



Prediction of combustion, performance, and emission parameters of ethanol powered spark ignition engine using ensemble Least Squares boosting machine learning algorithms

D. Jesu Godwin ^a, Edwin Geo Varuvel ^{a,b,*}, M. Leenus Jesu Martin ^a

^a Department of Automobile Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, 603203, Chengalpattu District, Tamil Nadu, India

^b Department of Mechanical Engineering, Faculty of Engineering and Natural Sciences, Istinye University, Istanbul, Turkey

ARTICLE INFO

Handling Editor: Giovanni Baiocchi

Keywords:

Machine learning
Dual fuel SI engine
Neat gasoline
Ethanol dual-fuel blends
Artificial neural network
Ensemble LS boosting

ABSTRACT

This research concentrates on the application of machine learning techniques to predict combustion, performance, and emission parameters in a dual-fuel spark ignition (SI) engine powered by neat gasoline and E20 ethanol dual fuel. The goal is to overcome the limitations posed by repeated engine experiments and nonlinear test results. In order to optimise engine parameters, the research seeks to develop efficient machine learning models with high generalizability and employ an optimisation strategy to determine the optimal engine settings. Input for training and evaluating machine learning algorithms, such as Artificial Neural Networks (ANN), and Ensemble LS Boosting was derived from experimental data from a combustion test engine, which includes Neat gasoline, and Ethanol dual fuel blend E20 at various load conditions. The dataset includes engine combustion, performance, and emission indices such as brake thermal efficiency (BTE), Exhaust Gas Temperature (EGT), Hydrocarbons (HC), Carbon monoxide (CO), Carbon dioxide (CO₂), and Nitrogen oxides (NOx), under various operating conditions. Load and brake-specific fuel consumption (BSFC) were training input attributes. Using a comprehensive experimental database of input-output engine parameters, the Artificial Neural Network (ANN) and Ensemble LS Boosting were constructed. The training data points were resampled to generate multiple training datasets for training different models. 50 test samples were used to evaluate the generalisation capability of the machine learning models, while BTE, EGT, CO, CO₂, HC and NOx, were the primary parameters subject to prediction. The optimal machine learning method was determined by comparing R-squared (R^2) values, root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE). Using multiple hyper-parameter tuning iterations, the agreement between actual and predicted values for diverse Ensemble LS Boost algorithms was evaluated. The Ensemble LS Boost model exhibits the maximum level of agreement between predicted and experimental engine parameters across all datasets when compared to the other ANN models. This finding was corroborated by additional research based on test datasets, specifically the test sample interpolation data, which measures generalisation ability. The study also focuses on developing and applying two unique, interactive Simulink models for the Spark Ignition (SI) engine that are tailored for Neat Gasoline and Ethanol E20 test fuels under all loads. The key component of the model-based development technique in MATLAB and Simulink was the incorporation of sophisticated machine learning algorithms, i.e., Ensemble Least-Squares (LS) Boosting, to the model-based development workflow which produced reliable results. Implementing an Ensemble LS Boost machine learning framework is therefore highly recommended as an efficient method for predicting and optimising the combustion, performance, and emission characteristics of dual-fuel gasoline engines utilising Ethanol-based dual-fuel blends.

* Corresponding author. Department of Automobile Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, 603203, Chengalpattu District, Tamil Nadu, India.

E-mail address: vedwingeo@gmail.com (E.G. Varuvel).

Nomenclature			
ANN	Artificial Neural Network	HC	Hydrocarbons
AST	After Spark Timing	HRR	Heat Release Rate
ASTM	American Society of Testing and Materials	LS	Least Squares
BP	Brake Power	MAE	Mean Absolute Error
BSFC	Basic Specific Fuel Consumption	MSE	Mean Square Error
BTE	Brake Thermal Efficiency	NDIR	Non-Dispersive Infrared
CA	Crank Angle	NOx	Oxides of Nitrogen
CO	Carbon Monoxide	E20	Ethanol 20% + Gasoline 80%
CO ₂	Carbon-dioxide	ppm	Parts Per Million
EGT	Exhaust Gas Temperature	RMSE	Root Mean Square Error
		rpm	Revolutions Per Minute
		SI	Spark Ignition

1. Introduction

Air pollution is still a serious problem on a global scale, having persisted over time and posing serious challenges worldwide (Baidya and Borken-Kleefeld, 2009). The transportation industry has become one of the largest emitters of pollutants into the atmosphere. Continuous vehicle population growth is accompanied by a corresponding rise in emissions, which not only worsen air quality but also have an adverse impact on human health (Julia et al., 2015). Additionally, these emissions significantly contribute to the phenomena of global warming, escalating the planet's warming trend (Guttikunda et al., 2019). As a result, it is essential to switch to cleaner, more sustainable energy sources. A complete and quick changeover might be difficult to achieve in a short span (Chu et al., 2017), but significant efforts have been made to reduce automotive emissions through the implementation of strict emission standards in response to the growing concern about air pollution on a worldwide scale. The investigation of alternative fuels that are more environmentally friendly is necessary since the combustion of fossil fuels produces excessive amounts of carbon dioxide and other greenhouse gases (Bae and Kim, 2017). The search for alternative fuels as alternatives to traditional petroleum-based fuels has drawn a lot of interest in reaction to increasingly rigorous worldwide restrictions aimed at limiting the use of fossil fuel-based energy sources (Covert et al., 2016). These alternative fuels present a potential remedy for reducing air pollution and the negative effects it has on the environment. Gasoline blends have been utilised extensively in SI engines for an extended period of time since the early 1970s. This increasing interest can be attributed to the substantial demand for hydrocarbons, which has led to the worldwide exploration and utilization of crude resources. Tetraethyl lead, one of the gasoline additives, is used to increase the octane rating. Nonetheless, this additive is a significant contributor to toxic air pollution and poses a substantial hazard to human health (Entenberg and Menard, 1966). Alcohols are widely acknowledged as a suitable fuel for SI engines, effectively substituting gasoline. This is primarily attributable to their advantageous properties, such as their high oxygen content, latent heat of vaporisation, flame speed, and capacity for complete combustion (Demirbas, 2007). However, there are still obstacles to overcome in the process of mixing alcohol and gasoline at lower temperatures, which may lead to stratification and knocking challenges (Agarwal et al., 2017).

Extensive research has been conducted to investigate possible alcohol blends with gasoline as the pilot fuel and alcohol as the secondary fuel, taking the physical and chemical properties into careful consideration. Previous research studies have examined the combination of lower alcohols, specifically methanol and ethanol, with gasoline extensively, resulting in a substantial number of experiments in this field (Yusri et al., 2017; Manzetti and Andersen, 2015).

Ethanol has been identified as a highly appropriate alcohol for use as a supplement to hydrocarbon fuels. Its remarkable ability to withstand impact and high heat of vaporisation makes it a favoured option. Due to

its low molecular weight, ethanol has become a popular substitute for traditional gasoline octane boosters (Surisetty et al., 2011).

In their study, Elfasakhany et al. (Elfasakhany, 2014) investigated the effects of blending ethanol with gasoline. The researchers found that the blends promote complete combustion, which can be attributed to the presence of oxygen molecules that improve combustion efficiency. Furthermore, the incorporation of ethanol in fuel blends leads to a decrease in atmospheric pollution due to its absence of sulphur content.

A study was conducted by Masum et al. (Masum et al., 2015) on the development of the physicochemical properties of alcohol-gasoline blends with varying ratios of alcohol. It was optimised to obtain three optimal blends. In comparison to gasoline, testing of the ethanol blends in a four-cylinder gasoline engine revealed greater brake torque and thermal efficiency but reduced specific fuel consumption. The optimised blends had lower CO and HC emissions than gasoline, while NOx emissions were higher. The study conducted by Elfasakhany et al. (Elfasakhany, 2016) contrasts the engine performance, combustion, and pollutant emissions of three alcohol-based ternary blended fuels. Experiments were conducted under comparable operating conditions with low blend contents and variable engine speeds. Compared to other blends, the ethanol-methanol-gasoline fuel blends exhibit superior engine performance of volumetric efficiency, torque, and brake power and lower pollutant emissions CO and UHC.

Farooq et al. (Farooq and Kumar, 2023) investigated the performance and emission characteristics of E20 and its equivalent Gasoline-Ethanol-Methanol blends in a spark ignition single-cylinder engine. The E20-equivalent Gasoline-Ethanol-Methanol blends exhibited a brake thermal efficiency comparable to that of E20 and superior performance compared to neat gasoline. The blended fuels produced lower CO and HC emissions compared to neat gasoline, while NOx emissions increased.

Liu et al. (2023) examined the effect of ethanol blending with various excess air ratios on knock, combustion characteristics, power output, thermal efficiency, and emissions in a spark ignition engine with a high compression ratio. In comparison to gasoline, the addition of ethanol reduced knock tendency, enhanced combustion stability, and increased Indicated Mean Effective Pressure and Indicated Thermal Efficiency. Nevertheless, the relationship between ethanol blending ratio and emissions (NOx, HC, and CO) was non-monotonic due to factors such as oxygen properties, latent heat of evaporation, intake-charge cooling effect, and combustion temperature.

Verma et al. (2022) evaluated the effectiveness of ethanol-gasoline blends in a spark-ignition engine. At varying speeds and loads, the operational parameters of the engine, including BSFC, BP, BTE, and BSFC, were evaluated for various ethanol concentrations. Performance parameters including BMEP, BTE, and BSFC improved up to 4200 rpm, while BP improved up to the maximal speed. However, as ethanol concentration increased, engine performance decreased.

Mohammed et al. (2021) investigated the impact of ethanol-gasoline blends on engine performance, fuel economy, and emissions. Using an

ultrasonic chamber, different proportions of ethanol and gasoline (10%, 20%, 30%, and 40%) were combined. Increasing the concentration of ethanol improved power, fuel consumption, and thermal efficiency while reducing hazardous emissions. However, it had a negative effect on volumetric efficiency. Ethanol enhanced engine performance and reduced emissions overall.

Iodice et al. (2018) conducted experiments in which increasing volumes (10%, 20%, and 30%) of gasoline were blended with ethanol derived from grape pomace. On a four-stroke SI engine, the emissions and energy consumption of the blends were evaluated. Higher concentrations of ethanol in the fuel blends resulted in significant reductions in cold emission factors compared to conventional gasoline, indicating enhanced performance.

Thakur et al. (2017) examined improvements in the efficacy of spark-ignition engines powered by gasoline-ethanol blends. Ethanol with lower concentrations (E5, E10, E20) increased engine torque by 2.31–4.16% and brake power by 0.29–4.76%. Brake-specific fuel consumption increased from 5.17 to 56% as the ethanol content of the blends increased. Performance metrics and BSFC were impacted by the compression ratio and speed. Blends of ethanol and gasoline improved brake thermal efficiency by 3.5% for E20, 2.5% for E10, and 6% for E40.

Masum et al. (2013) conducted a review on the influence of ethanol blends on emissions in SI engines. The analysis revealed that ethanol reduced CO and HC emissions, but its effects on NOx emissions were variable. It was found that fuel properties and composition played a significant impact, with higher octane ethanol blends allowing for higher compression ratios without increasing NOx emissions. It has been discovered that ethanol-gasoline blends made cold starts easier and reduced HC, CO, and NOx emissions.

In their investigation, Kumar et al. (2023) systematically evaluated the effects of various ethanol volumetric concentrations on engine performance. According to the findings, the administration of ethanol increased brake thermal efficiency. The maximum value was recorded for a 40% ethanol blend at 398 g/kWh. The concentrations of carbon monoxide and hydrocarbons were reduced by approximately 1.37 and 123 ppm, respectively. In contrast, carbon dioxide and nitrogen oxides increased by approximately 13.82% in volume and 496 ppm, respectively.

Significant scientific priorities include the search for sustainable and ecologically friendly fuels as well as the optimisation of operational parameters. However, experimenting with internal combustion engines can be difficult, costly, and time-consuming. Consequently, the engine and fuel research communities, as well as the industry, are actively pursuing dependable computational methods to meet these needs. Using experimental datasets and machine learning, it is now possible to estimate dual-fuel blend engine parameters. Machine learning models can improve engine combustion, and performance and reduce emissions by predicting test variables which hold good for dual fuel blends as well. Algorithms are capable of analysing real-time sensor data from the engine and adjusting fuel blends accordingly, resulting in enhanced performance and reduced emissions. By taking into account variables such as the percentage of fuel blend, the load fuel consumption, and other relevant parameters, machine learning systems can accurately predict engine emissions from a wide range of fuel blends.

In the experimental study conducted by Yücesu et al. (2007), ethanol-gasoline blends (E10, E20, E40, and E60) were tested in a spark-ignition engine to determine their effect on engine torque and petroleum consumption. Artificial neural networks (ANN) were utilised in the mathematical modelling analysis to predict torque and fuel consumption based on engine parameters. The R^2 values obtained were in the range of 0.99 for torque and fuel consumption in testing data, respectively, the ANN models attained high accuracy. The results indicate that both experimental and modelling approaches were effective for predicting engine performance.

In their study, Yang et al. (2022) compared three machine learning models for predicting Indicated Mean Effective Pressure, Artificial

Neural Network (ANN), Support Vector Regression (SVR), and Random Forest (RF). Using data from a validated one-dimensional computational fluid dynamics model, the machine learning models were trained, validated, and evaluated. Indicated Mean Effective Pressure was accurately predicted by both the SVR and ANN models, with R^2 values close to 1 and RMSE values close to 0.

Using artificial neural network (ANN) modelling, Tekin et al. (Tekin and Saridemir, 2019) intended to predict engine power, torque, and exhaust emissions of a spark ignition engine using gasoline and methanol blends. The ANN model trained with the back-propagation algorithm performed exceptionally well. Comparing predictions to experimental results, the ANN model demonstrated high accuracy, with correlation coefficients spanning from 0.9976 for hydrocarbon emissions to 0.9992 for exhaust temperature and brake-specific fuel consumption.

Liu et al. (2021) evaluated ANN, RF, SVM, and GBT predictions of exhaust temperature in a one-dimensional practicable computational fluid dynamics model. As model inputs, engine speed, equivalence ratio, and ignition timing were considered. Simulations of neural networks predicted the exhaust temperature more accurately than physical models. The algorithm for an Artificial Neural Network required hyperparameter adjustment.

Josephin et al. (2018) used rubber seed oil-ethanol dual-fuel engine test data in their research. An ANN with feed forward-back propagation predicted compression ignition engine performance and exhaust pollutants. BTE, BSEC, HC, CO, NOx, and smog were the emission parameters investigated. ANN output values complemented experimental data effectively. Engine emission regression models were evaluated using R^2 , MAE, MSE, and RMSE. Efficiency analysis determined the engine's optimum operating parameters.

Dey et al. (2020) investigated the application of an artificial neural network model for predicting the performance and emissions of a single-cylinder compression ignition engine operating on diesel-palm biodiesel-ethanol dual fuel blends. The ANN model was trained with the Levenberg-Marquardt backpropagation algorithm and achieves high accuracy with correlation coefficients (R^2) ranging from 0.99329% to 0.99875% and minimal mean square errors (MSE) between 2.5% and 4.5%. To optimise engine parameters, experimental and predicted data were compared using a fuzzy interface system.

Barboza et al. (2022) investigated on blending 10% ethanol with gasoline (E10) and emulsifying it with 1.5% H₂O₂. The combination increased BTE by 4.8% compared to neat gasoline and reduced CO, HC, and NOx emissions. Ethanol concentration, H₂O₂ concentration, engine speeds, and BTE, CO, HC, and NOx emissions were modelled using experimental data. Multi-objective particle swarm optimisation was used to optimise ANN models. The ideal engine speed was 2000 rpm, ethanol blending was 4%–6%, and H₂O₂ emulsification was identified to be 1.5%.

Bhowmik et al. (2017) examined the impact of adding 5% ethanol to diesel-kerosene blends in a single-cylinder indirect injection compression ignition engine. The addition of ethanol improved BTE, and BSEC also reduced NOx, HC, and CO emissions. Experimental data were used to develop an accurate artificial neural network (ANN) model, with a (3-6-5) topology and Levenberg-Marquardt backpropagation algorithm. The ANN model achieved a low MSE, MAPE, and high R^2 . The study also explored fuzzy-based analysis to optimise the ANN model's network topology.

Karunamurthy et al. (2023) carried out an in-depth study of the machine-learning techniques used to predict the performance and emission parameters of IC engines. Artificial neural networks were among the most commonly employed algorithms. The study highlighted the predictive capacity of machine learning and deep learning as decision-making aids for internal combustion engine parameters. It was also noted that artificial neural networks have limitations, such as unexplained network behaviour and an accuracy dependence on the number of hidden layers, which increases computational complexity.

Pla et al. (2023) used Singular Value Decomposition to extract combustion characteristics from the knock signal spectrogram in their study. In spark ignition engine tests, they established a correlation between these characteristics and an artificial neural network. In terms of predicting combustion phasing parameters such as spark advance, fuel mass, and engine speed, the performance of the ANN models was superior to that of engine control models. In addition, the introduction of unlearned circumstances resulted in a remarkable MAE.

AlNazr et al. (2023) developed an ANN-based machine-learning model that predicts the density and viscosity of petroleum fuels containing oxygenated chemical classes, such as alcohols, using the molecular structure of compounds as inputs. The model was trained using 164 pure compounds and 14 blends with known compositions. With density regression coefficients of 0.99 and viscosity regression coefficients of 0.98, the neural network model made accurate predictions.

Fagundez et al. (2020) investigated the use of artificial neural networks and particle swarm optimisation in conjunction to predict SI engine operation with negative valve overlap. The combined method increased the precision and generalizability of emission estimates and indicated efficiency. The best model outperformed experimental data, indicating optimisation potential for engine efficiency. The strategy reasonably decreased NOx emissions while maintaining comparable CO and HC emissions.

Ricci et al. (2023) used an ANN to predict internal combustion engine NOx emissions. A Nonlinear Autoregressive with Exogenous Input architecture using in-cylinder pressure data and flame front images replicated the experimental NOx concentration trend with a 2% error rate. RMSE and Coefficient of variance below the threshold showed that the ANN performed consistently under all engine operating conditions. The prediction model was most sensitive to Indicated Mean Effective Pressure.

Kodavasal et al. (2018) used machine learning to study spark-ignition gasoline engine cycle-to-cycle variation. Flame topology and pre-ignition flow-field metrics were calculated using three-dimensional data from test cycles. Then, peak cylinder pressure was correlated with these metrics using a Random Forest (RF) machine learning technique, which effectively captured metrics-peak cylinder pressure interactions. This study showed machine learning models can learn implicitly complex relationships between pre-ignition flow fields, flame shapes, and cycle outcomes.

The use of ensemble learning to predict CO₂, NOx, smoke, BTE, and HC emissions from engines was investigated by Bai et al. (Bai, 2023; Bai et al., 2023). Several ensemble learning techniques, including XGBoost, LGBM, CatBoost, and Random Forest, were utilised. The superior accuracy of CatBoost was minimised by the XGBoost, RF, and LightGBM models.

Mohan et al. (Mohan and Badra, 2023) developed a machine learning-based framework for the automated optimisation of internal combustion engine applications. A SuperLearner model, ensemble-based learners, optimisation algorithms, and an active learning strategy were included in the framework. To solve the multi-objective problem, it was simplified into a single-objective problem, resulting in a high merit score. Despite employing 96% fewer data, the merit value outperformed the highest values derived from both the reduced and original datasets. The study demonstrated the capability of the framework to generate optimal solutions with fewer experimental or simulation trials and fewer data points.

The effectiveness and environmental impact of a compression ratio diesel engine running on waste-derived biodiesel and biogas were assessed by Sharma et al. (Sharma and Sahoo, 2022) using Boosted Regression Trees (BRT) method. Compression ratio, load, and fuel injection duration were taken into consideration as model inputs. The brake fuel ratio, brake thermal efficiency, nitrogen oxides, carbon monoxide, and hydroxyl radicals were among the important factors that the BRT model was used to determine based on experimental data. The Enhanced Regression Tree model produced correlations between 0.9947

and 0.9997. BRT outperformed Artificial Neural Networks in all statistical parameters when compared under the same operational conditions.

This research centres on the application of machine learning techniques to predict the characteristics of a gasoline engine operating in dual-fuel mode with ethanol. It compares two prediction models in order to determine the most accurate model for predicting combustion, performance, and emission parameters in a dual-fuel gasoline engine using Ethanol-gasoline blends. Using Artificial Neural Networks and Ensemble LS Boost techniques, the paper provides predictions for the considered engine parameters. This comparison establishes which machine learning strategy is preferable. The study concludes with a performance and error-based evaluation of the optimal machine learning model. The results demonstrate that machine learning has the potential to revolutionise the field of internal combustion engines and contribute to the development of more efficient, dependable, and environmentally friendly engines fuelled by gasoline-ethanol blends.

2. Experimental setup and procedure

The experiments were conducted on a two-cylinder, four-stroke petrol engine, as depicted in Fig. 1. The engine had MPFI injection capability and electronic ignition and had a 624-cc displacement, two valves per cylinder, and a 12.54 kW output as depicted in Table 1. The engine was electronically fuel-injected and controlled by a closed-loop system utilising a lambda sensor. An eddy current dynamometer was shaft-coupled to the engine, allowing for the engine to be loaded. For data collection, the test configuration included a AVL high-speed data acquisition and storage system. Using the non-dispersive infrared (NDIR) principle, a gas analyzer from AVL was used to measure the engine's gaseous emissions. The primary focus was on HC, CO, CO₂, and NOx emissions. The primary sensor data collected consisted of cylinder pressure readings from a pressure sensor, crank angle position readings from an encoder, and temperature readings from a thermocouple. The configuration was used to measure all data for 100 consecutive cycles, and the collected data were consolidated and refined to eliminate nonlinearity in each cycle. Table 2 displays the uncertainty range of instruments and apparatus used. Constantly restricting the engine's speed to 2500 rpm, experimental trials were conducted. The engine was loaded from no load to maximum load for testing neat gasoline and E20 test fuels. The test fuels were subjected to experiments under steady-state conditions, without modifications to the engine's physical parameters. The engine was allowed to operate for 15 min prior to each test, and readings were recorded for each test conducted under varying loading conditions.

3. Error analysis and uncertainty

An uncertainty analysis was conducted to ensure the accuracy and robustness of the experimental results. The uncertainty analysis involved estimating potential errors in both the measured and calculated test outcomes, considering various factors such as instrument precision, human factors, and environmental influences. To minimize uncertainties, readings were taken after the engine had reached steady-state operation at each test load condition, and thrice the readings were recorded for each measurement related to the test fuels.

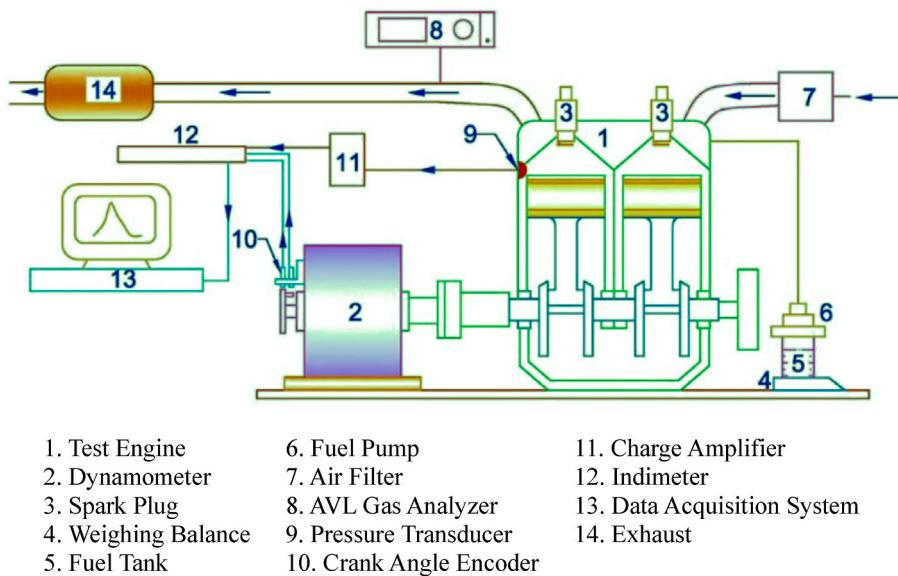
The uncertainty of the measured engine parameters is estimated using the Gaussian distribution method with a confidence interval of $\pm 2\sigma$ (where 95.45% of the measured data lies within this range from the mean). The uncertainty is determined using the following equation:

Uncertainty of any measured parameter

$$\Delta x_i = \frac{2\sigma_i}{x_i} \times 100 \quad (\text{Eq. 1})$$

Experiments are conducted to obtain the mean (\bar{x}_i) and standard deviation (σ_i) of any measured parameter (x_i) for a number of readings.

The sample standard deviation (σ) of three replicates (n) is calculated

**Fig. 1.** Schematic diagram of dual fuel SI engine experimental test setup.**Table 1**
Test engine specifications.

SI. No.	Specification	Value
1	Engine/BP	4 Stroke _SI _25 KW
2	Rated Speed	2500 rpm
3	No of Cylinders	2
4	Displacement	624 cc
5	Engine Bore Diameter	73.5 mm
6	Engine Stoke length	73.5 mm
7	Compression ratio	9.5:1
8	Spark Timing (MBT)	23° CA bTDC
9	Cooling System	Water Cooled
10	Fuel Injection Pressure	5 Bar
11	Injector	3 Holed
12	Valve Events	IVO 15 Degrees bTDC IVC 20 Degrees aTDC EVO 27 Degrees bTDC EVC 10 Degrees aTDC

given below:

Table 2
Uncertainty of test variables.

Variable	Technique	Accuracy	Uncertainty %
Speed	Pickup - Magnetic	±10 rpm	±1
Load	Load Cell - Strain Gauge	±0.1 Kg	±0.2
Time	Stop Watch	±01 s	±0.2
Crank Angle	Pickup - Magnetic	±1°	±0.2
Cylinder Pressure	Pickup - Piezoelectric	±0.5 bar	±1
Manometer Pressure	Balancing Column	±1 mm	±1
EGT	Thermocouple - K type	±1 °C	±0.15
NOx	Electrochemical Principle	±10 ppm	±1
CO	NDIR Principle	±0.02%	±0.2
HC	NDIR Principle	±20 ppm	±0.2
CO ₂	NDIR Principle	±0.03%	±0.15

as

$$\sigma = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (\text{Eq. 2})$$

To obtain realistic error limits for the computed quantity, the root-sum-square method is employed, and the magnitude of the error is determined by the following Eq. (3).

$$\Delta R = \sqrt{\left(\frac{\partial R}{\partial x_1} \Delta x_1\right)^2 + \left(\frac{\partial R}{\partial x_2} \Delta x_2\right)^2 + \dots + \left(\frac{\partial R}{\partial x_n} \Delta x_n\right)^2} \quad (\text{Eq. 3})$$

The overall uncertainty for the engine parameters can be calculated using the Root Sum Square method, which is represented by Eq. (4).

The estimated uncertainty values for each engine parameter are

$$U = \pm 2.5461\%$$

Thus, the overall uncertainty for the given parameters is approximately ±2.55%.

4. Experimental trials and dataset generation

4.1. Brake thermal efficiency (BTE)

The brake thermal efficiency of the Ethanol dual-fuel gasoline engine which represents the ability to convert the chemical energy of the fuel into useful mechanical output, fluctuates frequently based on the type of fuel and operating conditions. Notably, BTE is essential for determining the quantity of fuel required to maintain constant engine speed when

$$U = \sqrt{U_{BP} + U_{BSFC} + U_{BTE} + U_{HC} + U_{CO} + U_{BP} + U_{NOx} + U_{CO2} + U_{EGT} + U_{RPM} + U_{P-incy}} \quad (\text{Eq. 4})$$

$$= \sqrt{(1.08)^2 + (0.85)^2 + (0.88)^2 + (0.2)^2 + (0.2)^2 + (1)^2 + (0.25)^2 + (0.65)^2 + (1.12)^2 + (1)^2}$$

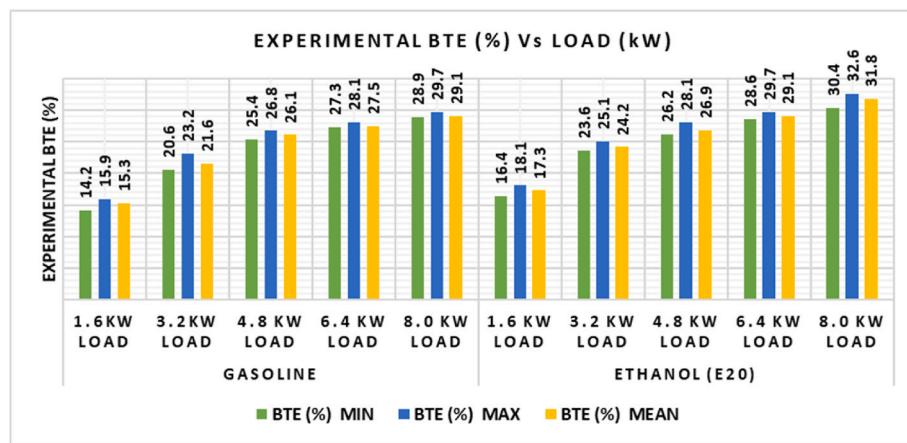


Fig. 2. Experimental BTE (%) at various loads across datasets.

adjusting the fuel supply. Fig. 2 indicates that the BTE of both the Gasoline and E20 blends tends to increase as the load increases. This is primarily attributable to a reduction in friction and other losses under heavy loads. During tests conducted under maximum load at 2500 rpm, the BTE ranged from 28.9% to 29.1% for gasoline and from 30.4% to 31.8% for the E20 blend, depending on the load. The addition of ethanol to gasoline significantly increased BTE at all load levels. This enhancement is attributable to a number of factors, including accelerated mixing, enhanced atomization, superior vapour quality, and the presence of oxygen molecules, which promote combustion. In addition, the higher calorific value of ethanol compared to gasoline significantly reduced fuel consumption, thereby enhancing BTE. In addition, the E20 blend demonstrated a higher rate of heat release than gasoline, which contributed to its superior BTE as shown in Fig. 2.

4.2. Carbon monoxide (CO)

Carbon Monoxide emissions are one of the by-products of incomplete combustion and are frequently present in biofuel-gasoline mixtures due to inefficient combustion. As a result of enhanced combustion characteristics and increased in-cylinder temperatures, both of which promote complete combustion, it is generally observed that CO emissions decrease as load increases. As depicted in Fig. 3., the CO emissions of the E20 ethanol-gasoline blend were consistently lower than those of neat gasoline. This decrease is predominantly attributable to the presence of oxygen molecules in fuel blends, which significantly boost CO oxidation within the combustion chamber. Typically, the highest CO emissions are

observed when the engine is operating in a mode with a rich mixture under initial load conditions. Inadequate mixing, local rich zones, and incomplete combustion led to CO emissions. CO emissions initially decreased for both petrol and E20, but after 6.4 kW of load, they began to increase, revealing an intriguing trend. At maximum load, the CO emissions were in the range of 0.74%–0.79% for neat gasoline and 0.81%–0.87% for E20. This trend is clarified by the fact that at reduced loads, fuel blends have sufficient time for thorough mixing and complete combustion. However, as the load increases, the time available for proper mixing decreases, and the likelihood of incomplete combustion rises, leading to an increase in CO emissions. Due to its higher oxygen content, which diminishes CO levels in the exhaust.

4.3. Carbon dioxide (CO₂)

Fig. 4 Illustrates an upward trend of increasing CO₂ emissions for the ethanol blend E20 and neat gasoline across a range of load conditions, from low to high loads. At maximal load, the CO₂ emissions for gasoline were in the range of 8.6%–8.8% and for the E20 blend they were 7%–7.2%. Typically, the amount of CO₂ emissions is determined by the carbon content of the fuel and the rate of fuel consumption. Intriguingly, the lowest CO₂ emissions were observed with the ethanol blend, which was considerably less than neat gasoline. This decrease in CO₂ emissions is attributable to the lower carbon content of ethanol blend E20.

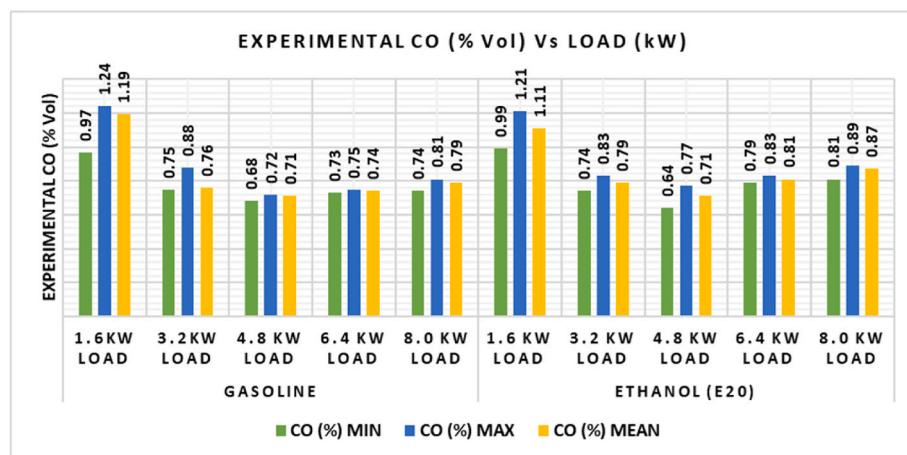


Fig. 3. Experimental CO (%) at various loads across datasets.

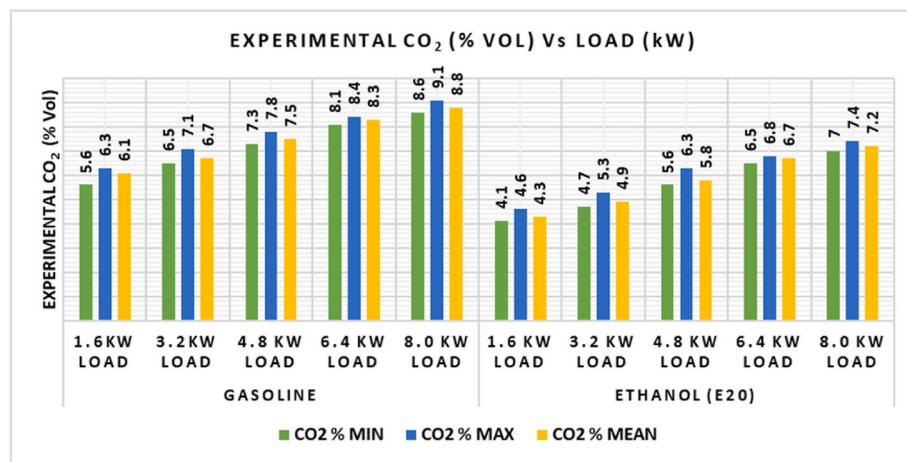
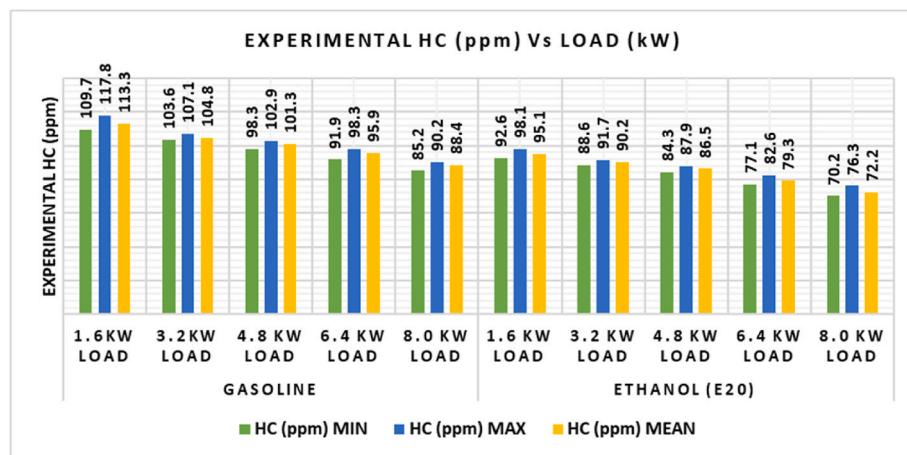
Fig. 4. Experimental CO₂ (%) at various loads across datasets.

Fig. 5. Experimental HC (ppm) at various loads across datasets.

4.4. Hydrocarbons (HC)

If carbon-based fuels do not completely combust, unburned hydrocarbons (HC) are released into the atmosphere. Due to enhanced combustion conditions, such as increased cylinder temperature, these HC emissions tend to be highest at low loads for all tested fuel types and decrease with increasing loads. Fig. 5 Depicts the variation in HC

emissions across a range of load conditions, from 1.2 kW to 8 kW, for all tested fuels. Notable is the fact that the maximum levels of HC emissions for both gasoline and the E20 blend were recorded at 1.6 kW load, ranging between 109.6 ppm to 113.3 ppm and 92.6 ppm–95.1 ppm respectively. This is due to the concentration of the air-fuel mixture at the engine's intake and the nature of fuel premixing at lower temperatures. Due to enhanced combustion processes, which effectively oxidise

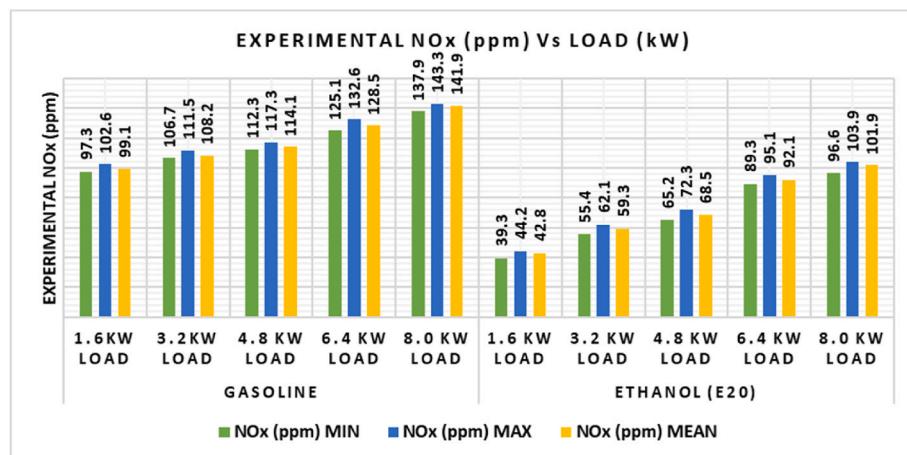


Fig. 6. Experimental NOx (ppm) at various loads across datasets.

unburned hydrocarbons, ethanol blends emit lesser hydrocarbons at a lower loading than gasoline. In addition, the oxygen content of the E20 blend facilitates faster combustion rates, resulting in reduced HC emissions compared to neat gasoline. As load increases, the concentration of the air-fuel mixture decreases HC emissions and increases combustion temperatures, which is especially evident at maximum load. The trend indicates figures ranging between 85.2 ppm and 88.4 ppm for petrol and 70.2 ppm–72.2 ppm for E20, respectively. Due to its higher latent heat of vaporisation and higher combustion rate, it can be inferred that the Ethanol E20 blend performs better than neat gasoline.

4.5. Oxides of nitrogen NO_x

NO_x emissions are produced by the reaction of nitrogen and oxygen molecules in a combustion chamber's high-temperature environment. These emissions are affected by the temperature and duration of fuel's presence in the combustion chamber. NO_x emissions increased uniformly for all tested fuels as the engine load increased. This increase in NO_x emissions is most likely due to improved combustion quality and increased cylinder temperatures under higher loads. Fig. 6 illustrates the variation in NO_x emissions for neat gasoline and E20 ethanol blends at various loads. NO_x emissions are primarily influenced by factors such as excess oxygen and cylinder temperature. At elevated temperatures, nitrogen bonds break down, allowing monatomic nitrogen to interact with oxygen molecules. In addition, the reaction rates of nitrogen molecules during bonding at elevated temperatures have a significant impact on NO_x formation. In contrast, nitrogen remains in diatomic form at lower temperatures. At maximum load, NO_x emissions were measured in the range of 137.9 ppm–141.9 ppm for gasoline and 96.6 ppm–101.9 ppm for the E20 blend. Among the two test fuels, neat gasoline test fuel had the highest NO_x concentration and the ethanol blend exhibited the lowest NO_x emissions. Despite the excess oxygen in the fuel, the higher latent heat of vaporisation of ethanol decreases the in-cylinder temperature, preventing NO_x formation.

4.6. Exhaust gas temperature (EGT)

In the realm of dual-fuel gasoline engines, the assessment of performance, efficiency, and durability is dependent upon the measurement of Exhaust Gas Temperature. Serving as an effective indicator of combustion efficiency, EGT is capable of detecting potential complications such as imbalances within the air-fuel mixture. Elevated EGT potentially denotes combustion inefficiencies, posing a risk of damage to engine components due to accumulated thermal stress. Furthermore, EGT plays an influential role in emission control mechanisms, notably catalytic converters, which necessitate specific temperatures for their optimal operation. Consequently, diligent monitoring of EGT is essential for

optimising engine performance, enhancing engine longevity, and efficiently managing emission control. Fig. 7 Provides a comparative analysis of EGT for the E20 blend and conventional gasoline test fuels under various load conditions. The data demonstrate a positive correlation between EGT and load for both test fuels. At no load to maximum load condition, the EGTs for gasoline and the E20 blend were recorded to be in the range of 648.2 °C–652.6 °C and 618.3 °C–624.4 °C respectively. Blending gasoline with ethanol appears to induce a reduction in EGT, attributable to improved combustion and an increased proportion of alcohol in the fuel blend. Ethanol's superior latent heat of vaporisation produces a cooling effect within the combustion chamber, primarily towards the peak of the induction phase. This subsequently impacts the in-cylinder temperature during the terminal stage of compression, resulting in a notable reduction in EGT. An additional contributing factor to the reduction in EGT is the shortened duration of combustion as a result of ethanol's superior laminar flame speed. This accelerated combustion process further lessens the exhaust gas temperature.

4.7. Machine learning procedure

Machine learning is a subfield of artificial intelligence that focuses on the development of statistical models and algorithms that enable computers to learn and interpret data. These systems are intended to foresee or make decisions without being explicitly programmed to do so. This study focuses predominantly on the application of supervised learning techniques to a dual-fuel SI engine supplied with neat gasoline and E20 ethanol blends. The application of machine learning algorithms for predicting the properties of dual-fuel internal combustion engines is a new and promising area of research with potential applications in engine design, control, and optimisation. Using machine learning, it is possible to predict various aspects of engines, such as their combustion properties, efficacy, and emission levels. These predictions can aid in the development of more efficient and effective engines for a variety of applications while assuring regulatory compliance. The machine learning methodology adopted in this work is depicted visually in Fig. 8.

The engine's combustion, performance, and emissions, namely BSFC, BTE, HC, CO, NO_x, and EGT, were measured under controlled conditions and diverse operational conditions, including engine load, and fuels. Data pre-processing involved the removal of outliers, missing values, and irrelevant data, thereby preparing the data for feature extraction. Fuel type, Load, and brake-specific fuel consumption were chosen as features for machine learning model training. Due to the magnitude of the dual fuel blend dataset, the accuracy requirements, and the complex relationship of engine parameters, Artificial Neural Network and Ensemble LS Boost were chosen as the primary machine learning methods for investigation.

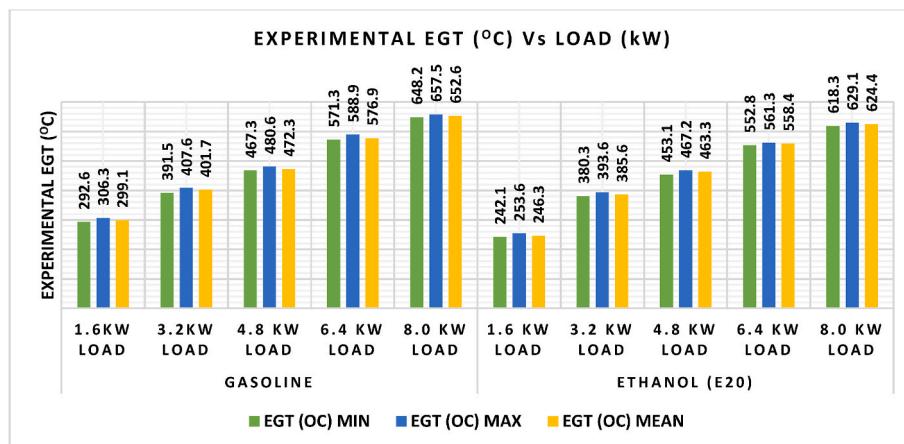


Fig. 7. Experimental EGT (%) at various loads across datasets.

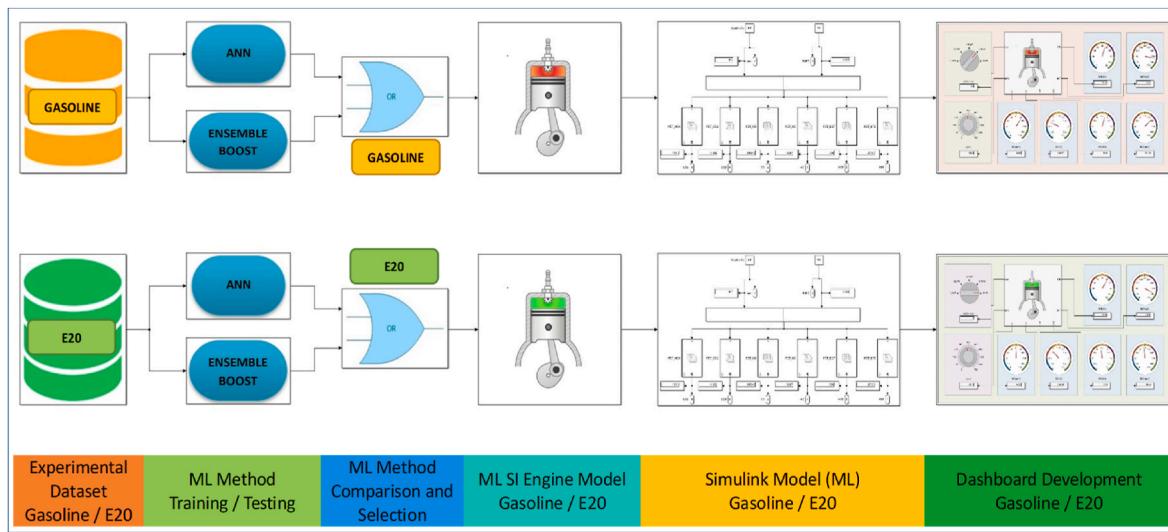


Fig. 8. Outline of the machine learning method for neat gasoline and E20 test fuels.

4.8. Artificial neural networks

A neural network, which is frequently compared to the human brain, is a type of parallel processor with a natural tendency for storing experimental knowledge through a learning process. Like the human brain, it acquires knowledge through a learning process and stores it as synaptic weights or inter-neuron connection strengths. Using back-propagation, an Artificial Neural Network modifies these weights and then stores the modifications as values. The number of neurons in the hidden layers can be affected by the problem's complexity and the provided dataset which caused the current study to choose two hidden layers.

The network design begins with the selection of suitable activation functions. Then, network parameters include setting up of the weights and biases. The subsequent steps entail defining the error target, the maximum number of epochs, and other training method-specific parameters. After the network has been initialised and trained, the model's output was evaluated by simulating the neural network's output using the evaluated input data and comparing it to the measured outputs. Validation, the final verification procedure, requires independent data. This study employed a multi-layer perception network to facilitate non-

linear mapping between input and output variables. The ANN models for all the engine parameters were trained using the back-propagation method, which involves randomly initializing network weights and biases at the beginning of the training phase and employing the supervised training method and the gradient descent rule for error reduction.

In the current experiment, two input parameters engine load and brake-specific fuel consumption (BSFC) and six output parameters (BTE, EGT, CO, CO₂, HC, and NOx) were used for prediction. The dataset was divided into a training set of 70% and a test set of 30%. There were two neurons in the input layer and six neurons in the output layer of the network architecture. The number of concealed layers and neurons in each layer can be modified based on the problem's and dataset's complexity. A network with two hidden layers, each containing 10 neurons, was finalized for this investigation. During the training procedure, the number of neurons was determined based on the correlation coefficient and error metrics. Fig. 9 Shows the graphical representation of the ANN's architecture utilised in this research work.

4.9. Ensemble LS boosting

The Ensemble LS Boost algorithm is an approach to machine learning

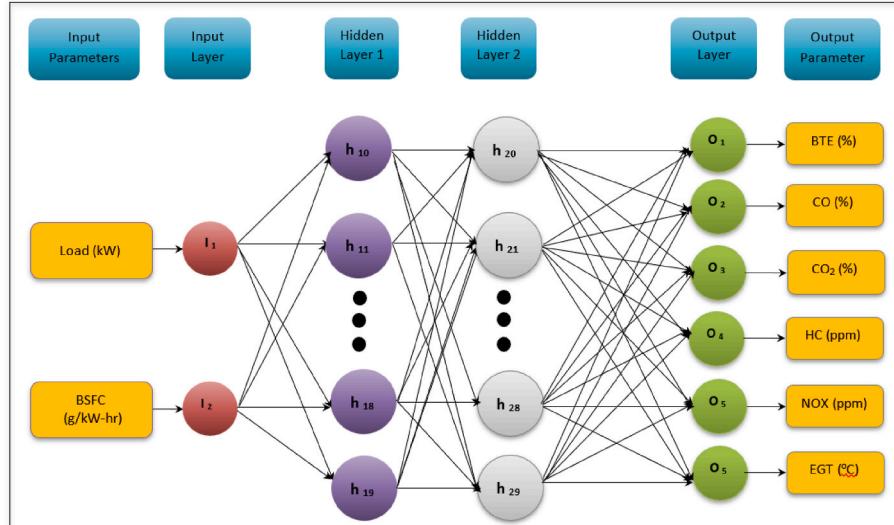


Fig. 9. Architecture of artificial neural method (ANN) for dual fuel SI engine.

intended for regression tasks. Its primary objective is to reduce the mean-square error (MSE) of the model's predictions by creating a set of weak learners, which are typically decision trees or tiny decision trees. To generate an initial prediction, LS Boost initiates a simple model, which typically predicts the mean or median of the target variable. This marks the commencement of a recursive procedure consisting of the following crucial steps. The initial phase is to calculate residuals. These are the differences between observed responses and aggregated predictions from all trained models. The subsequent phase is to fit a new weak learner to these residuals. This weak learner, which is typically a decision tree, is trained to predict residuals rather than the initial target variable. The objective is for this new weak learner to concentrate on instances that present the greatest difficulty for the existing ensemble model, specifically instances with large residuals. The final step entails incorporating the predictions of this new weak learner into the ensemble model. Consequently, the ensemble model's prediction is a synthesis of the predictions of all weak learners trained up to that point. This process is repeated a predetermined number of times or until the residuals can no longer be reduced significantly. To prevent overfitting, measures such as limiting the number of iterations or the complexity of weak learners were taken.

Fig. 10 Shows the graphical representation of the Ensemble LS Boost architecture utilised in this research work. The following steps illustrate the Ensemble LS Boost learning procedure. Both the E20 and neat gasoline datasets were divided into two distinct subsets, a training set and a test set. The test set was used to evaluate the performance of the final model, whereas the training set was used to train the ensemble models. The training procedure was the same for all samples, and the weights were set accordingly. Each boosting cycle involved the construction of a new model using training data. The sample weights were adjusted to place a greater emphasis on instances incorrectly categorised by the previous models. The current model's weight was then determined based on its performance on the training data. The models were then combined, typically by averaging their respective predictions. For validation, the k-fold cross-validation method with a k-value of 5 was utilised. The test data set was then used to evaluate the efficacy of the boosted ensemble model. When the Ensemble LS Boost was executed and the ensemble adjusts each new learner to the corresponding equation.

$$f(x) = y_n - \eta f(x_n) \dots \quad (\text{Eq. 5})$$

In this equation, x_n represents the number of previous learners, y_n represents the observed outcome, $f(x_n)$ represents the cumulative prediction for observation x_n based on all previously generated weak learners, and η represents the learning velocity. The hyperparameters were optimised, which is a crucial phase in determining the optimal model architecture. A broad range of parameter values that affect the learning process and performance of the model was examined to regulate the training procedure. **Table 3** summarises the specific values and

Table 3

Machine learning algorithms and the selected hyperparameters.

Artificial Neural Network	Ensemble LS Boost
Fully Connected Layers	Minimum leaf size
Layer Size	Number of learners
Activation Function	Learning rate
Iteration Limit	Predictors to sample
Regularization Strength	Optimizer and Iterations

their impact on the model.

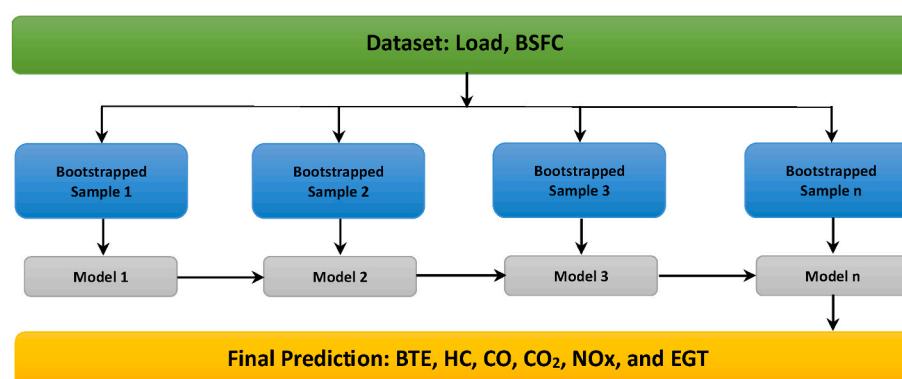
After pre-processing, selection of the above-mentioned machine learning methods ANN and Ensemble LS Boost, training, validation, and testing datasets, the accuracy was measured using metrics such as R-squared (R^2) value, Root Mean Square Error (RMSE), Mean Square Error (MSE), and Mean Absolute Error (MAE). Optimisation of hyperparameters, a crucial stage in discovering the optimal model architecture, was performed. Controlling the training process, a comprehensive range of parameter values affecting learning and model performance were evaluated. **Table 3** shows the hyperparameters utilised in the Ensemble LS Boost method. The Bayesian optimisation strategy was chosen over other refining methods due to its high prediction accuracy for the Ensemble LS Boost methods. This essential phase was executed despite being computationally expensive and time-consuming in order to optimise model performance. The machine learning models were refined, validated, and then evaluated on datasets that were not utilised during training. Using a k-fold validation technique, the data were divided into roughly equal subsets and the models for all the engine parameters were validated. After attaining the desired level of precision, the model was used to predict engine characteristics under various operational conditions.

Statistical metrics such as the Coefficient of Determination (R^2), Root Mean Square Error (RMSE), Mean Square Error (MSE), and Mean Absolute Error (MAE) were used to evaluate the model's performance.

R-squared value (R^2): In regression analysis, the R-squared value represents the proportion of the dependent variable's variance that can be explained by the independent variables (Bai, 2023). R-squared ranges from 0 to 1, and an ideal model is one that explains 100% of the variance in response data around the mean. The R-square value is determined by a specific formula Eq. (6).

$$R^2 = \frac{\text{Variance Explained by the Model}}{\text{Total Variance}} \quad (\text{Eq. 6})$$

MSE quantifies the average of the squared differences between actual and predicted values (Bai, 2023). Since differences are squared prior to being averaged, the MSE is never negative. An error-free, ideal model would produce an MSE of zero. The MSE calculation formula is provided in Eq. (7).

**Fig. 10.** Architecture of ensemble LS boosting for dual fuel SI engine.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2 \quad (\text{Eq. 7})$$

RMSE is a metric that calculates the square root of the average of squared differences between actual and predicted values. RMSE quantifies the dispersion of data points by measuring the distance between the best-fit line and the data points. The RMSE is computed with the aid of a specified equation Eq. (8).

$$RSME = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2} \quad (\text{Eq. 8})$$

Mean Absolute Error (MAE): Mean Absolute Error (MAE) computes the average absolute differences between actual and predicted values (Bai, 2023). It provides the average error magnitude across the entire dataset, regardless of direction. The MAE equation is shown in Eq. (9).

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}| \quad (\text{Eq. 9})$$

This study utilised empirical data collected from a spark-ignition engine powered by both gasoline and a gasoline-Ethanol blend E20 while operating in a dual-fuel mode at maximal engine load and 2500 rpm. Using two distinct techniques, Artificial Neural Networks and Ensemble LS Boost, the experimental data was utilised to develop machine learning models. Two input variables, Engine Load and Brake-Specific Fuel Consumption (BSFC), and six output variables, Brake Thermal Efficiency (BTE), Hydrocarbons (HC), Carbon Monoxide (CO), Carbon Dioxide (CO₂), Nitrogen Oxides (NOx), and Exhaust Gas Temperature (EGT), were included in the model creation process. The development of these predictive machine learning models required lab-based engine experiments; 70% of the total acquired data was designated for training the models, while the remaining data were used for

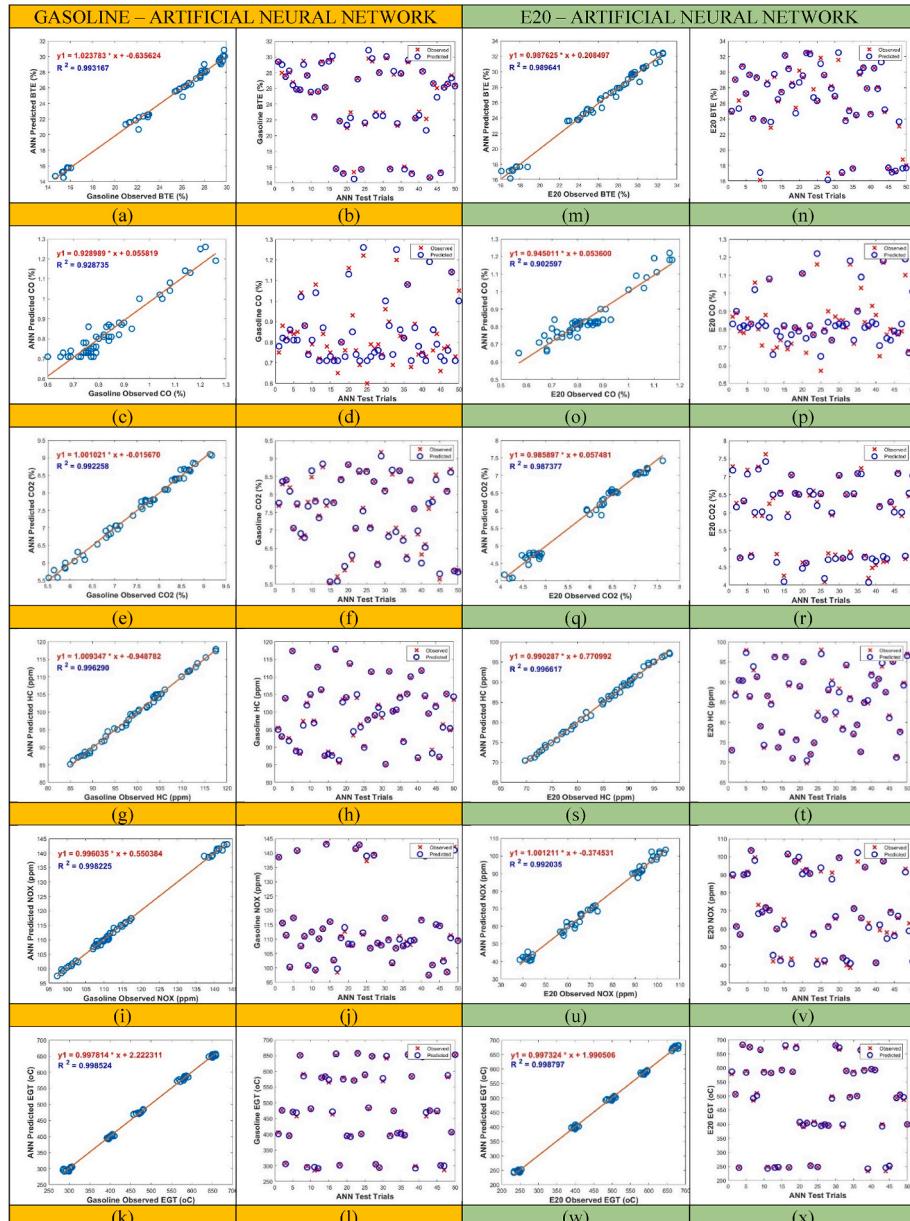


Fig. 11. Gasoline ANN model. (a,b,c,d,e,f) Observed vs ANN predicted BTE, CO, CO₂, HC, NOx and EGT (g,h,i,j,k,l) ANN prediction for all data points of BTE, CO, CO₂, HC, NOx and EGT. E20 ANN model. (m,n,o,p,q,r) Observed vs ANN predicted BTE, CO, CO₂, HC, NOx and EGT (s,t,u,v,w) ANN prediction for all data points of BTE, CO, CO₂, HC, NOx and EGT.

testing. In the concluding phase of the study, a comprehensive comparative analysis was conducted using statistical metrics and graphical representations to evaluate the performance and predictive capabilities of the machine learning models. Further two Simulink models were developed and tested for dual-fuel SI engine for neat gasoline and E20 test fuels.

5. Results and discussion

5.1. Artificial neural network-based prediction model (ANN)

Analysing the combustion, performance, and emission characteristics of a dual-fuel Spark Ignition (SI) engine, a predictive model was constructed using Artificial Neural Network (ANN) principles. This model was used to predict the performance, combustion and emission levels for numerous input scenarios. The ANN chosen possesses two

hidden layers having 10 neurons each. The input layer has two neurons and the output layer has six neurons.

Fig. 11 depicts a comparison between the predicted model values and the actual observed data for BTE, CO, CO₂, HC, NOx, and EGT. This comparison revealed a strong correlation between the model-predicted BTE values and the experimentally determined BTE values. On the graph, the majority of data points were evenly distributed along the line of best fit, validating the model's predictive ability. **Fig. 11** also depicts the graphical representation of the predicted values of the model for each test run.

Fig. 13 displays the outcomes of the ANN algorithm's evaluation of the relationships between key engine parameters BTE, CO, CO₂, HC, NOx, and EGT. The R-squared R² values of neat gasoline for the aforementioned gasoline engine parameters were 0.9931, 0.9287, 0.9922, 0.9962, 0.9982, and 0.9985. For the E20 blend, the ANN method returned R² values of 0.9896, 0.9025, 0.9873, 0.9966, 0.9920, and

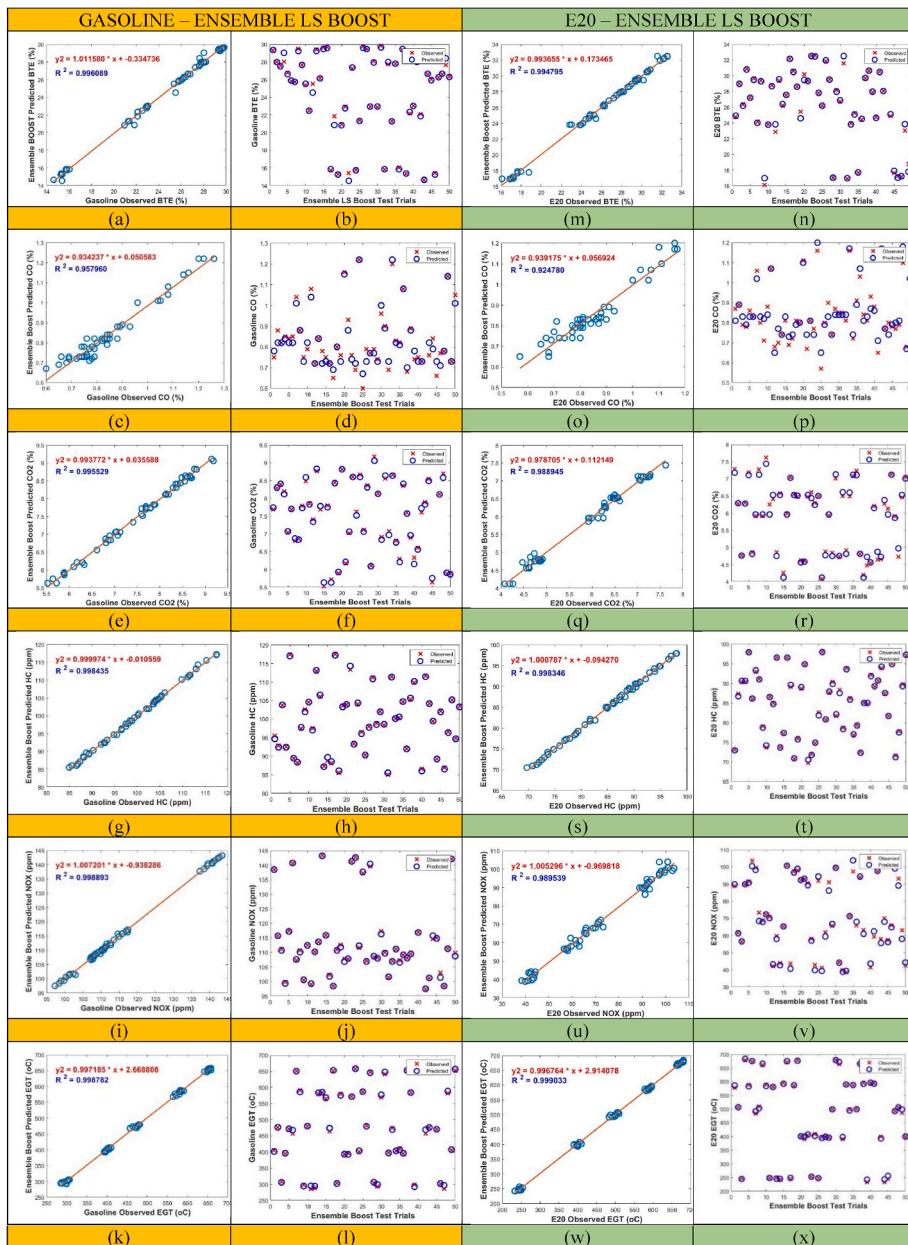


Fig. 12. Gasoline Ensemble LS Boost Model. (a,b,c,d,e,f) Observed vs Ensemble LS Boost predicted BTE, CO, CO₂, HC, NOx and EGT (g,h,i,j,k,l) Ensemble LS Boost prediction for all data points of BTE, CO, CO₂, HC, NOx and EGT. **E20 Ensemble LS Boost Model.** (m,n,o,p,q,r) Observed vs Ensemble LS Boost predicted BTE, CO, CO₂, HC, NOx and EGT (s,t,u,v,w) Ensemble LS Boost prediction for all data points of BTE, CO, CO₂, HC, NOx and EGT.

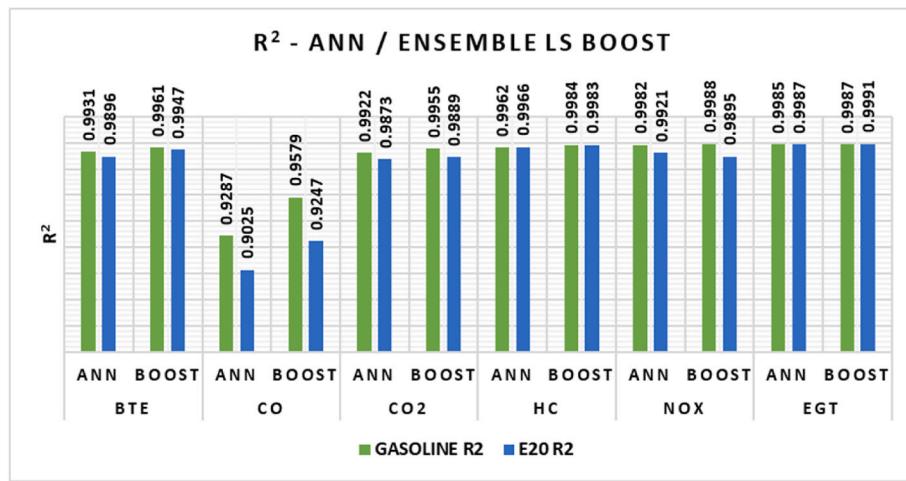
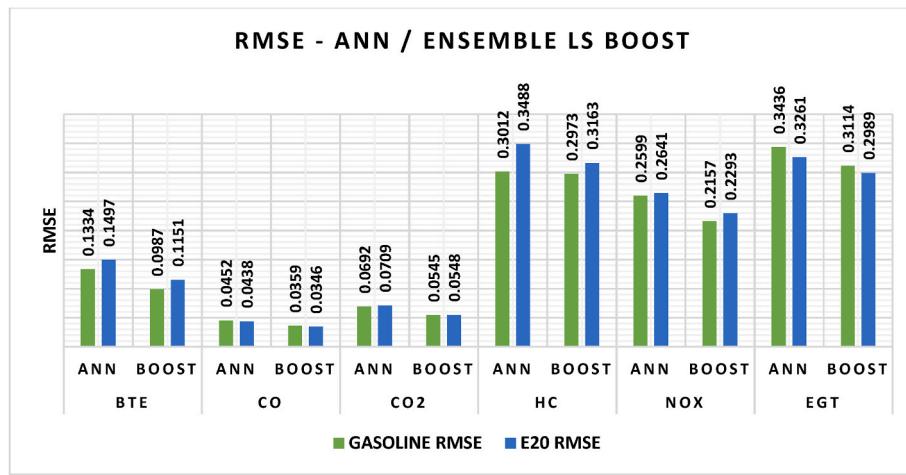
Fig. 13. R² - combustion performance and emission parameters.

Fig. 14. RMSE - combustion performance and emission parameters.

0.9987. As indicated by R² values approaching 1, the ANN method achieved near-perfect accuracy in predicting virtually all engine parameters for both unleaded petrol and E20 test fuels.

Root Mean Square Error (RMSE) values, which measure the average magnitude of prediction errors, followed uniformly, highlighting the model's predictive accuracy. In addition, as shown in Fig. 14 The root mean square errors for petrol test values BTE, HC, CO, NOx, and EGT were, respectively, 0.1334, 0.0452, 0.0692, 0.3012, 0.2599, and 0.3436. The E20 dual-fuel blend had respective values of 0.1497, 0.0438, 0.0709, 0.3488, 0.2641, and 0.3261.

In light of these excellent correlation coefficients and small prediction errors, the ANN-based model appears to be a reliable and robust forecasting tool for predicting the combustion, performance, and emission parameters of the dual-fuel SI engine powered by ethanol blends.

5.2. Ensemble boosting based prediction model (ensemble LS BOOST)

For both neat gasoline and E20 dual-fuel blends, the Ensemble LS Boost model was used to forecast engine combustion, performance, and emission levels under a variety of input situations. The values predicted by the Ensemble LS Boost model and the engine parameters as obtained empirically are shown in comparison in Fig. 12. The plots also provide a side-by-side comparison of the predicted and actual values for each of the testing data points, illuminating a significant association. The outcomes of the Ensemble LS Boost algorithm's analysis of the connections

between significant engine parameters BTE, HC, CO, CO₂, NOx, and EGT. The R-squared R² values were 0.9960, 0.9579, 0.9955, 0.9984, 0.9988, and 0.9987 for predicting the aforementioned engine parameters for neat gasoline. For the E20 blends, the method returned R² values of 0.9947, 0.9247, 0.9889, 0.9983, 0.9895, and 0.9990. The Ensemble LS Boost method almost perfectly predicted nearly all engine parameters for both neat petrol and E20 test fuels, as shown by R² values close to 1. The root mean square errors for the petrol test values for BTE, HC, CO, NOx, and EGT were 0.0987, 0.0359, 0.0545, 0.2973, 0.2157, and 0.3114, respectively, as shown in Fig. 14. For the E20 dual-fuel blend, the corresponding values were 0.1151, 0.0346, 0.0548, 0.3163, 0.2293, and 0.2989. For both test fuels, the error values for E20 and petrol were remarkably near 0. The optimal hyperparameters for the Ensemble LS Boost algorithm have been determined through cross-validation. These hyperparameters include a minimum leaf size of 4, a learning rate of 0.1, and 100 iterations in total. After the initial 100 iterations, Bayesian optimisation was employed to further refine the hyperparameters and improve the model's performance.

Comparison of ANN and Ensemble LS Boost Models for Ethanol Dual-Fuel SI Engine Characteristics Prediction.

Artificial Neural Network (ANN) and Ensemble LS Boost models were compared in order to forecast the optimal combustion, performance, and emission values of a dual-fuel SI engine. This comparison is displayed in Figs. 13 and 14, which depict the R² values and statistical error indicator RMSE for the predicted outcomes. Throughout the test

trials, Ensemble LS Boost outperformed the ANN models marginally, demonstrating superior efficacy and generalisation capabilities. This conclusion is supported by a comparison between the coefficient of determination R^2 and statistical error indicators including RMSE.

The Ensemble LS Boost model yielded higher R^2 values for BTE, CO, CO₂, HC, NOx, and EGT compared to the ANN model. Specifically, for neat gasoline, the R^2 values were 0.9961, 0.9579, 0.9955, 0.9984, 0.9988, and 0.9987, while for E20, they were 0.9947, 0.9247, 0.9889, 0.9983, 0.9895, and 0.9991. In contrast, the ANN model's R^2 values for neat gasoline were 0.9931, 0.9287, 0.9922, 0.9962, 0.9982, and 0.9985, and for E20, they were 0.9896, 0.9025, 0.9873, 0.9966, 0.9921, and 0.9987, as shown in Fig. 13.

Furthermore, the Ensemble LS Boost model outperformed the ANN model in terms of statistical error metrics, as shown in Fig. 14. The RMSE values for BTE, CO, CO₂, HC, NOx, and EGT were 0.0987, 0.0359, 0.0545, 0.2973, 0.2157, and 0.3114 for gasoline and 0.1151, 0.0346, 0.0548, 0.3163, 0.2293, and 0.2989 for E20. In comparison, the ANN-derived RMSE values for neat gasoline were 0.1334, 0.0452, 0.0692, 0.3012, 0.2599, and 0.3436, and for the E20 dual-fuel blend, they were 0.1497, 0.0438, 0.0709, 0.3488, 0.2641, and 0.3261. The lower RMSE values obtained by the Ensemble LS Boost model demonstrate its higher predictive accuracy and superior performance in estimating engine characteristics for both petrol and E20 blends.

The superiority of Ensemble LS Boosting over Artificial Neural Networks (ANN) can be attributed to its efficiency in leveraging multiple weak learners. Since the dataset had a medium number of training samples, ANN captured and generalised intricate patterns adequately, while Ensemble LS Boosting harnessed the collective strength of multiple weak learners to create a more robust model than ANN. The sensitivity of ANN models to noise and outliers in smaller datasets also made the ANN model marginally inferior. The boosting mechanism of Ensemble LS Boosting prioritised challenging data points, mitigating the risk of overfitting and improving the overall predictive accuracy. The capability of Ensemble LS Boosting to handle nonlinear relationships was particularly advantageous when engine characteristics and emission parameters had complex dependencies on input features. This adaptability enabled for accurate modelling of fundamental relationships, which led to enhanced predictive performance. The optimisation of Ensemble LS Boosting's hyperparameters enhanced its capabilities by fine-tuning the model, thereby making it more adaptive and data-specific. Ensemble LS Boosting proved to be a highly effective and reliable method for predicting the combustion, performance and emission parameters in Ethanol-gasoline dual-fuel engine.

5.3. Simulink Model and dashboard development

The research further concentrates on creating and implementing the two distinct, interactive Simulink models with the Ensemble LS Boost Model for the Spark Ignition (SI) engine, each specifically designed for Neat Gasoline and Ethanol E20 test fuels across all loads. Figures Figs. 15–18 depict the Simulink models and dashboards developed for both Neat Gasoline and Ethanol E20 fuels. These representations of Simulink models offer a complete view of the system architecture and the dynamic, interactive dashboard interface, which enables a thorough investigation of engine combustion, performance, and emission parameters. The integration of advanced machine learning algorithms, particularly Ensemble Least-Squares (LS) Boosting, was a crucial aspect of the model-based development methodology, in MATLAB and Simulink. The Ensemble Least-Squares Boosting technique, which is characterised by a collection of weak learners, such as decision trees, excels at capturing the complex, nonlinear relationships innate to engine performance parameters. Ensemble LS Boosting performed marginally better than Artificial Neural Network (ANN) models in the preliminary experiments, which led to its selection. After training, the ensemble models were imported into the MATLAB workspace for use in the Regression Ensemble Predict block's real-time prediction. This integration was essential for concluding the model-based development workflow and launching the final Simulink model.

Each model was independently trained using a comprehensive dataset that embodied a wide variety of engine operational conditions unique to both the test fuels. Ensemble LS Boosting was used for the prediction of all the parameters, effectively minimising prediction error and maximising model interpretability by optimising critical hyperparameters such as the maximum tree depth, number of trees in the ensemble, and node splitting etc. For both Neat Gasoline and Ethanol E20 dual fuel blend test fuels, the Simulink models of the SI engine, which form the basis of the Ensemble LS Boost algorithms, demonstrated outstanding performance in predicting all parameters BTE, CO, CO₂, HC, NOx, and EGT. Segregated validation datasets were utilised to ensure the model's robust generalisation capability and prevent overfitting. During operation, the fundamental parameters Load and Brake Specific Fuel Consumption BSFC were modified using interactive knobs on the Simulink dashboard for all load settings between 1.6 kW and 8 kW across the range of BSFC values. This dynamic method permits the modification of engine Load and BSFC conditions during simulation. In turn, the trained models predict performance outputs, such as Brake Thermal Efficiency (BTE), Carbon Monoxide (CO), Carbon Dioxide (CO₂), Hydrocarbons (HC), Nitrogen Oxides (NOx), and Exhaust Gas

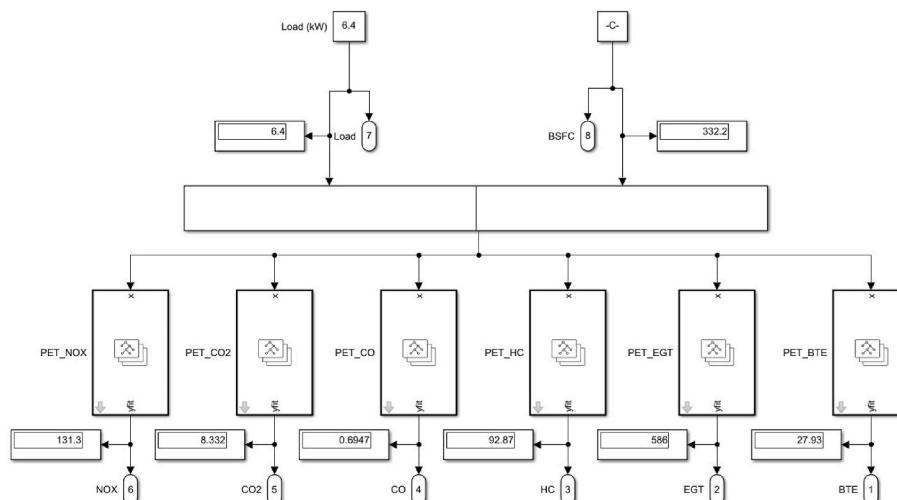


Fig. 15. Simulink model with inputs and outputs for neat gasoline test fuel.

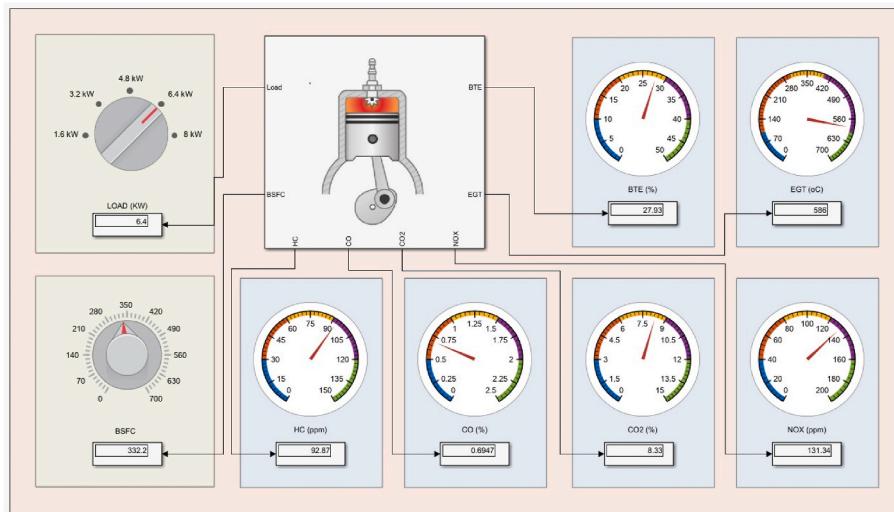


Fig. 16. Simulink dashboard for neat gasoline test fuel.

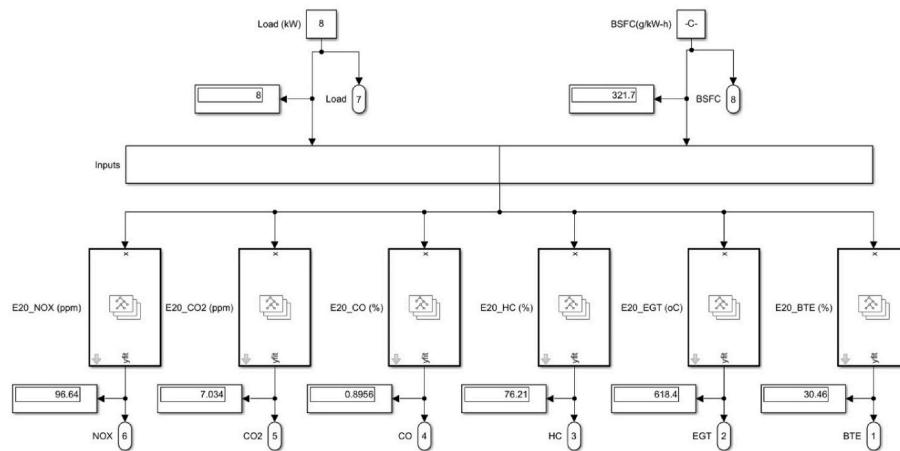


Fig. 17. Simulink model with inputs and outputs for E20 test fuel blend.

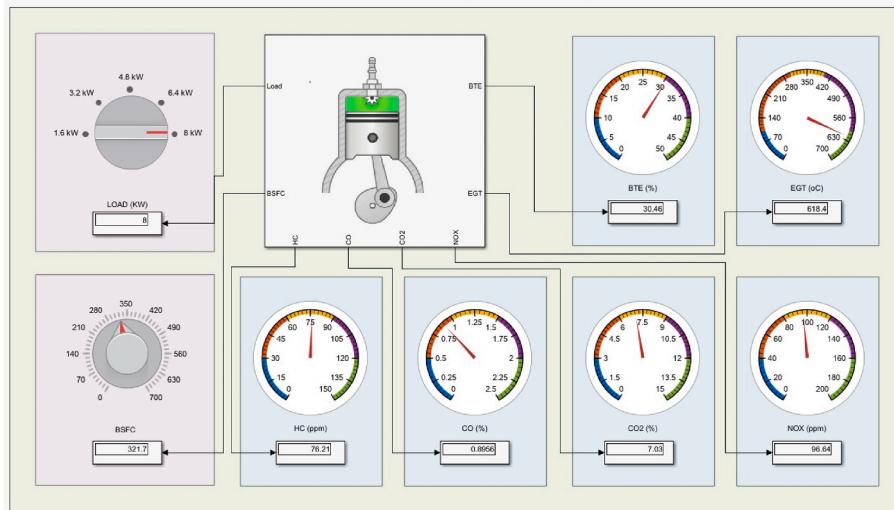


Fig. 18. Simulink dashboard for E20 test fuel blend.

Temperature (EGT), providing a real-time simulation of engine performance for both neat gasoline and E20 Ethanol dual fuel blend.

These outputs, which are represented by dashboard dials and displays, provide real-time performance monitoring and facilitate an interactive investigation of the interdependence of engine parameters. The successful validation of both models, coupled with their interactive and intuitive design, confirms their potential as valuable tools for real-time model-based engine analysis and optimisation, highlighting the extensive applicability of machine learning in the development of models for complex ethanol-based dual-fuel SI engine systems.

6. Conclusions

This study utilised two machine learning techniques, Artificial Neural Networks (ANN) and Ensemble LS Boosting, to predict the combustion, performance, and emission parameters of an internal combustion engine operating on gasoline and ethanol blends. These parameters included BTE, CO, CO₂, HC, NO_x, and EGT.

The regression analysis results on the data from the engine experiments showed a complex relationship between the independent and dependent variables. Both machine learning models exhibited good predictive capabilities and demonstrated good accuracy in forecasting engine characteristics.

The ANN model, which is based on neural networks, showed that the predicted and measured values of key engine parameters, such as BTE, CO, CO₂, HC, NO_x, and EGT, were in a good relationship. These parameters' R² values were close to 1, indicating near-perfect precision for both neat gasoline and E20 test fuels. The consistent and small root mean square errors (RMSE) of the ANN-based model further validated its dependability and robustness in predicting engine behaviour.

However, the Ensemble LS Boosting model showed superior congruence with experimental engine parameters across all datasets. In terms of predicting engine parameters, Ensemble LS Boosting models were marginally more accurate than ANN models. Its ensemble nature, resistance to overfitting, ability to manage nonlinear relationships, and optimised hyperparameters contribute to its outstanding performance in predicting the combustion, performance, and emission characteristics of the dual-fuel SI engine utilising ethanol blends.

The Simulink models dashboard generated with Ensemble LS Boost Machine Learning models for neat gasoline and E20 test fuels offers a comprehensive view of the system architecture and the dynamic, interactive dashboard interface, allowing for an in-depth assessment of engine combustion, performance, and emission parameters.

While the ANN model displayed dependable performance and accuracy, the Ensemble LS Boosting model excelled at attaining an even higher level of concordance with experimental data. Overall, the Ensemble LS Boost machine learning technique proved to be a robust tool, demonstrating significant promise for predicting the combustion and performance and reducing the emissions characteristics of the Ethanol-gasoline dual-fuel engine.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

D. Jesu Godwin: Conceptualization, Investigation, Writing – original draft. **Edwin Geo Varuvel:** Validation, Formal analysis, Supervision, Writing – review & editing. **M. Leenus Jesu Martin:** Conceptualization, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

- Agarwal, Avinash Kumar, Singh, Akhilendra Pratap, Maurya, Rakesh Kumar, 2017. Evolution, challenges and path forward for low temperature combustion engines. *Prog. Energy Combust. Sci.* 61, 1–56.
- AlNazr, H.A., Ahmad, N., Ahmed, U., Mohan, B., Jameel, A.G.A., 2023. Predicting physical properties of oxygenated gasoline and diesel range fuels using machine learning. *Alex. Eng. J.* 76, 193–219.
- Bae, Choongsik, Kim, Jaeheun, 2017. Alternative fuels for internal combustion engines. *Proc. Combust. Inst.* 36 (3), 3389–3413.
- Bai, Femilda Josephin Joseph Shobana, 2023. A Machine Learning Approach for Carbon di oxide and Other Emissions Characteristics Prediction in a Low Carbon Biofuel-Hydrogen Dual Fuel Engine. *Fuel* 341, 127578.
- Bai, Femilda Josephin Joseph Shobana, Shamugaiyah, Kaliraj, Sonthalia, Ankit, Devarajan, Yuvarajan, Varuvel, Edwin Geo, 2023. "Application of machine learning algorithms for predicting the engine characteristics of a wheat germ oil-Hydrogen fuelled dual fuel engine". *Int. J. Hydrogen Energy* 48 (60), 23308–23322.
- Baidya, S., Borken-Kleefeld, J., 2009. Atmospheric emissions from road transportation in India. *Energy Pol.* 37 (10), 3812–3822.
- Barboza, Augustine BV., Mohan, Sooraj, Dinesha, P., 2022. On reducing the emissions of CO, HC, and NO_x from gasoline blended with hydrogen peroxide and ethanol: optimization study aided with ANN-PSO. *Environ. Pollut.* 310, 119866.
- Bhowmik, S., Panua, R., Debroy, D., Paul, A., 2017. Artificial neural network prediction of diesel engine performance and emission fueled with diesel-kerosene-ethanol blends: a fuzzy-based optimization. *J. Energy Resour. Technol.* 139 (4), 042201.
- Chu, Steven, Cui, Yi, Liu, Nian, 2017. The path towards sustainable energy. *Nat. Mater.* 16 (1), 16–22.
- Covert, Thomas, Greenstone, Michael, Knittel, Christopher R., 2016. Will we ever stop using fossil fuels? *J. Econ. Perspect.* 30 (1), 117–138.
- Demirbas, Ayhan, 2007. Progress and recent trends in biofuels. *Prog. Energy Combust. Sci.* 33 (1), 1–18.
- Dey, S., Reang, N.M., Majumder, A., Deb, M., Das, P.K., 2020. A hybrid ANN-Fuzzy approach for optimization of engine operating parameters of a CI engine fueled with diesel-palm biodiesel-ethanol blend. *Energy* 202, 117813.
- Elfasakhany, A., 2014. The effects of ethanol-gasoline blends on performance and exhaust emission characteristics of spark ignition engines. *Int. J. Automotive Eng.* 4 (1), 609–620.
- Elfasakhany, Ashraf, 2016. Performance and emissions of spark-ignition engine using ethanol-methanol-gasoline, n-butanol-iso-butanol-gasoline and iso-butanol-ethanol-gasoline blends: a comparative study. *Eng. Sci. Technol. Int. J.* 19 (4), 2053–2059.
- Entenberg, Robert D., Menard Jr., Albert L., 1966. Future octane number requirements for future market demand. *J. Market.* 30 (1), 28–32.
- Fagundez, J.L.S., Lanzanova, T.D.M., Martins, M.E.S., Salau, N.P.G., 2020. Joint use of artificial neural networks and particle swarm optimization to determine optimal performance of an ethanol SI engine operating with negative valve overlap strategy. *Energy* 204, 117892.
- Farooq, S., Kumar, D.V., 2023. Experimental investigation of gasoline ethanol methanol iso-stoichiometric blends on SI engine. *Mater. Today Proc.* <https://doi.org/10.1016/j.mtpr.2023.06.138>.
- Gulia, S., Nagendra, S.S., Khare, M., Khanna, I., 2015. Urban air quality management-A review. *Atmos. Pollut. Res.* 6 (2), 286–304.
- Guttikunda, Sarath K., Nishadh, K.A., Jawahar, Puja, 2019. Air pollution knowledge assessments (APnA) for 20 Indian cities. *Urban Clim.* 27, 124–141.
- Iodice, P., Langella, G., Amoresano, A., 2018. Ethanol in gasoline fuel blends: effect on fuel consumption and engine out emissions of SI engines in cold operating conditions. *Appl. Therm. Eng.* 130, 1081–1089.
- Josephin, J.F., Geo, V.E., Sonthalia, A., Bharatiraja, C., Aloui, F., 2018. Predicting the different engine parameters of a rubber seed oil-ethanol dual fuel engine using artificial neural networks. *Int. J. Glob. Warming* 16 (4), 485–506.
- Karunamurthy, K., Janvekar, A.A., Palaniappan, P.L., Adhitya, V., Lokeshwar, T.T.K., Harish, J., 2023. Prediction of IC engine performance and emission parameters using machine learning: a review. *J. Therm. Anal. Calorim.* 1–23.
- Kodavasal, J., Abdul Moiz, A., Ameen, M., Som, S., 2018. Using machine learning to analyze factors determining cycle-to-cycle variation in a spark-ignited gasoline engine. *J. Energy Resour. Technol.* 140 (10), 102204.
- Kumar, T.S., Ashok, B., Saravanan, B., 2023. Calibration of flex-fuel operating parameters using grey relational analysis to enhance the output characteristics of ethanol-powered direct injection SI engine. *Energy*, 128340.
- Liu, J., Huang, Q., Ulishney, C., Dumitrescu, C.E., 2021. Machine learning assisted prediction of exhaust gas temperature of a heavy-duty natural gas spark ignition engine. *Appl. Energy* 300, 117413.
- Liu, S., Lin, Z., Zhang, H., Fan, Q., Lei, N., Wang, Z., 2023. Experimental study on combustion and emission characteristics of ethanol-gasoline blends in a high compression ratio SI engine. *Energy* 274, 127398.

- Manzetti, Sergio, Andersen, Otto, 2015. A review of emission products from bioethanol and its blends with gasoline. Background for new guidelines for emission control. *Fuel* 140, 293–301.
- Masum, B.M., Masjuki, H.H., Kalam, M.A., Fattah, I.R., Palash, S.M., Abedin, M.J., 2013. Effect of ethanol–gasoline blend on NO_x emission in SI engine. *Renew. Sustain. Energy Rev.* 24, 209–222.
- Masum, B.M., Masjuki, H.H., Kalam, M.A., Palash, S., Habibullah, M., 2015. Effect of alcohol–gasoline blends optimization on fuel properties, performance and emissions of a SI engine. *J. Clean. Prod.* 86, 230–237.
- Mohammed, M.K., Balla, H.H., Al-Dulaimi, Z.M.H., Kareem, Z.S., Al-Zuhairy, M.S., 2021. Effect of ethanol–gasoline blends on SI engine performance and emissions. *Case Stud. Therm. Eng.* 25, 100891.
- Mohan, B., Badra, J., 2023. An automated machine learning framework for piston engine optimization. *Appl. Energy Combust. Sci.* 13, 100106.
- Pla, B., De la Morena, J., Barres, P., Aramburu, A., 2023. A supervised machine learning technique for combustion diagnosis using a vibration sensor signal. *Fuel* 343, 127869.
- Ricci, F., Petrucci, L., Mariani, F., 2023. Using a machine learning approach to evaluate the NO_x emissions in a spark-ignition optical engine. *Information* 14 (4), 224.
- Sharma, Prabhakar, Sahoo, Bibhuti B., 2022. Precise prediction of performance and emission of a waste derived Biogas–Biodiesel powered Dual–Fuel engine using modern ensemble Boosted regression Tree: a critique to Artificial neural network. *Fuel* 321, 124131.
- Surisetty, Venkateswara Rao, Kumar Dalai, Ajay, Kozinski, Janusz, 2011. Alcohols as alternative fuels: an overview. *Appl. Catal. Gen.* 404 (1–2), 1–11.
- Tekin, Mehmet, Saridemir, Suat, 2019. Prediction of engine performance and exhaust emissions with different proportions of ethanol–gasoline blends using artificial neural networks. *Int. J. Ambient Energy* 40 (5), 470–476.
- Thakur, A.K., Kaviti, A.K., Mehra, R., Mer, K.K.S., 2017. Progress in performance analysis of ethanol–gasoline blends on SI engine. *Renew. Sustain. Energy Rev.* 69, 324–340.
- Verma, A., Dugala, N.S., Singh, S., 2022. Experimental investigations on the performance of SI engine with Ethanol–Premium gasoline blends. *Mater. Today: Proc.* 48, 1224–1231.
- Yang, Ruomiao, Xie, Tianfang, Liu, Zhentao, 2022. The application of machine learning methods to predict the power output of internal combustion engines. *Energies* 15 (9), 3242.
- Yücesu, H., Serdar, et al., 2007. Comparative study of mathematical and experimental analysis of spark ignition engine performance used ethanol–gasoline blend fuel. *Appl. Therm. Eng.* 27 (2–3), 358–368.
- Yusri, I.M., Mamat, R., Najafi, G., Razman, A., Awad, O.I., Azmi, W.H., et al., 2017. Alcohol based automotive fuels from first four alcohol family in compression and spark ignition engine: a review on engine performance and exhaust emissions. *Renew. Sustain. Energy Rev.* 77, 169–181.