Nigerian Journal of Physics (NJP)
ISSN online: 3027-0936
ISSN print: 1595-0611
DOI: https://doi.org/10.62292/njp.v35(s).2026.734
Volume 35(S), 2026

# Development of a Multiple Linear Regression Model for Predicting Key Fuel Properties of *Jatropha curcas* and *Azadirachta indica*-Based Biodiesel

*[1]Polycarp Ulea Tanko, [2]Joseph Aidan, [2]Pascal Timtere and [3]Moses Tara Langkuk

[1]Department of Physics Education, Federal College of Education (Technical) Gombe
[2]Department of Physics, Modibbo Adama University, Yola, Adamawa State, Nigeria
[3]Chemistry Advanced Research Centre (CARC), Sheda Science and Technology Complex (SHESTCO), Abuja, Nigeria.

*Corresponding Author's Email: uleapt@gmail.com

**Keywords**:
Biodiesel,
Multiple Linear Regression,
*Jatropha curcas*,
*Azadirachta indica*,
Hybrid Feedstock,
Kinematic Viscosity,
Calorific Value,
Fuel Property Prediction

**ABSTRACT**

Accurate prediction of biodiesel fuel properties is essential for reducing experimental cost and improving the efficiency of fuel formulation. This study developed Multiple Linear Regression (MLR) models to predict the kinematic viscosity and calorific value of biodiesel–diesel blends produced from pure and hybrid *Jatropha curcas* and *Azadirachta indica* feedstocks. Biodiesel was produced from five feedstock formulations comprising pure *Jatropha curcas*, pure *Azadirachta indica*, and three intermediate hybrid mixtures, after which each biodiesel was blended with petroleum diesel at B5–B30 ratios. Experimental measurements of the selected fuel properties were used to establish regression relationships based on two predictor variables: Jatropha feedstock composition and biodiesel blend ratio. The developed models showed reasonable predictive capability, with coefficients of determination ($R^2$) of 0.9860 for kinematic viscosity and 0.9948 for calorific value. Prediction errors were consistently low, with RMSE values of 0.0505 $mm^2\ s^{-1}$ and 0.0446 $MJ\ kg^{-1}$, and MAE values of 0.0413 $mm^2\ s^{-1}$ and 0.0356 $MJ\ kg^{-1}$ for kinematic viscosity and calorific value, respectively. Leave-One-Out Cross-Validation further confirmed good model performance, with cross-validated $R^2$ values of 0.9810 for kinematic viscosity and 0.9932 for calorific value. Residual diagnostics showed constant error variance for both models, while normality was better satisfied for the calorific-value model than for the kinematic-viscosity model. The findings show that simple MLR equations can provide useful estimates of key fuel properties from feedstock composition and biodiesel blend ratio within the experimental range investigated, thereby reducing the need for repeated laboratory analyses. The proposed models therefore provide a practical and computationally efficient approach for biodiesel formulation, fuel quality evaluation and preliminary engineering analysis, while offering a useful basis for future predictive studies involving other non-edible biodiesel feedstocks.

**Submitted**: 07 Aug. 2026
**Accepted**: 11 Aug. 2026
**Published**: 16 Sept. 2026

## INTRODUCTION

The increasing global demand for energy, depletion of fossil fuel reserves and growing environmental concerns have intensified the search for cleaner and more sustainable energy sources. Consequently, renewable fuels have attracted considerable research interest because of their potential to enhance energy security and mitigate environmental impacts, particularly in developing countries (Atsuwe et al., 2025). Among these, biodiesel has emerged as a promising alternative due to its renewability, biodegradability and compatibility with conventional compression-ignition engines. Produced through the transesterification of oils and fats into fatty acid methyl esters (FAME), biodiesel can be used either as a neat fuel or blended with petroleum diesel, thereby reducing dependence on fossil-based fuels (Vilas Bôas & Mendes, 2022).

Growing interest in sustainable biodiesel has increased attention on non-edible feedstocks such as *Jatropha curcas* and *Azadirachta indica* (Neem), which can support renewable energy development without directly competing with food resources (Riayatsyah et al., 2022). Among the properties that determine biodiesel quality, kinematic viscosity and calorific value are particularly important because they affect fuel injection, atomisation, combustion and energy output. Their laboratory determination, however, can become time-consuming and costly when several feedstock compositions and blend ratios are considered (Santos et al., 2022; Xiao et al., 2023).

Mathematical modelling offers a practical way of estimating biodiesel properties while reducing repeated laboratory measurements, and data-driven methods are increasingly being applied in biodiesel research (Adejuwon et al., 2026; Ishola et al., 2024). Recent studies have further demonstrated that compositional information can be successfully incorporated into data-driven models for predicting important biodiesel fuel properties (Díez Valbuena et al., 2024). Multiple Linear Regression (MLR) is particularly attractive because it is simple, computationally efficient and easy to interpret. However, most existing studies have focused on single feedstocks or fatty acid composition, with relatively limited attention given to models that combine hybrid non-edible feedstock composition and biodiesel blend ratio (Yahya & Aghel, 2021; Riayatsyah et al., 2022). Although predictive techniques have recently been applied to biodiesel production from blended or multi-feedstock systems, comparatively less attention has been given to simple and interpretable models that simultaneously incorporate hybrid feedstock composition and biodiesel–diesel blend ratio (Amenaghawon et al., 2024).

This study therefore developed MLR models for predicting the kinematic viscosity and calorific value of biodiesel–diesel blends produced from pure and hybrid *Jatropha curcas* and *Azadirachta indica* feedstocks. Experimental data from five feedstock formulations and different blend ratios were used, while model performance was assessed using $R^2$, RMSE and MAE. The resulting equations provide a simple and efficient approach for estimating these fuel properties and supporting biodiesel formulation and fuel quality assessment.

## MATERIALS AND METHODS

### Experimental Data

The experimental data were obtained from biodiesel produced from *Jatropha curcas* (J) and *Azadirachta indica* (N) seed oils using five feedstock formulations: 100J, 75J:25N, 50J:50N, 25J:75N and 100N. Biodiesel was produced through NaOH-catalysed transesterification using a 6:1 methanol-to-oil molar ratio and 1.0 wt.% NaOH catalyst. The reaction was carried out at 60–70 °C for 60 min with continuous stirring at 600 rpm. After the reaction, the biodiesel was separated, washed with distilled water and dried before characterization.

Each biodiesel was blended with petroleum diesel at B5, B10, B15, B20, B25 and B30. Density was determined using a pycnometer and a Mettler Toledo ME-series analytical balance. Kinematic viscosity was measured at 38 °C using an NDJ-5S viscometer, with ASTM D445-24 used as the reference standard for comparison (ASTM International, 2024), while calorific value was determined using a Parr 6100 bomb calorimeter in dynamic mode. All measurements were carried out in triplicate (n = 3), and the mean values were used for the analysis.

The five feedstock formulations and six blend ratios provided 30 observations for modelling. Jatropha composition (J) and biodiesel blend ratio (B) were used as predictor variables, while kinematic viscosity and calorific value were the response variables in the MLR models.

### Model Variables

Multiple Linear Regression (MLR) was adopted to establish mathematical relationships between the selected predictor variables and the fuel properties investigated. Owing to its simplicity, robustness and ability to generate explicit predictive equations, MLR has been widely applied to multivariable modelling in applied physics and engineering (Sheu et al., 2026; Osman et al., 2024).

*Jatropha curcas* composition (J) and biodiesel blend ratio (B) were selected as the predictor variables, while kinematic viscosity (KV) and calorific value (CV) served as the response variables. These variables adequately describe the blend composition and the fuel properties investigated.

The proportions of *Azadirachta indica* (N) and petroleum diesel (D) were excluded because they are not independent variables. Their values are completely determined by the compositional relationships

$$J + N = 100 \quad (1)$$

$$B + D = 100 \quad (2)$$

where J and N represent the percentages of *Jatropha curcas* and *Azadirachta indica* in the biodiesel feedstock, while B and D denote the percentages of biodiesel and petroleum diesel in the fuel blend.

Including all four composition variables would introduce perfect multicollinearity because N and D are linear

functions of J and B, respectively. This violates a fundamental assumption of MLR and produces unstable regression coefficients. Therefore, only J and B were retained as predictor variables to avoid perfect multicollinearity and ensure that the regression coefficients could be estimated and interpreted correctly.

**Formulation of the Multiple Linear Regression Model**

Multiple Linear Regression (MLR) was employed to establish mathematical relationships between the selected predictor variables and the measured fuel properties. The method was chosen because it provides a simple and statistically robust means of quantifying the combined effects of feedstock composition and biodiesel blend ratio. Furthermore, MLR generates explicit regression equations that facilitate straightforward interpretation and prediction within the experimental domain.

The general form of the Multiple Linear Regression model is expressed as

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad (3)$$

where:

$\hat{Y}$= Predicted response variable (kinematic viscosity or calorific value)

$\beta_0$= Intercept (constant term)

$\beta_1, \beta_2,$= Regression coefficients

$X_1, X_2$= Feedstock composition (%)

$\varepsilon$= Random error term

For the present study, the percentage composition of *Jatropha curcas* (J) and the biodiesel blend ratio (B) were selected as the independent variables because they fully describe the experimental blend compositions. Separate regression models were developed for kinematic viscosity (KV) and calorific value (CV), which are expressed as

$$\widehat{KV} = \beta_0 + \beta_1 J + \beta_2 B + \varepsilon \quad (4)$$

$$\widehat{CV} = \beta_0 + \beta_1 J + \beta_2 B + \varepsilon \quad (5)$$

The regression coefficients were estimated using the Ordinary Least Squares (OLS) technique, which determines the line of best fit by minimizing the sum of the squared differences between the experimental measurements and the corresponding model predictions. The predictive ability of the developed models was subsequently evaluated using the coefficient of determination ($R^2$), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

**Model Performance Evaluation**

The predictive performance of the developed Multiple Linear Regression (MLR) models was evaluated by comparing the predicted values with the corresponding experimental observations. Model accuracy was assessed using the coefficient of determination ($R^2$), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Together, these statistical measures provide a comprehensive assessment of model fit and prediction accuracy.

The coefficient of determination ($R^2$), which indicates the proportion of variation in the observed data explained by the regression model, is given by

$$\widehat{R^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (6)$$

where:

$y_i$= experimental (observed) value,

$\hat{y}_i$= predicted value,

$\bar{y}$= mean of the experimental values, and

$n$= total number of observations.

The Root Mean Square Error (RMSE), which measures the average magnitude of the prediction error, was calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (7)$$

where the variables $y_i$ and $\hat{y}_i$ are as defined in Equation (6) above.

The Mean Absolute Error (MAE), which represents the average absolute difference between the observed and predicted values, was determined using:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \quad (8)$$

where the variables $y_i$ and $\hat{y}_i$ are as defined in Equation (6) above.

Models with higher $R^2$ values and lower RMSE and MAE values were considered to exhibit better predictive performance.

Leave-One-Out Cross-Validation (LOOCV) was used to further assess the performance of the developed MLR models. The analysis was carried out using the 30 experimental observations. In each iteration, one observation was removed from the dataset, the model was fitted using the remaining 29 observations, and the excluded observation was then predicted. This process was repeated until every observation had been used once as the validation sample. The cross-validated predictions were evaluated using the coefficient of determination ($R^2$CV), root mean square error (RMSECV), and mean absolute error (MAECV). The 30 observations arise from the five feedstock formulations evaluated at six biodiesel blend ratios.

## RESULTS AND DISCUSSION

**Descriptive Statistics of the Experimental Dataset**

The descriptive statistics presented in Table 1 summarize the experimental dataset used to develop the Multiple

Linear Regression (MLR) models. The predictor variables cover the full range of feedstock compositions and biodiesel blend ratios investigated, with *Jatropha curcas* composition varying from 0 to 100% and biodiesel blend ratio from 5 to 30%. This broad experimental coverage provides a sound basis for developing and evaluating the predictive models (Bukkarapu & Krishnasamy, 2024a).

The response variables also show sufficient variability for model development. Kinematic viscosity ranged from 2.80 to 4.36 $mm^2\ s^{-1}$ (mean = 3.46 $mm^2\ s^{-1}$), while calorific value varied from 41.80 to 44.10 $MJ\ kg^{-1}$ (mean = 43.05 $MJ\ kg^{-1}$). The corresponding standard deviations indicate a reasonable spread of the observations, confirming that the dataset adequately captures the influence of feedstock composition and biodiesel blend ratio on the measured fuel properties. Comparable variations in viscosity and other thermophysical properties of blended vegetable oils have been reported by Uwaoma et al. (2025), while recent studies have demonstrated the effectiveness of data-driven models for reliable prediction of biodiesel fuel properties (Arif et al., 2025).

Overall, the descriptive statistics indicate that the experimental data are well distributed across the investigated design space, providing a reliable foundation for developing and evaluating the proposed Multiple Linear Regression models.

**Table 1: Descriptive Statistics of the Variables Used for MLR Model Development**

| Variable | Unit | Minimum | Maximum | Mean | Standard Deviation |
|---|---|---|---|---|---|
| *Jatropha curcas* composition (J) | % | 0 | 100 | 50.00 | 35.96 |
| Biodiesel blend ratio (B) | % | 5 | 30 | 17.50 | 8.69 |
| Kinematic viscosity (KV) | $mm^2\ s^{-1}$ | 2.80 | 4.36 | 3.46 | 0.43 |
| Calorific value (CV) | $MJ\ kg^{-1}$ | 41.80 | 44.10 | 43.05 | 0.63 |

The results show that the dataset adequately represents the range of feedstock compositions and biodiesel blend ratios investigated in this study. The observed variation in kinematic viscosity and calorific value reflects the influence of changes in blend composition on the measured fuel properties. Such variability is essential for regression modelling because it enables meaningful relationships to be established between the predictor variables and the corresponding fuel properties. The distribution of the experimental observations therefore provides an appropriate basis for developing and cross-validating the proposed MLR models.

**Correlation Analysis**

Pearson correlation analysis was conducted to examine the relationships between the predictor variables and the selected fuel properties before developing the Multiple Linear Regression (MLR) models. The correlation coefficients are presented in Table 2**.**

**Table 2: Pearson Correlation Matrix for the Variables Used in the MLR Models**

| Variable | J | B | KV | CV |
|---|---|---|---|---|
| J | 1.000 | 0.000 | -0.348 | 0.312 |
| B | 0.000 | 1.000 | 0.930 | -0.947 |
| KV | -0.348 | 0.930 | 1.000 | -0.996 |
| CV | 0.312 | -0.947 | -0.996 | 1.000 |

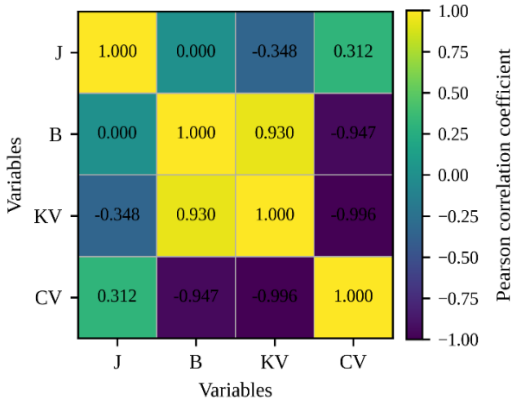


Figure 1: Pearson Correlation Heat Map of the Predictor and Response Variables Used in the MLR Analysis

The Pearson correlation matrix presented in Table 2 and visualized in Figure 1 provides an overview of the relationships between the predictor variables and the investigated fuel properties. Among the predictors, biodiesel blend ratio showed the strongest association with the response variables. Its strong positive correlation with kinematic viscosity (r = 0.930) indicates that viscosity increased with increasing biodiesel content, whereas it's strong negative correlation with calorific value (r = –0.947) indicates that higher biodiesel proportions were associated with lower energy content. These trends are consistent with the experimental results and show that biodiesel blend ratio had the strongest association with the measured fuel properties among the predictor variables considered.

The relationships between *Jatropha curcas* composition and the response variables were comparatively weaker. The moderate negative correlation with kinematic viscosity (r = –0.348) suggests that increasing the proportion of *Jatropha curcas* slightly reduced viscosity, while the moderate positive correlation with calorific value (r = 0.312) indicates a modest increase in fuel energy content. Although these effects were less pronounced than those of blend ratio, they indicate that feedstock composition is also associated with variations in the measured fuel properties. A nearly perfect inverse relationship was observed between kinematic viscosity and calorific value (r = –0.996), indicating that blends with higher viscosity generally exhibited lower calorific values. This strong negative relationship agrees with the trend observed in the biodiesel–diesel blends investigated in this study.

The zero Pearson correlation indicates that J and B were orthogonal in the experimental design, thereby minimizing linear multicollinearity. This was further supported by VIF and tolerance values of 1.000 for both predictors, indicating no evidence of linear multicollinearity. Overall, the correlation analysis shows that the selected predictor variables are suitable for developing the MLR models used to estimate the kinematic viscosity and calorific value of the biodiesel–diesel blends.

**Development of the Multiple Linear Regression Models**

Multiple Linear Regression (MLR) analysis was used to develop predictive models for kinematic viscosity and calorific value based on *Jatropha curcas* composition (J) and biodiesel blend ratio (B**).** The developed equations describe the relationship between the selected predictor variables and the fuel properties investigated in this study.

***MLR Model for Kinematic Viscosity***

The developed Multiple Linear Regression model for predicting the kinematic viscosity of the biodiesel–diesel blends is expressed as

$$KV = 2.858067 - 0.004193J + 0.046377B \quad (9)$$

where (KV) denotes the predicted kinematic viscosity ($mm^2\ s^{-1}$), (J) represents the percentage composition of *Jatropha curcas* in the biodiesel feedstock, and (B) is the biodiesel blend ratio (%).

The regression coefficients reveal that both predictor variables influence the kinematic viscosity of the biodiesel–diesel blends, although to different extents. The positive coefficient of the biodiesel blend ratio indicates that increasing the biodiesel content leads to a corresponding increase in viscosity. This observation is consistent with the experimental results and reflects the inherently higher viscosity of biodiesel relative to petroleum diesel.

Conversely, the negative coefficient associated with *Jatropha curcas* composition suggests that increasing its proportion in the feedstock produces a slight reduction in the predicted kinematic viscosity. Although this effect is relatively small, it indicates that feedstock composition contributes to the overall variation in viscosity alongside the biodiesel blend ratio.

Overall, the regression model describes the combined influence of feedstock composition and biodiesel blend ratio on kinematic viscosity. Its simple mathematical form allows viscosity to be estimated readily within the experimental range investigated and provides a practical basis for biodiesel formulation, fuel quality assessment and preliminary engineering analysis.

***MLR Model for Calorific Value***

The Multiple Linear Regression model developed for predicting the calorific value of the biodiesel–diesel blends is given by:

$$CV = 43.977333 + 0.005467J - 0.068800B \quad (10)$$

where (CV) is the predicted calorific value ($MJ\ kg^{-1}$), (J) denotes the percentage composition of *Jatropha curcas* in the biodiesel feedstock, and (B) represents the biodiesel blend ratio (%).

The regression coefficients indicate that both feedstock composition and biodiesel blend ratio contribute to variations in the calorific value of the biodiesel–diesel blends. The positive coefficient associated with *Jatropha curcas* composition suggests that increasing its proportion in the feedstock produces a slight increase in the predicted calorific value. This finding indicates that blends derived from *Jatropha curcas* tend to possess a marginally higher energy content than those containing a greater proportion of *Azadirachta indica*.

Conversely, the negative coefficient for biodiesel blend ratio shows that the calorific value decreases as the proportion of biodiesel in the blend increases. This trend is consistent with the experimental observations and

reflects the relatively lower heating value of biodiesel compared with conventional petroleum diesel. The larger magnitude of the blend ratio coefficient further demonstrates that biodiesel concentration exerts a greater influence on calorific value than variations in feedstock composition within the experimental range considered. Recent applications of composition-based predictive methods in biodiesel studies have also shown the importance of fuel composition in estimating key fuel properties (Díez-Valbuena et al., 2025). In the present study, the regression equation shows how changes in feedstock composition and biodiesel blend ratio affect the calorific value within the experimental range considered. Its simple form makes it easy to estimate calorific value and can support biodiesel formulation, fuel quality assessment and further modelling studies.

**Table 3: Regression Coefficients and Statistical Significance of the Developed MLR Models**

| Model | Term | Coefficient | Standard Error | t-value | p-value | 95% Confidence Interval |
|---|---|---|---|---|---|---|
| Kinematic viscosity | Intercept | 2.858067 | 0.026066 | 109.647 | <0.001 | 2.804583–2.911550 |
| | Jatropha composition (J) | −0.004193 | 0.000275 | −15.262 | <0.001 | −0.004757 to −0.003630 |
| | Biodiesel blend ratio (B) | 0.046377 | 0.001138 | 40.767 | <0.001 | 0.044043–0.048711 |
| Calorific value | Intercept | 43.977333 | 0.023017 | 1910.654 | <0.001 | 43.930107–44.024560 |
| | Jatropha composition (J) | 0.005467 | 0.000243 | 22.532 | <0.001 | 0.004969–0.005964 |
| | Biodiesel blend ratio (B) | −0.068800 | 0.001005 | −68.489 | <0.001 | −0.070861 to −0.066739 |

The regression analysis showed that both Jatropha composition and biodiesel blend ratio were significantly associated with the two fuel properties investigated ($p < 0.001$). For kinematic viscosity, Jatropha composition had a negative coefficient (−0.004193), indicating a slight decrease in viscosity as the proportion of Jatropha increased, while biodiesel blend ratio had a positive coefficient (0.046377), showing that viscosity increased with increasing biodiesel content. For calorific value, Jatropha composition had a positive coefficient (0.005467), whereas biodiesel blend ratio had a negative coefficient (−0.068800). This indicates that increasing Jatropha content slightly increased the calorific value, while increasing the biodiesel blend ratio reduced it. The 95% confidence intervals for all predictor coefficients excluded zero, further supporting their statistical significance within the experimental range.

The overall regression models were also statistically significant. The kinematic viscosity model gave an F-statistic of 947.43 ($p < 0.001$) and an adjusted $R^2$ of 0.9849, while the calorific value model gave an F-statistic of 2599.22 ($p < 0.001$) and an adjusted $R^2$ of 0.9945. These results show that Jatropha composition and biodiesel blend ratio, when considered together, accounted for a substantial proportion of the variation observed in both fuel properties.

**Performance Evaluation of the Developed MLR Models**

The predictive performance of the developed Multiple Linear Regression (MLR) models was evaluated using the coefficient of determination ($R^2$), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). As shown in Table 4, both models provided a good fit to the experimental data.

The kinematic viscosity model gave an $R^2$ value of 0.9860, indicating that 98.6% of the variation in the experimental data was explained by the model. The calorific value model gave a slightly higher $R^2$ value of 0.9948, showing a closer agreement between the experimental and predicted values.

The low RMSE and MAE values also show that the differences between the experimental and predicted values were small across the feedstock compositions and biodiesel blend ratios investigated. Of the two models, the calorific value model showed slightly better predictive performance, as indicated by its higher $R^2$ value and lower prediction errors.

The LOOCV results provided further support for the performance of the developed models. The $R^2CV$ values of 0.9810 for kinematic viscosity and 0.9932 for calorific value were only slightly lower than the corresponding $R^2$ values obtained from the full dataset. Similarly, the small increases in RMSE and MAE under LOOCV show that both models maintained good predictive performance when individual observations were excluded from model fitting.

**Table 4: Performance Statistics of the Developed MLR Models**

| Performance Metric | Kinematic Viscosity | Calorific Value |
|---|---|---|
| $R^2$ | 0.9860 | 0.9948 |
| Adjusted $R^2$ | 0.9849 | 0.9945 |
| RMSE | 0.0505 | 0.0446 |
| MAE | 0.0413 | 0.0356 |

**Table 5: LOOCV Performance of the Developed MLR Models**

| Property | $R^2CV$ | RMSECV | MAECV |
|---|---|---|---|
| Kinematic viscosity | 0.9810 | 0.0587 mm² s⁻¹ | 0.0471 mm² s⁻¹ |
| Calorific value | 0.9932 | 0.0513 MJ kg⁻¹ | 0.0403 MJ kg⁻¹ |

The LOOCV results were generally consistent with those obtained from the full dataset. The kinematic viscosity model gave an $R^2CV$ of 0.9810, with RMSECV and MAECV values of 0.0587 and 0.0471 mm² s⁻¹, respectively. The calorific value model gave an $R^2CV$ of 0.9932, with corresponding RMSECV and MAECV values of 0.0513 and 0.0403 MJ kg⁻¹. The closeness of these values to the full-dataset results indicates that both models maintained good predictive performance when individual observations were excluded during cross-validation.

The predictive performance of the developed Multiple Linear Regression (MLR) models was further evaluated by comparing the experimental and predicted values of kinematic viscosity and calorific value, as presented in Figure 2(a) and (b).

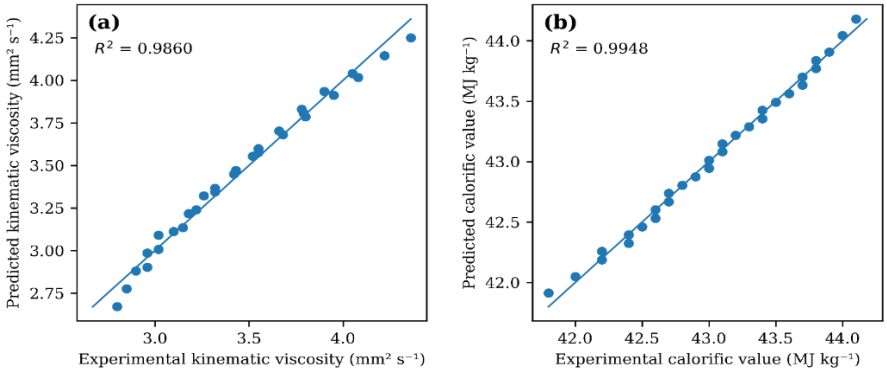


Figure 2: Comparison of the Experimental and Predicted Values Obtained Using the Developed MLR Models: (a) Kinematic Viscosity and (b) Calorific Value

Figure 2(a) and (b) show close agreement between the experimental and predicted values for both response variables. The clustering of the data points around the 1:1 reference line indicates that the developed MLR models adequately captured the observed variations in kinematic viscosity and calorific value, consistent with the performance statistics presented in Table 4. Similar agreement between experimental and predicted values has been reported in recent predictive modelling studies (Otunla & Ayegboyin, 2026; Bukkarapu & Krishnasamy, 2024a). This observation is also consistent with recent findings showing that data-driven models can capture composition-dependent variations in the physical properties of biodiesel-containing fuel blends (Yeşilova et al., 2025).

To further evaluate the adequacy of the developed Multiple Linear Regression (MLR) models, residuals were plotted against the corresponding fitted values. The residual plots for kinematic viscosity and calorific value are presented in Figure 3(a) and (b), respectively.

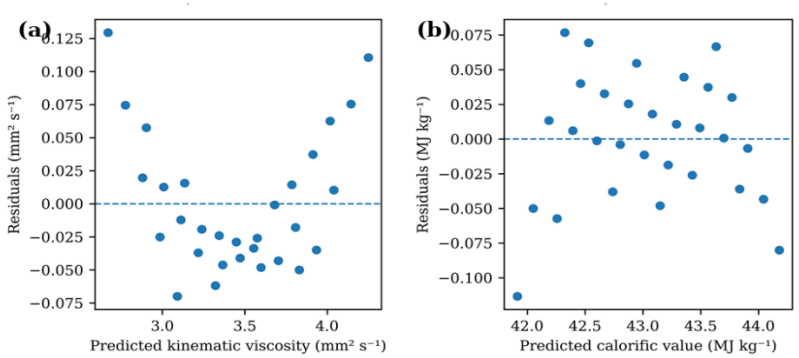


Figure 3: Residuals Versus Fitted Values for the Developed MLR Models: (a) Kinematic Viscosity and (b) Calorific Value

Figure 3(a) and (b) show that the residuals are generally scattered around the zero reference line without any clear pattern or noticeable change in spread. This suggests that the error variance remained reasonably constant across the predicted values. The Breusch–Pagan test supports this observation, as the results were non-significant for both kinematic viscosity (LM = 0.1791, p = 0.9144) and calorific value (LM = 2.3979, p = 0.3015), indicating no evidence of heteroscedasticity. Similar residual patterns have been reported in biodiesel modelling studies as an indication of satisfactory model behaviour (Kodua et al., 2024; Zohmingliana et al., 2024).

Residual normality was also examined using the Q–Q plots presented in Figure 4(a) and (b), together with the Shapiro–Wilk test.

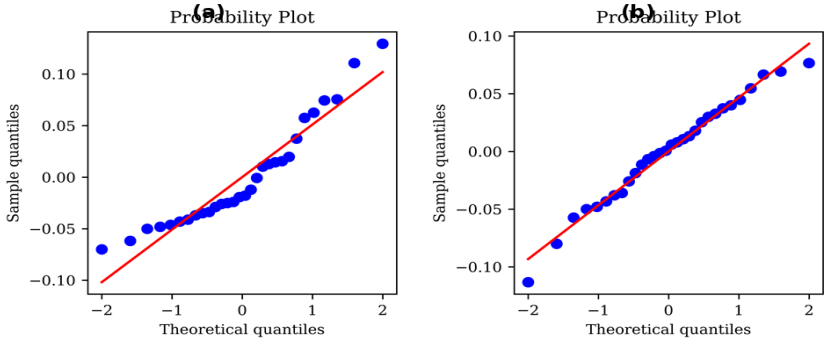


Figure 4: Normal Probability (Q–Q) Plots of the Residuals for the Developed MLR Models: (a) Kinematic Viscosity and (b) Calorific Value

The kinematic-viscosity residuals showed noticeable curvature and some departure from the reference line, particularly toward the tails. This was supported by the Shapiro–Wilk result (W = 0.9093, p = 0.0143), indicating a significant departure from normality. In contrast, the calorific-value residuals followed the reference line more closely, with the Shapiro–Wilk test showing no significant departure from normality (W = 0.9807, p = 0.8448). Overall, both models showed constant error variance, although the normality assumption was better satisfied for the calorific-value model than for the kinematic-viscosity model.

**Practical Implications of the Developed Models**

The regression models developed in this study provide a simple means of estimating the kinematic viscosity and calorific value of biodiesel–diesel blends prepared from pure and hybrid *Jatropha curcas* and *Azadirachta indica* feedstocks. By relating these properties to feedstock composition and biodiesel blend ratio, the equations allow the two fuel properties to be estimated without repeated laboratory measurements for every blend formulation. Similar applications of predictive modelling have also shown the usefulness of data-driven approaches for estimating biodiesel thermophysical properties and reducing the experimental effort required for fuel characterization (Hussain et al., 2025).

The close agreement between the predicted and experimental values shows that the proposed equations can be useful for preliminary fuel formulation and quality assessment within the experimental range investigated. Their simple mathematical form also makes them easy to apply using standard statistical packages or spreadsheet software, making them suitable for routine engineering analysis and related research applications.

Furthermore, the regression equations developed in this study provide a useful basis for extending predictive modelling to other biodiesel feedstocks and blend formulations. They can also serve as reference models for comparing more advanced predictive approaches, including Artificial Neural Networks and other machine-learning techniques (Bukkarapu & Krishnasamy, 2024b; Mwenge & Rutto, 2025). The findings therefore show that relatively simple statistical models can provide useful estimates of key fuel properties while reducing the time, cost and experimental effort associated with repeated laboratory measurements.

The developed MLR models are limited to the conditions investigated in this study, covering *Jatropha curcas* composition (J) from 0 to 100% and biodiesel blend ratio (B) from 5 to 30%. Since the models were not evaluated outside these ranges, predictions beyond these conditions should be treated with caution.

## CONCLUSION

This study developed Multiple Linear Regression (MLR) models for predicting the kinematic viscosity and calorific value of biodiesel–diesel blends produced from pure and hybrid *Jatropha curcas* and *Azadirachta indica* feedstocks. Using Jatropha composition and biodiesel blend ratio as predictors, the models clearly described how changes in blend composition influenced the two fuel properties.

The high coefficients of determination and low RMSE and MAE values, together with the LOOCV results, showed that the models provided good predictive performance within the experimental conditions investigated. The regression equations are simple to apply and can support the estimation of kinematic viscosity and calorific value during biodiesel formulation, fuel quality assessment and preliminary engineering analysis.

The models are limited to the experimental ranges considered in this study, covering *Jatropha curcas* composition (J) from 0 to 100% and biodiesel blend ratio (B) from 5 to 30%. Since their performance outside these ranges was not evaluated, predictions beyond these conditions should be treated with caution. Future studies could examine other biodiesel feedstocks, wider blend ranges and larger independent datasets to further assess the applicability of the models.

## REFERENCES

Adejuwon, S. O., Ogunlana, R., Aremu, O. A., & Makinde, O. S. (2026). Artificial neural network modelling of performance and emissions in turbulent biodiesel combustion within a compression ignition engine. *Nigerian Journal of Physics, 35*(1), 210–219. https://doi.org/10.62292/njp.v35i1.2026.491

Amenaghawon, A. N., Omede, M. O., Ogbebor, G. O., Eshiemogie, S. A., Igemhokhai, S., Evbarunegbe, N. I., Ayere, J. E., Osahon, B. E., Oyefolu, P. K., Eshiemogie, S. O., Anyalewechi, C. L., Okedi, M. O., Chinemerem, B. A., Kusuma, H. S., Darmokoesoemo, H., & Okoduwa, I. G. (2024). Optimized biodiesel synthesis from an optimally formulated ternary feedstock blend via machine learning-informed methanolysis using a composite biobased catalyst. *Bioresource Technology Reports, 25*, 101805. https://doi.org/10.1016/j.biteb.2024.101805

Arif, M., Alalawy, A. I., Zheng, Y., Koutb, M., Kareri, T., Salama, E.-S., & Li, X. (2025). Artificial intelligence and machine learning models application in biodiesel optimization process and fuel properties prediction. *Sustainable Energy Technologies and Assessments, 73*, 104097. https://doi.org/10.1016/j.seta.2024.104097

ASTM International. (2024). *ASTM D445-24: Standard test method for kinematic viscosity of transparent and opaque liquids (and calculation of dynamic viscosity).* ASTM International. https://doi.org/10.1520/D0445-24

Atsuwe, B. A., Atser, R. A., & Okoh, T. (2025). Ethanol as renewable energy for powering telecommunications devices in Benue State: A comprehensive review. *Nigerian Journal of Physics, 34*(2). https://doi.org/10.62292/njp.v34i2.2025.367

Bukkarapu, K. R., & Krishnasamy, A. (2024a). *Biodiesel composition based machine learning approaches to predict engine fuel properties*. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 238*(7), 1844–1860. https://doi.org/10.1177/09544070231158240

Bukkarapu, K. R., & Krishnasamy, A. (2024b). *Investigations on the applicability of machine learning algorithms to optimize biodiesel composition for improved engine fuel properties*. *International Journal of Engine Research, 25*(7), 1299–1314. https://doi.org/10.1177/14680874241227540

Díez Valbuena, G., García Tuero, A., Díez, J., Rodríguez, E., & Hernández Battez, A. (2024). Application of machine learning techniques to predict biodiesel iodine value. *Energy, 292*, 130638. https://doi.org/10.1016/j.energy.2024.130638

Díez-Valbuena, G., García-Tuero, A., Rodríguez, E., Hernández Battez, A., & Díez, J. (2025). Modeling biodiesel properties by preference learning: Case study of cetane number. *Engineering Applications of Artificial Intelligence, 144*, 110107. https://doi.org/10.1016/j.engappai.2025.110107

Hussain, S. S., Ali, S. A., Husain, D., & Sharma, M. (2025). A machine learning model for the computation of thermophysical properties of WCO biodiesel mixed with multiwalled carbon nanotubes. *Science and Technology for Energy Transition*, *80*, 40. https://doi.org/10.2516/stet/2025021

Ishola, N. B., Epelle, E. I., & Betiku, E. (2024). Machine learning approaches to modeling and optimization of biodiesel production systems: State of art and future outlook. *Energy Conversion and Management: X*, *23*, 100669. https://doi.org/10.1016/j.ecmx.2024.100669

Kodua, S. T., Boadi, N. O., Badu, M., & Al Akraa, I. M. (2024). Optimization of the reaction conditions in biodiesel production: The case of baobab seed oil as alternative feedstock. *Journal of Chemistry*, *2024*, Article 1498240. https://doi.org/10.1155/2024/1498240

Mwenge, P., & Rutto, H. (2025). Machine learning-based predictive modelling of biodiesel production from animal fats catalysed by a blast furnace slag geopolymer. *Results in Engineering*, *25*, 104126. https://doi.org/10.1016/j.rineng.2025.104126

Osman, A. I., Nasr, M., Farghali, M., Rashwan, A. K., Abdelkader, A., Al-Muhtaseb, A. A. H., Ihara, I., & Rooney, D. W. (2024). Optimizing biodiesel production from waste with computational chemistry, machine learning and policy insights: A review. *Environmental Chemistry Letters*, *22*, 1005–1071. https://doi.org/10.1007/s10311-024-01700-y

Otunla, T. A., & Ayegboyin, E. O. (2026). Day-ahead hourly forecasting of solar radiation using a physics-based hybrid machine learning model in selected locations in Nigeria. *Nigerian Journal of Physics*, *35*(2). https://doi.org/10.62292/njp.v35i2.2026.503

Riayatsyah, T. M. I., Sebayang, A. H., Silitonga, A. S., Padli, Y., Fattah, I. M. R., Kusumo, F., Ong, H. C., & Mahlia, T. M. I. (2022). Current progress of *Jatropha curcas* commoditisation as biodiesel feedstock: A comprehensive review. *Frontiers in Energy Research*, *9*, 815416. https://doi.org/10.3389/fenrg.2021.815416

Santos, S. M., Maciel, M. R. W., & Fregolente, L. V. (2022). Application of multivariate exploratory techniques to predict kinematic viscosity of biodiesel from vegetable and algae oils. *Chemical Engineering Transactions*, *92*, 739–744. https://doi.org/10.3303/CET2292124

Sheu, A. L., Alade, M. O., Olabisi, O., Adewumi, A. S., Aremu, O. A., & Azeez, I. A. (2026). Modeling tropospheric effects on mobile network performance using KPI-based metrics. *Nigerian Journal of Physics*, *35*(1), 73–83. https://doi.org/10.62292/njp.v35i1.2026.479

Uwaoma, C. J., Ugwu, J. U., Joseph, U., & Nduaka, C. E. (2025). Specific heat capacity, density and viscosity determined using different combinations of coconut oil and palm kernel oil. *Nigerian Journal of Physics*, *34*(2), 194–209. https://doi.org/10.62292/njp.v34i2.2025.443

Vilas Bôas, R. N., & Mendes, M. F. (2022). A review of biodiesel production from non-edible raw materials using the transesterification process with a focus on the influence of feedstock composition and free fatty acids. *Journal of the Chilean Chemical Society*, *67*(1), 5433–5444. https://doi.org/10.4067/S0717-97072022000105433

Xiao, H., Wang, W., Bao, H., Li, F., & Zhou, L. (2023). Biodiesel–diesel blend optimized via leave-one cross-validation based on kinematic viscosity, calorific value, and flash point. *Industrial Crops and Products*, *191*, 115914. https://doi.org/10.1016/j.indcrop.2022.115914

Yahya, S. I., & Aghel, B. (2021). Estimation of kinematic viscosity of biodiesel–diesel blends: Comparison among accuracy of intelligent and empirical paradigms. *Renewable Energy*, *177*, 318–326. https://doi.org/10.1016/j.renene.2021.05.092

Yeşilova, K., Yücel, Ö., & Ergan, B. T. (2025). Modeling prediction of physical properties in sustainable biodiesel–diesel–alcohol blends via experimental methods and machine learning. *Processes*, *13*(7), 2310. https://doi.org/10.3390/pr13072310

Zohmingliana, H., Ruatpuia, J. V. L., Anal, J. M. H., Chai, F., Halder, G., Dhakshinamoorthy, A., & Rokhum, S. L. (2024). Metal-organic framework-derived solid catalyst for biodiesel production from *Jatropha curcas* oil: Kinetic study and process optimization using response surface methodology. *International Journal of Energy Research*, *2024*, Article 6336617. https://doi.org/10.1155/2024/6336617