

Research Paper

Integrating Synthetic Data with Deep Learning for Predictive Modelling and Optimization of Diesel Engine Performance on Waste Plastic Oil Blends

Fitra Hidiyanto¹✉, Rizqon Fajar¹, Fauzi Dwi Setiawan¹, Sigit Tri Atmaja¹, Heru Priyanto¹, Muhammad Samsul Maarif^{1,2}, Yaaro Telaumbanua¹

¹Research Center for Transportation Technology, National Research and Innovation Agency of Indonesia, Tangerang Selatan 15314, Indonesia

²Mechanical Precision Engineering, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia (UTM), Kuala Lumpur 54100, Malaysia

✉ fitr009@brin.go.id

🌐 <https://doi.org/10.31603/ae.14360>

Published by Automotive Laboratory of Universitas Muhammadiyah Magelang

Abstract

Article Info

Submitted:

30/07/2025

Revised:

09/10/2025

Accepted:

01/11/2025

Online first:

06/02/2026

The scarcity of experimental data for diesel engines fueled by waste plastic oil (WPO) is a critical obstacle to optimizing engine performance. In this study, only 42 experimental data points covering six blend ratios and seven load conditions were available. To overcome this limitation, 121 synthetic data points were generated by training a suite of machine-learning models—Random Forest, Gradient Boosting, and AdaBoost—on the original dataset and then predicting outputs across a grid of WPO blend ratios (0–50% in 5% increments) and engine loads (0–100% in 10% increments). The synthetic data were rigorously validated using Kolmogorov–Smirnov tests, kernel density estimation, and principal component analysis to ensure statistical similarity with the original measurements. Subsequently, a Multi-Input Multi-Output (MIMO) deep neural network was trained on the combined real and synthetic dataset to predict four key performance metrics—power, torque, specific fuel consumption (SFC) and brake thermal efficiency (BTE)—and its hyperparameters were fine-tuned using Bayesian optimization via Optuna, achieving coefficients of determination (R^2) above 0.95. Optimization analysis indicated that a 17% WPO blend at 82% load delivers the best trade-off between power, efficiency and fuel consumption for non-road applications. This integrated framework demonstrates how synthetic data generation, rigorous validation and deep-learning modelling can effectively mitigate data scarcity and provide actionable insights for performance optimization of plastic pyrolysis oil in diesel engines.

Keywords: Waste plastic oil; Pyrolysis; Diesel engine performance; Synthetic data generation; Engine optimization, Deep learning modelling

1. Introduction

Waste plastic pyrolysis has emerged as an effective strategy to mitigate plastic waste accumulation while producing a liquid fuel known as waste plastic oil (WPO) [1]. Laboratory investigations show that pyrolysis oils comprise predominantly middle-distillate hydrocarbon chains ($\approx C_{12}$ – C_{20}) similar to diesel fuel, yielding high calorific values and allowing direct use in compression-ignition engines without extensive

modifications [2]. However, WPO exhibits higher viscosity, density and lower cetane index than petroleum diesel, which impairs atomisation and prolongs ignition delay during injection [3]. These physicochemical differences lead to longer premixed combustion phases and higher diffusion-burn durations [4], reducing thermal efficiency and increasing pollutants such as unburned hydrocarbons, carbon monoxide, smoke and nitrogen oxides [4]. Despite these



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

challenges, moderate blending of WPO with diesel has demonstrated promising performance improvements. Experimental studies with low-density plastic pyrolysis oils (10–50% WPO) reported increases in brake thermal efficiency (BTE) and reductions in specific fuel consumption (SFC), identifying 30% WPO as a suitable compromise for diesel substitution [5]. Similar work on high-density polyethylene WPO blends (5–15%) found that BTE improved at low blend levels and CO₂ emissions decreased, highlighting the potential of WPO for sustainability [6]. Yet other investigations note that higher WPO ratios ($\geq 40\%$) increase SFC and emissions due to poor spray atomisation [7]; thus, identifying optimal blend levels and operating conditions remains non-trivial.

The existing literature reveals a twofold limitation. First, experimental datasets are small because engine testing across multiple blend ratios and load conditions is costly and time-consuming; typical studies report fewer than fifty data points [1]. This scarcity of training data constrains the predictive power of conventional statistical models and hinders the application of advanced machine-learning algorithms requiring extensive data. Second, most modelling efforts rely solely on empirical data. For instance, a study using artificial neural networks (ANNs) to predict WPO-diesel engine performance achieved impressive regression coefficients ($R \approx 0.99$) [8], yet the model was trained on a narrow dataset (0–30% WPO blends), limiting its generalisability. Response-surface methodologies (RSM) have also been applied to optimize WPO ratio and engine load; a recent RSM study using XLPE-derived WPO determined an optimum blend of 19.6% at 2600 W load (BTE 22.3%, SFC 332.3 g kW⁻¹ h⁻¹, $R^2 > 0.98$) [9]. Although informative, these models cannot extrapolate to untested conditions. The small-data constraint therefore represents a critical research gap: without larger datasets, the community cannot fully explore the parameter space (blend ratio, load, injection timing, etc.) necessary for practical implementation of WPO fuels.

Synthetic data offer a promising solution to overcome data scarcity. In fields such as power-transmission systems, researchers have generated large synthetic datasets to train machine-learning models when real operational

data are inaccessible [10]. Similarly, manufacturing researchers use synthetic datasets created from computer-aided designs and object-detection algorithms to train visual inspection models, thereby reducing manual data collection [11]. Mechanical design researchers have developed deep-learning-based generative pipelines to produce synthetic structural datasets that expand the design space and improve surrogate modelling accuracy [12]. Guidelines for synthetic data generation in engineering emphasise the need for statistical validation and careful sampling to ensure that synthetic points reflect realistic behaviours [13]. In the context of energy systems, a large synthetic dataset was published for power-grid load forecasting to address the lack of open operational data [10]. These precedents demonstrate that synthetic datasets can augment limited experimental observations, enabling robust machine-learning models and design optimization. To date, however, little work has incorporated synthetic data into modelling diesel engines fuelled by WPO. The few engine studies using synthetic data remain confined to physics-based simulations rather than data-driven approaches.

This paper seeks to fill that gap by integrating synthetic data generation with deep learning for predictive modelling and performance optimization of WPO-diesel engines. Our approach begins with a limited experimental dataset comprising six WPO blend ratios (0–50%) and seven load conditions (0–100%), yielding 42 observations. We train multiple regression models—Random Forest, Gradient Boosting, and AdaBoost—on these data and use the best performers to generate 121 synthetic samples that span the continuous space of blend ratios and loads. The synthetic data are statistically validated against the original dataset using Kolmogorov–Smirnov tests, kernel-density estimates and principal-component analysis to ensure distributional similarity (no significant differences at the 5% level). Combining real and synthetic data, we develop a multi-input multi-output deep neural network (MIMO-DNN) to predict four key performance metrics: power, torque, SFC and BTE. Bayesian optimization (Optuna) is employed to tune the DNN architecture, learning rate and regularization parameters. This hybrid framework enables reliable extrapolation across

the blend-load space, outperforming prior ANN models trained on limited datasets. Furthermore, we apply the trained model to identify optimal WPO blend ratios and operating points for non-road applications (e.g., agricultural tractors, construction machinery), evaluating trade-offs between performance and emissions. Our study contributes methodological innovations—combining synthetic data augmentation, rigorous statistical validation and deep-learning optimization—and provides actionable insights for deploying WPO as a sustainable diesel substitute in real-world engines.

2. Literature Review

2.1. Physicochemical Challenges of Waste Plastic Oil and the Role of AI-Based Modelling in Diesel Engine Optimization.

Waste plastic oil (WPO), derived from plastic pyrolysis, presents significant physicochemical challenges as a diesel engine fuel, especially in non-road applications. In general, WPO exhibits lower viscosity and density and a lower cetane index than conventional diesel [2], [14]. These properties are associated with the efficiency of fuel atomization, which directly relates to the quality of air-fuel mixing. A longer ignition delay at the start of combustion was obtained with WPO, resulting in the higher peak of In-Cylinder Pressure and Rate of Heat Release compared to that of diesel fuel [3]. Consequently, combustion stability is degraded – engines running on WPO tend to experience less complete, more irregular combustion. Empirical studies report elevated emissions of unburned hydrocarbons (HC) and carbon monoxide (CO) with WPO, attributed to insufficient combustion time and the fuel's low cetane and heavy fractional composition [3], [14].

Particulate smoke can also increase due to the higher aromatic and sulphur content in WPO [14]. The effect on NO_x emissions is more complex: cooler, prolonged ignition phenomena with WPO can suppress peak NO_x formation in some cases, whereas in other cases the abrupt premixed burn after a long delay raises in-cylinder temperatures and increases NO_x production. These combustion and emission instabilities underscore the difficulty of optimizing engine performance with WPO using traditional approaches alone. Recent literature therefore highlights the necessity of artificial intelligence (AI) and machine-learning

(ML) techniques – particularly artificial neural networks (ANNs) – to predict and optimize diesel engine behaviours on such alternative fuels under varied operating conditions as done by A. Wiangkham *et al.* who uses NSGA-II with GRNNs to optimize the process of using WPO in diesel engines [15].

Data-driven ANN models have demonstrated superior capability in capturing the nonlinear relationships between fuel properties, engine parameters, and performance outcomes, outperforming conventional empirical models in prediction accuracy [16]. An ANN trained on experimental diesel-WPO blend data achieved near-perfect correlation ($R \approx 0.99$) in predicting engine performance and emissions. This high accuracy surpasses traditional regression models, demonstrating AI's ability to extrapolate engine responses across various conditions. AI models significantly reduce the need for extensive physical testing by learning from limited datasets and rapidly evaluating multiple scenarios. However, their predictive reliability is constrained by small experimental datasets. To address this, this study augments experimental data with statistically validated synthetic data, enhancing model robustness and generalization.

In summary, advanced ML techniques are becoming indispensable for harnessing WPO in diesel engines, enabling data-driven optimization of combustion stability and emissions performance that would be difficult to achieve with conventional methods alone.

2.2. Synthetic Data Augmentation for Limited Experiments

When experimental data are limited, one emerging approach is to generate synthetic data to enrich the training set for AI models. Synthetic data generation can involve statistical sampling, physics-based simulations, or ML-generated predictions to cover untested conditions. In automotive and engine domains, techniques such as design of experiments (DoE) combined with response surface modelling have long been used to interpolate between sparse test points, as done by D. Huri *et al.* whose research using DoE and RSM to optimize the shape of a rubber bumper in the automotive domain [17]. More recently, generative ML models (e.g., Gaussian Process regressors or even GANs) have been used to

simulate plausible data points that are consistent with known trends [18].

While synthetic data augmentation is widely used in aerospace diagnostics and medical imaging, its application in diesel engine modelling remains largely unexplored. Most studies focus on using AI with existing data rather than generating new training datasets. Drawing parallels from aerospace research, where synthetic data improves rare event detection [19], this study pioneers its use in WPO-diesel blends. By bridging experimental data gaps with ML-generated synthetic data, the approach enhances predictive accuracy and expands the applicability of AI-driven engine modelling [20]. The success of those efforts suggests that strategically generated synthetic data can improve model training, provided the synthetic points are statistically representative of real phenomena, as have shown by Jason W. A. *et al.*, who found that models trained on datasets augmented with synthetic images were more accurate than those trained on real images alone [21]. Ensuring this representativeness is crucial – techniques like the Kolmogorov–Smirnov (KS) test have been used to compare distributions of real and synthetic data [22].

In our context, we leverage multiple regression models (Random Forest [23], Gradient Boosting [24], etc.) trained on the original WPO experiment to predict engine performance at new blend ratios and operating points. By doing so, we fill the gaps between measured data points with plausible values. Each candidate synthetic dataset is rigorously evaluated against the original data distribution (using KS statistic, density plots, and other distance measures) to select a synthetic data pool that mimics the real data's statistical characteristics. This approach, to the best of our knowledge, has not been reported in prior engine studies and represents a novel contribution of this work: bridging experimental data gaps with ML-generated data as a form of intelligent augmentation for deep learning model training.

2.3. Research Gap and Novelty

From the above review, two key gaps can be identified. First, while WPO has been tested in engines experimentally, the parameter space (blend ratios, engine loads, etc.) explored in literature is limited, and results vary across studies. There is a lack of comprehensive models

that can generalize engine performance for any given WPO blend and operating condition, which is crucial for engine designers and policymakers considering WPO usage. Second, AI models (especially ANNs) have shown promise in this domain, but prior works typically train these models on the limited experimental datasets available, without attempting to expand the data volume. No study so far has combined data augmentation with AI modelling for WPO fuel. Moreover, integrating such a model into a practical optimization framework (e.g., finding the best blend for a tractor engine duty cycle) has not been addressed.

This study integrates ML-based synthetic data generation with deep learning and Bayesian Optimization to overcome data limitations and improve predictive accuracy. By fine-tuning hyperparameters, the model ensures reliable performance predictions while enabling practical optimization of WPO blends for real-world applications, such as agricultural and construction machinery. By doing so, this study moves beyond simply modelling engine performance – it provides a pathway to optimize and apply the model for real-world scenarios (such as non-road engines in agriculture or construction).

In summary, our work fills the gap between experimental investigations of WPO in engines and the growing need for data-driven tools to predict and optimize performance for sustainable fuel applications. The following sections will detail how this gap is addressed through our methodology and present the outcomes that underscore the contributions of this novel approach.

3. Method

3.1. Experimental Setup

The experiments were conducted using a single-cylinder, four-stroke diesel engine with direct injection. The engine specifications are as shown in Table 1. The fuel blends tested consisted of diesel fuel mixed with waste plastic oil (WPO) derived from Low-Density Polyethylene (LDPE) through pyrolysis. Six volumetric WPO-diesel blend ratios (0%, 10%, 20%, 30%, 40%, and 50%) were tested. The WPO was produced by pyrolyzing LDPE plastic waste at 400 °C – 700 °C in an oxygen-free environment [25]. The resulting pyrolysis oil had physicochemical properties com-

Table 1. Test diesel engine specification [25]

Parameter	Specifications
Engine model/type	: Diesel LC178F (D), single-cylinder 4- stroke, direct injection
Displacement	: 349 cc
Bore x Stroke	: 78 mm x 73 mm
Compression ratio	: 21:1
Maximum power	: 4.5 kW at 3000 rpm
Cooling system	: Water-cooled
Lubrication system	: Forced lubrication
Fuel supply temperature	: Maintained at 18 ± 1 °C

parable to diesel fuel, including similar calorific value, density, and cetane number, which made it suitable as a diesel substitute or additive.

The engine was tested at constant speed (2035 rpm) under varying loads (0-100%) to evaluate power, torque, SFC, and thermal efficiency [25]. The load was varied from no load to full load to assess performance across a range of operating conditions. Each WPO blend was tested independently to observe its effect on engine performance compared to pure diesel fuel.

3.2. Data Collection and Pre-processing

3.2.1. Experimental data

From the four plots in the paper by Sunaryo *et al.* [25] — namely Engine Power vs. Load, Engine Torque vs. Load, Specific Fuel Consumption (SFC) vs. Load, and Brake Thermal Efficiency (BTE) vs. Load — we systematically sampled the data using WebPlotDigitizer, a widely recognized tool for accurate data extraction from graphical representations. The digitized data, one of which is illustrated in Figure 1, were subsequently used for model training and validation.

WebPlotDigitizer is a free, web-based software developed by Ankit Rohatgi that enables accurate quantitative extraction of data from a wide variety of graphical plots through axis calibration and image processing algorithms. As demonstrated by Marin *et al.* [26], WebPlotDigitizer was successfully employed to retrieve ultraviolet spectropolarimetric data from historical observations of the Seyfert-2 galaxy NGC 1068. Similarly, Cicala *et al.* [27] utilized the software to extract sonic log data from vintage well logs in the ViDEPI project for geological applications, while Dashti *et al.* [28] applied it to extract experimental data for modelling pesticide photodegradation in environmental research.

3.2.2. Data Extraction

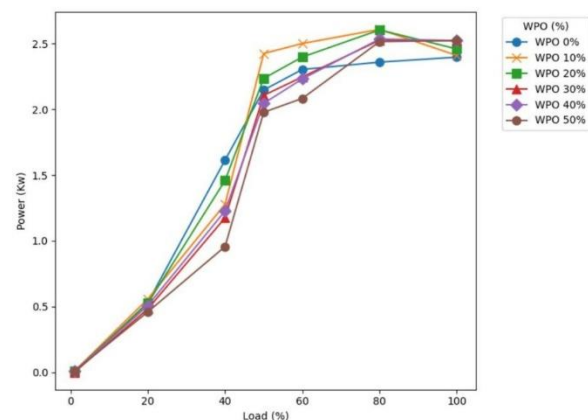
In this study, the author sampled data only at specific experimental points, namely at 7 load levels of 0%, 20%, 40%, 50%, 60%, 80%, and 100%. For each of these load conditions, all variations of WPO were included, specifically WPO levels of 0%, 10%, 20%, 30%, 40%, and 50%, resulting in a total of 42 data points as shown in Figure 1.

Five-fold cross-validation was applied to the 42 standardized observations to identify suitable models. Candidate regression models—including Random Forest, Gradient Boosting, AdaBoost, Neural Network, Linear Regression and k-nearest neighbours—were trained on the training subsets and evaluated on the validation subsets. The performance metrics used to compare models were the coefficient of determination (R^2), mean squared error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE), averaged over the five folds. This cross-validation strategy mitigates overfitting and ensures a fair comparison despite the small sample size. The models achieving the highest average R^2 and lowest error metrics were selected for each output variable (power, torque, SFC and BTE).

3.3. Predicted-Based Synthetic Data

3.3.1. Model architecture

This study utilizes Orange Data Mining, an open-source platform, to train machine learning models (e.g., Gradient Boosting, Random Forest) on experimental diesel-WPO blend data [29]. Unlike conventional interpolation, this approach ensures that synthetic data maintains the functional relationships inherent in experimental measurements, improving predictive accuracy. This predicted-based method involves selecting a

**Figure 1.** Engine Diesel load vs Engine Power (KW)

regression model that yields the highest accuracy on experimental data and then using it to extrapolate new data points under unexplored engine load and fuel blend conditions.

Six ML models (Random Forest [16], AdaBoost [30], Gradient Boosting [31], Neural Network [32], Linear Regression [33], and K-Nearest Neighbours [34]) were evaluated for each output using R^2 , MSE, MAE, and MAPE [35]. The best-performing model generated synthetic data across all WPO-load combinations. This approach ensures a more comprehensive representation of engine performance across a broader range of operational conditions not fully captured by the initial experiments.

This architecture ensures that the generated synthetic data maintains the functional relationships inherent in the experimental measurements, while allowing for scalable exploration of engine behaviour beyond the original dataset scope. By using predicted outputs rather than generative sampling, the method offers a transparent and interpretable approach suited to engineering performance modelling [36].

3.3.2. Synthetic Data Generation

The synthetic data were generated by applying the best-performing supervised machine learning models, trained on the original experimental dataset [37], to a set of predefined engine load and fuel blend scenarios. Rather than replicating statistical distributions, this predictive approach maintains the functional and multivariate relationships between engine performance parameters and fuel composition, as learned directly from the experimental data. The synthetic dataset was generated to provide a complete coverage of WPO values ranging from 0% to 50% in 5% increments, and engine load conditions from 0% to 100% in 10% increments.

The synthetic data was validated using Kolmogorov-Smirnov (KS) tests [22], [38], Kernel Density Estimation (KDE) [39], heatmap correlation [38], [40], to confirm statistical similarity with experimental data [41]. Distribution distance metrics [42] and spatial mapping validated the model's ability to capture data variability, preparing MSE and R^2 for regression evaluations [36].

3.4. Deep Learning Modelling, Optimization and Metrics Performance

3.4.1. Deep learning modelling

In this study, we developed a machine learning model employing a Deep Learning Neural Network with a Multi-Input Multi-Output (MIMO) architecture [43], as illustrated by the flowchart Figure 2a and MIMO model in Figure 2b. The model was trained on a combined dataset consisting of both real experimental data and synthetic data.

This model is employed to predict key engine performance metrics such as power (kW), torque (Nm), specific fuel consumption (SFC, kg/kWh), and brake thermal efficiency (BTE, %) [43] based on the blend ratio of plastic pyrolysis oil (WPO)

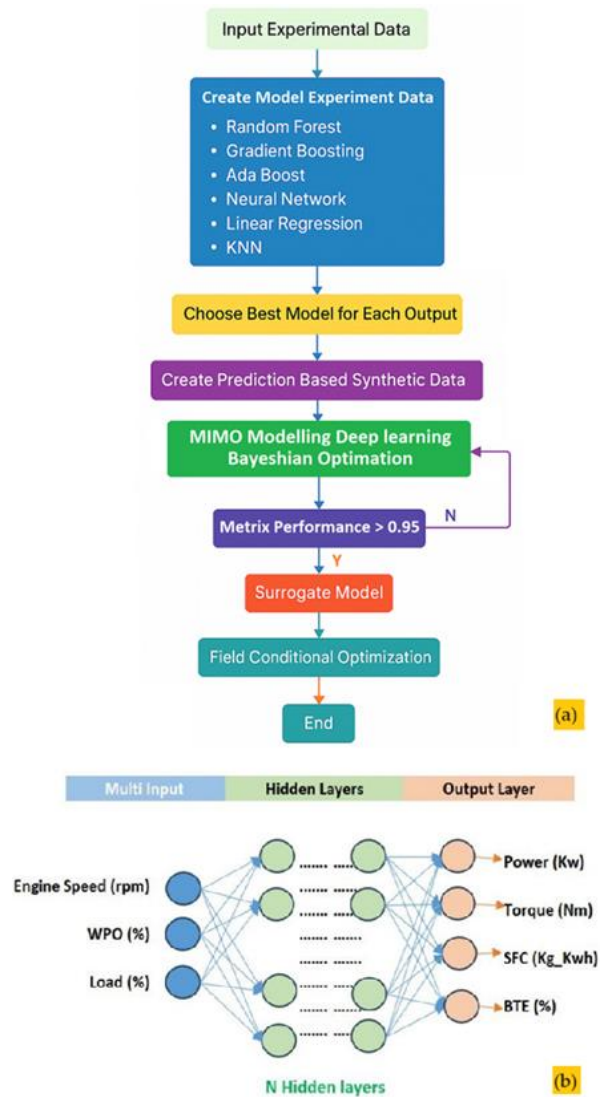


Figure 2. (a) Flowchart of the integrated modelling and optimization framework; (b) Schematic of the multi-input multi-output (MIMO) deep learning architecture

and operational conditions such as load, while maintaining a constant engine speed of 2035 RPM. By leveraging the generalization capabilities of deep learning to model nonlinear and multivariate relationships among parameters, the system provides performance estimations that closely approximate real experimental results across various alternative fuel scenarios.

The process involves training a deep neural network with N hidden layers, where each input is processed through these layers to generate the corresponding outputs as shown in Figure 2b. The system aims to optimize engine performance by learning complex, non-linear relationships in the data, ensuring high prediction accuracy [44]. Bayesian optimization is employed to fine-tune the hyperparameters of the model, ensuring that the metrics of performance reach a good and acceptable coefficient of determination (R^2).

3.4.2. Hyperparameter optimization

To optimize the hyperparameters of the deep neural network model and ensure the development of a high-performance surrogate model, this study employed the Bayesian optimization method [23] using the Optuna library an open-source Python library widely recognized for its efficiency and popularity in hyperparameter tuning [45], [46]. Optuna leverages the Tree-structured Parzen Estimator (TPE), which serves as its core algorithm, offering effective optimization for non-linear, adaptive models with fast convergence capabilities [47], [48].

Optuna was selected over Keras Tuner [46] due to superior computational efficiency despite similar optimization performance. The hyperparameter optimization was carried out over 500 epochs and 150 trials, with the following search space: number of layers ranging from 1 to 5, number of units per layer from 32 to 128, dropout rates between 0.1 and 0.5, and learning rates from 1e-5 to 1e-2 [49].

3.4.3. Metrics performance

The performance evaluation of the digital twin model for a diesel engine operating on plastic pyrolysis oil was conducted using four key metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the Coefficient of

Determination (R^2). These metrics collectively offer a robust assessment of both the predictive accuracy and generalization ability of the model.

MSE [50] Eq. (1), calculates the average of the squared differences between the predicted values $\hat{y}_i$ and the actual values y_i , and is expressed as:

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

MAE [51] Eq. (2) [52], represents the average of the absolute differences between predicted and actual values:

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

MAPE [53] Eq. (3), provides a percentage-based measure of prediction accuracy and is calculated as [54]:

$$MAPE = \frac{1}{n} \times \sum \left| \frac{x - y}{y} \right| \times 100 \quad (3)$$

R^2 [55] Eq. (4), quantifies how well the predicted values approximate the actual data, and is given by:

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

Among the four-performance metrics, the coefficient of determination (R^2) was prioritized as the primary evaluation metric, followed by mean squared error (MSE) as the secondary consideration.

3.4.4. Multi-criteria optimization for non-road diesel engines

To support diverse non-road applications—such as agriculture, construction, and mining—this study introduces a multi-criteria optimization framework that tailors diesel engine performance using WPO blends based on specific operational needs. A performance score function was formulated to balance power, torque, and specific fuel consumption (SFC), with tuneable weights (α , β , γ) [56], reflecting priority differences across sectors (Agriculture, manufacture, mining, Infrastructure) as done by Wiangkham *et al.* [55], who employed a weighted sum approach by weighting torque, power, and emissions to apply multi-criteria decision making (MCDM) with weights that can be adjusted for different applications (such as gen-sets, transportation, etc.), :

$$\text{Score} = \frac{(\alpha \times \text{Power}) + (\beta \times \text{Torque})}{\gamma \times \text{SFC}} \quad (5)$$

For example, agricultural engines might prioritize torque, while portable generators focus on maximizing power per fuel unit. This approach allows flexible trade-off analysis using real and synthetic data predicted by the deep learning model.

Such adaptive, data-driven frameworks are increasingly supported in the literature as effective tools for maximizing engine efficiency while minimizing emissions, as demonstrated by Wiangkham *et. al* [55] in the context of multi-objective optimization using biofuel blends. The present study extends this approach by combining synthetic data and deep learning, offering an interpretable and practical tool for operational tuning of WPO-fuelled engines.

4. Results and Discussion

4.1. Machine Learning-Based Synthetic Data Generation and Generalization Assessment

Synthetic data analysis confirms that models from Orange Data Mining preserve the original dataset's statistical distributions and relationships (Table 2). While AdaBoost, Gradient Boosting, and Random Forest all showed high performance ($R^2 > 0.97$), the final selection for each output was based on a multi-criteria decision prioritizing R^2 , followed by MAPE and MSE.

AdaBoost was chosen for Power due to its highest R^2 (0.98) and lowest MAPE (1.00%), ensuring superior predictive accuracy with minimal percentage error. For Torque, Gradient

Boosting was selected for its highest R^2 (0.98) combined with the lowest MSE (0.31) and MAPE (0.11%), signifying the best overall precision. Random Forest was optimal for SFC and BTE, delivering the highest R^2 values (0.95 and 0.98) and the lowest MAPE for BTE (2.28%), demonstrating robustness for these critical efficiency metrics. A total of 121 high-fidelity synthetic samples were generated using these selected models.

The Kolmogorov–Smirnov test (Table 3) showed no significant distributional differences ($p > 0.4$) much greater than standard value ($p > 0.05$), supporting the synthetic data's reliability. Furthermore, a comparison of the central tendency and dispersion of the original and synthetic datasets (Table 4) confirms their statistical similarity.

Figure 3a to Figure 3f shows that synthetic and original distributions overlap consistently across Load, WPO, Power, Torque, SFC, and BTE. Moreover, the synthetic data spans a wider value range, enriching the dataset with new combinations valuable for assessing engine behaviours under diverse operating conditions.

Heatmap visualizations of feature correlations in Figure 4 revealed that the relationships among key parameters, such as Torque, BTE, and Power, were successfully preserved in the synthetic data. These observations align with the findings of I. E. Livieris *et al.* [42], who emphasized the importance of preserving multivariate relationships in synthetic data for effective generalization.

Table 2. Predicted based model synthetic data by orange-data-mining tool best model result

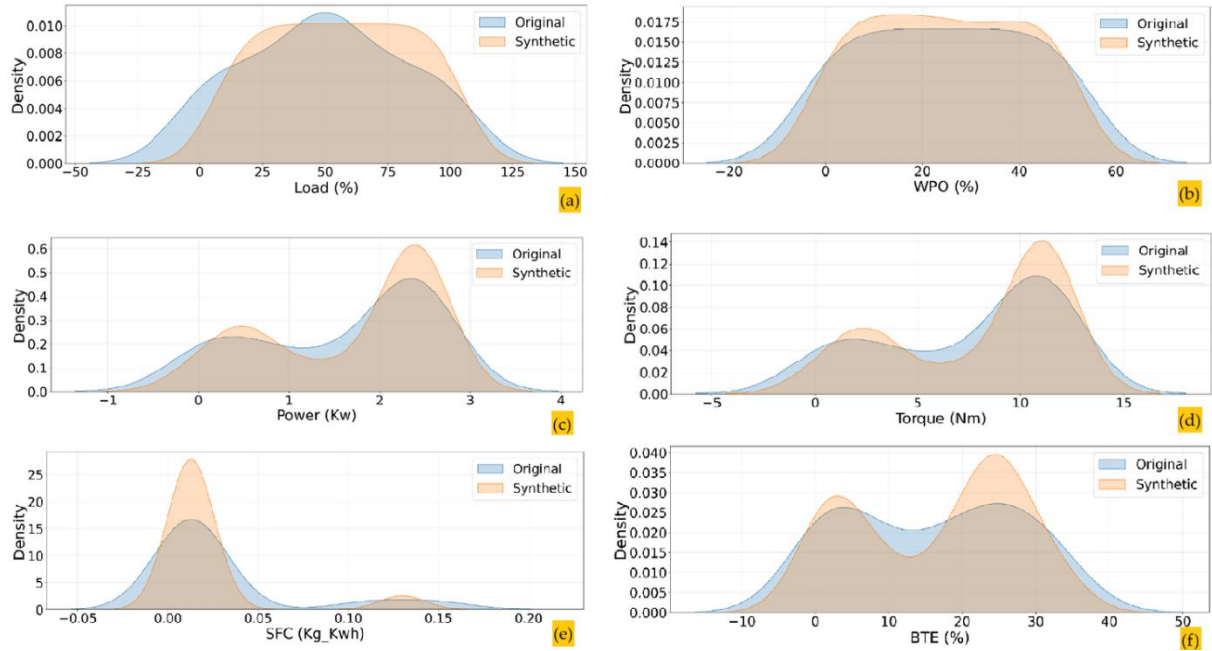
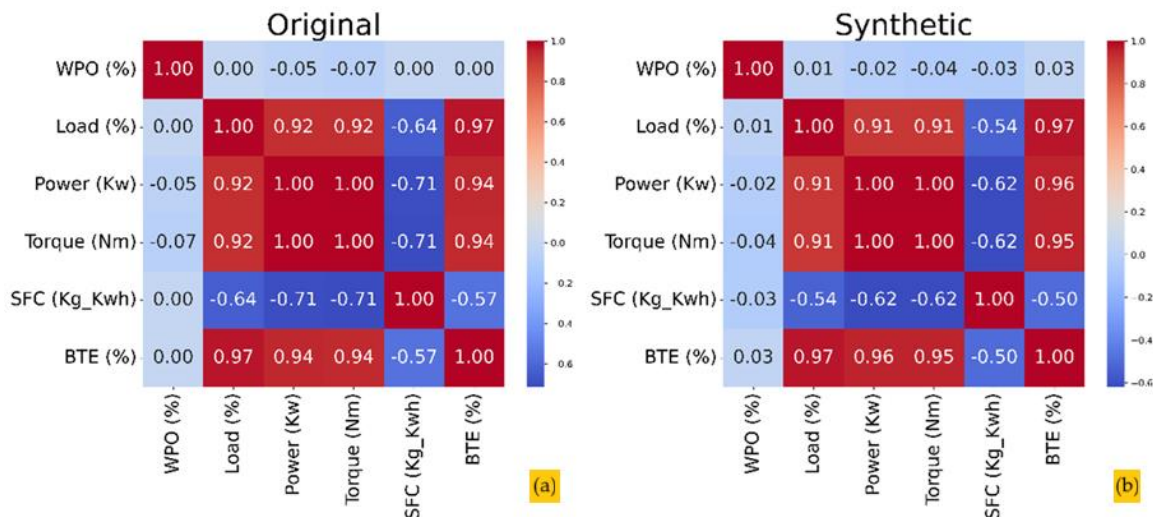
Model	MSE				MAPE				R^2			
	Power	Torque	SFC	BTE	Power	Torque	SFC	BTE	Power	Torque	SFC	BTE
Random Forest	0.02	0.54	0.00	2.85	14.78	0.18	0.17	2.28	0.98	0.97	0.95	0.98
Gradient Boosting	0.02	0.31	0.00	3.18	5.21	0.11	0.17	1.05	0.98	0.98	0.87	0.97
AdaBoost	0.02	0.44	0.00	3.24	1.00	0.11	0.20	0.87	0.98	0.98	0.88	0.98
Neural Network	0.02	0.36	0.00	4.87	1.23	0.13	0.78	1.18	0.98	0.98	0.83	0.96
Linear Regression	0.16	2.98	0.00	8.27	38.12	0.64	1.57	4.90	0.83	0.83	0.28	0.94
kNN	0.06	1.09	0.00	8.96	22.47	0.56	0.39	3.68	0.93	0.94	0.66	0.93

Table 3. KS (Kolmogorov-Smirnov) and P-value distribution result test

Feature	KS Statistic	P-value	Interpretation
WPO (%)	0.0741	0.9911	Similar
Load (%)	0.1429	0.5166	Similar
Power (kW)	0.0794	0.9813	Similar
Torque (Nm)	0.1019	0.8766	Similar
SFC (kg/kWh)	0.1481	0.4709	Similar
BTE (%)	0.1283	0.6497	Similar
WPO (%)	0.0741	0.9911	Similar

Table 4. Comparative statistical analysis of original and synthetic datasets

Variable	Original_Mean	Synthetic_Mean	Original_Std	Synthetic_Std
WPO (%)	25.000000	24.861111	17.285268	15.994426
Load (%)	50.142857	55.833333	31.778633	28.453109
Power (Kw)	1.606588	1.695394	0.961576	0.921389
Torque (Nm)	7.584986	7.980257	4.281273	4.108889
SFC (Kg_Kwh)	0.029733	0.022562	0.042466	0.032674
BTE (%)	15.863779	16.568430	11.560770	10.989773

**Figure 3.** Distribution plot of original vs. synthetic all parameter, which (a) KDE plot - load (%), (b) KDE plot - WPO (%), (c) KDE plot - power (kW), (d) KDE plot - torque (Nm), (e) KDE plot - SFC (kg/kWh), (f) KDE plot - BTE (%)**Figure 4.** Pearson heatmap correlation matrix: (a) Original data correlation; (b) Synthetic data correlation

4.2. Deep Learning Modelling, Optimization and Metrics Performance

4.2.1. Surrogate model performance

A MIMO deep learning model with Bayesian-optimized hyperparameters (Optuna) was trained

on the combined dataset. Optimal hyperparameters: MAE loss, 3 layers [68,68,4], dropout [0.194,0.194], learning rate 0.0064776. The results demonstrate strong model performance as shown in [Figure 5](#), with coefficient of

determination (R^2) values of 0.982 for power shown in [Figure 5a](#), 0.994 for torque shown in [Figure 5b](#), 0.993 for specific fuel consumption (SFC) shown in [Figure 5c](#), and 0.982 for brake thermal efficiency (BTE) shown in [Figure 5d](#).

4.2.2. Comparison with experimental data

[Table 5](#) presents the performance metrics of the proposed model before synthetic data and after synthetic data augmentation. Although no study has directly generated synthetic WPO–Load data with an expanded range, similar methodologies have been used in engine performance modelling as [\[9\]](#) demonstrated performance parameter distribution broadening via WPO–ethanol blends under varying loads.

From [Table 5](#), we can observe that after using synthetic data generated through the predicted-based synthetic data method with the Orange Data Mining tool, the model performance improved, achieving R^2 values of 0.982 for Power, 0.994 for Torque, 0.993 for SFC, and 0.982 for BTE. This result was done using 42 engine experimental data points and 121 synthetic data points, totalling 163 data points. In this MIMO Deep learning

modelling author used 50% data experimental for training, 25% validation and 25% test data, and used 121 synthetic data as additional in training data.

4.3. Adaptive Optimization of WPO-Diesel Blends for Non-Road Diesel Engine Applications

The model ($R^2 > 0.98$) aligns with Zhang *et al.* [\[57\]](#) and Dutta *et al.* [\[43\]](#) enabling optimization of WPO blends for maximum power/torque and minimum fuel consumption. The optimization results for three non-road application scenarios namely generator sets, agricultural/plantation operations, and manufacturing industries are presented in [Table 6](#).

Using Eq. (5) for all condition, the optimal score was obtained at WPO 17% and load 82% [\[55\]](#). The 3D plot shown in [Figure 6](#) illustrates the results of a multi-objective optimization of diesel engine performance using Waste Plastic Oil (WPO) blend ratio and engine load as input variables. This visualization facilitates the identification of fuel blend ratios and load conditions that yield optimal engine performance

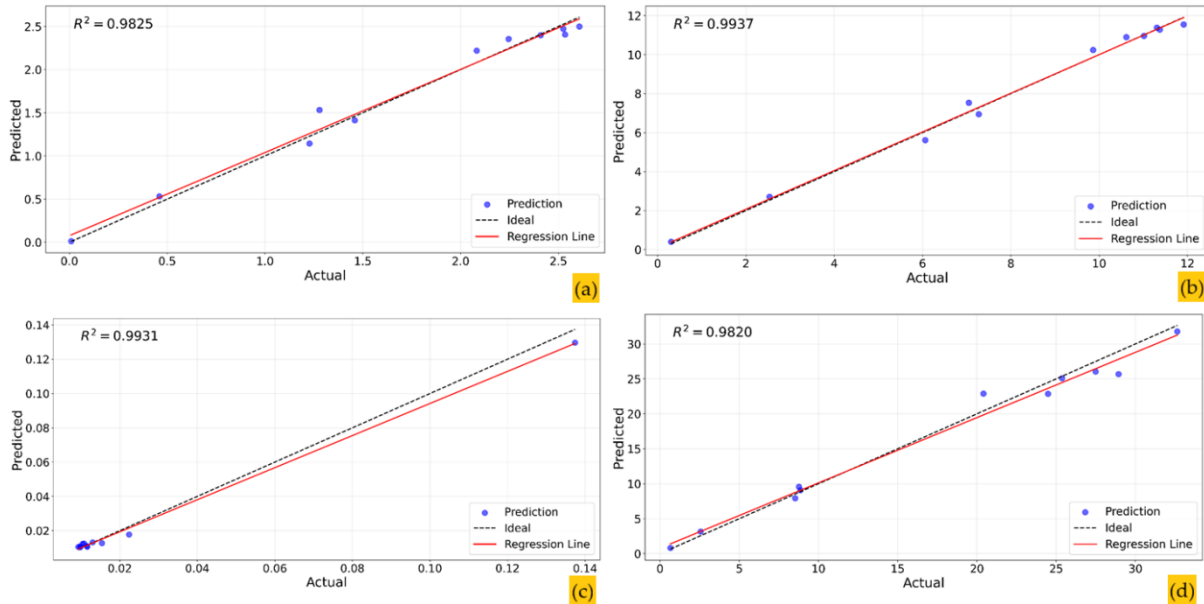


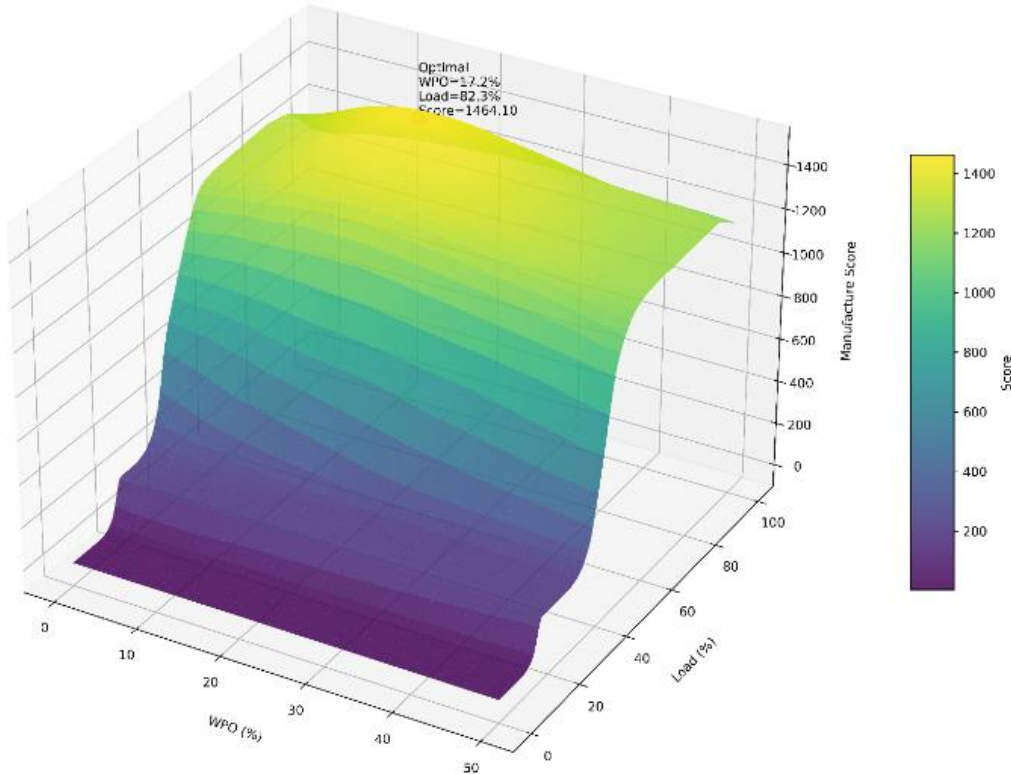
Figure 5. Real vs. predicted plot of diesel engine performance: (a) Power, (b) Torque, (c) SFC, (d) BTE

Table 5. Model performance comparison before and after synthetic data augmentation

Engine Output	MSE	MAE	MAPE	R^2	MSE	MAE	MAPE	R^2
	Before Synthetic data				After Synthetic data			
Power (kW)	0.020	0.127	104.836	0.973	0.013	0.091	8.717	0.982
Torque (Nm)	0.389	0.508	20.220	0.972	0.087	0.252	6.110	0.994
SFC (Kg/kWh)	0.000	0.003	13.059	0.988	0.000	0.002	11.022	0.993
BTE (%)	1.729	1.045	14.933	0.986	2.159	1.119	9.510	0.982

Table 6. Optimal WPO (%) and Load (%) values for three application scenarios

Category	α	β	γ	WPO	Load	Score
Generator	0.5	0.3	0.2	16.92	82.88	2319.30
Agriculture	0.3	0.5	0.2	17.46	82.29	3205.89
Manufacture	0.4	0.3	0.3	17.19	82.29	1464.15

**Figure 6.** 3D Surface Plot for Multi-Objective Engine Performance Optimization using WPO-Diesel Blends

with maximum fuel efficiency, thereby supporting decision-making in fuel blending and alternative fuel testing.

Figure 7 heatmap shows that load is positively correlated with power and torque, and negatively correlated with SFC [58]. **Figure 8** presents a spider web plot showing that 17% WPO achieves near-maximum scores across applications, whereas 50% WPO demonstrates poor performance due to its high SF value. Furthermore, **Figure 9** chord plot as used by Beltran, D. *et al.* [59], Hao, R. *et al.* [60], Zimta, A. *et al.* [61] shows 82% load as optimal, with 35-38% showing moderate gains and 7% yielding minimal performance, while loads between 82% and 100% dominate in terms of power and efficiency.

4.4. Discussion

This study demonstrates that ML-based synthetic data generation (Random Forest, Gradient Boosting, AdaBoost) successfully

expands limited experimental datasets while preserving statistical characteristics [62]. The MIMO deep learning model, optimized via Bayesian Optimization (Optuna), achieved $R^2 > 0.98$ for all parameters, consistent with Zhang *et al.* [57] in capturing nonlinear multivariate relationships.

KS tests and KDE validation [63] confirm the model captures inter-parameter relationships beyond memorization.

Optimization identified 17% WPO at 82% load as optimal for non-road applications, corroborating literature [64] on ~20% blends.

This modelling framework addresses data limitations in alternative fuel research and offers a scalable tool for optimizing diesel engine performance using WPO blends. Acting as a digital twin, it enables regulators and OEMs to virtually assess blend scenarios, streamline experimental efforts, and accelerate the adoption of sustainable fuels in non-road applications.

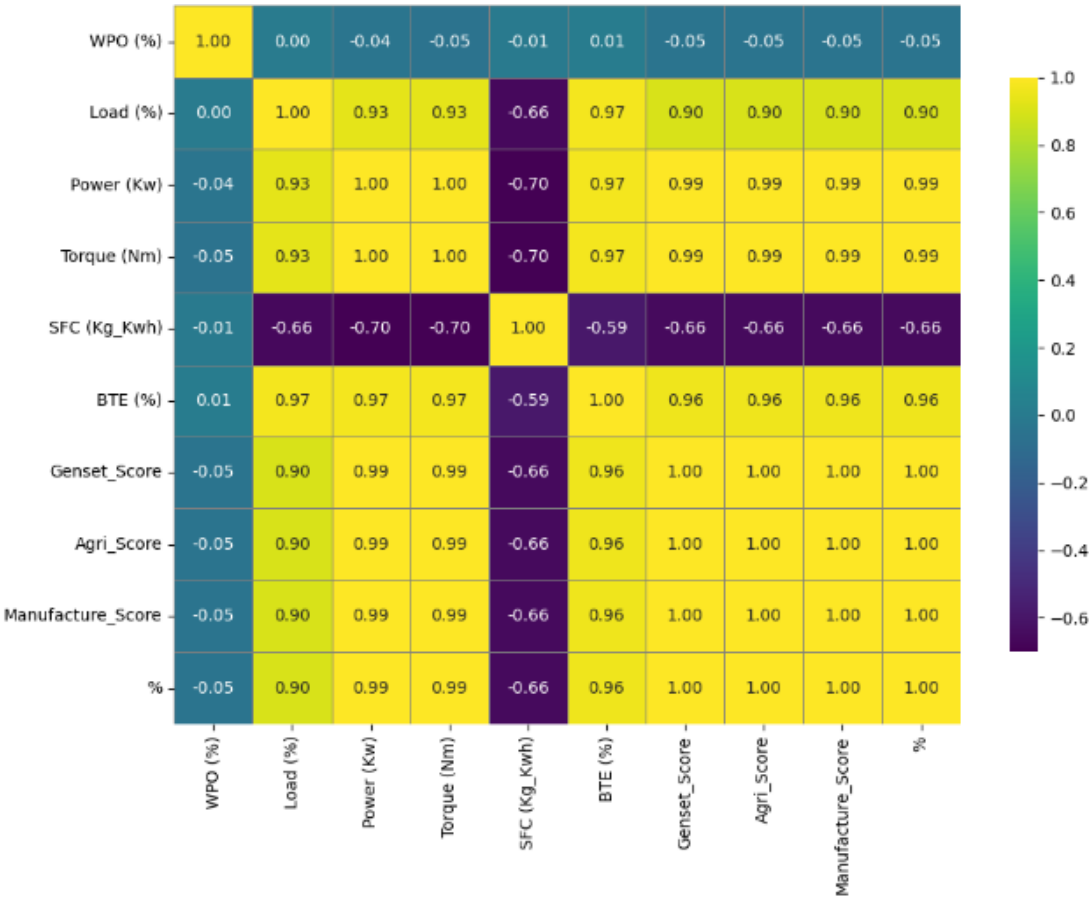


Figure 7. Heatmap correlation of WPO and Load to Score and performance

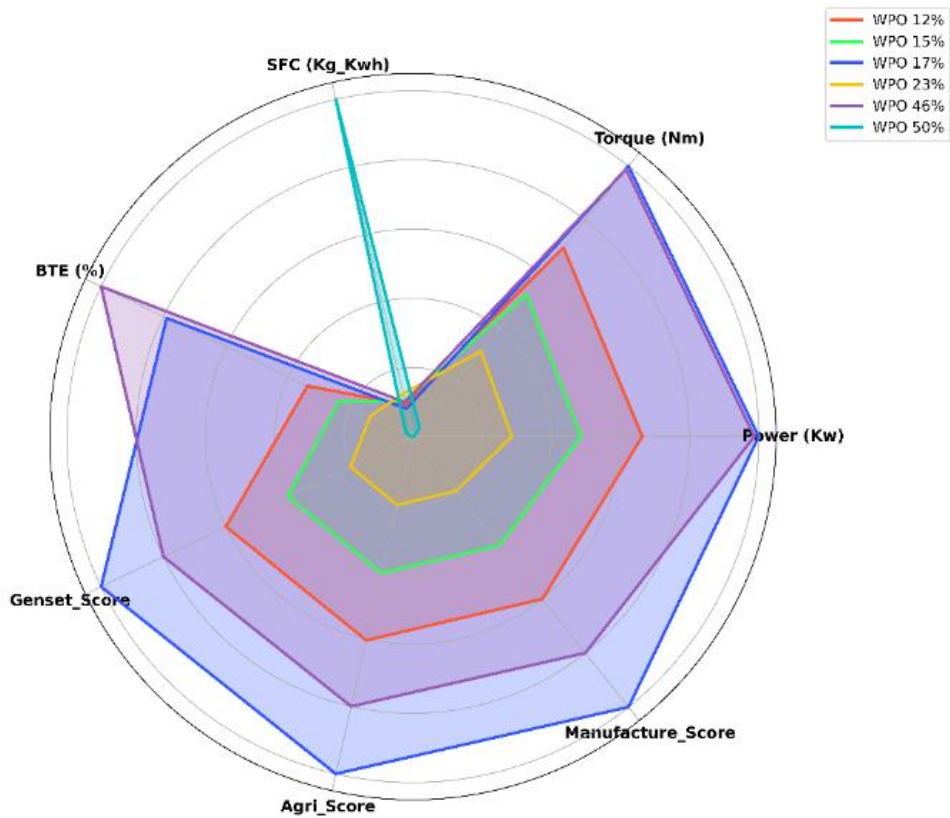


Figure 8. Radar Plot WPO to Score and performance

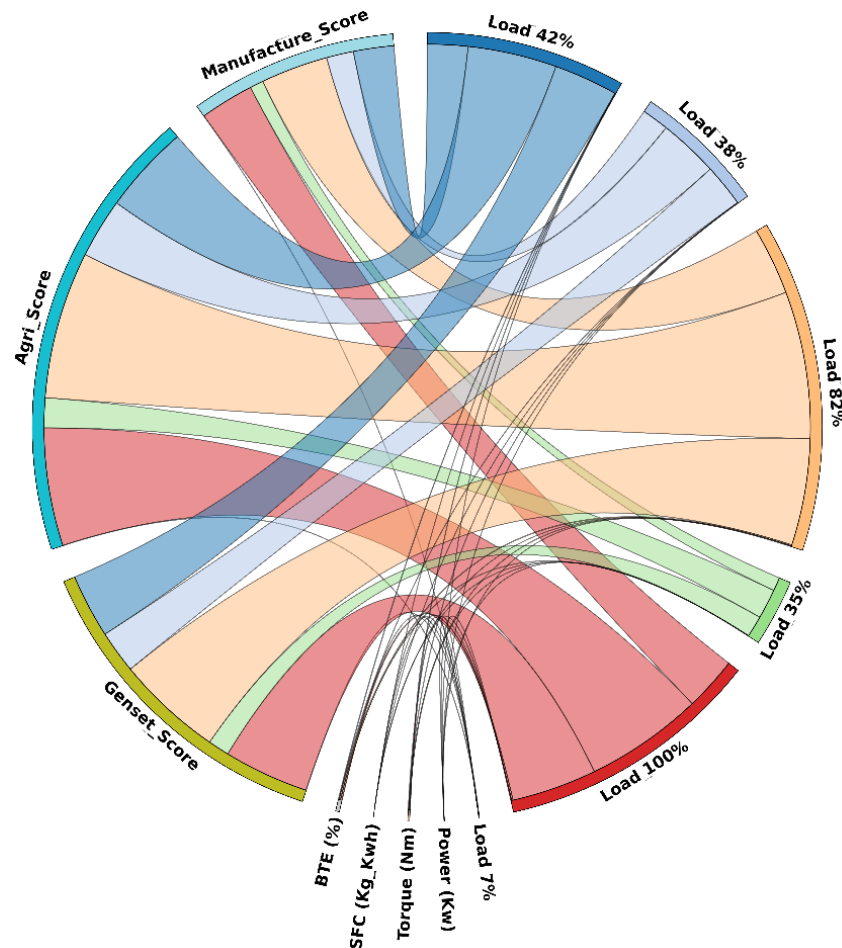


Figure 9. Chord plot of Load to performance and score

5. Conclusion

This study successfully developed a synthetic data-driven approach employing machine learning algorithms to evaluate and optimize the performance of diesel engines fueled with Waste Plastic Oil (WPO) blends. Utilizing a combination of Random Forest, Gradient Boosting, and AdaBoost models, the generated synthetic data was found to statistically resemble the characteristics of experimental observations. The deep learning-based Multi-Input Multi-Output (MIMO) model, optimized through Bayesian Optimization (Optuna), demonstrated high predictive accuracy ($R^2 > 0.98$) across all performance parameters, namely power, torque, specific fuel consumption (SFC), and brake thermal efficiency (BTE).

Validation via the Kolmogorov–Smirnov test and Kernel Density Estimation confirmed that the model did not merely memorize training data but effectively captured complex interdependencies among engine parameters. Optimization results

indicated that a WPO blend ratio of 17% and an engine load of 82% yielded the highest performance scores. These findings highlight the significant potential of WPO as a sustainable alternative fuel and provide a robust solution to the common challenge of limited experimental datasets in alternative fuel research.

Despite its limitations, this study offers a practical and scalable framework for optimizing WPO blends in diesel engines. The approach demonstrates strong potential to support regulators and OEMs by enabling pre-screening of feasible blends, guiding emission policy, and reducing development cycles. With further validation across diverse engine types and real-world conditions, this framework can accelerate the sustainable adoption of WPO as an alternative fuel in non-road applications.

Acknowledgements

We would like to express our sincere gratitude to the authors of the paper "Performance and Fuel

Consumption of Diesel Engine Fueled by Diesel Fuel and Waste Plastic Oil Blends: An Experimental Investigation" for granting us permission to use the data and graphical representations displayed in their study. Their work provided valuable insights and data that greatly contributed to the development of this research. Artificial Intelligence (AI)-assisted tools were used solely to enhance language clarity and to assist in graphical analysis and data interpretation stages. The authors reviewed and approved all content generated and take full responsibility for the integrity and accuracy of the final manuscript.

Author's Declaration

Authors' contributions and responsibilities

The authors made substantial contributions to the conception and design of the study. The authors took responsibility for data analysis, interpretation and discussion of results. The authors read and approved the final manuscript.

Funding

This paper is supported by funding from the research program Development of Generative Artificial Intelligence Models for the Performance of Diesel Engines Using Plastic Pyrolysis Fuel, funded by the Energy and Manufacturing Research Organization, National Research and Innovation Agency of Indonesia.

Availability of data and materials

All data are available from the paper by Sunaryo, P. A. Sesotyo, E. Saputra, and A. P. Sasmito, "Performance and fuel consumption of diesel engine fueled by diesel fuel and waste plastic oil blends: An experimental investigation," *Automot. Exp.*, vol. 4, no. 1, pp. 20–26, 2021, doi: 10.31603/ae.3692.

Competing interests

The authors declare no competing interest.

Additional information

No additional information from the authors.

References

- [1] W. Arjarn, P. Liplap, S. Maithomklang, K. Thammakul, S. Chuepeng, and E. Sukjit, "Distilled Waste Plastic Oil as Fuel for a Diesel Engine: Fuel Production, Combustion Characteristics, and Exhaust Gas Emissions," *ACS Omega*, vol. 7, no. 11, pp. 9720–9729, 2022, doi: 10.1021/acsomega.1c07257.
- [2] K. Wathakit *et al.*, "Characterization and impact of waste plastic oil in a variable compression ratio diesel engine," *Energies*, vol. 14, no. 8, 2021, doi: 10.3390/en14082230.
- [3] A. Pumpuang, N. Klinkaew, K. Wathakit, A. Sukhom, and E. Sukjit, "The influence of plastic pyrolysis oil on fuel lubricity and diesel engine performance," *RSC Advances*, vol. 14, no. 14, pp. 10070–10087, 2024, doi: 10.1039/D3RA08150H.
- [4] M. J. Szwaja and A. Szymanek, "Combustion comparative analysis of pyrolysis oil and diesel fuel under constant-volume conditions," *Combustion Engines*, vol. 195, no. 4, pp. 90–96, Aug. 2023, doi: 10.19206/CE-169805.
- [5] J. Senthil, M. Prabhakar, C. Thiagarajan, S. Prakash, and M. S. Kumar, "Evaluating Diesel Engine Performance with Blended Fuels from Low-Density Plastic Pyrolysis Oil: A Sustainable Approach to Diesel Engine Operation," *International Journal of Mechanical Engineering*, vol. 10, no. 10, pp. 26–36, Oct. 2023, doi: 10.14445/23488360/IJME-V10I10P103.
- [6] H. Yaqoob, H. M. Ali, U. Sajjad, and K. Hamid, "Investigating the potential of plastic pyrolysis oil-diesel blends in diesel engine: Performance, emissions, thermodynamics and sustainability analysis," *Results in Engineering*, vol. 24, no. October, p. 103336, Dec. 2024, doi: 10.1016/j.rineng.2024.103336.
- [7] S. Maithomklang, E. Sukjit, J. Srisertpol, N. Klinkaew, and K. Wathakit, "Pyrolysis Oil Derived from Plastic Bottle Caps: Characterization of Combustion and Emissions in a Diesel Engine," *Energies*, vol. 16, no. 5, 2023, doi: 10.3390/en16052492.
- [8] A. K. Panda, S. K. Rout, and A. K. Das, "Optimization of diesel engine performance and emission using waste plastic pyrolytic oil by ANN and its thermo-economic assessment," *Environmental Science and Pollution Research*, vol. 31, no. 27, pp. 38893–38907, Apr. 2023, doi: 10.1007/s11356-023-26891-9.
- [9] S. Şen and M. Bahattin Çelik, "Modeling the effect of plastic oil obtained from XLPE cable waste on diesel engine performance and emission parameters with the response surface method," *Science and Technology for Energy Transition*, vol. 79, p. 58, Aug. 2024, doi: 10.2516/stet/2024059.
- [10] M. Gillioz, G. Dubuis, and P. Jacquod, "A

- large synthetic dataset for machine learning applications in power transmission grids," *Scientific Data*, vol. 12, no. 1, pp. 1–17, 2025, doi: 10.1038/s41597-025-04479-x.
- [11] J. Werheid, S. He, A. Gannouni, A. Abdelrazeq, and R. H. Schmitt, "A Synthetic Data Pipeline for Supporting Manufacturing SMEs in Visual Assembly Control," in *2025 2nd International Generative AI and Computational Language Modelling Conference (GACLM)*, Sep. 2025, pp. 1–7. doi: 10.48550/arXiv.2509.13089.
- [12] S. Hong, Y. Kwon, D. Shin, J. Park, and N. Kang, "DeepJEB: 3D Deep Learning-Based Synthetic Jet Engine Bracket Dataset," *Journal of Mechanical Design*, vol. 147, no. 4, Apr. 2025, doi: 10.1115/1.4067089.
- [13] C. Picard, J. Schiffmann, and F. Ahmed, "Dated: Guidelines for Creating Synthetic Datasets for Engineering Design Applications," in *Volume 3A: 49th Design Automation Conference (DAC)*, American Society of Mechanical Engineers, Aug. 2023, pp. 1–13. doi: 10.1115/DETC2023-111609.
- [14] S. Maithomklang, K. Wathakit, E. Sukjit, B. Sawatmongkhon, and J. Srisertpol, "Utilizing Waste Plastic Bottle-Based Pyrolysis Oil as an Alternative Fuel," *ACS Omega*, vol. 7, no. 24, pp. 20542–20555, Jun. 2022, doi: 10.1021/acsomega.1c07345.
- [15] A. Wiangkham, N. Klinkaew, P. Aengchuan, P. Liplap, A. Ariyarit, and E. Sukjit, "Experimental and optimization study on the effects of diethyl ether addition to waste plastic oil on diesel engine characteristics," *RSC Advances*, vol. 13, no. 36, pp. 25464–25482, 2023, doi: 10.1039/d3ra04489k.
- [16] M. S. Aswathanrayan *et al.*, "Prediction of performance and emission features of diesel engine using alumina nanoparticles with neem oil biodiesel based on advanced ML algorithms," *Scientific Reports*, vol. 15, no. 1, p. 12683, Apr. 2025, doi: 10.1038/s41598-025-97092-2.
- [17] D. Huri and T. Mankovits, "Automotive Rubber Product Design Using Response Surface Method," *Periodica Polytechnica Transportation Engineering*, vol. 50, no. 1, pp. 28–38, Dec. 2021, doi: 10.3311/PPtr.16280.
- [18] M. Sheyyab, P. T. Lynch, E. K. Mayhew, and K. Brezinsky, "Optimized synthetic data and semi-supervised learning for Derived Cetane Number prediction," *Combustion and Flame*, vol. 259, p. 113184, Jan. 2024, doi: 10.1016/j.combustflame.2023.113184.
- [19] M. G. De Giorgi, S. Campilongo, and A. Ficarella, "A diagnostics tool for aero-engines health monitoring using machine learning technique," *Energy Procedia*, vol. 148, pp. 860–867, Aug. 2018, doi: 10.1016/j.egypro.2018.08.109.
- [20] I. Pan, L. Mason, and O. Matar, "Data-Centric Engineering: integrating simulation, machine learning and statistics. Challenges and Opportunities," *Chemical Engineering Science*, vol. 249, p. 117271, Nov. 2021, doi: 10.1016/j.ces.2021.117271.
- [21] J. W. Anderson, M. Ziolkowski, K. Kennedy, and A. W. Apon, "Synthetic Image Data for Deep Learning," *arXiv preprint arXiv:2212.06232*, Dec. 2022, doi: 10.48550/arXiv.2212.06232.
- [22] M. Stenger, R. Leppich, I. Foster, S. Kounev, and A. Bauer, "Evaluation is key: a survey on evaluation measures for synthetic time series," *Journal of Big Data*, vol. 11, no. 1, p. 66, May 2024, doi: 10.1186/s40537-024-00924-7.
- [23] M. Chaibi, E. M. Benghoulam, L. Tarik, M. Berrada, and A. El Hmadi, "Machine Learning Models Based on Random Forest Feature Selection and Bayesian Optimization for Predicting Daily Global Solar Radiation," *International Journal of Renewable Energy Development*, vol. 11, no. 1, pp. 309–323, Feb. 2022, doi: 10.14710/ijred.2022.41451.
- [24] P. Whig, K. Gupta, N. Jiwani, H. Jupalle, S. Kouser, and N. Alam, "A novel method for diabetes classification and prediction with Pycaret," *Microsystem Technologies*, vol. 29, no. 10, pp. 1479–1487, 2023, doi: 10.1007/s00542-023-05473-2.
- [25] S. Sunaryo, P. A. Sesotyo, E. Saputra, and A. P. Sasmito, "Performance and Fuel Consumption of Diesel Engine Fueled by Diesel Fuel and Waste Plastic Oil Blends: An Experimental Investigation," *Automotive Experiences*, vol. 4, no. 1, pp. 20–26, 2021, doi: 10.31603/ae.3692.
- [26] F. Marin, A. Rohatgi, and S. Charlot, "WebPlotDigitizer, a polyvalent and free

- software to extract spectra from old astronomical publications: application to ultraviolet spectropolarimetry," *arXiv preprint arXiv:1708.02025*, Aug. 2017, doi: 10.48550/arXiv.1708.02025.
- [27] M. Cicala, V. Festa, and A. Marrone, "Data extraction from vintage well sonic log graphs in the ViDEPI project (offshore the Apulia, southern Italy): A multi-useful dataset," *Data in Brief*, vol. 46, p. 108814, Feb. 2023, doi: 10.1016/j.dib.2022.108814.
- [28] A. Dashti, A. H. Navidpour, F. Amirkhani, J. L. Zhou, and A. Altaee, "Application of machine learning models to improve the prediction of pesticide photodegradation in water by ZnO-based photocatalysts," *Chemosphere*, vol. 362, no. January, p. 142792, Aug. 2024, doi: 10.1016/j.chemosphere.2024.142792.
- [29] Z. Dobesova, "Evaluation of Orange data mining software and examples for lecturing machine learning tasks in geoinformatics," *Computer Applications in Engineering Education*, vol. 32, no. 4, 2024, doi: 10.1002/cae.22735.
- [30] Á. Palma, M. Antunes, J. Bernardino, and A. Alves, "Multi-Class Intrusion Detection in Internet of Vehicles: Optimizing Machine Learning Models on Imbalanced Data," *Future Internet*, vol. 17, no. 4, pp. 1–14, 2025, doi: 10.3390/fi17040162.
- [31] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, doi: 10.1214/aos/1013203451.
- [32] S. Patnaik, N. Khatri, and E. R. Rene, "Artificial neural networks-based performance and emission characteristics prediction of compression ignition engines powered by blends of biodiesel derived from waste cooking oil," *Fuel*, vol. 370, p. 131806, Aug. 2024, doi: 10.1016/j.fuel.2024.131806.
- [33] K. Sunil Kumar et al., "Performance, Combustion, and Emission analysis of diesel engine fuelled with pyrolysis oil blends and n-propyl alcohol-RSM optimization and ML modelling," *Journal of Cleaner Production*, vol. 434, p. 140354, Jan. 2024, doi: 10.1016/j.jclepro.2023.140354.
- [34] F. Hidiyanto and A. Halim, "KNN Methods with Varied K, Distance and Training Data to Disaggregate NILM with Similar Load Characteristic," in *Proceedings of the 3rd Asia Pacific Conference on Research in Industrial and Systems Engineering 2020*, New York, NY, USA: ACM, Jun. 2020, pp. 93–99. doi: 10.1145/3400934.3400953.
- [35] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science*, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.
- [36] J. Xiong, O. Fink, J. Zhou, and Y. Ma, "Controlled physics-informed data generation for deep learning-based remaining useful life prediction under unseen operation conditions," *Mechanical Systems and Signal Processing*, vol. 197, 2023, doi: 10.1016/j.ymssp.2023.110359.
- [37] S. M. Aithal and P. Balaprakash, "MaLTESE: Large-Scale Simulation-Driven Machine Learning for Transient Driving Cycles," in *Lecture Notes in Computer Science*, 2019, pp. 186–205. doi: 10.1007/978-3-030-20656-7_10.
- [38] L. Hansen, A. Petrovic, N. Seedat, and M. Van Der Schaar, "Reimagining Synthetic Tabular Data Generation through Data-Centric AI: A Comprehensive Benchmark," in *Advances in Neural Information Processing Systems 36*, San Diego, California, USA: Neural Information Processing Systems Foundation, Inc. (NeurIPS), Oct. 2023, pp. 33781–33823. doi: 10.52202/075280-1466.
- [39] J. Heine, E. E. E. Fowler, A. Berglund, M. J. Schell, and S. Eschrich, "Techniques to produce and evaluate realistic multivariate synthetic data," *Scientific Reports*, vol. 13, no. 1, pp. 1–28, 2023, doi: 10.1038/s41598-023-38832-0.
- [40] J. Siswanto et al., "Short-Term Prediction of Bus Station Fleet Number Using a Combination of BiLSTM Models," *Automotive Experiences*, vol. 8, no. 1, pp. 205–215, Apr. 2025, doi: 10.31603/ae.13402.
- [41] S. Alabdulwahab, Y. T. Kim, and Y. Son, "Privacy-Preserving Synthetic Data Generation Method for IoT-Sensor Network IDS Using CTGAN," *Sensors*, vol. 24, no. 22,

- pp. 1–18, 2024, doi: 10.3390/s24227389.
- [42] I. E. Livieris, N. Alimpertis, G. Domalis, and D. Tsakalidis, “An Evaluation Framework for Synthetic Data Generation Models,” *IFIP Advances in Information and Communication Technology*, vol. 713 IFIPAI, pp. 320–335, 2024, doi: 10.1007/978-3-031-63219-8_24.
- [43] S. Dutta, M. N. U. Khan, and M. E. Hoque, “An Artificial Neural Network-based Model for Predicting Diesel Engine Performance and Emissions using Pyrolytic Plastic Oil Blend with Diesel Fuel,” *SSRN Electronic Journal*, no. December, pp. 1–6, 2024, doi: 10.2139/ssrn.4859475.
- [44] M. Zhang, N. Tsoulakos, P. Kujala, and S. Hirdaris, “A deep learning method for the prediction of ship fuel consumption in real operational conditions,” *Engineering Applications of Artificial Intelligence*, vol. 130, p. 107425, Apr. 2024, doi: 10.1016/j.engappai.2023.107425.
- [45] S. Watanabe, “Tree-Structured Parzen Estimator: Understanding Its Algorithm Components and Their Roles for Better Empirical Performance,” *arXiv preprint arXiv:2304.11127*, pp. 1–74, 2023, doi: 10.48550/arXiv.2304.11127.
- [46] K. P. K. K. Mohamad Zaim Awang Pon, “Hyperparameter Tuning of Deep learning Models in Keras,” *Sparklinglight Transtions on Artificial Intelligence and Quantum Computing(STAIQC)*, vol. 1, no. 2021, pp. 36–40, doi: 10.55011/staiqc.2021.1104.
- [47] S. Tao, P. Peng, Y. Li, H. Sun, Q. Li, and H. Wang, “Supervised contrastive representation learning with tree-structured parzen estimator Bayesian optimization for imbalanced tabular data,” *Expert Systems with Applications*, vol. 237, no. November 2022, 2024, doi: 10.1016/j.eswa.2023.121294.
- [48] S. Watanabe and F. Hutter, “c-TPE: Tree-Structured Parzen Estimator with Inequality Constraints for Expensive Hyperparameter Optimization,” *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2023-Augus, pp. 4371–4379, 2023, doi: 10.24963/ijcai.2023/486.
- [49] S. Sieradzki and J. Mańdziuk, “Modified Adaptive Tree-Structured Parzen Estimator for Hyperparameter Optimization,” *arXiv preprint arXiv:2502.00871*, pp. 1–21, Feb. 2025, doi: 10.48550/arXiv.2502.00871.
- [50] L. Qiao, X. Li, Q. Umer, and P. Guo, “Deep learning based software defect prediction,” *Neurocomputing*, vol. 385, pp. 100–110, 2020, doi: 10.1016/j.neucom.2019.11.067.
- [51] A. M. Abdullah, R. S. A. Usmani, T. R. Pillai, M. Marjani, and I. A. T. Hashem, “An Optimized Artificial Neural Network Model using Genetic Algorithm for Prediction of Traffic Emission Concentrations,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 797–807, 2021, doi: 10.14569/IJACSA.2021.0120693.
- [52] D. Bouabdallaoui, T. Haidi, M. Derri, I. Hbiak, and M. El Jaadi, “Multi-temporal forecasting of wind energy production using artificial intelligence models,” *International Journal of Renewable Energy Development*, vol. 14, no. 3, pp. 505–517, 2025, doi: 10.61435/ijred.2025.61086.
- [53] J. Castresana, G. Gabiña, L. Martin, A. Basterretxea, and Z. Uriondo, “Marine diesel engine ANN modelling with multiple output for complete engine performance map,” *Fuel*, vol. 319, no. March, 2022, doi: 10.1016/j.fuel.2022.123873.
- [54] I. Tyass, T. Khalili, M. Rafik, B. Abdelouahed, A. Raihani, and K. Mansouri, “Wind Speed Prediction Based on Statistical and Deep Learning Models,” *International Journal of Renewable Energy Development*, vol. 12, no. 2, pp. 288–299, 2023, doi: 10.14710/ijred.2023.48672.
- [55] C. Kaewbuddee et al., “Effects of Alcohol-Blended Waste Plastic Oil on Engine Performance Characteristics and Emissions of a Diesel Engine,” *Energies*, vol. 16, no. 3, 2023, doi: 10.3390/en16031281.
- [56] R. T. Marler and J. S. Arora, “The weighted sum method for multi-objective optimization: New insights,” *Structural and Multidisciplinary Optimization*, vol. 41, no. 6, pp. 853–862, 2010, doi: 10.1007/s00158-009-0460-7.
- [57] M. Zhang, N. Tsoulakos, P. Kujala, and S. Hirdaris, “A deep learning method for the prediction of ship fuel consumption in real operational conditions,” *Engineering*

- Applications of Artificial Intelligence*, vol. 130, no. October 2023, p. 107425, 2024, doi: 10.1016/j.engappai.2023.107425.
- [58] K. Sunil Kumar *et al.*, "Performance, Combustion, and Emission analysis of diesel engine fuelled with pyrolysis oil blends and n-propyl alcohol-RSM optimization and ML modelling," *Journal of Cleaner Production*, vol. 434, no. December 2023, p. 140354, 2024, doi: 10.1016/j.jclepro.2023.140354.
- [59] D. Beltrán and M. L. L. Dantas, "CACHAI's First Module: A Fully Customizable Chord Diagram for Astronomy and Beyond," *Research Notes of the AAS*, vol. 9, no. 8, p. 216, 2025, doi: 10.3847/2515-5172/adf8df.
- [60] R. Hao, Y. Wang, and W. Zhang, "Gene expression profiling of extraocular muscles in primary inferior oblique overaction," *PeerJ*, vol. 13, pp. 1–17, 2025, doi: 10.7717/peerj.19474.
- [61] A. A. Zimta, A. B. Tigu, C. Braicu, C. Stefan, C. Ionescu, and I. Berindan-Neagoe, "An Emerging Class of Long Non-coding RNA With Oncogenic Role Arises From the snoRNA Host Genes," *Frontiers in Oncology*, vol. 10, no. April, 2020, doi: 10.3389/fonc.2020.00389.
- [62] M. Goyal and Q. H. Mahmoud, "A Systematic Review of Synthetic Data Generation Techniques Using Generative AI," *Electronics*, vol. 13, no. 17, p. 3509, Sep. 2024, doi: 10.3390/electronics13173509.
- [63] V. K. Kaimal and P. Vijayabalan, "A study on synthesis of energy fuel from waste plastic and assessment of its potential as an alternative fuel for diesel engines," *Waste management*, vol. 51, pp. 91–96, 2016, doi: 10.1016/j.wasman.2016.03.003.
- [64] H. Yaqoob, H. M. Ali, U. Sajjad, and K. Hamid, "Investigating the potential of plastic pyrolysis oil-diesel blends in diesel engine: Performance, emissions, thermodynamics and sustainability analysis," *Results in Engineering*, vol. 24, no. November, 2024, doi: 10.1016/j.rineng.2024.103336.