10 (1): 24-31 (2026)

Journal of Aviation

https://dergipark.org.tr/en/pub/jav

e-ISSN 2587-1676

Prediction of Biofuel Density Obtained by Blending JP-5 Jet Fuel at Different Ratios with Machine Learning Algorithms

Bülent Kurt[1]

[1]Balıkesir University, Edremit Civil Aviation College, Edremit, Balıkesir, 10300, Türkiye. (bulent.kurt@balikesir.edu.tr)

Article Info

Received: 26 March 2026
Revised: 27 January 2026
Accepted: 23 February 2026
Published Online: 26 February 2026

Keywords:
Aviation
Machine learning algorithms
Aviation safety
Environment

Corresponding Author: *Bülent Kurt*

RESEARCH ARTICLE

https://doi.org/10.30518/jav.1665932

Abstract

In the aviation industry, sustainable biofuels are emerging as a crucial alternative to reduce dependence on fossil fuels and mitigate harmful greenhouse gas emissions. However, determining the physicochemical properties of biofuel blends traditionally relies on expensive and time-consuming laboratory experiments. This study proposes a highly accurate, data-driven computational approach to predict the density of biofuel blends obtained by mixing 14 different plant and animal-based oils with JP-5 jet fuel at various ratios. To ensure robust generalization and eliminate overfitting risks on the experimental dataset (71 samples), six advanced machine learning architectures—Artificial Neural Networks (ANN), Gaussian Process Regression (GPR), Regression Trees, Random Forest, LSBoost, and Support Vector Machines (SVM)—were comprehensively evaluated using a rigorous 5-fold cross-validation strategy. The results demonstrated that the Artificial Neural Network optimized with the Bayesian Regularization algorithm (ANN-BR) achieved the highest predictive performance. Specifically, the ANN-BR model yielded a Cross-Validation Coefficient of Determination (R^2) of 0.9820, a Correlation Coefficient (R) of 0.9910, and a minimal Mean Squared Error (MSE) of 0.00121 on the unseen test folds. The Regression Tree and GPR models also exhibited exceptional accuracy, closely following the ANN. Ultimately, this study proves that predictive machine learning modeling can reliably supplement and accelerate conventional fuel characterization tests, offering significant time and cost advantages for the aviation sector.

1. Introduction

Today, airplanes and jets need fossil fuels to fly. Fossil fuels consumed in aircraft and jet engines increase both harmful emissions and costs. The main harmful emissions from aircraft engines are carbon dioxide (CO2), nitrogen oxides (NOx), sulfur oxides (SOx), lead compounds, hydrocarbons (HC), particulate matter (PM), and carbon monoxide (CO). Since airplanes and jets burn fossil fuels in their engines, they contribute to global warming by releasing CO_2 and other greenhouse gases into the atmosphere (Nojoumi et al., 2009). The combustion process releases airborne pollutants such as PM, NO_x, and sulfates, causing local air pollution (Carslaw and Taylor, 2009). Furthermore, emissions from airplanes and jets at high altitudes can indirectly affect the ozone layer in the stratosphere (Emmons et al., 2012; Lazic et al., 2023).

Reducing harmful emissions is among the objectives of the International Civil Aviation Organization (ICAO). ICAO supports new studies and projects to achieve this goal. In order to reduce emissions from aviation activities, the search for alternative liquid fuels has emerged. Liquid fuels derived from plant or animal waste are called biofuels. The widespread use of biofuels in the aviation sector will contribute to reducing the greenhouse gas effect and air pollution. In addition, the use of biofuels provides economic advantages to the sector by reducing fuel costs. For these reasons, new studies on biofuel blends are needed. In this regard, the development and increased use of biofuels are regarded as a strategic step toward achieving sustainable aviation goals. Moreover, from the perspective of energy policies and international competitiveness, the expansion of biofuels plays a critical role in shaping the future transformation of the aviation industry.

1.1. Related Work

There are many different studies on jet fuels and harmful emissions in the literature. Initially, experimental approaches were dominant. For instance, Solmaz et al. (2014) evaluated the effects of Jet A1 fuel mixed with diesel fuel on engine performance. Boomadevi et al. (2021) investigated the use of biofuels from microalgae in micro gas turbine engines. Zhou et al. (2015) examined the effects of chemical and physical properties of different biofuels on engine thermodynamic cycle parameters and combustion chamber performance. Similarly, other studies experimentally analyzed the production of biodiesel from grape seed oil (Fernández et al., 2010), and the performance and emission characteristics of JP-5 and biodiesel blends (Korres et al., 2008; Uyumaz et al., 2014).

In recent years, to overcome the high costs and time consumption of laboratory experiments, the focus has shifted towards mathematical and computational modeling. Early modeling efforts proposed empirical equations and mathematical correlations for predicting biodiesel density (Lapuerta et al., 2010; Verduzco, 2013). Sajjadi et al. (2016) conducted a comprehensive literature review predicting these properties using traditional methods. However, modern regression-based Machine Learning (ML) studies have proven to be more robust in capturing the highly nonlinear physical relationships inherent in fuel blends. For instance, Comesana et al. (2026) developed a structured ML framework for predicting sustainable aviation fuel properties, while Alkharashi et al. (2025) applied ensemble learning models such as XGBoost and random forest to assess biofuel blend performance and emissions. Giwa et al. (2023) compared ANN and RT models using FAME composition-based approaches to predict hybrid biodiesel (HB) fuel properties, and demonstrated that RT achieved higher accuracy. Mohammad et al. (2025) developed three machine learning models—Gaussian Process Modeling (GPM), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Decision Tree (DT)—to predict the density of waste frying oil biodiesel–alcohol blends using a large experimental dataset, and found that GPM provided the highest accuracy. Ibrahim et al. (2025) developed predictive models for the density of biodiesel–alkane mixtures using experimental data and various machine learning techniques, showing that CNN and SVR achieved the highest accuracy compared to other methods and traditional correlations. Wang et al. (2026) collected experimental density data, developed new predictive models, identified the most influential parameters, and achieved high accuracy and stability in fuel density predictions. In a recent symposium study using the same dataset, Kurt (2024) investigated the compatibility and correlation between various input parameters and the density of biofuel blends. Despite these advancements, there is still a lack of comprehensive comparative studies evaluating multiple advanced ML algorithms—particularly ensemble learning and probabilistic models—on aviation biofuel blends.

1.2. Aim and objectives

In order for biodiesel to replace diesel fuels, it must be competitive with petroleum diesel in terms of quality. When production parameters are optimized, it is possible to produce biodiesel with properties similar to or in some ways superior to petroleum diesel (Tolu, 2023). The main important properties for biofuels include viscosity, freezing point, cetane number, flash point, calorific value, and density. Among these parameters, fuel density is important in terms of biofuel efficiency and economy.

The main objective of this study is to improve the efficiency of biodiesel and to show that the density of biodiesel blends can be predicted in a computer environment using machine learning methods. Prak et al. (2022) recorded the viscosity, density, flash point, speed of sound, and surface tension of blends of 14 different biofuels and JP-5 jet fuel at different ratios to develop new models. This study involves the prediction of fuel density by machine learning method models developed with the recorded biofuels.

To address the limitations of existing literature and provide a comprehensive comparative analysis, this study expands the scope beyond traditional methods by evaluating six distinct machine learning approaches: Artificial Neural Networks (ANN), Support Vector Regression (SVR), Regression Trees (Fine Tree), Random Forest (RF), Least Squares Boosting (LSBoost), and Gaussian Process Regression (GPR). The rationale behind selecting these specific models is to compare traditional architectures (ANN, SVR) with modern ensemble techniques (RF, LSBoost) and probabilistic approaches (GPR). GPR is particularly emphasized and hypothesized to perform exceptionally well, as its Bayesian foundation makes it highly suitable for modeling uncertainties in small datasets (71 samples) where deeper neural networks might struggle with generalization. Furthermore, to prevent the overfitting risks associated with limited data and to ensure maximum reliability and reproducibility, a rigorous 5-fold cross-validation strategy with fixed random seeds is implemented for all models, replacing the conventional single-split approach. By comparing the models on error performance criteria, it is aimed to develop the model that predicts the fuel density with the least error and the highest accuracy.

By comparing the models on error performance criteria, it is aimed to develop the model that predicts the fuel density with the least error and the highest accuracy. The results obtained provide the following contributions:

- Unlike the studies on biodiesel, it has been shown that biofuel density can be estimated in a computer environment using the mass fraction and surface tension values of biodiesel in JP-5.
- Studies analyzing different properties of biofuel will increase the use of biofuel. Thus, the damage of fossil fuels will be reduced.
- With the developed model, biofuel analysis will be advantageous in terms of both time and cost.

The article is organized as follows: the second section provides information about the data and machine learning methods used to estimate biofuel density. In the third section, the details of the developed machine learning models and the results obtained over the selected criteria are given. In the last section, the conclusions and significance of the study are presented.

2. Materials and Methods

In order for biofuels to burn efficiently, the blend ratios and fuel types must be determined correctly. Researchers are trying to obtain high quality biofuels by using blends of different types and ratios. Studies on biofuels are usually carried out in the laboratory. These experiments are both costly and time consuming. This study aims to bring an alternative and new perspective to the studies on biofuels by estimating fuel density in a computer environment using advanced Machine Learning (ML) techniques.

2.1. Data and Validation Strategy

In this study, the dataset created by Prak et al. was used. In this dataset, 14 different plant and animal-based oils (avocado oil, grape seed oil, bacon fat, canola oil, etc.) were investigated as biofuels. Surface tension, density, flash point, viscosity, and speed of sound values of the biofuel obtained by mixing each oil with JP-5 jet fuel were measured at different temperatures. In Table 1, the measured values of the blend of olive oil and JP-5 fuel are presented as an example. A total of 71 different biofuels were used to develop the models. To ensure the mathematical stability of the machine learning algorithms and improve their convergence speed, all input and output variables were normalized to a range of [0, 1] using the min-max scaling technique prior to model training.

Table 1. An example data set showing mixing ratios

Type	Mass fraction of biodiesel in JP-5	Surface tension 21.0 °C (mN/m)	Density 15.0 °C (kg/m³)
JP-5 jetfuel	0.0000	26.0	805.5115
Olive oil biodiesel	1.0000	28.9	878.674
80% Olive oil biodiesel + 20% JP-5	0.7963	27.9	862.6455
60% Olive oil biodiesel + 40% JP-5	0.6010	27.5	847.8465
40% Olive oil biodiesel + 60% JP-5	0.4013	26.8	830.943
20% Olive oil biodiesel + 80% JP-5	0.2004	26.2	819.0225

Given the limited number of samples (71 data points) in the dataset, relying on a conventional single train-test split makes machine learning models highly susceptible to data variance and overfitting. To overcome this limitation, ensure robust model evaluation, and guarantee mathematical reproducibility, a 5-fold Cross-Validation (CV) strategy with a fixed random seed was strictly employed in this study. In this procedure, the dataset is systematically partitioned so that exactly 80% of the data is used for training the models, while the remaining 20% is reserved for testing in each iteration. This process is repeated 5 times, ensuring that every single data point is evaluated as part of the unseen test set exactly once. The final performance metrics are calculated as the average of these 5 folds, providing a highly reliable and unbiased representation of the models' generalization capability.

2.2. Machine Learning Algorithms

In supervised learning, models are developed that establish the relationship between input parameters (mass fraction of biodiesel in JP-5 and surface tension) and the output parameter (fuel density). To provide a comprehensive comparison, six different ML architectures were employed:

- **Artificial Neural Networks (ANN):** A multi-layer perceptron model optimized with the Bayesian Regularization (BR) training algorithm. BR is highly resistant to overfitting on small datasets compared to standard Levenberg-Marquardt approaches.
- **Support Vector Regression (SVR):** SVR aims to find a hyperplane that best fits the data within a specified tolerance margin (ε). In this study, different kernel functions (Linear, Polynomial, and Gaussian) were tested to capture nonlinear relationships.
- **Regression Tree (Fine Tree):** A decision tree technique that explores the connections between input parameters and output variables by splitting the data into hierarchical branches. To maintain high precision, a "Fine Tree" structure was utilized.
- **Gaussian Process Regression (GPR):** GPR is a nonparametric, Bayesian approach to regression. It is particularly advantageous for small datasets as it provides uncertainty measurements for its predictions.
- **Random Forest (RF):** An ensemble learning method that constructs a multitude of decision trees during training and outputs the average prediction. It effectively corrects the habit of decision trees overfitting to their training set.
- **Least Squares Boosting (LSBoost):** A sequential ensemble technique where each new tree attempts to correct the residual errors of the previous trees. It is highly effective for complex, non-linear regression tasks.

2.3. Hyperparameter Optimization Setup

For a fair and scientifically rigorous comparison, all algorithms were trained and tested using the exact same 5-fold CV splits controlled by fixed random seeds. The fixed parameters and the search ranges used during the systematic hyperparameter optimization (Grid Search) for each machine learning algorithm are summarized in Table 2.

Table 2. Hyperparameter optimization setup and search ranges for the evaluated machine learning algorithms.

Machine Learning Algorithm	Hyperparameter	Search Range / Fixed Value
Artificial Neural Network (ANN)	Training Algorithm	Bayesian Regularization (Fixed)
	Max Epochs	1000 (Fixed)
Support Vector Regression (SVR)	Kernel Function	Gaussian, Linear, Polynomial (Order 3)
	BoxConstraint (C)	Logarithmic scale search (0.01-100)
	Epsilon (ε)	Iterative search
Gaussian Process Regression (GPR)	Basis Function	Constant, Linear, pureQuadratic
	Kernel Function	SquaredExponential, Matern32, Matern52
Regression Tree (Fine Tree)	MinLeafSize	1 to 20 (Grid Search)
Random Forest (RF)	Ensemble Method	Bagging (Fixed)
	Number of Trees	100 (Fixed)
	MinLeafSize	1 to 20 (Grid Search)
Least Squares Boosting (LSBoost)	Ensemble Method	LSBoost (Fixed)
	Number of Trees	100 (Fixed)
	LearnRate	0.01 to 0.50 (Grid Search)
	MaxNumSplits (Tree Depth)	1 to 10 (Grid Search)

2.4. Performance Evaluation Metrics

To evaluate and compare the prediction performance of the ML models, several error and correlation metrics were utilized. Mean Squared Error (MSE), and Mean Absolute Error (MAE) evaluate the magnitude of the prediction errors. Additionally, the Correlation Coefficient (R) and the Coefficient of Determination (R^2) were used to measure the goodness-of-fit. It is important to clarify that R measures the linear correlation between the predicted and actual values, whereas R^2 indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. The mathematical expressions are presented below:

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

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (4)$$

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \quad (5)$$

Where:

- n is the total number of biofuel samples in the dataset,
- y_i represents the actual (measured) fuel density for the i^{th} sample,
- $\hat{y}_i$ represents the predicted fuel density by the machine learning model for the i^{th} sample,
- $\bar{y}$ is the mean of the actual (measured) fuel densities,
- $\bar{\hat{y}}$ is the mean of the predicted fuel densities.

3. Results and Discussion

In this study, six different advanced machine learning models (ANN, Regression Tree, GPR, LSBoost, Linear SVM, and Random Forest) were comprehensively evaluated to predict the density of biofuel blends. To completely eliminate the risk of overfitting on the limited dataset (71 samples) and to ensure robust generalization, all models were trained and tested using a strictly controlled 5-fold Cross-Validation (CV) strategy with normalized data.

3.1. Performance of the Machine Learning Models

The predictive performance of the optimized models on both the training set and the unseen CV test folds is summarized in Table 3. The models were evaluated based on their Correlation Coefficient (R), Coefficient of Determination (R^2), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

.

Table 3. Comprehensive performance comparison of the optimized ML models (5-Fold CV Results).

ANN (Bayesian Regularization) - (Neurons: 61)					
Data Phase	**Correlation (R)**	**Determination (R²)**	**MSE**	**RMSE**	**MAE**
Training	0.9951	0.9903	0.00065	0.0255	0.0183
Test (CV)	0.9910	0.9820	0.00121	0.0348	0.0263
Regression Tree (Fine Tree) - (MinLeafSize: 4)					
Data Phase	**Correlation (R)**	**Determination (R²)**	**MSE**	**RMSE**	**MAE**
Training	0.9929	0.9858	0.00095	0.0309	0.0213
Test (CV)	0.9908	0.9814	0.00125	0.0353	0.0241
Gaussian Process Regression (Matern52) - (Basis: pureQuadratic)					
Data Phase	**Correlation (R)**	**Determination (R²)**	**MSE**	**RMSE**	**MAE**
Training	0.9921	0.9843	0.00105	0.0325	0.0233
Test (CV)	0.9900	0.9800	0.00134	0.0366	0.0264
Gradient Boosting (LSBoost) - (LearnRate: 0.05, Depth: 9)					
Data Phase	**Correlation (R)**	**Determination (R²)**	**MSE**	**RMSE**	**MAE**
Training	0.9944	0.9887	0.00076	0.0276	0.0187
Test (CV)	0.9892	0.9778	0.00149	0.0387	0.0276
Support Vector Machine (Linear) (BoxConstraint: 0.6579)					
Data Phase	**Correlation (R)**	**Determination (R²)**	**MSE**	**RMSE**	**MAE**
Training	0.9902	0.9804	0.00132	0.0363	0.0284
Test (CV)	0.9890	0.9781	0.00147	0.0384	0.0295
Random Forest (RF) (MinLeafSize: 5)					
Data Phase	**Correlation (R)**	**Determination (R²)**	**MSE**	**RMSE**	**MAE**
Training	0.9797	0.9589	0.00277	0.0526	0.0326
Test (CV)	0.9751	0.9490	0.00344	0.0586	0.0387

As seen in Table 3, all developed models exhibited remarkable success in predicting the fuel density. The Artificial Neural Network (ANN) trained with Bayesian Regularization (ANN-BR) demonstrated the highest predictive capability, achieving a correlation coefficient (R) of 0.9910 and an R^2 of 0.9820 on the unseen test folds. Surprisingly, the highly optimized Regression Tree (Fine Tree) model with a MinLeafSize of 4 followed very closely with a test R of 0.9908, proving that when node splitting is strictly regulated, a single tree architecture can effectively map the non-linear relationship between mass fractions, surface tension, and fuel density. The probabilistic GPR model also showed exceptional performance (R = 0.9900), validating its suitability for capturing uncertainties in smaller datasets. Ensemble methods like LSBoost (R = 0.9892) and Random Forest (R = 0.9751) also performed well but were slightly outperformed by the perfectly regularized ANN architecture.

When all the results are evaluated, the ANN-BR model provided the highest predictive accuracy on both the training and cross-validation test datasets, demonstrating a stronger generalization capability compared to the other algorithms. Although the Regression Tree and GPR models also performed remarkably well and followed closely, models such as the Polynomial SVM significantly lagged behind in capturing the non-linear relationships of the test data. These results strongly suggest that the Artificial Neural Network optimized with Bayesian Regularization is the most effective and robust model for predicting biofuel density in this problem. The general characteristics of this optimized network are detailed in Table 4, while the comparative R^2 performance of all models and the specific estimation results of the ANN-BR model are illustrated in Figures 1 and 2, respectively.

3.2. Overfitting Analysis and Model Justification

A common challenge when applying complex machine learning algorithms to relatively small datasets (such as the 71 biofuel samples in this study) is the risk of overfitting. An overfitted model memorizes the training data, resulting in near-perfect training scores but exceptionally high error rates on unseen test data. To definitively address this, the training and cross-validation test errors were meticulously compared in Table 3.

The analysis clearly demonstrates that the models learned the underlying physical and thermodynamic rules rather than memorizing the data. For instance, the top-performing ANN-BR model achieved a training R^2 of 0.9903 and a test R^2 of 0.9820. The gap between training and testing performance is minimal, and the test MSE (0.00121) is remarkably low. This robust generalization is directly attributed to the use of the Bayesian Regularization algorithm, which mathematically penalizes overly complex neural weights, automatically forcing the network to maintain the simplest possible architecture that solves the problem.

Table 4. General Characteristics of the Optimized ANN-BR Model

Model Parameter Characteristic	Value / Configuration
Algorithm Type	Artificial Neural Network (ANN) - Multi-Layer Perceptron
Training Function	Bayesian Regularization (trainbr)
Validation Strategy	5-Fold Cross-Validation
Data Partitioning (per fold)	80% Training, 20% Testing (Unseen)
Network Architecture	2 Inputs → 61 Hidden Neurons →1 Output
Number of Hidden Neurons	61 (Fixed optimum)
Random Seed Initialization	27 (Fixed for reproducibility)
Total Dataset Size	71 samples
Early Stopping / Validation	Disabled internally (0%) to prevent data leakage during 5-Fold CV

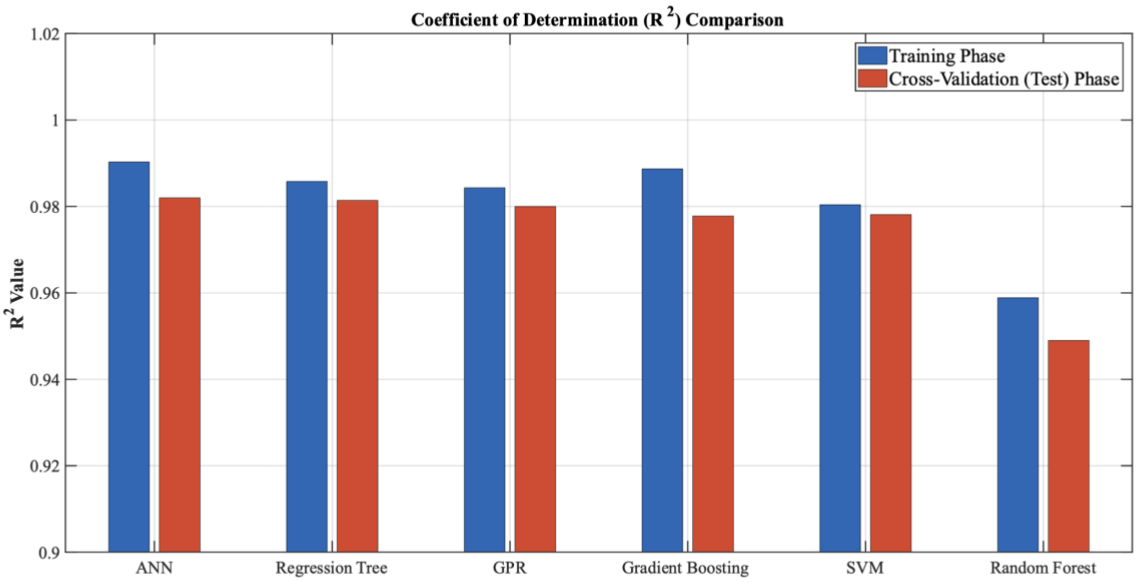

Figure 1. Comparison of the Coefficient of Determination (R^2) for the developed machine learning models during the training and 5-fold cross-validation phases.

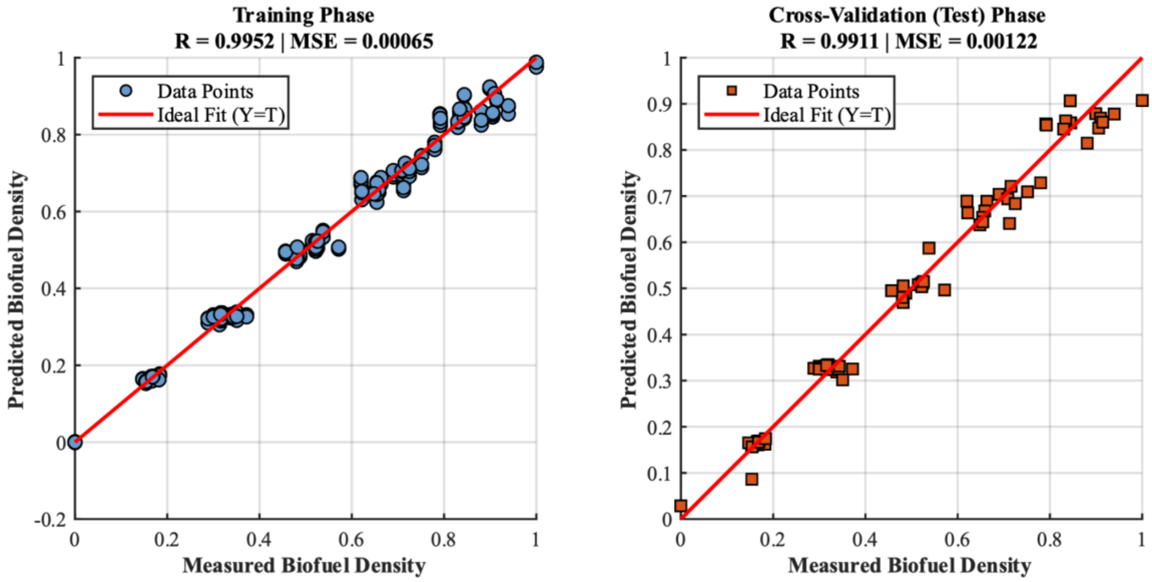

Figure 2. Estimation results of the developed ANN (Bayesian Regularization) model for the training and test phase

Similarly, the Regression Tree avoided overfitting because its minimum leaf size was constrained to 4 during hyperparameter optimization, preventing the tree from branching endlessly to memorize individual data points. The GPR model, by its Bayesian nature, inherently provides a probabilistic confidence bound, making it highly resistant to data noise.

3.3. Comparison with the Literature

When the results are compared with existing literature, the superiority of the applied ML optimization methodology becomes evident. Traditional empirical models proposed by Verduzco (2013) and Lapuerta et al. (2010) often rely on fixed equations that struggle to adapt to multi-variable blends containing 14 different plant and animal oils mixed with JP-5 jet fuel. Furthermore, contemporary research investigating fuel properties still predominantly relies on extensive, costly, and time-consuming experimental procedures. For instance, recent studies by Prak and Cowart (2024) continue to conduct rigorous laboratory measurements to determine the densities, speeds of sound, and viscosities of binary hydrocarbon mixtures. Similarly, other current works experimentally characterize the physicochemical properties of newly synthesized crude palm oil biodiesels (Balfas et al., 2024), evaluate the combustion emissions and particulate matter of bio-derived fuels in diesel engines (Sukjit et al., 2025), and perform thermal hazard assessments of specific fuel components through demanding physical setups (De Liso et al., 2024).

In contrast to these purely experimental approaches, the optimally tuned ML models developed in this study autonomously capture the complex physical interactions between the mass fraction of biodiesel and surface tension, predicting the final fuel density without the need for physical blending or manual recalibration. Compared to recent ML-based fuel studies that typically rely on single train-test splits or unoptimized default algorithms, the fixed-seed 5-fold CV methodology presented here offers much higher mathematical reliability. Achieving an $R^2 > 0.98$ computationally using ANN, Regression Tree, and GPR proves that the extensive laboratory tests typically required for characterizing aviation biofuel blends can be reliably supplemented, and in some aspects entirely replaced, by advanced predictive ML modeling, drastically reducing the associated time and financial burdens.

4. Conclusion

At present, fossil fuels are largely used to meet global aviation and energy needs. However, the continuous consumption of these fuels causes serious damage to both human health and the environment. In particular, negative impacts such as greenhouse gas emissions, air pollution, and the degradation of ecosystems heavily contribute to global problems like climate change. In order to mitigate these harmful effects, sustainable aviation biofuels are being extensively developed and promoted as a viable alternative to petroleum-derived jet fuels.

In this study, to accelerate the integration and optimization of these alternative fuels, advanced machine learning models were developed to predict the density of biofuel blends obtained by mixing 14 different types of plant and animal-based oils with JP-5 jet fuel. To ensure robust generalization and completely eliminate the risk of overfitting inherent to limited experimental datasets, six distinct computational architectures—including Artificial Neural Networks (ANN), Gaussian Process Regression (GPR), Regression Trees, ensemble methods (LSBoost, Random Forest), and Support Vector Machines (SVM)—were comprehensively evaluated using a rigorous 5-fold cross-validation strategy.

The evaluation revealed that the Artificial Neural Network optimized with the Bayesian Regularization algorithm (ANN-BR) provided the highest predictive accuracy, achieving an exceptional coefficient of determination (R^2) of 0.9820 and a minimal Mean Squared Error (MSE) of 0.00121 on the unseen test folds. The structurally optimized Regression Tree (MinLeafSize: 4) and the probabilistic GPR models also demonstrated outstanding performance, following the ANN closely. Conversely, standard parametric approaches like the Polynomial SVM showed the lowest capability in capturing the complex, non-linear thermodynamic interactions of the blends.

Ultimately, this study proves that the physical properties of complex aviation biofuel blends can be predicted computationally with extremely high precision ($R > 0.99$). By drastically reducing the time, material waste, and financial costs associated with traditional laboratory experiments, the developed ML models have the capacity to accelerate biofuel research, encourage their more widespread use in the aviation industry, and substantially contribute to the reduction of global carbon emissions.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgements

During the preparation of this work, the author used the Gemini AI tool to improve the readability and academic tone of the English language, as well as to format the methodological responses. The author assumes full responsibility for the integrity of the published work.

References

Alkharashi, A., Al Mazroa, A., M. Alashjaee, A., V, S., M, D. B., S, S., & S, V. (2025). Machine learning models for biogas potential in sustainable aviation: XGBoost, random forest and ridge regression. Aircraft Engineering and Aerospace Technology.

Balfas, R. N., Muhammad Syam, A., Muhammad, M., Setiawan, A., & Fithra, H. (2024). Characteristics of biodiesel produced from crude palm oil through non-alcohol synthesis route using dimethyl carbonate and immobilized eco-enzyme catalyst. Energies, 17(7), 1551.

Boomadevi, P., Paulson, V., Samlal, S., Varatharajan, M., Sekar, M., Alsehli, M., & Tola, S. (2021). Impact of microalgae biofuel on microgas turbine aviation engine: A combustion and emission study. Fuel, 302, 121155.

Carslaw, D. C., & Taylor, P. J. (2009). Analysis of air pollution data at a mixed source location using boosted regression trees. Atmospheric Environment, 43(22-23), 3563-3570.

Comesana, A. E., Chen, S., Niemeyer, K. E., & Rapp, V. H. (2026). A structured framework for predicting sustainable aviation fuel properties using liquid-phase FTIR and machine learning. Fuel, 408, 137588.

Emmons, L. K., Hess, P. G., Lamarque, J. F., & Pfister, G. G. (2012). Tagged ozone mechanism for MOZART-4, CAM-chem and other chemical transport models. Geoscientific Model Development, 5(6), 1531-1542.

Fernández, C. M., Ramos, M. J., Pérez, Á., & Rodríguez, J. F. (2010). Production of biodiesel from winery waste: extraction, refining and transesterification of grape seed oil. Bioresource technology, 101(18), 7019-7024.

Giwa, S. O., Taziwa, R. T., & Sharifpur, M. (2023). Dependence of composition-based approaches on hybrid biodiesel fuel properties prediction using artificial neural network and random tree algorithms. Renewable Energy, 218, 119324.

Ibrahim, S. K., Albadr, R. J., Doshi, H., Yadav, A., Ballal, S., Singh, A., ... & Mottaghi, M. (2025). Machine learning prediction of density of fatty acid methyl ester mixed with alkanes biodiesel over a wide range of operating conditions. Biomass and Bioenergy, 196, 107712.

Korres, D. M., Karonis, D., Lois, E., Linck, M. B., & Gupta, A. K. (2008). Aviation fuel JP-5 and biodiesel on a diesel engine. Fuel, 87(1), 70-78.

Kurt, B. (2024). Prediction of fuel density obtained from different biodiesel blends with jet fuel JP-5 using artificial neural network, in Ases II. International Bandırma Scientific Studies Conference, pp. 367–374.

Lapuerta, M., Rodríguez-Fernández, J., & Armas, O. (2010). Correlation for the estimation of the density of fatty acid esters fuels and its implications. A proposed biodiesel cetane index. Chemistry and physics of lipids, 163(7), 720-727.

Liso, D. B. A., Pio, G., & Salzano, E. (2024). Small scale pool fires: The case of toluene. Journal of Loss Prevention in the Process Industries, 92, 105430.

Lazic, D., Grujic, V., & Cvetkovic, D. (2023). Results of using sustainable aviation fuel in transport aircraft on different flight levels and their influence on carbon-dioxide Co2 emissions. Zeszyty Naukowe. Transport/Politechnika Śląska.

Mohammad, S. I., Owida, H. A., Vasudevan, A., Menon, S. V., Al-Hasnaawei, S., Ray, S., ... & Ranjbar, F. (2025). Thermophysical properties of used frying oil biodiesels blended with alcohols: Robust machine learning frameworks for density prediction. Industrial Crops and Products, 236, 121916.

Nojoumi, H., Dincer, I., & Naterer, G. F. (2009). Greenhouse gas emissions assessment of hydrogen and kerosene-fueled aircraft propulsion. International journal of hydrogen energy, 34(3), 1363-1369.

Prak, D. L., Hamilton, M., Banados, R., & Cowart, J. (2022). Density, viscosity, speed of sound, flash point, bulk modulus, and surface tension of mixtures of military jet fuel JP-5 and biodiesels dataset. Data in Brief, 41, 107849.

Prak, D. J. L., & Cowart, J. S. (2024). Binary Mixtures of n-Alkylcyclohexanes with Decane: Densities and Speeds of Sound between 288.15 and 333.15 K and Viscosities between 293.15 and 323.15 K at 0.1 MPa. Journal of Chemical & Engineering Data, 69(9), 3029-3043.

Sajjadi, B., Raman, A. A. A., & Arandiyan, H. (2016). A comprehensive review on properties of edible and non-edible vegetable oil-based biodiesel: Composition, specifications and prediction models. Renewable and Sustainable Energy Reviews, 63, 62-92.

Solmaz, H., Yamık, H., İçingür, Y., & Calam, A. (2014). Investigation of the effects of civil aviation fuel Jet A1 blends on diesel engine performance and emission characteristics.

Sukjit, E., Klinkaew, N., Pumpuang, A., Batsungnoen, K., Thumthae, C., & Maneedaeng, A. (2025). Particulate matter characteristics and emissions of waste-and bio-derived fuels in diesel engine applications. Cleaner Engineering and Technology, 101056.

Tolu, Ç.M. (2023). Biyodizel üretiminde yağ asidi bileşenlerinin yakıt özelliklerine etkisinin araştırılması. Ph.D. dissertation, Necmettin Erbakan Univ., Konya.

Wang, J., An, Q., Zhou, D., Gao, K., & Bi, S. (2026). Prediction methods for high pressure density of biofuels: A hybrid data-driven ML-PC-SAFT equation of state. Fuel, 418, 138721.

Uyumaz, A., Solmaz, H., Yılmaz, E., Yamık, H., & Polat, S. (2014). Experimental examination of the effects of military aviation fuel JP-8 and biodiesel fuel blends on the engine performance, exhaust emissions and combustion in a direct injection engine. Fuel Processing Technology, 128, 158-165.

Verduzco, L. F. R. (2013). Density and viscosity of biodiesel as a function of temperature: Empirical models. Renewable and Sustainable Energy Reviews, 19, 652-665.

Zhou, L., Liu, Z. W., & Wang, Z. X. (2015). Numerical study of influence of biofuels on the combustion characteristics and performance of aircraft engine system. Applied Thermal Engineering, 91, 399-407.

Cite this article: Kurt, B. (2026). Prediction of Biofuel Density Obtained by Blending JP-5 Jet Fuel at Different Ratios with Machine Learning Algorithms. Journal of Aviation, 10(1), 24-31.