**, *34*, 976–982. [[CrossRef](#)]
115. Pai, P.S.; Rao, B.S. Artificial neural network based prediction of performance and emission characteristics of a variable compression ratio CI engine using WCO as a biodiesel at different injection timings. *Appl. Energy* **2011**, *88*, 2344–2354.
116. Muralidharan, K.; Vasudevan, D. Applications of artificial neural networks in prediction of performance, emission and combustion characteristics of variable compression ratio engine fuelled with waste cooking oil biodiesel. *J. Braz. Soc. Mech. Sci. Eng.* **2015**, *37*, 915–928. [[CrossRef](#)]
117. Najafi, B.; Faizollahzadeh Ardabili, S.; Mosavi, A.; Shamshirband, S.; Rabczuk, T. An intelligent artificial neural network-response surface methodology method for accessing the optimum biodiesel and diesel fuel blending conditions in a diesel engine from the viewpoint of exergy and energy analysis. *Energies* **2018**, *11*, 860. [[CrossRef](#)]
118. Kannan, G.; Balasubramanian, K.; Anand, R. Artificial neural network approach to study the effect of injection pressure and timing on diesel engine performance fueled with biodiesel. *Int. J. Automot. Technol.* **2013**, *14*, 507–519. [[CrossRef](#)]
119. Jaliliantabar, F.; Ghobadian, B.; Najafi, G.; Yusaf, T. Artificial neural network modeling and sensitivity analysis of performance and emissions in a compression ignition engine using biodiesel fuel. *Energies* **2018**, *11*, 2410. [[CrossRef](#)]
120. Kurtgoz, Y.; Karagoz, M.; Deniz, E. Biogas engine performance estimation using ANN. *Eng. Sci. Technol. Int. J.* **2017**, *20*, 1563–1570. [[CrossRef](#)]
121. Aydogan, H. Prediction of diesel engine performance, emissions and cylinder pressure obtained using Bioethanol-biodiesel-diesel fuel blends through an artificial neural network. *J. Energy S. Afr.* **2015**, *26*, 74–83. [[CrossRef](#)]
122. Ilangkumaran, M.; Sakthivel, G.; Nagarajan, G. Artificial neural network approach to predict the engine performance of fish oil biodiesel with diethyl ether using back propagation algorithm. *Int. J. Ambient Energy* **2016**, *37*, 446–455. [[CrossRef](#)]

123. Tosun, E.; Aydin, K.; Bilgili, M. Comparison of linear regression and artificial neural network model of a diesel engine fueled with biodiesel-alcohol mixtures. *Alex. Eng. J.* **2016**, *55*, 3081–3089. [[CrossRef](#)]
124. Dharma, S.; Hassan, M.H.; Ong, H.C.; Sebayang, A.H.; Silitonga, A.S.; Kusumo, F.; Milano, J. Experimental study and prediction of the performance and exhaust emissions of mixed *Jatropha curcas*-*Ceiba pentandra* biodiesel blends in diesel engine using artificial neural networks. *J. Clean. Prod.* **2017**, *164*, 618–633. [[CrossRef](#)]
125. Najafi, B.; Akbarian, E.; Lashkarpour, S.M.; Aghbashlo, M.; Ghaziaskar, H.S.; Tabatabaei, M. Modeling of a dual fueled diesel engine operated by a novel fuel containing glycerol triacetate additive and biodiesel using artificial neural network tuned by genetic algorithm to reduce engine emissions. *Energy* **2019**, *168*, 1128–1137. [[CrossRef](#)]
126. Ozgur, T.; Tuccar, G.; Ozcanli, M.; Aydin, K. Prediction of emissions of a diesel engine fueled with soybean biodiesel using artificial neural networks. *Energy Educ. Sci. Technol. Part A Energy Sci. Res.* **2011**, *27*, 301–312.
127. ZareNezhad, B.; Aminian, A. Accurate prediction of surface tension of biodiesel fuels at different operating conditions using a neuro-fuzzy model. *J. Mol. Liq.* **2015**, *207*, 206–210. [[CrossRef](#)]
128. Gopalakrishnan, K.; Mudgal, A.; Hallmark, S. Neuro-fuzzy approach to predictive modeling of emissions from biodiesel powered transit buses. *Transport* **2011**, *26*, 344–352. [[CrossRef](#)]
129. Mostafaei, M. ANFIS models for prediction of biodiesel fuels cetane number using desirability function. *Fuel* **2018**, *216*, 665–672. [[CrossRef](#)]
130. Sakthivel, G.; Sivaraja, C.; Ikua, B.W. Prediction OF CI engine performance, emission and combustion parameters using fish oil as a biodiesel by fuzzy-GA. *Energy* **2019**, *166*, 287–306. [[CrossRef](#)]
131. Sakthivel, G. Prediction of CI engine performance, emission and combustion characteristics using fish oil as a biodiesel at different injection timing using fuzzy logic. *Fuel* **2016**, *183*, 214–229. [[CrossRef](#)]
132. Debnath, R.; Sastry, G.R.K.; Rai, R.N. An experimental investigation of performance and emission of thumba biodiesel using butanol as an additive in an IDI CI engine and analysis of results using multi-objective fuzzy-based genetic algorithm. *Environ. Sci. Pollut. Res.* **2019**, *26*, 2281–2296. [[CrossRef](#)] [[PubMed](#)]
133. Ardabili, S.; Najafi, B.; Shamshirband, D. Fuzzy Logic Method for the prediction of cetane number using carbon number, double bounds, iodine and saponification values of biodiesel fuels. *Environ. Progress Sustain. Energy* **2019**, *38*, 584–599. [[CrossRef](#)]
134. Silitonga, A.S.; Masjuki, H.H.; Ong, H.C.; Sebayang, A.H.; Dharma, S.; Kusumo, F.; Siswanto, J.; Milano, J.; Daud, K.; Mahlia, T.M.I.; et al. Evaluation of the engine performance and exhaust emissions of biodiesel-bioethanol-diesel blends using kernel-based extreme learning machine. *Energy* **2018**, *159*, 1075–1087. [[CrossRef](#)]
135. Silitonga, A.S.; Hassan, M.H.; Ong, H.C.; Kusumo, F. palm oil methyl ester blends as biodiesel. *Environ. Sci. Pollut. Res.* **2017**, *24*, 25383–25405. [[CrossRef](#)]
136. Wong, P.K.; Wong, K.I.; Vong, C.M.; Cheung, C.S. Modeling and optimization of biodiesel engine performance using kernel-based extreme learning machine and cuckoo search. *Renew. Energy* **2015**, *74*, 640–647. [[CrossRef](#)]
137. Wong, K.I.; Wong, P.K. Optimal calibration of variable biofuel blend dual-injection engines using sparse Bayesian extreme learning machine and metaheuristic optimization. *Energy Convers. Manag.* **2017**, *148*, 1170–1178. [[CrossRef](#)]
138. Wong, K.I.; Wong, P.K.; Cheung, C.S.; Vong, C.M. Modeling and optimization of biodiesel engine performance using advanced machine learning methods. *Energy* **2013**, *55*, 519–528. [[CrossRef](#)]
139. Wong, K.I.; Vong, C.M.; Wong, P.K.; Luo, J. Sparse Bayesian extreme learning machine and its application to biofuel engine performance prediction. *Neurocomputing* **2015**, *149*, 397–404. [[CrossRef](#)]
140. Alves, J.C.L.; Poppi, R.J. Biodiesel content determination in diesel fuel blends using near infrared (NIR) spectroscopy and support vector machines (SVM). *Talanta* **2013**, *104*, 155–161. [[CrossRef](#)] [[PubMed](#)]
141. Maheshwari, N.; Balaji, C.; Ramesh, A. A nonlinear regression based multi-objective optimization of parameters based on experimental data from an IC engine fueled with biodiesel blends. *Biomass Bioenergy* **2011**, *35*, 2171–2183. [[CrossRef](#)]
142. Shamshirband, S.; Tabatabaei, M.; Aghbashlo, M.; Yee, L.; Petković, D. Support vector machine-based exergetic modelling of a DI diesel engine running on biodiesel–diesel blends containing expanded polystyrene. *Appl. Therm. Eng.* **2016**, *94*, 727–747. [[CrossRef](#)]