https://doi.org/10.36375/prepare_u.iiche.a408

Machine Learning-Guided Optimization of Biofuel Blends for Enhanced Engine Efficiency and Emission Reduction

Prithwish Das[1], Aniket Biswas[2], Srijan Sardar[3], Bitopama Modak[4], Fahim Ahmed[5], Dr Sandip Kr Lahiri[*]

[1,2,3,4,5] *Jadavpur University, BE Chemical Engineering, Kolkata-700049, India*
** NIT Durgapur, BE Chemical Engineering, Durgapur-713209, India,*
Email address- ***sklahiri.che@nitdgp.ac.in***

Abstract

India heavily relies on imported foreign crude oil, prompting the need for effective solutions to reduce this dependency. One such solution is the blending of Biodiesel, Palm oil, and Ethanol with original diesel. The crucial concern lies in determining the optimal blending ratio that maximizes Engine Efficiency while maintaining reasonable levels of Oil Consumption and NOx emission. To address this, experimental data are collected from the paper [1] which systematically blends biodiesel. Experimental data involved three input parameters [Load, Palm Biodiesel, Ethanol] and three output parameters [Motor Brake Thermal Efficiency (BTE), Brake Specific Fuel Consumption (BSFC), and Nitrogen Oxides (NOx)], with 40 different runs. The prediction was accomplished using 26 Machine Learning Models, including Gaussian Process Regression, Support Vector regression, ANN, Tree and Linear Regression and others. Among the 26 models considered in the analysis, three models emerged as the top performers. The Stepwise Linear Regression Model [SLRM] yielded the highest Brake Thermal Efficiency (BTE), the Fine Tree Regression Model [FTRM] achieved the lowest Brake Specific Energy Consumption [BSFC], and the Matern 5/2 Gaussian Process Regression Model [MGPRM] demonstrated the lowest Nitrogen Oxide (NOx) emission. These models displayed a range of Root Mean Square Error (RMSE) and R-squared(validation) values: 0.02077–0.02333 and 0.99 for SLRM, 0.03789–0.03907 and 0.98 for FTRM & 0.02184–0.02296 and 0.99 for MGPRM. Moving forward, a multi-objective optimization approach has been undertaken to simultaneously maximize BTE while minimizing both BSFC and NOx emissions. To accomplish this, a Multi Objective Genetic Algorithm [MOGA] is employed to identify the Pareto Optimal Solution. The optimization process [MOGA] resulted in a series of 18 Pareto Optimal Solutions. These solutions provide insights on the appropriate blend ratios of Load, Palm Biodiesel and Ethanol in order to maximize Engine Thermal Efficiency while minimizing Fuel Consumption and NOx emissions.

Keywords: Palm biodiesel; Engine Thermal Efficiency; Machine Learning Model; Multi Objective Genetic Algorithm [MOGA]; Pareto Optimal solutions; Efficient Optimization.

https://doi.org/10.36375/prepare_u.iiche.a408

1. Introduction

The growing demand for fossil fuels and their application in diesel engines, especially in the industrial and automotive sectors, has increased significantly. However, this excessive use of conventional fuels has raised concerns over fuel depletion and the environmentally harmful emissions associated with diesel engines, motivating the exploration of alternative fuels over the past few decades [1,2]. The ongoing global industrial revolution has led to stricter regulations governing the use of diesel engines, largely driven by concerns about environmental pollution. As a response, alternative fuels have gained prominence as a means to mitigate harmful emissions and address global warming [3–5]. Over approximately a century, alternative fuels have been actively pursued as substitutes for traditional fossil diesel. Among these alternatives, biofuels have shown promise, with biodiesel emerging as a leading contender due to its similar properties to conventional diesel fuel [7–9]. Biodiesel is sulphur-free, less toxic, and more environmentally friendly, and it can be produced from a variety of readily available sources. Nevertheless, biodiesel presents some drawbacks, such as higher viscosity, greater density, and poor cold flow properties, which can hinder its effective use in diesel engines [10–12]. In such situations, alcohol-based additives are frequently utilized to modify the characteristics of biodiesel [13]. Notably, among these additives, ethanol, methanol, n-butanol, and diethyl ether are well-known and widely employed [14–18].

Researchers try to conduct experiments with diesel engines to find out at what proportions these additives can be blended so that engine performance, fuel consumptions and NOx emissions from diesel engines can be optimized. A diesel engine's performance is usually measured by three Metrics namely Brake Thermal Efficiency (BTE), Brake Specific Fuel Consumption (BSFC) and NOx (Nitrogen Oxides) emissions. Brake Thermal Efficiency (BTE) is a measure of how efficiently a diesel engine converts the heat energy from fuel into useful mechanical work. It is defined as the ratio of the brake power output to the energy input from the fuel. BTE is typically expressed as a percentage and can be calculated using the following formula: BTE (%) = (Brake Power Output / Fuel Energy Input) × 100.

On the other hand, Brake Specific Fuel Consumption (BSFC) is a measure of the fuel efficiency of a diesel engine. It quantifies the amount of fuel (typically expressed in mass or volume) required to produce one unit of brake power (usually one kilowatt or one horsepower) over a specific period of time. Lower BSFC values indicate better fuel efficiency. BSFC is defined using the following formula*:*

BSFC (g/kWh) = (Fuel Consumption in grams) / (Brake Power Output in kilowatts)

A lower BSFC value indicates that the engine is using less fuel to produce a given amount of power, which is a desirable characteristic for an efficient diesel engine. Engineers and manufacturers work to optimize engine design and operating conditions to minimize BSFC and improve fuel economy. NOx (Nitrogen Oxides) emissions from a diesel engine are a type of air pollution that consists of various nitrogen oxide compounds, primarily nitrogen dioxide (NO2) and nitric oxide (NO). These emissions are a result of the high-temperature combustion process in the diesel engine, which causes nitrogen and oxygen in the air to react and form NOx compounds.

NOx emissions are typically quantified and regulated in terms of concentration or mass per unit volume of exhaust gases. Common units for expressing NOx emissions include parts per million (ppm), grams per kilogram (g/kg), or grams per brake horsepower-hour (g/bhp·hr). The specific method and units used may vary depending on regional regulations and standards.

NOx emissions can be measured using various techniques, including exhaust gas analysers or emissions testing equipment. These measurements are usually taken at the exhaust

outlet of the engine or vehicle to determine the concentration of NOx in the exhaust gases. These values can be used to assess compliance with emissions standards and regulations. Reducing NOx emissions is a significant concern for diesel engine manufacturers and vehicle operators due to their adverse environmental and health effects.

The main problems researchers try to solve is to find the optimum blend ratio of biodiesel and various additives so that the above three conflicting objectives can be optimized.
Traditional methods of conducting engine experiments are now considered time-consuming and costly. As a result, new techniques utilizing computational analysis have been introduced to facilitate faster and more cost-effective experimentation [19–21]. Researchers, such as Krishnamoorthi et al. [22], have conducted experiments to optimize engine performance by varying injection pressure (IP), injection timing (IT), and compression ratio in a diesel engine, ultimately using Response Surface Methodology (RSM) to determine the optimal engine input parameters. They found a compression ratio of 18:1, a fuel injection pressure (FIP) of 250 bar, and an injection timing of 21 degrees before top dead center (bTDC) to be the best settings. Through the application of different design of experiments (DoE) and various RSM models, predictive models for engine responses, such as Brake Thermal Efficiency (BTE), Brake Specific Energy Consumption (BSFC), nitrogen oxides (NOx), hydrocarbons (HC), carbon monoxide (CO), and smoke opacity, have been developed, with coefficient of determination (R2) values for these parameters ranging between 0.9256 and 0.9991 [23–25].

Recent advances in computational analysis of diesel engine performance and emission parameters have introduced Artificial Neural Networks (ANN) as a promising alternative to RSM. ANN models are known for their ability to provide accurate predictions and strong correlations with experimental results. Shivakumar et al. [26] applied separate ANN models to predict engine responses within 8% mean error values (MRE). Taghavifar et al. [27] utilized ANN predictions with a feed-forward backpropagation (BP) learning algorithm and Levenberge-Marquardt transfer function for Computational Fluid Dynamics (CFD) simulated engine outputs, achieving high R2 values of 0.9951, 0.9976, and 0.9995 for NOx, CO2, and soot emissions, respectively. Rao et al. [28] investigated ANN prediction for nine different engine outputs with low error percentages (0.01–0.03) and high R2 values (0.980–0.999). Javed et al. [29] trained an ANN model for predicting engine performance and emissions, achieving an overall R value of 0.99360, Mean Squared Error (MSE) of 0.0011, and Mean Absolute Percentage Error (MAPE) of 4.863001%. Uslu et al. [30] accurately predicted engine responses using ANN models, observing R2 and mean relative error (MRE) values ranging from 0.964 to 0.9878 and 0.51–4.8%, respectively. Furthermore, several comparative studies have been conducted to assess the accuracy of prediction between RSM and ANN models [31,32]. In most cases, ANN models have outperformed RSM models due to their capacity to handle highly nonlinear behaviour during training [33–37]. Uslu et al. [38] developed a prediction model, obtaining an R2 value of 0.9 for RSM and 0.85–0.95 for ANN.

Despite the extensive literature focusing on prediction and optimization using RSM and ANN, and their combined RSM-ANN models, there remains a gap in the investigation of a comparative study between more advanced machine learning models in this domain. An effort has been made in the present study to apply various advanced machine learning models to predict engine performance parameters (BTE, BSFC and NOx) from engine input parameters such as Load %, Palm Biodiesel % and ethanol%. In the present study 2 advanced machine learning models have been applied and the best model is finally selected. After a reliable machine learning model is built,

the second objective is to optimize the blend ratio of Palm Biodiesel % and ethanol% so that a reasonable balance can be found in conflicting 3 objectives like BTE, BSFC and NOx.

The study's goals are bifurcated into two primary components. Firstly, the 26 advanced machine learning models are harnessed for predicting BTE, BSFC, and NOx emissions. Furthermore, diverse error types and correlation analyses have been executed to gauge and contrast the models' precision. On the other hand, leveraging the anticipated responses from the model, a multiobjective genetic algorithm system is effectively employed to determine the optimal engine input parameters and output responses.

2. Materials and Methodology

2.1. Experimental Setup:

In this study experimental data from reference [39] is taken. The reference [39] experimental investigation was carried out on a Kirloskar (TV1) type direct-injected diesel engine with a power rating of 3.5 kW. This engine is a single-cylinder, four-stroke, water-cooled unit with a variable compression ratio. The setup includes an eddy current dynamometer for measuring engine load, a crank angle sensor to monitor engine speed at every 1° crank angle rotation, K-type thermocouples for measuring various temperatures, and a piezoelectric transducer to record in-cylinder pressure. Additionally, a gas analyser was connected to the exhaust pipe to measure exhaust emissions, including NOx, CO, UHC (unburned hydrocarbons), CO2, and O2. Data acquisition and control were facilitated by a computerized system interfaced with specialized software.

2.2. Test Methodology:

The experimental procedure used in [39] for assessing engine performance and emissions involved testing the engine under varying loads, ranging from 20% to 100% in 20% increments, while maintaining a constant speed of 1500 rpm. Fuel consumption was measured using a fuel burette, and environmental conditions were documented. Different blends of diesel, palm biodiesel, and anhydrous ethanol were used in the experiments. 40 experimental runs were carried out at different engine loads, Palm Biodiesel % and ethanol% and 3 metrics of engine performance namely BTE, BSFC and NOx were measured in each run. The data are shown in table 1.

Table 1: Experimental data used for model building [39]

Test Run	Engine Load (%) A	Palm biodiesel (%) B	Ethanol (%) C	BTE	BSFC	NOx
1	40	10	10	0.34	0.51	0.40
2	60	10	5	0.55	0.33	0.60
3	100	15	10	0.81	0.12	0.82
4	40	15	5	0.34	0.55	0.40

https://doi.org/10.36375/prepare_u.iiche.a408

Test Run	Engine Load (%) A	Palm biodiesel (%) B	Ethanol (%) C	BTE	BSFC	NOx
5	60	10	10	0.50	0.34	0.64
6	80	15	10	0.63	0.24	0.74
7	80	15	5	0.69	0.22	0.74
8	40	20	10	0.30	0.55	0.37
9	20	5	5	0.18	0.81	0.12
10	60	15	10	0.48	0.36	0.60
11	100	10	5	0.87	0.12	0.82
12	100	20	10	0.76	0.15	0.78
13	20	10	5	0.17	0.81	0.11
14	20	15	10	0.17	0.77	0.13
15	20	5	10	0.20	0.72	0.11
16	100	20	5	0.90	0.10	0.83
17	80	20	10	0.58	0.27	0.73
18	40	5	5	0.39	0.49	0.37
19	100	5	10	0.86	0.10	0.90
20	60	20	10	0.45	0.38	0.59
21	20	20	5	0.17	0.81	0.10
22	80	5	10	0.69	0.20	0.83
23	80	20	5	0.71	0.21	0.73
24	60	5	5	0.56	0.33	0.56
25	60	20	5	0.53	0.34	0.58
26	80	10	10	0.64	0.23	0.80
27	40	20	5	0.40	0.48	0.38
28	100	15	5	0.87	0.12	0.83
29	100	10	10	0.85	0.11	0.86
30	40	10	5	0.39	0.49	0.39
31	20	20	10	0.10	0.90	0.11
32	40	15	10	0.36	0.49	0.35
33	40	5	10	0.36	0.49	0.45
34	80	10	5	0.73	0.20	0.76
35	80	5	5	0.75	0.19	0.72

https://doi.org/10.36375/prepare_u.iiche.a408

Test Run	Engine Load (%) A	Palm biodiesel (%) B	Ethanol (%) C	BTE	BSFC	NOx
36	20	15	5	0.16	0.84	0.12
37	100	5	5	0.90	0.11	0.84
38	60	15	5	0.51	0.36	0.60
39	20	10	10	0.15	0.81	0.15
40	60	5	10	0.53	0.32	0.68

https://doi.org/10.36375/prepare_u.iiche.a408

2.3. Development Of Machine Learning Model:

Experimental data shown in table 1 is used to develop various machine learning models. First 3 parameters in table-1 namely engine load, Palm Biodiesel % and ethanol%are taken as the input and next 3 engine performance parameters namely BTE, BSFC and NOx are taken as output of machine learning models. At a time one output was taken and 3 different models were built.

In the past decade, a multitude of machine learning models and algorithms have emerged in the literature. The goal is to assess how quickly and accurately these machine learning algorithms can comprehend the intricate nonlinear relationships between input and output data. Each algorithm has its own strengths and weaknesses. Given the absence of prior knowledge about which model suits the data best, this study explores 26 advanced machine learning models.

In this step, pre-processed data are inputted into diverse nonlinear advanced machine learning models such as decision trees, support vector regression, artificial neural regression, and Gaussian process regression. The objective is to evaluate and compare the predictive performance of each algorithm and their accuracy in predicting efficiency (BTE), Brake specific energy consumption (BSFC) and emissions (NOx). Initially, all algorithms were run with their default meta-parameter values in the MATLAB environment. To accurately assess the predictive capabilities of different machine learning models, the dataset was divided into training and validation sets using techniques like k-fold cross-validation. The model's performance was rigorously evaluated on the validation set to ensure its generalizability to unseen data.

Performance metrics, including Coefficient of Determination (R2), Root Mean Square Error (RMSE), and Average Error Percent (AEP), were employed to quantify the accuracy and predictive power of the models. R2 represents the proportion of variance in the dependent variables (output parameters) from the independent variables (input parameters), while RMSE signifies the average deviation between predicted values and actual observations. These metrics offer valuable insights into the model's ability to accurately capture underlying patterns in the data.

2.4. Modelling By Different Machine Learning Algorithms:

For the sake of brevity, a concise introduction to the different machine learning models [40] utilized in this study is provided below.

Linear Regression:

- ***Simple Linear Regression:*** This model involves a linear regression approach with the inclusion of an intercept term and linear predictors. It aims to establish a linear relationship between the predictor variables and the target variable.
- ***Interaction Linear Regression:*** In this linear regression variant, not only the intercept and linear predictors are considered, but also interaction terms among predictors are incorporated. This allows the model to capture more complex relationships between variables.
- ***Robust Linear Regression:*** This type of linear regression is designed to be resistant to the influence of outliers in the data. It incorporates an intercept and linear predictors, but it employs techniques that downplay the impact of extreme data points.
- ***Stepwise Linear Regression:*** This linear regression model employs a stepwise algorithm to determine which predictor variables should be included in the model. It automatically selects variables based on their contribution to the model's performance.

https://doi.org/10.36375/prepare_u.iiche.a408

Decision Trees:

- ***Fine Tree Regression:*** A regression tree with a fine structure, requiring a minimum leaf size of 4. This results in more detailed and potentially overfitted trees that can capture intricate patterns in the data.
- ***Medium Tree Regression:*** This regression tree strikes a balance by requiring a minimum leaf size of 12. It aims to capture meaningful patterns while avoiding excessive complexity that might lead to overfitting.
- ***Coarse Tree Regression:*** In this case, the regression tree is constrained by a minimum leaf size of 36. This promotes a simplified and general representation of the data, suitable for capturing broader trends.

Support Vector Regression:

- ***Linear SVM Regression:*** This support vector machine employs a linear kernel to capture a simple linear relationship in the data. It's relatively interpretable and is suitable for cases where the underlying relationship appears to be linear.
- ***Quadratic SVM Regression:*** Here, a support vector machine is used with a quadratic kernel, enabling the model to capture quadratic relationships between variables.
- ***Cubic SVM Regression:*** Similar to the quadratic variant, this model employs a cubic kernel to capture cubic relationships between variables.
- ***Fine Gaussian SVM Regression:*** This SVM is tailored to capture finely-detailed structures in the data using the Gaussian kernel. The kernel scale is adjusted based on the number of predictors.
- ***Medium Gaussian SVM Regression:*** This version of the SVM captures less intricate patterns in the data compared to the fine Gaussian variant. It still uses the Gaussian kernel but with a kernel scale determined by the number of predictors.
- ***Coarse Gaussian SVM Regression:*** This SVM identifies coarse structures in the data by using a Gaussian kernel with a larger kernel scale, adapted to the number of predictors.

Gaussian Process Regression:

- ***Rational Quadratic GPR:*** This Gaussian process regression model employs the rational quadratic kernel to capture complex relationships with varying scales and magnitudes.
- ***Squared Exponential GPR:*** Using the squared exponential kernel, this model excels at capturing smooth relationships in the data, making it suitable for cases where the underlying relationship is continuous and smooth.
- ***Matern 5/2 GPR:*** This Gaussian process regression model uses the Matern 5/2 kernel, which provides flexibility in capturing both smooth and abrupt changes in the data.
- ***Exponential GPR:*** The exponential kernel is utilized here to capture relationships with a focus on rapidly decaying correlations.

Kernel Approximation Regression:

- ***SVM Kernel Regression:*** In this approach, a Gaussian kernel is applied to perform regression on nonlinear data with numerous observations. The kernel maps predictors to a higher-dimensional space and fits a linear SVM model to the transformed predictors.
- ***Least Square Kernel Regression:*** Similar to the SVM kernel regression, this model uses a Gaussian kernel to regress nonlinear data. However, instead of an SVM, it employs an ordinary least squares linear regression model on the transformed predictors.

https://doi.org/10.36375/prepare_u.iiche.a408

Ensembles of Trees:

- ***Boosted Trees:*** This ensemble method involves combining multiple regression trees using the LS Boost algorithm. It is efficient in terms of time and memory usage, but may require a larger number of ensemble members for optimal performance.
- ***Bagged Trees:*** A bootstrap-aggregated ensemble of regression trees. While it often provides high accuracy, it can be resource-intensive in terms of computation and memory, particularly for large datasets.

Artificial Neural Network:

- ***Narrow ANN Regression:*** This regression neural network consists of a single fully connected layer with 10 neurons, excluding the final fully connected layer used for regression prediction.
- ***Medium ANN Regression:*** With a single fully connected layer containing 25 neurons, this neural network aims to capture moderately complex relationships between predictors and the target.
- ***Wide ANN Regression:*** In this neural network, a single fully connected layer with 100 neurons is used, suitable for capturing broader patterns and relationships in the data.
- ***Bilayered ANN Regression:*** This regression neural network incorporates two fully connected layers (excluding the final regression layer), allowing for more intricate feature extraction and representation.
- ***Trilayered ANN Regression:*** With three fully connected layers (excluding the final regression layer), this neural network aims to capture even more complex hierarchical patterns in the data.

In this study, experimental data of table 1 were applied to all 26 models mentioned above, and a comparison table summarizing their performance was created. Most accurate model (highest R2 and lowest RMSE and AEP) was chosen for further investigation and optimization.

2.5. Optimisation By Multi-Objective Genetic Algorithm:

Once a reliable and accurate model is shortlisted, the next objective of this study is to find the optimum value of 3 input parameters namely engine load, Palm Biodiesel % and ethanol% so that BTE is maximized and BSFC and NOx are minimized simultaneously. Since the 3 objectives are conflicting in nature, a multi-objective genetic algorithm is used here to strike a balance.

Multi-Objective Genetic Algorithms (MOGAs) [41] have emerged as powerful tools in the realm of optimization and decision-making. Unlike traditional single-objective optimization methods, MOGAs are designed to handle complex problems with multiple conflicting objectives, offering a versatile approach to finding a set of diverse and optimal solutions, known as the Pareto front [42]. Leveraging principles from natural evolution, MOGAs employ genetic operators such as selection, crossover, and mutation to iteratively evolve a population of candidate solutions. These algorithms emphasize the exploration of the solution space, balancing the trade-offs between competing objectives. Through generations of evolution, MOGAs guide the search towards the Pareto-optimal front, enabling decision-makers to make informed choices based on a range of optimal solutions rather than a single compromise. (MOGA)s are based on the principle of natural evolution to find optimal methods for problems with multiple conflicting goals. The process begins with initializing a population of solutions to a problem. Each solution is represented as a set, often called chromosomes. These individuals are evaluated using objective functions that

https://doi.org/10.36375/prepare_u.iiche.a408

measure how well they meet the various objectives of the problem. The main elements of MOGA include choice, competition, and change. During the selection period, individuals in the population are selected as parents based on their fitness, determined by the performance. Individuals with more positive goals are more likely to be selected as parents.

During crossover, two parents come together to produce offspring. By applying crossover operators to the parents' chromosomes, genetic information is exchanged to create new solutions. This mimics the process of genetic recombination in natural evolution. After the birth, some candidate solutions may develop mutations in which some of their chromosomes change. Mutations add genetic diversity to the population, preventing premature convergence to optimal solutions. When a new generation of offspring is created, a process called environmental selection is used. This step involves matching parents and offspring and selecting individuals for the next generation. Often, strategies such as elitism are used to ensure that the best available solutions are preserved for future generations. The algorithm repeats this step for several generations or until the fitness improvement is complete.

The goal of MOGA [43] is to strike a balance between exploration and implementation, exploring the overall solution space while using the contract space to converge to Pareto optimality, that is, a set of solutions where no other solution can improve an objective without harming the others.

The result of a multi-objective genetic algorithm is a set of variables that represent trade-offs between conflicting goals. By analysing the Pareto front, decision makers can gain a better understanding of the actual decision-making process by choosing the solutions that best suit their preferences and needs.

3. Results And Discussions:

3 different models are created for BTE, BSFC and NOx. For each model, 26 machine learning algorithms were used and out of them only the one that had the highest R-squared value was selected. Table 2, 3 and 4 summaries Model predictions performance of 26 different algorithms for BTE, BSFC and NOx respectively. The very high value of R2 and low value of RMSE signifies that all 3 models are very accurate.

Table 2: BTE Model predictions performance of different algorithms

Model Number	Model Type	RMSE (Validation)	MSE (Validation)	R-Squared (Validation)	MAE (Validation)
1	Simple Linear Regression	0.03	0.00	0.99	0.02
2	Interaction Linear Regression	0.02	0.00	0.99	0.02
3	Robust Linear Regression	0.03	0.00	0.99	0.02
4	Stepwise Linear Regression	0.02	0.00	0.99	0.02
5	Fine Tree Regression	0.05	0.00	0.97	0.04
6	Medium Tree Regression	0.15	0.02	0.64	0.13

https://doi.org/10.36375/prepare_u.iiche.a408

Model Number	Model Type	RMSE (Validation)	MSE (Validation)	R-Squared (Validation)	MAE (Validation)
7	Coarse Tree Regression	0.25	0.06	0.00	0.21
8	Linear SVM	0.03	0.00	0.99	0.02
9	Quadratic SVM	0.02	0.00	0.99	0.02
10	Cubic SVM	0.03	0.00	0.99	0.02
11	Fine Gaussian SVM	0.24	0.06	0.10	0.20
12	Medium Gaussian SVM	0.07	0.01	0.92	0.06
13	Coarse Gaussian SVM	0.05	0.00	0.96	0.04
14	Boosted Ensemble	0.09	0.01	0.87	0.08
15	Bagged Ensemble	0.09	0.01	0.88	0.07
16	Rational Quadratic GPR	0.02	0.00	0.99	0.02
17	Squared Exponential GPR	0.02	0.00	0.99	0.02
18	Matern 5/2 GPR	0.04	0.00	0.98	0.03
19	Exponential GPR	0.02	0.00	0.99	0.02
20	Narrow ANN	0.04	0.00	0.98	0.03
21	Medium ANN	0.03	0.00	0.99	0.02
22	Wide ANN	0.03	0.00	0.98	0.02
23	Bilayered ANN	0.05	0.00	0.96	0.04
24	Trilayered ANN	0.04	0.00	0.98	0.03
25	SVM Kernel	0.04	0.00	0.97	0.03
26	Least Square Kernel	0.07	0.01	0.92	0.06

Table 3: BSFC Model predictions performance of different algorithms

Model Number	Model Type	RMSE (Validation)	MSE (Validation)	R-Squared (Validation)	MAE (Validation)
1	Simple Linear Regression	0.07	0.00	0.92	0.06
2	Interaction Linear Regression	0.08	0.01	0.91	0.06
3	Robust Linear Regression	0.07	0.00	0.92	0.06
4	Stepwise Linear Regression	0.07	0.00	0.92	0.06
5	Fine Tree Regression	0.03	0.00	0.98	0.03

https://doi.org/10.36375/prepare_u.iiche.a408

Model Number	Model Type	RMSE (Validation)	MSE (Validation)	R-Squared (Validation)	MAE (Validation)
6	Medium Tree Regression	0.13	0.02	0.72	0.12
7	Coarse Tree Regression	0.25	0.06	0.00	0.21
8	Linear SVM	0.08	0.01	0.91	0.05
9	Quadratic SVM	0.04	0.00	0.97	0.03
10	Cubic SVM	0.04	0.00	0.98	0.03
11	Fine Gaussian SVM	0.24	0.06	0.07	0.19
12	Medium Gaussian SVM	0.10	0.01	0.85	0.07
13	Coarse Gaussian SVM	0.10	0.01	0.83	0.07
14	Boosted Ensemble	0.11	0.01	0.82	0.07
15	Bagged Ensemble	0.09	0.01	0.86	0.07
16	Rational Quadratic GPR	0.04	0.00	0.97	0.03
17	Squared Exponential GPR	0.04	0.00	0.98	0.03
18	Matern 5/2 GPR	0.06	0.00	0.95	0.04
19	Exponential GPR	0.04	0.00	0.97	0.03
20	Narrow ANN	0.07	0.01	0.92	0.05
21	Medium ANN	0.05	0.00	0.95	0.04
22	Wide ANN	0.05	0.00	0.97	0.03
23	Bilayered ANN	0.06	0.00	0.94	0.04
24	Trilayered ANN	0.05	0.00	0.95	0.04
25	SVM Kernel	0.06	0.00	0.94	0.04
26	Least Square Kernel	0.08	0.01	0.91	0.05

Table 4: BTE Model predictions performance of different algorithms

Model Number	Model Type	RMSE (Validation)	MSE (Validation)	R-Squared (Validation)	MAE (Validation)
1	Simple Linear Regression	0.06	0.00	0.94	0.06
2	Interaction Linear Regression	0.07	0.01	0.93	0.06
3	Robust Linear Regression	0.06	0.00	0.94	0.06
4	Stepwise Linear Regression	0.07	0.00	0.94	0.06

https://doi.org/10.36375/prepare_u.iiche.a408

Model Number	Model Type	RMSE (Validation)	MSE (Validation)	R-Squared (Validation)	MAE (Validation)
5	Fine Tree Regression	0.03	0.00	0.98	0.03
6	Medium Tree Regression	0.12	0.01	0.80	0.11
7	Coarse Tree Regression	0.27	0.07	0.00	0.24
8	Linear SVM	0.07	0.00	0.94	0.06
9	Quadratic SVM	0.03	0.00	0.98	0.03
10	Cubic SVM	0.03	0.00	0.98	0.03
11	Fine Gaussian SVM	0.25	0.06	0.11	0.22
12	Medium Gaussian SVM	0.07	0.00	0.94	0.05
13	Coarse Gaussian SVM	0.09	0.01	0.88	0.07
14	Boosted Ensemble	0.10	0.01	0.87	0.08
15	Bagged Ensemble	0.10	0.01	0.87	0.08
16	Rational Quadratic GPR	0.02	0.00	0.99	0.02
17	Squared Exponential GPR	0.02	0.00	0.99	0.02
18	Matern 5/2 GPR	0.04	0.00	0.98	0.03
19	Exponential GPR	0.02	0.00	0.99	0.02
20	Narrow ANN	0.03	0.00	0.98	0.02
21	Medium ANN	0.05	0.00	0.97	0.04
22	Wide ANN	0.03	0.00	0.99	0.02
23	Bilayered ANN	0.04	0.00	0.98	0.03
24	Trilayered ANN	0.05	0.00	0.96	0.04
25	SVM Kernel	0.05	0.00	0.96	0.04
26	Least Square Kernel	0.06	0.00	0.94	0.05

As discussed earlier, the model selection is based on 3 features – R-squared value, RMSE (Root Mean Square Error) and MAE (Mean Absolute Error).

All the 3 features for each model have been obtained from the results section of Regression Learner App in Matlab-2023. The model that has displayed the Highest R-squared value, Lowest RMSE value and Lowest MAE value has been selected in each of the cases.

In BTE model, Stepwise Linear Regression algorithm was selected as it gave an R-squared value of 0.99, RMSE value of 0.2234 and MAE value 0.01 which were comparatively the Highest, lowest, and lowest respectively.

https://doi.org/10.36375/prepare_u.iiche.a408

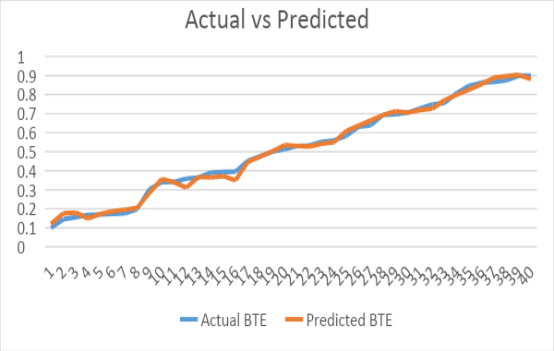

Fig 1. Actual vs model predicted outcome of BTE

In the BSFC model, Tree algorithm was selected as it gave R-squared value of 0.98, RMSE value of 0.034 and MAE value 0.025 which were comparatively the Highest, lowest, and lowest respectively.

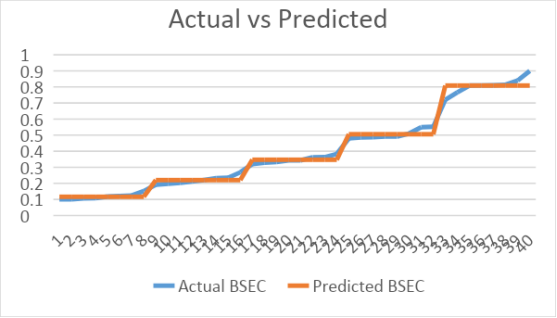

Fig 2. Actual vs model predicted outcome of BSFC

In the NOx model, Gaussian Process Regression-Matern 5/2 algorithm was selected as it gave R-squared value of 0.99, RMSE value of 0.00047 and MAE value 0.016 which were comparatively the Highest, lowest, and lowest respectively.

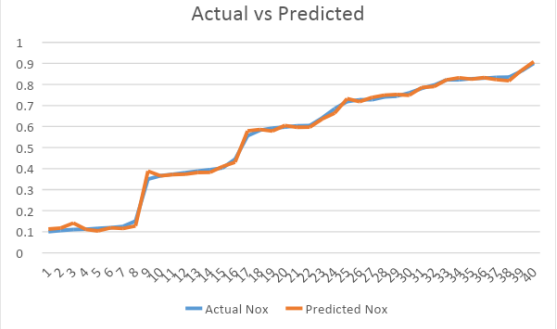

Fig 3. Actual vs model predicted outcome of NOx

Fig 1 ,2 and 3 shows the best model performance, namely experimental data with model predictions and they're almost overlapping for BTE, BSF and NOx respectively.

3.1. Optimization results:

Once a reliable and accurate model is shortlisted, the next objective of this study is to find the optimum value of 3 input parameters namely engine load, Palm Biodiesel % and ethanol% so that BTE is maximized and BSF and NOx are minimized simultaneously. Since the 3 objectives are conflicting in nature, a multi-objective genetic algorithm is used here to strike a balance and create a pareto optimal front.

The Pareto-optimal front, a fundamental concept in multi-objective optimization, represents a set of solutions where no individual solution can be improved in one objective without degrading at least one other objective. In other words, these solutions embody the best possible compromises between conflicting objectives, showcasing the trade-offs inherent in the problem at hand. The Pareto front is not a single solution but a collection of diverse, non-dominated solutions, each offering a unique balance between the competing criteria. Finding these solutions is crucial in decision-making processes where multiple, often conflicting, objectives need to be considered simultaneously. These objectives could be Fuel consumption minimization, Thermal efficiency maximization, or NOx reduction, and the Pareto front reveals the spectrum of solutions that optimally balance these goals. Determining the Pareto front involves utilizing optimization algorithms like Multi-Objective Genetic Algorithms (MOGAs), which iteratively evolve a population of potential solutions, ensuring that no solution dominates another in all objectives. The Pareto-optimal front is invaluable for decision-makers, providing a comprehensive understanding of the solution space and enabling them to make well-informed

https://doi.org/10.36375/prepare_u.iiche.a408

decisions based on the available trade-offs among conflicting objectives.

Fig 4 shows Pareto optimal solutions for BTE vs BSFC, Fig 5 shows Pareto optimal solutions for BTE vs NOx, Fig 6 shows Pareto optimal solutions for BSFC vs NOx

solutions after applying a multi-objective genetic algorithm.

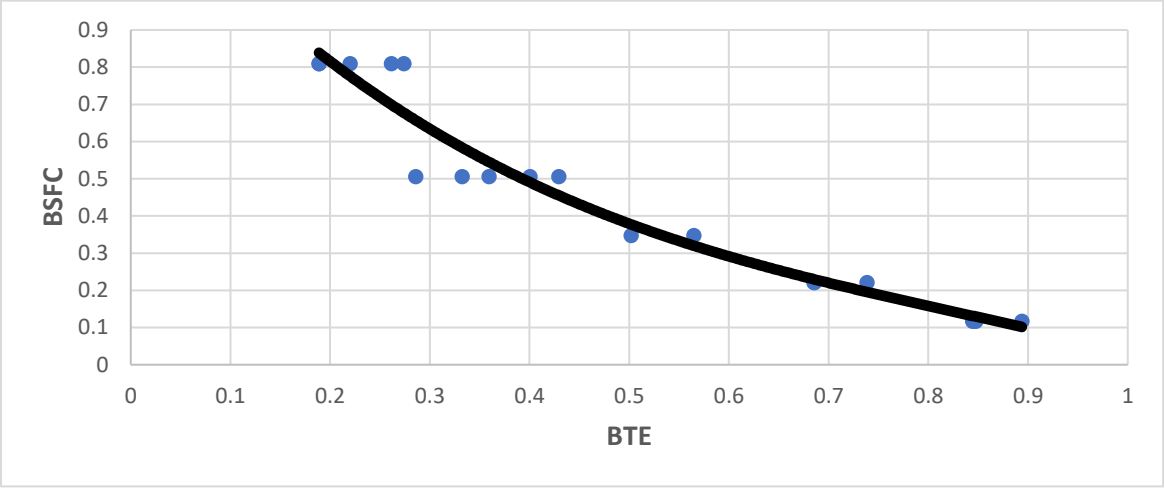

Fig 4: Pareto optimal solutions for BTE vs BSFC

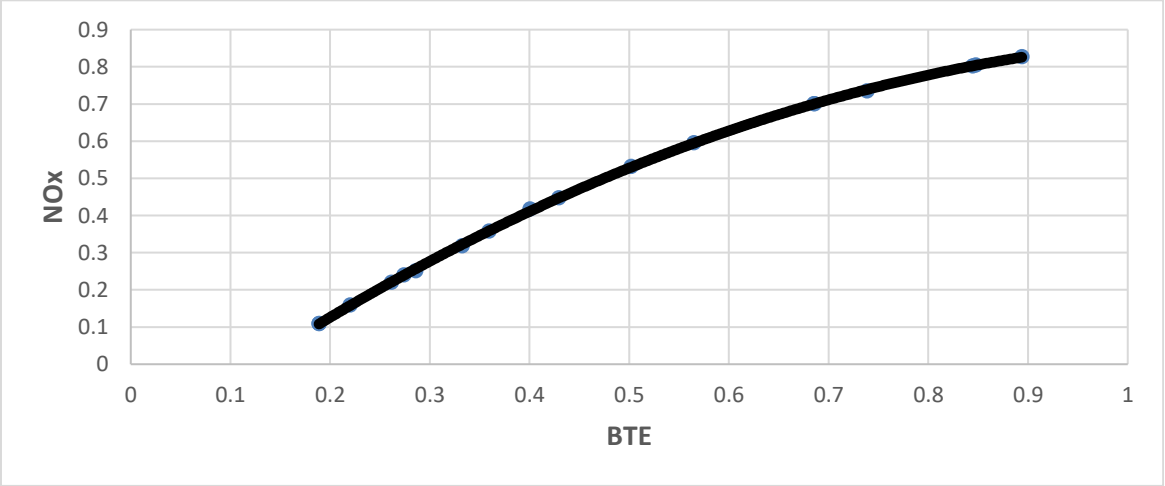

Fig 5: Pareto optimal solutions for BTE vs NOx

Table5 Summarizes the Pareto optimal

https://doi.org/10.36375/prepare_u.iiche.a408

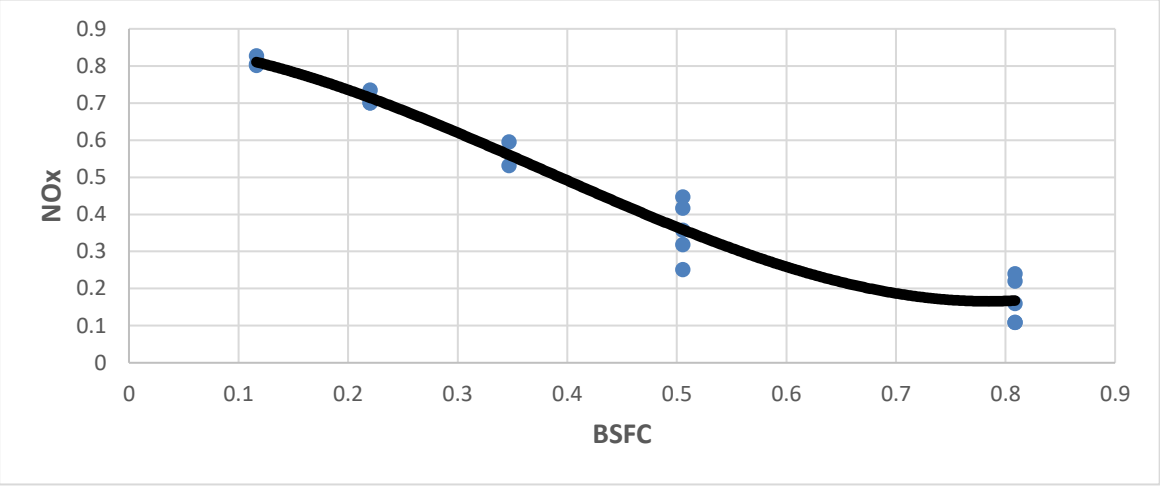

Fig 6: Pareto optimal solutions for BSFC vs NOx

Table 5: Pareto optimal solutions after applying multi-objective genetic algorithm

load% (A)	Palm-biodiesel (B)	Ethanol (C)	BTE	BSFC	NOx
20.00	5.00	5.00	0.19	0.81	0.11
61.89	5.08	5.29	0.57	0.35	0.60
75.34	5.78	5.08	0.69	0.22	0.70
20.00	5.00	5.00	0.19	0.81	0.11
93.46	6.09	5.05	0.85	0.12	0.81
28.11	5.55	5.07	0.26	0.81	0.22
98.75	6.83	5.03	0.89	0.12	0.83
43.71	6.16	5.21	0.40	0.51	0.42
35.98	5.23	5.09	0.33	0.51	0.32
23.47	5.31	5.20	0.22	0.81	0.16
30.79	5.09	5.01	0.29	0.51	0.25
81.15	5.43	5.02	0.74	0.22	0.74
29.55	5.88	5.07	0.27	0.81	0.24
54.97	6.26	5.03	0.50	0.35	0.53
38.98	5.21	5.27	0.36	0.51	0.36
46.85	6.11	5.03	0.43	0.51	0.45
92.97	5.11	5.15	0.85	0.12	0.80
75.39	5.52	5.12	0.69	0.22	0.70

4. Conclusion:

This study aims to determine the optimal proportions of palm biodiesel and ethanol in a biodiesel-ethanol (bio-diesohol) blend. It assesses how load, the shares of palm biodiesel and ethanol, impact Brake Thermal Efficiency (BTE), Brake Specific Energy Consumption (BSFC), and NOx emissions through a multi-objective genetic algorithm. Subsequently, comprehensive 26 machine learning models are established to precisely forecast engine performance. Lastly, a sophisticated hybrid modelling approach is employed to explore the multi-objective genetic algorithm optimization of engine input parameters and response variables.

5. References:

[1] Ghadikolaei MA, Cheung CS, Yung KF. Study of combustion, performance and emissions of diesel engine fueled with diesel/biodiesel/alcohol blends having the same oxygen concentration. Energy 2018;157:258–69. https://doi.org/10.1016/j.energy.2018.05.164.

[2] Agarwal AK, Dhar A. Experimental investigations of performance, emission and combustion characteristics of Karanja oil blends fuelled DICI engine. Renewable Energy2013;52:283–91. https://doi.org/10.1016/j.renene.2012.10.015.

[3] Agarwal AK, Gupta JG, Dhar A. Potential and challenges for large-scale application of biodiesel in automotive sector. Prog Energy Combust Sci 2017;61:113–49. https://doi.org/10.1016/j.pecs.2017.03.002.

[4] Dey S, Deb M, Das PK. Chemical Characterization and Tribological Behavior of Kitchen Chimney Dump Lard (KCDL) as a Bio-lubricant. Revue Des Composites et Des Materiaux Avances 2019;29. 10.18280/rcma.290303.

[5] Aydin H, Ilkiliç C. Effect of ethanol blending with biodiesel on engine performance and exhaust emissions in a CI engine. Appl Therm Eng 2010;30:1199–204. https://doi.org/10.1016/j.applthermaleng.2010.01.037.

7] Dey S, Reang NM, Das PK, Deb M. A comprehensive study on prospects of economy, environment, and efficiency of palm oil biodiesel as a renewable fuel. J Cleaner Prod 2020. https://doi.org/10.1016/j.jclepro.2020.124981.

[8] Guido C, Beatrice C, Napolitano P. Application of bioethanol / RME / diesel blend in a Euro5 automotive diesel engine: potentiality of closed loop combustion control technology. Appl Energy 2013;102:13–23. https://doi.org/10.1016/j.apenergy.2012.08.051

[9] Gulum M, Bilgin A. An experimental optimization research of methyl and ethyl esters production from safflower oil. Environ Climate Technol 2018;22:132–48. https://doi.org/10.2478/rtuect-2018-0009.

[10] Mohiddin MN, Saleh AA, Reddy ANR, Hamdan S. A study on chicken fat as an alternative feedstock: Biodiesel production, fuel characterisation, and diesel engine performance analysis. Int J Automotive and Mech Eng 2018;15:5535–46.10.15282/ijame.15.3.2018.10.0425.

[11] Ali OM, Mamat R, Abdullah NR, Abdullah AA. Analysis of blended fuel properties and engine performance with palm biodiesel-diesel blended fuel. Renewable Energy 2015;86:59–67. https://doi.org/10.1016/j.renene.2015.07.103.

[12] Zaharin MSM, Abdullah NR, Najafi G, Sharudin H, Yusaf T. Effects of physicochemical properties of biodiesel fuel blends with alcohol on diesel engine performance and exhaust emissions: a review. Renew Sustain Energy Rev 2017;79: 475–93.

https://doi.org/10.1016/j.rser.2017.05.035.

[13] Di Y, Cheung CS, Huang Z. Comparison of the effect of biodiesel-diesel and ethanol-diesel on the gaseous emission of a direct-injection diesel engine. Atmos Environ 2009;43:2721–30. https://doi.org/10.1016/j.atmosenv.2009.02.050.

[14] Ren Y, Huang ZH, Jiang DM, Li W, Liu B, Wang XB. Effects of the addition of ethanol and cetane number improver on the combustion and emission characteristics of a compression ignition engine. Proc Inst Mech Eng, Part D: J Automobile Eng 2008;222:1077–87. https://doi.org/10.1243/09544070JAUTO516.

[15] EL-Seesy AI, Kayatas Z, Takayama R, He Z, Kandasamy S, Kosaka H. Combustion and emission characteristics of RCEM and common rail diesel engine working with diesel fuel and ethanol/hydrous ethanol injected in the intake and exhaust port: Assessment and comparison. Energy Conversion and Management 2020;205: 112453. 10.1016/j.enconman.2019.112453.

[16] Dey S, Reang NM, Deb M, Das PK. Study on performance-emission trade-off and multi-objective optimization of diesel-ethanol-palm biodiesel in a single cylinder CI engine: a Taguchi-fuzzy approach ARTICLE HISTORY. Energy Sources Part A 2020;1–21. https://doi.org/10.1080/15567036.2020.1767234.

[17] Yesilyurt MK, Aydin M. Experimental investigation on the performance, combustion and exhaust emission characteristics of a compression-ignition engine fueled with cottonseed oil biodiesel/diethyl ether/diesel fuel blends. Energy Convers Manage 2020;205:112355. https://doi.org/10.1016/j.enconman.2019.112355.

[18] Qi DH, Chen H, Geng LM, Bian YZ. Effect of diethyl ether and ethanol additives on the combustion and emission characteristics of biodiesel-diesel blended fuel engine. Renewable Energy 2011;36:1252–8. https://doi.org/10.1016/j.renene.2010.09.021

[19] Tosun E, Aydin K, Bilgili M. Comparison of linear regression and artificial neural network model of a diesel engine fueled with biodiesel-alcohol mixtures. Alexandria Engineering Journal 2016;55:3081–9.

https://doi.org/10.1016/j.aej.2016.08.011

[20] Atmanli A, Yüksel B, Ileri E, Deniz Karaoglan A. Response surface methodology based optimization of diesel-n-butanol -cotton oil ternary blend ratios to improve engine performance and exhaust emission characteristics. Energy Convers Manage

2015;90:383–94. https://doi.org/10.1016/j.enconman.2014.11.029

[21] Sakthivel R, Ramesh K, Joseph John Marshal S, Sadasivuni KK. Prediction of performance and emission characteristics of diesel engine fuelled with waste biomass pyrolysis oil using response surface methodology. Renewable Energy 2019;136:91–103.10.1016/J.RENENE.2018.12.109.

[22] Krishnamoorthi M, Malayalamurthi R, Mohamed Shameer P. RSM based optimization of performance and emission characteristics of DI compression ignition engine fuelled with diesel/aegle marmelos oil/diethyl ether blends at varying compression ratio, injection pressure and injection timing. Fuel 2018;221: 283–97.

https://doi.org/10.1016/j.fuel.2018.02.070

[23] Fang W, Kittelson DB, Northrop WF. Optimization of reactivity-controlled compression ignition combustion fueled with diesel and hydrous ethanol using response surface methodology. Fuel 2015;160:446–57. https://doi.org/10.1016/j.fuel.2015.07.055

[24] Hirkude JB, Padalkar AS. Performance optimization of CI engine fuelled with waste fried oil methyl ester-diesel blend using response surface methodology. Fuel 2014; 119:266–73. https://doi.org/10.1016/j.fuel.2013.11.039.

[25] Ileri E, Karaoglan AD, Atmanli A. Response surface methodology based prediction of engine performance and exhaust emissions of a diesel engine fuelled with canola oil methyl ester. J Renewable Sustainable Energy 2013;5:033132. https://doi.org/10.1063/1.4811801.

[26] Shivakumar, Srinivasa Pai P, Shrinivasa Rao BR. Artificial Neural Network based prediction of performance and emission characteristics of a variable compression ratio CI engine using WCO as a biodiesel at different injection timings. Applied Energy 2011;88:2344–54.10.1016/j.apenergy.2010.12.030.

[27] Taghavifar HH, Taghavifar HH, Mardani A, Mohebbi A, Khalilarya S, Jafarmadar S. Appraisal of artificial neural networks to the emission analysis and prediction of CO2, soot, and NOx of n-heptane fueled engine. J Cleaner Prod 2016;112:1729–39. https://doi.org/10.1016/j.jclepro.2015.03.035

[28] Prasada Rao K, Victor Babu T, Anuradha G, Appa Rao BV. IDI diesel engine performance and exhaust emission analysis using biodiesel with an artificial neural network (ANN). Egypt J Pet 2017;26:593–600. https://doi.org/10.1016/j.ejpe.2016.08.006

[29] Javed S, Satyanarayana Murthy YV V, Baig RU, Prasada Rao D. Development of ANN model for prediction of performance and emission characteristics of hydrogen dual fueled diesel engine with Jatropha Methyl Ester biodiesel blends. J Natural Gas Sci Eng 2015;26:549–57. doi:10.1016/j.jngse.2015.06.041.

[30] Uslu S, Celik MB. Prediction of engine emissions and performance with artificial neural networks in a single cylinder diesel engine using diethyl ether. Eng Sci Technol Int J 2018;21:1194–201. https://doi.org/10.1016/j.jestch.2018.08.017

[31] Aryasomayajula Venkata Satya Lakshmi SB, Subramania Pillai N, Khadhar Mohamed MSB, Narayanan A. Biodiesel production from rubber seed oil using calcined eggshells impregnated with Al2O3 as heterogeneous catalyst: a comparative study of RSM and ANN optimization. Brazilian J Chem Eng 2020;37: 351–68. https://doi.org/10.1007/s43153-020-00027-9

[32] Hariharan N, Senthil V, Krishnamoorthi M, Karthic SV. Application of artificial neural network and response surface methodology for predicting and optimizing dual-fuel CI engine characteristics using hydrogen and bio fuel with water injection. Fuel 2020;270:117576. https://doi.org/10.1016/j.fuel.2020.117576

[33] Hammoudi A, Moussaceb K, Belebchouche C, Dahmoune F. Comparison of artificial neural network (ANN) and response surface methodology (RSM) prediction in compressive strength of recycled concrete aggregates. Constr Build Mater 2019;209:425–36. https://doi.org/10.1016/j.conbuildmat.2019.03.119

[34] Samuel OD, Okwu MO. Comparison of Response Surface Methodology (RSM) and Artificial Neural Network (ANN) in modelling of waste coconut oil ethyl esters production. Energy Sources Part A 2019;41:1049–61. https://doi.org/10.1080/15567036.2018.1539138.

[35] Astray G, Gullon ´ B, Labidi J, Gullon ´ P. Comparison between developed models using response surface methodology (RSM) and artificial neural networks (ANNs) with the purpose to optimize oligosaccharide mixtures production from sugar beet pulp. Ind Crops Prod 2016;92:290–9. https://doi.org/10.1016/j.indcrop.2016.08.011

[36] Ayoola AA, Hymore FK, Omonhinmin CA, Babalola PO, Fayomi OSI, Olawole OC, et al. Response surface methodology and artificial neural network analysis of crude palm kernel oil biodiesel production. Chem Data Collect 2020;28:100478. https://doi.org/10.1016/j.cdc.2020.100478

[37] Chouaibi M, Daoued K Ben, Riguane K, Rouissi T, Ferrari G. Production of bioethanol from pumpkin peel wastes: comparison between response surface methodology (RSM) and artificial neural networks (ANN). Industrial Crops and Products 2020;155:112822.10.1016/j.indcrop.2020.112822.

[38] Uslu S. Optimization of diesel engine operating parameters fueled with palm oil-diesel blend: Comparative evaluation between response surface methodology (RSM) and artificial neural network (ANN). Fuel 2020;276:117990. https://doi.org/10.1016/j.fuel.2020.117990

[39]. Dey, Suman, Narath Moni Reang, Pankaj Kumar Das, and Madhujit Deb. "Comparative study using RSM and ANN modelling for performance-emission prediction of CI engine fuelled with bio-diesohol blends: A fuzzy optimization approach." Fuel 292 (2021): 120356. https://doi.org/10.1016/j.fuel.2021.120356

[40]. Zheng, M., Wang, F., Hu, X., Miao, Y., Cao, H., & Tang, M. (2022). A Method for Analyzing the Performance Impact of Imbalanced Binary Data on Machine Learning Models. *Axioms*, *11*(11). https://doi.org/10.3390/axioms11110607

[41]. Wang, Z., Zhang, X., Zhang, Z. K., & Sheng, D. (2022). Credit portfolio optimization: A multi-objective genetic algorithm approach. Borsa Istanbul Review, 22(1). https://doi.org/10.1016/j.bir.2021.01.004

[42]. Giordano, P. C., Pereyra, V., Beccaria, A. J., Vero, S., & Goicoechea, H. C. (2020). Application of pareto-optimal front as an option to desirability function for the optimization of a microbiological process. *Microchemical Journal*, *155*. https://doi.org/10.1016/j.microc.2020.104682

[43]. Yu, W., Li, B., Jia, H., Zhang, M., & Wang, D. (2015). Application of multi-objective genetic algorithm to optimize energy efficiency and thermal comfort in building design. *Energy and Buildings*, *88*. https://doi.org/10.1016/j.enbuild.2014.11.063