INTERNATIONAL JOURNAL OF AUTOMOTIVE AND MECHANICAL ENGINEERING
ISSN: 2229-8649 e-ISSN: 2180-1606
VOLUME 23, ISSUE 1, 2026, 13339 - 13355
DOI: https://doi.org/10.15282/ijame.23.1.2026.13.1011

RESEARCH ARTICLE

A comparative analysis of regression modelling and artificial neural networks for diesel engine performance prediction

Muhammad Asad[1]*, Umer Iftikhar[1], Daniyal Abbasi[2], Muhammad Bilal Jamil[2]

[1]Mechanical Engineering Department, University of Engineering and Technology, G.T. Road, Lahore 54890, Pakistan
[2]Mining Engineering Department, University of Engineering and Technology, G.T. Road, Lahore 54890, Pakistan

Abstract – Diesel engines are inherently difficult to analyze due to the complex and nonlinear interactions among engine speed, throttle position, and load. This study investigates the key factors affecting the performance of a four-stroke diesel engine. It compares the predictive capabilities of multiple linear regression and Artificial Neural Networks (ANN) in modeling five critical performance indicators: power (kW), torque (Nm), Brake Thermal Efficiency (BTE, %), Brake Specific Fuel Consumption (BSFC, kg/kWh), and Air-to-Fuel (A/F) ratio. Experimental data were collected from a TD23 diesel engine tested on an engine dynamometer across a range of operating conditions, with engine speeds from 1100 to 1600 RPM, throttle positions from 10% to 40%, and varying loads, yielding a dataset of 24 observations. Regression models were developed using Minitab, while ANN models were implemented in MATLAB. Results show that engine speed and load exert the strongest influence on performance, whereas throttle position has a relatively minor effect. Although regression achieves slightly lower Root Mean Square Error (RMSE) for power (0.1406 kW vs. 0.3561 kW) and torque (1.0937 Nm vs. 1.4698 Nm), likely due to the small dataset favoring simpler linear fits, the ANN consistently demonstrates superior coefficient of determination (R^2) values for nonlinear responses. It improves R^2 by 44.38% for BSFC (0.9143 vs. 0.4705), 35.26% for BTE (0.8186 vs. 0.4660), 19.31% for torque (0.8136 vs. 0.6205), and 2.63% for A/F ratio (0.8589 vs. 0.8326). Notably, BSFC exhibits extremely small RMSE values due to its unit scale (kg/kWh) and low data variability, underscoring the importance of clear unit reporting. Overall, the ANN proves more effective at capturing the complex, nonlinear behaviour of diesel engine performance, particularly when sufficient data diversity is present, while regression remains competitive for near-linear outputs in data-limited scenarios.

Article History
Received : 27 June 2025
Revised : 28 August 2025
Accepted : 30 January 2026
Published : 12 March 2026

Keywords
Artificial neural network
4-Stroke diesel engine
RPM variation
Throttle position
Comparative analysis

1. Introduction

Continuous research and innovation aim to enhance the performance and emission characteristics of internal combustion (IC) engines [1]. With the depletion of fossil fuel resources, there is a pressing need to optimize the performance of internal combustion engines with conventional fuels. However, the optimal performance characteristics are still unknown for some engines. These determinations, made through experimentation, require significant time and money. Though theoretical analysis offers an alternative, it involves lengthy calculations and is prone to inaccuracies in nonlinear systems. Most combustion systems are therefore nonlinear, and analysis is exact only under special conditions [2]. If there are no assumptions in the model, the overall set of nonlinear equations will become too difficult to solve or require substantial computational effort. Mathematical modeling has enhanced the understanding of combustion processes, but its practical use for diagnosis and optimization remains limited. As a result, artificial intelligence techniques such as Artificial Neural Networks (ANN) [3], fuzzy logic [4], support vector machines [5], and genetic algorithms [6] are increasingly used in real applications to bridge these gaps. It is observed that machine learning techniques can effectively model the performance characteristics of an internal combustion engine, including nonlinear behaviors and patterns. These methods are useful for predicting a range of parameters, including brake power, nitrogen oxides (NOx) emissions, brake torque, unburnt hydrocarbons, brake mean effective pressure (BMEP) [7], and brake-specific fuel consumption (BSFC) [8].

The performance of internal combustion engines can be predicted using several algorithms, including ANN, Response Surface Methodology, Relevance Vector Machine, and Genetic Programming [9-10]. ANNs are widely employed models for predicting the performance and emission characteristics of internal combustion engines [11]. One advantage of the ANN algorithm is that it can be modeled, trained, and tested without prior knowledge of the subject. Instead of lengthy, iterative calculations, an ANN selects the most suitable neural topology from the predefined options. However, an ANN cannot perform tasks without proper training. Typically, 80% of the data is used to train the model, while the remaining 20% is used for testing. The training and test data are obtained through experiments, using equipment such as a dynamometer to measure engine brake power, an airflow meter to measure air consumption, a laminar flow meter to measure fuel consumption, and an exhaust gas analyzer to measure the engine's emission characteristics during fuel testing [12-13]. The ANN model typically consists of an input layer, an output layer, and multiple hidden layers. The number of hidden layers varies depending on the complexity and relationships within the data. Each layer is composed of several interconnected nodes, both within the same layer and across different layers [13]. Let us assume the fuel blend, engine load, and engine speed as input nodes, and the output nodes as the characteristics of the output performance (brake power, torque, etc.). Thus, hidden layers consist of nodes that link the given input to the desired output, i.e., they form the topology that connects the inputs to the outputs. These fast algorithms fall into two categories: the first uses heuristic

*CORRESPONDING AUTHOR | M. Asad | ✉ muhammadasad202125@gmail.com

techniques and a variable learning rate, and resilient backpropagation is in this category. This category uses numerical optimization techniques such as Levenberg–Marquardt (trainlm) and scaled conjugate gradient (trainscg) [14].

The input parameters can vary, affecting the output parameters, as each input, ranging from idle to acceleration, has a distinct significance. The influence of one variable or node on another is determined by the weights or masses assigned to it. For instance, during idling, fuel consumption tends to be higher at lower engine speeds, leading to lower efficiency and economic losses. In contrast, at a steady speed, both engine speed and fuel consumption differ from those during idling, and as a result, the ANN model's output will reflect these variations [15]. To ensure equal contributions from each input, the inputs were pre-processed and scaled to the range -1 to 1. As mentioned above, the model's accuracy increases with the number of hidden layers. To assess the model's accuracy, a regression analysis was conducted between the network's response and the desired target, and the results were very close. An artificial neural network learns via forward or backward propagation, depending on the direction of learning [16]. The trainlm algorithm is considered a faster method for training moderately sized feedforward neural networks. The backpropagation algorithm in an ANN proved effective in determining torque, power, and emissions, as these parameters exhibited high correlation coefficients. Additionally, integrating fuzzy logic with neural networks helps manage data noise [17]. Among the two neuro-fuzzy logic techniques, namely the Adaptive Neuro-Fuzzy Inference System (ANFIS) and the Dynamic Evolving Neuro-Fuzzy Inference System (DENFIS), the ANFIS model demonstrated greater accuracy than the DENFIS model in predicting the emission characteristics of a fuel blend in an engine [18].

Kiani et al. [19] used an ANN to predict power, torque, and emissions in a spark-ignition engine fuelled with ethanol-gasoline blends. They got high R^2 values, like 0.99 for torque, indicating that the ANN fits the experimental data well. Nevertheless, their work focused on gasoline engines, not diesel engines, and they did not compare their results with Minitab's regression analysis. They missed how throttle positions affect the diesel-air fuel mixture nonlinearly and why regression might fail in that region. This study fills this gap by focusing on diesel with throttle at 10%-40% and RPM at 1100-1600, showing that the ANN achieves a 35.26% higher R^2 for efficiency than regression analysis. However, this study focuses on diesel with throttle from 10% to 40% and RPM 1100 to 1600. They missed regression analysis and Pareto charts. Roy et al. [20] investigated the prediction of performance and emission characteristics of a single-cylinder compression-ignition CI engine assisted by common-rail direct injection and coupled with exhaust gas recirculation using an ANN algorithm. The study focused on predicting BSFC, carbon monoxide (CO), particulate matter (PM), and nitrogen oxide (NOx) emissions. His team predicted emissions using an ANN, but this study also includes comparisons of efficiency and BSFC. This fills a gap in the literature by showing that ANN handles complex data better than linear regression in Minitab. They also applied Minitab's regression analysis to predict BSFC, CO, PM, and NOx in a single-cylinder diesel engine with exhaust gas recirculation. They found the ANN to be accurate, with low error, but, again, there was no direct comparison with regression analysis or other statistical methods. They said ANN captures interactions, but did not demonstrate this with linear models for parameters such as power and A/F ratio. This gap means we do not know whether an ANN is needed or whether simple regression suffices for diesel performance under varying loads.

Arumugam et al. [21] used artificial intelligence to predict the performance and emission characteristics of a compression ignition engine running on rapeseed methyl ester-diesel blends under various operating loads and conditions. The study showed that the root-mean-square error was very low, indicating minimal deviation from the actual results. This confirmed the close alignment between the predicted and actual data. Today, diesel engines are widely used in cars, trucks, and machinery. They help in moving goods and people. However, fuel is getting scarcer, and pollution is a big problem. This study examines how to improve diesel engine performance by adjusting speed and throttle. The use of regression modeling and ANN to predict quantities such as power, torque, and fuel consumption is common nowadays. This is important for the market because car makers want engines that save fuel and produce less smoke. For example, in Europe and the USA, there are strict rules on emissions, such as Euro 6 and Environmental Protection Agency standards. Our work can help meet these requirements by predicting engine behavior with fewer tests. This saves companies time and money. Also, with ANN, we get better predictions for nonlinear phenomena that older methods like regression analysis may miss. This means better engine design for electric hybrid systems, too, where diesel is still used. This study contributes to existing works by comparing conventional regression analysis and ANN on real data from a 4-stroke diesel engine. Many studies use ANN for emissions, but few compare with statistical tools like Minitab for RPM and throttle changes. For example, Hoang et al. [22] reviewed ANNs for biodiesel engines but did not compare them with regression analysis. This paper shows that ANNs achieve higher R^2 values for most parameters. This helps engineers choose the right tools.

This study also aims to improve the analysis of diesel engine performance. Diesel engines are quite challenging to study because many factors, such as speed, throttle position, and load, simultaneously influence performance. The problem is that tools such as regression analysis may not accurately capture complex relationships, leading to inaccurate results. This paper aims to identify the key factors influencing engine performance and to compare the accuracy of regression analysis and an ANN in predicting metrics such as efficiency, fuel consumption, power, torque, and air-to-fuel (A/F) ratio. It is hypothesized that an ANN will achieve higher accuracy than regression analysis because it can handle nonlinear data. The new part of this research is using an ANN to evaluate the engine's performance and compare it with older methods, such as regression analysis. ANN understands the engine's behavior much better. This new study introduces a more accurate and dependable method for testing and analyzing an engine. Many studies use ANNs for engine predictions but skip side-by-side comparisons with traditional regression. For example, Hoang et al. [22] reviewed the use of ANNs for diesel engines fuelled with biodiesel. They said ANN predicts performance and emissions well, such as BSFC and NOx, because it handles nonlinear data from experiments. However, they did not compare ANN to any other regression analysis or any linear regression tool. They focused solely on the advantages of ANNs over mathematical models, without showing how statistical software falls short in real diesel data. This leaves a gap because engineers need to know whether an ANN is truly better than simpler tools, such as regression analysis, for complex factors like speed and throttle. This paper fixes that by showing that ANNs achieve higher R^2 values for most tasks. For efficiency, ANN beats regression analysis by 35.26% in R^2. For BSFC, it wins by 44.38%. Torque gets 19.31% better, and the A/F ratio is 2.63% higher. Power has an R^2 of 0.89%, which is higher than the ANN's, but we choose the ANN because it captures nonlinear patterns in the engine better.

Other works, such as Kurtgoz et al. [23], estimated biogas engine performance using an ANN and reported better generalization than regression, but their data were small and lacked a diesel focus. Zhang et al studied diesel ethanol blends and found speed affects power, but used no artificial intelligence or stats comparison. Akkoli et al. [24] tested injection parameters with producer gas in a diesel engine and observed a drop in brake thermal efficiency (BTE) at high speeds, but relied solely on experimental data. Zare [25] shows that a simple neural network often beats linear regression on tricky datasets. It can detect curved and complex patterns in the data, whereas linear regression can only find straight-line relationships. This makes the neural network much stronger when the data has unusual points or sudden shifts. Because it is so flexible and can adapt to many different situations, it usually makes more accurate predictions, especially when the link between cause and effect is complicated and messy. Sai et al. [26] and their team recently demonstrated that, for big data tasks, complex ensemble methods such as Gradient Boosting can deliver superior accuracy, but this comes at a steep computational cost. In contrast, straightforward models like Linear Regression yield results much more quickly and are easier to interpret. Their key conclusion is that the best algorithm depends entirely on the project's specific priorities, whether those are raw predictive power, operational speed, or effective scalability. Grebovic et al. [27] and other researchers showed that ANNs are far better than older statistical methods, such as linear regression, for handling complex data with complex patterns. In their tests, they found that ANN could be up to 20% more accurate for tasks such as sorting items into categories or predicting numerical values. The authors noted that these networks require more computing power, yet they still strongly recommended using them for important tasks where achieving the most accurate prediction is the main goal. They see ANN as a powerful replacement for traditional models.

Li et al. [28] developed a tensor neural network that decomposes complex data relationships into simpler, more maintainable components. The new approach outperformed traditional neural networks on high-dimensional data, achieving substantially lower prediction errors across many test cases. The model also enabled efficient computation of derivatives and integrals, offering many useful insights for sensitivity analysis. This work shows a promising direction in handling complex regression tasks while maintaining computational feasibility in high-dimensional spaces. In a study, Koscik et al. [29] tested two forecasting methods for office lighting: a traditional polynomial regression and an Artificial Neural Network. The results were clear that the neural network was the winner by a significant margin. It cut the regression model's prediction error in half, showing a much stronger ability to handle the complex, real-world relationship between outdoor brightness and power use. It makes neural networks a much tougher and more dependable choice for the design of smart systems. Rahman et al. [30] showed that ANNs are a powerful and efficient tool for predictive problems. In their tests on a standard car mileage dataset, the neural network consistently made more accurate predictions than other complex optimization algorithms. A major benefit was its speed; the network delivered these superior results much faster and with less computational power, proving it to be a highly effective and practical choice for real-world forecasting tasks. Oğuz et al. [31] aimed to create an ANN model to forecast diesel engine performance parameters with biofuels. Their goal was to address the expense and time demands of conventional experiments. They made fuel mixtures using diesel, biodiesel, B20, and bioethanol (5, 10, 15, and 20). Engine tests were conducted to collect reference data on power, torque, hourly fuel consumption, and specific fuel consumption. A feedforward ANN was then trained on 74 out of 96 data points. The results showed that ANN predictions closely matched experimental values, with a reliability of 99.94%. This indicates the model efficiently estimates performance. Wang et al. [32] developed an integrated model combining an improved seagull optimization algorithm with a backpropagation neural network and extreme gradient boosting (XGBoost) to predict diesel particulate filter regeneration states and engine performance metrics such as nitrogen oxides (NOx), oxygen (O_2), smoke, fuel consumption, and exhaust temperature during regeneration. They conducted engine bench tests to collect data, optimized the neural network, and fused the models to enhance predictive accuracy. The results demonstrated high precision, with R^2 values ranging from 0.97 to 0.99 and mean absolute percentage error reductions of up to 14.93%, closely matching the experimental data and supporting improved DPF regeneration. Similarly, Dave et al. [33] investigated diesel engine combustion, performance, and emissions using neat diesel blended with di-tert-butyl peroxide as a cetane improver. Unlike Wang et al. [32], who focus on algorithmic integration for regeneration-state prediction, Dave's study emphasized the effects of fuel formulation on engine performance. Key metrics, peak pressure, fuel consumption, NOx, and smoke were modeled using ANNs and MLELMs based on data from experiments at fixed load and varying speeds. Their results highlighted that higher speeds increased fuel consumption and smoke by 14–29%, while blended fuels reduced fuel consumption by 5% and NOx emissions by 16%. The predictive models achieved fit coefficients (R^2) of 0.95–1.00 with minimal error, demonstrating strengths in modelling blended-fuel impacts rather than regeneration processes. Unlike Wang et al. [32], this study emphasized the effects of a cetane improver and its prediction using different machine learning frameworks.

Odufuwa et al. [34] developed an ANN model to predict the efficiency and emissions of a mini-diesel engine, focusing on outputs including carbon dioxide (CO_2), NOx, CO, and PM. The model was trained using dynamometer test data in MATLAB. Results demonstrated strong predictive capability, with correlation coefficients above 0.9, low prediction errors, a 12% improvement in engine efficiency, and a 40% reduction in emissions. Pallicheruvu and Gnanasekaran [35] tested fish oil biodiesel blends in dual-fuel engines with biogas, using machine learning for vibration monitoring and neural networks to forecast performance, emissions, and combustion. They found a 25% blend optimal, with 97% accuracy and high fit coefficients of 0.97 to 0.98. Khujamberdiev and Cho [36] conducted a review of ANN applications for modeling the performance and emissions of biodiesel-fuelled diesel engines. They emphasized the effectiveness of ANNs in accurately predicting engine parameters, optimizing efficiency, and lowering pollutant emissions. However, they also identified challenges related to data quality, model generalization, and NOx emission prediction. The study underscores the potential of ANNs to advance sustainable automotive technologies. Paramasivam et al. [37] investigated the optimization of a dual-fuel engine operating on biogas and algal biodiesel, aiming to enhance performance and minimize emissions. They used Response Surface Methodology and applied machine learning models, including Extreme Gradient Boosting (XGBoost) and Random Forest, to analyze experimental data generated by varying engine load and compression ratio. Optimal conditions were identified at 82.71% engine load and a compression ratio of 18.5, resulting in a BTE of 16.98% and reductions in CO, NOx, and hydrocarbon emissions. Among the models, XGBoost exhibited the highest prediction accuracy. Dong et al. [38] sought to optimize diesel engine torque and NOx emissions through ANN predictions. Their approach involved simulating an engine model, generating a comprehensive dataset, and training ANNs optimized with a dung beetle algorithm. The trained models demonstrated excellent predictive accuracy, achieving correlation coefficients above 0.99. Subsequent multi-objective optimization effectively enhanced engine performance while adhering to emission limits. Mohammedali et al. [39] aimed to optimize the performance and emissions of a syngas–diesel dual-fuel engine by combining machine learning and genetic algorithms. They applied a Bayesian-optimized ANN and other predictive models to estimate the indicated thermal efficiency, CO, and NOx from experimental data. The Bayesian-optimized ANN demonstrated the highest prediction accuracy, and coupling it with the Non-dominated Sorting Genetic Algorithm-III facilitated the identification of optimal operating conditions. This approach achieved an indicated thermal efficiency of 41.8% while maintaining controlled emission levels.

These studies show ANN is popular for engines but rarely tied to why stats tools like regression analysis fail on complex relationships in diesel with round per minute (RPM) and throttle variations. Most assume ANNs are superior without proof on real metrics, like our 44.38% higher R^2 for BSFC. This paper bridges the gap by directly comparing TD23 engine data, proving that the ANN captures nonlinear patterns that conventional regression analysis misses. This helps fill a gap in the literature for better engine design. This study introduces new ideas for testing a 4-stroke diesel engine by varying the RPM and throttle. Real data from the TD23 engine were used, and regression analysis was compared with an ANN to estimate quantities such as power, torque, efficiency, BSFC, and A/F ratio. This stands out because older studies mostly use ANNs only for smoke or engine operation, but they skip comparing them to statistical tools like regression when speed and throttle shift. Many prior studies have examined ANNs for biodiesel or gas engines. This study includes all main measures and uses a full factorial design of experiments (DOE) with 24 runs to cover every combination. ANN with two hidden layers and 100 neurons each learns hard without guesses. Diesel engines are tricky to study because speed, throttle, and load blend nonlinearly. Old tools like regression analysis might skip this, but ANN nabs it. This study shows that the ANN achieves the highest accuracy for true engine actions. This helps engineers choose the best tool and reduce cut-test time. For business, it leads to top designs for rules like Euro 6, but with many attempts. No other work combines conventional regression analysis with an ANN for this engine type. This study introduces a new right-check method. This search shines by using the same info in both tools and showing that ANN is best for nonlinear engine data prediction, as evidenced by its performance evaluation.

2. Materials and Methods

2.1 Experimental Setup

Outputs from the engine test bed in the internal combustion engine laboratory were recorded to determine BSFC, BTE, A/F ratio, torque, and power. Sensors recorded data well, which was depicted and saved on the digital display for easy viewing, as shown in Figure 1. The TD23 is a 4-cylinder direct injection diesel engine made by Nissan. We chose it because it is strong and common in trucks and vans like the Nissan Urvan. It delivers up to 44 kW of power and works well across a range of 1100 to 1600 RPM. It makes it perfect for our tests on throttle and load changes. The engine has a 2289 cc displacement, a bore of 89 mm, a stroke of 92 mm, a high compression ratio of 21.9:1, and other engine specifications are briefly listed in Table 1. Different conditions, including changes in throttle position, engine speed, and other loads, were used to provide ample readings, as shown in the experimental data in Table 2. This setup provided a controlled environment for running the experiments and ensured the results could be repeated and trusted. These help study real engine behavior without big issues. This model is widely used in many studies for its ease of setup and consistent results. For example, a study by Yokomizo et al. [40] used the TD23 to test diesel exhaust particles. They ran it at 3000 RPM and found it makes good flow for emission checks. The engine showed stable torque and a low particle size of 20-500 nm. This proves it is good for performance work like this study.

The engine test bed used sensors to measure key things. For engine speed RPM, we had a magnetic pickup sensor on the flywheel. It picks up teeth signals for accurate reading. Calibrated with a hand tachometer to match known speeds. The error is ±1% from three tests. Torque was measured by a load cell on the dynamometer, rated at 98.7 N, and calibrated with standard weights before each run. Uncertainty is ±2% based on repeat loads. Power was calculated as torque times RPM divided by a constant, so its error combines both about ±2.5%. Fuel consumption was measured using a laminar-flow meter calibrated by timing the flow of known amounts of fuel from the tank. Accuracy is ±3% from trials. Airflow was measured with a 51 mm nozzle and a differential pressure sensor. Calibrated with a manometer for pressure drops. Error is ±2%. The A/F ratio came from an exhaust gas analyzer that measures oxygen and CO_2, and it is calibrated weekly with standard gas bottles, such as 21 percent oxygen. Uncertainty is ±4% due to gas variations. BSFC and BTE were calculated using fuel power and efficiency formulas. Overall, the uncertainty for BSFC is ±5% and for BTE, ±4% due to sensor errors. Sensors recorded data well, giving steady readings after these checks, with no large jumps during the tests.

Figure 1. Experimental Setup

Table 1. Specifications of a 4 Stroke Diesel Engine

Engine Specifications		Engine Specifications	
Engine Model	TD23	Link Ratio (L/R)	3.46
Engine Type	Diesel	Compression Ratio	21.9:1
No. of Cylinder	4	Diameter of Air Nozzle	51 mm
No. of Strokes	4	Load Cell	98.7 N
Bore	89 mm	Max Output	44.13kW
Stroke	92 mm	Starting Method	Cell Motor Drive
Piston Displacement	2289 cc	Dynamo Model	EWS-150-L

Table 2. Experimental Data

Serial No.	Engine Speed (RPM)	Governor Level Position (%)	Dynamometer Load (N)	Air-to-Fuel Ratio A/F	Torque (Nm)	Power (kW)	Brake Specific Fuel Consumption (kg/kJ)	Brake Thermal Efficiency (%)
1	1100	10	317	56.1600	90.8205	10.4618	0.0001065	20.6396
2	1100	20	319	51.4671	91.3935	10.5278	0.0001106	19.8806
3	1100	30	328	53.1177	93.9720	10.8248	0.0001088	20.1989
4	1100	40	327	55.9762	93.6855	10.7918	0.0001078	20.3884
5	1200	10	313	57.5496	89.6745	11.2689	0.0001079	20.3767
6	1200	20	315	59.7769	90.2475	11.3409	0.0001066	20.6242
7	1200	30	316	62.2299	90.5340	11.3769	0.0001052	20.8956
8	1200	40	318	57.2204	91.1070	11.4489	0.0001068	20.5838
9	1300	10	322	59.0091	92.2530	12.5590	0.0001090	20.1658
10	1300	20	322	59.0091	92.2530	12.5590	0.0001090	20.1658
11	1300	30	322	58.2489	92.2530	12.5590	0.0001104	19.9060
12	1300	40	322	56.9819	92.2530	12.5590	0.0001129	19.4730
13	1400	10	326	55.4356	93.3990	13.6930	0.0001090	20.1575

Table 3. Continued

Serial No.	Engine Speed (RPM)	Governor Level Position (%)	Dynamometer Load (N)	Air-to-Fuel Ratio A/F	Torque (Nm)	Power (kW)	Brake Specific Fuel Consumption (kg/kJ)	Brake Thermal Efficiency (%)
14	1400	20	328	54.1049	93.9720	13.7770	0.0001110	19.7943
15	1400	30	327	54.9163	93.6855	13.7350	0.0001097	20.0299
16	1400	40	328	54.5268	93.9720	13.7770	0.0001102	19.9487
17	1500	10	329	50.4887	94.2585	14.8061	0.0001080	20.3413
18	1500	20	329	49.0951	94.2585	14.8061	0.0001111	19.7798
19	1500	30	329	49.2534	94.2585	14.8061	0.0001108	19.8436
20	1500	40	332	48.8417	95.1180	14.9411	0.0001107	19.8571
21	1600	10	328	47.3846	93.9720	15.7452	0.0001083	20.3015
22	1600	20	328	46.3711	93.9720	15.7452	0.0001106	19.8672
23	1600	30	329	45.5475	94.2585	15.7932	0.0001123	19.5739
24	1600	40	330	47.9368	94.5450	15.8412	0.0001116	19.7017

2.2 Design of Experiment

Design of experiments is a statistical technique to solve different engineering problems and achieve results with less variability. Several approaches and principles are used in DOE, including Factorial and Taguchi experimentation, Regression, Replication, and Randomization. The main purpose of the DOE is to obtain meaningful results by determining the number of experiments needed to observe the influence of various input parameters on desired output variables. Hence, it has diverse applications in engineering, production, manufacturing processes, resource optimization, and product quality improvement. Several variables regulate processes and procedures. Similarly, the performance optimization of a 4-stroke diesel engine is influenced by design parameters, including combustion type, engine size, fuel injection type and pressure, and compression ratio, as well as by operating parameters, such as ambient air conditions, engine speed (RPM), and dynamometer load.

2.3 Full Factorial DOE

The full factorial design of experiments is one of the most effective DOE methods, providing a clear understanding of the effects of multiple factors on various response parameters by testing all possible combinations of factor levels. This approach captures information about the main effects of individual factors and the interactions among multiple factors, and is also appropriate for analyzing nonlinear interactions. The following formula determines the number of experiments in the Full Factorial method, Eq. (1), where factors are the input parameters and levels are the values for each factor.

$$\textbf{Number of Experiments} = \textbf{Levels}^{\textbf{Factors}} \tag{1}$$

In this study, two factors are taken as inputs: RPM and throttle position, with 6 and 4 levels, respectively, and the experiments to be performed are specified by Eq. (2).

$$\textbf{Number of Experiments} = \mathbf{6^1 \times 4^1 = \ 24\ experiments} \tag{2}$$

The steps for creating a full factorial DOE in Minitab software are shown in Figure 2. Although it provides detailed information on the parameters, it is time-consuming and cost-ineffective when the number of experiments is too large or when performing them is difficult due to resource constraints. In this case, other DOE methods, such as Taguchi or Fractional Factorial, are suitable for minimizing the number of experiments and making the experiment resource- and cost-effective.

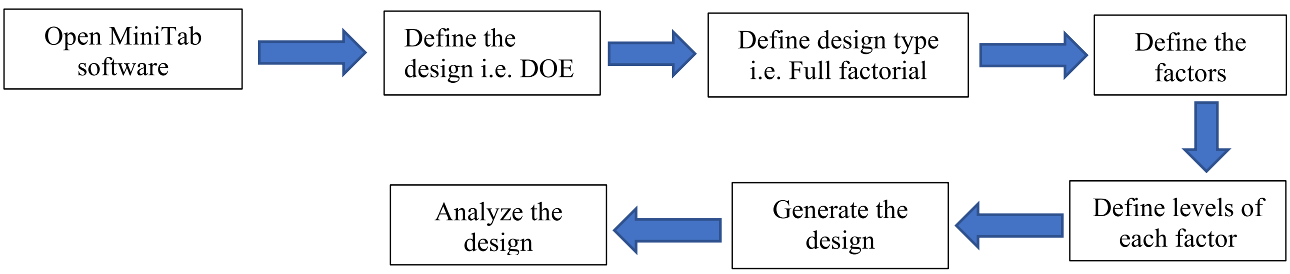


Figure 2. Full Factorial DOE Steps Chart

2.4 Linear Regression

Regression analysis is a statistical technique used to determine the relationship between one or more independent variables (inputs) and a desired dependent or output variable. The analysis may be linear or nonlinear, such as quadratic, cubic, or polynomial regression, depending on the data and results. In linear regression, the response (output) is expressed as a function of predictors in Eq. (3).

$$\mathbf{y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n} \tag{3}$$

where y is the chosen output, x_1 to x_n show the input variables and $a_1$to a_n. These are the coefficients for each independent variable in the linear regression. The quality of linear regression can be assessed for the given data using the coefficient of determination (R^2), which indicates the amount of variation in y explained by the model.

2.5 Data Analysis

Minitab statistical software is used to conduct a Full-Factorial DOE and perform linear regression. The mathematical equations representing the relationships between the dependent variables and the selected independent variables are obtained through regression analysis. The mathematical relationship between torque and the input parameters is shown in Eq. (4).

A = Engine Speed (RPM); B = Throttle Position (%)

$$\mathbf{Torque\ (Nm) = 92.5 - 0.0111\ A + 0.231\ B + 0.000008\ A * A - 0.000144\ A * B} \tag{4}$$

The mathematical relationship between the A/F ratio and the input parameters is shown in Eq. (5).

$$\frac{A}{F} = -98.4 - 0.2507\ A + 0.050\ B + 0.000100\ A * A + 0.00182\ B * B - 0.000043\ A * B \tag{5}$$

The mathematical relationship between the output power and the input parameters is shown in Eq. (6).

$$\text{Power (kW)} = 0.57 - 0.00741\ A + 0.0233\ B + 0.000001\ A * A + 0.000018\ B * B - 0.000014\ A * B \tag{6}$$

The mathematical relationship between the BTE and the input parameters is shown in Eq. (7).

$$\text{BTE} = 22.42 - 0.00249\ A + 0.0126\ B + 0.000001\ A * A + 0.000573\ B * B - 0.000038\ A * B \tag{7}$$

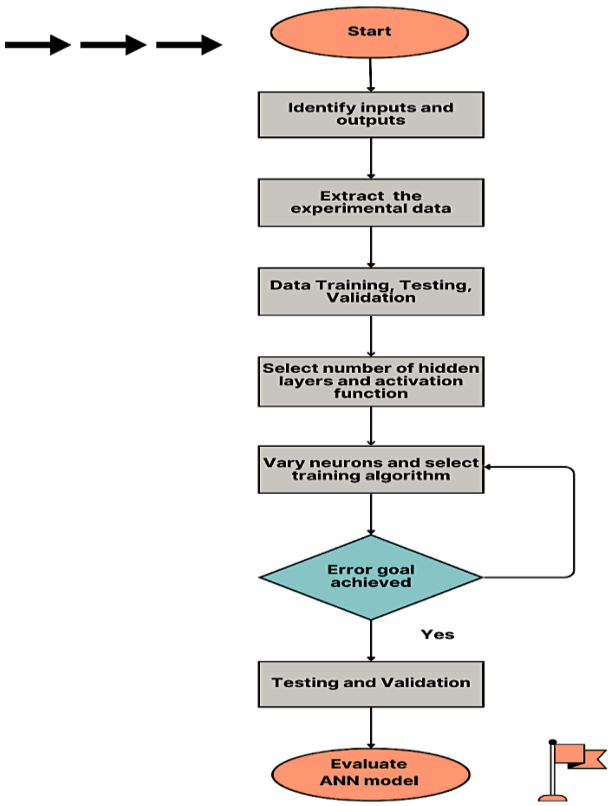


Figure 3. ANN Steps Chart

2.6 Artificial Neural Network

An Artificial Neural Network is the most common type of machine learning algorithm, inspired by the functionality and operation of the human neural system. An ANN model has a network of nodes, similar in structure to the neurons in the human brain. The nodes are interconnected to transfer information between layers. There is an input layer with input nodes (or neurons), one or more hidden layers (depending on the data), and an output layer. The working of the ANN starts from the input layer, in which nodes receive inputs, multiply them with the weights showing the importance of each input, and transfer them to the nodes of the hidden layer, where computations are performed to obtain the results, and the output layer shows the achieved results, like in an image recognition system. A bias is added before the hidden layer, allowing the model to make better predictions and achieve more accurate results. There is an activation function in the computational layer that determines how computations take place. The activation function can be linear, sigmoidal, Tangent Hyperbolic (Tanh), or Rectified Linear Unit (ReLU), with different specified output ranges. The forward uniflow of the process from the input to the output layer is called forward propagation. The ANN model needs to be trained; usually, 80% of the data is used for training, and 20% for testing the trained model. The backward flow from the output layer to the input layer for better prediction, adjusting bias and weights with fewer errors, is called Backpropagation.

Figure 3 illustrates how to create and train an artificial neural network, a type of computer program that mimics the brain in solving problems. The ANN model will take input data and produce the predicted values of the responses we need for our analysis. Second, one collects experimental data for training the network. This is divided into three parts: training (to teach the ANN), testing (to check its learning), and validation (to confirm its accuracy). Next, determine how many hidden layers the ANN needs. These are similar to the problem-solving steps in the brain. You also pick an activation function, which helps the ANN decide which information is important. Then, the number of neurons (the tiny parts of the network that process data) and the training algorithm (how the ANN learns) are selected. The process is repeated until it meets an error goal, that is, until it becomes accurate enough. Once the error goal is reached, the ANN is retested to confirm it performs well. Finally, the ANN is evaluated to ensure it is ready for real-world tasks. ANNs are like intelligent calculators that learn patterns in data to make predictions, anything from recognizing faces to forecasting the weather. ANN has 2 input parameters: engine speed and throttle position. It has 2 hidden layers. The first hidden layer has 100 neurons, and the second has 100 as well. The output layer has 1 neuron for predicted values like power, torque, etc. The pictorial representation of the whole ANN is shown in Figure 4.

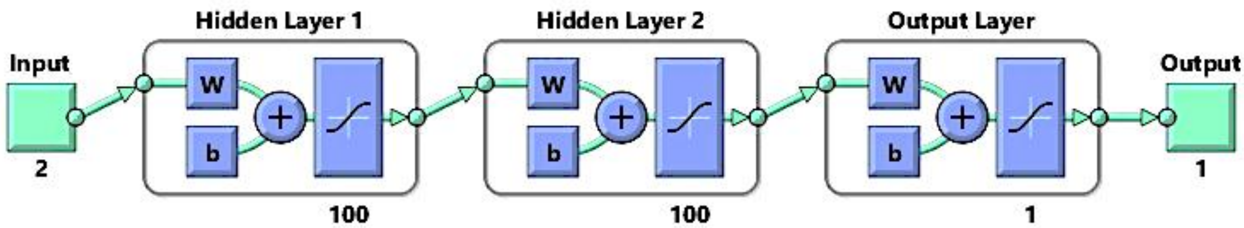


Figure 4: Number of hidden layers and neurons in each layer for ANN

3. Results and Discussion

3.1 Regression Analysis

The following Figure 5 shows the effect of various parameters, i.e., Engine Speed (RPM), Throttle Position (%), and Dynamometer Load (N), on performance metrics that are BTE (%), BSFC (kg/kJ), Power(kW), Torque (Nm), and A/F. From Figure 5(a), it seems that dynamometer load has the most effect, engine speed has a moderate effect, and throttle position has the least effect on BTE. The engine is most fuel-efficient at low speed, and as speed increases, efficiency decreases slightly. As the throttle position increases, efficiency decreases slightly, and the engine is most fuel-efficient under moderate load. The BTE is observed to be 20.8% at a moderate load of 316N, 20.6% at 1200 rpm, and 20.3% at 10% throttle position. From Figure 5(b), it seems that the dynamometer has the most, engine speed has moderate, and 53.75 at 10% throttle position. Throttle position has a minimal effect on BSFC. BSFC is highest at medium RPMs, which provide the most fuel-efficient engine performance. As the throttle position increases, the BSFC rises slightly, and at moderate load, it is highest. The BSFC is observed 1.115×10^{-4} kg/kJ at 330N of dynamometer load, 1.104×10^{-4} kg/kJ at 1300 rpm, and 1.10×10^{-4} kg/kJ at 40% of Throttle position. From Figure 5(c), it appears that dynamometer load, engine speed, and throttle position have the greatest, moderate, and least effect on power (kW). Power output increases consistently with RPM, indicating a direct relationship between power and engine speed. There is no significant variation in power with throttle position. Power increases with dynamometer load, reaching its highest value at high load. The Power is observed at 15.9kW at 1600rpm and 330N and 13.2 kW at 40% of throttle position. From Figure 5(d), it appears that engine speed has the greatest effect on Torque, the dynamometer has a moderate effect, and throttle position has a minimal effect. Torque decreases initially, then increases as RPMs increase. Torque increases moderately and consistently with throttle position. The Torque is observed to be 95.2 Nm at 332N load, 94.5 Nm at 1600 rpm, and 93.4 at 40% of throttle position. From Figure 5(e), it appears that dynamometer load, engine speed, and throttle position have the greatest, moderate, and least effect on A/F ratio. Initially, it increases with engine speed, then decreases as RPMs increase. The A/F ratio remains relatively constant with throttle position. It increases slightly, peaks, and then decreases as the load increases. The A/F ratio is observed to be 62.5 at 316N of load, 58.75 at 1200 rpm, and 53.75 at 10% throttle position.

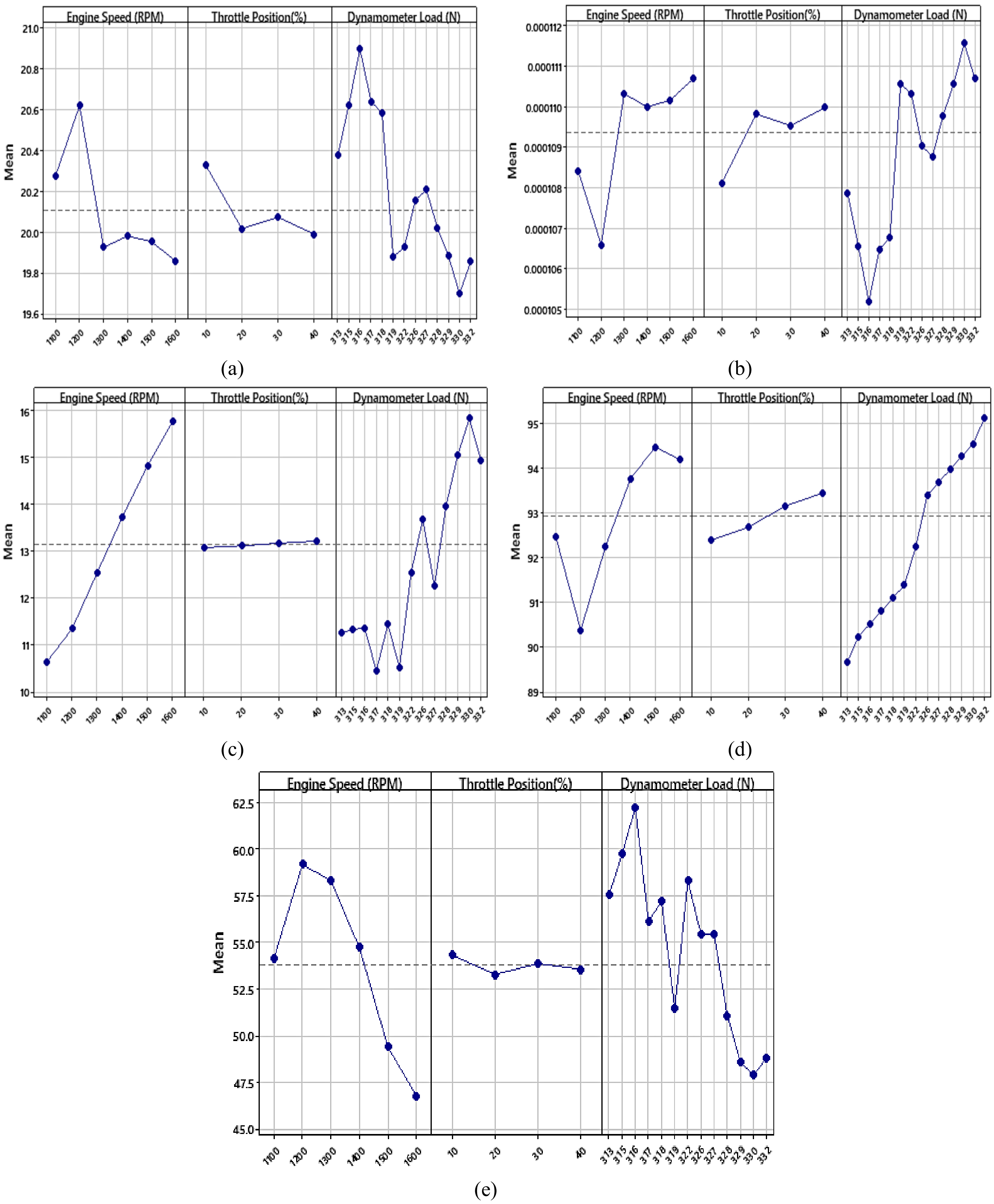


Figure 5 Main effect plots for output responses (a) Brake thermal efficiency, (b) Brake specific fuel consumption, (c) Power, (d) Torque, (e) Air-to-fuel ratio

Figure 6 shows the interaction plots of throttle position (%) with the parameters BTE, BSFC, power (kW), Torque (Nm), and A/F ratio. Here is the detailed discussion of the trend and interaction at different RPMs for each graph. From Figure 6(a), it appears that engine speeds of 1200 rpm, 1300 rpm, 1600 rpm, 1400 rpm, and 1500 rpm have the greatest, moderate, and minimal effects on BTE, respectively, remaining relatively constant across throttle positions. At 1100 rpm, efficiency increases slightly with throttle and peaks at 30%. At 1200 rpm, efficiency decreases initially, then rises to 30% at 1300 rpm. At 1400 rpm, efficiency increases slightly. At 1500 rpm, efficiency remains constant and decreases significantly at 1600 rpm. The highest BTE is 20.875 at 30% throttle position, which occurs at 1200 rpm. From Figure 6(b), the most significant interaction occurs at 1600 rpm, with a peak at 30% throttle position, while a moderate interaction occurs at 1300 rpm, with noticeable variation. The least interaction is at 1500 rpm, where efficiency remains nearly unchanged at all throttle positions. The highest BTE is 1.128 $\times 10^{-4}$ kg/kJ at 40% throttle position, observed at 1300 rpm.

From Figure 6(c), it appears that at 1100 rpm, power increases gradually with throttle position, and this is the most significant interaction. At 1200 rpm, there is a slight increase, indicating a moderately significant interaction. At 1300, 1400, 1500, and 1600 rpm, the power remains constant and shows minimal interaction. The maximum power across all speeds is 20.875kW at 40% throttle position, at 1600 rpm. From Figure 6(d), it appears that at 1100 rpm, torque increases consistently with throttle position, indicating a strong interaction between engine speed and throttle position. At 1200 rpm, torque increases moderately with increasing throttle position, indicating moderate interaction. At 1300, 1400, 1500, and 1600 rpm, the torque shows only slight variation, indicating minimal interaction. The maximum torque is 95.2 Nm at 40% throttle position, at 1500 rpm. At 1100 rpm, the A/F ratio initially decreases and then increases, indicating the most significant interaction with throttle position. At 1200 rpm, the A/F ratio gradually increases as the throttle position rises, indicating moderate interaction. At 1300, 1400, 1500, and 1600 rpm, the A/F ratio decreases slightly, indicating the least interaction at higher engine speeds. Among these, 1200 rpm has the highest A/F ratio, and 62.2 occurs at 30% throttle position.

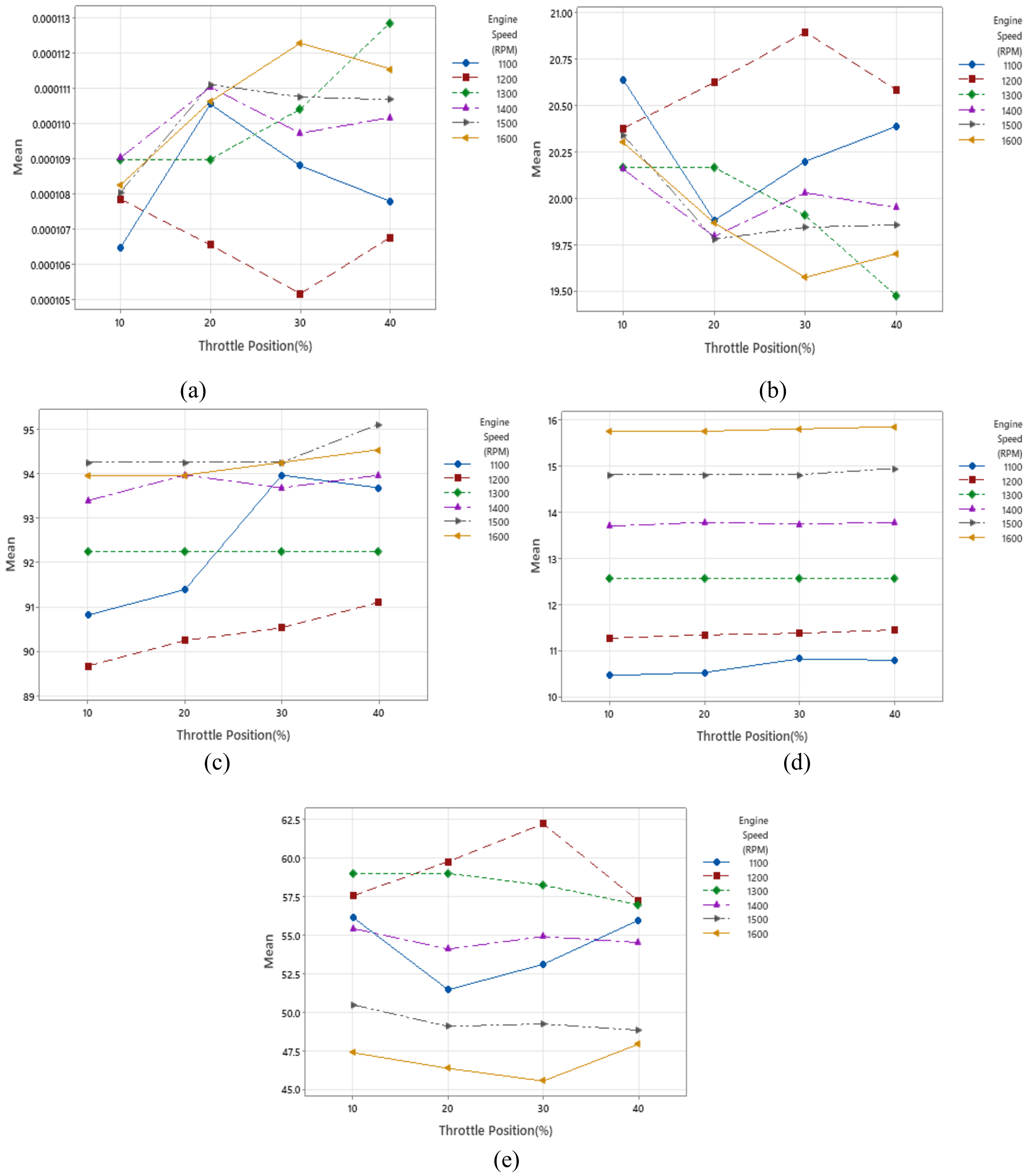
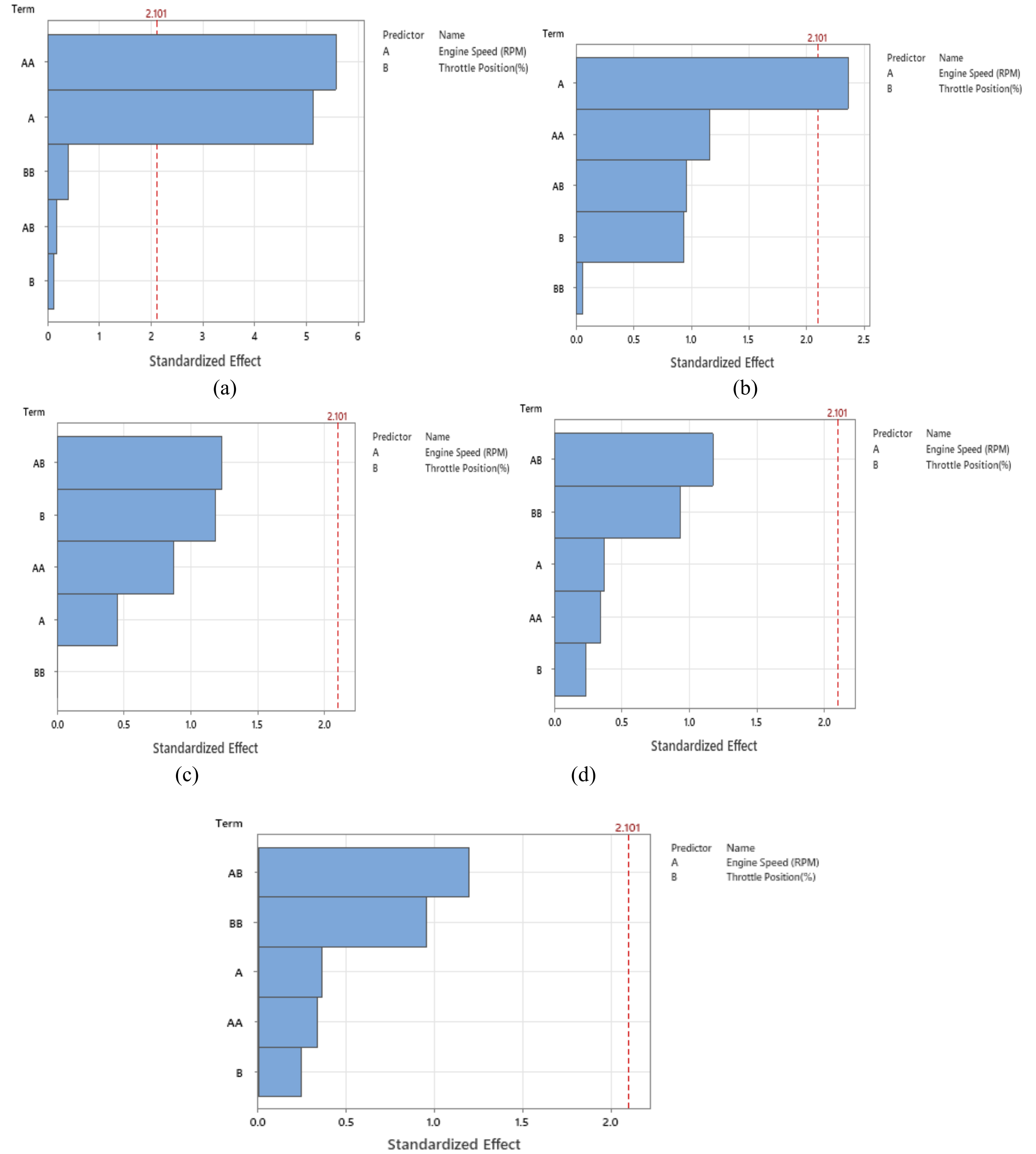


Figure 6 Interaction plots for output responses (a) Brake specific fuel consumption (b) Brake thermal efficiency (c) Torque (d) Power (e) Air-to-fuel ratio

Figure 7 shows the effect of each parameter on the specified output. A Pareto chart is a simple bar chart used to identify and prioritize problems or causes. It helps focus on the most important issues. The chart has bars and a line. The bars represent the size of each problem or cause, arranged from biggest to smallest. The line represents the total percentage as you add each bar. It is based on the "80/20 rule," which means 80% of the problems come from 20% of the causes. As illustrated in Figure 7(a), the most important component affecting this reaction is the quadratic term of engine speed (AA), which has the biggest standardized effect on the A/F ratio. Although it is slightly smaller than the quadratic component, the linear term of engine speed (A) also has a considerable impact, showing the significance of both linear and nonlinear interactions of engine speed. However, because they are below the threshold value of 2.101, the effects of throttle position (B), its quadratic component (BB), and the interaction between engine speed and throttle position (AB) are all insignificant. This demonstrates that engine speed is the primary determinant of the A/F ratio. As illustrated in Figure 7(b), the Pareto chart shows that engine speed (A) has the strongest impact on power. The squared term for engine speed (AA) and its interaction with throttle position (AB) also have significant effects. Throttle position (B) and its squared term (BB) have the least and statistically negligible effects.



(a) (b) (c) (d) (e)

Figure 7 Pareto charts for output responses (a) Air-to-fuel ratio, (b) Power, (c) Torque, (d) Brake thermal efficiency, (e) Brake specific fuel consumption

Figure 7(c) shows that torque is most affected by the combined effect of engine speed and throttle position (AB), with a standardized effect above the threshold value of 2.101. Throttle position (B) has a moderate effect, while engine speed (A) and its curve (AA) have the least effect. The curve of throttle position (BB) has a negligible effect. From Figure 7(d), the Pareto chart for BTE shows that the interaction between Engine Speed (A) and Throttle Position (B), represented as

(AB), has the most effect but is not significant. The quadratic effect of throttle position (BB) and the linear effect of engine speed (A) are both medium in magnitude but fall short of significance. The quadratic effect of engine speed (AA) and the linear effect of throttle position (B) have the least effect. From Figure 7(e), the Pareto chart for BSFC shows that the interaction between engine speed (A) and throttle position (B), as represented by (AB), has a strong and significant effect but is not statistically significant. The quadratic effect of Throttle Position (BB) is the second most effective, followed by the linear effect of engine speed (A) and the quadratic effect of engine speed (AA), both with a smaller effect. The linear effect of throttle position (B) is the least. None of the predictors significantly affects BSFC, though AB and BB show relatively higher effects. For the A/F ratio in Figure 7(a), the quadratic engine speed AA has the largest effect because airflow rises nonlinearly with speed, as volumetric efficiency drops at high RPMs due to intake limits and turbulence. This makes the quadratic term key, as speed squared impacts air intake. Linear speed A is strong too for basic air increase. The lowest are throttle B, quadratic BB, and interaction AB, as diesel controls power by fuel, not air throttle, so these add little change. In Figure 7(b), the linear engine, speed A is highest because power equals torque times RPM, so speed increases directly when torque is held constant. Quadratic AA and interaction AB matter, as speed curves and throttle-speed mix affect output. Lowest throttle B and quadratic BB, because at a given load, throttle plays a small role in power. For torque in Figure 7(c), the interaction AB is at the top, as torque depends on fuel from the throttle at certain speeds for best combustion and pressure. Throttle B is moderate for fuel control. Lowest speed A, quadratic AA, and throttle BB, with speed alone or throttle curve, have minor torque influence. For BTE in Figure 7(d), the strongest interaction is between speed (A) and throttle (B), although it was not statistically significant. The efficiency is highest at specific speed and throttle settings, as these conditions give better heat utilization and cleaner combustion. The quadratic throttle-term influence, BB, and the main speed effect, A, are moderate, showing a curvilinear relationship with fuel use. The quadratic speed term, AA, and throttle effect, B, have the smallest impact and contribute the least to efficiency. BSFC, in Figure 7(e) interaction, AB is strongest, as fuel per power changes with speed due to load efficiency gains at high speeds. Quadratic BB follows the speed A and AA for curve impacts. Lowest throttle B, since linear throttle shifts little on BSFC.

This regression analysis shows how engine speed, throttle position, and dynamometer load affect performance metrics, including BTE, BSFC, power, torque, and A/F. It was found that dynamometer load has the greatest effect on BTE and BSFC, while engine speed has a greater effect on power and torque. Throttle position has less impact but still changes things at different RPMs. For example, BTE peaks at 20.8% with a moderate load of 316 N and 1200 RPM with 30% throttle. Power reaches 15.9 kW at 1600 RPM and 330N. Torque hits 95.2 Nm at 332N load and 1500 RPM with 40% throttle. A/F ratio reaches 62.5 at 316N and 58.75 at 1200 RPM. These trends align with what others have observed in diesel engines. Some studies support the findings presented in this study. Zhang et al. [8] studied diesel-ethanol blends and found that engine speed and load strongly affect power and torque, as shown. They noted power rises with RPM, which matches our 15.9 kW at 1600 RPM. Akkoli et al. [24] tested a diesel engine with producer gas and observed a drop in BTE at high speeds, consistent with our drop at 1600 RPM. Their BTE was around 20% at moderate loads, like our 20.8%. This shows our results fit with past work. In this study, interaction plots also show engine speed and throttle position work together, especially at 1100-1200 RPM for BTE and A/F ratio. This is similar to Syta et. al. [2], who found speed changes affect torque and emissions in aircraft engines. This study stands out because we used a full factorial design with 24 tests covering 6 RPM levels and 4 throttle positions. This gives clear data on interactions that many studies miss. For example, Roy et al. [20] used an ANN for a single-cylinder engine but did not present detailed regression plots like ours. Our Pareto charts highlight engine speed (A) and its quadratic term (AA) as key factors for A/F and power matching (Norouzi et al. [7]), who found that speed dominates NOx prediction. However, our BSFC shows no strong effect from any factor, unlike Zhang et al. [8], who reported a strong effect of the load. This might be due to our TD23 engine's older design or test range. Compared to the literature, our work adds new value by combining Minitab's statistical view with real data trends. Past studies like Hoang et al [22] focused on ANN, but we show that regression can still guide initial analysis. Our detailed plots and 80/20 rule in Pareto charts give a fresh look at diesel performance. Future work can test wider loads to confirm BSFC trends. We think this makes our paper stronger and useful for engine research. Quadratic regression in Minitab is used because engine data, such as torque and power, exhibit nonlinear trends. Linear models are too simple and miss curves in speed and throttle effects. For example, power rises with RPM but slows at high speeds due to friction. Quadratic terms like A×A capture this bend without making the model too complex for 24 data points. Higher-order terms, such as cubic functions, can overfit and lead to poor predictions on new data. The interaction AB is also added to show how speed and throttle interact in the outputs. This is common in engine studies as combustion is not straight.

3.2 ANN Analysis

This work compares the predictive performance of the ANN and Minitab regression models in analyzing the main parameters of the 4-stroke diesel engine: power, A/F ratio, efficiency, BSFC, and torque. The comparison was performed using two metrics: the RMSE, which represents the prediction error, and the coefficient of determination, which represents the quality of fit between the predicted and actual values. The lower the RMSE and the higher the R^2, the better the performance. Figure 8(a) shows the model performance across training, validation, and test sets, and across all datasets, with varying regression slopes and intercepts for the A/F ratio. Similarly, Figures 8(b-e) show the model's performance for outputs such as power, torque, BTE, and BSFC. In power prediction, regression analysis performed better, with a lower RMSE of 0.1406, compared to 0.3561 for the ANN. Moreover, regression analysis yielded a slightly higher R^2 of 0.9955, whereas ANN yielded 0.9867.

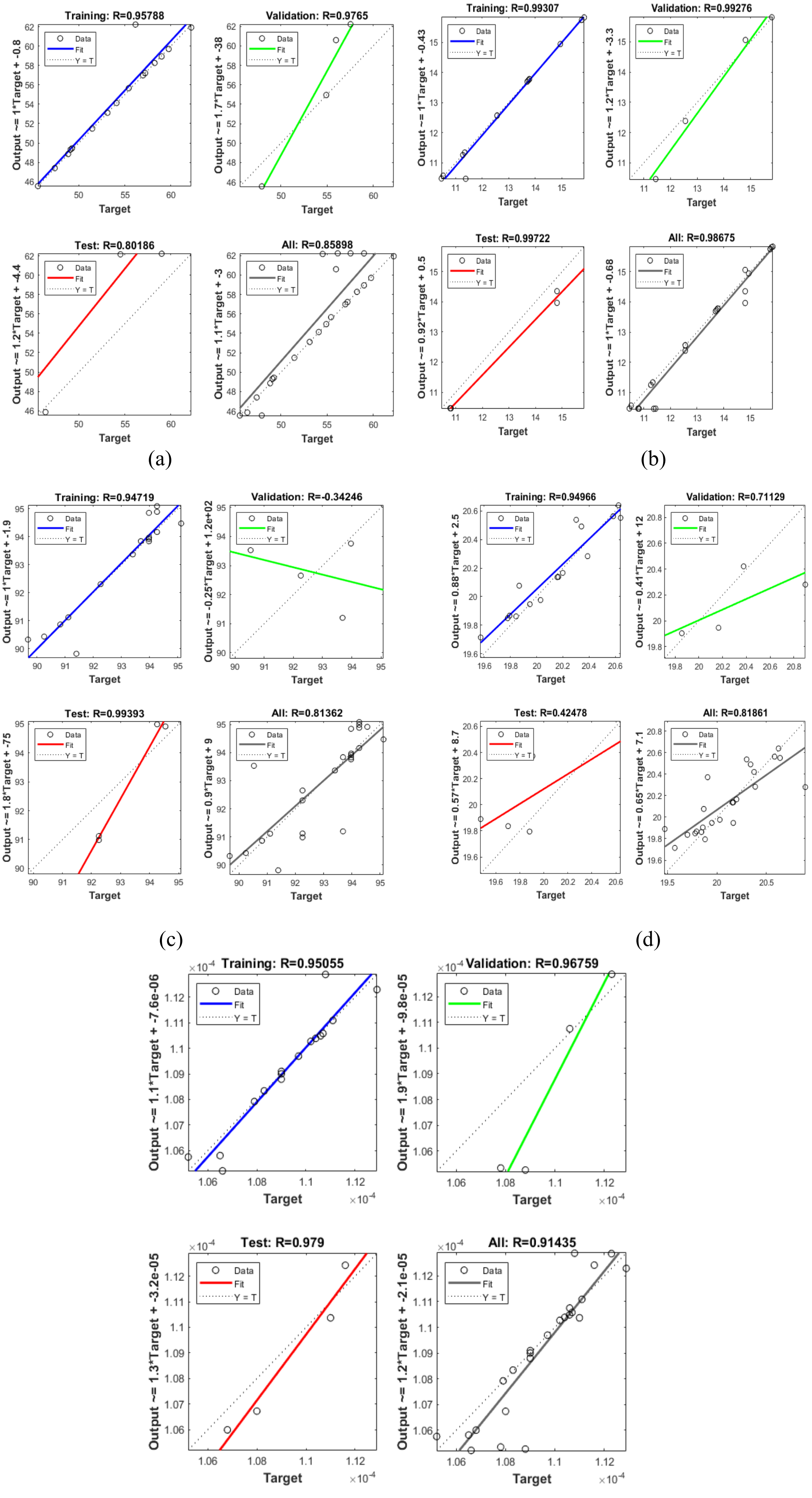


Figure 8 ANN regression for output responses (a) Air-to-fuel ratio, (b) Power, (c) Torque, (d) Brake thermal efficiency, (e) Brake specific fuel consumption

Therefore, regression analysis shows better accuracy in predicting power for this data set; however, the ANN's performance is quite good and competitive, given real-world applications with nonlinear relationships and data variations. When predicting the A/F ratio, the ANN outperformed the regression model, with an R^2 of 0.8589 compared to 0.8326, indicating a stronger correlation between the predicted and actual values. ANN also achieved an RMSE of 3.3352, slightly higher than the regression model's 2.1875. Despite the minor difference in RMSE, the ANN's higher R^2 indicates its ability to model complex relationships in A/F ratio prediction, making it more reliable for this parameter. For efficiency predictions, ANN outperformed regression analysis by a significant margin. ANN recorded an RMSE of 0.2066 and an R^2 of 0.8186, while regression analysis has a higher RMSE of 0.2998 and a much lower R^2 of 0.4660. The considerable difference in R^2 highlights ANN's ability to capture the intricate relationships that influence engine efficiency, whereas Minitab's regression model shows limitations in accurately representing these complexities. The BSFC results further highlighted ANN strength, with an RMSE of 1.1159×10^{-6} and an R^2 of 0.9143. Regression analysis, on the other hand, has a higher RMSE of 1.6×10^{-6} and a significantly lower R^2 of 0.4705. The above findings clearly show that ANN predicts more accurate values and fits the actual data well; thus, it is more reliable when applied to parameters where accuracy is not compromised, such as fuel consumption. In torque prediction, the ANN has an RMSE of 1.4698, slightly higher than the RMSE of 1.0937 obtained by regression analysis. However, the ANN achieved a much higher R^2 of 0.8136 than the regression analysis, which achieved 0.6205. The RMSE values suggest that conventional regression analysis offers slightly better numerical accuracy; however, the much higher R^2 value of the ANN indicates it is better at capturing the underlying relationship affecting torque and, in practical scenarios, provides more reliable predictions. The analysis shows that although regression analysis sometimes yields lower RMSE values, ANN consistently yields higher R^2 values across most parameters, indicating stronger predictive relationships. The ability of ANNs to adapt to nonlinear patterns and provide strong predictions across a wide range of engine performance parameters makes them a very reliable tool for analyzing complex datasets. Table 3 shows the RMSE and R^2 values for different output parameters. Power regression analysis has a lower RMSE of 0.1406 and a higher R^2 of 0.9955 than ANN, with RMSEs of 0.3561 and 0.9867, respectively. For the A/F ratio, ANN has a higher RMSE (3.3352) than regression analysis (2.1875), but a better R^2 (0.8589) than regression analysis (0.8326). We still prefer ANN because it captures nonlinear relations in engine data better. Engine performance, such as power and A/F ratio, involves complex nonlinear factors from combustion and airflow. Regression analysis uses linear or quadratic regression, which fits well to small datasets like our 24 points, but may not generalize to new conditions. ANNs learn patterns without a fixed form, so they are more reliable for real-world engines that change. Regression analysis is better in RMSE for power because power may be more linear with speed in our range, so a simple model fits tightly with low error. For the A/F ratio, regression analysis has a lower RMSE, but the ANN has a higher R^2, indicating it explains more variance despite larger errors at some points. ANN is weak on RMSE here due to a small dataset, which makes training difficult and can lead to overfitting or higher prediction variance. With more data, the ANN would improve. A recent study supports this. Kurtgoz et al. [23] used an ANN over regression for a spark-ignition (SI) engine and noted that the ANN was preferred for generalization despite small errors in fit metrics on test data.

Table 4. Comparison Analysis of ANN and Regression

Outputs	ANN		Regression		Preference based on capturing non-linearities
	RMSE	R^2	RMSE	R^2	
Power	0.3561	0.9867	0.1406	0.9955	ANN
A/F	3.3352	0.8589	2.1875	0.8326	ANN
Efficiency	0.2066	0.8186	0.2998	0.4660	ANN
BSFC	0.0000011159	0.9143	0.0000016	0.4705	ANN
Torque	1.4698	0.8136	1.0937	0.6205	ANN

4. Conclusions

Dynamometer loads have the greatest impact on performance parameters. Engine speed has a stronger influence, and throttle position has a minor effect. Brake thermal efficiency and brake-specific fuel consumption are best at moderate loads and at lower to medium speeds. Power and torque rise with load and speed. Air-to-fuel ratio changes with load and speed, too. Interactions show that engine speed strongly affects metrics at some throttle positions. Brake thermal efficiency and air-to-fuel ratio peak at 30% throttle position. Power and torque interactions peak at lower speeds with constant variations at higher speeds. The Pareto chart shows that engine speed has the greatest effect on airflow and air-to-fuel ratio. Torque and power are mainly affected by how engine speed and throttle position interact. Throttle position affects fuel use and efficiency, but only slightly. Results show that ANN performed better than conventional regression analysis in most cases. For power, ANN has an R^2 that is 0.89% lower than that of regression analysis, but we prefer ANN for nonlinear data. For the air-to-fuel ratio, the ANN has an R^2 that is 2.63% higher than that of regression analysis. For efficiency, the ANN has an R^2 that is 35.26% higher than that of regression analysis. For brake-specific fuel consumption, ANN has an R^2 that is 44.38% higher than that of regression analysis. For torque, ANN has an R^2 that is 19.31% higher than that of regression analysis. This shows that ANNs are more accurate than methods like regression analysis at understanding complex engine behavior by capturing nonlinear patterns more effectively.

Acknowledgements
The authors thank the University of Engineering and Technology, Pakistan, for providing the laboratory facilities that enabled this work.

Funding
This study was not supported by any grants from funding bodies in the public, private, or not-for-profit sectors.

Declaration of Competing Interest
The author declares no conflicts of interest.

CRediT Authorship Contribution Statement
Muhammad Asad (Conceptualisation; Methodology; Formal analysis; Software; Writing - original draft; Supervision; Project administration)
Umer Iftikhar (Data curation; Software; Validation; Investigation; Visualisation)
Daniyal Abbasi (Formal analysis;Writing - review & editing; Visualisation)
Muhammad Bilal Jamil (Data curation; Writing)

Availability of Data and Materials
The data supporting this study's findings are available on request from the corresponding author.

Ethics Declarations
This study did not involve human participants or animals. Ethical approval was therefore not required.

Generative Artificial Intelligence Declarations
The authors stated that generative AI was not used to generate content, ideas, or theories. We have just utilized AI to enhance readability and refine the language. This was used with extreme human control and oversight. The authors take full responsibility for reviewing and approving the content.

References

[1] K.M. Akkoli, N.R. Banapurmath, M.M. Shivashimpi, M.E.M. Soudagar, I.A. Badruddin, M.A. Alazwari, et al., "Effect of injection parameters and producer gas derived from redgram stalk on the performance and emission characteristics of a diesel engine," *Alexandria Engineering Journal*, vol. 60, no. 3, pp. 3133-3142, 2021.

[2] A. Syta, J. Czarnigowski, P. Jakliński, "Detection of cylinder misfire in an aircraft engine using linear and nonlinear signal analysis," *Measurement*, vol. 174, p. 108982, 2021.

[3] B. Mehlig, "*Machine learning with neural networks: an introduction for scientists and engineers,"* UK: Cambridge University Press, 2021.

[4] G. Sakthivel, "Prediction of CI engine performance, emission, and combustion characteristics using fish oil as a biodiesel at different injection timing using fuzzy logic," *Fuel*, vol. 183: pp. 214-229, 2016.

[5] D. Hao, R.K. Mehra, S. Luo, Z. Nie, X. Ren, M. Fanhua, "Experimental study of hydrogen-enriched compressed natural gas (HCNG) engine and application of support vector machine (SVM) on prediction of engine performance at specific condition," *International Journal of Hydrogen Energy*, vol. 45, no. 8, pp. 5309-5325, 2020.

[6] R. Dancila, R. Botez, "New flight trajectory optimisation method using genetic algorithms," *The Aeronautical Journal,* vol. 125, no. 1286, pp. 618-671, 2021.

[7] A. Norouzi, M. Aliramezani, C.R. Koch, "A correlation-based model order reduction approach for a diesel engine NOx and brake mean effective pressure dynamic model using machine learning," *International Journal of Engine Research*, vol. 22, no. 8, pp. 2654-2672, 2021.

[8] Z. Zhang, J. Li, J. Tian, G. Xie, D. Tan, B. Qin, et al., "Effects of different diesel-ethanol dual fuel ratio on performance and emission characteristics of diesel engine," *Processes*, vol. 9, no. 7, p. 1135, 2021.

[9] D.O. Onukwuli, C. Esonye, A.U. Ofoefule, R. Eyisi, "Comparative analysis of the application of artificial neural network-genetic algorithm and response surface methods-desirability function for predicting the optimal conditions for biodiesel synthesis from chrysophyllum albidum seed oil," *Journal of the Taiwan Institute of Chemical Engineers*, vol. 125, pp. 153-167, 2021.

[10] N. Gammoudi, M. Mabrouk, T. Bouhemda, K. Nagaz, A. Ferchichi, "Modeling and optimization of capsaicin extraction from Capsicum annuum L. using response surface methodology (RSM), artificial neural network (ANN), and Simulink simulation," *Industrial Crops and Products*, vol. 171, p. 113869, 2021.

[11] A.T. Hoang, S. Nižetić, H.C. Ong, W. Tarelko, V.V. Pham, T.H. Le,, et al., "A review on application of artificial neural network (ANN) for performance and emission characteristics of diesel engine fueled with biodiesel-based fuels," *Sustainable Energy Technologies and Assessments*, vol. 47, p. 101416, 2021.

[12] A.V. Prabhu, A. Alagumalai, A. Jodat, "Artificial neural networks to predict the performance and emission parameters of a compression ignition engine fuelled with diesel and preheated biogas–air mixture," *Journal of Thermal Analysis and Calorimetry*, vol. 145, pp. 1935-1948, 2021.

[13] M.I.L. Galdo, J.T. Miranda, J.M.R. Lorenzo, C.G. Caccia, "Internal modifications to optimize pollution and

emissions of internal combustion engines through multiple-criteria decision-making and artificial neural networks," *International Journal of Environmental Research and Public Health*, vol. 18, no. 23, p. 12823, 2021.

[14] A.N. Bhatt, N. Shrivastava, "Application of artificial neural network for internal combustion engines: a state of the art review," *Archives of Computational Methods in Engineering*, vol. 29, no. 2, p. 897-919, 2022.

[15] X.H. Fang, N. Papaioannou, F. Leach, M.H. Davy, "On the application of artificial neural networks for the prediction of NOx emissions from a high-speed direct injection diesel engine," *International Journal of Engine Research*, vol. 22, no. 6, p. 1808-1824, 2021.

[16] M. Aliramezani, C.R. Koch, M. Shahbakhti, "Modeling, diagnostics, optimization, and control of internal combustion engines via modern machine learning techniques: A review and future directions," *Progress in Energy and Combustion Science*,vol. 88, p. 100967, 2022.

[17] W. Yu, F. Zhao, "Predictive study of ultra-low emissions from dual-fuel engine using artificial neural networks combined with genetic algorithm," *International Journal of Green Energy*, vol. 16, no. 12, p. 938-946, 2019.

[18] I. Veza, M.F. Muhamad Said, Z. Abdul Latiff, M.A. Abas, "Application of Elman and Cascade neural network (ENN and CNN) in comparison with adaptive neuro fuzzy inference system (ANFIS) to predict key fuel properties of ABE-diesel blends," *International Journal of Green Energy*, vol. 18, no. 14, p. 1510-1522, 2021.

[19] M.K.D. Kiani, B. Ghobadian, T. Tavakoli, A.M. Nikbakht, G. Najafi, "Application of artificial neural networks for the prediction of performance and exhaust emissions in SI engine using ethanol-gasoline blends," *Energy*, vol. 35, no. 1, p. 65-69, 2010.

[20] S. Roy, R. Banerjee, P.K. Bose, "Performance and exhaust emissions prediction of a CRDI-assisted single cylinder diesel engine coupled with EGR using artificial neural network," *Applied Energy*, vol. 119, p. 330-340, 2014.

[21] S. Arumugam, G. Sriram, P.S. Subramanian, "Application of artificial intelligence to predict the performance and exhaust emissions of diesel engine using rapeseed oil methyl ester," *Procedia engineering*, vol. 38, p. 853-860, 2012.

[22] A.T. Hoang, S. Nižetić, H.C. Ong, W. Tarelko, V.V. Pham, T.H. Le, et al., "A review on application of artificial neural network (ANN) for performance and emission characteristics of diesel engine fueled with biodiesel-based fuels," *Sustainable Energy Technologies and Assessments,* vol. 47, p. 101416, 2021.

[23] Y. Kurtgoz, M. Karagoz, E. Deniz, "Biogas engine performance estimation using ANN," *Engineering Science and Technology, An International Journal*, vol. 20, 6, p. 1563-1570, 2017.

[24] K.M. Akkoli, N.R. Banapurmath, M.M. Shivashimpi, M.E.M. Soudagar, I.A. Badruddin, M.A. Alazwari, et al., "Effect of injection parameters and producer gas derived from redgram stalk on the performance and emission characteristics of a diesel engine," *Alexandria Engineering Journal*, vol. 60, no. 3, p. 3133-3142, 2021.

[25] M. Zare, "Compare neural network and linear regression when there exist outliers or sensitive data: Advantages and disadvantages. 2024.

[26] R.N. Prasad, P.N. Devi "A comparative analysis of machine learning algorithms for big data applications in predictive analytics," *International Journal of Scientific Research and Management*, vol. 12, no. 10, pp. 1608-30 2024.

[27] M. Grebovic, L. Filipovic, I. Katnic, M. Vukotic, T. Popovic, "Overcoming limitations of statistical methods with artificial neural networks," *in 2022 International Arab Conference on Information Technology*, IEEE, 2022.

[28] Y. Li, Y. Wang, Z. Lin, H. Xie, "An efficient approach to regression problems with tensor neural networks". arXiv preprint arXiv:2406.09694, 2024.

[29] P. Belany, P. Hrabovsky, S. Sedivy, N. Cajova Kantova, Z. Florkova, "A comparative analysis of polynomial regression and artificial neural networks for prediction of lighting consumption," *Buildings*, vol. 14, no. 6, p. 1712, 2024.

[30] M, Rahman, M. Asadujjaman, "Implementation of artificial neural network on regression analysis," *In 2021 5th Annual Systems Modelling Conference*, IEEE, 2021.

]31] H. Oğuz, I. Sarıtas, H.E. Baydan, "Prediction of diesel engine performance using biofuels with artificial neural network," *Expert Systems with Applications*, vol. 37, no. 9, p. 6579-6586, 2010.

[32] Y. Wang, J. Li, G. Wang, G. Chen, S. He, "Prediction of diesel particulate filter regeneration conditions and diesel engine performance under regeneration mode using AMSO-BPNN and combined with XGBoost," *Applied Energy*, vol. 377, p. 124341, 2025.

[33] H. Dave, V. Vakharia, H. Panchal, M.I.H. Siddiqui, D. Dobrotă, "ANN and multilayer-ELM based prediction of combustion, performance and emission characteristics of a diesel engine fuelled with Diesel-DTBP blends," *Case Studies in Thermal Engineering*, vol. 72, p. 106323, 2025.

[34] O. Odufuwa, L. Tartibu, K. Kusakana, "Artificial neural network modelling for predicting efficiency and emissions in mini-diesel engines: Key performance indicators and environmental impact analysis," *Fuel*, vol. 387, p. 134294, 2025.

[35] N.K. Pallicheruvu, S. Gnanasekaran, "ANN-driven prediction of optimal machine learning models for engine performance in a dual-fuel mode powered by biogas and fish oil biodiesel," *Energy Conversion and Management*, vol. 25, p. 100827, 2025.

[36] R. Khujamberdiev, H.M. Cho, "Artificial intelligence in automotives: ANNs' impact on biodiesel engine performance and emissions," *Energies*, vol. 18, no. 2, p. 438, 2025.

[37] P. Paramasivam, K. Alnamasi, A.M.A. Alsharif, P.K. Kanti, "Performance and emission analysis of a dual-fuel engine using biogas and algal biodiesel: Machine learning prediction and response surface optimization," *Case*

Studies in Thermal Engineering, p. 107227, 2025

[38] A. Dong, H. Cui, C. Zhao, Y. Guan, "Neural network prediction based on dung beetle optimization algorithm and engine performance emission optimization using multi-objective rime optimization algorithm," *Energy*, p. 137553, 2025

[39] A.A.M. Mohammedali, A. Albadwi, A.A.M. Omara, K. Ali, M.I.H. Ali, "Machine learning prediction and evolutionary multi-objective optimization on performance and emissions of a syngas-diesel dual-fuel engine," *International Journal of Hydrogen Energy*, vol. 162, p. 150781, 2025.

[40] H. Kawakami, A. Zukeran, K. Yasumoto, M. Kubojima, Y. Ehara, T. Yamamoto, et al., "Effect of AC electrostatic precipitator on removal diesel exhaust particles," *IEEJ Transactions on Fundamentals and Materials*, vol. 131, no. 3, p. 192-198, 2011.