Korean J. Chem. Eng., 40(12), 2941-2956 (2023)
DOI: 10.1007/s11814-023-1489-9

pISSN: 0256-1115
eISSN: 1975-7220

Application of machine learning and genetic algorithms to the prediction and optimization of biodiesel yield from waste cooking oil

Aqueel Ahmad*,†, Ashok Kumar Yadav, and Achhaibar Singh***

*Mechanical Engineering Department, Netaji Subhas University of Technology, New Delhi, India
**Department of Mechanical Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad, India
(*Received 7 February 2023 • Revised 19 April 2023 • Accepted 3 May 2023*)

Abstract−The synthesis and usage of biodiesel have become the focus of extensive research due to the ever-increasing emphasis on the development of sustainable and renewable sources. As biodiesel yield and quality depend on the feedstocks used in transesterification, numerous process variables must be controlled at optimal levels to ensure high productivity throughout the biodiesel synthesis process. This study provides three machine learning-based approaches (Gradient boosting, eXtreme gradient boosting (XGB), and light gradient boosting machine (LGBM) regression) for prediction and genetic algorithm (GA) for the optimization of biodiesel yield from waste cooking oil. Throughout the modeling, training, testing, and cross-validation processes, supervised machine learning methods were used to analyze the datasets exhaustively. Four performance indicators (MAE, RMSE, MAPE, and R^2) were used to evaluate and contrast the precision of the algorithms. Based on the findings of the coefficient of determination, the LGBM algorithm has the highest prediction accuracy of 0.94048, followed by XGB with 0.8631 accuracy and gradient boosting with 0.8547accuracy. The optimal condition for producing biodiesel through GA was a molar ratio of 7.62 : 1 with a catalyst concentration of 0.5 wt%, a reaction temperature of 65 °C, and a reaction time of 105 min and the corresponding value of the biodiesel yield was 98.98% (by wt.). Experiments confirmed this prediction, and with acceptable error, all results are close to the model's predicted values. The LGBM coupled with GA can be used as a strategic decision-support and optimization tool for the production of high-quality biodiesel.

Keywords: Biodiesel, Machine Learning, Gradient Boosting Regression, XGBoost Regression, LightGBM Regression, Genetic Algorithm

INTRODUCTION

Renewable energy has garnered significant attention due to the depletion of global petroleum resources, increased energy consumption, growing environmental pollution, and rising oil prices [1,2]. While researchers are exploring ways to produce renewable energy sources, such as solar, wind, tidal, ocean, and hydro, none of these sources currently fulfill the criteria for substituting traditional fossil fuels. Therefore, it is essential to explore viable alternative fuels to meet the current energy shortage. Biofuels, including biodiesel, bioethanol, and biogas, are gaining traction as feasible alternatives to meet impending energy needs. Biodiesel, composed of alkyl esters generated from vegetable and animal fat, has been considered a practical substitute for petroleum fuels due to its similarities [3,4]. The use of biodiesel results in significant reductions in emitted particles, hydrocarbons, and carbon monoxide compared to petroleum-based diesel [5]. However, the commercial success of biodiesel has been limited in several countries, notably India, due to a shortage of adequate feedstock. Biodiesel can be produced from a variety of plant-based oils and fats, including sunflower, palm oil, soybean, waste vegetable oil, animal fats, and algae, using minimally damaging processing techniques [6]. One of the most commonly used methods for producing biodiesel is through transesterification, in which triglyceride molecules undergo a reaction with methanol via homogeneous or heterogeneous catalysts to form methyl esters [7].

Biodiesel yield can be significantly influenced by various process constraints, such as molar ratio, catalyst, reaction time, and temperature, which must be optimized to achieve better yields [8,9]. In this regard, AI techniques have been widely employed in the energy industry for prediction, modeling, and optimization purposes. Machine learning, deep learning, genetic algorithms, PSO, RSM, ANOVA, fuzzy logic, ANN, ANFIS, and logistic regression, among others, are some of the most popular AI approaches utilized in the energy sector [10-14]. The application of machine learning in biodiesel production has the potential to significantly enhance the efficiency and effectiveness of the process, leading to improved yields and reduced costs. In recent years, machine learning (ML) has emerged as a powerful tool for modeling and prediction driven by data, making it a subfield of AI technology. Unlike traditional modeling approaches that rely on mathematical equations, ML techniques view physical objects as black-box models based on input and output data, enabling a high level of prediction capacity based on inferences [15]. This makes ML-based techniques preferable over methodological concepts when the target system is more complex or nonlinear due to various input variables [16]. As a result, there has been a significant increase in ML-based research in recent years.

†To whom correspondence should be addressed.
E-mail: aqueel.me20@nsut.ac.in

This development can be attributed not only to the industry's growing interest in ML tools but also to the widespread availability and increasing sophistication of AI techniques.

Machine learning (ML) techniques have been applied in biodiesel production to predict and optimize process parameters for improved biodiesel yield and quality. Gupta et al. conducted a study utilizing machine learning techniques to develop a predictive model for the biodiesel production process. The study involved analyzing the response features as a function of process parameters and comparing the performance of three prominent machine learning algorithms: linear regression, random forest regression, and AdaBoost regression. The results showed that both random forest regression and AdaBoost regression were highly accurate in estimating biodiesel yield [17]. Similarly, Moayedi et al. explored the viability of four unique machine learning methods in estimating biodiesel purity. The study used various input constraints such as oil type, catalyst type, reaction temperature, catalyst concentration, reaction time, methanol-to-oil ratio, frequency, and amplitude. The response parameter was the fatty acid methyl ester output. The study employed well-established statistical measures to evaluate the predictive outputs of these methods and concluded that the alternating model tree was the most effective predictive network [18]. Li et al. used a support vector machine (SVM) algorithm to predict the optimal conditions for the transesterification reaction of waste cooking oil with methanol. The SVM model achieved a prediction accuracy of 97.7% for the biodiesel yield, demonstrating the effectiveness of the ML approach [19]. Wang et al. a deep belief network (DBN) was used to predict the optimal conditions for biodiesel production from soybean oil. The DBN model was trained on a dataset of 30 experiments and achieved a prediction accuracy of 96.5% for the biodiesel yield [20]. Shen et al. used a random forest (RF) algorithm to predict the optimal reaction conditions for biodiesel production from waste cooking oil. The RF model achieved a prediction accuracy of 93.6% for the biodiesel yield, and the predicted conditions were experimentally validated to confirm their accuracy [21]. Moreover, several other authors have explored the application of machine learning technology to renewable energy and chemical engineering research, as evidenced by studies conducted by them [22-26]. Overall, these studies demonstrate the potential of ML-based prediction models in biodiesel production, enabling the optimization of process parameters and improving biodiesel yield and quality. ML-based models can reduce the need for time-consuming and costly experimental trials, making biodiesel production more efficient and economically viable. Genetic algorithms have been widely used for biodiesel yield optimization. The genetic algorithm is an optimization technique based on the principles of natural selection and genetics. It involves generating a population of solutions, selecting the fittest individuals, and then applying genetic operators, such as crossover and mutation, to generate new solutions. In the context of biodiesel production, genetic algorithms have been used in several research studies to optimize the process parameters and improve the biodiesel yield [27]. For instance, Sajjadi et al. utilized a two-stage optimization approach that combined statistical methods such as CCD and RSM to design an experimental plan and assess the regression model equation for five independent variables with an artificial neural network-based genetic algorithm to analyze the results. The ANN-GA-based optimization focused on evaluating the influence of ultrasonic irradiation, which emerged as a crucial factor affecting the biodiesel synthesis process. Additionally, an RSM-based study demonstrated that the molar and temperature requirements for the reaction significantly affect the biodiesel yield. The optimization procedures employing RSM and ANN-GA resulted in biodiesel yields of 95.2% and 95.1%, respectively [28]. A genetic algorithm was utilized to optimize the transesterification reaction conditions for biodiesel production from jatropha oil with the objective of maximizing the biodiesel yield. The results showed that the genetic algorithm was able to find the optimal conditions, which resulted in a biodiesel yield of 92.5%. Machine learning techniques, including genetic algorithms, have also been applied to biodiesel production to optimize process parameters, predict biodiesel yield, and identify important factors that affect biodiesel production. Li et al. explored genetic algorithms to optimize the process parameters for biodiesel production from waste cooking oil using machine learning techniques. The results showed that the genetic algorithm was able to find the optimal conditions, which resulted in a biodiesel yield of 92.6% [29]. Similarly, Xia et al. conducted a study utilizing a genetic algorithm to identify the most important factors that affect biodiesel production from soybean oil using machine learning techniques. The results showed that the genetic algorithm was able to identify the most important factors and provide insights into the biodiesel production process [30].

In light of the above literature, biofuels derived from renewable and non-edible sources are appealing since they are environmentally friendly and sustainable. However, biodiesel commercialization is plagued with difficulties. Biodiesel producers are continually striving to lower production costs on a wide scale; however, maintaining the ASTM standard for quality remains a challenge. Since biodiesel yield can be varied by changing numerous parameters, there could be no other way to determine ideal combinations of biodiesel production parameters except by conducting large experimental trials. A large number of experiments require a significant expenditure of money and time. Thus, prediction and optimization methods are required for achieving the optimal combinations of operating parameters to attain the objective. The studies cited above focused on the effects of operating parameters by using an RSM-based desirability approach, but very limited significant work has been done on biodiesel production by using a machine-learning-coupled genetic algorithm approach. To address this issue, the principal objective of this study was to develop a robust machine learning-based prediction model for biodiesel production from waste cooking oil. Using a multitude of machine learning approaches, a robust, comprehensive, and advanced model has been created, and its accuracy has been enhanced by utilizing the most recent, inclusive, and well-structured datasets. The intended datasets were comprehensive, since all relevant features were taken into account. This study provides three machine learning-based approaches (Gradient boosting, XGBoost, and LightGBM regression) for predicting biodiesel yield. We use a variety of error metrics to provide a comprehensive evaluation of the machine learning regression predictive models. Furthermore, the GA model is coupled with ML predicted values to determine the optimal operating parameters for biodiesel yield.

MATERIAL AND METHODOLOGY

1. Raw Material

Waste cooking oil was selected as the feedstock for biodiesel production. Prior to the transesterification reaction, the waste cooking oil was collected from a nearby restaurant and underwent a purification process to remove impurities. Typically, waste oil contains soil, food residue, and water. Therefore, the waste cooking oil was initially passed through a filtration process to eliminate large food particles and impurities. Subsequently, the water was left to settle for an hour to allow it to become clean. The presence of water in the feedstock could lead to the hydrolysis of esters to form a long chain of free fatty acids, which could increase the chances of a saponification process. To remove the water, a mixture of used frying oil, sulfuric acid, and methanol was heated at 60 °C for 50 minutes. The methanol-water mixture was extracted from the solution's surface. The stimulated blend was then created by combining KOH catalyst with methanol at 60 °C. The solvent and catalyst streams were subsequently introduced into a transesterification reactor. The measured density and kinematic viscosities (KVs) of the waste cooking oil were 911 kg/m^3 and 38.43 cSt, respectively. To synthesize the waste cooking oil methyl ester (WCOME), methanol (99.8%), Paratoluene Sulfonic Acid (PTSA), Phenolphthalein Indicator, 0.1 N (NaOH), filter paper of 125 mm size, and KOH (85% purity) pellets were utilized.

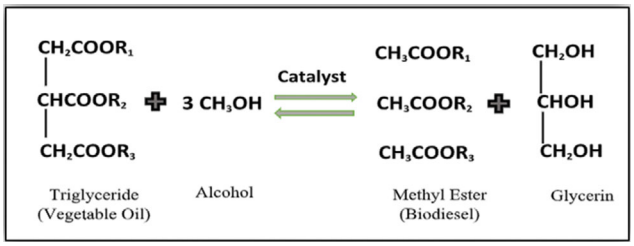

Fig. 1. Transesterification reaction.

2. Biodiesel Synthesis and Purification

The FFA content of the WCO sample utilized in this experiment was determined to be 2.5%. One-step esterification was performed to minimize FFA levels. P-Toluene sulfonic acid (0.5% w/w) and methanol (20% w/w) were added to 1liter of oil at 60 °C. The mixture was stirred on a magnetic stirrer at 450 rpm for two hours. After esterification, the FFA value for WCO was measured to be 1.3. WCOME produced by the transesterification process as explained in Fig. 1, in which present triglycerides reacted with methanol in presence of KOH as a base catalyst.

The transesterification reaction is controlled by varying values of the independent parameters methanol-oil molar ratio (3 : 1-

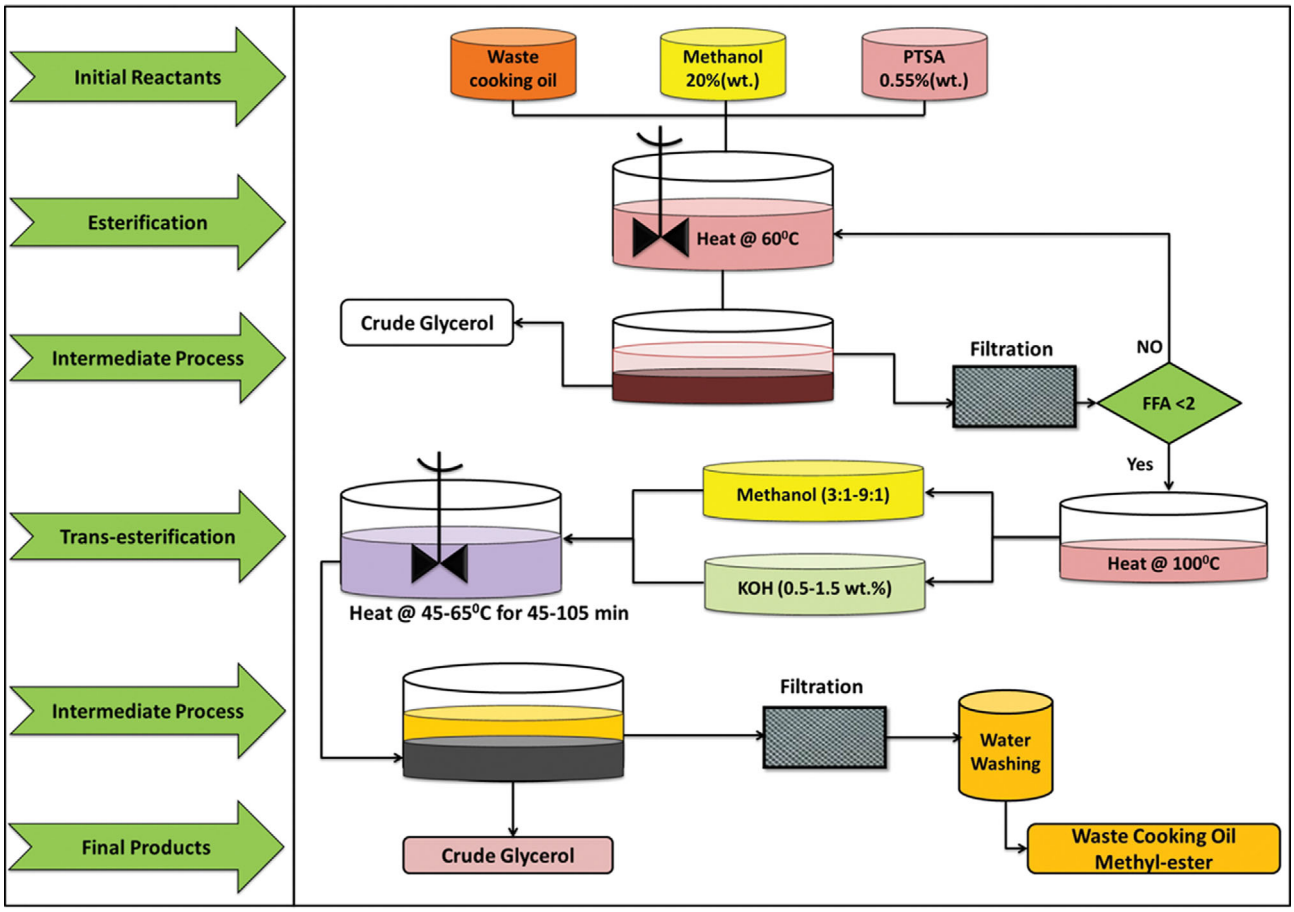

Fig. 2. Flow diagram of Waste cooking oil biodiesel production.

9 : 1), catalyst concentration (0.5-1.5 wt%), reaction temperature (45-65 °C), and reaction time (45-105 min). Subsequently, the outcome of the transesterification reaction is channeled into a separation unit, which is composed of two distinct layers. The lowermost layer comprises glycerol, an undesirable byproduct, which is purged accordingly. The uppermost layer, on the other hand, contains the desired biodiesel, which is then transported to a storage container. A comprehensive depiction of the entire process configuration is presented in Fig. 2.

3. Biodiesel Characterization

The FFA ratio present in waste cooking oil is assessed through the process of titration. To begin with, 1 gram of WCO and 10 milliliters of isopropyl alcohol are placed into a jar and thoroughly mixed. A few drops of the pH indicator phenolphthalein are then added to the mixture, following which a titration with N/10 NaOH is carried out to determine the FFA content. The amount of N/10 NaOH consumed is then measured, and the concentration of the solution is adjusted until it turns pink. It is important to ensure that the solution maintains a steady pink color for at least 10 seconds. The acid value (AV) is subsequently expressed by Eq. (1).

$$AV = \frac{VOl_{NaOH} \times Conc_{NaOH}(M) \times XMW_{KOH}}{\text{Sample weight (gr)}} \quad (1)$$

The fatty acid composition of biodiesel was analyzed by means of gas chromatography (GC) utilizing a built-in flame ionization detector. A sample of 0.2 g of WCOME was injected into the GC chamber, which was furnished with a split/split-less injection port with a split ratio of 50 : 1. The temperature was initially maintained at 120 °C for duration of 2 minutes, following which it was ramped up at a rate of 40 °C per minute until it reached its final temperature of 240 °C, at which it was held for 7 minutes. Nitrogen gas was used as a carrier gas at a flow rate of 8 ml/min. During the operation of the GC, five peaks corresponding to palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3) were detected with GC area yields of 13.0264%, 3.350%, 24.1820%, 51.7410%, and 7.700%, respectively, as presented in Table 1. The biodiesel production yield was calculated using Eq. (2).

$$\%\ \text{Yield} = \frac{\text{Weight of biodiesel produced} \times 100}{\text{Weight of raw oil}} \quad (2)$$

4. Machine Learning (ML) Models

Prediction is an effective ML technique for predictive data analytics. It utilizes historical or training data to draw fresh input archives to relevant dependent output parameters according to the values of independent factors [31]. It is essential to employ novel prediction algorithms to achieve the highest level of precision due to their capacity to manage complex-related factors and their efficacy in controlling linked variables [32]. In this study, three prediction algorithms were applied (Gradient Boosting, XGBoost, and LightGBM regression). XGBoost and LightGBM are regarded as two of the newest and most effective ML-based prediction algorithms. A detailed literature analysis of alternative machine learning prediction methodologies led to the selection of the accurate and reasonably quick XGBoost and LightGBM algorithms to enable more regularized model establishment and enhanced over-fitting management. Training and testing sets were separated from the datasets. This step is crucial for assessing the efficacy of the implemented ML methods. The adopted algorithms are developed using the training component of the datasets, while the testing portion is utilized to demonstrate the model's reaction to newly processed data. The datasets used in this analysis were split into training and testing portion of 80% and 20%, respectively. Cross-validation was used to check that the proposed prediction models were accurate, robust, and valid. Following is a discussion of the properties of each approach, complete with illustrations and implementation instructions.

4-1. Gradient Boosting Regression (GBR)

Boosting is a powerful technique for combining many base classifiers to construct a consensus whose performance can be far superior to that of any basis classifier. “Boosting” is used to add models to an ensemble in a systematic manner. The ensemble's error is used to inform the training of a new, weak base-learner model at each iteration [33]. Like other boosting methods, gradient boosting builds the model incrementally and generalizes it by permitting the optimization of an arbitrarily differentiable loss function [34]. Gradient descent is used to address the minimization problem, and the resulting prediction model is an ensemble of low-fidelity prediction methods, typically decision trees. The authors have used the gradient boosting regressor (GBR) to estimate the biodiesel yield prediction, as detailed in the introduction. GBR comprises three elements: a loss function (which must be minimized), a weak learner (used to make predictions), and an additive model (to add weak learners to reduce the loss function) [35]. The key idea of this technique is to generate new base learners that are optimally associated with the negative gradient of the ensemble-wide loss function. The applied loss functions can be chosen freely, although for the purpose of understanding, if the error function is the typical squared-error loss, the learning strategy will result in successive error-fitting. In the broad-spectrum, it is up to the researcher to choose the loss function. He or she can choose from a number of loss functions

Table 1. Fatty acid composition of waste cooking oil

S. No.	Fatty acid	Carbon structure	Waste cooking oil biodiesel (wt%)	Peak width
1	Palmitic acid methyl ester	C16:0	13.0264	0.396
2	Stearic acid methyl ester	C18:0	3.3501	0.442
3	Oleic acid methyl ester	C18:1	24.1820	0.505
4	Linoleic acid methyl ester	C18:2	51.7410	0.412
5	linolenic acid methyl ester	C18:3	7.7003	0.369

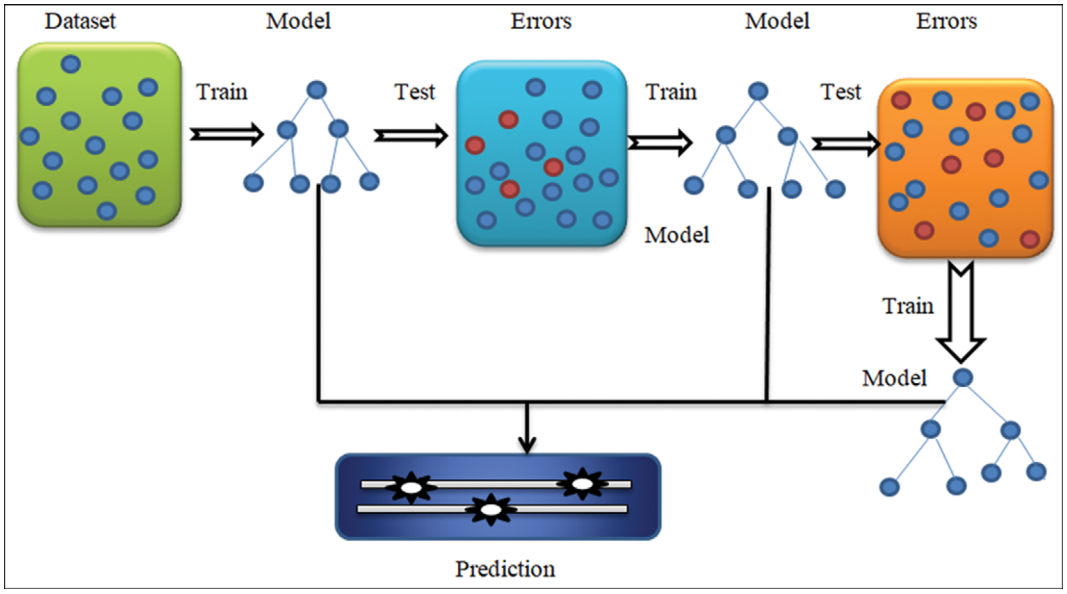

Fig. 3. Gradient boosting (GB) regression.

that have already been made, or they can make a loss function that is specific to the task. The following is the precise mathematical formulation of GBR with M number of trees:

$$f_M(Xi) = \Sigma_m^M \gamma_m h_m(Xj) \quad (3)$$

where h_m is a weak learner with poor individual performance, γ_m is a scaling factor that adds the tree involvement in the model. GBR employs the gradient descent loss function to decrease errors by updating the starting estimate with the new estimate. Thus, a final model is produced by combining all early estimates with appropriate weights. This test's GBR model is based on the gradient-boosting regression (GBR) approach given by scikit-learn. Fig. 3 depicts the GBR algorithm. GBR has been applied for the prediction of biodiesel yield from various feedstocks and process conditions. The performance of GBR has been compared with other regression techniques in different studies, and has been shown to be a promising tool for biodiesel yield prediction. Singh et al. used GBR to predict the yield of biodiesel from waste cooking oil. The study compared the performance of GBR with other regression techniques and found that GBR performed better in predicting the biodiesel yield [36]. Li et al. aimed to predict biodiesel yield from different vegetable oils using GBR. The study compared the performance of GBR with other regression techniques and found that GBR outperformed the other techniques in predicting biodiesel yield [37]. In a recent study by Shukla et al., GBR was used to predict the yield of biodiesel from jatropha oil. The study compared the performance of GBR with other regression techniques and found that GBR performed better in predicting the biodiesel yield. These studies demonstrate the effectiveness of GBR in predicting biodiesel yield. The use of GBR for biodiesel yield prediction can aid in the optimization of biodiesel production processes and the selection of feedstocks for biodiesel production [38].

4-2. eXtreme Gradient Boosting (XGBoost) Regression

eXtreme Gradient Boosting (XGBoost) is a popular machine learning algorithm that has been applied to a wide range of applications, including biodiesel yield prediction. The method's organization, portability, and adaptability make it suitable for a variety of applications [39]. XGBoost combines a cause-based decision tree (CBDT) and GBM to create an efficient algorithm that improves the speed and accuracy of the tree-boosting method for processing virtually any dataset. This approach is effective in constructing a prediction model when regression and classification techniques are applied to a given dataset. Furthermore, XGBoost can handle large datasets with multiple variables and classifications. When balancing the tradeoffs between efficiency and precision, this method provides effective strategies for tackling new optimization challenges. The XGBoost algorithm calculates the first- and second-order gradients for each training case's "squared error" objective function at each boosting iteration. To construct the model, XGBoost, which is compatible with Scikit-Learn, was utilized. Tree-based learners in XGBoost performed optimally with 100 boosts, 0.08 for "colsample by tree", and 0.04 for the learning rate. Fig. 4 depicts the XGBoost regression algorithm. Kocyigit et al. utilized XGBoost for the prediction of biodiesel yield from waste cooking oil. They used several input parameters, including temperature, reaction time, and catalyst amount, to predict the yield. The results showed that XGBoost outperformed other machine learning algorithms in terms of prediction accuracy, with a mean absolute error (MAE) of 1.77%. The study concluded that XGBoost is an effective tool for predicting biodiesel yield and can be used in the optimization of biodiesel production processes [40].

4-3. Light Gradient Boosting Machine (LGBM) Regression

Machine learning algorithms have been gaining popularity in various fields, including the prediction of biodiesel yield. Among these algorithms, the light gradient boosting machine (LGBM) has been shown to offer exceptional efficiency, accuracy, and speed, making it an attractive option for biodiesel yield prediction. LGBM integrating two unique data sampling and categorization approaches

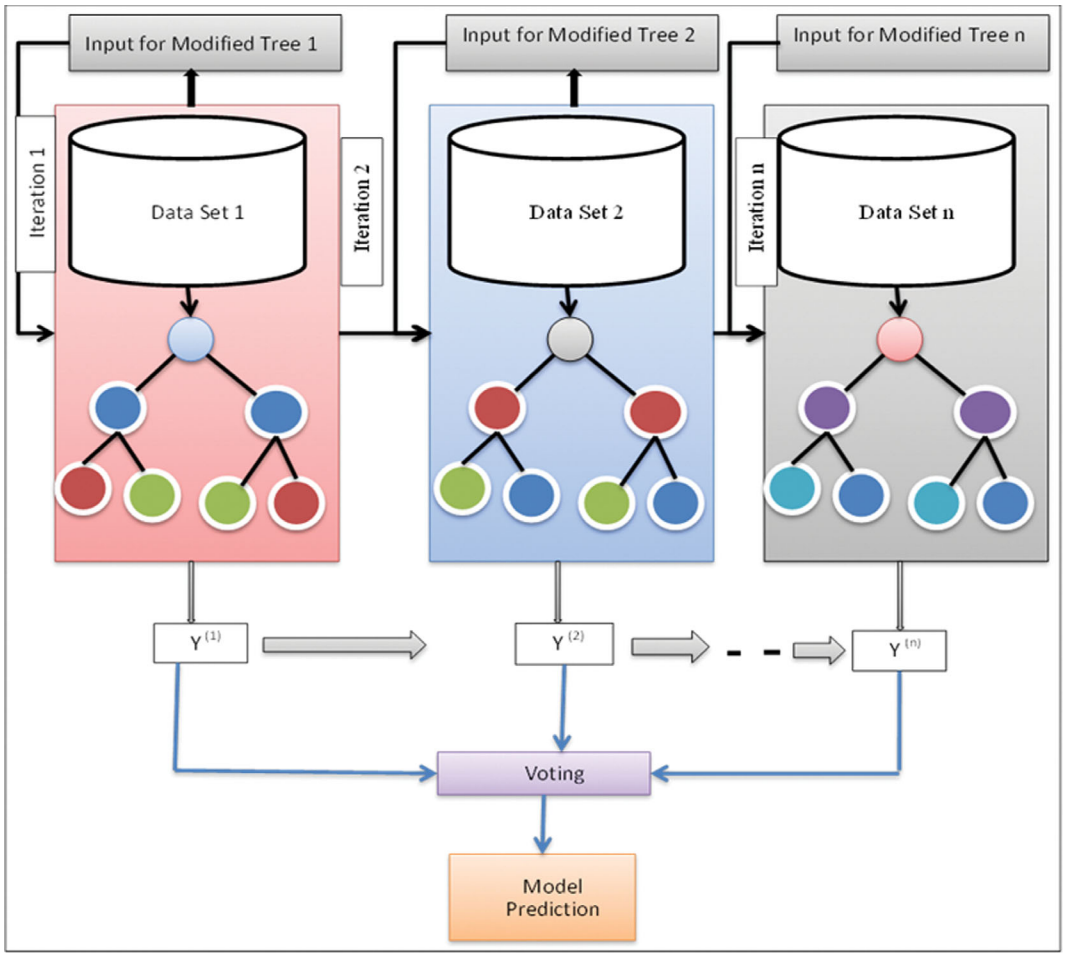

Fig. 4. eXtreme Gradient Boosting (XGBoost) regression.

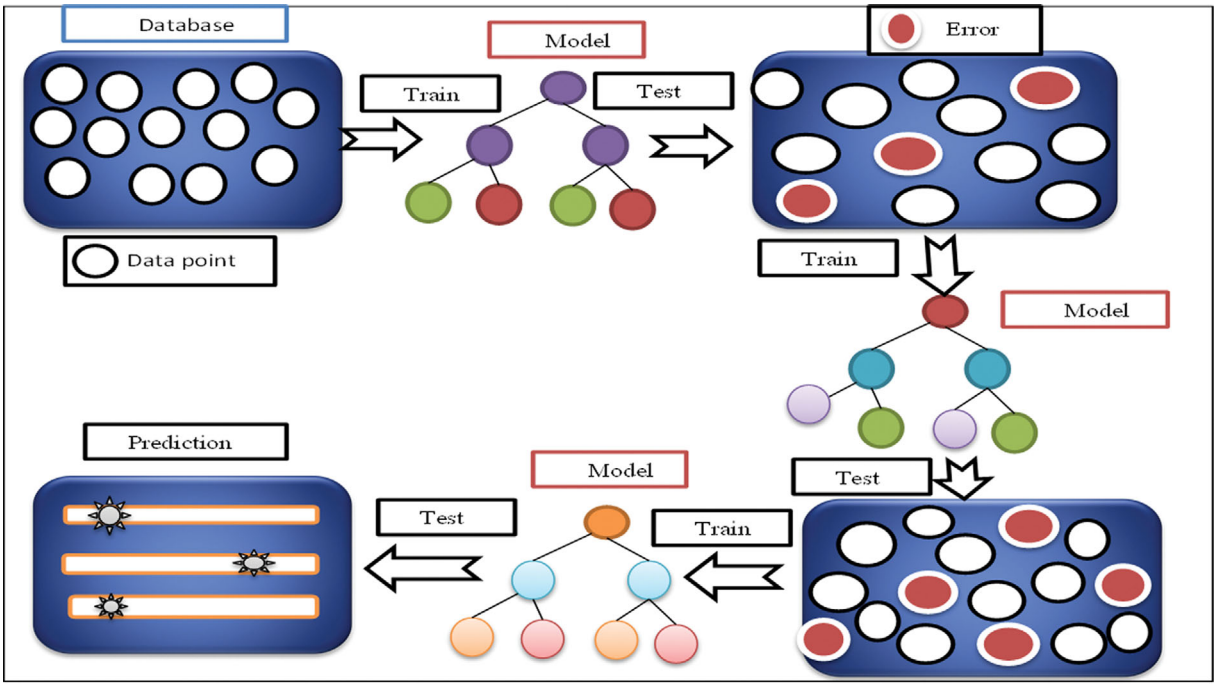

Fig. 5. Light Gradient Boosting Machine (LightGBM) regression.

Table 2. Biodiesel production process parameters and biodiesel yield

S.N	Molar ratio (X1)	Catalyst concentration (X2: Wt%)	Reaction temperature (X3: °C)	Reaction time (X4: Min)	Biodiesel yield (Y: Wt%)
1	6	0.5	55	75	75.79
2	4.5	0.75	60	90	89.36
3	6	1	55	75	91.69
4	6	1	65	75	97.02
5	7.5	0.75	50	90	86.21
6	7.5	1.25	50	90	92.13
7	4.5	0.75	60	60	81.37
8	4.5	1.25	50	60	81.34
9	6	1	45	75	89.16
10	6	1	55	75	91.87
11	4.5	1.25	60	75	75.57
12	9	1	55	75	97.49
13	4.5	1.25	50	90	80.01
14	6	1	55	45	83.93
15	6	1	55	105	98.17
16	4.5	1.25	60	60	75.38
17	6	1	55	75	91.48
18	6	1	55	75	91.37
19	7.5	0.75	60	60	89.6
20	7.5	0.75	60	90	95.69
21	7.5	1.25	60	60	92.03
22	4.5	0.75	50	90	83.18
23	9	1	55	75	97.35
24	7.5	0.75	50	60	79.71
25	3	0.5	45	45	55.37
26	7.5	1.25	60	90	93.095
27	6	1	55	75	91.73
28	6	1.5	55	75	88.21
29	7.5	1.25	50	60	90.09
30	4.5	1.25	50	60	81.81

has resulted in a newly developed technique. With such a combination of characteristics, data scanning, sampling, grouping, and categorization can be performed with higher efficiency and accuracy than with analogous methods. LightGBM is a wise idea when evaluating training speed, efficiency, memory usage, processing speed, arithmetic speed, ICU utilization, sufficient accuracy, parallelism, and the ability to process massive amounts of data with minimal delays. The mechanism of light-GBM regression is depicted in Fig. 5. Singh et al. utilized LGBM for biodiesel yield prediction using data collected from experiments on a transesterification process. The results showed that LGBM outperformed other machine learning algorithms, such as artificial neural networks (ANN) and support vector regression (SVR), with an accuracy of 99.81%. This demonstrates the potential of LGBM for predicting biodiesel yield with high accuracy [41]. Wu et al. compared the performance of LGBM to that of other machine learning algorithms, including XGBoost and GBR, in predicting biodiesel yield using data from different feedstocks. The results indicated that LGBM had higher accuracy and was more efficient in terms of memory usage and processing time [42].

5. Description of Datasets

This section discusses the dataset and the methods for concentrating on the algorithms that result in the proposed models. The data set has 30 data points, four inputs, and one output as shown in Table 2. The inputs are X_1=the molar ratio (methanol to waste cooking oil), X_2=catalyst concentration, X_3=reaction temperature, and X_4=reaction times and the output Y=methyl ester production yield (wt%). There are six primary stages to the modeling process.

Phase 1: Data Collection and Description: 30 data points of biodiesel yield datasets were obtained from the RSM central composite design (CCD) the specified data was exported into Excel documents.

Phase 2: Data Preparation: A preprocessing phase was performed in advance of the development of each model in order to improve its predictive abilities.

Phase 3: Feature Selection: This is the practice of determining the most important input parameters that influence biodiesel yield.

Phase 4: ML Modeling and Comparison: for predicting biodiesel

yield, gradient boosting, XGBoost, and LightGBM regression modeling techniques are used.

Phase 5: Model Performance Evaluation and Validation: the accuracy of all methods is determined by comparing actual and predicted values. In addition, the most appropriate model was determined using five accuracy assessment criteria (i.e., R^2, RMSE, MAE, and MAPE).

Phase 6: Output Exemplification and Discussion, in which modeling outcomes are evaluated and analogized to experimental value for a robust approach.

6. Evaluation Method

There are a multitude of measurement metrics available for conducting performance evaluations. For the purposes of this investigation, four capability evaluation measuring indicators were utilized to assess continuous attributes: R^2, RMSE, MAE, and MAPE. These metrics were calculated for all testing datasets in order to evaluate the appropriateness and reliability of the created prediction models. R^2, which indicates the precision of the prediction, was especially important in achieving greater accuracy in the prediction as it approached 1. When evaluating the performance of models with

Table 3. Mathematical representation for performance metrics

Measure	Formula
Mean square error (MSE)	$MSE=\frac{1}{n}\Sigma_{i=1}^{n}(y_i-\hat{y}_t)^2$
Mean absolute error (MAE)	$MAE=\frac{1}{n}\Sigma_{i=1}^{n}\lvert(y_i-\overline{y}_t)\rvert$
Root mean square error (RMSE)	$RMSE=\sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(y_i-\hat{y}_t)^2}$
Mean absolute percentage error (MAPE)	$MAPE=\frac{1}{n}\Sigma_{i=1}^{n}\left\lvert\frac{y_i-\overline{\hat{y}}_t}{y_i}\right\rvert\times 100$
Coefficient of determination (R^2)	$R^2=1-\frac{\Sigma_{i=1}^{n}(y_i-\hat{y}_t)^2}{\Sigma_{i=1}^{n}(y_i-\overline{y}_t)^2}$

where - n is sample size, y_i is the actual value of the i^{th} case, $\hat{y}_t$ is the predicted value of i^{th} case, $\overline{y}_t$ is the arithmetic mean of y.

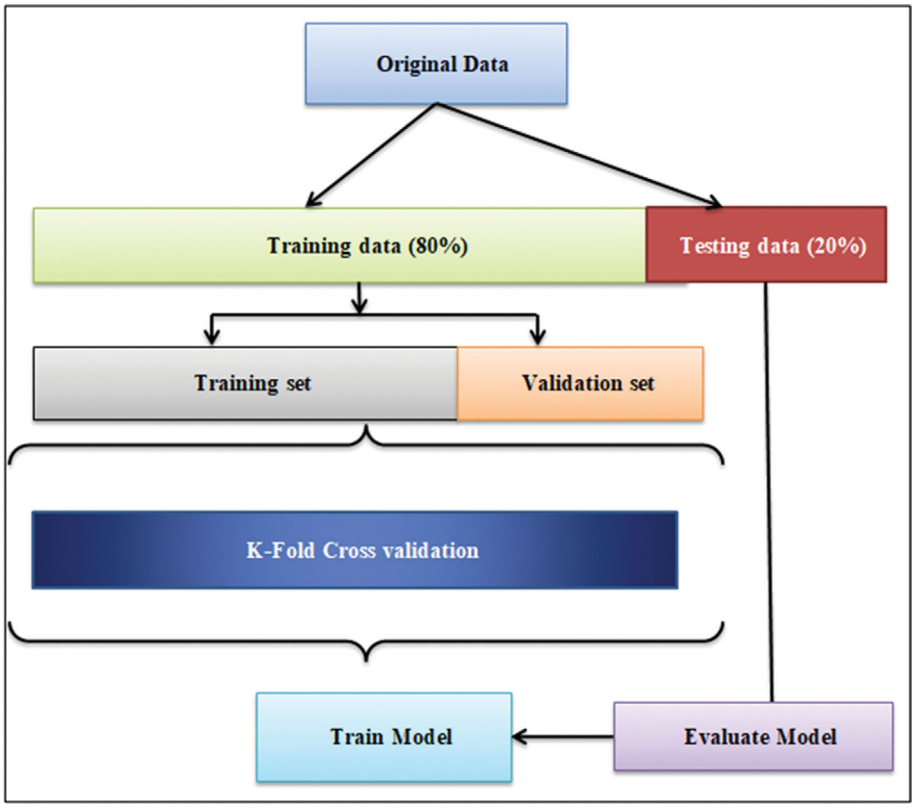

Fig. 6. K-fold cross validation representation.

a continuous dependent parameter, popular measures such as MAE, RMSE, and MAPE are often employed. The objective of this research was to achieve the most precise prediction possible while also acknowledging the potential for errors in such predictions. As such, these metrics were deemed superior measurements of accuracy as they provided insight into the potential errors of the prediction. This metric was used to evaluate the trained models on the test dataset, and the mathematical formulation of the adopted procedures can be found in Table 3.

7. Validation of Predictive Model

Upon the separation of datasets into training and testing subgroups, the modeling process for three prediction models was carried out. This stage prepares for the cross-validated internal confirmation test. To ensure effectiveness and consistency, stratified k-fold cross-validation was employed for validations. This technique involves the arbitrary division of each dataset into k folds, where k−1 folds are utilized for training, and the remaining fold is used for testing. The three prediction models were subjected to iterative evaluations using stratified k-fold cross-validation, ranging from k=1 to k=10. Stratified 10-fold cross-validation is a well-known approach for assessing the validity of models and the competence of their developers. Remarkably, the best results for each of the three algorithms were obtained using 10-fold cross-validation. Fig. 6 depicts a diagrammatic representation of the k-fold cross-validation method.

8. Optimization through Genetic Algorithm (GA)

Genetic algorithms (GA) are powerful optimization tools that can be used to optimize the biodiesel yield. GA is a search algorithm that mimics the process of natural selection and can be used to find the optimal set of input parameters for a given objective function. Since genetic algorithms can successfully solve discontinuous fitness functions, they have eclipsed the popularity of conventional optimization methods [43]. In recent years, numerous investigations have been conducted using GAs. Ighose et al. optimized biodiesel production from Thevetia Peruviana oil using GA and RSM with an adaptive neuro-fuzzy inference system. The highest FAME yields for GA and RSM were 99.8% and 98.8%, respectively [44]. Bhaskar et al. used GA to optimize the reaction conditions for biodiesel production from waste cooking oil. They found that the optimal reaction conditions were a temperature of 65 °C, a molar ratio of methanol to oil of 6 : 1, and a reaction time of 90 minutes, which resulted in a biodiesel yield of 96.2% [45]. Kalam et al. used GA to optimize the feedstock composition for biodiesel production from Jatropha curcas oil. They found that the optimal feedstock composition was a blend of 40% Jatropha curcas oil, 30% palm oil, and 30% canola oil, which resulted in a biodiesel yield of 95.2% [46]. Onukwuli et al. compared the use of ANN-GA and RSM's desirability function for biodiesel synthesis from Chrysophyllum albidum seed oil and found that ANN-GA was 9.40% faster, 15.88% less energy-intensive, and 16.82 percent less catalyst-intensive than RSM [47].

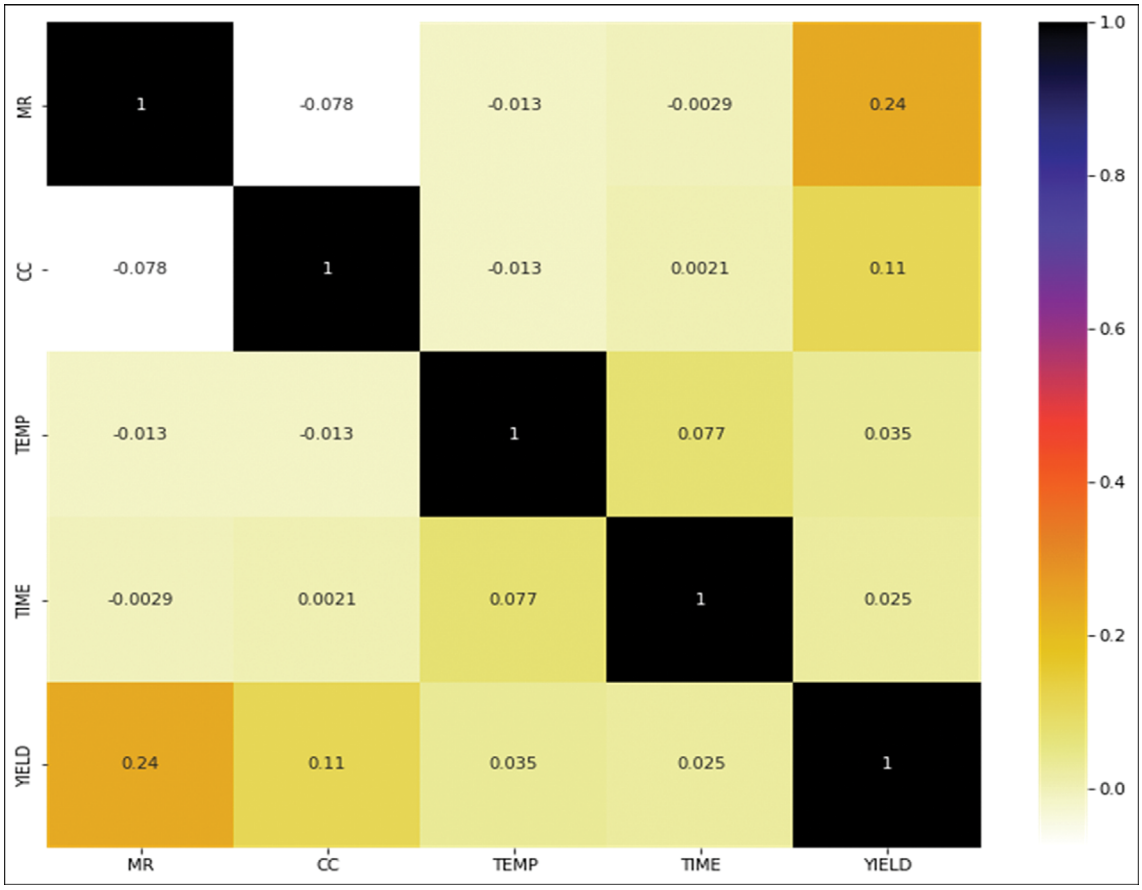

Fig. 7. The Kendall correlation map of biodiesel yield.

Table 4. Validation of regression models and performance evaluation

Regression model	Performance evaluation metrics			
	Coefficient of determination (R^2)	RMSE	MAE	MAPE (%)
Gradient boosting regression (GBR)	0.85831	10.858	8.433	18.15
XGB regression	0.86479	9.702	7.799	17.83
LGBM regression	0.94048	8.929	6.722	13.60

RESULTS AND DISCUSSION

Before initiating the prediction process, it is imperative to identify the most significant variables and correlations that can expedite the procedure and prevent prospective over fitting. This can be achieved by minimizing the number of variables considered. Furthermore, a performance evaluation of the models is presented, where the accuracy of the prediction outcomes is validated and the competency of the three algorithms is compared for various k-fold cross-validations. This evaluation serves to demonstrate the effectiveness of the models and enables the selection of the optimal algorithm for the given task.

1. Correlation Analysis

In correlation analysis, a substantial linear relationship between at least two continuous variables or several variables and a dependent variable is determined. The correlation coefficient quantifies this connection. In a negative correlation, one variable surges while the other decreases, but in a positive correlation both variables vary in the same way. However, correlation analysis does not provide information regarding the connection between variables and their causes. Before performing a successful analysis, it is necessary to test for normality and linearity using graphs and computations. In this investigation, Kendall rank correlation analyses are examined, and the outcomes are presented in Fig. 7. Fig. 7 examines the relationship between the various process factors and the yield. A negligible connection exists between the process parameters, confirming the absence of multicollinearity. The relationship between yield and the molar ratio (MR) is moderately strong (0.24). An insignificant relationship was observed between yield, reaction temperature, and reaction time. Though, a very modest positive correlation of 0.11 occurs between catalyst concentration and yield. This suggests that, on average, low levels of catalyst concentration will not result in excessive biodiesel production. Therefore, the interpretation of the experimental statistics on biodiesel yield, as depicted in Fig. 7, indicates that a brute-force approach to finding the optimal combination of process parameters to boost yield would be exceedingly expensive. Due to the absence of a discernible pattern in the scatter plots of this figure, it may be necessary to conduct a comprehensive experimental investigation of all potential process parameter arrangements. This emphasizes the requirement for robust machine learning techniques to develop prediction models for quantifying yield as a function of process variables.

2. ML Model's Performance Evaluation

After arbitrarily designating 80% of the datasets for the training of the model, the residual 20% were used to evaluate the model's competency. This study used four measures of capability evaluation (RMSE, MAE, MAPE, and R^2) that were very useful for evaluating continuous attributes. All performance indicators for the three prediction methods utilized in this research to quantify the difference between observed and predicted values were generated to ensure consistent and accurate performance estimation. As performance metric values decline, the model's precision increases. R^2 measures the precision of a prediction. In addition, the RMSE, MAE, and MAPE metrics can be used to compare the predictive accuracy of the three algorithms. Table 3 details the performance evaluation findings for the regression model. The k-fold cross-validation analysis was performed repetitively for k=10. Cross-validation with a 10-fold sample size yielded the highest prediction accuracy. Table 4 depicts a group of random, non-overlapping, partitioned folds used as training and test datasets for k=10 along with their respective performance evaluation measures.

3. Comparison of Various ML Predictive Models

In this research, the LGBM regression model employed in 10-fold cross-validation demonstrated a high degree of precision in predicting biodiesel yield, with an accuracy level of 94.048%. The XGB regression and gradient boosting regression followed with accuracy levels of 86.45% and 85.82%, respectively. Notably, the utilization of 10-fold cross-validation resulted in the highest level of precision for all three algorithms. These accuracy outcomes were further supported by the RMSE, MAE, and MASE data. A model's accuracy is rated as excellent when MAPE is less than 10, good when $10 \leq MAPE \leq 20$, and acceptable when $20 < MAPE \leq 50$. The study's findings showed that the LGBM regression model had a MAPE of 13.60, indicating comparable predictive capacity. On the other hand, the XGB regression and gradient boosting regression models had MAPEs of 17.83 and 18.15, respectively, demonstrating strong predictive abilities. The evaluation results suggest that the unique LGBM regression model can be utilized to aid in biodiesel yield prediction decisions, given its high predictive performance and efficiency. Specifically, LGBM was the fastest model, with a processing time of 163 milliseconds, followed by XGB regression at 166 milliseconds and gradient boosting at 283 milliseconds. The 10-fold cross-validation performance evaluation results are comprehensively presented in Fig. 8. Fig. 9 depicts the machine-learning prediction models' performance across multiple error measures, showing that the LGBM regression model outperforms the XGB regression and gradient boosting regression models.

4. Optimization of Process Parameters through GA

An effort was created to identify the optimum input process factors in order to achieve the highest feasible biodiesel produce (%). A powerful optimization method, the genetic algorithm (GA), was developed to solve the investigation's optimization problem.

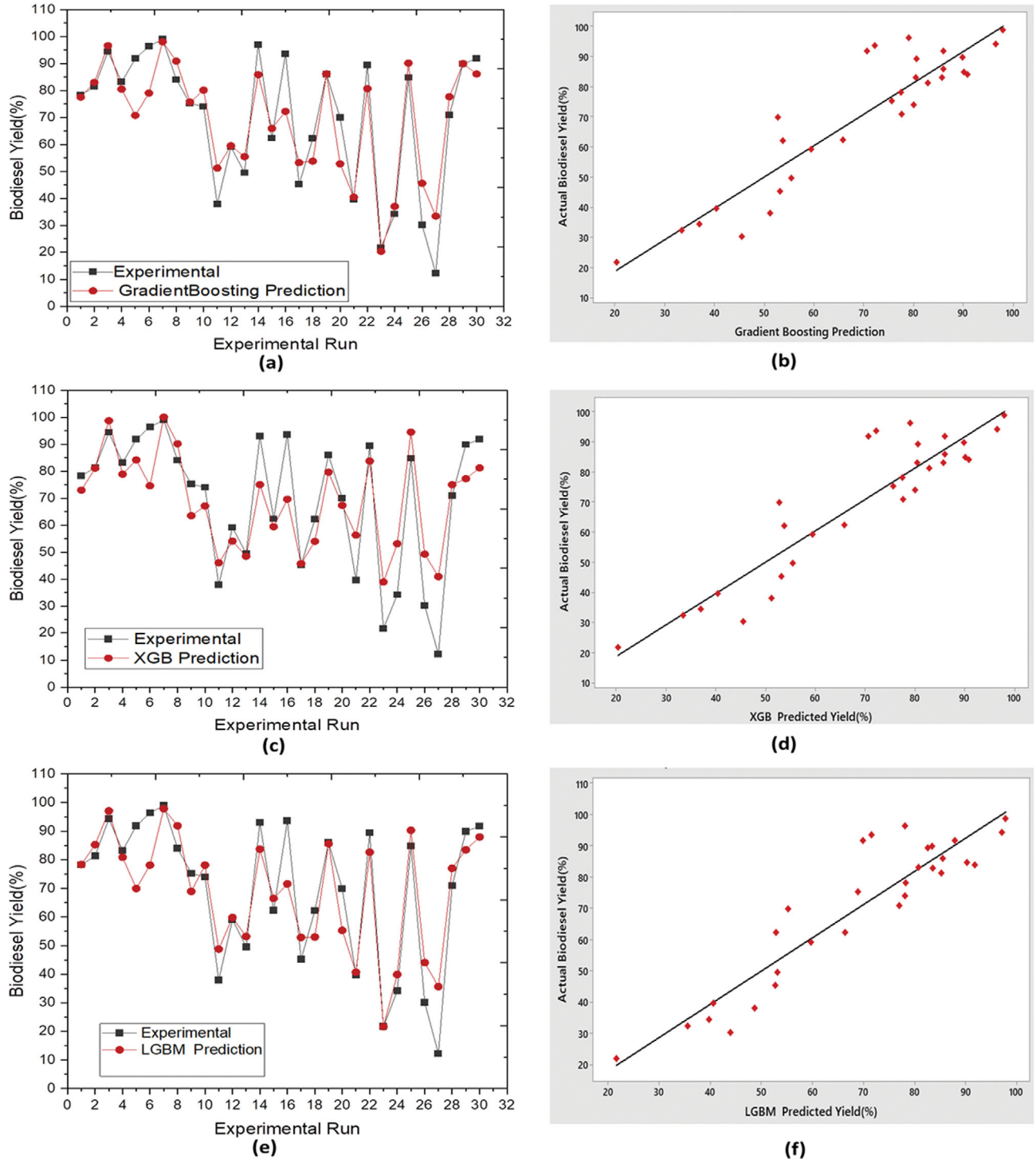

Fig. 8. ML-based performance model ((a) and (b)): observed vs. gradient boosting predicted ((c) and (d)) observed vs. XGB predicted ((e) and (f)) Observed vs. LGBM predicted biodiesel yield.

Following are the important GA steps:

Step I: Develop a random chromosome community

Step II: Decipher all the genes of the chromosomes, comprising the molar ratio (x_1), catalyst concentration (x_2), reaction temperature (x_3), and reaction time (x_4).

Step III: By means of the Design expert developed RSM models and the input parameters provided, compute the expected value of biodiesel yield (%).

Step IV: Define the suitability of each chromosome as well as the level of fitness (suitability) that is the highest (fitmax).

Step V: If fitmax is less than the required level of fitness, then carry out the subsequent genetic actions:

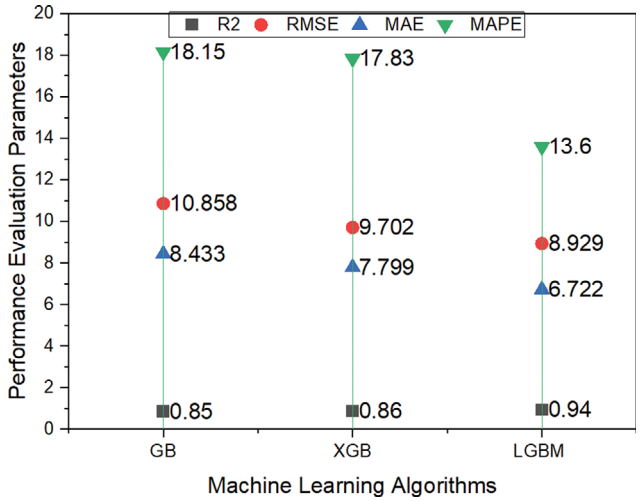

Fig. 9. Hierarchical views of the 10-fold cross-validation performance evaluation results.

1. Technique of choice centered on expected number control,
2. Crossover, and
3. Mutation to generate a new chromosome population; if not, advance to step 2; else, stop.

The following is the recommended mathematical format for presenting the problem of maximizing biodiesel produce: Locate: x_1, x_2, x_3, and x_4.

Maximize: Biodiesel Yield (%) with respect to the restraints

$3 \le x_1 \le 9$

$0.5\% \le x_2 \le 1.5$ wt%

$45 \le x_3 \le 65$ °C

$45 \le x_4 \le 105$ min

The genetic algorithm utilizes a mathematical equation as the input parameter, and a quadratic RSM is employed to model the response surface based on CCD response values at a 95% confidence level. GA determines minimum and maximum values for input process parameters, with mutation rate, generations, population size, and cross-over rate serving as GA parameters. The fitness functions are described by Eq. (4). Table 5 shows GA factor settings, whereas Table 6 shows experimental and GA findings. The GA algorithm offers the best and mean values (Fig. 10) and the best individuals based on variables.

Fitness function=−(197.61559−2.07911*x1+116.98840*x2
+6.34164*x3+1.18373*x4+4.80656*x1.*x2
+0.24655*x1.*x3+0.033586*x1.*x4−1.69226*x2.*x3
−0.37584*x2.*x4−0.00557045*x3.*x4−1.30339*x1.^2
−12.01230*x2.^2−0.050331*x3.^2−0.00415141*x4.^2); (3)

Table 5. GA parameter settings

Parameter	Values
Population type	Double vector
Population size	50
Creating function	Constraint dependent
Scaling function	Rank
Selecting function	Stochastic uniform
Elite count	0.05*population size
Crossover fraction	0.8
Mutation function	Constraint dependent
Crossover function	Constraint dependent
Direction	Forward
Hybrid function	None
Stopping criteria: generation	(100*Number of variables)=400
Stopping criteria: Time	Infinite
Stopping criteria: Fitness limit	Infinite
Stopping criteria: stall generation	50

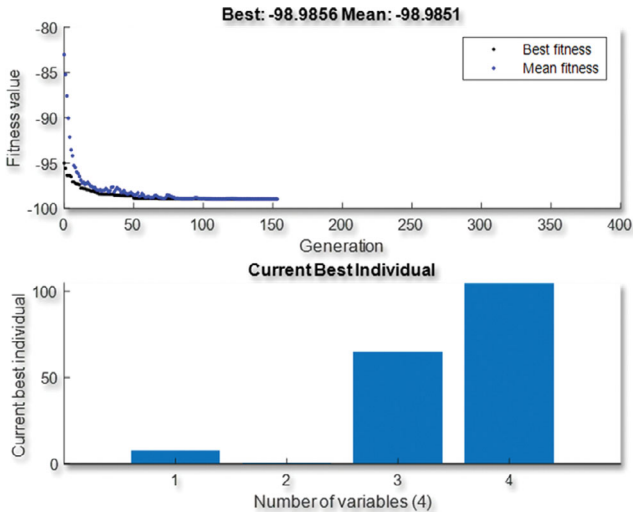

Fig. 10. Results of best fitness, mean fitness and current best individual for biodiesel yield.

5. Experimental Validation

By specifying independent variables for the GA approach, exper-

Table 6. Validation test

Input variable	Lower level	Higher level	Optimum value	
Molar ratio	5	25	7.62 : 1	
Catalyst concentration (%wt)	4	8	0.50	
Reaction temperature (°C)	30	50	65	
Reaction time (min)	0	40	105	
Response	Goal	GA predicted	Observed	Error (%)
Biodiesel yield (wt%)	maximize	98.98	96.83	2.17

imentation was conducted to validate the optimization results. Using the optimal parameter combinations, the confirmatory tests demonstrate that the prediction is accurate. Table 6 shows the optimal input process factors forecasted by GA for a maximum biodiesel yield of 98.98%. The error rate found to be the highest is 2.17%. Consequently, the expected and observed values are closely related.

Table 7. Machine learning algorithms and optimum conditions for biodiesel production via transesterification in published literature

Feedstock	Machine learning (ML) models	Most accurate and efficient model	Important features and performance	Optimum condition	Yield (%)	Reference
Waste cooking oil	Support vector machine (SVM) model	-	-	M.R=6 : 1, CC=1.25 wt%, T=60 °C, t=1.5 h	97.7%	[19]
Soybean oil	Deep belief network (DBN) model	-	-	M.R=6 : 1, CC=1.5 wt%, T=60 °C, t=1.5 h	96.5%	[20]
Waste cooking oil	Random forest (RF) model	-	-	M.R=6 : 1, CC=1.2 wt%, T=60 °C, t=1.5 h	93.6%	[21]
Flaxseed oil	Artificial neural networks (ANNs), support vector machines (SVMs), decision trees (DTs), and random forests (RFs)	Random forests (RFs)	Molar ratio, catalyst concentration, reaction temperature, and reaction time	-	Not reported	[17]
Waste cooking oil	Artificial neural networks (ANNs) and support vector regression (SVR)-GA	ANN	-	M.R=6 : 1, CC=1.2 wt%, T=60 °C, t=1.5 h	92.6%	[29]
Soybean oil	Decision trees (DT), support vector regression (SVR), and artificial neural networks (ANNs)-GA	ANN	Molar ratio, reaction time, and reaction temperature	-	Not reported	[30]
Waste cooking oil	Gradient boosting] machine (GBM)	-	Methanol-to-oil ratio and the reaction time R^2=0.985, MAPE=1.33%	-	-	[36]
Rapeseed oil, soybean oil, and palm oil	Gradient boosting regression (GBR)	-	Reaction time and temperature MAPE of less than 3%	-	-	[37]
Jatropha oil	Decision tree regression (DTR), random forest regression (RFR), support vector regression (SVR), and artificial neural network (ANN)	Random forest regression (RFR)	R^2=0.979, RMSE=1.71%	-	-	[38]

Table 7. Continued

Feedstock	Machine learning (ML) models	Most accurate and efficient model	Important features and performance	Optimum condition	Yield (%)	Reference
Waste cooking oil	Decision tree, random forest, support vector regression, k-nearest neighbor, artificial neural network, and gradient boosting regression	Gradient boosting regression	R^2=0.996, MSE=0.0005	-	-	[40]
Soybean oil, Rapeseed oil, Waste cooking oil, and Palm oil	Decision tree, random forest, and artificial neural network	Random forest algorithm	Amount of alcohol, catalyst, and reaction temperature R^2=0.997	-	-	[42]
Waste cooking oil	Gradient boosting regression (GBR), XGB regression, LGBM regression and genetic algorithm	LGBM regression	Molar ratio, catalyst concentration, reaction temperature, and reaction time R^2=0.94048, RMSE=8.929, MAE=6.722, MAPE (%)=13.60	MR=7.62 : 1, CC=0.5 wt%, T=65 °C, T=105 min	98.98% (GA Predicted) 96.83% (Observed)	Present survey

In Table 7, we present a comparison of the different machine learning algorithms and optimal operating conditions obtained in this study with those reported in the literature using various criteria and optimization techniques.

6. Physiochemical Properties of Produced Biodiesel

Biodiesel produced for use in diesel engines as an alternative fuel must satisfy fuel standards. Physical and chemical properties were calculated and compared to ASTM criteria to evaluate the excel-

Table 8. Physico-chemical properties of biodiesel under optimum operating conditions

Property	Produced biodiesel	Diesel	Equipment used	ASTM standard limit	Test method	Uncertainty (%)	Aghbashlo et al. study [24]
Density@15 °C (g/cc)	0.871	0.830	Pyknometer	0.860-0.900	D 1298	0.1%	0.872
Kinematic viscosity @40 °C (cst)	5.81	2.3	Brookfield viscometer	1.9-6.0	D 445	0.05%	3.54
Flash point (°C)	154	53	Pensky-Martens apparatus	Min 120	D 93	0.05%	-
Pour point (°C)	1	−18	Pensky-Martens apparatus	−15 to 16	D 97	0.05%	−3.62
Cloud point (°C)	10.4	−40	Pensky-Martens apparatus	Max 18	D 95	0.05%	2.95
Acid value (mgKOH/g oil)	0.46	0.08	Chemical titration	Max 0.5	D 664	0.05%	-
Calorific value (MJ/Kg)	37.30	42	Isothermal Bomb calorimeter	-	D 240	0.1%	38.48
Saponification value (mg KOH/g fat)	278		Chemical titration	Max 320	D5558	0.05%	197.18
Iodine value (g I^2/100 g oil)	63.4	-	Chemical titration	-	D5554	0.1%	119.17
Cetane number	51.48	40-55	Cooperative fuel research (CFR) test engine model F-5	Min 47	D613	0.1%	47.17

lence of biodiesel. The acid value was determined by means of the method specified in Standard EN ISO 662:2000, and the iodine value was determined by means of the standards defined in Standard EN 14111:2003. Comparing the results of each parameter to their respective standard values shows that the biodiesel produced in this investigation conforms to ASTM specifications. The WCOME properties obtained as per the ASTM standard are given in Table 8. In addition, one column of the table lists the biodiesel (methyl ester) properties of WCO from distinct study Aghbashlo et al. [24]. By comparing these results with the values obtained in this study, it is evident that the qualities of the biodiesel generated are superior to that of the prior study in terms of density, Iodine value, and Cetane number.

CONCLUSION

Developing predictive models for biodiesel yield prediction based on controlled factors is a feasible objective that can significantly impact environmental sustainability. Predictive machine learning algorithms provide a robust, flexible, and transparent method to predict the methyl-ester yield from waste cooking oil. In this article, we modeled biodiesel yield as a function of molar ratio, catalyst concentration, reaction temperature, and reaction time. To predict biodiesel yield, we utilized gradient-boosting regression (GBR), LGBM, and XGBoost regression models. Additionally, we demonstrated how to use a GA-based optimization technique to maximize biodiesel yield (%) using the RSM-developed regression model. The proposed model's performance was evaluated using prediction accuracy, RMSE, MAE, MAPE, and R^2 performance assessment measures.

- After analyzing the R^2 results, the LGBM model demonstrated the most accurate prediction with a score of 0.94048, surpassing XGBoost, which achieved a score of 0.86479, and gradient boosting, which had a score of 0.85831. Additionally, the use of 10-fold cross-validation resulted in the highest degree of precision across all three algorithms. Furthermore, the MAE, RMSE, and MAPE metrics confirm the accuracy of the LGBM prediction, with an MAE of 7.722, RMSE of 8.929, and MAPE of 13.60, which are superior to those of XGBoost and gradient boosting. These findings suggest that the LGBM model possesses excellent predictive capabilities and can assist in making informed decisions regarding biodiesel yield prediction. In terms of computational efficiency, the LGBM model exhibited the quickest performance with a time of 163 milliseconds, followed by XGB regression at 166 milliseconds and Gradient Boosting at 283 milliseconds.
- For the GA approach, the most optimal conditions were found to be a molar ratio of 7.62 : 1, a catalyst concentration of 0.50 wt%, a reaction temperature of 65 °C, and a reaction time of 105 min. The fitness function values in GA offer a quick optimal solution based on the set of experimental results, and the maximum yield obtained under optimal conditions was 99.98%.

NOMENCLATURE

ML : machine learning
GBR : gradient boosting regression
XGB : eXtreme gradient boosting
LGBM : light gradient boosting machine
RSM : response surface methodology
CCD : central composite design
ANN : artificial neural network
GA : genetic algorithm
M.R : molar ratio
C.C : catalyst concentration
R Temp : reaction temperature
R Time : reaction time
KOH : potassium hydroxide
FAME : fatty acid methyl ester
FFA : free fatty acid
PP : pour point
FP : flash point
MSE : mean square error
MAE : mean absolute error
RMSE : root mean square error
MAPE : mean absolute percentage error
R^2 : coefficient of determination

FUTURE SCOPE

In light of the ongoing developments in the biofuel research domain, it is the authors' firm belief that the proposed model would enable users to generate more accurate predictions by fitting accessible datasets. This would prove to be a valuable prerequisite for training machine-learning-based models. Nevertheless, future studies will incorporate additional ML-based prediction and regression models to enhance the performance of the proposed model. Furthermore, there is a need to adopt improved data collection and assortment procedures to record, categorize, analyze, and regenerate gathered datasets for more precise outputs. Additionally, the proposed modeling algorithms could be further extended and refined to achieve even higher performance and more precise prediction results. These future research directions are anticipated to lead to significant advancements in the biofuel research field, ultimately contributing to the development of more efficient and sustainable fuel production processes.

Declarations

Ethics approval and consent to participate: Not Applicable

Consent for publication: Yes

Availability of data and materials: Not Applicable

Conflict of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding: No funding was received for conducting this study

REFERENCES

1. J. M. Cunha, A. S. Faria, T. Soares, Z. Mourão and J. Nereu, *Clean. Energy Syst.*, **1**, 100005 (2022).
2. J. C. Oraegbunam, N. B. Ishola, B. A. Sotunde, L. M. Latinwo and E. Betiku, *Green Technol. Sustainability*, **1**(1), 100007 (2023).

3. D. K. Jana, S. Bhattacharjee, S. Roy, P. Dostál and B. Bej, *Clean. Energy Syst.*, **3**, 100033 (2022).
4. N. N. Mahamuni and Y. G. Adewuyi, *Energy Fuels*, **24**, 3 (2010).
5. E. Betiku, O. R. Omilakin, S. O. Ajala, A. A. Okeleye, A. E. Taiwo and B. O. Solomon, *Energy*, **72**, 266 (2014).
6. V. G. Deshmane and Y. G. Adewuyi, *Fuel*, **107**, 474 (2013).
7. A. Sharma, P. Kodgire and S. S. Kachhwaha, *J. Clean. Prod.*, **259**, 120982 (2020).
8. A. Attari, A. Abbaszadeh-Mayvan and A. Taghizadeh-Alisaraie, *Biomass Bioenergy*, **158**, 106357 (2022).
9. G. R. Moradi, S. Dehghani, F. Khosravian and A. Arjmandzadeh, *Renew. Energy*, **50**, 915 (2013).
10. D. K. Jana, S. Roy, P. Dey and B. Bej, *Clean. Chem. Eng.*, **2**, 100010 (2022).
11. D. N. Thoai, C. Tongurai, K. Prasertsit and A. Kumar, *Int. J. Appl. Eng. Res.*, **13**(10), 7529 (2018).
12. A. Sarve, S. S. Sonawane and M. N. Varma, *Ultrason. Sonochem.*, **26**, 218 (2015).
13. M. A. Ahmadi, *Math. Probl. Eng.*, **2015**, 1 (2015).
14. S. R. Moosavi, D. A. Wood, M. A. Ahmadi and A. Choubineh, *Nat. Resour. Res.*, **28**, 1619 (2019).
15. M. Ahmadi and Z. Chen, *J. Pet. Explor. Prod. Technol.*, **10**, 2873 (2020).
16. D. De Clercq, Z. Wen, F. Fei, L. Caicedo, K. Yuan and R. Shang, *Sci. Total Environ.*, **712**, 134574 (2020).
17. K. K. Gupta, K. Kalita, R. K. Ghadai, M. Ramachandran and X. Z. Gao, *Energies*, **14**(4), 1122 (2021).
18. H. Moayedi, B. Aghel, L. K. Foong and D. T. Bui, *Fuel*, **262**, 116498 (2020).
19. Y. Li, Y. Huang and F. Gao, *J. Clean. Prod.*, **291**, 125973 (2021).
20. M. Wang, S. Fan, Y. Cheng, J. Wang and J. Wang, *Energy Conv. Manag.*, **249**, 114601 (2021).
21. Y. Shen, Z. Li, H. Li and X. Li, *Fuel*, **278**, 118344 (2020).
22. C. R. Khudsange and K. L. Wasewar, *Int. J. Chem. Reactor Eng.*, **15**(3) (2017).
23. M. Aghbashlo, W. Peng, M. Tabatabaei, S. A. Kalogirou, S. Soltanian, H. Hosseinzadeh-Bandbafha, O. Mahian and S. S. Lam, *Prog. Energy Combust. Sci.*, **85**, 100904 (2021).
24. M. Aghbashlo, S. Hosseinpour, M. Tabatabaei and M. M. Soufiyan, *Fuel*, **235**, 100 (2019).
25. M. A. Ahmadi, R. Soleimani, M. Lee, T. Kashiwao and A. Bahadori, *Petroleum*, **1**(2), 118 (2015).
26. M. Ramezanizadeh, M. A. Ahmadi, M. H. Ahmadi and M. Alhuyi Nazari, *J. Therm. Anal. Calorim.*, **137**, 307 (2019).
27. T. A. Khan, A. K. Tasmeem and K. Y. Ashok, *Environ. Sci. Pollut. Res.*, **29**(32), 49465 (2022).
28. B. Sajjadi, M. Davoody, A. R. Abdul Aziz and S. Ibrahim, *Chem. Eng. Commun.*, **204**(3), 365 (2017).
29. C. Li, Q. Li, Y. Li and Y. Wang, *Energy Conv. Manag.*, **214**, 173 (2020).
30. Z. Xia, X. Xiong, Z. Yu and C. Xu, *J. Clean. Prod.*, **244**, 118856 (2020).
31. M. C. Chiu, C. Y. Wen, H. W. Hsu and W. C. Wang, *Sust. Energy Technol. Assessments*, **52**, 102223 (2022).
32. S. Chi, S. J. Suk, Y. Kang and S. P. Mulva, *Adv. Eng. Informatics*, **26**(3), 574 (2012).
33. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg and J. Vanderplas, *J. Machine Learning Res.*, **12**, 2825 (2011).
34. N. Bagalkot, A. Keprate and R. Orderløkken, *Vibration*, **4**(1), 248 (2021).
35. M. Uluskan and M. G. Karşı, International Journal of Lean Six Sigma (2022).
36. R. Singh, A. Kumar, S. Kumar and A. Kumar, *J. Renew. Energy*, **134**, 85 (2019).
37. Y. Li, Z. Zhang, H. Xu and Y. Qian, *Energies*, **13**(4), 860 (2020).
38. A. Shukla, A. K. Singh and S. Singh, *Energy Sources, Part A: Recovery, Util., Environ. Eff.*, **44**(2), 234 (2022).
39. S. Ramraj, N. Uzir, R. Sunil and S. Banerjee, *Int. J. Control Theory Appl.*, **9**, 40 (2016).
40. F. Kocyigit, S. I. Kirbaslar, E. Yavuz and M. Kocaarslan, *Energy Sources, Part A: Recovery, Util, Environ. Eff.*, **42**(16), 1947 (2020).
41. D. Singh, A. Kumar and V. Sachdeva, *Int. J. Renew. Energy Res. (IJRER)*, **10**(3), 1468 (2020).
42. S. Wu, Y. Chen, M. Wang and Y. Huang, *J. Clean. Prod.*, **312**, 127774 (2021).
43. A. Ahmad, A. K. Yadav, A. Singh and A. Pal, *Proc. Inst. Mech. Eng., Part E: J. Process Mech. Eng.*, 09544089231159832 (2023).
44. B. O. Ighose, I. A. Adeleke, M. Damos, H. A. Junaid, K. E. Okpalaeke and E. Betiku, *Energy Conv. Manag.*, **132**, 231 (2017).
45. T. Bhaskar, T. Balusamy and R. Karthikeyan, *Fuel*, **179**, 259 (2016).
46. M. A. Kalam, H. H. Masjuki, N. W. M. Zulkifli, A. Alabdulkarem and Y. H. Teoh, *J. Clean. Prod.*, **172**, 3552 (2018).
47. D. O. Onukwuli, C. Esonye, A. U. Ofoefule and R. Eyisi, *J. Taiwan Inst. Chem. Eng.*, **125**, 153 (2021).