Article citation info:
Naryanto RF, Setiadi R, Prabantara DR, Hadromi H. Diesel engine performance prediction using fuel blends with a support vector machine. Combustion Engines. 0000;XXX(X):xx-xx. https://doi.org/10.19206/CE-220543

Rizqi Fitri NARYANTO
Rizki SETIADI
Diffa Rakhman PRABANTARA
Hadromi HADROMI

Combustion Engines
Polish Scientific Society of Combustion Engines

Diesel engine performance prediction using fuel blends with a support vector machine

ARTICLE INFO

Received: 25 November 2025
Revised: 5 April 2026
Accepted: 10 April 2026
Available online: 6 May 2026

The urgent need for environmentally friendly alternative fuels arises from concerns over fossil fuel depletion and the harmful effects of exhaust gas emissions. One promising solution involves blending biodiesel with natural additives, such as essential oils, to improve combustion efficiency in diesel engines. However, conventional engine performance testing is often time-consuming and repetitive, requiring a more efficient approach such as predictive modeling using Machine Learning (ML). This study investigates the effect of biodiesel and essential oil blend ratios on engine performance and develops predictive models to estimate engine power and torque. A total of 1045 experimental data points were collected from tests using varying compositions of biosolar, Dexlite, and essential oil additives at different RPMs. The research applied a quantitative method, utilizing Linear Regression and Support Vector Machine (SVM) algorithms implemented in RStudio. Model accuracy was evaluated using MAE, RMSE, and R^2 metrics. The results indicate that variations in biodiesel blend ratios and essential oil additives had no statistically significant effect on engine performance. However, ML-based predictive modeling proved highly effective. The SVM model achieved superior accuracy, with R^2 values of 0.993 for power and 0.973 for torque. In contrast, the Linear Regression model yielded much lower R^2 values – 0.671 for power and 0.093 for torque. These findings demonstrate the potential of Machine Learning, particularly SVM, as a reliable tool for predicting diesel engine performance with biodiesel and additive blends, offering a faster, more accurate alternative to conventional testing methods.

Key words: *diesel engine, performance, linear regression, support vector machine, prediction*

1. Introduction

Fossil fuels, such as coal, crude oil, and natural gas, have historically been a critical component of the global energy infrastructure, accounting for 81% of the world's primary energy supply in 2019, according to the International Energy Agency [22]. In Indonesia, fuel oil consumption increased by 3.19% in the third quarter of 2021, reaching 48.59×10^6 m^3 compared to the same period the previous year [28]. While these fuels have driven economic growth and urban development, their use has led to significant environmental issues, primarily the rise in carbon dioxide emissions that fuels climate change [27]. This urgency, coupled with the need for energy security, has spurred discussions about the future of energy sources. Renewable energies like solar, wind, hydro, geothermal, and biomass are playing a vital role in this transition, representing a shift toward sustainable energy production and a balanced relationship with the environment [5, 24, 29].

As industries remain dependent on power generation and society on transportation and fossil fuel-based products, the search for alternative energy sources continues. Biodiesel, known as fatty acid methyl ester (FAME), has garnered attention as a viable alternative to traditional fuels, particularly for transportation, due to its potential to reduce greenhouse gas emissions [7]. Biodiesel can reduce carbon monoxide (CO) emissions by 8–41% [9]. This renewable fuel can be produced from various biological materials, including used cooking oil, vegetable seed oils, animal fats, and microalgal oil [11]. Importantly, modern diesel engines can operate on biodiesel without requiring modifications [1]. Researchers recommend using biodiesel as a full or partial substitute for petroleum diesel, as it effectively reduces exhaust emissions [6, 12].

Challenges in the transportation sector include higher nitrogen oxide emissions compared to diesel, reduced engine performance in cold weather, sediment deposition that can clog storage tanks and fuel lines, and lower fuel density, which necessitates greater fuel consumption for equivalent power output [9]. The physicochemical properties of biodiesel can vary significantly depending on the feedstock and production method, thereby directly affecting its performance in diesel engines [7]. Key properties such as viscosity, density, calorific value, and oxidation stability affect combustion efficiency, engine start-up performance, fuel consumption, emission characteristics, and long-term reliability [6].

In a diesel engine, the combustion process of biodiesel influences thermal efficiency, power output, and exhaust gas emissions. The higher oxygen content in biodiesel promotes more complete combustion, leading to increased Brake Thermal Efficiency (BTE) and effectively reducing emissions from four-stroke compression engines, particularly unburned hydrocarbons, carbon monoxide, and nitrogen oxides [6]. Therefore, it is essential to consider both the physicochemical properties and combustion characteristics of biodiesel during engine optimization, and to establish accurate predictive models to evaluate its overall performance [13]. Determining optimal biodiesel composition through experimental methods is time-consuming and labor-intensive. Advances in computational approaches, particularly computational fluid dynamics (CFD) and Machine Learning (ML), enable faster and more reliable prediction of biodiesel properties [1, 7]. These methods pro-

vide a robust basis for optimizing biodiesel formulations, improving efficiency and engine performance [9, 17, 21, 32].

The development of predictive models plays a key role in improving production efficiency, product quality, scalability, cost, and sustainability in modern industrial applications [6]. However, their effectiveness depends strongly on the selected algorithm and data characteristics. Linear regression offers simplicity and interpretability but is limited to linear relationships. Support Vector Machines (SVMs) perform well on small to medium datasets and can handle non-linear patterns via kernel functions, though they require careful parameter tuning. In contrast, Artificial Neural Networks (ANNs) are better suited for capturing complex non-linear relationships between biodiesel properties and engine performance, but they require large datasets and higher computational costs [2, 8]. Therefore, the choice of model depends on criteria such as data size, nonlinearity, interpretability, and computational resources. By selecting appropriate methods, Machine Learning models can accurately predict key fuel properties, including calorific value, viscosity, and oxidation stability, enabling more efficient optimization of biodiesel production [1].

2. Experimental

2.1. Data source and experimental setup

This study utilized a quantitative approach based on a dataset derived from primary experiments. The data was sourced from engine performance tests conducted at the Engine Performance Laboratory, Department of Mechanical Engineering, Universitas Negeri Semarang, Indonesia. The engine specifications are shown in Table 1. In this study, the diesel engine is a vehicle engine. The engine's fuel system is not naturally aspirated (NA), but rather uses a common-rail system, a modern fuel-injection system in which diesel fuel is stored in a single high-pressure pipe (rail) and distributed electronically to each injector. This system allows for more precise injection control, increasing efficiency and power while reducing emissions.

Table 1. Diesel engine specifications in vehicles

Engine model	Toyota Diesel 2 KD-FTV
Engine type	4 stroke, in-line type, 16 valve DOHC
Cylinder number	4-cylinder
Injection system	Common Rail
Displacement	2494 cm^3
Bore and stroke	92 mm × 93.8 mm
Compression ratio	18.5:1

The population for this study consisted of all experimental data on diesel engine performance using biodiesel (biosolar and Dexlite) fuel blends and an essential oil additive. The fuel was mixed in a stationary-bed reactor at 300 rpm and 60°C for 30 minutes. The fuel mixture is used in the vehicle and then tested for performance, namely, torque and power, on a chassis dynamometer.

A total of 1045 data samples were selected for the analysis using purposive sampling. The selection criteria required the data to include parameters for rpm-engine power and rpm-torque. Data analysis using R Studio. The research variables were defined as follows:

– Independent Variables: The primary independent variables were fuel process, consisting of ten biodiesel–Dexlite blend ratios with and without essential oil additives. The 10 fuel blend variations used in the experiments are detailed in Table 2. The table shows the fuel used to collect engine performance data. A total of 1,045 data points were collected from performance tests, including rpm and engine power. Density testing uses the ASTM D1298 standard, and viscosity testing uses the ASTM D445 standard
– Dependent Variables: The dependent variables were the key engine performance metrics: torque [Nm] and power [kW]
– Control Variables: To ensure consistency across experiments, the test engine, engine speed [rpm], and engine load were held constant.

Table 2. Composition of experimental fuel blends

Blend No.	Biodiesel [%]	Dexlite [%]	Essential oil additive [%]	Density [kg/m^3]	Viscosity [cSt]
1	90	10	0.015	922.7	5.035
2	80	20	0.015	930.7	4.934
3	70	30	0.015	925.3	4.755
4	60	40	0.015	928	4.425
5	50	50	0.015	926.7	4.336
6	90	10	0	934.7	5.328
7	80	20	0	932.7	5.341
8	70	30	0	930.7	5.328
9	60	40	0	928	5.226
10	50	50	0	926.7	5.022

2.2. Machine learning approach

A predictive modeling workflow was designed to estimate engine performance based on the fuel composition. The procedure involved data preprocessing, model training, and evaluation. To provide a clearer overview of the research workflow, a flowchart is presented in Fig. 1. The raw experimental dataset underwent several preprocessing steps to prepare it for Machine Learning analysis. This included data cleaning to handle any incomplete or irrelevant entries and normalization to scale the numerical features to a uniform range. The dataset was then partitioned into 75% for training and 25% for testing and validation.

To guarantee that the visualization accurately and distortion-free depicts the experimental results, the labels and data representation were kept clear. Repeated measurements were used to ensure the dataset's reliability; each test was performed three times under carefully controlled operational conditions to ensure consistency. The alignment of the observed trends with recognized engine performance metrics further supports the data's veracity. The non-linear relationship between input parameters (fuel mix ratio, engine speed, and additive condition) and output performance (power and torque) was modeled using support vector regression (SVR). The model's efficacy in precisely estimating engine performance was confirmed by evaluating its prediction performance using RMSE, MAE, and R^2.

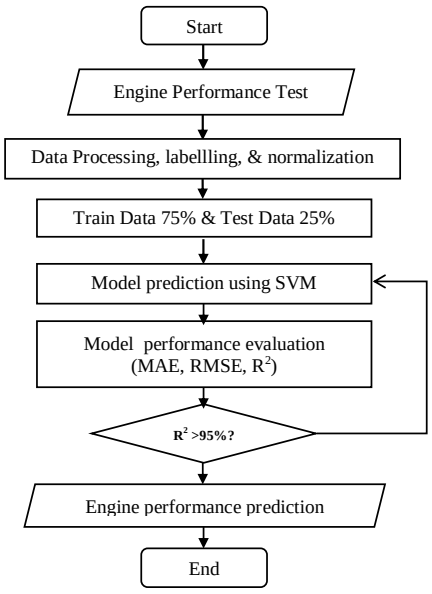

Fig. 1. Machine learning workflow for engine performance prediction

2.3. Support vector regression (SVR) model

A supervised learning approach called Support Vector Regression (SVR) applies Support Vector Machines to regression issues. Finding a function f(x) that minimizes model complexity while approximating the target values within a given tolerance level – defined by the ε-insensitive loss function – is the goal of SVR. By defining a tolerance margin in which errors are not penalized, the parameter ε enables the model to ignore minor deviations and thereby improve generalization.

The fundamental goal is to find a regression function that is as flat as possible. For non-linear data, this function takes the form:

$$f(x) = w^T x + b \tag{1}$$

Here, w^T represents the weight vector, b is the bias term, and x is a non-linear function that maps the input features, x, into a higher-dimensional space where the relationship may become linear. Instead of solving for w directly in the high-dimensional space, SVR solves a more manageable convex optimization problem known as the dual problem. This is achieved using Lagrange multipliers, where the goal is to maximize the following objective function, L, with respect to the Lagrange multipliers α_i and α_i^*:

$$L = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) - \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) \tag{2}$$

This maximization is performed under the following constraints:

$$\sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \tag{3}$$

where, $0 \le \alpha_i, \alpha_i^* \le C, \text{and for } i = 1, \dots, N$. In this formulation, C is a user-defined hyperparameter that controls the penalty for data points that lie outside the $\epsilon - \text{tube}$, balancing model complexity against prediction error.

The term $K(x_i, x_j)$ represents the Kernel function, a crucial component that calculates the similarity or inner product between two data points, x_i and x_j, in the higher-dimensional space without explicitly computing the transformation x. This study utilized the Radial Basis Function (RBF) kernel, which is defined as:

$$K(x_i, x_j) = \exp\left(-\gamma\left\|x_i - x_j\right\|^2\right) \tag{4}$$

Here, γ is a kernel hyperparameter that defines the influence of a single training example. After solving the optimization problem, the data points for which the corresponding Lagrange multipliers (α_i or α_i^*) are non-zero and are designated as Support Vectors (SV). These are the only data points that contribute to the final model and define the boundaries of the regression function.

Once the non-zero Lagrange multipliers are found, the bias term, b, can be determined. This is achieved by selecting a Support Vector, x_k, that lies on the boundary of the $\epsilon - \text{tube}$ (i.e., its corresponding multiplier satisfies $0 < \alpha_k < C$) and solving for b using the following relationship:

$$b = y_k - \sum_{i \in SVs}(\alpha_i - \alpha_i^*)K(x_i, x_k) \pm \tag{5}$$

The choice of using $+\epsilon \; or \; -\epsilon$ depends on whether the support vector lies on the upper or lower boundary of the margin.

With the Support Vectors, their corresponding weights ($\alpha_i - \alpha_i^*$), and the bias term (b) is determined, the model is complete. The prediction for a new, unseen data point, xnew, is then calculated using the final SVR function:

$$f(x_{new}) = \sum_{i \in SVs}(\alpha_i - \alpha_i^*)K(x_i, x_{new}) + b \tag{6}$$

This function generates a prediction by computing a weighted sum of the similarities (as defined by the kernel function) between the new data point and each established Support Vector.

2.4. Linear regression model

Linear regression is a fundamental statistical and machine learning algorithm used to model the relationship between a dependent variable and one or more independent variables. In this study, Multiple Linear Regression (MLR) was employed as a baseline model to predict engine performance. The core assumption of MLR is that there is a linear relationship between the input features and the output variable. The following equation represents the mathematical model for Multiple Linear Regression:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon \tag{7}$$

where: Y is the dependent (target) variable, representing either engine power or torque; $X_1, X_2, \dots, X_p$ are the independent (predictor) variables, which include fuel composition and engine speed (RPM); β_0 is the intercept, representing the predicted value of Y when all predictor variables are zero; $\beta_1, \beta_2, \dots, \beta_p$ are the regression coefficients, each coefficient β_j represents the expected change in Y for a one-unit change in the corresponding predictor X_j, while holding all other predictors constant; ϵ is the model's error term, which captures the random variation in Y that cannot be explained by the linear relationship with the predictors.

The primary goal during the model training phase is to estimate the optimal values for the coefficients

$(\beta_0, \beta_1, \ldots, \beta_p)$. This is typically achieved using the Ordinary Least Squares method. OLS works by finding the coefficients that minimize the sum of the squared differences between the observed actual values (y_i) and the values predicted by the model ($\hat{y}i$). This is also known as minimizing the Residual Sum of Squares (RSS):

$$\text{Minimize RSS} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\Big(y_i - \big(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}\big)\Big)^2 \quad (8)$$

Once the optimal coefficients $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)$ have been estimated from the training data, the model can be used to predict new data. The prediction function is simply the regression equation without the error term:

$$\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1}X_1 + \widehat{\beta_2}X_2 + \cdots + \widehat{\beta_p}X_p \quad (9)$$

While computationally simple and highly interpretable, Linear Regression's performance is fundamentally limited by its core assumption of linearity. It is best suited for problems where the underlying relationship between inputs and outputs is indeed linear.

2.5. Model evaluation

The performance of the trained models was assessed using K-Fold cross-validation to ensure generalization and prevent overfitting. The predictive accuracy of the Linear Regression and SVM models was quantitatively compared using the following standard regression metrics:

– Mean Absolute Error (MAE): Measures the average absolute difference between the predicted and actual values

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (10)$$

– Root Mean Squared Error (RMSE): Represents the square root of the average of squared differences, providing an error measure in the same units as the target variable

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (11)$$

– R-Squared (R^2): Indicates the proportion of the variance in the dependent variable that is predictable from the independent variables

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (12)$$

3. Results

This section presents the performance evaluation of the linear regression and support vector machine (SVM) models for predicting engine power and torque. The analysis includes a quantitative comparison using standard metrics, a visual inspection of predictive accuracy, and a discussion of the results in the context of existing literature. The results of the engine performance test are shown in Fig. 2. The graphs and data show that engine power increases with increasing revs from 2000 to approximately 5000 rpm, then decreases at higher revs (≥ 5200 rpm) for all fuel variations. Lowest power occurs at 2000 rpm, while maximum power is in the range of 66 to 70 kW, with variations 6 to 10 generally producing higher power than variations 1 to 5 across most of the rev range.

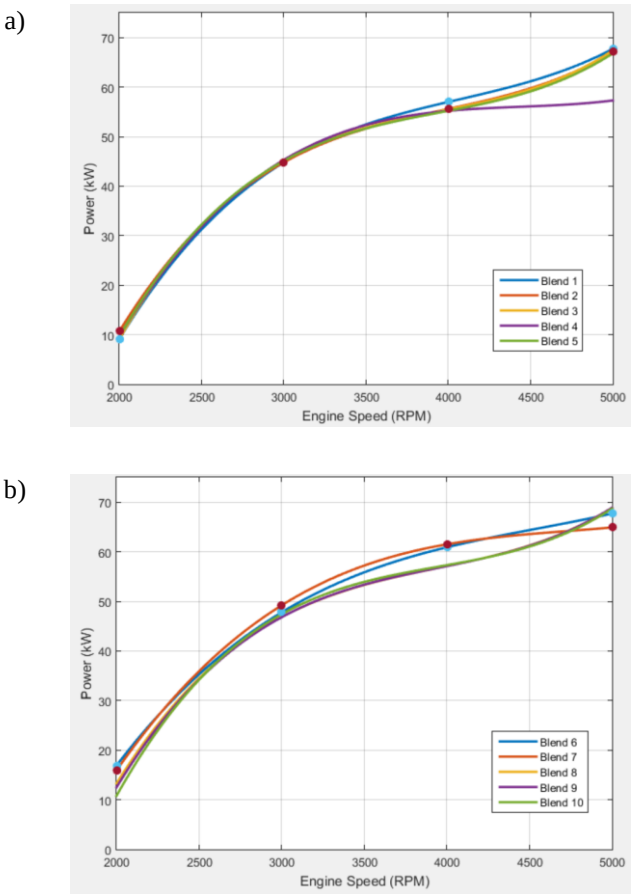

Fig. 2. Engine power trends with rising engine speed for various fuel blends: (a) blends 1–5; (b) blends 6–10

3.1. Statistical analysis of fuel blend effects

To assess the impact of fuel composition on engine performance, a series of nonparametric statistical tests was conducted. Assumption tests, including the Shapiro-Wilk test for normality and Levene's test for homogeneity of variances, were performed first to determine the appropriate inferential tests. The data for both power and torque were found not to be normally distributed ($p > 0.001$), necessitating the use of nonparametric methods. A summary of the key statistical test results is provided in Table 3.

The Wilcoxon rank-sum test confirmed these observations, yielding p-values of 0.732 for power and 0.840 for torque, both well above the 0.05 significance threshold. Furthermore, the calculated effect sizes were negligible for both power (Cohen's d = –0.016) and torque (Cohen's d = 0.002), indicating that the additive's practical impact was insignificant. The boxplot visualizations in Fig. 3 and Fig. 4 further illustrate the similarity in data distribution between the two groups, supporting the statistical findings.

Table 3. Summary of statistical test results

Analysis	Variable	Test	p-value	Finding
Additive Effect	Power (BHP)	Wilcoxon	0.732	Not Significant
	Torque (Nm)	Wilcoxon	0.840	Not Significant
Blend Ratio Effect	Power (BHP)	Kruskal-Wallis	0.782	Not Significant
	Torque (Nm)	Kruskal-Wallis	0.0504	Not Significant

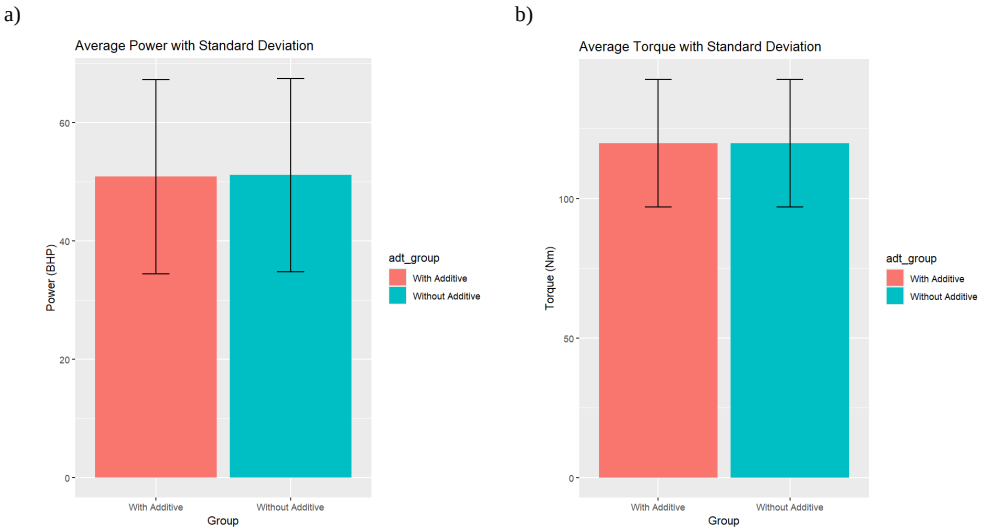

Fig. 3. Statistical comparison of engine performance parameters, showing mean power (a) and torque (b) with standard deviation for fuel blends under additive and non-additive conditions

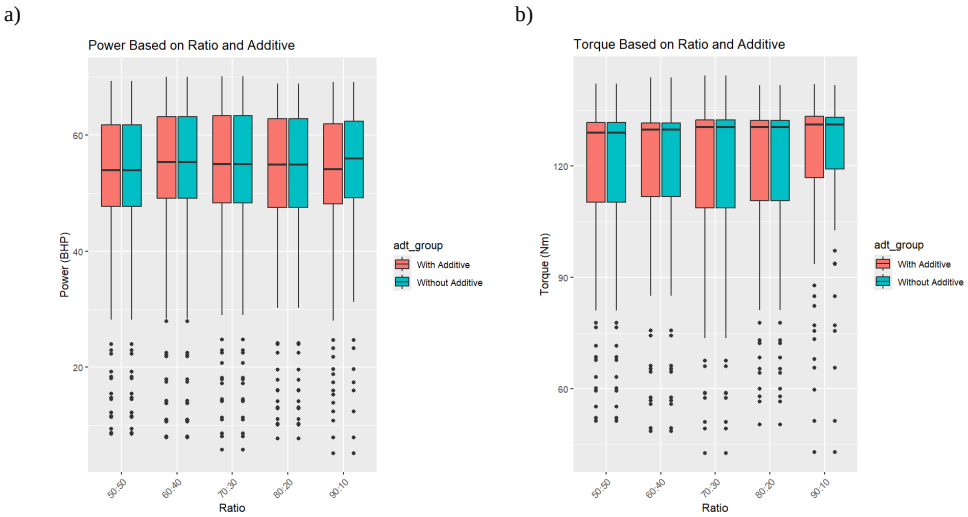

Fig. 4. Statistical distribution of engine performance parameters, showing boxplot comparisons of power (a) and torque (b) across fuel blend ratios under additive and non-additive conditions

This lack of a significant effect aligns with previous research. For example, [3] noted that the addition of essential oils to biodiesel does not consistently improve diesel engine performance across all conditions. Similarly, [4, 19, 20], stated that the effectiveness of essential oil bio-additives is highly dependent on the dosage and their compatibility with the base fuel. The findings are also consistent with [14, 25], who argued that many additives have a more pronounced effect on thermal efficiency and emissivity than on primary performance metrics such as power and torque.

The influence of varying the biodiesel-dexlite blend ratio was also found to be not statistically significant. The Kruskal-Wallis test yielded a p-value of 0.782, indicating no significant differences among the five blend ratios. For torque, the test yielded a p-value of 0.0504, which is marginally significant. While this borderline result might suggest a subtle trend, it is not statistically strong enough to reject the null hypothesis at a 95% confidence level. The boxplots in Fig. 4 and Fig. 5, which detail performance by ratio, visually support these findings. No clear trend or significant difference in the distribution of power or torque is observable across the different blend ratios. These results are consistent with prior studies [18, 26]. Both found that variations in biodiesel blend ratios (from B20 to B100) often lead to minor, inconsistent fluctuations in power that are not statistically significant. Likewise [15] observed no significant power trend when testing various corn biodiesel blends.

Across various fuel blend percentages, the boxplots show non-linear trends, variability, and outliers. Support Vector Regression (SVR), which is ideal for handling data variability and modeling non-linear relationships, is justified by these features. The significance of adding an additive as a crucial input parameter in the predictive model is further shown by the substantial difference between additive and non-additive situations. In this situation, the figure is a crucial exploratory data analysis tool for identifying influential variables, analyzing data distributions, and assessing whether SVR is appropriate for modeling. It supports the creation of a reliable SVR model for precise diesel engine performance prediction by providing a basic understanding of the relationships between input variables (fuel mix ratio and additive) and output responses (torque and power).

3.2. Model performance evaluation

The predictive accuracy of each model was evaluated on the test dataset using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). To predict the two distinct engine performance parameters, separate models were trained for each target output. A summary of the evaluation metrics is presented in Table 4.

The results clearly indicate that the SVM model significantly outperformed the Linear Regression model for both prediction tasks. The SVM model demonstrated superior performance, with an MAE of 1.119 and an RMSE of 1.490, approximately 7–8 times lower than the errors from the Linear Regression model. Critically, the SVM achieved an R^2 of 0.993, indicating that it can explain 99.3% of the variance in the engine power data. In contrast, the Linear Regression model's R^2 of 0.671 suggests a much weaker fit, explaining only 67.1% of the variance. For Torque Prediction: The performance gap between the models was even more pronounced. The SVM model yielded an R^2 of 0.974, whereas the Linear Regression model yielded an R^2 of 0.093. This extremely low R^2 value indicates that the linear model almost completely failed to capture the underlying patterns in the torque data.

3.3. Visual analysis of predictive accuracy

To further interpret the models' performance, the predicted values were plotted against the actual experimental values. In these plots, the dashed diagonal line ($y = x$) represents a perfect prediction. The visual results, shown in Fig. 5 and Fig. 6, corroborate the findings from the evaluation metrics. The SVM predictions for both power and torque show data points tightly clustered around the identity line, confirming high accuracy across the entire range of values. Conversely, the Linear Regression plots exhibit significant deviations. For power prediction, the points form a curved pattern that deviates from the diagonal, indicating the model's inability to capture the nonlinearity. For torque prediction (Fig. 9), the model fails, with predictions scattered far from the ideal line, consistent with its near-zero R^2 value.

The significant outperformance of the SVM model can be attributed to its ability to handle complex, non-linear relationships via the Radial Basis Function (RBF) kernel. Diesel engine performance is governed by intricate interactions among fuel composition, engine speed, and combustion dynamics, resulting in non-linear data patterns. A simple linear model is fundamentally incapable of effectively capturing this complexity, leading to its poor performance, especially for the torque parameter. SVM, by mapping the data into a higher-dimensional feature space, successfully identified the complex regression function that accurately fits the data.

Table 4. Summary of model evaluation results

Parameter	Model	MAE	RMSE	R^2
Power	Linear Regression	8.5454	10.2548	0.6712
	SVM	1.1187	1.4900	0.9930
Torque	Linear Regression	18.1674	22.6074	0.09286
	SVM	2.3506	3.8527	0.9736
Note: MAE and RMSE are in the original units of the parameters (BHP and Nm). R^2 is on a scale of 0–1.				

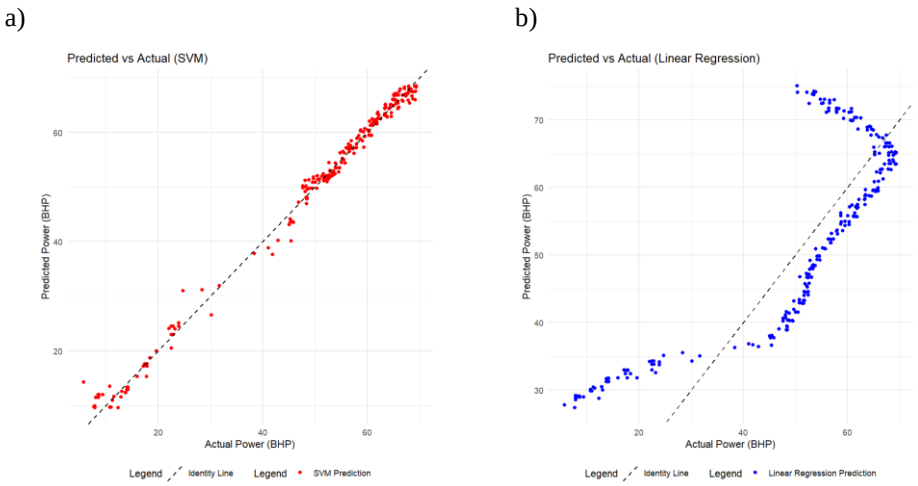

Fig. 5. (a) Predicted vs. actual values for power prediction with SVM; (b) predicted vs. actual values for power prediction with linear regression

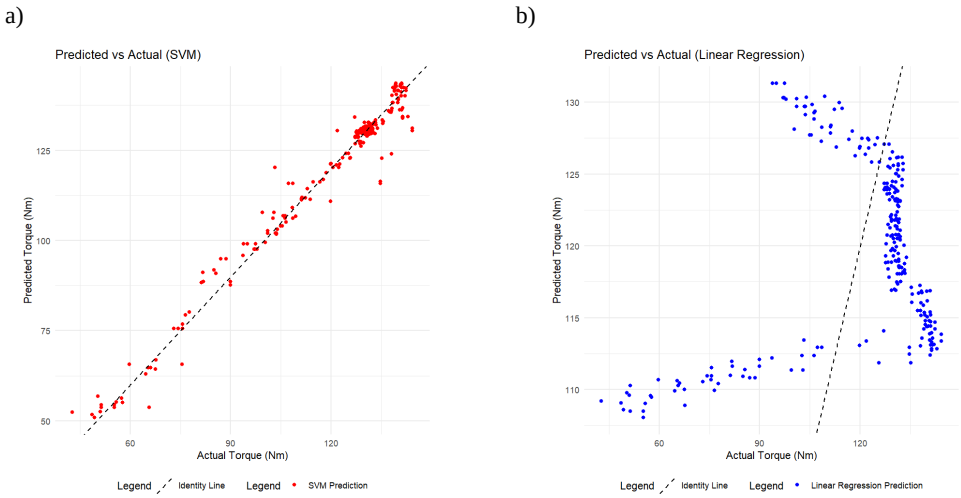

Fig. 6. (a) Predicted vs. actual values for torque prediction with SVM; (b) predicted vs. actual values for torque prediction with linear regression

The high accuracy achieved by the SVM model in this study (R^2 = 0.993 and 0.973 for torque) is highly competitive compared with recent literature in the field. In this model, both are treated as different outputs, even with the same input (in power learning, torque is not included as a variable). High accuracy occurs because the SVM parameters are tuned (to maximize accuracy across the model and dataset), and the data used is actual experimental data. This makes the relationship between variables clearer and more consistent for the model to learn. These results are consistent with, and in some cases exceed, the findings of other studies that used SVMs for engine parameter prediction. For instance [31] used an SVR model to predict the performance and emissions of a spark-ignition engine, achieving a validation R^2 of 0.9944 for fuel consumption rate and 0.9889 for NO_x emissions, which are comparable to our model's accuracy. Similarly [23] reported strong results predicting diesel engine performance with limited samples, achieving the best R^2 values of 0.9887 for maximum in-cylinder pressure and 0.9773 for engine efficiency, which align closely with our torque prediction R^2 of 0.974. A study [30] of predicting fuel consumption in marine diesel engines reported a test R^2 of 0.961 for their SVR model, a high value slightly surpassed by the R^2 values for both power and torque in our study. Even when predicting different parameters, such as exhaust gas temperature, [16] found that SVR was a strong predictor, with a validation R^2 of 0.82. Furthermore, a study by [10] on predicting biodiesel yield reported a Pearson Correlation Coefficient of approximately 0.75, underscoring the exceptional performance of our SVM model for this task.

While the high R^2 values indicate an excellent model fit for the given dataset, it is important to acknowledge the limitations. High performance on a test set partitioned from the same source does not guarantee perfect generalization to different engine types, substantially different operating conditions, or fuel variations not represented in the training data. Nonetheless, the results strongly validate that for the specific dataset of biodiesel-essential oil blends used, SVM is a highly effective and efficient modeling tool. This success highlights the significant potential of machine learning to accelerate research and development in alternative fuels by providing rapid and accurate performance predictions, thereby reducing the reliance on time-consuming and expensive physical experiments.

4. Conclusions

Based on the test results, it can be concluded that engine power tends to increase with rpm until it reaches an optimum range of 4800 to 5000 rpm, then decreases at higher rpm. The observed peak in power at approximately 5000 rpm, followed by a decline at higher speeds, is mainly due to reduced volumetric efficiency, shorter combustion duration, and increased frictional losses. These factors, acting together, reduce torque more rapidly than the increase in engine speed, leading to a decrease in overall engine performance. All fuel variations show a consistent trend, but variations 6 to 10 have produced relatively higher power because they have higher density and viscosity.

This study investigated the impact of an essential oil additive and varying biodiesel blend ratios on diesel engine performance, using both statistical analysis and machine learning. The statistical analysis revealed that neither the additive nor the blend ratio had a statistically significant main effect on engine power or torque ($p > 0.05$). Despite the lack of statistically significant group differences, the machine learning approach yielded clear results. A Support Vector Machine (SVM) model, developed alongside a Linear Regression model, demonstrated exceptional predictive capability. The SVM model achieved outstanding accuracy, with an R^2 of 0.993 for engine performance.

Further research using the SVM model can focus on the following developments. First, expand the input variables beyond RPM and fuel composition to include other operational parameters, such as torque, fuel flow rate, exhaust gas temperature, and injection pressure, so the model can better represent combustion conditions. Second, optimizing the SVM hyperparameters (C, γ, and ε) using metaheuristic or grid search methods more systematically is necessary to improve prediction accuracy, especially in high-speed regions that exhibit power degradation. Third, the application and comparison of various kernels (linear, polynomial, and RBF) are recommended to evaluate the model's ability to capture the non-linear relationship between input variables

and output power. Finally, model validation using a broader dataset and a variety of fuels, as well as testing on new experimental data, will improve the generalizability and reliability of the SVM model in diesel engine performance prediction applications.

Bibliography

[1] Abbasi KR, Shahbaz M, Zhang J, Irfan M, Alvarado R. Analyze the environmental sustainability factors of China: the role of fossil fuel energy and renewable energy. Renew Energy. 2022;187:390-402. https://doi.org/10.1016/j.renene.2022.01.066

[2] Brockway PE, Owen A, Brand-Correa LI, Hardt L. Estimation of global final-stage energy-return-on-investment for fossil fuels with comparison to renewable energy sources. Nat Energy. 2019;4(7):612-621. https://doi.org/10.1038/s41560-019-0425-z

[3] Doppalapudi AT, Azad AK, Khan MMK. Exergy, energy, performance, and combustion analysis for biodiesel NO_x reduction using new blends with alcohol, nanoparticle, and essential oil. J Clean Prod. 2024;15:467. https://doi.org/10.1016/j.jclepro.2024.142968

[4] Fitri N, Khofifah, Royhan H, Prakoso NI, Firdaus F, Sukri Q et al. Essential oils and their fractions as biodiesel additives: a review on quality and performance enhancement. Int J Des Nat Ecodynamics. 2025;20(8):1847-1864. https://doi.org/10.18280/ijdne.200815

[5] Grab-Rogaliński KA, Tutak W, Jamrozik A, Kociszewski A. Comparative analysis of IC diesel engine performance fueled with diesel/hydrogen and diesel/ammonia mixtures. Combustion Engines. 2026;204(1):176-183. https://doi.org/10.19206/CE-211069

[6] Hassan QH, Al-Abboodi H. Experimental investigation of the impact of methanol-diesel blends on diesel engine emissions and performance. Combustion Engines. 2025;201(2): 150-157. https://doi.org/10.19206/CE-204494

[7] Holechek JL, Geli HME, Sawalhah MN, Valdez R. A global assessment: can renewable energy replace fossil fuels by 2050? Sustainability. 2022;14(8):4792. https://doi.org/10.3390/su14084792

[8] Iskandar R, Arlinwibowo J, Setiadi R, Mujaki A, Naryanto RF, Setiyawan A et al. Impact of biodiesel blends on specific fuel consumption: A meta-analysis. IOP Conf Ser: Earth Environ Sci. 2024;1381:012033. https://doi.org/10.1088/1755-1315/1381/1/012033

[9] Jackson RB, Quéré CL, Andrew RM, Canadell JG, Korsbakken JI, Liu Z et al. Global energy growth is outpacing decarbonization. Environ Res Lett. 2018;13:120401. https://doi.org/10.1088/1748-9326/aaf303

[10] Jin X, Li S, Ye H, Wang J, Wu Y, Zhang D et al. Investigation and optimization of biodiesel production based on multiple machine learning technologies. Fuel. 2023;348:128546. https://doi.org/10.1016/j.fuel.2023.128546

[11] Johnsson F, Kjärstad J, Rootzén J. The threat to climate change mitigation posed by the abundance of fossil fuels. Clim Policy. 2019;19(2):258-274. https://doi.org/10.1080/14693062.2018.1483885

[12] Kalair A, Abas N, Saleem MS, Kalair AR, Khan N. Role of energy storage systems in energy transition from fossil fuels to renewables. Energy Storage. 2021;3(1):e135. https://doi.org/10.1002/est2.135

[13] Karunathilake H, Witharana S. Fossil fuels and global energy economics. Encyclopedia of Renewable Energy, Sustainability and the Environment. 2024;107-120. https://doi.org/10.1016/B978-0-323-93940-9.00050-5

[14] Kumar V, Choudhary AK. Prediction of the performance and emission characteristics of diesel engine using diphenylamine antioxidant and ceria nanoparticle additives with biodiesel based on machine learning. Energy. 2024;301: 131746. https://doi.org/10.1016/j.energy.2024.131746

[15] Lee K, Cho HM. Effects of castor and corn biodiesel on engine performance and emissions under low-load conditions. Energies. 2024;17:3349:1-12. https://doi.org/10.3390/en17133349

[16] Liu J, Huang Q, Ulishney C, Dumitrescu CE. Machine learning assisted prediction of exhaust gas temperature of a heavy-duty natural gas spark ignition engine. Appl Energy. 2021;300:117413. https://doi.org/10.1016/j.apenergy.2021.117413

[17] Longwic R, Tatarynow D, Kuszneruk M, Wozniak-Borawska G. Preliminary tests of a diesel engine powered by diesel and hydrogen. Combustion Engines. 2023;195(4): 35-39. https://doi.org/10.19206/CE-169485

[18] Mu Z, Fu J, Zhou F, Yuan K, Yu J, Huang D et al. A comparatively experimental study on the performance and emission characteristics of a diesel engine fueled with Tung oil-based biodiesel blends (B10, B20, B50). Energies. 2023; 16(14):5577. https://doi.org/10.3390/en16145577

[19] Naryanto RF, Delimayanti MK, Adi R, Abdurrahman A, PK Nashiroh, Sukoco I et al. Prediction of power output on external combustion engine using regression models. J Theor Appl Inf Technol. 2024;102(24):9233-9243. https://jatit.org/volumes/Vol102No24/34Vol102No24.pdf

[20] Naryanto RF, Delimayanti MK, Naryaningsih A, Adi R, Setiawan BA. Fault detection in diesel engines using artificial neural networks and convolutional neural networks. J Theor Appl Inf Technol. 2024;102(2):680-692. https://www.jatit.org/volumes/Vol102No2/23Vol102No2.pdf

[21] Naryanto RF, Hadromi H, Musyono ADNI, Setiadi R, Caesarendra W. Diesel engine power prediction based on fuel blends using neural network. Rep Mech Eng. 2025;l(6): 127-135. https://www.rme-journal.org/index.php/asd/article/view/492/109

[22] Newell RG, Raimi D, Aldana G. Global energy outlook 2019: The next generation of energy. 2019. https://www.rff.org/geo

[23] Niu X, Yang C, Wang H, Wang Y. Investigation of ANN and SVM based on limited samples for performance and emissions prediction of a CRDI-assisted marine diesel engine. Appl Therm Eng. 2017;111:1353-1364. https://doi.org/10.1016/j.applthermaleng.2016.10.042

[24] Overland I, Juraev J, Vakulchuk R. Are renewable energy resources more evenly distributed than fossil fuels? Renew Energy. 2022;200:379-386. https://doi.org/10.1016/j.renene.2022.09.046

[25] Pugazhendhi A., Kamarudin SK, Chinnathambi A, Alharbi SA, Ramya G. Investigation of bionano additives in red algae Cyanidioschyzon merolae ultrasonified MgO/MWCNT catalyzed biodiesel in optimized engine performance. Environ Res. 2024;258.119352. https://doi.org/10.1016/j.envres.2024.119352

[26] Putra AEE. The effect of heating B30 biodiesel on performance and combustion in a TV-1 type diesel engine. Jurnal Teknik Mesin Sinergi. 2021;19(1):128-135. https://doi.org/10.31963/sinergi.v19i1.2776

[27] Rashedi A, Khanam T, Jonkman M. On reduced consumption of fossil fuels in 2020 and its consequences in global

environment and exergy demand. Energies. 2020;13(22):1-14. https://doi.org/10.3390/en13226048
[28] Setiyawan A, Fitriyana DF, Bahatmaka A, Firmansyah HN, Darsono FB, Aryadi W et al. Diesel Engine Performance using Pertamina Dex and Biodiesel (B30) on rpm and fuel consumption. IOP Conf Ser: Earth Environ Sci. 2023;1203: 012010. https://doi.org/10.1088/1755-1315/1203/1/012010
[29] Woś P, Krzemiński A, Kuszewski H, Jakubowski M, Jaworski A, Hunicz J et al. Problems of filtration and standardization on parameter conformity of diesel fuels containing decarbonization components and processing impurities. Combustion Engines. 2026;204(1):184-191. https://doi.org/10.19206/CE-211450
[30] Yuksel O, Bayraktar M, Sokukcu M. Comparative study of machine learning techniques to predict fuel consumption of a marine diesel engine. Ocean Eng, 2023;286(1):115505. https://doi.org/10.1016/j.oceaneng.2023.115505
[31] Zhang Y, Wang Q, Chen X, Yan Y, Yang R, Liu Z et al. The prediction of spark-ignition engine performance and emissions based on the SVR algorithm. Processes. 2022; 10(2):312. https://doi.org/10.3390/pr10020312
[32] Zimakowska-Laskowska M, Laskowski P. Comparison of pollutant emissions from various types of vehicles. Combustion Engines. 2024;197(2):139-145. https://doi.org/10.19206/CE-181193

Rizqi Fitri Naryanto, DEng. – Mechanical Engineering, Universitas Negeri Semarang, Indonesia.
e-mail: *rizqi_fitri@mail.unnes.ac.id*

Diffa Rakhman Prabantara – Mechanical Engineering, Universitas Negeri Semarang, Indonesia.
e-mail: *difarahman413@gmail.com*

Rizki Setiadi – Mechanical Engineering, Universitas Negeri Semarang, Indonesia.
e-mail: *rizkisetiadi@mail.unnes.ac.id*

Prof. Hadromi Hadromi – Mechanical Engineering, Universitas Negeri Semarang, Indonesia.
e-mail: *hadromi@mail.unnes.ac.id*