Arabian Journal for Science and Engineering (2025) 50:19421–19441
https://doi.org/10.1007/s13369-025-10227-5

RESEARCH ARTICLE-COMPUTER ENGINEERING AND COMPUTER SCIENCE

A New Approach for Improving Biodiesel Conversion Efficiency: A Stacking Ensemble Model Based on Linear Regression Approach with GAN-Enhanced

Ahmet Karaoğlu[1] · Hüseyin Söyler[2]

Received: 31 December 2024 / Accepted: 14 April 2025 / Published online: 13 May 2025

Abstract

This study employs a Linear Regression-based stacking ensemble learning approach as a novel method to enhance biodiesel conversion efficiency. Initially, a dataset derived from the literature was used to train an ensemble model that combines predictions from Random Forest, XGBoost, and Deep Neural Network (DNN) through a Linear Regression-based fusion approach. This model outperformed individual models (Random Forest: − 0.16, XGBoost: − 0.67, and DNN: 0.36) by achieving an R^2 score of 0.45. To further improve model performance, 4900 synthetic data samples were generated and integrated into the dataset. Leveraging the stacking ensemble learning approach with this expanded dataset, the model demonstrated a significant improvement in predictive accuracy, achieving an R^2 score of 0.81. This corresponds to an approximate 4% increase in performance compared to individual models (Random Forest: 0.78, XGBoost: 0.78, and DNN: 0.77), highlighting the effectiveness of ensemble learning in optimizing biodiesel conversion efficiency. Additionally, the model exhibited high accuracy with low error rates (MAE: 1.16 and MAPE: 1.24%), effectively compensating for the weaknesses of individual models and providing more stable and generalized predictions. To the best of our knowledge, this is the first study to incorporate a Linear Regression-based stacking method to enhance biodiesel conversion efficiency. These findings underscore the potential of ensemble learning techniques and synthetic data integration in improving renewable fuel efficiency. Future research can further enhance model performance by incorporating larger datasets and exploring more advanced ensemble strategies.

Keywords Biodiesel conversion · Stacking ensemble · Linear regression · Synthetic data · Machine learning

1 Introduction

Alternative fuels have gained importance due to depleting fossil fuel sources and environmental concerns [1]. These fuels, including hydrogen, LPG, CNG, biodiesel, and synthetic fuels, offer potential solutions for various applications such as internal combustion engines, aviation, and power generation [2–4]. The development of alternative fuels has been influenced by public policies and economic factors [5]. However, challenges remain in terms of production, infrastructure, and technological readiness [4]. The selection of optimal alternative fuels requires consideration of multiple criteria, including environmental impact, energy efficiency, and economic factors [6]. Ongoing research focuses on characterizing and improving the performance of these fuels to support the transition to cleaner energy systems [7].

Biodiesel is a renewable fuel produced from vegetable oils, animal fats, or recycled cooking oils through transesterification [8, 9]. It can be used pure or blended with conventional diesel in unmodified engines [10]. Biodiesel offers several advantages, including improved lubricity, higher cetane number, and reduced emissions of particulate matter, carbon monoxide, and unburned hydrocarbons [8]. It is biodegradable, non-toxic, and environmentally beneficial [11, 12]. The use of biodiesel can enhance energy security, reduce greenhouse gas emissions, and provide socioeconomic benefits to rural areas [9, 13]. However, challenges such as feedstock competition with food crops and land-use changes need to be addressed [14]. Ongoing research focuses

✉ Hüseyin Söyler
hsoyler@sinop.edu.tr

1 Department of Computer Engineering, Sinop University, 57600 Sinop, Turkey

2 Gerze Vocational School, Sinop University, 57000 Sinop, Turkey

on improving production processes, exploring alternative feedstocks, and optimizing engine performance to further enhance biodiesel's potential as a sustainable fuel alternative [12, 14]. There is a need for optimization in biodiesel production to achieve the highest conversion yield at the lowest cost and in the shortest possible time.

Recent studies have explored the application of machine learning (ML) techniques to optimize biodiesel production processes. Various ML algorithms, including random forest, AdaBoost, and gradient boosting, have demonstrated high accuracy in predicting biodiesel yield and quality [15–17]. These models consider multiple input variables such as reaction time, temperature, catalyst concentration, and methanol-to-oil ratio [18, 19]. Ensemble learning methods, particularly boosting algorithms, have shown superior performance in modeling biodiesel production, with some studies reporting yields of up to 98% [20, 21]. ML approaches have been applied to various aspects of biodiesel production, including process optimization, quality estimation, and real-time monitoring [22]. These techniques offer advantages in handling complex, nonlinear relationships between process parameters and outcomes, potentially reducing the need for extensive physical experiments [15, 16].

Recent studies have explored machine learning approaches to optimize biodiesel and biogas production. Ensemble methods, particularly stacking and boosting, have shown promising results in predicting and maximizing yields [23] developed a stacking ensemble model integrating LightGBM, CatBoost, and evolutionary strategies to accurately predict biogas yield. For biodiesel production, Ahmad et al. combined boosting with polynomial chaos expansion to predict product composition and quantity while accounting for process uncertainties [24]. Almohana et al. compared three boosted ensemble models (Huber Regression, Decision Trees, and Gaussian Process) for optimizing biodiesel production from waste cooking oil, with the boosted Gaussian Process model achieving the highest accuracy [25]. Similarly, Lu and Zhang evaluated AdaBoost Regression Tree, Extra Trees, and Gradient Boosting Regression Tree for predicting biodiesel yield from palm oil, with Gradient Boosting Regression Tree demonstrating superior performance [21]. These studies highlight the effectiveness of ensemble learning techniques in improving prediction accuracy and optimizing biofuel production processes.

In the field of biodiesel production optimization, existing literature largely relies on traditional methods that require extensive physical experimentation and manual optimization of processes. While machine learning-based optimization studies exist, they primarily focus on individual models, often overlooking the potential improvements offered by combining these models into ensemble approaches.

This study addresses this gap by employing the Voting Ensemble method to integrate the strengths of individual models, resulting in more balanced, reliable, and interpretable predictions. The main contributions of this approach in filling the research gap are as follows:

1. Mitigating Individual Model Limitations: Individual models often excel under specific conditions but may underperform in others. Ensemble Model based on Linear Regression reduces these inconsistencies by combining the strengths of different models, creating a more robust predictive framework.
2. Improved Overall Performance: The Ensemble Model based on Linear Regression method in this study increased the average R^2 score of individual models by 7.35%, demonstrating superior accuracy and interpretability. This highlights the effectiveness of ensemble methods as a tool for biodiesel production optimization.
3. Enhancing Efficiency: The high accuracy achieved by the Ensemble Model based on Linear Regression method reduces the need for extensive physical experiments, enabling faster and more cost-effective optimization of production processes.
4. Effective Use of Augmented Data: The dataset expanded through Generative Adversarial Networks (GAN) allowed the Ensemble Model based on Linear Regression method to better capture complex and multidimensional relationships, further improving its predictive capabilities.

In conclusion, the Ensemble Model based on Linear Regression method emerges as a solution to the shortcomings of individual models and a powerful approach for optimizing biodiesel production processes. By demonstrating the applicability and advantages of ensemble methods, this study provides a foundation for future research and wider adoption of advanced machine learning techniques in biodiesel production.

1.1 Literature Review

The production of biodiesel from various oils is influenced by several key factors. Methanol-to-oil molar ratio significantly affects biodiesel yield, with optimal ratios ranging from 3:1 to 16:1 [26, 27]. Catalyst type and concentration play crucial roles, with both homogeneous (NaOH, KOH) and heterogeneous catalysts being effective [28]. Reaction temperature typically ranges from 55 to 65 °C, while optimal reaction times vary from 2 to 2.5 h [26, 29, 30]. Stirring speed also impacts biodiesel yield, with speeds ranging from 250 to 700 rpm [29]. Ultrasonic cavitation has shown faster reaction rates compared to mechanical stirring [30]. These factors collectively influence biodiesel yield, which can reach up to 98% under optimized conditions [26]. The resulting biodiesel often meets international standards for fuel quality [31, 32].

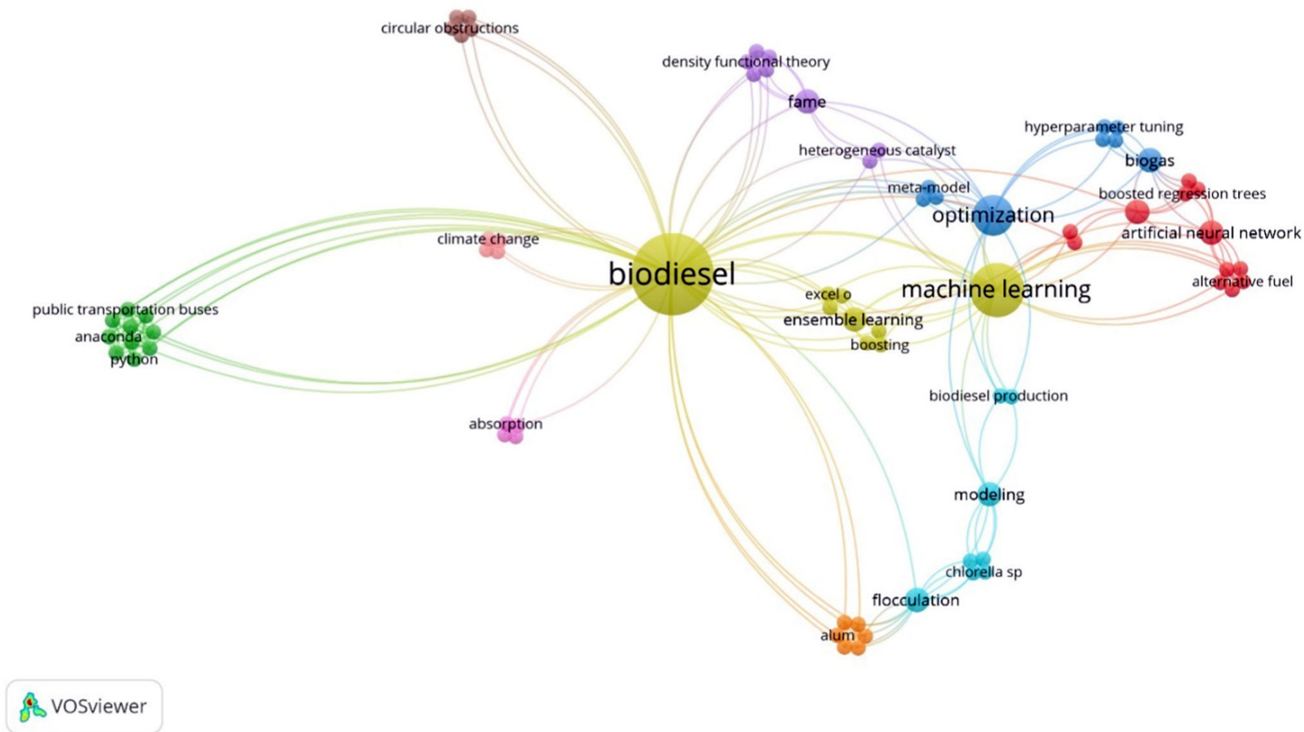


Fig. 1 Network visualization of key research topics related to biodiesel

In conclusion, achieving the highest conversion yield in biodiesel production depends on numerous factors. Hundreds of experiments may be required to achieve the highest yield using classical methods. By combining machine learning models with voting ensemble learning models, it is aimed to minimize this number. Figure 1 shows a network visualization of the main research topics related to biodiesel.

Figure 1 shows the network visualization of key research topics related to biodiesel production. The interconnected clusters highlight the multidisciplinary nature of the field, with significant associations to machine learning, optimization, and modeling processes. Concepts such as artificial neural networks, ensemble learning, and boosted regression trees are prominent in the context of machine learning applications. Additionally, terms such as biogas, climate change, and public transportation emphasize the broader environmental and practical implications of biodiesel research. This network analysis demonstrates the integration of advanced computational methods and sustainability concerns, reflecting the evolving trends and priorities in biodiesel studies.

For this study, 250 articles published in Elsevier in 2024 were reviewed. Only publications that produced biodiesel using the transesterification method were selected. From these, 100 articles were chosen where the data on the type of oil used, type and amount of catalyst, methanol-to-oil molar ratio, reaction temperature, reaction time, and conversion yield were complete. The dataset for this study consists of these 72 selected articles. The articles used as the dataset in the study are presented in Table 1.

2 Methodology

This study presents an integrated approach combining data augmentation and machine learning methods to optimize biodiesel production processes and analyze the relationship between conversion efficiency and production parameters. The methodology involves the use of Generative Adversarial Network (GAN)-based data augmentation and Ensemble Model based on Linear Regression method. GAN increases data diversity by generating synthetic data for limited and small datasets, while Ensemble Model based on Linear Regression method aims to achieve more balanced and accurate predictions by combining the strengths of different models. Within the scope of the study, important parameters used in biodiesel production processes (e.g., methanol-to-oil molar ratio, catalyst ratio, reaction time, and reaction temperature) and conversion efficiency as a target variable are considered. The detailed analysis of these processes and the preparation of the dataset used formed the basis for training

Table 1 Sources used in the creation of the dataset

Raw material	Catalyst name	Methanol molar ratio	Catalyst radio	Reaction time	Reaction temperature	Conversion efficiency	Refs
Soybean oil	KOH	6	0.5	180	65	97.5	[33]
Waste cooking oil	KOH	6.9	1	20	30	97.76	[34]
Soybean oil	GL-850	17	6	180	25	95.8	[35]
Palm oil	Cocoa pod husk	15	7	20	60	96.2	[36]
Waste animal fats	Solid carbon catalyst	12	3	150	70	94.4	[37]
Spent cooking oil	Geopolymer catalyst	9	2	120	50	99.2	[38]
Microalgae	Biochar	12	5	240	50	94.9	[39]
Waste biomass	Bio-nano-CaO	10	3	180	55	96.8	[40]
Dairy sludge	CaO	9	10	120	70	74	[41]
Jatropha–Karanja blend	KOH	8	1	120	70	85.26	[42]
Ulva lactuca	Cu-BTC@AC	15	3.96	140	65	92.56	[43]
Microalgae	NaOH	9	2	120	50	93.5	[44]
Sour cherry kernel oil	CaO	6	6	360	56.3	97.3	[45]
Nannochloropsis gaditana	CH_3ONa	1:12	1.5	75	50	87.25	[46]
Soybean oil	ZnO/TiO_2	8	20	60	55	90	[47]
Brassica carinata tohumu yağı	CaO	14.85	13.77	180	65	90	[48]
Azolla filiculoides	NaOH	12	1.6	145	55	89.07	[49]
Ulva lactuca	NaOH	3	1.5	61	55	86.17	
WFO palm	NaOH	12	1	185	69	99.15	
WFO sunflower	NaOH	9	1.2	230	73	98.95	
WFO corn	NaOH	4	3	236	73	99.18	
Castor oil	TiO_2@CaO	12	5	135	60	94.27	[50]
Carthamus tinctorius L. seeds	B_2O_3	6	0.69	100	75	94	[51]
Castor oil	CLPA@Fe_3O_4	30 wt	6	180	60	97.92	[52]
Mealworm oil (Tenebrio molitor)	Sodium phosphate (Na_3PO_4)	16	3.2	114	65	96.98	[53]
Jatropha curcas oil	CaO (Calcium oxide)	18	4	330	60	96	[54]

Table 1 (continued)

Raw material	Catalyst name	Methanol molar ratio	Catalyst radio	Reaction time	Reaction temperature	Conversion efficiency	Refs
Soybean oil	$NaAlO_2/CuFe_2O_4$	13	8	60	95	95.9	[55]
Melia azedarach fruits	LAC-NH2 catalyst	6	2	180	60	98	[56]
Waste cooking oil	$CaO/MgO/Fe_3O_4$	9	1	360	60	95.64	[56]
Waste cooking oil	$KOH/ZnAl_2O_4$	9	4	90	40	95.3	[57]
Castor oil	Red mud-based potassium composite (KRM)	18	5	150	65	94.52	[58]
Waste cooking oil (WCO)	NaOH	7	1	90	65	64.9	[59]
Soybean oil	KOH on activated carbon (C1)	12	10	120	65	91.8	[60]
Soybean oil	Sodium aluminate	28	3	60	65	95	[61]
Soybean oil	Psidium guajava L. leaves	17	6	180	25	95.8	[35]
Waste cooking oil	CoFe/biochar/CaO	16	3	157	73	99.21	[62]
Paper-pulp industrial wastewater	NaOH	1.82	4	384	65	92.10	[63]
Microalgal biomass	Lewis acidic deep eutectic solvent (LADES)	5	5	90	120	39.86	[64]
Waste catfish oil (CFW)	Fish bone	6	1	30	110	96.85	[65]
Oil blend (soybean. jatropha. pongamia oils)	Areca nut leaf ash-K_2CO_3	9	15	201	65	96.57	[66]
Triolein and methanol	NaOH and H_2SO_4	12	1.5	75	50	98	[67]
Waste frying oil (WFO)	MoO_3/RHA-$CoFe_2O_4$	20	6	180	160	94.6	[68]
Various agricultural residues	Biochar and activated carbon	6	5	180	60	92.4	[69]
Waste cooking oil (WCO)	(Ca@CS)	6	25	360	60	100	[70]
Waste frying oil	K_2CO_3/Orange peel-derived hydrochar	6.6	4.4	108	50	89.2	[71]
Castor oil	Swan eggshell-derived CaO	12	5	90	60	95.64	[72]
Green seed canola oil	12-tungstophosphoric heteropoly acid	22	5	300	60	88	[73]
Soybean oil	CaO@PC-900	12	10	240	60	98.6	[74]
Castor oil	Li-doped salmon fish bone-supported La_2O_3	6	5	90	65	99.38	[75]
Palm Oil	Sargassum plagiophyllum-derived heterogeneous catalyst	6	3	120	60	93	[76]
Sunflower Oil. Jatropha curcas Oil. Waste Cooking Oil	HCP-SB-K (Hypercrosslinked Polymer)	6	3	60	60	99.9	[77]

Table 1 (continued)

Raw material	Catalyst name	Methanol molar ratio	Catalyst radio	Reaction time	Reaction temperature	Conversion efficiency	Refs
Soybean Oil	Magnetic biochar from reed straw and electric furnace dust	15	7	240	74.15	99.89	[78]
Soybean Oil	MGO@SnO (Magnetic graphene oxide supported tin oxide)	10	5	180	120	96.5	[79]
Dunaliella tertiolecta	NaOH	12	1	150	146	80.22	[80]
Waste cooking oil	NaOH	12	1.5	30	59.4	82.8	[81]
Rapeseed oil	NiW/rGO	12	3	240	55	91	[82]
Non-edible oil seeds	CaO	10	1.5	120	70	94.5	[83]
Agro waste oilseeds	Nano catalyst	5.99	3.96	72.42		99.06	[84]
Quercus ballot	ZnO nanocatalyst	6	0.5	80	65	98.06	[85]
Waste oil	NaOH	6	0.5	60	55	96	[86]
Waste cooking oil	$ZnFe_2O_4$/CaOPS	12.96	5.06	212.22	63.69	96.89	[87]
Waste olive oil	Sulfonated carbon (SC)	15	4	300	60	94	[88]
Palm oil	La_2O_3-Al_2O_3	30	5	300	200	93.12	[89, 90]
Waste cooking oil	CuO/UiO-66	7	0.5	180	65	90.1	[90]
African pear seed oil	Activated palm fruit bunch	10.02	3.18	190.2	62.21	90.1	[91]
Waste cooking oil	NaOH	9.1	3.6	60	40	87.1	[92]
Low-quality non-edible feedstocks	CaO catalyst	6.1	2	300	65	95.9	[93]
Used cooking oil	CaO/Fe_3O_4	5	0.5	120	60	91	[94]
Waste cooking soybean oil	Mo/Silicalite-1	12	5	120	65	80.73	[95]
Acer truncatum Bunge seed oil	Alkaline ionic liquid	12	1.3	137	71	97.3	[96]
Soybean oil	CaO	6	0.5	120	60	96.1	[97]
Opium poppy seed oil	KOH	3	0.25	90	60	94.87	[98]
Almond oil	Cs-TPA/Ce-KIT-6	12	4	60	50	98	[98]
Chicken gizzard oil	KOH	6	0.01	300	60	94	[99]
Soybean oil	Snail shells-derived mixed oxide catalyst	12	2	90	65	93	[100]

Table 2 Description and hyperparameters of the regression models used

Model	Description	Hyperparameters
Random forest regressor (RF)	Regression performed using a random forest algorithm	n_estimators: 100, random_state: 42
XGBoost regressor (XGB)	Regression performed using an XGBoost model	n_estimators: 100, random_state: 42
Deep neural network (DNN)	Deep neural network model with multiple layers	Dense(64, relu) BatchNormalization Dropout(0.2) Dense(32, relu) Dense(1)
Ensemble model (Linear regression)	Linear regression model combining RF, XGB, and DNN predictions	No hyperparameters, standard linear regression used

and evaluating the model. The description and hyperparameters of the regression models used are given in Table 2.

2.1 Data Augmentation Process

The GAN (Generative Adversarial Network) model is based on a competitive learning process between its two core components: the generator and the discriminator. The process begins with the generation of random input vectors in the Latent Space (Random Noise). These random inputs are processed by the generator to produce Generated Data (Synthetic Samples), which are fake data instances. The generator strives to create synthetic data that share the same statistical properties as real data.

The GAN (Generative Adversarial Network)-based data augmentation process is realized by integrating synthetic data samples generated with random input with real data. As shown in Fig. 2, the feedback loop between the generator and the discriminator in this process increased the similarity of the synthetic data to the real data, and an augmented dataset was created by integrating these data. This augmented dataset was used to train and evaluate machine learning models to improve model performance.

The generated synthetic data are then evaluated by the discriminator alongside real data. The discriminator's task is to distinguish whether the input data are real data or synthetic data produced by the generator. Feedback from the discriminator is sent back to the generator, enabling it to improve and produce more realistic synthetic data. This iterative optimization process continues between the generator and the discriminator.

As shown in Fig. 3, the architectures of the generator and discriminative networks of the GAN model are detailed. The generator network is designed using dense layers starting with an input layer and including LeakyReLU activation functions to generate synthetic samples from random inputs. The discriminator network has a similar dense layer structure to distinguish between synthetic and real data. In addition, the model is optimized with an integration network that includes a two-layer sequential structure. This architecture is organized in such a way that data generation and discrimination are performed efficiently.

The synthetic data produced by the GAN and the real data are combined in the Data Integration + Model Training step. This combination results in an augmented and enriched dataset. The enhanced dataset is then used for training and evaluating machine learning models. Finally, in the Enhanced Dataset + Model Evaluation step, the improved dataset and analysis outcomes are obtained.

This process is particularly effective in scenarios with small and limited datasets, as it helps expand the dataset and improve model performance. The synthetic data generated by the GAN increases data diversity by closely resembling real data, thereby enhancing the generalization capacity of the model.

2.2 Data Preparation

The GAN-augmented dataset included biodiesel production parameters (methanol_ratio, catalyst_ratio, reaction_time, and reaction_temperature) and the target variable conversion_efficiency. Missing values were imputed with columnwise means, and the entire dataset was normalized to the [0, 1] range using MinMaxScaler. The dataset was randomly split into training (80%) and testing (20%) subsets. As shown in Fig. 4, the correlation matrix reveals the relationships between the parameters in the biodiesel production process.

Figure 5a shows a multimodal distribution of the methanol/oil molar ratio, with peaks at 6 and 16, indicating that these are commonly used ratios. Figure 5b highlights the proportion of catalysts, with frequent values at 1, 3, and 4, while higher ratios are less common. Figure 5c depicts response times, showing a right-skewed distribution with a peak at 180, concentrated between 60 and 180. Figure 5d illustrates reaction temperatures, with a bimodal structure peaking between 60 and 65 and declining toward 70. Figure 5e shows conversion efficiency, with a peak around 96 and concentration between 88 and 100, suggesting high overall efficiency. The box plot for the methanol/oil molar ratio confirms a balanced distribution, with most data between 8 and 14, a median of 11, and no outliers. These distributions highlight key patterns and preferences in biodiesel production processes. In particular, the positive correlation between methanol/oil molar ratio and reaction temperature (46%) indicates that controlling the methanol ratio can positively affect the reaction temperature. Furthermore, the

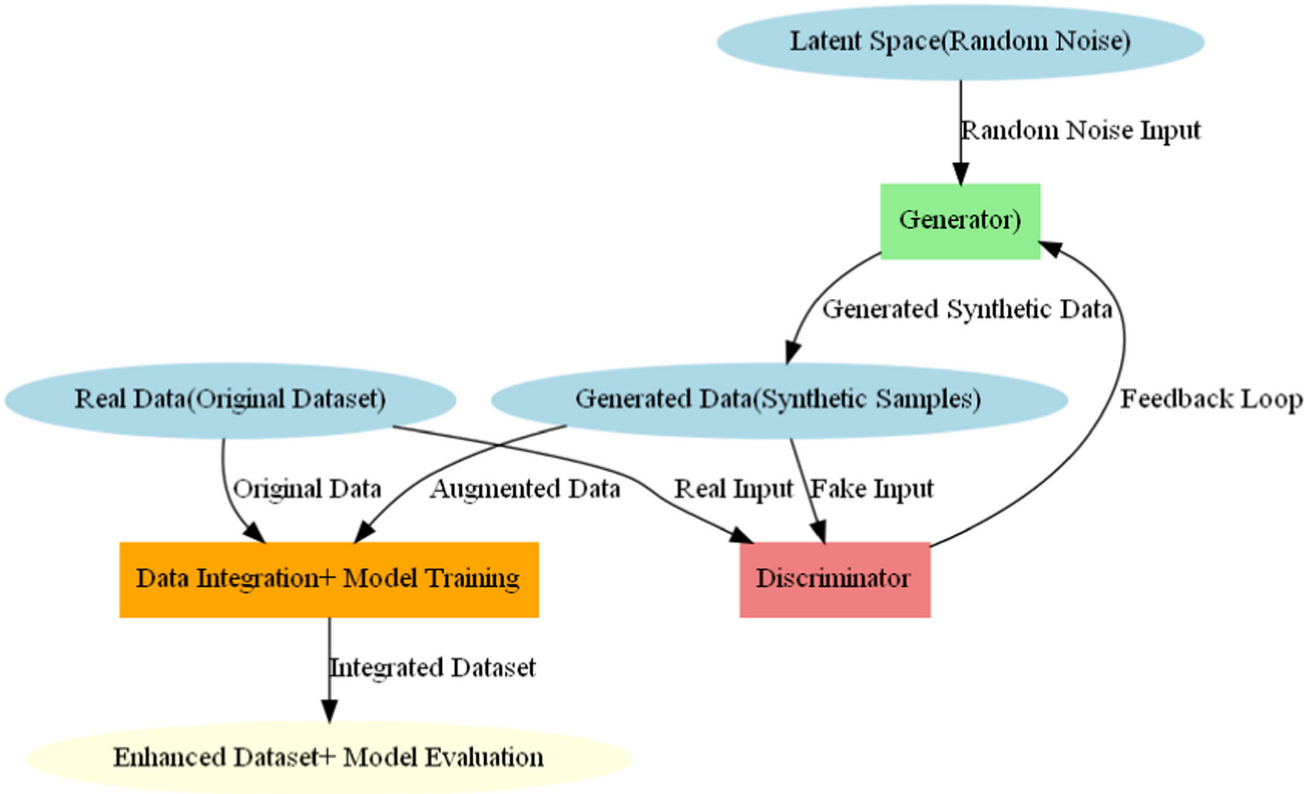


Fig. 2 GAN (Generative Adversarial Network)-based data augmentation process

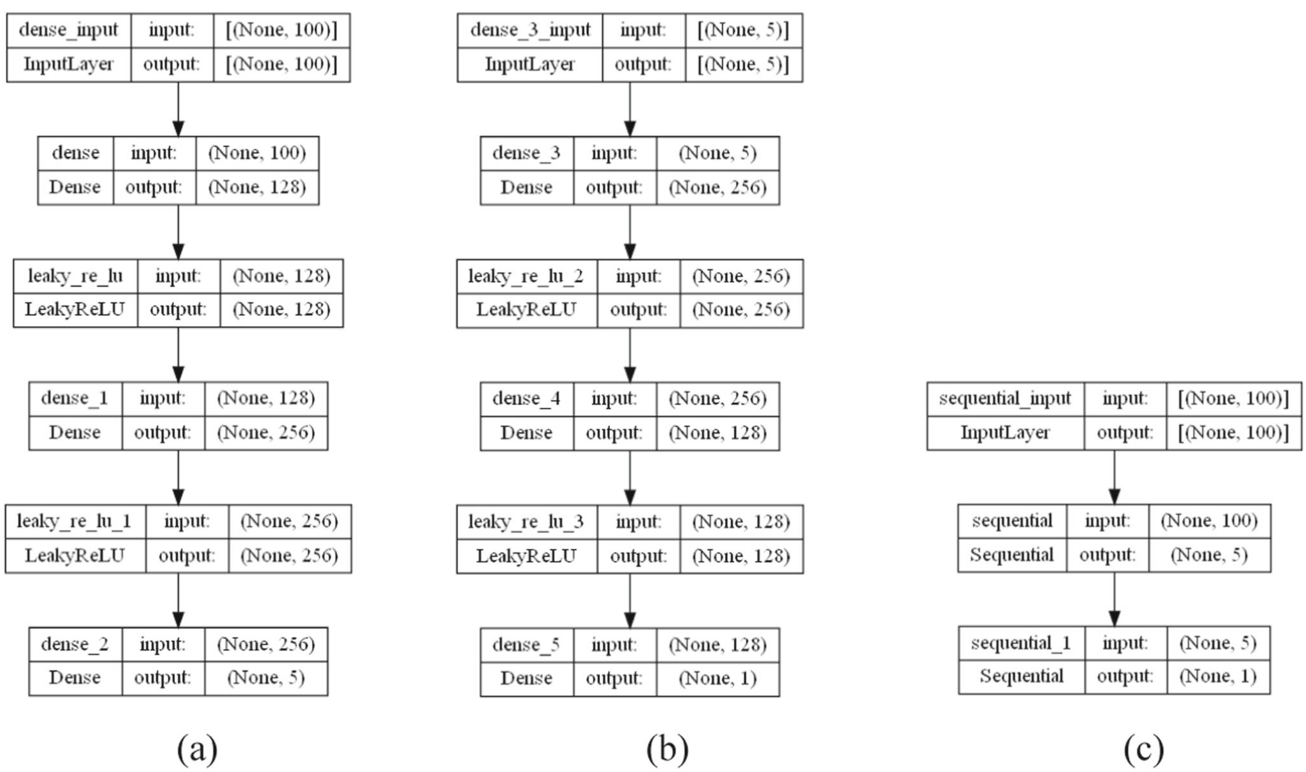


Fig. 3 Generative Adversarial Network (GAN) architecture. **a** The generator network. **b** The discriminator network. **c** The combined model

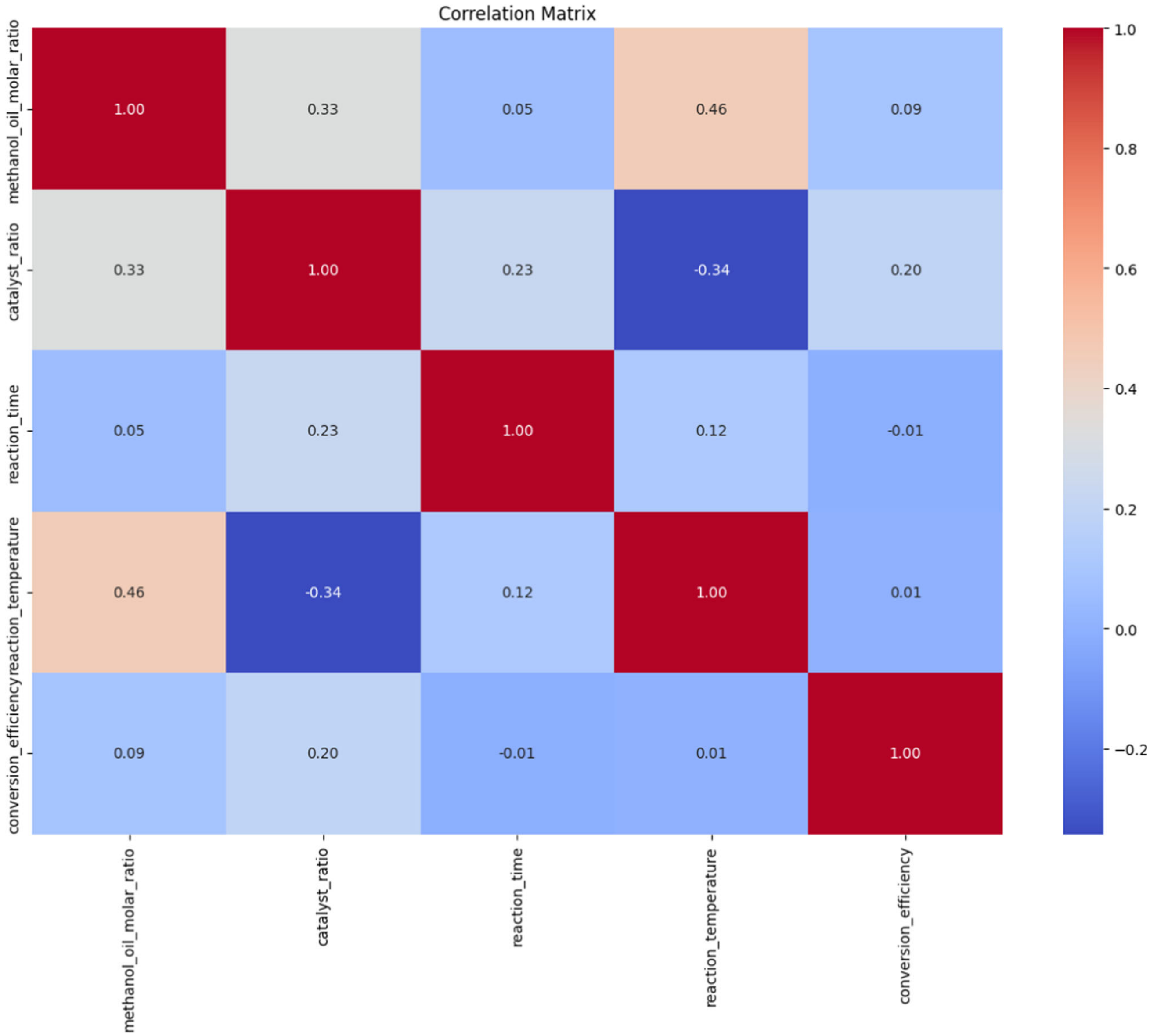


Fig. 4 Correlation matrix showing the relationships between parameters in the biodiesel production process

relationships between catalyst ratio and conversion efficiency provide opportunities for optimization at various stages of the process. These findings suggest that effective management of independent variables in biodiesel production can improve process efficiency. Figure 6 shows the main parameters of biodiesel production in the form of histograms and box plots.

The box plots in Fig. 6a–e provide a summary of key parameters in biodiesel production. Figure 6a shows the distribution of the methanol/oil molar ratio, where most data fall between 8 and 14, with a median of 11 and no outliers. Figure 6b illustrates the catalyst ratio, where values are concentrated between 1 and 4, with a median of 3 and limits at 0 and 7, showing no outliers. In Fig. 6c, the reaction time distribution indicates most data are between 100 and 160, with a median of 120 and boundaries at 60 and 180, again with no outliers.

Figure 6d highlights the reaction temperature, concentrated between 60 and 70 with a median of 65, though one outlier exists below the lower boundary of 55. Finally, Fig. 6e shows the conversion efficiency, where values are tightly clustered between 92 and 98, with a median of 96 and no outliers observed. Overall, the box plots demonstrate that the data for key parameters are generally well-distributed, with minimal outliers, providing a clear representation of trends in biodiesel production.

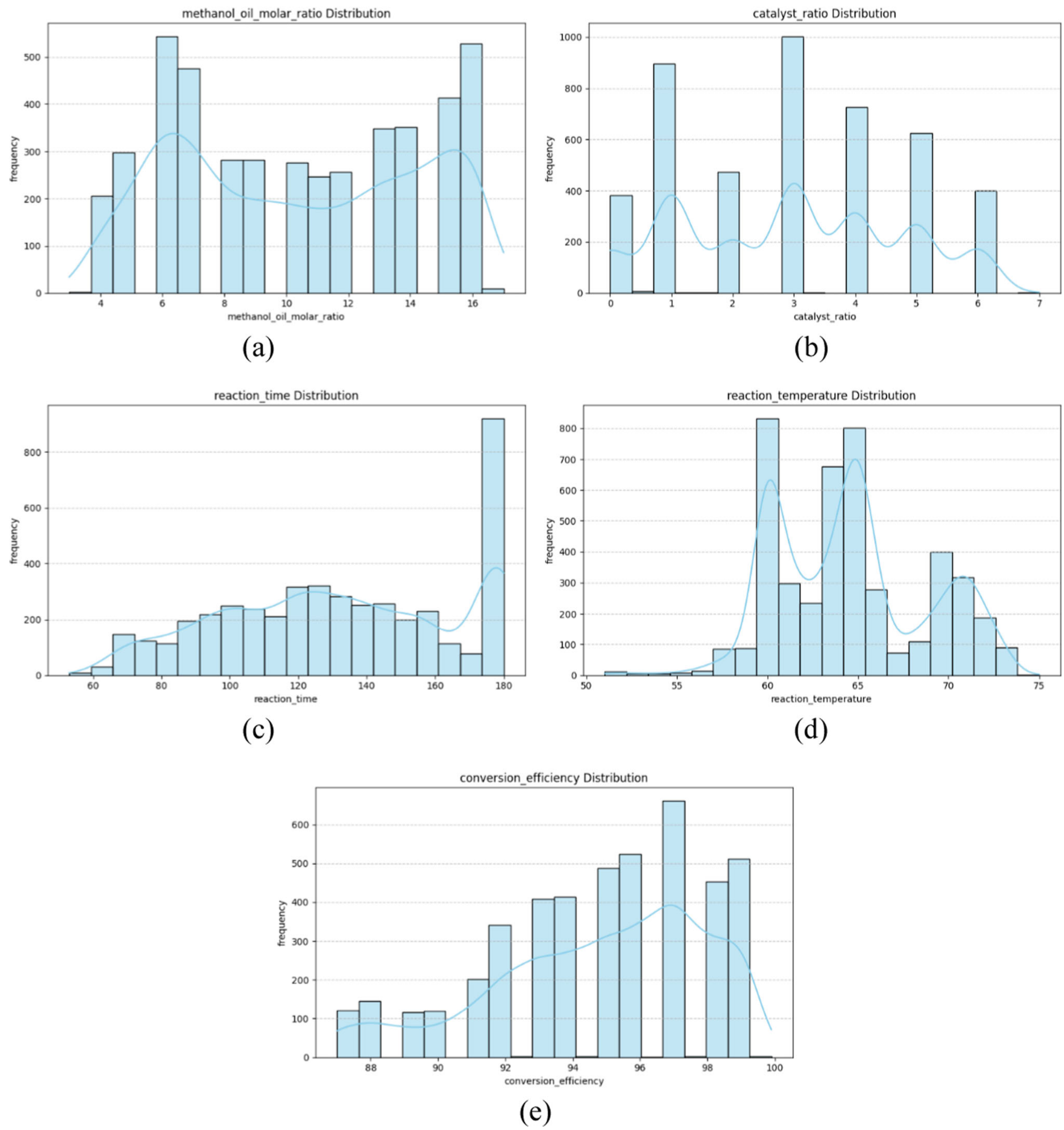


Fig. 5 Histograms with density curves showing the distribution of key parameters in biodiesel production

2.3 Machine Learning Models and Hyperparameter Optimization

Several machine learning models were utilized to predict conversion efficiency, with a focus on optimizing performance through hyperparameter tuning. The Random Forest Regressor was implemented with 100 estimators and a fixed random state to ensure reproducibility while maintaining model stability. Similarly, the XGBoost Regressor was applied with 100 estimators and a fixed random state to enhance predictive performance through boosted decision trees. The Deep Neural Network model was structured with multiple layers, including dense layers with ReLU activation, batch normalization, and dropout regularization to prevent overfitting. To improve predictive accuracy, hyperparameter

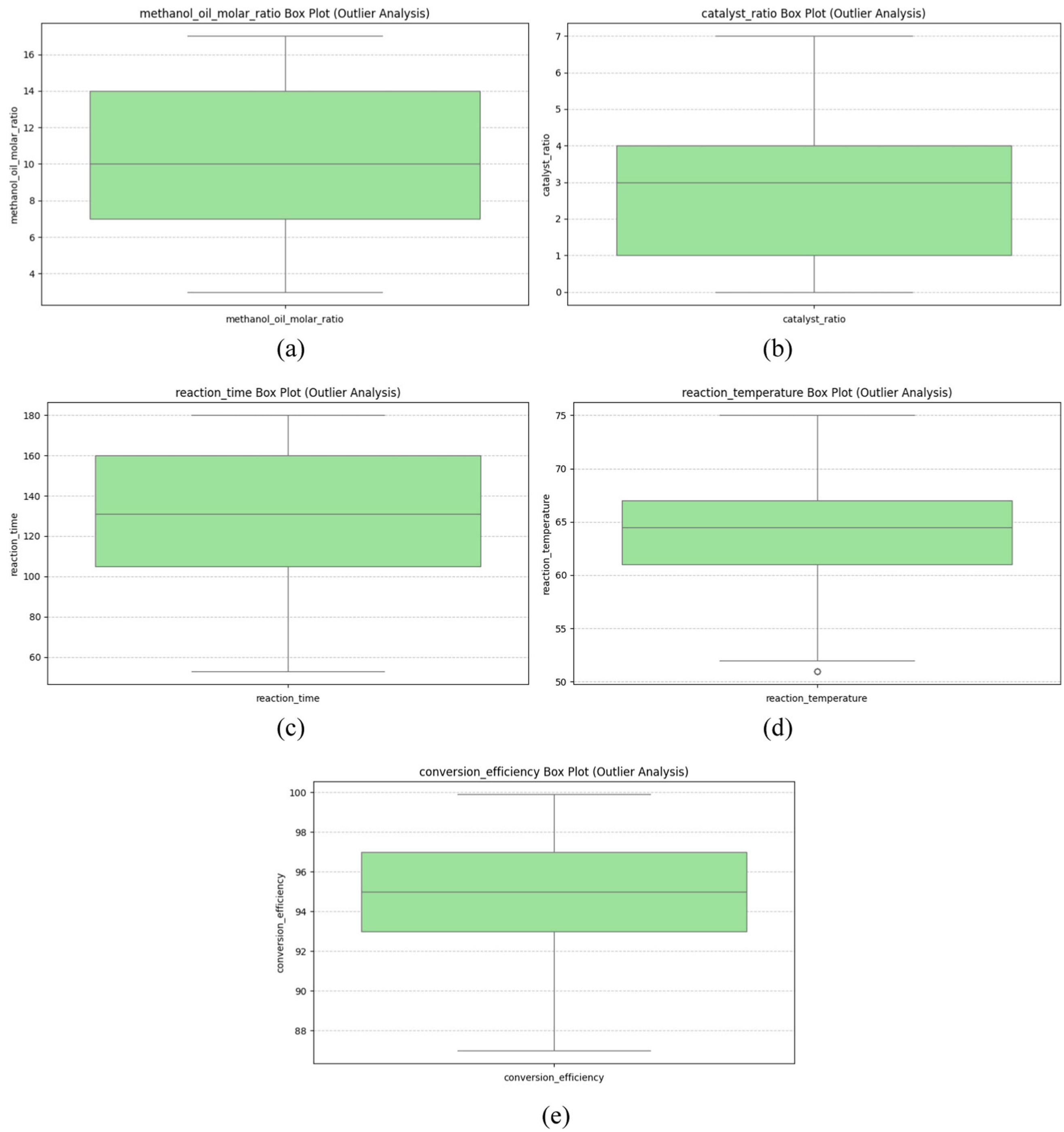


Fig. 6 Box plots of key parameters in biodiesel production

tuning was conducted to identify the most effective parameter configurations for each model.

2.4 Ensemble Model: Linear Regressor

In this study, three different regression models were utilized to analyze the dataset and predict the target variable. The models included Random Forest Regressor, XGBoost Regressor, Deep Neural Network, and an Ensemble Model based on Linear Regression. The Random Forest and XGBoost models were implemented with 100 estimators, and a fixed random state was applied to ensure reproducibility. The Deep Neural Network model was structured with multiple layers, including a 64-unit dense layer with ReLU activation, batch normalization, dropout regularization, a 32-unit dense layer with ReLU activation, and a final output

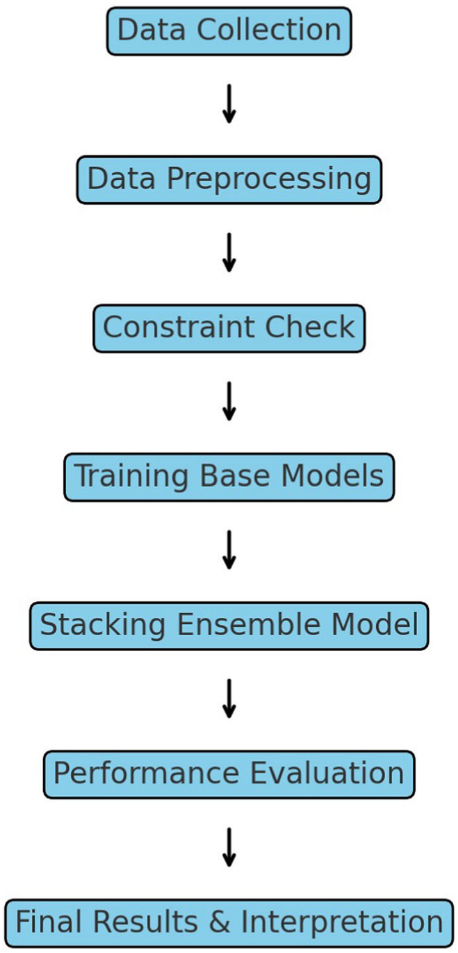


Fig. 7 Constraint handling flowchart for biodiesel model

layer. To enhance predictive performance, an ensemble learning approach was adopted by integrating the predictions of all models through a standard linear regression model. This strategy was designed to leverage the advantages of individual models while reducing their limitations, leading to improved overall prediction accuracy.

To ensure that the biodiesel model adheres to predefined constraints, a structured constraint handling approach was implemented. As illustrated in Fig. 7, the process begins with data collection and data preprocessing, ensuring that raw data is appropriately formatted and cleaned. Subsequently, a constraint check is conducted to verify whether the data and model outputs comply with the predefined feasibility criteria. Following constraint verification, base models are trained and further optimized through a stacking ensemble model to enhance predictive performance. The performance evaluation step assesses the model's accuracy and robustness, ensuring compliance with technical and environmental constraints. Finally, the results are interpreted to derive meaningful insights. This systematic approach improves model reliability and ensures that constraint violations are effectively managed.

In this study, a Stacking Ensemble Learning approach was employed to enhance predictive accuracy by combining multiple machine learning models. The base learners consisted of Random Forest Regressor, XGBoost Regressor, and Deep Neural Network, each trained on the dataset to generate independent predictions. These diverse models were selected to leverage their unique strengths—tree-based models for capturing nonlinear relationships and deep learning for feature extraction. To aggregate the outputs from the base learners, a meta-learning approach was implemented using Linear Regression as the meta-learner. The predictions generated by the base models were used as inputs to the meta-model, which learned to optimize their combination to produce the final prediction. This methodology aligns with the Stacking Ensemble Learning framework, where multiple heterogeneous models contribute to improving overall predictive performance by reducing individual model biases and variances.

Unlike Bagging, which trains multiple instances of the same algorithm on different subsets of the data, or Boosting, which iteratively improves weak learners by assigning higher weights to misclassified instances, Stacking focuses on learning an optimal combination of diverse model outputs through a higher-level learner. By utilizing Linear Regression as the meta-learner, the ensemble model effectively combines predictions in a weighted manner, allowing it to generalize better than any single base learner. This approach ensures that the model capitalizes on the complementary strengths of individual algorithms while mitigating their respective weaknesses. Thus, the Stacking-based ensemble framework used in this study offers an efficient and robust predictive modeling strategy, integrating different machine learning paradigms to enhance accuracy and generalization performance. The proposed stacking ensemble model diagram is shown in Fig. 8.

2.5 Model Training and Evaluation

Each model was trained on the training set and evaluated on the test set using multiple approaches to ensure robust performance assessment. The primary evaluation metric was the R^2 score, calculated as Eq. 1:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{1}$$

where y_i represents the actual values, $\hat{y}_i$ represents the predicted values, and $\overline{y}$ is the mean of the actual target values. The R^2 score provides a measure of how effectively the model explains the variance in the target variable, with values closer to 1 indicating stronger predictive capabilities. To further analyze the prediction accuracy, scatter plots of actual versus

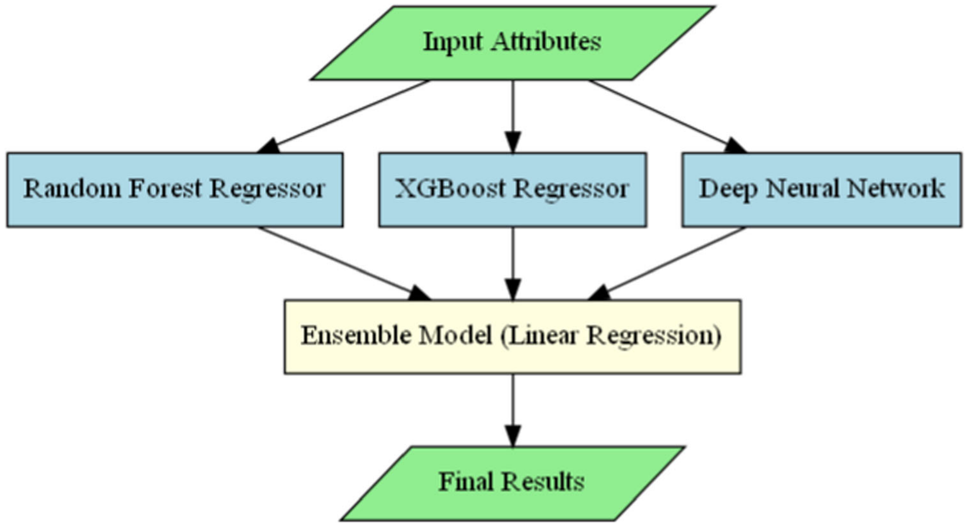


Fig. 8 Ensemble model based on linear regression

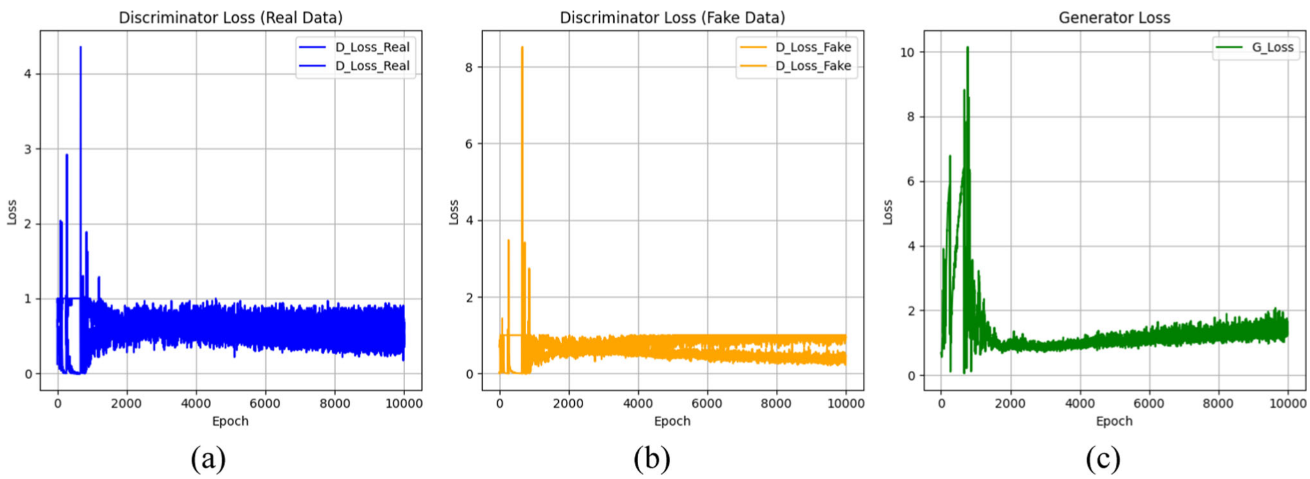


Fig. 9 Loss curves during GAN training

predicted values were generated for each model, accompanied by a reference line representing perfect predictions. Additionally, the R^2 scores for all models were compared and presented in a bar chart, offering a clear visual depiction of performance differences. This comprehensive evaluation approach, combining numerical metrics and visual tools, allowed for an in-depth understanding of each model's effectiveness and reliability in predicting the target variable.

3 Results

This study highlights the successful integration of GAN-based data augmentation and ensemble learning for analyzing and optimizing biodiesel production processes. The proposed approach provides a robust framework for improving the prediction and optimization of conversion efficiency in biodiesel production. To evaluate the effectiveness of the method, the performance of the Voting Ensemble Learning model was compared against individual machine learning models using the R^2 score as the evaluation metric. The models included in the comparison were the Random Forest Regressor, the XGBoost Regressor, Deep Neural Network, and the Ensemble Model based on Linear Regression. This comparative analysis demonstrated the superior performance of the ensemble approach in leveraging the strengths of individual models to achieve more accurate predictions. Figure 9 shows the loss curves during GAN training.

The data augmentation process using GAN significantly enhanced the dataset, leading to improved performance of the machine learning models. The Voting Regressor demonstrated the best overall performance by effectively combining the strengths of individual models. Among the individual models, Random Forest, and Gradient Boosting emerged as top performers, showcasing their robust predictive capabilities. Additionally, Random Forest Regressor, XGBoost

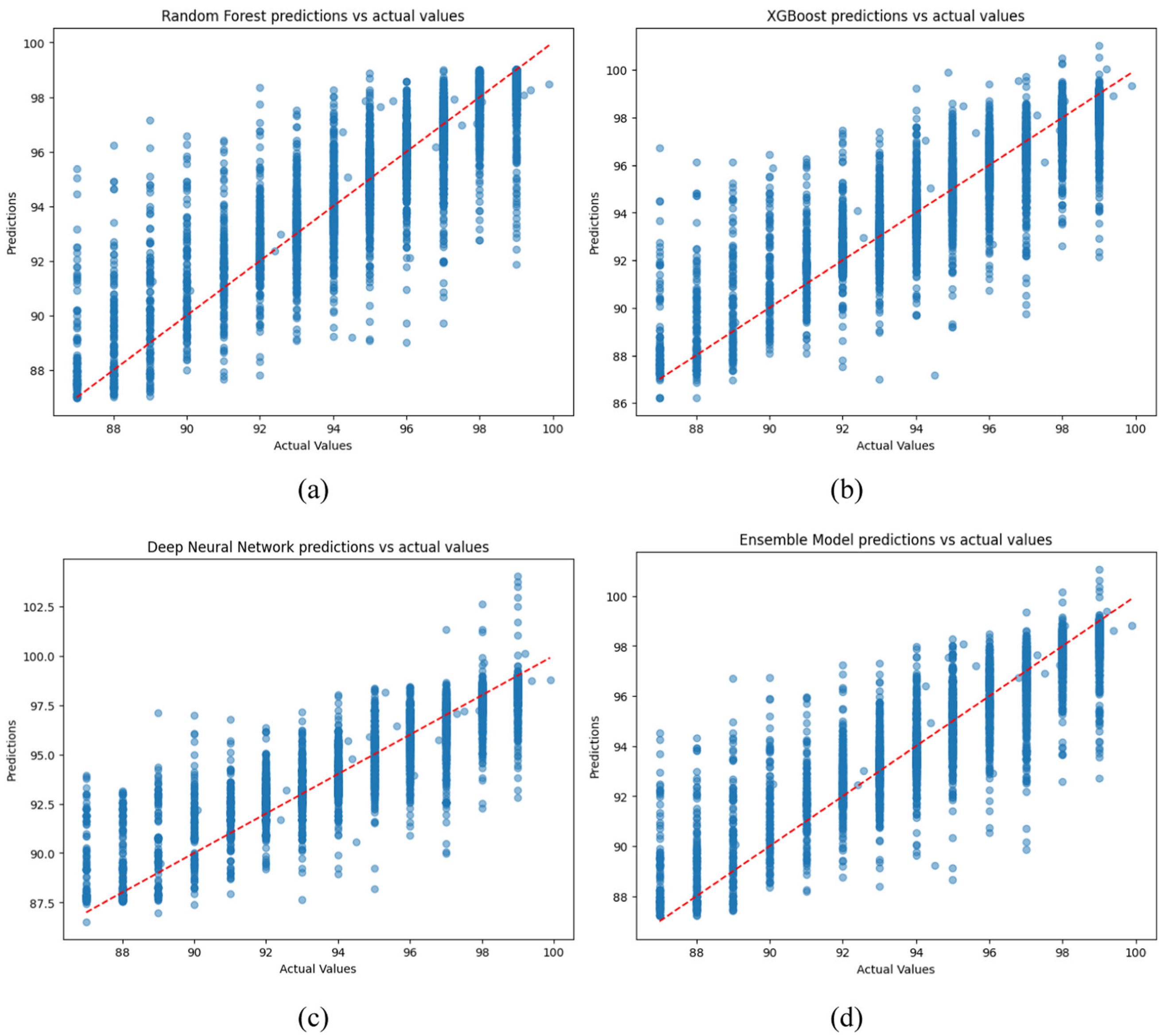


Fig. 10 Scatter plots of actual vs. predicted values for regression models **a** Random Forest predictions, **b** XGBoost predictions, **c** Deep Neural Network predictions, and **d** Ensemble Model predictions

Regressor, and Deep Neural Network provided reasonable predictions, highlighting their utility and reliability in this specific context. This demonstrates the effectiveness of both ensemble methods and individual models in optimizing biodiesel production processes.

Figure 9 shows in detail the loss during the training process of Generative Adversarial Network (GAN) model. The discriminative loss for real data fluctuated initially but stabilized as the process progressed. Similarly, the discriminative loss for the fake data fluctuated initially, then decreased significantly and stabilized. Producer loss showed a high fluctuation at the beginning of the training process but decreased and stabilized over time. The slight increase in the producer loss in the later stages of the training process indicates that the model is trying to produce more realistic data. Overall, the graphs show that the GAN model is effectively trained by successfully balancing between discriminant and producer.

Figure 10 presents the prediction performance of the Random Forest Regressor, XGBoost Regressor, Deep Neural Network, and Ensemble Model based on Linear Regression by comparing their predicted values with actual values. All models exhibit predictions that align closely with the true values, particularly in higher target ranges. In Fig. 10a, the Random Forest Regressor generally provides a strong predictive performance, though some deviations from the reference line are noticeable. Similarly, in Fig. 10b, the XGBoost Regressor demonstrates a comparable level of accuracy, with slightly reduced scatter in the predicted values. Figure 10c

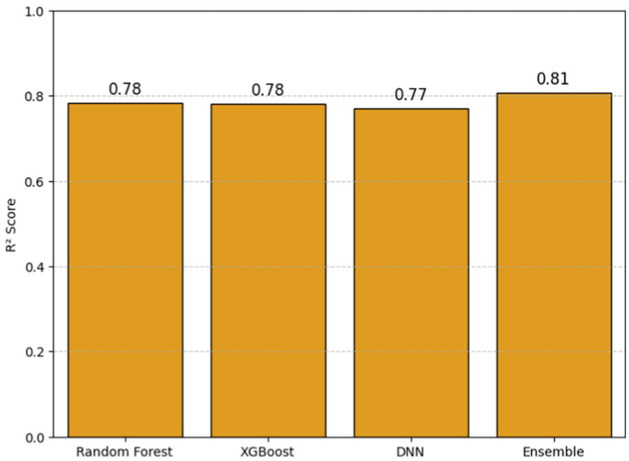


Fig. 11 Bar chart of R^2 scores for the selected and ensemble models

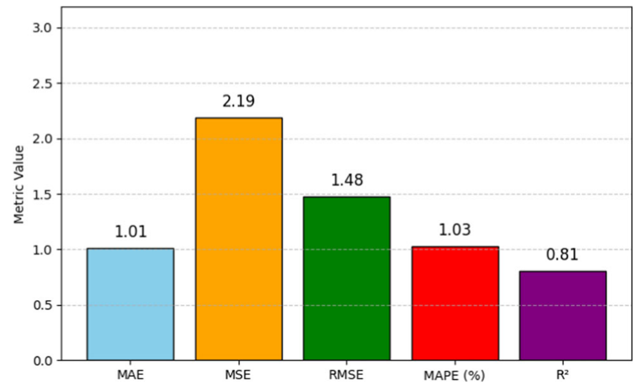


Fig. 12 Evaluation metrics for the ensemble model based on linear regression

shows the Deep Neural Network model, which maintains a reasonable alignment with the actual values but exhibits slightly larger variations at lower prediction ranges. Finally, in Fig. 10d, the Ensemble Model effectively integrates the predictions of the three models, resulting in a more stable and generalized output. These visualizations confirm that the selected models are effective in predicting the target variable, with ensemble learning demonstrating an advantage in improving overall predictive performance.

To facilitate the comparison of model performance, Fig. 11 presents a bar chart visually depicting the R^2 scores of the individual models and the ensemble model on the test dataset. The Random Forest Regressor and XGBoost Regressor each achieved an R^2 score of 0.78, while the Deep Neural Network model obtained a slightly lower score of 0.77. The Ensemble Model based on Linear Regression, which integrates predictions from these models, demonstrated the highest performance with an R^2 score of 0.81. By combining multiple base models, the ensemble approach effectively leveraged their strengths to enhance overall predictive accuracy. This result highlights the advantage of ensemble learning techniques in improving regression performance, making them a valuable approach for predictive modeling tasks.

4 Discussion

In this study, a regression analysis was performed for the prediction of conversion efficiency in the biodiesel production process using the Ensemble Model based on Linear Regression. The performance of the model was analyzed using various evaluation metrics. Figure 12 shows the evaluation metrics for the Voting Regressor model.

As shown in Fig. 12, the Mean Absolute Error (MAE) was calculated as 1.01, indicating that the Ensemble Model based on Linear Regression provides predictions that closely align with the actual values. The Mean Squared Error (MSE) was measured at 2.19, reflecting the presence of some moderate deviations. The Root Mean Squared Error (RMSE), recorded as 1.48, offers an interpretable measure of error magnitude in the units of the target variable, reinforcing the model's predictive reliability. Additionally, the Mean Absolute Percentage Error (MAPE) was determined as 1.03%, highlighting the model's accuracy in percentage terms. The R^2 score of 0.81 further demonstrates the model's strong explanatory power, confirming its effectiveness in capturing the variability of the target variable.

The results presented in the R^2 score comparison illustrate the effectiveness of the ensemble learning approach in enhancing predictive accuracy. While individual models such as Random Forest and XGBoost each achieved an R^2 score of 0.78, and the Deep Neural Network (DNN) slightly lower at 0.77, the Ensemble Model outperformed them all with an R^2 score of 0.81. This improvement can be attributed to the ensemble method's ability to combine the strengths of different models while mitigating their individual weaknesses. By aggregating the predictions of Random Forest, XGBoost, and DNN, the ensemble approach capitalizes on the diverse learning strategies of these models—leveraging the robust feature selection of tree-based methods and the deep feature extraction of neural networks. As a result, the ensemble model reduces variance and improves generalization, leading to a more accurate and stable predictive performance. These findings highlight the advantage of ensemble learning techniques in regression tasks, reinforcing their suitability for complex prediction problems where individual models may struggle to achieve optimal accuracy independently.

Based on the evaluation metrics presented in Figs. 11 and 12, the Ensemble Model based on Linear Regression demonstrated superior predictive performance compared to the individual models. The Mean Absolute Percentage Error (MAPE) was found to be 1.03%, indicating a low deviation from the actual values and highlighting the model's high accuracy. The R^2 score of 0.81 suggests that the majority of the variance in the target variable is well explained by the model.

Additionally, the Ensemble Model based on Linear Regression outperformed the individual models, including the Random Forest, XGBoost, and Deep Neural Network. While the R^2 scores for Random Forest and XGBoost were 0.78, and the Deep Neural Network achieved 0.77, the Ensemble Model improved upon these results by achieving the highest R^2 value of 0.81. This increase demonstrates the effectiveness of combining multiple models to enhance predictive performance.

The rate of improvement achieved was calculated using the following in Eq. 2.

$$\text{Improvement Rate}(\%) = \frac{\text{Ensemle Model } R^2 - \text{Average Base Model } R^2}{\text{Average Base Model } R^2} \times 100 \quad (2)$$

Taking the average base model R^2 value as approximately 0.78, the improvement rate of the Ensemble Model based on Linear Regression was found to be around 3.85%. This result further supports the hypothesis that ensemble learning enhances prediction accuracy by leveraging the strengths of different models while mitigating their weaknesses.

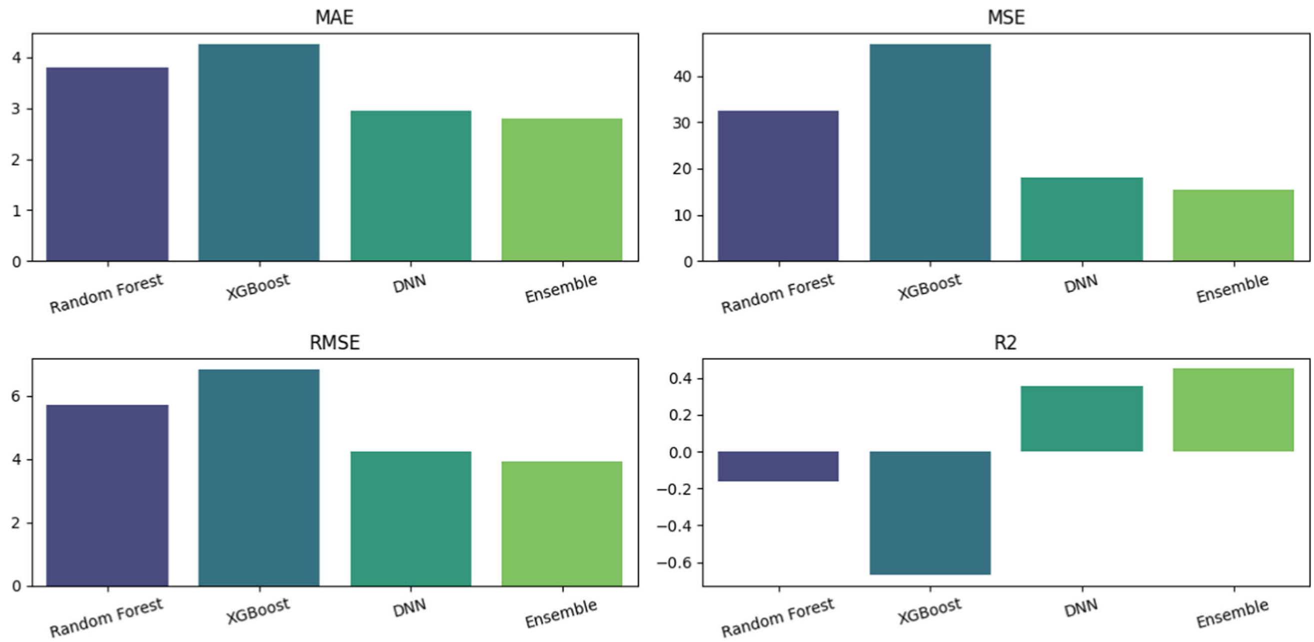
In conclusion, the Ensemble Model based on Linear Regression effectively improves predictive accuracy and generalization in estimating conversion efficiency. However, further optimizations such as hyperparameter tuning, feature selection, and the integration of additional variables could enhance performance even further. Future studies may also explore more complex ensemble techniques or deep learning architectures to further refine the prediction models for biodiesel production processes. These findings contribute to data-driven decision-making strategies for optimizing biodiesel conversion efficiency and advancing sustainable energy solutions.

The comparison of model performance between the original and synthetic datasets, as illustrated in Fig. 13a and b, provides valuable insights into the generalizability and robustness of the applied machine learning models. In the original dataset, the ensemble model demonstrates the best overall performance across multiple evaluation metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2. Notably, the XGBoost model exhibits the highest error values in the original dataset, whereas the deep neural network (DNN) and ensemble model achieve relatively lower error rates. The ensemble model's ability to integrate multiple predictions effectively leads to improved accuracy and reduced error, reinforcing its advantage in complex regression tasks.

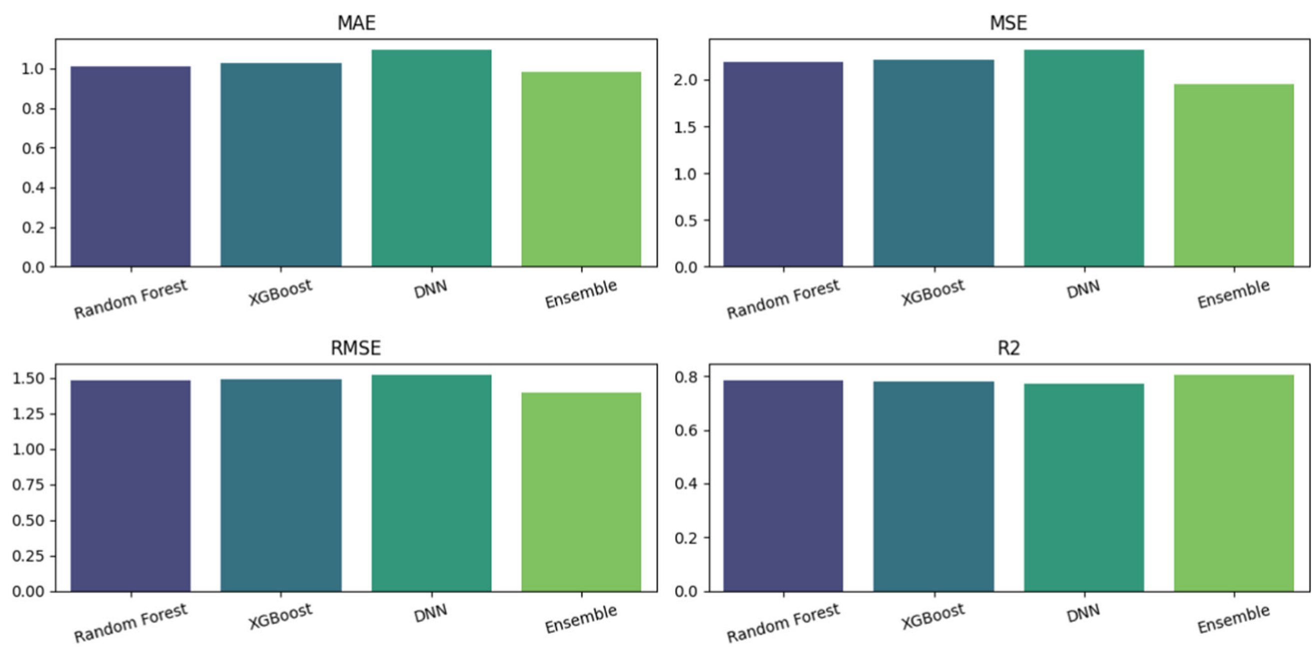
In contrast, when trained on the synthetic dataset, the performance of all models becomes more uniform, as shown in Fig. 13b. The MAE, MSE, and RMSE values for Random Forest, XGBoost, DNN, and the ensemble model converge, indicating that the models respond similarly to the synthetic data distribution. Furthermore, the R^2 values exhibit higher consistency across all models, suggesting that the synthetic dataset lacks the same level of complexity or variability as the original data. While this uniformity implies that the synthetic dataset can serve as a reasonable approximation for model evaluation, it also raises concerns regarding its ability to fully capture the intricacies of real-world data.

The results indicate that while synthetic data can be used for model validation and preliminary testing, it may not always reflect the complexities and patterns present in real-world datasets. The ensemble model continues to exhibit slightly superior performance in both datasets, further supporting the effectiveness of combining multiple models for improved predictive accuracy. However, future studies should explore techniques to enhance the representativeness of synthetic datasets to ensure that model evaluations remain meaningful and applicable to real-world scenarios. Additionally, advanced data augmentation methods or generative approaches, such as Generative Adversarial Networks (GANs), could be investigated to generate more realistic synthetic datasets that better capture the statistical properties of the original data.

Future studies can focus on improving model performance by utilizing larger and more diverse datasets while optimizing hyperparameters. Additionally, the proposed ensemble approach, which has been applied in this study for the first time, can be further explored in high-accuracy prediction tasks to enhance its effectiveness. Implementing this methodology in different domains or refining it with more complex architectures may lead to even greater improvements in predictive accuracy. Moreover, the integration of real-time prediction systems can contribute to the continuous monitoring and optimization of the biodiesel production process. Enhancing model interpretability through techniques such as SHAP or LIME can provide deeper insights into feature importance, facilitating more transparent decision-making. Additionally, analyzing both the environmental and economic impacts of biodiesel production using machine learning-driven optimization approaches could offer a multidimensional perspective for process enhancement. These recommendations not only contribute to academic research



(a) Original data set metrics.



(b) Sentetic data set metrics.

Fig. 13 Comparison of model performance metrics on original and synthetic datasets. **a** Evaluation metrics for models trained on the original dataset, including MAE, MSE, RMSE, and R^2. **b** Evaluation metrics for models trained on the synthetic dataset

but also offer significant practical implications for industrial applications, ensuring that machine learning-driven predictive models are continuously improved and effectively applied in sustainable energy solutions.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

References

1. Staš, M.; Kroufek, J.; Hlinčík, T.; Šimáček, P.: Properties and analysis of gaseous alternative fuels I: hydrogen and l iquefied petroleum gases. Paliva (2022). https://doi.org/10.35933/paliva.2022.02.03
2. Talupula, N.M.; Rao, P.S.; Kumar, B.S.; Praveen, C.: Alternative fuels for internal combustion engines: overview of current research. SSRG Int. J. Mech. Eng. **4**(4), 17–26 (2017). https://doi.org/10.14445/23488360/IJME-V4I4P105
3. Vinoth Kanna, I.; Arulprakasajothi, M.; Eliyas, S.: A detailed study of IC engines and a novel discussion with comprehensive view of alternative fuels used in petrol and diesel engines. Int. J. Ambient Energy **42**(15), 1794–1802 (2021). https://doi.org/10.1080/01430750.2019.1614994
4. Cabrera, E.; de Sousa, J.M.: Use of sustainable fuels in aviation—A review. Energies **15**(7), 2440 (2022). https://doi.org/10.3390/en15072440
5. Teixeira, A.C.; Machado, P.G.; Borges, R.R.; Mouette, D.: Public policies to implement alternative fuels in the road transport sector. Transp. Policy **99**, 345–361 (2020). https://doi.org/10.1016/j.tranpol.2020.08.023
6. Kuznetsov, G.; Antonov, D.; Piskunov, M.; Yanovskyi, L.; Vysokomornaya, O.: Alternative liquid fuels for power plants and engines for aviation, marine, and land applications. Energies **15**(24), 9565 (2022). https://doi.org/10.3390/en15249565
7. Seljak, T.; Vanhove, G.; Mounaïm-Rousselle, C.; Dias, V.; Contino, F.: Fundamental characterization and performance of alternative fuels. Front. Mech. Eng. **6**, 609504 (2020). https://doi.org/10.3389/fmech.2020.609504
8. Jon HVG, Charles LP, Carroll EG, Carroll E. Biodiesel: An Alternative Fuel for Compression Ignition Engines n.d.
9. Demirbas, A.: The importance of bioethanol and biodiesel from biomass. Energy Sour. Part B **3**(2), 177–185 (2008). https://doi.org/10.1080/15567240600815117
10. Bajpai, D.; Tyagi, V.K.: Biodiesel: source, production, composition, properties and its benefits. J. OLEo Sci. **55**(10), 487–502 (2006). https://doi.org/10.5650/jos.55.487
11. Balat, M.: Fuel characteristics and the use of biodiesel as a transportation fuel. Energy Sour. Part A **28**(9), 855–864 (2006). https://doi.org/10.1080/009083190951474
12. Luque, R.; Lovett, J.C.; Datta, B.; Clancy, J.; Campelo, J.M.; Romero, A.A.: Biodiesel as feasible petrol fuel replacement: a multidisciplinary overview. Energy Environ. Sci. **3**(11), 1706–1721 (2010). https://doi.org/10.1039/c0ee00085j
13. V. K, P. K. BIODIESEL: BENEFICIAL FOR ENVIRONMENT AND HUMAN HEALTH n.d.
14. Khairati M. Biodiesel: An Overview. International Journal of Research and Review n.d.;10:228–36. https://doi.org/10.52403/ijrr.20231127.
15. Gupta, K.K.; Kalita, K.; Ghadai, R.K.; Ramachandran, M.; Gao, X.Z.: Machine learning-based predictive modelling of biodiesel production—A comparative perspective. Energies **14**(4), 1122 (2021). https://doi.org/10.3390/EN14041122
16. Ahmad, I.; Ayub, A.; Ibrahim, U.; Khattak, M.K.; Kano, M.: Data-based sensing and stochastic analysis of biodiesel production process. Energies **12**(1), 63 (2018). https://doi.org/10.3390/EN12010063
17. Jin, X.; Li, S.; Ye, H.; Wang, J.; Wu, Y.; Zhang, D.; Ma, H.; Sun, F.; Pugazhendhi, A.; Xia, C.: Investigation and optimization of biodiesel production based on multiple machine learning technologies. Fuel **348**, 128546 (2023). https://doi.org/10.1016/j.fuel.2023.128546
18. Xing, Y.; Zheng, Z.; Sun, Y.; Agha, A.M.: A review on machine learning application in biodiesel production studies. Int. J. Chem. Eng. **2021**(1), 2154258 (2021). https://doi.org/10.1155/2021/2154258
19. Awogbemi, O.; Kallon, D.V.: Application of machine learning technologies in biodiesel production process—A review. Front. Energy Res. **11**, 1122638 (2023). https://doi.org/10.3389/fenrg.2023.1122638
20. Sukpancharoen, S.; Katongtung, T.; Rattanachoung, N.; Tippayawong, N.: Unlocking the potential of transesterification catalysts for biodiesel production through machine learning approach. Bioresour. Technol. **378**, 128961 (2023). https://doi.org/10.1016/j.biortech.2023.128961
21. Liu, P.; Zhang, Y.: Optimization of biodiesel production from oil using a novel green catalyst via development of a predictive model. Arab. J. Chem. **16**, 104785 (2023). https://doi.org/10.1016/j.arabjc.2023.104785
22. Abdelbasset, W.K.; Elkholi, S.M.; Opulencia, M.J.; Diana, T.; Su, C.H.; Alashwal, M.; Zwawi, M.; Algarni, M.; Abdelrahman, A.; Nguyen, H.C.: Development of multiple machine-learning computational techniques for optimization of heterogenous catalytic biodiesel production from waste vegetable oil. Arab. J. Chem. **15**(6), 103843 (2022). https://doi.org/10.1016/j.arabjc.2022.103843
23. Mukasine, A.; Sibomana, L.; Jayavel, K.; Nkurikiyeyezu, K.; Hitimana, E.: Maximizing biogas yield using an optimized stacking ensemble machine learning approach. Energies (Basel) **17**, 364 (2024). https://doi.org/10.3390/en17020364
24. Ahmad, I.; Ayub, A.; Ibrahim, U.; Khattak, M.K.; Kano, M.: Data-based sensing and stochastic analysis of biodiesel production process. Energies (Basel) **12**, 63 (2018). https://doi.org/10.3390/en12010063
25. Almohana, A.I.; Almojil, S.F.; Kamal, M.A.; Alali, A.F.; Kamal, M.; Alkhatib, S.E., et al.: Theoretical investigation on optimization of biodiesel production using waste cooking oil: machine learning modeling and experimental validation. Energy Rep. **8**, 11938–11951 (2022). https://doi.org/10.1016/j.egyr.2022.08.265
26. Qiu, F.; Li, Y.; Yang, D.; Li, X.; Sun, P.: Heterogeneous solid base nanocatalyst: preparation, characterization and application in biodiesel production. Bioresour. Technol. **102**(5), 4150–4156 (2011). https://doi.org/10.1016/j.biortech.2010.12.071
27. Okullo, A.; Temu, A.K.; Ntalikwa, J.W.; Ogwok, P.: Optimization of biodiesel production from jatropha oil. Int. J. Eng. Res. Afr. **3**, 62–73 (2010). https://doi.org/10.4028/www.scientific.net/jera.3.62
28. Yusuff, A.S.; Owolabi, J.O.: Synthesis and characterization of alumina supported coconut chaff catalyst for biodiesel production from waste frying oil. S. Afr. J. Chem. Eng. **30**, 42–49 (2019). https://doi.org/10.1016/j.sajce.2019.09.001
29. Ayu, D.; Aulyana, R.; Astuti, E.W.; Kusmiyati, K.; Hidayati, N.: Catalytic transesterification of used cooking oil to biodiesel: effect of oil-methanol molar ratio and reaction time. Autom. Exp. **2**(3), 73–77 (2019). https://doi.org/10.31603/ae.v2i3.2991

30. Yadav, A.K.; Khan, M.E.; Pal, A.; Singh, B.: Ultrasonic-assisted optimization of biodiesel production from Karabi oil using heterogeneous catalyst. Biofuels **9**(1), 101–112 (2018). https://doi.org/10.1080/17597269.2016.1259522
31. Sadia, H.; Ahmad, M.; Zafar, M.; Sultana, S.; Azam, A.; Khan, M.A.: Variables effecting the optimization of non edible wild safflower oil biodiesel using alkali catalyzed transesterification. Int. J. Green Energy **10**(1), 53–62 (2013). https://doi.org/10.1080/15435075.2011.647367
32. Kapilan, N.; Babu, T.A.; Reddy, R.P.: Study of variables affecting the synthesis of biodiesel from Madhuca Indica oil. Eur. J. Lipid Sci. Technol. **112**(2), 180–187 (2010). https://doi.org/10.1002/ejlt.200800284
33. Yang, Y.; Harper, J.: Adaptive state feedback controller design for efficient biodiesel production under kinetic uncertainty. Chem. Eng. Res. Des. **209**, 81–93 (2024). https://doi.org/10.1016/j.cherd.2024.07.056
34. Corral-Bobadilla, M.; Lostado-Lorza, R.; Sabando-Fraile, C.; Íñiguez-Macedo, S.: An artificial intelligence approach to model and optimize biodiesel production from waste cooking oil using life cycle assessment and market dynamics analysis. Energy **307**, 132712 (2024). https://doi.org/10.1016/j.energy.2024.132712
35. Sarkar, A.; Nayak, B.; Joshi, V.; Paul, B.; Puzari, A.: An economic, and environmentally benign *Psidium guajava* L. leaves catalyst for biodiesel production at room temperature: process optimization, and life cycle cost analysis. Ind. Crops Prod. **222**, 119609 (2024). https://doi.org/10.1016/j.indcrop.2024.119609
36. Tarigan, J.B.; Perangin-angin, S.; Debora, S.; Manalu, D.; Sari, R.N.; Ginting, J.; Sitepu, E.K.: Application of waste cocoa pod husks as a heterogeneous catalyst in homogenizer-intensified biodiesel production at room temperature. Case Stud. Chem. Environ. Eng. **10**, 100966 (2024). https://doi.org/10.1016/j.cscee.2024.100966
37. Malik, M.A.; Zeeshan, S.; Khubaib, M.; Ikram, A.; Hussain, F.; Yassin, H.; Qazi, A.: A review of major trends, opportunities, and technical challenges in biodiesel production from waste sources. Energy Convers. Manag.: X **23**, 100675 (2024). https://doi.org/10.1016/j.ecmx.2024.100675
38. Siyal, A.A.; Mohamed, R.M.; Musa, S.; Rassem, H.H.; Khamidun, M.H.: A review of the performance of geopolymer catalysts for biodiesel production. Biomass Bioenergy **190**, 107373 (2024). https://doi.org/10.1016/j.biombioe.2024.107373
39. Srikumar, K.; Tan, Y.H.; Kansedo, J.; Tan, I.S.; Mubarak, N.M.; Ibrahim, M.L.; Yek, P.N.; Foo, H.C.; Karri, R.R.; Khalid, M.: A review on the environmental life cycle assessment of biodiesel production: selection of catalyst and oil feedstock. Biomass Bioenergy **185**, 107239 (2024). https://doi.org/10.1016/j.biombioe.2024.107239
40. Sarangi, P.K.; Singh, A.K.; Ganachari, S.V.; Pengadeth, D.; Gunda, M.; Aminabhavi, T.M.: Biobased heterogeneous renewable catalysts: production technologies, innovations, biodiesel applications and circular bioeconomy. Environ. Res. **261**, 119745 (2024). https://doi.org/10.1016/j.envres.2024.119745
41. Abeysinghe, S.; Jeong, W.G.; Kwon, E.E.; Baek, K.: Biodiesel production, calcium recovery, and adsorbent synthesis using dairy sludge. Bioresour. Technol. **413**, 131494 (2024). https://doi.org/10.1016/j.biortech.2024.131494
42. Husaini, S.; Kadire, A.; Verma, R.K.; Pydimalla, M.: Biodiesel production from blended feedstocks and by-product utilization for achieving sustainability. Mater. Today Proc. (2024). https://doi.org/10.1016/j.matpr.2024.05.116
43. Yameen, M.Z.; Juchelková, D.; Naqvi, S.R.; Noor, T.; Ali, A.M.; Shahzad, K.; Rashid, M.I.; Mahpudz, A.B.: Biodiesel production from marine macroalgae Ulva lactuca lipids using novel Cu-BTC@ AC catalyst: parametric analysis and optimization. Energy Convers. Manag.: X **23**, 100628 (2024). https://doi.org/10.1016/j.ecmx.2024.100628
44. Pandey, S.; Narayanan, I.; Selvaraj, R.; Varadavenkatesan, T.; Vinayagam, R.: Biodiesel production from microalgae: a comprehensive review on influential factors, transesterification processes, and challenges. Fuel **367**, 131547 (2024). https://doi.org/10.1016/j.fuel.2024.131547
45. Kostić, M.D.; Đorđević, B.S.; Miladinović, M.R.; Stamenković, O.S.: Biodiesel production from sour cherry kernel oil: kinetics, thermodynamics, and optimization. Energy Convers. Manag. **317**, 118858 (2024). https://doi.org/10.1016/j.enconman.2024.118858
46. Sanjurjo, C.; Oulego, P.; Bartolomé, M.; Rodríguez, E.; Gonzalez, R.; Battez, A.H.: Biodiesel production from the microalgae Nannochloropsis gaditana: optimization of the transesterification reaction and physicochemical characterization. Biomass Bioenergy **185**, 107240 (2024). https://doi.org/10.1016/j.biombioe.2024.107240
47. Emeji, I.C.; Patel, B.: Box-Behnken assisted RSM and ANN modelling for biodiesel production over titanium supported zinc-oxide catalyst. Energy **308**, 132765 (2024). https://doi.org/10.1016/j.energy.2024.132765
48. Mekuriaw, T.A.; Abera, M.K.: CaO -catalyzed trans-esterification of brassica carinata seed oil for biodiesel production. Heliyon (2024). https://doi.org/10.1016/j.heliyon.2024.e33790
49. Senusi, W.; Ahmad, M.I.; Khalil, H.A.; Shakir, M.A.; Binhweel, F.; Shalfoh, E.; Alsaadi, S.: Comparative assessment for biodiesel production from low-cost feedstocks of third oil generation. Renew. Energy **236**, 121369 (2024). https://doi.org/10.1016/j.renene.2024.121369
50. Khan, R.I.; Ghosh, N.; Ahmad, S.; Varshney, K.; Nasir, M.; Halder, G.: Deciphering the application efficacy of titanium dioxide mediated calcareous nanocatalyst towards biodiesel production. Mater. Today Sustain. **27**, 100911 (2024). https://doi.org/10.1016/j.mtsust.2024.100911
51. Abbasi, T.U.; Ahmad, M.; Alsahli, A.A.; Asma, M.; Mussagy, C.U.; Abdellatief, T.M.; Pastore, C.; Mustafa, A.: Eco-friendly production of biodiesel from *Carthamus tinctorius* L. seeds using bismuth oxide nanocatalysts derived from *Cannabis sativa* L. Leaf extract. Process. Saf. Environ. Prot. **191**, 710–722 (2024). https://doi.org/10.1016/j.psep.2024.08.108
52. Ghosh, N.; Gouda, S.P.; Rokhum, S.L.; Halder, G.: Elucidating the scale-up potential of biogenic magnetized nanocatalyst towards intensified biodiesel production. Chem. Eng. J. **495**, 152998 (2024). https://doi.org/10.1016/j.cej.2024.152998
53. Siow, H.S.; Sudesh, K.; Ganesan, S.: Insect oil to fuel: optimizing biodiesel production from mealworm (*Tenebrio molitor*) oil using response surface methodology. Fuel **371**, 132099 (2024). https://doi.org/10.1016/j.fuel.2024.132099
54. Ruatpuia, J.V.; Halder, G.; Vanlalchhandama, M.; Lalsangpuii, F.; Boddula, R.; Al-Qahtani, N.; Niju, S.; Mathimani, T.; Rokhum, S.L.: Jatropha curcas oil a potential feedstock for biodiesel production: a critical review. Fuel **370**, 131829 (2024). https://doi.org/10.1016/j.fuel.2024.131829
55. dos Santos, H.C.; Gonçalves, M.A.; da Silva, L.V.; da Cas, V.A.; da Rocha Filho, G.N.; da Conceição, L.R.: $NaAlO_2/CuFe_2O_4$ as a novel basic magnetic heterogeneous catalyst for effective biodiesel production: Synthesis, characterization, and optimization via RSM-FCCD modeling approach. Process. Saf. Environ. Prot. **189**, 228–245 (2024). https://doi.org/10.1016/j.psep.2024.06.047
56. Mohammadpour, M.; Haghighi, M.; Shokrani, R.; Khajehlou, M.E.; Abdollahizadeh, Z.: Influence of carbon template and ultrasound irradiation on combustion design of nanostructured calcium manganese oxides used in biodiesel production. Surf. Interfaces **51**, 104812 (2024). https://doi.org/10.1016/j.surfin.2024.104812

57. El-Khair, M.A.; El Naga, A.O.; Morshedy, A.S.: Rapid and low-temperature biodiesel production from waste cooking oil: kinetic and thermodynamic insights using a $KOH/ZnAl_2O_4$ nanocatalyst derived from waste aluminum foil. Energy Convers. Manag. **318**, 118898 (2024). https://doi.org/10.1016/j.enconman.2024.118898
58. Liu, K.; Wei, G.; Zhu, Y.; Zhang, L.; He, Z.: A clean route of biodiesel production using red mud-based potassium catalyst. J. Environ. Chem. Eng. **11**(5), 111015 (2023). https://doi.org/10.1016/j.jece.2023.111015
59. Emmanouilidou, E.; Lazaridou, A.; Mitkidou, S.; Kokkinos, N.C.: A comparative study on biodiesel production from edible and non-edible biomasses. J. Mol. Struct. **1306**, 137870 (2024). https://doi.org/10.1016/j.molstruc.2024.137870
60. Ali, A.H.; Wanderlind, E.H.; Almerindo, G.I.: Activated carbon obtained from malt bagasse as a support in heterogeneous catalysis for biodiesel production. Renew. Energy **220**, 119656 (2024). https://doi.org/10.1016/j.renene.2023.119656
61. Ponce, S.; Gangotena, P.A.; Anthony, C.; Rodriguez, Y.; Bazani, H.A.; Keller, M.H.; Souza, B.S.; Vizuete, K.; Debut, A.; Mora, J.R.: Aluminum-based catalysts prepared in the presence of pectin for low-energy biodiesel production. Fuel **361**, 130691 (2024). https://doi.org/10.1016/j.fuel.2023.130691
62. Xia, S.; Tao, J.; Zhao, Y.; Men, Y.; Chen, C.; Hu, Y.; Zhu, G.; Chu, Y.; Yan, B.; Chen, G.: Application of waste derived magnetic acid-base bifunctional CoFe/biochar/CaO as an efficient catalyst for biodiesel production from waste cooking oil. Chemosphere **350**, 141104 (2024). https://doi.org/10.1016/j.chemosphere.2023.141104
63. Bagchi, S.K.; Patnaik, R.; Rawat, I.; Prasad, R.; Bux, F.: Beneficiation of paper-pulp industrial wastewater for improved outdoor biomass cultivation and biodiesel production using *Tetradesmus obliquus* (Turpin) Kützing. Renew. Energy **222**, 119848 (2024). https://doi.org/10.1016/j.renene.2023.119848
64. Alam, M.A.; Deng, L.; Ngatcha, A.D.; Fouegue, A.D.; Wu, J.; Zhang, S.; Zhao, A.; Xiong, W.; Xu, J.: Biodiesel production from microalgal biomass by Lewis acidic deep eutectic solvent catalysed direct transesterification. Ind. Crops Prod. **206**, 117725 (2023). https://doi.org/10.1016/j.indcrop.2023.117725
65. Agu, C.M.; Ani, K.A.; Abiazieije, P.O.; Omeje, J.A.; Ekuma, J.C.; Umelo, U.E., et al.: Biodiesel production from waste cat fish oil using heterogeneous catalyst from cat fish born: a viable waste management approach, and ANN modeling of biodiesel yield. Waste Manag. Bull. **1**, 172–181 (2024). https://doi.org/10.1016/j.wmb.2023.11.002
66. Boro, S.; Das, B.; Brahma, S.; Basumatary, B.; Basumatary, S.F.; Basumatary, S.: Biodiesel production using areca nut (Areca catechu L.) leaf ash-K_2CO_3 catalyst via transesterification from an oil blend of three different feedstocks. Sustain. Chem. Environ. **8**, 100164 (2024). https://doi.org/10.1016/j.scenv.2024.100164
67. Galusnyak, S.C.; Petrescu, L.; Cosprundan, A.M.; Cormos, C.C.: Biodiesel production using various methanol sources and catalytic routes: a techno-environmental analysis. Renew. Energy **232**, 121051 (2024). https://doi.org/10.1016/j.renene.2024.121051
68. Gonçalves, M.A.; dos Santos, H.C.; Ribeiro, T.S.; da Cas, V.A.; da Rocha Filho, G.N.; da Conceição, L.R.: Boosting biodiesel production of waste frying oil using solid magnetic acid catalyst from agro-industrial waste. Arab. J. Chem. **17**(2), 105521 (2024). https://doi.org/10.1016/j.arabjc.2023.105521
69. Soodesh, C.Y.; Seriyala, A.K.; Chattopadhyay, P.; Rozhkova, N.; Michalkiewicz, B.; Chatterjee, S.; Roy, B.: Carbonaceous catalysts (biochar and activated carbon) from agricultural residues and their application in production of biodiesel: a review. Chem. Eng. Res. Des. **203**, 759–788 (2024). https://doi.org/10.1016/j.cherd.2024.02.002
70. Aloia, A.; Izzi, M.; Rizzuti, A.; Casiello, M.; Mastrorilli, P.; Cioffi, N.; Nacci, A.; Picca, R.A.; Monopoli, A.: Chitosan-supported calcium hydroxide hybrid material as new, efficient, and recyclable catalyst for biodiesel production. Mol. Catal. **560**, 114128 (2024). https://doi.org/10.1016/j.mcat.2024.114128
71. Mawlid, O.A.; Abdelhady, H.H.; Abd El-Moghny, M.G.; Hamada, A.; Abdelnaby, F.; Kased, M.; Al-Bajouri, S.; Elbohy, R.A.; El-Deab, M.S.: Clean approach for catalytic biodiesel production from waste frying oil utilizing K_2CO_3/Orange peel derived hydrochar via RSM Optimization. J. Clean. Prod. **442**, 140947 (2024). https://doi.org/10.1016/j.jclepro.2024.140947
72. Khan, M.R.; Singh, H.N.: Clean biodiesel production approach using waste swan eggshell derived heterogeneous catalyst: an optimization study employing Box-Behnken-response surface methodology. Ind. Crops Prod. **220**, 119181 (2024). https://doi.org/10.1016/j.indcrop.2024.119181
73. Esmi, F.; Dalai, A.K.; Hu, Y.: Comparison of various machine learning techniques for modeling the heterogeneous acid-catalyzed alcoholysis process of biodiesel production from green seed canola oil. Energy Rep. **12**, 321–328 (2024). https://doi.org/10.1016/j.egyr.2024.06.029
74. Zhang, M.; Jiao, X.; Ren, D.; Huang, C.; Li, Y.; Chen, D.; Huo, Z.: Design of anti-saponification Ca-based catalyst with porous-shell structure from crustacean waste for biodiesel production. Fuel **363**, 131006 (2024). https://doi.org/10.1016/j.fuel.2024.131006
75. Piran, S.; Kazemeini, M.; Mohebolkhames, E.; Tamtaji, M.: Experimental and computational investigations of biodiesel production from castor oil using Li-doped salmon fish bone-supported lanthanum oxide catalyst in a micro-reactor. Renew. Energy **236**, 121455 (2024). https://doi.org/10.1016/j.renene.2024.121455
76. Farobie, O.; Santosa, N.F.; Fatriasari, W.; Karimah, A.; Amrullah, A.; Suseno, S.H.; Nandiyanto, A.B.; Hartulistiyoso, E.: Harnessing macroalgae Sargassum plagiophyllum-derived heterogeneous catalyst for biodiesel production. Bioresour. Technol. Rep. **25**, 101768 (2024). https://doi.org/10.1016/j.biteb.2024.101768
77. Señorans, S.; Rangel-Rangel, E.; Maya, E.M.; Díaz, L.: Hyper-crosslinked porous polymer as catalyst for efficient biodiesel production. React. Funct. Polym. **202**, 105964 (2024). https://doi.org/10.1016/j.reactfunctpolym.2024.105964
78. Wang, F.P.; Kang, L.L.; Wang, Y.J.; Wang, Y.R.; Wang, Y.T.; Li, J.G.; Jiang, L.Q.; Ji, R.; Chao, S.; Zhang, J.B.; Fang, Z.: Magnetic biochar catalyst from reed straw and electric furnace dust for biodiesel production and life cycle assessment. Renew. Energy **227**, 120570 (2024). https://doi.org/10.1016/j.renene.2024.120570
79. Peng, L.; Bahadoran, A.; Sheidaei, S.; Ahranjani, P.J.; Kamyab, H.; Oryani, B.; Arain, S.S.; Rezania, S.: Magnetic graphene oxide supported tin oxide (SnO) nanocomposite as a heterogeneous catalyst for biodiesel production from soybean oil. Renew. Energy **224**, 120050 (2024). https://doi.org/10.1016/j.renene.2024.120050
80. Tizvir, A.; Molaeimanesh, G.R.; Zahedi, A.R.; Labbafi, S.: Optimization of biodiesel production from microalgae and investigation of exhaust emissions and engine performance for biodiesel blended. Process Safety Environ. Prot. **175**, 319–340 (2023). https://doi.org/10.1016/j.psep.2023.05.056
81. Yusuf, H.A.; Abdulla, A.F.; Radhi, F.A.; Hussain, Z.J.: Optimization of biodiesel production in a high throughput branched microreactor. Energy Nexus **13**, 100276 (2024). https://doi.org/10.1016/j.nexus.2024.100276
82. Hasannia, S.; Kazemeini, M.; Seif, A.: Optimizing parameters for enhanced rapeseed biodiesel production: a study on acidic and basic carbon-based catalysts through experimental and DFT

evaluations. Energy Convers. Manag. **303**, 118201 (2024). https://doi.org/10.1016/j.enconman.2024.118201
83. Kumar, S.; Singh, P.; Shrivastava, V.; Kaur Sohal, J.; Sharma, A.; Kumar, M., et al.: Potential non-edible oil seeds for biodiesel production: an Indian context. Mater. Today Proc. (2023). https://doi.org/10.1016/j.matpr.2023.05.419
84. Nyorere, O.; Oluka, S.I.; Onoji, S.E.; Nwadiolu, R.; Adepoju, T.F.: Production of biodiesel from biocatalysis of agro-wastes in acidic environment. Sci. Afr. **24**, e02154 (2024). https://doi.org/10.1016/j.sciaf.2024.e02154
85. Khan, U.N.; Shah, G.M.; Inayat, A.; Abbas, S.M.: Quercus ballot as an innovative feedstock for biodiesel production using ZnO nanocatalyst. Fuel **381**, 133307 (2025). https://doi.org/10.1016/j.fuel.2024.133307
86. Asaad, S.M.; Inayat, A.; Ghenai, C.; Shanableh, A.: Response surface methodology in biodiesel production and engine performance assessment. Int. J. Thermofluids **21**, 100551 (2024). https://doi.org/10.1016/j.ijft.2023.100551
87. Fu, H.; Xiao, Y.; Abulizi, A.; Okitsu, K.; Maeda, Y.; Ren, T.: Surfactant-modified $ZnFe_2O_4$/CaOPS porous acid-base bifunctional catalysts for biodiesel production from waste cooking oil and process optimization. J. Environ. Chem. Eng. **12**(6), 114234 (2024). https://doi.org/10.1016/j.jece.2024.114234
88. Elnasr, T.A.; Al-Enezi, A.T.; Hussein, M.F.; Bielal, H.; Alhumaimess, M.S.; El-Ossaily, Y.A.; Hassan, H.M.; AlNahwa, L.H.; Aldawsari, A.M.; Alsohaimi, I.H.: Sustainable biodiesel production from waste olive oil: utilizing olive pulp-derived catalysts for environmental and economic benefits. Sustain. Chem. Pharm. **37**, 101426 (2024). https://doi.org/10.1016/j.scp.2024.101426
89. Kingkam, W.; Nuchdang, S.; Phalakornkule, C.; Suwanmanee, U.; Rattanaphra, D.: Synergistic effect of La_2O_3-Al_2O_3 based catalysts for efficient biodiesel production. J. Ind. Eng. Chem. (2024). https://doi.org/10.1016/j.jiec.2024.08.023
90. Yusuf, B.O.; Oladepo, S.A.; Ganiyu, S.A.; Peedikakkal, A.M.: Synthesis and evaluation of CuO/UiO-66 metal-organic frameworks as a solid catalyst for biodiesel production from waste cooking oil. Mol. Catal. **564**, 114313 (2024). https://doi.org/10.1016/j.mcat.2024.114313
91. Umeagukwu, O.E.; Onukwuli, D.O.; Ude, C.N.; Chizoo, E.; Nnamdi Ekwueme, B.; Asadu, C.O., et al.: Techno-economic analysis of the statistically optimized biodiesel production process using African pear seed oil and activated empty palm fruit bunch biocatalyst. Waste Manag. Bull. **2**, 95–112 (2024). https://doi.org/10.1016/j.wmb.2024.03.006
92. Rahim, M.; Liu, D.; Du, W.; Zhao, X.: Techno-economic assessment of co-production of biodiesel and epoxy fatty acid methyl ester as renewable fuel and plasticizer from waste cooking oil. Chem. Eng. J. **499**, 156363 (2024). https://doi.org/10.1016/j.cej.2024.156363
93. Kivevele, T.; Kichonge, B.: Techno-economic evaluation of transesterification processes for biodiesel production from low quality non-edible feedstocks: process design and simulation. Energy **297**, 131302 (2024). https://doi.org/10.1016/j.energy.2024.131302
94. Lani, N.S.; Ngadi, N.; Haron, S.; Inuwa, I.M.; Opotu, L.A.: The catalytic effect of calcium oxide and magnetite loading on magnetically supported calcium oxide-zeolite catalyst for biodiesel production from used cooking oil. Renew. Energy **222**, 119846 (2024). https://doi.org/10.1016/j.renene.2023.119846
95. He, P.; Zhang, W.; Fu, G.; Xie, E.; Wang, W.; Zhang, Z.; Zhang, T.; Wu, G.: Two-dimensional Mo catalysts supported on the external surface of planar Silicalite-1 zeolite for biodiesel production from waste cooking soybean oil: transesterification experiment and kinetics. Fuel **360**, 130610 (2024). https://doi.org/10.1016/j.fuel.2023.130610
96. Zhang, Y.; Li, Y.; Sun, S.: Unveiling the innovation: optimized biodiesel production from emerging Acer truncatum Bunge seed oil using novel and highly effective alkaline ionic liquid catalyst. Chem. Eng. J. **487**, 150603 (2024). https://doi.org/10.1016/j.cej.2024.150603
97. Das, S.; Anal, J.M.H.; Kalita, P.; Saikia, L.; Rokhum, S.L.: Upcycling waste snail shells into high-performance nanocatalyst for optimized biodiesel production: a sustainable approach. Invent. Discl. **4**, 100024 (2024). https://doi.org/10.1016/j.inv.2024.100024
98. Razaq, Z.; Tousif, M.I.; Noureen, S.; Hussain, S.U.; Saleem, M.; Mehmood Khan, F., et al.: Utilization of opium poppy seed oil for biodiesel production: a parametric characterization and statistical optimization. Heliyon **10**, e36851 (2024). https://doi.org/10.1016/j.heliyon.2024.e36851
99. Ashouri, R.; Jafari, D.; Esfandyari, M.; Vatankhah, G.; Mahdavi, M.: Valorization of slaughterhouse wastes through transesterification for sustainable biodiesel production using potassium hydroxide as a heterogeneous catalyst. J. Clean. Prod. **447**, 141596 (2024). https://doi.org/10.1016/j.jclepro.2024.141596
100. Ouafi, R.; Haldhar, R.; Mehdaoui, I.; Asri, M.; AlObaid, A.A.; Warad, I.; Taleb, M.; Rais, Z.; Kim, S.C.: Waste snail shells-derived mixed oxide catalyst for efficient transesterification of vegetable oil: towards sustainable biodiesel production. Mater. Today Commun. **39**, 109128 (2024). https://doi.org/10.1016/j.mtcomm.2024.109128