Review **Open Access**

Research Advances on Machine Learning Technologies for Enhanced Biodiesel Production: A Comprehensive Review

Joshua Ajao[1], Opeyemi Akanbi[2*], Israel Akinwole[1, 3], Abiodun Adedokun[1], Oluwakunle Ogunsakin[4], Akinsanmi Ige[2]*

[1] *Department of Chemical Engineering, Ladoke Akintola University of Technology, Ogbomoso, Nigeria*
[2] *Department of Pure and Applied Physics, Ladoke Akintola University of Technology, Ogbomoso, Nigeria*
[3] *Department of Chemical Nano-Engineering, Aix-Marseille, University, 13013, Marseille, France*
[4] *Department of Geology, Geological Engineering and Petroleum Engineering, Missouri, USA*

Received July 25, 2023; Accepted November 13, 2023

Abstract

The consumption of fossil fuels has experienced a significant increase in recent decades, despite the various challenges it poses, including air pollution, environmental degradation, health issues, and limited resources. In order to mitigate these concerns and capitalize on the environmental benefits and availability of alternative energy sources, biofuels, such as biodiesel, have emerged as viable substitutes for fossil fuels. Biodiesel production, however, is a complex process that involves identifying intricate nonlinear relationships between input and output data. To effectively design, manage, control, optimize, and monitor biodiesel production systems, accurate and efficient modeling tools like machine learning (ML) and artificial intelligence (AI) are necessary. Among the different modeling methods used in biodiesel production, machine learning has shown great potential in providing highly accurate predictions. Inspired by the autolearning and self-improving capabilities of the human brain, machine learning techniques offer superior performance in solving the complex challenges associated with (trans)esterification processes, physicochemical properties, and real-time monitoring of biodiesel systems. Applications of machine learning in the production phase encompass optimization and estimation of biodiesel quality, determination of process conditions and quantities, estimation of emissions composition and temperature, and analysis of motor performance. Key input parameters include oil and catalyst types, methanol-to-oil ratio, catalyst concentration, reaction time, domain, and frequency, while the output parameter of interest is fatty methyl acid ester. This paper presents a comprehensive review that discusses the advantages, disadvantages, and diverse applications of machine learning technologies in biodiesel production. It primarily focuses on recently published articles spanning from 2010 to 2021, with the aim of providing valuable insights for decision-making, optimization, modeling, control, monitoring, and forecasting of biodiesel production processes. By leveraging machine learning techniques, biodiesel producers can enhance their operational efficiency, improve product quality, and make informed decisions to ensure the sustainability and effectiveness of biodiesel production.

Keywords: *Biodiesel production; Machine learning; Artificial intelligence; Biofuel; Energy conversion efficiency.*

1. Introduction

In recent years, biodiesel has attracted a lot of attention and importance as a sustainable and eco-friendly substitute for traditional fossil fuels [1-2]. Reducing greenhouse gas emissions, improving energy security, and using sustainable feedstock sources are just a few of the possible advantages of producing and using biodiesel [3-5]. In comparison to conventional petroleum-based diesel fuel, biodiesel is noted for having fewer carbon dioxide (CO_2) emissions. It

is created using resources that can be replenished, like vegetable oils and animal fats, which can help to lessen greenhouse gas emissions and slow down climate change [6-7]. Additionally, dangerous chemicals like sulfur and particulate matter, which are linked to air pollution and related health hazards, are present at lower amounts in biodiesel. Production of biodiesel lessens reliance on fossil fuels and can help countries achieve energy security [8,112]. Biodiesel production can improve domestic energy output and lessen reliance on imported petroleum by using locally accessible feedstock sources, such as agricultural crops or waste materials [111-113]. Additionally, sustainable biodiesel feedstock production can help agricultural communities flourish economically and in terms of rural development [9-10]. In recent decades, achieving universal social, economic, and industrial progress has been significantly hampered by the ongoing need for affordable and efficient energy to meet rising consumption [11-12]. It is clear that the current energy supply situation cannot sustainably cover the needs of the entire world community [111-113]. Fossil-based fuels, notably coal, oil, and gas, have dominated the world's energy supply for more than 150 years [13].

The overall amount of energy consumed worldwide has significantly increased, going from 109,858 TW-hours (TWh) in 2000 to 127,0232 TWh, 151,100 TWh, and 163,709 TWh in 2005, 2015, and 2021, respectively [14]. Global population increased from 6.1 billion in 2000 to 6.5 billion in 2005, 7.4 billion in 2015, and 7.9 billion in 2021, accordingly (UN, 2022). The sustained increase in global energy use can be partly attributed to the faster population growth. Oil, coal, and gas contributed 33.1 percent, 27 percent, and 24.3 percent, respectively, to the world's energy supply as of 2019, which accounts for approximately 84.3 percent of it. Renewable energy sources accounted for only 11.3 percent of the global energy supply [14]. Waste products and byproducts from various industries, such as used cooking oil, animal fats from slaughterhouses, or crop residues, can be used in the production of biodiesel. This presents a chance to lessen waste production, support recycling and resource efficiency, and support the circular economy's guiding principles. Biodiesel manufacturing leads to a more sustainable and effective use of resources by transforming waste materials into a valuable fuel source [15-16]. Around the world, governments and regulatory organizations have recognized the potential of biodiesel production and have put in place laws and incentives to encourage its use. These consist of mix requirements, tax breaks, research funding, and renewable energy targets. An environment that is beneficial for research, development, and application of cutting-edge technology in biodiesel production processes is created by the support of policymakers and the availability of supportive regulatory frameworks. The ability to optimize and enhance different areas of biodiesel production is the relevance of integrating machine learning technologies into the process [17-18]. Machine learning methods can aid improve biodiesel quality prediction, reaction parameter optimization, process management, and feedstock selection [18-19]. Machine learning can help make biodiesel production more cost-effective and sustainable by utilizing data-driven models and predictive analytics. This will increase yield, lower costs, and guarantee consistent product quality [18-19].

Technologies like machine learning have become effective tools in many businesses, altering procedures and decision-making through data-driven methods [20]. Due to its ability to increase process efficiency, optimize production parameters, and improve overall product quality, machine learning's use in the manufacturing of biodiesel has attracted a lot of attention recently [19]. This in-depth overview examines how machine learning technologies are used to improve biodiesel production, emphasizing their importance and potential advantages. The study of creating algorithms and models that can learn from data and make predictions or judgments without being explicitly programmed is known as machine learning, a subfield of artificial intelligence (AI). Systems can use it to automatically recognize patterns, draw conclusions, and generate predictions or judgments based on incoming data. Large-scale data processing, complicated pattern detection, and important insights from machine learning algorithms can all lead to advancements and improvements. The use of machine learning technology to improve the efficiency, sustainability, and quality of biodiesel production processes has a huge potential [17]. Large datasets from multiple biodiesel manufacturing phases can be

examined by machine learning algorithms, which can find hidden patterns, discover correlations, and produce useful predictive models [18]. Operators are then able to adjust process parameters, increase overall production performance, and make data-informed decisions. At several stages of the biodiesel production process, machine learning can be used to address particular problems and spur advancements. These applications include:

- Feedstock selection: Machine learning algorithms may assess various feedstocks' qualities and forecast their potential for biodiesel synthesis, optimizing feedstock choice and lowering costs.
- Reaction optimization: By anticipating the best catalyst concentrations, reaction conditions, and durations, machine learning can enhance transesterification reactions to increase biodiesel yield and decrease undesirable byproducts.
- Process monitoring and control: In order to maintain process stability and improve performance, machine learning algorithms can monitor key process parameters in real-time, spot abnormalities or deviations, and offer proactive control measures.
- Quality prediction: The quality of the end product can be improved by using machine learning models to forecast biodiesel quality metrics such as viscosity, density, and cetane number based on input data.

While machine learning technologies have many advantages, there are several obstacles to their successful application in the biodiesel industry. These include the accessibility of high-quality and representative datasets, integration with current production systems, interpretability of complicated machine learning models, and the requirement for domain expertise to properly use and interpret the results [21-22].

The effectiveness and potential of machine learning in the manufacture of biodiesel have been demonstrated in a number of studies, contributing to significant improvements in the industry. Notably, the research done by Garg and Jain and Moayedi *et al.* [23-24] has made major contributions to the use of machine learning technologies in the manufacture of biodiesel. A study by Garg *et al.* [23] examined how machine learning models may be used to predict the characteristics of biodiesel. To forecast crucial biodiesel qualities like density, viscosity, and flash point, they created regression models based on different machine learning algorithms. The models were trained to correctly predict these qualities based on input variables using a large dataset of biodiesel samples with known properties. The study showed that machine learning models could accurately predict the qualities of biodiesel, providing a quicker and more affordable substitute for conventional laboratory testing procedures. On the other hand, Ong *et al.* [25] looked at the use of machine learning to optimize the production parameters for biodiesel. To identify the ideal circumstances for transesterification processes, they used an optimization algorithm with machine learning methods. They created prediction algorithms that may direct the selection of the ideal reaction conditions by examining a sizable dataset of process parameters and biodiesel yield. The study showed how machine learning may be used to optimize biodiesel production, increase yield, and lower production costs. These studies demonstrate how machine learning has been successfully applied to the manufacturing of biodiesel and show how it has the potential to improve a number of different areas of the procedure [19,26]. Researchers can create precise forecasts, enhance process parameters, and boost overall production efficiency by utilizing machine learning algorithms and models. These developments have a big impact on the biodiesel business and have the potential to boost output, cut costs, and improve sustainability. While the works by Garg and Jain and Kolakoti *et al.* [23-24] have made significant contributions, it is crucial to note that there is still continuing research in the area of machine learning in biodiesel production. In order to further increase the efficiency and effectiveness of machine learning technologies in the biodiesel manufacturing processes, efforts are continuously made to investigate new methodologies, algorithms, and data-driven approaches.

This comprehensive overview aims to give a thorough analysis of the application of machine learning technologies in the production of biodiesel. It investigates various machine learning techniques and algorithms, applications, and potential effects on process improvement, cost cutting, and sustainability. It also talks about the field's difficulties, constraints, and future

directions while indicating areas that could use more study and improvement. The manufacture of biodiesel can gain from data-driven insights, increased process effectiveness, and improved product quality by utilizing the power of machine learning technologies. The thorough overview seeks to aid in the biodiesel industry's comprehension and implementation of machine learning, supporting its transition into a more effective, long-lasting, and environmentally friendly energy source. Therefore, it is essential to examine how machine learning technologies are being used in the biodiesel production process in order to comprehend the developments, difficulties, and potential advantages that these technologies can bring to the industry, ultimately assisting in the shift to a more sustainable and cleaner energy future.

2. Biodiesel production process

2.1. Overview of traditional biodiesel production methods

Traditional biodiesel manufacturing techniques have been used for many years and consist of a number of tried-and-true procedures. These processes typically involve a transesterification reaction to produce biodiesel from spent cooking oil, animal fats, or vegetable oils [1,28]. An outline of the main steps in producing traditional biodiesel is provided below: Preparing the feedstock is the first stage in the typical biodiesel production process. Obtaining vegetable oils, animal fats, or used cooking oil from reputable sources is required for this. To guarantee the quality and purity of the oil, feedstock may go through pre-treatment procedures such as filtration, degumming, or eliminating contaminants. The transesterification reaction is the main step in the classic biodiesel synthesis process. Triglycerides (found in the feedstock) are chemically transformed into biodiesel through a reaction with an alcohol, often methanol or ethanol, in the presence of a catalyst like sodium hydroxide or potassium hydroxide [1,28]. As a byproduct of this reaction, glycerin is separated, and biodiesel is created. To make it easier to separate the glycerin and biodiesel after the transesterification procedure, the mixture is allowed to settle. Centrifugation or gravity settling can be used to improve separation. After being separated, the biodiesel goes through washing procedures to get rid of any contaminants, catalyst residues, or water that may still be present [29-30]. After washing, the biodiesel is normally dried to remove any moisture that may have remained. Drying can be accomplished in a number of ways, including vacuum drying and desiccants. Additionally, the biodiesel is often filtered to get rid of any contaminants or particulates that could degrade its quality. Traditional biodiesel production techniques incorporate quality control checks to guarantee the final product satisfies the required criteria. To ascertain whether biodiesel samples match regulatory criteria and industry standards, important metrics including viscosity, density, flash point, acid value, and other specifications are examined [31-32]. Then, for usage in a variety of applications, certified biodiesel can be utilized as a renewable fuel or combined with petroleum diesel [33]. The biodiesel industry is built on the widespread adoption of traditional biodiesel production techniques [34-35]. Even though these techniques have been proven successful, ongoing research and development is being done with the aim of improving the production procedure, increasing efficiency, and investigating alternative feedstocks and catalysts to further increase the sustainability and commercial viability of biodiesel production.

2.2. Challenges and limitations of conventional approaches

Challenges and limitations of conventional biodiesel production approaches:
Although standard biodiesel production methods have been utilized extensively, they also have some drawbacks that may restrict their effectiveness, sustainability, and scalability [36]. The following are some of the main issues with using traditional biodiesel production techniques:

2.2.1. Feedstock availability and cost

For the manufacturing of biodiesel, conventional methods mainly rely on edible vegetable oils like soybean oil, canola oil, or palm oil. The economic viability and sustainability of biodiesel production are threatened by the scarce supply and unstable prices of these feedstocks.

Concerns regarding the moral ramifications and potential effects on food security are also raised by the competition between the production of food and fuel [37-39].

2.2.2. Feedstock diversity and quality variations

The composition and quality of different feedstocks vary, which can have an impact on the effectiveness of the transesterification reaction and the overall production of biodiesel. Having trouble getting the best conversion and product quality might be caused by inconsistent feedstock quality, such as fluctuating levels of free fatty acids, moisture content, or contaminants [40]. For conventional biodiesel production, securing a steady and high-quality feedstock supply continues to be a considerable difficulty.

2.2.3. Energy and resource intensity

The transesterification process, which involves heating, mixing, and separation procedures, is one phase in the traditional biodiesel production process that frequently requires considerable energy inputs. These energy-intensive procedures add to the overall carbon footprint of the manufacture of biodiesel [36,41]. Additionally, using a lot of water for washing and separating might put a burden on water supplies in some areas, especially when water shortage is an issue.

2.2.4. Waste disposal and glycerin utilization

Glycerin is a byproduct of the conventional transesterification process. It can be difficult to properly dispose of or use this glycerin. Inadequate methods of treatment or disposal might result in environmental pollution and problems with waste management [42-43]. For conventional biodiesel production, it is still difficult to come up with economically sensible and long-lasting uses for glycerin byproducts.

2.2.5. Production efficiency and scale-up

It might be challenging to scale up biodiesel production using traditional methods. Maintaining continuous product quality, enhancing reaction conditions for higher volumes, and controlling intricate supply chains are challenges [44]. The rising demand for biodiesel might not always be supplied by conventional techniques, which might necessitate substantial expenditures in infrastructure and technological advancements to obtain increased production capabilities [45-46].

2.2.6. Environmental impact

Although standard biodiesel manufacturing methods still have some negative effects on the environment, biodiesel is typically regarded as a more environmentally friendly fuel than fossil fuels [47]. Land use changes, potential deforestation brought on by the development of feedstock crops, waste disposal-related water contamination, and energy use in manufacturing processes are a few of these. The sustainability of biodiesel production over the long run depends on reducing these environmental effects and implementing sustainable procedures.

2.3. Potential benefits of integrating machine learning technologies

There are a number of potential advantages to incorporating machine learning technology into the biodiesel manufacturing process, which can greatly improve productivity, sustainability, and overall performance. Among the main advantages are:

2.3.1. Process optimization

Large amounts of data from the biodiesel production processes can be analyzed using machine learning algorithms, which can then be used to spot intricate patterns and relationships that may be difficult to spot using more conventional techniques [19]. Machine learning can increase conversion efficiency, optimize biodiesel yield, and decrease the generation of undesirable byproducts by optimizing process variables like reaction conditions, catalyst concentrations, and feedstock compositions [48]. Cost reductions, lower resource usage, and enhanced process performance can all result from this optimization.

2.3.2. Predictive analytics

Based on input factors including feedstock composition, reaction parameters, and process circumstances, machine learning models can be created to forecast biodiesel quality characteristics like viscosity, density, flash point, and cetane number [19]. Due to the early discovery of potential quality problems made possible by these predictive models, timely changes and interventions are possible. Machine learning increases the marketability and dependability of biodiesel while also maintaining constant product quality, resulting in higher consumer satisfaction and a greater degree of regulatory compliance. Machine learning models can be developed to forecast the qualities of biodiesel, such as viscosity, density, flash point, and cetane number. The models can be generated based on input information such as feedstock composition, reaction parameters, and process circumstances [19]. These prediction models allow for the early identification of potential quality issues, which enables prompt adjustments and interventions. Machine learning boosts the biodiesel's marketability and dependability while also preserving a consistent level of product quality, leading to increased consumer satisfaction and a higher level of regulatory compliance.

2.3.3. Real-time monitoring and control

To find anomalies, deviations, or unfavorable situations, machine learning algorithms can examine real-time data from sensors and process monitoring systems. As a result, proactive decision-making and control strategies are made possible, enabling operators to quickly resolve problems and improve process performance [49]. Process stability, downtime, and overall operational efficiency are all improved through real-time monitoring and control.

2.3.4. Data-driven decision-making

Data-driven decision-making is made possible in the manufacture of biodiesel using machine learning algorithms. Operators can make more educated decisions by using historical data to acquire insights into production trends, patterns, and correlations [19,50]. This can involve maintaining inventory levels, forecasting maintenance needs, and fine-tuning manufacturing schedules, among other things. Data-driven decision-making increases resource allocation, cuts costs, and operational efficiency [51].

2.3.5. Advanced fault detection and maintenance

Equipment used in the manufacture of biodiesel can benefit from using machine learning techniques for proactive maintenance and issue finding [19]. Machine learning algorithms can find patterns suggestive of probable equipment breakdowns or maintenance requirements by examining sensor data and historical maintenance records. This makes it possible to schedule preventive maintenance, which lowers unplanned downtime, increases equipment reliability, and lowers repair costs [52].

2.3.6. Sustainability and feedstock optimization

The selection and use of biodiesel feedstocks can be sustainably optimized with the use of machine learning [18]. Machine learning algorithms can suggest the most environmentally friendly and economically advantageous feedstock sources by examining data on feedstock availability, composition, and market circumstances. This increases the sustainability of biodiesel production overall by encouraging responsible sourcing, lowering dependency on finite resources, and supporting the use of non-food feedstocks.

The industry could be revolutionized by incorporating machine learning technology into biodiesel manufacturing processes that increase process efficiency, improve product quality, optimize resource use, and lower operational costs [53]. These advantages make machine learning an appealing tool for the biodiesel business, enabling its transformation towards a more effective, eco-friendly, and financially viable energy source. They also come with the promise for enhanced scalability and sustainability.

3. Fundamentals of machine learning and applications in biodiesel production

3.1. Introduction to machine learning concepts and algorithms

Within the broader subject of artificial intelligence, machine learning has become a potent discipline that enables computers to learn from data and make predictions or judgments without being explicitly programmed [54]. It entails the creation of algorithms and models that can recognize patterns, draw conclusions, and decide wisely based on given data. Machine learning is based on the idea of developing models that can make precise predictions or judgments using new, unforeseen data by training algorithms on old data. In order to produce insightful knowledge, this learning process entails drawing out significant patterns, correlations, and trends from the data.

3.1.1. Supervised learning

One of the fundamental subfields of machine learning is supervised learning. In this method, input examples are linked to appropriate target labels or outcomes, and the algorithm is trained on labeled data. The algorithm picks up generalization skills from the labeled data and uses them to predict or categorize brand-new, untainted data points.

3.1.2. Unsupervised learning

Another crucial area of machine learning is unsupervised learning, which focuses on finding structures or patterns in unlabeled data. Contrary to supervised learning, the algorithm is not given any predetermined labels or results. Instead, without any external instruction, the algorithm probes the data to find underlying structures, groupings, or linkages.

3.1.3. Feature selection and data preprocessing

A critical stage in machine learning is feature selection, which entails locating and choosing pertinent features or variables from the input data that are most helpful to the learning job. In order to ensure that the data is adequate for machine learning algorithms, appropriate data preprocessing techniques are frequently used, such as addressing missing data, normalization, or scaling.

3.1.4. Common machine learning algorithms

There are various machine learning algorithms used for different types of learning tasks. Some commonly used algorithms include:

- Decision trees: Tree-like models that make decisions based on a series of questions or conditions.
- Random forests: Ensemble models composed of multiple decision trees, providing improved accuracy and robustness.
- Support vector machines (SVM): Algorithms that classify data points by finding an optimal hyperplane that separates different classes.
- Naive bayes: Probabilistic algorithms based on Bayes' theorem, commonly used for classification tasks.
- K-nearest neighbors (KNN): An algorithm that classifies data points based on the majority vote of their nearest neighbors.

3.1.5. Neural networks and deep learning

Neural networks are effective machine learning models because they were inspired by the way the human brain is organized. Neural networks with numerous hidden layers are the focus of the machine learning field of deep learning. Deep learning has excelled in a number of fields, including speech recognition, natural language processing, and image recognition. For data analysis, pattern identification, and decision-making, machine learning principles and algorithms offer a potent toolkit. Businesses and researchers may create accurate predictions, automate challenging operations, and gain useful insights from their data by utilizing these algorithms. Machine learning is a rapidly evolving subject that is continually developing new algorithms and methods to solve ever-more-complex issues and increase the precision and effectiveness of data-driven decision-making.

3.2. Supervised, unsupervised, and reinforcement learning techniques

The manufacture of biodiesel can make use of supervised, unsupervised, and reinforcement learning approaches in a variety of ways [55]. Here is a description of each sort of machine learning approach and how it can be used:

3.2.1. Supervised learning

In order to produce biodiesel, supervised learning methods often include training models on labeled data, where both the input data and the intended output are given. Some uses of supervised learning in the creation of biodiesel include:

- Feedstock composition prediction: The feedstock compositions and accompanying biodiesel attributes of labeled datasets can be used to train supervised learning algorithms. The best feedstock composition can then be predicted using these models to achieve desired biodiesel qualities, such as viscosity, density, or flash point.
- Biodiesel quality assessment: Supervised learning models can be used to evaluate the quality of biodiesel generated, forecast attributes like viscosity, density, or cetane number, and categorize biodiesel samples according to quality standards by training on labeled datasets of biodiesel samples with known quality parameters [56].

3.2.2. Unsupervised learning

When there are no established labels or expected outputs for the data, unsupervised learning algorithms are used. Techniques for unsupervised learning are helpful for exploratory research and identifying structures or trends in data on biodiesel production [57]. Unsupervised learning is used in the creation of biodiesel in a variety of ways:

- Clustering analysis: Based on similarity in feedstock composition, process variables, or product quality, biodiesel production data can be grouped using unsupervised learning methods like k-means clustering or hierarchical clustering. This can offer information on specific groups or categories of the data.
- Anomaly detection: To locate anomalies or outliers in data on biodiesel production, unsupervised learning algorithms can be used. These anomalies may be a sign of system malfunctions, equipment failures, or data errors, allowing for the proactive mitigation of potential problems.

3.2.3. Reinforcement learning

A type of machine learning called reinforcement learning teaches an agent to make decisions based on interactions with the outside world and feedback in the form of rewards or penalties. For tasks like process management and optimization, reinforcement learning can be used in the manufacture of biodiesel. Some potential uses of reinforcement learning in the creation of biodiesel include:

- Optimal process control: In order to maximize the performance of the biodiesel production process, reinforcement learning algorithms can be trained to regulate crucial process parameters like temperature, pressure, or catalyst dosage. The agent adapts its decision-making process in response to feedback it receives regarding the quality of the biodiesel it has created.
- Resource optimization: To maximize the use of resources, such as feedstock, energy, or catalysts, in the manufacture of biodiesel, reinforcement learning techniques can be applied. The agent gains the ability to decide in a way that meets quality standards and maximizes resource efficiency.

Processes used in the manufacturing of biodiesel can gain from enhanced prediction accuracy, process optimization, quality assessment, and resource efficiency by utilizing supervised, unsupervised, and reinforcement learning techniques. By enabling data-driven decision-making, these machine learning techniques have the potential to increase the productivity, sustainability, and cost-effectiveness of biodiesel production.

3.3. Feature selection and data preprocessing for biodiesel production

Preparing data for machine learning algorithms in the biodiesel production process requires important stages like feature selection and data preprocessing. These procedures assist in ensuring the accuracy of predictions, enhancing the performance of machine learning models, and guaranteeing the quality and relevance of input data. Here is a summary of feature selection and data preparation in relation to the creation of biodiesel:

3.3.1. Feature selection

Finding and choosing the most pertinent and instructive features (input variables) from the available dataset is feature selection. The objective is to select characteristics that significantly affect the output or target variable. Feature selection can be used in the manufacturing of biodiesel to pinpoint the critical variables that influence yield, quality, or process effectiveness. Some factors to take into account when choosing features for biodiesel production include:

- Domain knowledge: The choice of crucial elements can be influenced by expert understanding of the biodiesel production process. Identifying the pertinent variables to include in the feature set can be made easier by having a thorough understanding of the chemical reactions, catalysts, feedstock compositions, and process parameters.
- Correlation analysis: Correlation analysis, for example, is a statistical technique that may be used to evaluate the correlations between variables and pinpoint highly associated aspects. To lessen computing complexity and prevent multicollinearity problems, redundant or highly correlated features could be eliminated.
- Feature importance: Decision trees and random forests are two examples of machine learning techniques that can shed light on the relative relevance of various features. Less useful features can be omitted from the analysis based on these metrics.

3.3.2. Data preprocessing

Data preparation entails preparing the data for machine learning algorithms by altering and organizing it. It seeks to address problems like missing data, outliers, scaling, and normalizing. Among the crucial phases in data preprocessing for the creation of biodiesel are:

- Handling missing data: Datasets on the manufacturing of biodiesel frequently have missing data. The gaps can be filled in using methods like imputation, where missing values are inferred based on other available data [57]. Alternatively, if it is determined that missing data significantly affects the outcomes, records with incomplete data may be omitted from the study.
- Outlier detection and treatment: Extreme values that differ greatly from the rest of the dataset are known as outliers, and they can have an impact on the precision of machine learning models [58]. By deleting them or substituting more typical data for them, outliers can be addressed after being identified using statistical techniques or visualization techniques.
- Normalization and scaling: The sizes and ranges of various dataset features may vary. To make sure that all features are on a same scale, normalization and scaling approaches can be used, such as min-max scaling or z-score normalization [59]. This assists in preventing the dominance of some features over the learning process due to their greater magnitude.
- Encoding categorical variables: Machine learning methods require categorical variables in datasets to be numerically represented, such as the type of feedstock or catalyst utilized [60]. Techniques like label encoding and one-hot encoding can be used to accomplish this.

Datasets related to the manufacture of biodiesel can be adequately prepared for machine learning algorithms by performing feature selection and data preprocessing. As a result, predictions, process optimizations, and quality assessments for the production of biodiesel will be more accurate and dependable. This guarantees that the input data is pertinent, representative, and in an appropriate format [61].

3.4. Prediction of optimal feedstock composition using machine learning

A useful strategy in the manufacture of biodiesel is the prediction of the ideal feedstock composition using machine learning, which provides knowledge and direction to enhance the production process [62]. Here is a thorough explanation of this approach:

A dataset containing a wide range of feedstock compositions, together with matching biodiesel characteristics and process variables, is first created. Taking into account their possible influence on feedstock composition and biodiesel qualities, pertinent features are chosen from the dataset. This could include impurity levels, moisture content, free fatty acid content, impurity patterns, and other compositional traits. To assure data quality and applicability for machine learning algorithms, data preparation techniques are used. Outliers are dealt with, missing data is addressed, and scaling or normalization is done as necessary. If present, categorical variables are numerically encoded. The qualities of the dataset, the difficulty of the challenge, and the required interpretability all influence the choice of machine learning algorithm. It is possible to use neural networks, support vector regression (SVR), decision tree-based models like random forest regression, and linear regression algorithms. Using a training set, the machine learning model is taught to understand the connections between the input properties (feedstock composition, process parameters), and the output variable (optimal feedstock composition). The model's performance is assessed using a different testing set in order to gauge its capacity for generalization and guard against overfitting. For model evaluation, methods like cross-validation might be utilized. The machine learning model can be used to forecast the ideal feedstock composition for new, unforeseen data points after being trained and validated. The model can be used to input desired biodiesel qualities or process goals, and it subsequently outputs the appropriate ideal feedstock composition. To meet specific biodiesel quality goals or process improvements, these forecasts help decision-makers choose the right feedstock types and their proportions. The model can be improved continuously by being updated and retrained with fresh data as it becomes available. Over time, accuracy and relevance are ensured by routine updates. Overall, machine learning's ability to forecast the ideal feedstock composition allows producers of biodiesel to make decisions based on data. It aids in maximizing resource use, enhancing the effectiveness of processes, and achieving desirable biodiesel qualities [63]. Biodiesel manufacturers may select feedstocks intelligently, improve product quality, and streamline the entire biodiesel manufacturing process by utilizing machine learning algorithms [17].

3.5. Modeling and optimization of transesterification reactions

In order to produce biodiesel, transesterification reactions must be modeled and optimized, and machine learning technologies have shown to be helpful in this regard [64]. Researchers and industry professionals can create models and optimize the transesterification process to increase efficiency and improve biodiesel yield and quality by utilizing machine learning methods and techniques [19]. Machine learning approaches are used in the modeling phase to create models that accurately forecast the conversion efficiency and biodiesel production depending on numerous parameters including feedstock composition, catalyst concentration, temperature, reaction time, and mixing intensity [26]. These models accurately depict the intricate interaction between the different input variables and the desired outcome, offering insightful information about the transesterification procedure. A dataset including a wide range of experimental or simulated data points reflecting various combinations of input factors and accompanying transesterification reaction outcomes is developed in order to train these models [65]. To account for all possible situations, the dataset includes changes in feedstock types, catalysts, reaction conditions, and process parameters.

Taking into account their influence on the transesterification process, pertinent features are chosen from the dataset. These characteristics could be the content of the feedstock, the concentration of the catalyst, the temperature, the reaction duration, and other pertinent variables [66]. Machine learning models perform better when preparing data using methods in-

cluding addressing missing data, removing outliers, and standardization. Following model selection, the dataset is prepared for training, during which the model discovers the correlations between the input variables and the intended transesterification results. To guarantee the correctness and generalizability of the model, its performance is assessed using appropriate evaluation metrics and cross-validation methods. In the optimization phase, machine learning models and optimization algorithms are integrated to find the best circumstances for reactions that would maximize biodiesel yield or produce particular attributes in the biodiesel [67]. The choice of feedstock composition, catalyst dosage, reaction temperature, and reaction duration can be guided by genetic algorithms, gradient-based approaches, or other optimization techniques. The models must be updated and modified consistently. Real-time process information and feedback loops enable adaptive control and long-term model improvement. As new data becomes available or as the process conditions change, the models can be updated and enhanced to ensure their relevance and accuracy.

In conclusion, biodiesel producers can increase process efficiency, maximize biodiesel yield, and achieve desired biodiesel qualities by modeling and optimizing transesterification processes using machine learning approaches [68]. By offering data-driven insights and decision-making tools for process optimization and control, these applications increase biodiesel production.

3.6. Monitoring and control of critical process parameters using ML algorithms

Using machine learning (ML) algorithms to monitor and regulate crucial process parameters is a potent strategy in the manufacture of biodiesel. By doing so, producers may preserve product quality, ensure the best possible process performance, and increase overall productivity [17]. The use of ML algorithms for monitoring and controlling crucial process parameters during the synthesis of biodiesel is fully explained here: Using sensors and monitoring systems, the process first gathers information on crucial process parameters. Temperature, pressure, catalyst concentration, feedstock composition, reaction duration, and other factors are examples of these parameters. For ML algorithms, the acquired data is used as the input. Preprocessing is done on the gathered data to deal with missing values, get rid of outliers, and normalize or scale the data as needed. Data preparation enhances the accuracy and reliability of the input data, making it ready for ML algorithms to analyze. Regression, neural networks, and support vector machines are a few examples of machine learning (ML) methods that are used to create models that can recognize patterns and relationships between process parameters and desired results. The historical data used to train these models contains measurements of the process parameter values as well as the related target values (e.g., biodiesel yield, product quality) [69]. The models can capture the intricate dynamics of the biodiesel generation process thanks to the training process. The ML models can be used for real-time monitoring of crucial process parameters once they have been trained. The models use the incoming sensor data to forecast and evaluate the process's current status. It is possible to identify deviations from expected values or predetermined thresholds, which can then set off alarms or messages calling for rapid attention.

ML algorithms are very good at spotting abnormalities or differences from typical process behavior. Anomalies can be found by comparing real-time sensor data with the trained model's predictions. These abnormalities could point to faulty machinery, process irregularities, or possible problems that could hinder the production of biodiesel. Insights into the underlying causes of abnormalities are provided by ML algorithms, facilitating problem detection and directing corrective measures. ML algorithms can be used for process optimization and control. The ML models can offer suggestions or adjustments to crucial process parameters by examining real-time sensor data [70]. This entails adjusting the temperature, pressure, catalyst dosage, and other factors in order to maximize the yield, quality, and energy efficiency of biodiesel. Control techniques can be iteratively improved using reinforcement learning algorithms based on feedback from the process. Algorithms for machine learning can continuously learn and get better. They can incorporate fresh data and adjust to changing process circum-

stances to function more effectively. The models are adaptable to changes in feedstock composition, equipment performance, or environmental conditions since they can be routinely retrained using current datasets. Continuous learning facilitates continual process optimization and enhances the predictive power of the models.

Biodiesel manufacturers may increase process stability, product quality, resource usage, and operational costs by utilizing ML algorithms for monitoring and controlling crucial process parameters [19]. These tools support the deployment of advanced process control strategies, encourage data-driven decision-making, and contribute to more effective and sustainable biodiesel production processes.

3.7. Quality prediction and assessment of biodiesel using machine learning

The prediction and evaluation of biodiesel quality have a lot of potential when using machine learning (ML) approaches. According to Rocabruno-Valdés *et al.* [70], ML algorithms can examine intricate correlations between input parameters and biodiesel quality indicators to make precise predictions and offer insightful information. Here is a thorough explanation of how machine learning is used in the manufacture of biodiesel to anticipate and evaluate quality:

A large dataset of biodiesel samples with related quality parameters is gathered in order to create ML models for quality prediction [56]. The collection contains information on the chemical makeup of biodiesel, reaction conditions, production parameters, and feedstock composition [71]. The handling of missing values, outliers, and normalization of the data all require the use of data preparation techniques. Based on their influence on biodiesel quality, relevant features are chosen or engineered from the dataset. Fatty acid content, kinematic viscosity, density, oxidative stability, cetane number, and other physicochemical characteristics are a few examples of these characteristics. The performance and interpretability of the ML models are improved by feature selection by removing unnecessary or redundant information. Regression algorithms, random forests, support vector machines, and neural networks are just a few examples of the ML techniques that can be used to forecast quality. The dataset is used to train the chosen algorithm with the biodiesel quality metrics as the goal variables. The relationship between the input features and the required quality indicators is learned by the model. The performance and generalization capacities of the trained ML model are evaluated using suitable evaluation criteria. Techniques for cross-validation can be used to ensure resilience and reduce overfitting. The model is verified by evaluating how well it performs on a separate test dataset that represents hypothetical biodiesel samples. The ML model can be used to forecast the quality of biodiesel for new, untested samples once it has been trained and validated. The model can calculate quality characteristics including kinematic viscosity, density, oxidative stability, and cetane number by feeding in the pertinent features. Additionally, ML models can categorize biodiesel samples in accordance with quality standards to determine whether they adhere to particular industrial or regulatory norms.

In biodiesel quality data, ML systems can identify anomalies or outliers, highlighting samples that drastically depart from the mean. These abnormalities could be a sign of process flaws, feedstock contaminants, or other elements impacting biodiesel quality. Such outlier analyses can shed light on potential problems and direct quality-improvement initiatives. By including fresh data and iteratively improving the models, machine learning algorithms for predicting the quality of biodiesel can be improved continually. As new samples and high-quality data become available, regular updates and retraining of the models assure their accuracy and relevance.

In biodiesel quality data, ML systems can identify anomalies or outliers, highlighting samples that drastically depart from the mean. These abnormalities could be a sign of process flaws, feedstock contaminants, or other elements impacting biodiesel quality. Such outlier analyses can shed light on potential problems and direct quality-improvement initiatives. By including fresh data and iteratively improving the models, machine learning algorithms for predicting the quality of biodiesel can be improved continually. As new samples and high-quality data become available, regular updates and retraining of the models assure their accuracy and relevance.

4. Machine learning models for biodiesel production

4.1. Regression models for feedstock composition prediction

Various regression algorithms are employed to develop models that accurately predict the feedstock composition. Linear regression, support vector regression (SVR), decision tree-based models like random forest regression, or more advanced techniques like neural networks can be utilized. The selected algorithm is trained using the dataset, with the feedstock composition as the target variable. The model learns the relationships between the input features and the corresponding composition. The trained regression model is evaluated using appropriate evaluation metrics to assess its performance and generalization capabilities. Cross-validation techniques may be applied to ensure robustness and mitigate overfitting. The model is validated by assessing its performance on an independent test dataset, representing unseen feedstock samples. Once the regression model is trained and validated, it can be used to predict the composition of new, unseen feedstock samples. By inputting the relevant features into the model, it can estimate the composition of different constituents in the feedstock. This prediction capability aids in feedstock selection, process optimization, and meeting specific quality targets.

Regression models for feedstock composition prediction can be continuously improved by incorporating new data and refining the models over time. Regular updates and retraining of the models ensure their accuracy and relevance as new feedstock samples and data become available. The use of regression models for feedstock composition prediction empowers biodiesel producers to make informed decisions regarding feedstock selection, process optimization, and achieving desired biodiesel quality targets [19]. These models provide valuable insights into the relationships between input variables and feedstock composition, enhancing the efficiency and effectiveness of biodiesel production processes.

4.2. Classification algorithms for biodiesel quality assessment

Various classification algorithms are employed to develop models that accurately classify biodiesel samples based on quality standards. Logistic regression, support vector machines (SVM), random forests, and neural networks are commonly used algorithms [72]. The selected algorithm is trained using the dataset, with the biodiesel quality indicators serving as the target variables. The model learns the relationships between the input features and the corresponding quality categories. The trained classification model is evaluated using appropriate evaluation metrics to assess its performance and generalization capabilities. Cross-validation techniques may be applied to ensure robustness and mitigate overfitting. The model is validated by assessing its performance on an independent test dataset, representing unseen biodiesel samples.

Once the classification model is trained and validated, it can be utilized to assess the quality of new, unseen biodiesel samples. By inputting the relevant features into the model, it can classify the samples into different quality categories or indicate whether they meet specific regulatory or industry criteria. This capability aids in quality control, compliance with standards, and decision-making regarding product acceptance or rejection. Classification algorithms can also detect anomalies or outliers in biodiesel quality data, flagging samples that deviate significantly from the norm [73]. These anomalies may indicate process deviations, impurities in the feedstock, or other factors affecting biodiesel quality. Analyzing such outliers can provide insights into potential issues and guide quality improvement measures. Classification models for biodiesel quality assessment can be continuously improved by incorporating new data and refining the models over time [72] Regular updates and retraining of the models ensure their accuracy and relevance as new samples and quality data become available.

The use of classification algorithms for biodiesel quality assessment enables producers to make informed decisions, optimize processes, and ensure compliance with quality standards [72]. These models provide valuable insights into the relationships between input variables and biodiesel quality, enhancing the efficiency and effectiveness of quality control processes in biodiesel production.

4.3. Neural networks and deep learning for process optimization

Neural networks and deep learning techniques offer significant potential for process optimization in biodiesel production [57]. These advanced machine learning methods can analyze complex relationships within biodiesel production data and provide valuable insights to optimize various aspects of the process. Here is a comprehensive insight into the use of neural networks and deep learning for process optimization in biodiesel production:

The neural network is trained using the prepared dataset, with the process parameters and input variables as the input features and the output variables as the target variable for optimization. The training process adjusts the weights and biases of the neural network based on the training data, gradually improving its performance. The model's performance is evaluated using validation data to ensure its generalization capabilities.

Once the neural network is trained and validated, it can be used for process optimization. The network takes input data representing process parameters and input variables and predicts the corresponding output variables. By iteratively adjusting the process parameters within specified constraints, the neural network can guide decision-making to optimize biodiesel yield, quality, energy efficiency, or other desired objectives. Deep learning techniques, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), can be applied for specific process optimization tasks in biodiesel production. CNNs are suitable for analyzing image or spectroscopic data, such as analyzing the biodiesel production process using images from sensors. RNNs are useful for time series analysis, such as optimizing process parameters over time. Neural networks and deep learning models can continuously learn and adapt to changes in the biodiesel production process. By incorporating real-time process data and feedback loops, the models can update and refine their predictions and optimization strategies, ensuring they remain relevant and accurate as the process conditions evolve. Neural networks and deep learning models can be integrated with optimization algorithms, such as genetic algorithms or particle swarm optimization, to further enhance process optimization. These hybrid approaches combine the learning capabilities of neural networks with the search and optimization capabilities of the optimization algorithms, enabling more efficient and effective process optimization.

The application of neural networks and deep learning for process optimization in biodiesel production allows for data-driven decision-making, improved process efficiency, and enhanced product quality [57]. These techniques enable producers to optimize process parameters, resource allocation, and energy consumption, leading to more sustainable and economically viable biodiesel production processes.

4.4. Ensemble learning approaches for enhanced accuracy

Ensemble learning approaches can be effectively used to enhance the accuracy of biodiesel production predictions and models. Ensemble learning combines multiple individual models to make more accurate and robust predictions by leveraging the strengths of each model . Here is a comprehensive insight into the use of ensemble learning approaches for enhanced accuracy in biodiesel production.

4.4.1. Bagging

Bagging, short for bootstrap aggregating, is an ensemble learning approach where multiple models are trained on different subsets of the training data. In the context of biodiesel production, multiple models, such as decision trees or neural networks, are trained on different subsets of the biodiesel production dataset [74]. The predictions from each model are then combined to make the final prediction. Bagging helps reduce the variance of the predictions and improves the overall accuracy of biodiesel production models.

4.4.2. Random Forests

Random forests is a popular ensemble learning technique that combines multiple decision tree models. Each decision tree is trained on a random subset of the training data and makes

predictions independently. The final prediction is determined by aggregating the predictions from all the decision trees in the forest. Random Forests are known for their robustness, ability to handle complex datasets, and feature importance analysis.

4.4.3. Boosting

Boosting is an ensemble learning technique that focuses on training models sequentially, where each subsequent model is trained to correct the errors of the previous models. Boosting algorithms, such as AdaBoost or Gradient Boosting, assign weights to the training instances to emphasize the misclassified samples and improve the overall model performance. In biodiesel production, boosting can be applied to enhance the accuracy of models predicting biodiesel yield, quality indicators, or process parameters.

4.4.4. Stacking

Stacking is an ensemble learning approach that combines multiple models by training a meta-model that takes the predictions of individual models as input [75]. In biodiesel production, multiple models, such as regression models or neural networks, are trained on the biodiesel production dataset [70]. The predictions from each model are then used as features for training the meta-model, which makes the final prediction. Stacking allows for more sophisticated learning and can improve the accuracy of biodiesel production models.

4.4.5. Voting

Voting is a simple yet effective ensemble learning approach where multiple models are trained independently, and the final prediction is determined based on a majority vote or weighted average of the predictions. In biodiesel production, different models with varying algorithms or parameters can be trained, and their predictions can be combined through voting to improve accuracy. Voting can be implemented as hard voting (majority) or soft voting (weighted average based on confidence scores).

Ensemble learning approaches, such as bagging, random forests, boosting, stacking, and voting, offer improved accuracy and robustness in biodiesel production predictions and models. By combining the predictions of multiple models, these approaches mitigate biases, reduce variance, and enhance the overall accuracy of biodiesel production models. Biodiesel producers can benefit from ensemble learning by making more reliable decisions, optimizing process parameters, and improving the efficiency and quality of biodiesel production.

5. Performance evaluation and validation

5.1. Evaluation metrics for assessing the performance of machine learning models

Different evaluation criteria can be used to gauge the efficiency and accuracy of machine learning models for the manufacture of biodiesel [76]. These metrics offer information on the models' accuracy in foretelling biodiesel yield, quality indicators, or other pertinent results. The effectiveness of machine learning models employed in the manufacture of biodiesel can be evaluated using some of the metrics listed below:

5.1.1. Mean absolute error (MAE)

The average absolute difference between the expected and actual values is measured by MAE. It provides a clue as to the typical size of the prediction mistakes. Better model performance, with less differences between predicted and actual values, is indicated by a lower MAE.

5.1.2. Mean squared error (MSE)

MSE calculates the average of the squared differences between the predicted and actual values. It amplifies larger errors, making it more sensitive to outliers. MSE provides a measure of the average squared deviation between predicted and actual values, and a lower MSE implies a better model fit.

5.1.3. Root mean squared error (RMSE)

The RMSE, which is the square root of the MSE, is frequently employed as an evaluation statistic to determine the overall amount of the predictions' errors. It offers a more understandable measurement of the typical difference between expected and actual values. A lower RMSE indicates better model performance.

5.1.4. R-squared (R^2)

R-squared is a statistical measure that shows how much of the variance in the dependent variable (like the yield of biodiesel) can be accounted for by the independent variables (e.g., input parameters). It has a value between 0 and 1, with higher values suggesting a better model fit and the capacity to explain the variation in the target variable.

5.1.5. Accuracy

Accuracy is a common criterion for evaluating classification tasks in the production of biodiesel. It determines the percentage of correctly identified samples in respect to the overall sample count. Higher accuracy denotes better classification performance [77].

5.1.6. Precision, Recall, and F1-score

To evaluate the effectiveness of machine learning models in binary classification tasks, these metrics are frequently utilized. The F1-score combines accuracy and recall into a single metric, representing the proportion of true positive predictions out of all positive forecasts, recall (sensitivity) measuring the proportion of true positive predictions out of all real positive cases. These indicators are helpful for assessing how well models work when forecasting certain biodiesel quality classes or complying with requirements.

5.1.7. Receiver operating characteristic (ROC) curve and area under the curve (AUC)

The trade-off between true positive rate and false positive rate at various categorization criteria is represented by the ROC curve. The area under the ROC curve, or AUC, serves as a general indicator of the discrimination capacity of a model. Better categorization performance is indicated by a higher AUC.

These evaluation metrics offer quantitative ways to rate the efficiency of machine learning models used in the creation of biodiesel. Researchers and practitioners can choose the relevant metrics to assess and contrast various models, assisting in model choice, optimization, and decision-making in the biodiesel manufacturing process.

5.2. Cross-validation techniques for model validation

The cross-validation techniques are frequently employed for model validation in the manufacturing of diesel in order to evaluate the effectiveness and generalizability of machine learning models. Here are a few cross-validation methods that are frequently used in diesel production.

5.2.1. Holdout validation

The dataset is divided into a training set and a validation set for holdout validation. The model is tested on the validation set after being evaluated on the training set. This technique gives a rapid and easy evaluation of the model's performance, but it may be sensitive to the particular instances included in each batch.

5.2.2. K-Fold cross-validation

The dataset is divided into K folds of equal size using K-fold cross-validation. The validation set is utilized once for each fold, while the remaining folds are used for training. The model is trained and tested K times [52]. To assess total performance, the performance values from each fold are averaged. K-fold cross-validation strengthens evaluation by removing reliance on a particular data split.

5.2.3. Stratified K-Fold cross-validation

For classification problems in diesel manufacturing, stratified K-fold cross-validation is highly helpful, especially when working with unbalanced datasets. As a result, each fold will have a comparable number of samples from each class and the class distribution will be kept. The accuracy of performance estimations for various classes is improved by this method.

5.2.4. Leave-One-Out cross-validation (LOOCV)

In LOOCV, the model is trained on all samples except one, and its performance is then assessed on the sample that was left out [55]. Each sample in the dataset is subjected to this method once more. Despite the fact that LOOCV offers an unbiased measure of model performance, it can be computationally expensive, especially for bigger datasets.

5.2.5. Time Series cross-validation

Time series cross-validation is applied to diesel production data with a temporal component [78]. When dividing the data into training and validation sets, this method takes into account the temporal order of the data. The model is evaluated on future data after being trained on previous data, simulating the real-world situation where predictions are made about unknowable future events.

By revealing information about how well machine learning models generalize to new data, these cross-validation techniques aid in evaluating the performance of those models in the diesel manufacturing process [77]. These strategies enable the detection of potential problems like overfitting or underfitting by testing models on various subsets of the dataset. The characteristics of the diesel production dataset and the particular needs of the modeling task determine the cross-validation technique to be used [77].

5.3. Case studies in biodiesel production

The following processes are involved in making biodiesel from renewable sources: oil extraction, pretreatment of the feedstock, transesterification reaction, product separation, recovery of unreacted alcohol, neutralization of glycerin, washing, and biodiesel purification [17,24]. The five key stages of biodiesel production—soil, feedstock, production, consumption, and emissions—were attempted to be categorized and reviewed in this section [66,78].

All five steps can benefit from ML technology to improve the accuracy of estimates. This study primarily focuses on the first three steps because there are many reviews of research on the use of machine learning technology in modeling biodiesel-fueled engines and combustion techniques.

5.3.1. ML Applications in soil stage

Numerous studies on the plot and tree cases applying ML have been reported in the soil stage of the biofuels' life cycle. The most common ML methods in the soil stage are Random Forest (RF), Gaussian Process Model (GPM), and Support Vector Machines (SVM).

The sorghum crop is useful for creating healthy biofuels, seeds, and feed from aboveground biomass [79]. In order to predict future trends in sorghum bicolor yield under four different greenhouse gas (GHG) emission scenarios and two different watering regimes, Huntington *et al.* [80] employed the RF approach. The most accurate predictors of sorghum productivity were vapor pressure deficit, duration, and irrigation methods. The RF model was able to create accurate predictions by precisely training and classifying data samples according to year and country. Habyarimana *et al.* [79] used a variety of machine learning (ML) techniques to predict the yield of sorghum biomass in a study based on satellite images of sorghum fields. These techniques included radial basis kernel (SVM-R), nonlinear kernel (SVM-G), PCA discriminant analysis (PCA-DA), PLS discriminant analysis (PLS-DA), and SVM with linear classifier.

Tsai *et al.* [81] examined the Linear Mixed-effects Regression (LME), Cubist, Support Vector Regression (SVR), and Random Forest (RF) approaches to estimate biomass in a moderately dense forest with 40 to 60 percent canopy closure, where SVR delivered the most accurate

biomass model. In a four-scenario context emissions-based study, Lee *et al.* [82] estimated the environmental impacts of corn production from 2022 to 2100 using the boosted regression tree (BRT) model. The BRT model had a correlation coefficient of 0.82 for estimating eutrophication impacts and 0.78 for global warming. To get more precise estimations, [83] used the Gaussian Process Model (GPM), a Bayesian inference technique, in a two-stage machine learning process. GPM crops yield downscaling came first, followed by an RF model's predicted yield. Typically, input parameters include soil properties, solar radiation, average precipitation, wind speed, and temperature, while output parameters include biomass yield and future life cycle environmental impact. To comprehend the efficient strategy used in each investigation, Table 1 summarizes soil phase studies.

Table 1: Various ML applications in the soil phase of biodiesel production outline.

Reference	Applied model	Field	Results
[79]	GBL, GBD, GBT, ANN, RF, SVR, SVM, SVM-P, SVM-R, SVM-G, PCA-DA, PLS-DA	Predict sorghum crop yield	GBT
[80]	RF	Predict sorghum crop yield	RF
[81]	LME, SVR, RF	Predict biomass yield in forest	SVR
[82]	BRT	Estimate corn production environmental impacts	BRT
[83]	GPM, RF	Land productivity	GPM

5.3.2. ML Applications in Feedstock

The most widely used techniques for machine learning in the feedstock phase investigations include ANN, multiple linear regression, statistical regression, and multiple nonlinear regression models. The usual input parameters include blend composition, temperature, mixing duration, and speed. The usual output parameters include viscosity, flash point, oxidation stability, density, methane fraction, higher heating values, and cetane number. Mairizal *et al.* [84] looked at biodiesels produced from a variety of sources, including hydrogenated coconut oil, hydrogenated copra oil, beef tallow, walnut oil, sunflower oil, peanut oil, rapeseed oil, and rapeseed oil, to predict higher heating value, viscosity, flashpoint, oxidative stability, and density using multiple linear regressions. The results revealed that by including PU/MU (mono- and polyunsaturated fatty acids balance) as an independent parameter, prediction performance improves. The content of polyunsaturated fatty acids in the feedstock, as well as its iodine and saponification values, were model inputs. Another study used the ANN approach to calculate the cetane number, density, kinematic viscosity, and flashpoint of several biodiesels made from fatty acids [85]. The results for cetane number (1.637 percent; 96.6 percent), flash point (0.997 percent; 99.07 percent), kinematic viscosity (1.638 percent; 95.80 percent), and density (1.638 percent; 95.6 percent) represent the average absolute deviation and the estimation accuracy of the model, respectively (0.101percent; 99.40 percent).

5.3.3. ML applications in production

Tchameni *et al.* [86] predicted the rheological parameters of waste vegetable oil using multiple ANN and nonlinear regression (MNLR). Results showed that the ANN model performed better than the MNLR technique. A significant association between the potentials of methane biomass and chemical composition was found when estimating methane yield in biomass structural components using single linear regressions and multiple linear regressions. To identify the effective method ₃and study aims, Table 2 summarizes feedstock phase investigations [85-86].

Table 2: Various ML applications in the feedstock phase of biodiesel production outline.

Reference	Applied model	Aim
[84]	Multiple linear regressions	To predict HHV, viscosity, FP, oxidative stability, density
[85]	ANN	To estimate C, density, kinematic viscosity, FP
[86]	ANN, MNLR, single and multiple linear regressions	To estimate oil rheological properties

Quality prediction

The most popular machine learning technique for predicting quality is an algorithm called Annotated Neural Networks (ANN), which was developed by regression models and uses input variables such as reaction temperature, reaction time, calcination temperature, pressure, and flow rate and output variables such as FAME (fatty acid methyl ester) content, viscosity, composition, quantity, cetane number, and density.

When distilling Palm Fatty Acids (PFADs) into esters, Soltani *et al.* [88] used a sulfonated mesoporous zinc oxide SO_3HZnO catalyst. They used an Artificial Neural Network (ANN) to model the effects of different reaction parameter effects, such as calcination temperature, metal ratio, reaction time, and reaction temperature. The evaluated ideal parameters for forecasting a 56.41 nm SOH-ZnO nanocrystalline catalyst size were 160°C reaction temperature, 700°C calcine temperature, and 0.004 mole of Zn concentration over 18 min of reaction time. The most and least efficient parameters are generally acknowledged to be zinc concentration and reaction time.

In order to predict quantity, quality, flow rate, the cetane number of fatty acid methyl esters (FAME), and composition in the process of producing vegetable oil-based biodiesel, Ahmad *et al.* [89] combined an ensemble learning technique called Least Squares Boosting (LSBoost) with the Polynomial Chaos Expansion method (PCE). Using Mean Absolute Deviation Percent (MADP), predicted values exhibited 1 percent uncertainty in all process parameters, demonstrating the proposed model's excellent accuracy in outcome prediction and quantification of uncertainty effect in the process. The PCA approach was used to calculate the relative density, viscosity, and percentage of vegetable oil conversion to methyl esters during the biodiesel synthesis process from vegetable oil. PCA is a useful tool for differentiating and separating pure biodiesel, pure diesel, and waste oil.

5.4. Pure diesel, waste oil, and biodiesel

The production of biodiesel from sesame oil using barium hydroxide as a basic catalyst was predicted by Bharadwaj *et al.* [90] using an Artificial Neural Network (ANN) and response surface methodology (RSM) based on a central composite design (CCD). The best feasible set of values for the optimum conditions were the methanol-to-oil molar ratio (6.69:1), the reaction time (40.30 min), the catalyst concentration (1.79 wt%), and the temperature (31.92°C), which produced a 98.6 percent FAME content. According to the study, the key factor affecting the amount of FAME in the finished product is the catalyst concentration. ANN has a better correlation coefficient, root mean square error (R^2), and standard error of prediction, which improve its capacity to forecast the FAME content. According to the study, the key factor affecting the amount of FAME in the finished product is the catalyst concentration. In comparison to RSM, ANN performs better in terms of correlation coefficient, root mean square error (R^2), standard error of prediction (SEP), and relative percent deviation (RPD).

5.4.1. Yield estimation

Numerous studies focused on the use of ML techniques to forecast the production of biodiesel from oils that are not edible, such as anaerobic sludge, castor oil, and jatropha algae [91].

Using jatropha-algae oil blends as inputs, Kumar *et al*. [91] trained an ANN model with the Levenberg-Marquardt (LM) algorithm and backpropagation learning method to predict biodiesel yield in the transesterification process. The effectiveness of the ANN technique was demonstrated by the R-square value of 0.9976 when compared to the experimental findings.

Gandhi and Gogate as well as Álvarez-Mateos. Castor oil and methanol transesterification with H_2SO_4 acid catalyst employed the ANN and CCD model to forecast the percent fatty acid methyl ester content [92-93]. Using the experimental and computational data, they also developed a kinetic model. The rate constants of a kinetic model have also been computed using experimental outputs and anticipated data generated by ANNs. The input parameters are the temperature, catalyst concentration, and methanol-to-oil molar ratio. A percent fatty acid methyl ester yield with an 8 percent variation was predicted by the ANN model.

Chollom *et al.* [94] modelled and estimated the anaerobe thermophilic upflow sludge blanket digester biodiesel and biogas production rate using the ANN method and multilayer neural networks topology. In both constant and abnormal situations, trained and tested experimental data were assessed; a high correlation coefficient indicated ANN optimistic findings for online monitoring of the thermophilic reactors. In a trial using a jatropha-algae oil combination, ANN outperformed RSM [95]. RSM and ANN were used to model a biodiesel synthesis process from waste goat tallow that contained notable free fatty acids (FFAs) in order to determine the ideal parametric parameters that led to the greatest FA conversion. Response surface methodology (RSM) and ANN demonstrated comparable predictive performance under ideal circumstances [96].

In a different study, a Levenberg-Marquardt learning algorithm-based ANN model and a linear regression (LR) model were created to estimate the yield of biodiesel made from soybean oil. The ANN outperformed the LR model [97]. To forecast biodiesel yield, a variety of circumstances for the transesterification of soybean oil to biodiesel have been explored [98]. In this study, a multilayer feedforward neural network and kinetic models are used using artificial neural networks. The outcomes demonstrated that the ANN model was superior than kinetic modeling in terms of accuracy and clarity. Talaghat *et al.* [26] estimated the biodiesel production yield as a function of methanol/oil ratio, pressure, reaction time, and temperature in the noncatalytic supercritical methanol (SCM) method using an adaptive neurofuzzy interference system (ANFIS) method based on a statistical learning theory. The ANFIS model's influence on forecasting biodiesel yield can be shown by the strong R-squared values of the data. In order to forecast and simulate the effectiveness of these approaches in estimating the transesterification yield, Mostafa *et al.* [99] contrasted the Adaptive Neurofuzzy Inference System (ANFIS) and response surface methodology (RSM). The effect of independent variables on the conversion of fatty acid methyl esters was examined using the Box-Behnken design of RSM and two ANFIS approaches (hybrid and backpropagation optimization methods) (FAME). The significant R^2 value for the RSM was 0.9669, while it was 0.9812 and 0.9808 for the two ANFIS models, suggesting the superiority of the ANFIS models over the RSM model for modeling and optimizing. Artificial neural network (ANN) and response surface methodology (RSM) efficiency were evaluated by Meran *et al.* [100] to predict and simulate muskmelon oil-based biodiesel yield. Research conducted by the Central Composite Rotatable Design (CCRD) compared the ANN model to the RSM model. FAME conversion via Multilayer Perceptron (MLP) neural network and RSM is influenced by the catalyst concentration, reaction time, reaction temperature, and methanol-to-oil molar ratio. The R^2 value for the RSM was 0.869, while the ANN model's value was 0.991, demonstrating the ANN model's superiority over the RSM in terms of modeling and optimizing FAME production.

5.4.2. Quality and yield estimation

Numerous studies have focused on biodiesel quality and yield optimization. Bobadilla *et al.* [101] used a set of Support Vector Machines (based on radial basic function kernel, linear kernel, and polynomial kernel) and linear regression methods to predict and improve biodiesel yield of particular properties like turbidity, higher heating value (HHV) with decreased viscosity, and density. Appling genetic algorithms to the regression models obtained more accurate biodiesel optimization scenarios to identify the best combination of independent and dependent variables.

Pustokhina *et al.* [102] developed a GA-ESIM method which is the combination of Evolutionary Support Vector Machine Inference Model (ESIM) and K-means Chaotic Genetic Algorithm (KCGA) to predict precisely and optimize biodiesel mixture properties. They found GA-ESVM

better than ANN-GA and SVM. Obtained results demonstrate that the GA-ESIM model performance in prediction is more accurate than other AI-based tools.

Muna *et al.* [103] used ANN-GA-based and RSM models to predict and optimize the biodiesel yield in *Simarouba glauca* transesterification. They used a gas chromatography-mass spectroscopic (GC-MS) analysis oil to observe Free Fatty Acid (FFA) level, and alcohol ratio, reaction time, and reaction temperature were input variables.

Musa *et al.* [104] focused on an RSM optimization tool alongside the ANFIS model to predict and optimize the biodiesel yield in the *Thevetia peruviana* seed oil transesterification process. In addition to ANFIS and RSM model, using GA resulted in higher *Thevetia Peruviana* Methyl Esters yield (TPME) in less time. The results determined the priority of ANFIS prediction capability over the RSM model. Kumar. [74] applied ANN and GA combination in polanga oil-based biodiesel production to predict and optimize reaction variables to maximize the transesterification process. The input variables are the ethanol-to-oil molar ratio, the reaction temperature, the catalyst concentration, the reaction time, and the stirring speed. Outputs were combined with GA to optimize reaction conditions resulting in 92% by weight biodiesel yield.

5.4.3. Estimation and optimization of process conditions and efficiency

Pauline *et al.* [105] used a multiobjective analysis in order to assess the FAME concentration and exergetic efficiency in waste cooking oil transesterification (WCO) for biodiesel synthesis. To obtain 95.7 percent projected FAME content, the following factors have been optimized: water concentration, reaction duration, immobile lipase, and methanol concentration. The 35 percent catalyst concentration, 12 percent water content, 6.7 molar ratio of methanol to WCO, 20-hour production time, 86 percent FAME content, and 80.1% energy efficiency are the corresponding input variables.

In order to simulate and compare the esterification and transesterification processes of palm waste cooking oil, Kolakoti *et al.* [4] employed nondominated sorting GA-II (NSGA-II) multiobjective optimization. They also optimized heat duty, profit, and organic waste. The profit improved as the heat duty grew, which led to an increase in organic waste. Rouchi *et al.* [106] processed the analysis and directed the reaction parameters in the desired direction using a Multivariate Curve Resolution Alternative Least Square (MCR-ALS). To produce biodiesel from the soybean process, the Multiple Scatter Correction preprocessing technique and MCR-ALS analyze concentrations, the kind of component, and spectra. Kolakoti *et al.* [4] optimized heat duty, profit, and organic waste while simulating and comparing palm waste cooking oil esterification and transesterification reactions using Nondominated Sorting GA-II (NSGA-II) multiobjective optimization. The profit rose as the heat duty rose, which raised the volume of organic waste. In order to process the study and steer the reaction parameters in the desired direction, Rouchi *et al.* [106] applied a Multivariate Curve Resolution Alternative Least Square (MCR-ALS). In order to produce biodiesel from the soybean process, multiple scatter correction preprocessing approach and MCR-ALS assess concentrations, the kind of component, and spectra.

In order to optimize operating conditions as a function of inputs, Aghbashlo *et al.* [107] created an ANFIS model integrated with linear interdependent fuzzy multiobjective (ALIFMO) techniques and nondominated sorting genetic algorithm (NSGA-II). Reaction temperature, methanol/oil molar ratio, and residence time were the input parameters. For the best Conversion Efficiency (CE), which is more than 96.5 percent biodiesel content, optimization decreased Normalized Exergy Destruction (NED) and maximized functional exergy efficiency (FEE) and universal exergy efficiency (UEE) output parameters. With an R^2 of 1, applied ANFIS models accurately predicted the FEE, UEE, NED, and CE parameters.

In order to optimize the production of biodiesel, Aung *et al.* [108] examined the analysis sensitivity, predictability and generalizability, and parametric effects of ANN and RSM. At the optimal temperature, ethanol-to-oil molar ratio, beginning CO_2 pressure, reaction time, and temperature—where the temperature was the most effective-97.42 percent of the Fatty Acid Ethyl Ester (FAEE) content were attained. When it came to predicting the FAEE content of mahua oil and data fitting, ANN model outperformed RSM.

By improving crucial process parameters in the synthesis of biodiesel from vegetable oil, Kumar *et al.* [48] were able to maximize the purification of crucial compounds while lowering energy needs. The process model requires inputs for the following variables: dryer temperature, flux ratio, water mass flow rate, water temperature, flash temperature, number of trays, and water temperature. Among all other designs, the one that proves the lowest specific energy consumption and meets the required biodiesel quality parameters was identified. All of the current two-phase equilibrium between liquids in the biodiesel production system, including those between glycerol, low molecular weight alcohols, water, fatty acids, and biodiesel, were predicted and validated by Noriega *et al.* [109] using Group Interaction Parameters (GIP). According to the findings, the amount of carbon, the presence of hydroxyl groups, and unsaturated bonds all have an impact on liquid-liquid equilibrium. The overall mass fraction of the distributed components was also found to be the most important parameter, followed by the length of the alcohol chain.

Hassan *et al.* [110] used virtual sensors and an extended Kalman filter (EKF) to assess and estimate operating conditions variables, regulate performance, and track the response. Performance study was done to assess the jatropha oil-based biodiesel utilizing alcohol, triglycerides (TG), methyl ester, diglycerides (DG), glycerol (GL), and monoglycerides (MG) concentrations because there are only a few quantifiable factors, such as PH and temperature. Talaghat *et al.* [26] developed an ANN superstructure model to pinpoint the best biodiesel production site and operating conditions. The ANN model served as a workable replacement for the thermodynamics, unit operation, and mixing models by offering a less intricate illustration of the synthesis procedure. As previously mentioned, the SO_3HZnO catalyst was used by Soltani *et al.* [88] to apply ANN to model various reaction parameter effects. The 18-minute reaction time, the reaction temperature of 160°C, the calcine temperature of 700°C, and the Zn concentration of 0.004 moles were determined to be the ideal conditions. The zinc content and reaction time were the variables that worked best and worst.

6. Conclusions

The most popular ML techniques in the soil stage, according to the machine learning applications in this study, are Random Forest, Gaussian Process Model, and Support Vector Machines. The most widely used techniques in feedstock phase investigations include ANN, multiple linear regression, statistical regression, and multiple nonlinear regression models. The usual input parameters include blend composition, temperature, mixing duration, and speed. The usual output parameters include viscosity, flash point, oxidation stability, density, methane fraction, higher heating values, and cetane number. The most popular machine learning technique for predicting quality is an ANN developed by a regression model, which uses input variables such as reaction temperature, reaction time, calcination temperature, pressure, and flow rate and output variables such as FAME content, viscosity, composition, quantity, cetane number, and density. The most popular ML method for yield estimation is ANN combined with ANFIS, using the following parameters: temperature, reaction time, catalyst concentration, total volatile fatty acid of the effluent, and methanol-to-oil molar ratio. Additionally, regular output variables include percent FAME yield, biogas production rate estimation, yield, and production of biodiesel. ANN, along with GA-based ANFIS and SVM, is the most popular machine learning technique for maximizing yield and quality section. The most often utilized input parameters are the methanol-to-oil molar ratio, stirring rate, catalyst concentration, reaction time, and reaction temperature. The output variables that are most frequently used include FAME yield, biodiesel yield, high heating value density, and oil's ultimate acid value. In the process efficiency and optimization section, ANN and ANFIS are the most important ML techniques. The following are frequently used input variables: reaction time, concentration, water content, methanol-to-oil molar division, and temperature. The following are commonly used output variables: CE, Universal Exergy Efficiency (UEE), FAME content, biodiesel yield, and functional exergy efficiency. ANN, ANFIS, ELM, and SVM Machine Learning methods were employed to study consumption, engine performance, and emission.

References

[1] Pan H, Li H, Zhang H, Wang A, Yang S. Functional nanomaterials-catalyzed production of bio-diesel. Curr. Nanosci., 2020; 16(3): 376–91. http://dx.doi.org/10.2174/1573413715666190411142820

[2] Singh D, Sharma D, Soni SL, Inda CS, Sharma S, Sharma PK, et al. A comprehensive review of biodiesel production from waste cooking oil and its use as fuel in compression ignition engines: 3rd generation cleaner feedstock. J. Clean Prod., 2021; 307(127299): 127299. http://dx.doi.org/10.1016/j.jclepro.2021.127299

[3] Noreen S, Sahar I, Masood N, Iqbal M, Zahid M, Nisar J, et al. Thermodynamic and kinetic approach of biodiesel production from waste cooking oil using nano-catalysts. Z. Phys. Chem., 2021; 235(12): 1673–88. http://dx.doi.org/10.1515/zpch-2020-1644

[4] Kolakoti A, Setiyo M, Waluyo B. Biodiesel production from waste cooking oil: Characterization, modeling and optimization. Mech. Eng. for Soc. and Ind., 2021; 1(1): 22–30. http://dx.doi.org/10.31603/mesi.5320

[5] Sronsri C, Sittipol W, U-yen K. Performance of CaO catalyst prepared from magnetic-derived $CaCO_3$ for biodiesel production. Fuel, 2021; 304(121419): 121419. http://dx.doi.org/10.1016/j.fuel.2021.121419

[6] Ranjith, Velmurugan V, Thanikaikarasan S. Prediction of Exhaust Gas Emission characteristics using Neem oil blended bio- diesel in diesel engine. Mater. Today, 2020;21:870–5. http://dx.doi.org/10.1016/j.matpr.2019.07.706

[7] Soly Peter A, Alias MP, Iype MP, Jolly J, Sankar V, Joseph Babu K, et al. Optimization of biodiesel production by transesterification of palm oil and evaluation of biodiesel quality. Mater. Today, . 2021; 42: 1002–7. http://dx.doi.org/10.1016/j.matpr.2020.11.995

[8] Ogunkunle O, Ahmed NA. Overview of biodiesel combustion in mitigating the adverse impacts of engine emissions on the sustainable human–environment scenario. Sustainability, 2021; 13(10): 5465. http://dx.doi.org/10.3390/su13105465

[9] Zulqarnain, Ayoub M, Yusoff MHM, Nazir MH, Zahid I, Ameen M, et al. A comprehensive review on oil extraction and biodiesel production technologies. Sustainability, 2021; 13(2): 788. http://dx.doi.org/10.3390/su13020788

[10] Embong NH, Hindryawati N, Bhuyar P, Govindan N, Rahim MHA, Maniam GP. Enhanced biodiesel production via esterification of palm fatty acid distillate (PFAD) using rice husk ash $(NiSO_4)/SiO_2$ catalyst. Appl. Nanosci., 2023; 13: 2241-2249.http://dx.doi.org/10.1007/s13204-021-01922-4

[11] Asif M, Bashir S, Khan S. Impact of non-renewable and renewable energy consumption on economic growth: evidence from income and regional groups of countries. Environ. Sci. Pollut. Res. Int., 2021; 28(29): 38764–73. http://dx.doi.org/10.1007/s11356-021-13448-x

[12] Millward-Hopkins J, Steinberger JK, Rao ND, Oswald Y. Providing decent living with minimum energy: A global scenario. Glob. Environ. Change, 2020; 65(102168): 102168. http://dx.doi.org/10.1016/j.gloenvcha.2020.102168

[13] Roy B, Schaffartzik A. Talk renewables, walk coal: The paradox of India's energy transition. Ecol. Econ., 2021; 180(106871): 106871. http://dx.doi.org/10.1016/j.ecolecon.2020.106871

[14] Yavuz V, Altiner A, Bozkurt E. The effects of globalization on energy consumption: Evidence from EAGLEs. Bus. Manag. Stud. Int. J., 2022; 10(3) :885–94. http://dx.doi.org/10.15295/bmij.v10i3.2081

[15] Kurczyński D, Wcisło G, Leśniak A, Kozak M, Łagowski P. Production and testing of butyl and methyl esters as new generation biodiesels from fatty wastes of the leather industry. Energies, 2022; 15(22):8744. http://dx.doi.org/10.3390/en15228744

[16] Wang X, Lu H, He Y, Sun K, Feng T, Zhu X. Listening to 15 Hz binaural beats enhances the connectivity of functional brain networks in the mental fatigue state—an EEG study. Brain Sci., 2022; 12(9): 1161. http://dx.doi.org/10.3390/brainsci12091161

[17] Aghbashlo M, Peng W, Tabatabaei M, Kalogirou SA, Soltanian S, Hosseinzadeh-Bandbafha H, Mahian O, Lam SS. Machine learning technology in biodiesel research: A review. Prog. Energy Combust. Sci., 2021; 85: 100904. http://dx.doi.org/10.1016/j.pecs.2021.100904

[18] Cui Z, Huang S, Wang M, Nie K, Fang Y, Tan T. Improving the CFPP property of biodiesel via composition design: An intelligent raw material selection strategy based on different machine learning algorithms. Renew. Energy, 2021; 170: 354–63. http://dx.doi.org/10.1016/j.renene.2021.02.008

[19] Gupta KK, Kalita K, Ghadai RK, Ramachandran M, Gao X-Z. Machine learning-based predictive modelling of biodiesel production—A comparative perspective. Energies, 2021; 14(4): 1122. http://dx.doi.org/10.3390/en14041122

[20] Mijwil M, Salman Shukur B. A scoping review of machine learning techniques and their utilisation in predicting heart diseases. Ibn AL- Haitham J. Pure Appl. Sci., 2022; 35(3): 175–89. http://dx.doi.org/10.30526/35.3.2813

[21] Carbonell P, Le Feuvre R, Takano E, Scrutton NS. *In silico* design and automated learning to boost next-generation smart biomanufacturing. Synth. Biol., 2020; 5(1): ysaa020. http://dx.doi.org/10.1093/synbio/ysaa020

[22] Liu Y, Hong W, Cao B. Machine learning for predicting thermodynamic properties of pure fluids and their mixtures. Energy, 2019; 188: 116091. http://dx.doi.org/10.1016/j.energy.2019.116091

[23] Garg A, Jain S. Process parameter optimization of biodiesel production from algal oil by response surface methodology and artificial neural networks. Fuel, 2020; 277(118254): 118254. http://dx.doi.org/10.1016/j.fuel.2020.118254

[24] Moayedi H, Aghel B, Foong LK, Bui DT. Feature validity during machine learning paradigms for predicting biodiesel purity. Fuel, 2020; 262: 116498. http://dx.doi.org/10.1016/j.fuel.2019.116498

[25] Ong HC, Milano J, Silitonga AS, Hassan MH, Shamsuddin AH, Wang C-T, et al. Biodiesel production from Calophyllum inophyllum-Ceiba pentandra oil mixture: Optimization and characterization. J. Clean Prod., 2019; 219: 183–98. http://dx.doi.org/10.1016/j.jclepro.2019.02.048

[26] Reza Talaghat M, Mokhtari S, Saadat M. Modeling and optimization of biodiesel production from microalgae in a batch reactor. Fuel, 2020; 280: 118578. http://dx.doi.org/10.1016/j.fuel.2020.118578

[27] Ritchie H, Roser M, Rosado P. Energy. Our World in Data. 2022 [cited 2023 Jul 21]; https://ourworldindata.org/energy-mix

[28] Jung S, Kim M, Lin K-YA, Park Y-K, Kwon EE. Biodiesel synthesis from bio-heavy oil through thermally induced transesterification. J. Clean Prod., 2021; 294: 126347. http://dx.doi.org/10.1016/j.jclepro.2021.126347

[29] Encinar JM, Nogales S, González JF. Biodiesel and biolubricant production from different vegetable oils through transesterification. Eng Rep, 2020; 2(12): e12190. http://dx.doi.org/10.1002/eng2.12190

[30] Al Basir F, Roy PK. Study on enzyme inhibition in biodiesel synthesis: effect of stepwise addition of methanol and removal of glycerol. Energy Ecol. Environ., 2019; 4(2): 75–84. http://dx.doi.org/10.1007/s40974-019-00112-x

[31] Cardeño F, Lapuerta M, Rios L, Agudelo JR. Reconsideration of regulated contamination limits to improve filterability of biodiesel and blends with diesel fuels. Renew. Energy, 2020; 159: 1243–51. http://dx.doi.org/10.1016/j.renene.2020.06.079

[32] Dinesh K, Tamilvanan A, Vaishnavi S, Gopinath M, Mohan KSR. Biodiesel production using *Calophyllum inophyllum* (Tamanu) seed oil and its compatibility test in a CI engine. Biofuels, 2019; 10(3): 347–53. http://dx.doi.org/10.1080/17597269.2016.1187543

[33] Saravana Sathiya Prabhahar R, Nagaraj P, Jeyasubramanian K. Promotion of bio oil, H_2 gas from the pyrolysis of rice husk assisted with nano silver catalyst and utilization of bio oil blend in CI engine. Int. J. Hydrogen Energy, 2020; 45(33): 16355–71. http://dx.doi.org/10.1016/j.ijhydene.2020.04.123

[34] Almeida FLC, Travália BM, Gonçalves IS, Forte MBS. Biodiesel production by lipase-catalyzed reactions: bibliometric analysis and study of trends. Biofuel. Bioprod. Biorefin., 2021; 15(4): 1141--1159. http://dx.doi.org/10.1002/bbb.2183

[35] Thushari I, Babel S, Samart C. Biodiesel production in an autoclave reactor using waste palm oil and coconut coir husk derived catalyst. Renew. Energy, 2019; 134: 125–34. http://dx.doi.org/10.1016/j.renene.2018.11.030

[36] Ngoie WI, Oyekola OO, Ikhu-Omoregbe D, Welz PJ. Valorisation of edible oil wastewater sludge: Bioethanol and biodiesel production. Waste Biomass Valorization, 2020; 11(6): 2431–40. http://dx.doi.org/10.1007/s12649-019-00633-w

[37] Asaad SM, Inayat A, Jamil F, Ghenai C, Shanableh A. Optimization of biodiesel production from waste cooking oil using a green catalyst prepared from glass waste and animal bones. Energies, 2023; 16(5): 2322. http://dx.doi.org/10.3390/en16052322

[38] Hassan HMA, Alhumaimess MS, Alsohaimi IH, Essawy AA, Hussein MF, Alshammari HM, Aldosari OF. Biogenic-mediated synthesis of the cs_2O–MgO/MPC nanocomposite for biodiesel production from Olive oil. ACS Omega, 2020; 5(43): 27811-22. http://dx.doi.org/10.1021/acsomega.0c02814

[39] Abdul Hakim Shaah M, Hossain MS, Salem Allafi FA, Alsaedi A, Ismail N, Ab Kadir MO, Ahmad MI. A review on non-edible oil as a potential feedstock for biodiesel: physicochemical properties and production technologies. RSC Adv., 2021; 11(40): 25018–37. http://dx.doi.org/10.1039/d1ra04311k

[40] Joven JMO, Gadian JT Jr, Perez MA, Caingles JG, Mansalaynon AP, Ido AL, Arazo RO. Optimized ultrasonic-assisted oil extraction and biodiesel production from the seeds of *Maesopsis eminii*. Ind. Crops Prod., 2020; 155:112772. http://dx.doi.org/10.1016/j.indcrop.2020.112772

[41] Taher H, Giwa A, Abusabiekeh H, Al-Zuhair S. Biodiesel production from *Nannochloropsis gaditana* using supercritical CO2 for lipid extraction and immobilized lipase transesterification: Economic and environmental impact assessments. Fuel Process Technol., 2020; 198:106249. http://dx.doi.org/10.1016/j.fuproc.2019.106249

[42] Smythers AL, Napier EG, Higginbotham EL, Holland AT, Stephenson AA, Kolling DRJ. Direct incorporation of exogenous glycerol leads to increased triacylglycerol formation in *Chlorella vulgaris*. Energy Fuels, 2019; 33(11): 11125–34. http://dx.doi.org/10.1021/acs.energyfuels.9b02653

[43] Firemichael D, Hussen A, Abebe W. Production and characterization of biodiesel and glycerine pellet from macroalgae strain: Cladophora glomerata. Bull. Chem. Soc. Ethiop., 2020; 34(2): 249–58. http://dx.doi.org/10.4314/bcse.v34i2.4

[44] Yin Z, Zhu L, Li S, Hu T, Chu R, Mo F, et al. A comprehensive review on cultivation and harvesting of microalgae for biodiesel production: Environmental pollution control and future directions. Bioresour. Technol., 2020; 301:122804. http://dx.doi.org/10.1016/j.biortech.2020.122804

[45] Kanan M, Habib MS, Shahbaz A, Hussain A, Habib T, Raza H, Raza H, Abusaq Z, Assaf R. A grey-fuzzy programming approach towards Socio-economic optimization of second-generation biodiesel supply chains. Sustainability, 2022; 14(16): 10169. http://dx.doi.org/10.3390/su141610169

[46] Farvardin M, Hosseinzadeh Samani B, Rostami S, Abbaszadeh-Mayvan A, Najafi G, Fayyazi E. Enhancement of biodiesel production from waste cooking oil: ultrasonic- hydrodynamic combined cavitation system. Energy Sources Recovery Util. Environ. Eff., 2022; 44(2): 5065-5079. http://dx.doi.org/10.1080/15567036.2019.1657524

[47] Emaish H, Abualnaja KM, Kandil EE, Abdelsalam NR. Evaluation of the performance and gas emissions of a tractor diesel engine using blended fuel diesel and biodiesel to determine the best loading stages. Sci. Rep., 2021; 11(1). http://dx.doi.org/10.1038/s41598-021-89287-0

[48] Kumar V, Kalita K, Madhu S, Ragavendran U, Gao X-Z. A hybrid genetic programming–gray wolf optimizer approach for process optimization of biodiesel production. Processes, 2021; 9(3): 442. http://dx.doi.org/10.3390/pr9030442

[49] Mohammad Fuad II, Ngah Demon MF, Husni H. Automated real time anomaly detection model for operation and production data at scale. In: Day 1 Mon, November 09, 2020. SPE; 2020.

[50] Basha SM, Rajput DS. Survey on evaluating the performance of machine learning algorithms: Past contributions and future roadmap. In: Deep Learning and Parallel Computing Environment for Bioengineering Systems. Elsevier; 2019. p. 153–64.

[51] Battisti E, Shams SMR, Sakka G, Miglietta N. Big data and risk management in business processes: implications for corporate real estate. Bus. Proc. Management J., 2019; 26(5): 1141–55. http://dx.doi.org/10.1108/bpmj-03-2019-0125

[52] Chuang S-Y, Sahoo N, Lin H-W, Chang Y-H. Predictive maintenance with sensor data analytics on a Raspberry Pi-based experimental platform. Sensors, 2019; 19(18): 3884. http://dx.doi.org/10.3390/s19183884

[53] Silitonga A, Mahlia T, Shamsuddin A, Ong H, Milano J, Kusumo F, Sebayang AH, Dharma S, Ibrahim H, Husin H, Mofijur MAshrafur Rahman, SM. Optimization of Cerbera manghas biodiesel production using artificial neural networks integrated with ant colony optimization. Energies, 2019; 12(20): 3811. http://dx.doi.org/10.3390/en12203811

[54] Datta BN, Sahoo B. Machine learning, regression and optimization. In: Data Science and SDGs. Singapore: Springer Singapore; 2021. p. 177–97.

[55] Awogbemi O, Kallon DVV. Application of tubular reactor technologies for the acceleration of biodiesel production. Bioengineering, 2022; 9(8): 347. http://dx.doi.org/10.3390/bioengineering9080347

[56] Constantino AF, Cubides-Román DC, dos Santos RB, Queiroz LHK Jr, Colnago LA, Neto ÁC, Barbosad LL, Romãoe W, de Castroa EVR, Filgueirasa PR, Lacerda Jr. V. Determination of physicochemical properties of biodiesel and blends using low-field NMR and multivariate calibration. Fuel, 2019; 237: 745–52. http://dx.doi.org/10.1016/j.fuel.2018.10.045

[57] Silitonga Baron D. Machine Learning in Astronomy: a practical overview. arXiv: Instrumentation and Methods for Astrophysics. 2019. https://doi.org/10.48550/arXiv.1904.07248

[58] Pusdiktasari ZF, Fitriani R, Sumarminingsih E. Simulation study using average difference algorithm to analyze the outlierness degree of spatial observations. In: 2020 4th International Conference on Informatics and Computational Sciences (ICICoS). IEEE; 2020.

[59] Ozsahin DU, Taiwo Mustapha M, Mubarak AS, Said Ameen Z, Uzun B. Impact of feature scaling on machine learning models for the diagnosis of diabetes. In: 2022 International Conference on Artificial Intelligence in Everything (AIE). IEEE; 2022.

[60] Wang S, Shen Z, Shen Z, Dong Y, Li Y, Cao Y, Zhang Y, Guo S, Shuai J, Yang Y, Lin C, Chen X, Zhang X, Huang Q. Machine-learning micropattern manufacturing. Nanotoday, 2021; 38: 101152. http://dx.doi.org/10.1016/j.nantod.2021.101152

[61] Oraegbunam JC, Oladipo B, Falowo OA, Betiku E. Clean sandbox (Hura crepitans) oil methyl esters synthesis: A kinetic and thermodynamic study through pH monitoring approach. Renew. Energy, 2020; 160: 526–37. http://dx.doi.org/10.1016/j.renene.2020.06.124

[62] Ayadi I, Belghith H, Gargouri A, Guerfali M. Utilization of wheat bran acid hydrolysate by *Rhodotorula mucilaginosa* Y-MG1 for microbial lipid production as feedstock for biodiesel synthesis. Biomed. Res. Int., 2019; 2019: 1–11. http://dx.doi.org/10.1155/2019/3213521

[63] Anwar M. Potential vs prevalent vs popular vs proven biodiesel feedstocks: A critical 4P selection process. Fuel, 2021; 298: 120712. http://dx.doi.org/10.1016/j.fuel.2021.120712

[64] Polat E, Yüksel E, Altınbaş M. Mutual effect of sodium and magnesium on the cultivation of microalgae Auxenochlorella protothecoides. Biomass Bioenergy, 2020;132:105441. http://dx.doi.org/10.1016/j.biombioe.2019.105441

[65] Zhu Y-H, Hu J, Ge F, Li F, Song J, Zhang Y, Yu D-J. Accurate multistage prediction of protein crystallization propensity using deep-cascade forest with sequence-based features. Brief. Bioinform., 2021; 22(3). http://dx.doi.org/10.1093/bib/bbaa076

[66] Hasnain M, Abideen Z, Naz S, Roessner U, Munir N. Biodiesel production from new algal sources using response surface methodology and microwave application. Biomass Convers. Biorefin., 2023; 13(7): 6213–28. http://dx.doi.org/10.1007/s13399-021-01560-4

[67] Kumar S, Jain S, Kumar H. Performance evaluation of adaptive neuro-fuzzy inference system and response surface methodology in modeling biodiesel synthesis from jatropha–algae oil. Energy Sources Recovery Util. Environ. Eff., 2018; 40(24):3 000–8. http://dx.doi.org/10.1080/15567036.2018.1515277

[68] Yesilyurt MK, Arslan M, Eryilmaz T. Application of response surface methodology for the optimization of biodiesel production from yellow mustard (*Sinapis alba,* L.) seed oil. Int. J. Green Energy, 2019; 16(1): 60–71. http://dx.doi.org/10.1080/15435075.2018.1532431

[69] Daniyan I, Bello E, Ogedengbe T, Mpofu K. Use of central composite design and artificial neural network for predicting the yield of biodiesel. Procedia CIRP., 2020; 89: 59–67. http://dx.doi.org/10.1016/j.procir.2020.05.119

[70] Rocabruno-Valdés CI, González-Rodriguez JG, Díaz-Blanco Y, Juantorena AU, Muñoz-Ledo JA, El-Hamzaoui Y, Hernández JA. Corrosion rate prediction for metals in biodiesel using artificial neural networks. Renew. Energy, 2019; 140: 592–601. http://dx.doi.org/10.1016/j.renene.2019.03.065

[71] Chi NTL, Duc PA, Mathimani T, Pugazhendhi A. Evaluating the potential of green alga Chlorella sp. for high biomass and lipid production in biodiesel viewpoint. Biocatal. Agric. Biotechnol., 2019; 17: 184–8. http://dx.doi.org/10.1016/j.bcab.2018.11.011

[72] Mehrizadeh M, Nikbin-Fashkacheh H, Zand N, Najafi-Marghmaleki A. Accurate prediction of kinematic viscosity of biodiesels and their blends with diesel fuels. J. Am. Oil Chem. Soc., 2020; 97(10): 1083–94. http://dx.doi.org/10.1002/aocs.12421

[73] Xie L, Jia Y, Xiao J, Gu X, Huang J. GMDH-based outlier detection model in classification problems. J. Syst. Sci. Complex., 2020; 33(5): 1516–32. http://dx.doi.org/10.1007/s11424-020-9002-6

[74] Kumar S. Estimation capabilities of biodiesel production from algae oil blend using adaptive neuro-fuzzy inference system (ANFIS). Energy Sources Recovery Util. Environ. Eff., 2020; 42(7): 909–17. http://dx.doi.org/10.1080/15567036.2019.1602203

[75] Cui S, Yin Y, Wang D, Li Z, Wang Y. A stacking-based ensemble learning method for earthquake casualty prediction. Appl. Soft. Comput., 2021;101:107038. http://dx.doi.org/10.1016/j.asoc.2020.107038

[76] Gidon D, Pei X, Bonzanini AD, Graves DB, Mesbah A. Machine learning for real-time diagnostics of cold atmospheric plasma sources. IEEE Trans. Radiat. Plasma. Med. Sci., 2019; 3(5): 597-605. http://dx.doi.org/10.1109/trpms.2019.2910220

[77] Vogado L, Veras R, Aires K, Araújo F, Silva R, Ponti M, Tavares JMRS. Diagnosis of leukaemia in blood slides based on a fine-tuned and highly generalisable deep learning model. Sensors, 2021; 21(9): 2989. http://dx.doi.org/10.3390/s21092989

[78] Hu H, Tang M, Bai C. DATSING: Data augmented time series forecasting with adversarial domain adaptation. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York, NY, USA: ACM; 2020.

[79] 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(4): 203. http://dx.doi.org/10.3390/agronomy9040203

[80] Huntington T, Cui X, Mishra U, Scown CD. Machine learning to predict biomass sorghum yields under future climate scenarios. Biofuel Bioprod. Biorefin., 2020; 14(3): 566–77. http://dx.doi.org/10.1002/bbb.2087

[81] Tsai Y-Z, Hsu K-S, Wu H-Y, Lin S-I, Yu H-L, Huang K-T, Hu M-C, Hsu S-Y. Application of random forest and ICON models combined with weather forecasts to predict soil temperature and water content in a greenhouse. Water, 2020; 12(4): 1176. http://dx.doi.org/10.3390/w12041176

[82] Lee EK, Zhang W-J, Zhang X, Adler PR, Lin S, Feingold BJ, Khwaja HA,. Romeiko XX. Projecting life-cycle environmental impacts of corn production in the U.S. Midwest under future climate scenarios using a machine learning approach. Sci. Total Environ., 2020; 714:136697. http://dx.doi.org/10.1016/j.scitotenv.2020.136697

[83] 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(7): 074013. http://dx.doi.org/10.1088/1748-9326/ab865f

[84] Mairizal AQ, Awad S, Priadi CR, Hartono DM, Moersidik SS, Tazerout M, Andres Y. Experimental study on the effects of feedstock on the properties of biodiesel using multiple linear regressions. Renew Energy, 2020; 145: 375–81. http://dx.doi.org/10.1016/j.renene.2019.06.067

[85] Giwa SO, Adekomaya SO, Adama KO, Mukaila MO. Prediction of selected biodiesel fuel properties using artificial neural network. Front. Energy, 2015; (4): 33–45. http://dx.doi.org/10.1007/s11708-015-0383-5

[86] Tchameni AP, Zhao L, Ribeiro JXF, Li T. Predicting the rheological properties of waste vegetable oil biodieselmodified water-based mud using artificial neural network. Geosystem Engineering. 2019; 22(2): 101–11.https://doi.org/10.1080/12269328.2018.1490209

[87] Liu Z, Baghban A. Application of LSSVM for biodiesel production using supercritical ethanol solvent. Energy Sources Recovery Util. Environ. Eff., 2017; 39(17): 1869–74. http://dx.doi.org/10.1080/15567036.2017.1380732

[88] Soltani S, Rashid U, Roodbar Shojaei T, Nehdi IA, 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(1): 33–47. http://dx.doi.org/10.1080/00986445.2018.1471399

[89] Ahmad I, Ayub A, Ibrahim U, Khattak M, Kano M. Data-based sensing and stochastic analysis of biodiesel production process. Energies, 2018; 12(1): 63. http://dx.doi.org/10.3390/en12010063

[90] Bharadwaj AVSLS, Niju S, Meera Sheriffa Begum KM, Narayanan A. Performance and evaluation of calcined limestone as catalyst in biodiesel production from high viscous nonedible oil. Environ. Prog. Sustain. Energy, 2020; 39(3). http://dx.doi.org/10.1002/ep.13342

[91] 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: Recovery, Utilization, and Environmental Effects, 2019; 41: 1285–95.https://doi.org/10.1080/15567036.2018.1548507

[92] Gandhi SS, Gogate PR. Intensified transesterification of castor oil using ultrasonic horn: response surface methodology (RSM) based optimization. Int. J. Green Energy, 2021; 18(14): 1523–35. http://dx.doi.org/10.1080/15435075.2021.1911808

[93] Álvarez-Mateos P, García-Martín JF, Guerrero-Vacas FJ, Naranjo-Calderón C, Barrios-Sánchez CC, Pérez-Camino MC. Valorization of a high-acidity residual oil generated in the waste cooking oils recycling industries. Grasas Aceites, 2019; 70(4): 335. http://dx.doi.org/10.3989/gya.1179182

[94] Chollom MN, Rathilal S, Swalaha FM, Bakare BF, Tetteh EK. Comparison of response surface methods for the optimization of an upflow anaerobic sludge blanket for the treatment of slaughterhouse wastewater. Environ. Eng. Res., 2019; 25(1): 114–22. http://dx.doi.org/10.4491/eer.2018.366

[95] Agu CM, Menkiti MC, Ekwe EB, Agulanna AC. Modeling and optimization of Terminalia catappa L. kernel oil extraction using response surface methodology and artificial neural network. Artificial Intelligence in Agriculture, 2020; 4: 1–11. http://dx.doi.org/10.1016/j.aiia.2020.01.001

[96] Desai NN, S. Soraganvi V, Madabhavi VK. Solar photocatalytic degradation of organic contaminants in landfill leachate using TiO_2 nanoparticles by RSM and ANN. Nat. Environ. Pollut. Technol. 2020; 19(02): 651–62. http://dx.doi.org/10.46488/nept.2020.v19i02.019

[97] Kumar S. Comparison of linear regression and artificial neural network technique for prediction of a soybean biodiesel yield. Energy Sources Recovery Util. Environ. Eff., 2020; 42(12): 1425–35. http://dx.doi.org/10.1080/15567036.2019.1604858

[98] Cavalheiro LF, Misutsu MY, Rial RC, Viana LH, Oliveira LCS. Characterization of residues and evaluation of the physico chemical properties of soybean biodiesel and biodiesel: Diesel blends in different storage conditions. Renew. Energy, 2020; 151: 454–62. http://dx.doi.org/10.1016/j.renene.2019.11.039

[99] 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: 626–36. http://dx.doi.org/10.1016/j.energy.2016.09.028

[100] Prakash Maran J, Priya B. Modeling of ultrasound assisted intensification of biodiesel production from neem (Azadirachta indica) oil using response surface methodology and artificial neural network. Fuel, 2015; 143: 262–7. http://dx.doi.org/10.1016/j.fuel.2014.11.058

[101] Bobadilla M, Fernandez Mart ınez R, Lorza R, Gomez FS, Vergara Gonz E, Alez. Optimizing biodiesel production from waste cooking oil using genetic algorithm-based support vector machines. Energies. 2018;11(11). https://doi.org/10.3390/en11112995

[102] Pustokhina I, Seraj A, Hafsan H, Mostafavi SM, Alizadeh SM. Developing a robust model based on the Gaussian Process Regression approach to predict biodiesel properties. Int. J. Chem. Eng., 2021; 2021: 1–12. http://dx.doi.org/10.1155/2021/5650499

[103] Muna NL, Mu'alimah AA, Pridiana DB, Widodo AK, Adiyar SR, Elinda EH. The production of biodiesel from waste cooking oil (simultaneous esterification and transesterification using Fe/Zeolite catalysts from waste geothermal). Mater. Sci. Forum, 2020; 990: 161–7. http://dx.doi.org/10.4028/www.scientific.net/msf.990.161

[104] Musa H, Eric D, Hassan A, Usman B, Isa NF, Alhassan AS. Optimized Biodiesel Production from *jatropha curcas* oil. Energy Res. J., 2021; 12(1): 39–44. http://dx.doi.org/10.3844/erjsp.2021.39.44

[105] Mercy Nisha Pauline J, Sivaramakrishnan R, Pugazhendhi A, Anbarasan T, Achary A. Transesterification kinetics of waste cooking oil and its diesel engine performance. Fuel,. 2021; 285:119108. http://dx.doi.org/10.1016/j.fuel.2020.119108

[106] Babaei Rouchi M, Khorrami M, Garmarudi AB, 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 (MCRALS). Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019; 213: 347–53. http://dx.doi.org/10.1016/j.saa.2019.01.046

[107] Aghbashlo M, Hosseinpour S, Tabatabaei M, Mojarab Soufiyan 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: 100–12. http://dx.doi.org/10.1016/j.fuel.2018.07.095

[108] Aung T, Kim S-J, Eun J-B. A hybrid RSM-ANN-GA approach on optimisation of extraction conditions for bioactive component-rich laver (Porphyra dentata) extract. Food Chem., 2022; 366: 130689. http://dx.doi.org/10.1016/j.foodchem.2021.130689

[109] Noriega MA, Narváez PC. UNIFAC correlated parameters for liquid-liquid equilibrium prediction of ternary systems related to biodiesel production process. Fuel, 2019; 249: 365–78. http://dx.doi.org/10.1016/j.fuel.2019.03.124

[110] Hassan AEO, Mohammed TAA, Demirkol A. Fault detection in a three-tank hydraulic system using unknown input observer and extended Kalman filter. J. Eng. Res., 2021; 9(4A). http://dx.doi.org/10.36909/jer.8985

[111] Atadashi IM, Aroua MK, Aziz AA. Biodiesel separation and purification: A review. Renew. Energy, 2011; 36(2): 437–43. http://dx.doi.org/10.1016/j.renene.2010.07.019

[112] Mathew GM, Raina D, Narisetty V, Kumar V, Saran S, Pugazhendi A, Sindhu R, Pandey A, Binod P. Recent advances in biodiesel production: Challenges and solutions. Sci. Total Environ., 2021; 794:148751.http://dx.doi.org/10.1016/j.scitotenv.2021.148751

[113] van Gerpen J, Shanks B, Pruszko R, Knothe G. Biodiesel Production Technology. National Renewable Energy Laboratory 1617 Cole Boulevard, NREL/SR-510-36244.

To whom correspondence should be addressed: Joshua Ajao, Department of Chemical Engineering, Ladoke Akintola University of Technology, Ogbomoso, Nigeria, E-mail: ajaojoshua123@gmail.com