POLISH MARITIME RESEARCH 2 (130) 2026 Vol. 33; pp. 85-97
10.2478/pomr-2026-0023

Ensemble Machine Learning and Simulated Annealing for Behaviour Prediction of a Diesel Engine Fuelled with Dimethyl Ether

Dinh Trung Pham [1,]
Diep Ngoc [2,]
Joanna Grochowalska [3]
Lan Huong Nguyen [4,*]
Duc Pham [5]
Van Hung Bui [6]
Thanh Hieu Chau [7]
Thanh Nam Dang [7]
Huu Cuong Le [7]

[1] Yersin University of Da Lat, Viet Nam / [2] Long Huynh, HUTECH University, Viet Nam / [3] Gdansk University of Technology, Poland
[4] Vietnam Maritime University, Viet Nam / [5] Dong Nai Technology University, Viet Nam / [6] The University of Danang, Viet Nam
[7] Ho Chi Minh City University of Transport, Viet Nam

* Corresponding author: *huongnlvck@vimaru.edu.vn (Lan Huong Nguyen)*

ABSTRACT

This study introduces a novel hybrid approach based on ensemble machine learning prediction and simulated annealing for the optimisation and minimisation of emissions from dimethyl ether (DME)-diesel fuel combustion compression ignition engines. Experimental data were obtained from a commercial, four-cylinder, direct injection diesel engine (Model 4113) under different load conditions, with varying blends of DME and diesel. Key emissions (NOx, CO, and HC), the brake mean effective pressure and the blend ratio were measured with high precision by AVL analysers. Multiple regression models were built and compared, including support vector regression (SVR) and XGBoost, with rigorous hyperparameter tuning. SVR yielded better performance for NOx (test $R^2 = 0.820$) and HC (test $R^2 = 0.854$), although both models achieved good accuracy for CO (SVR test $R^2 = 0.854$; XGBoost test $R^2 = 0.731$). A correlation analysis showed that there was a strong positive link between engine load and NOx, and good HC reduction with higher DME content. The trained ensemble models were implemented as objective functions in a simulated annealing algorithm to achieve multi-objective optimisation of critical engine parameters. The proposed approach has a high level of predictive reliability and effective global search capability, and provides a computationally efficient pathway for emission control in alternative fuel engines. These findings provide support for the broader use of DME as a clean, renewable fuel for compliance with stringent future emission standards.

Keywords: Dimethyl ether; Alternative Fuel; Diesel engine; Machine Learning; Simulated Annealing Optimisation

INTRODUCTION

Global transportation is heavily reliant on internal combustion engines (ICEs), despite the growing concern over air pollution and greenhouse gas (GHG) emissions [1,2]. Indeed, ICEs running on fossil fuels have been found to release very significant quantities of nitrogen oxides (NOx), particulate matter (PM), carbon monoxide (CO), and unburned hydrocarbons (HC) under normal operation cycles [3–5]; these pollutants are very dangerous to human health, as they worsen respiratory diseases and create layers of smog in cities. The increasing regulatory pressure in most countries has created an urgent need to find alternative fuels that are much cleaner and can significantly reduce the detrimental emissions from exhaust [6,7]. For these reasons, researchers around the world are exploring options for alternative fuels such as biodiesel [8,9], hydrogen [10], ammonia [11], alcohol [12,13], and gaseous fuels (biogas, producer gas, syngas, natural gas, etc.) [14–16] to power ICEs, in order to reduce GHGs and increase sustainability. Dimethyl ether (DME) is one promising oxygen-rich fuel for compression-ignition engines that requires relatively few changes [17,18]. As a synthetic compound, DME has a high cetane number and results in very low soot emissions, since the

molecular structure has no bonds between its carbon atoms and carbon atoms [19], [20]. Decades of research have demonstrated that DME combustion is associated with almost zero PM emissions and lower levels of HC and CO compared to typical diesel, although the formation of NOx tends to be heightened in situations of high engine load, and advanced control measures are required to become more efficient [21]. Arcoumanis et al. [22] highlighted the outstanding prospects of DME in terms of its low-emission performance when combined with suitable approaches to engine calibration. Subsequent research has solidified these results, and has demonstrated that DME is effective at increasing thermal efficiency whilst significantly reducing CO and HC levels in optimally designed setups [23–26]. Regardless of these obvious advantages, however, challenges are still encountered during practical implementations, which are associated with the fuel pressure demand, injector lubrication, and the special storage infrastructure that is necessary for on-vehicle applications. Recent studies of heavy-duty engines [27–30] have shown that the combination of an exhaust gas recirculation system and a selective catalytic reduction system is an effective way to manage the NOx emissions from a DME engine. Computational fluid dynamics models have enhanced the understanding of in-cylinder flow and reaction dynamics, but such models require significant computational resources to conduct predictive tasks [31,32].

Machine learning methodologies have previously been adopted in an attempt to achieve reliable predictions of engine performance and exhaust properties, based on few experimental measurements [33,34]. The use of artificial neural networks has been found in multiple works to provide a great deal of effectiveness in predicting the emissions and performance of diesel engines, with high correlation coefficients, under different operating conditions [35,36]. Ensemble learning methods, and especially random forest and extreme gradient boosting algorithms, can increase the robustness of prediction and reduce the risk of overfitting and nonlinear dependencies, which would otherwise underperform with standalone models [37–39]. Such ensemble techniques have provided high performance in the prediction of PM emissions in gasoline direct-injection engines and the CO_2 concentrations in a wide range of vehicle fleets. In spite of this development, however, the use of ensemble machine learning for the prediction of emissions from engines running on pure DME as the sole fuel still has remarkably few applications [40,41]. Optimisation algorithms play an essential role in identifying the best parameters for engines in terms of reducing pollutant emissions while ensuring that they remain competitive in terms of their efficiency and power production [42,43]. The simulated annealing (SA) algorithm was inspired by controlled metallurgical cooling [44], and it can be particularly useful in achieving global optimisation in the complex and multimodal calibration surfaces of engines. Previous researchers have reported the use of various optimisation methods to enhance the efficiency of DME synthesis through a significant improvement in yield parameters [45,46], but a combination of these approaches with predictive emission models to achieve direct minimisation for DME-fuelled engines has not been studied sufficiently in the existing literature. Traditional methods of optimising these engines usually rely on gradient-based local search methods or genetic algorithms, which can easily become trapped in poor-quality solutions. There are therefore important gaps in research into the synergistic use of ensemble machine learning in situations where the precise forecasting of emissions, followed by global parameter refinement, takes place by SA within the context of DME.

There are few publications that have addressed the issue of multi-objective optimisation to realise a balanced reduction in NOx, CO, and HC emissions. The limited availability of detailed data on pure DME in large operating regimes also limits the applicability of the model and its application to real-world environments. The present work attempts to address these key gaps by presenting a new hybrid approach in which ensemble machine learning prediction is combined with SA optimisation. The proposed ensemble of extreme gradient boosting models will be compared with a standalone support vector regressor (SVR) trained on massive experimental datasets from DME-powered compression-ignition engines. The aim of this work is to provide high-fidelity forecasts of primary emissions in the form of nitrogen oxides, carbon monoxide, and unburned hydrocarbons under varying load and speed conditions. The SA and artificial bee colony (ABC) algorithms are used for optimisation of control parameters to reduce emissions. The main research goal is to attain a high level of prediction, although an additional objective is to achieve a reduction in nitrogen oxides with optimised settings. The results will be used to design the next generation of low-emission engines that will be powered using DME. Finally, this approach can lead to carbon reductions for transportation, without compromising the reliability of operation or the level of performance.

MATERIALS AND METHODS

TEST SETUP AND DATA COLLECTION

A commercially available direct-injection diesel engine, type 4113, produced by the Wuxi Diesel Engine Factory, was used in this study by Ying et al. [47]. It had four cylinders, water cooling, a four-stroke cycle, and a natural aspiration design. Other than the necessary modifications to the fuel distribution system, the base engine structure was left unaltered. Since DME corrodes rubber materials, copper tubing was used in place of the original rubber fuel lines. To reduce vaporisation losses of the volatile DME component, a customised LPG-style container was used in place of the conventional diesel tank. Fig. 1 provides a comprehensive schematic of the entire experimental setup, including all its associated parts. The fuel exit is positioned at the base of the vessel, and the recycled surplus blend is directed back into the top inlet. Before delivery, the DME/diesel mixture is pressurised to around 700 kPa using an electric pump, and this pressurised mixture enters the combustion chambers via conventional injectors. To keep the system stable, excess fuel is returned to the storage tank via a specialised regulating valve. An exhaust gas analysis system (AVL brand) was used to quantify the gaseous exhaust components, and to ensure reliability, each experimental trial was repeated. The extensive dataset obtained from these controlled engine tests served

as the basis for building sophisticated machine learning models and carrying out SA-based optimisation processes with the aim of efficiently forecasting and reducing emissions.

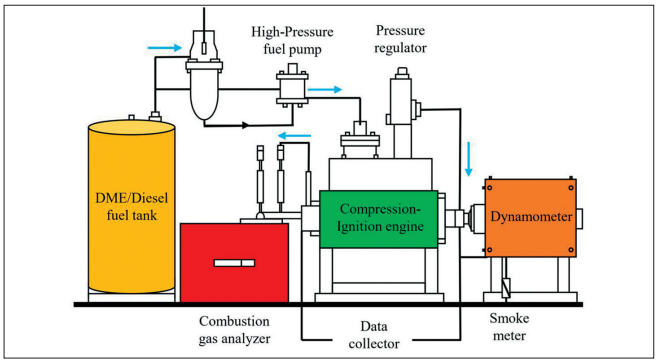

Fig. 1. Test setup modified from [47]

METAHEURISTIC METHOD

Metaheuristic approaches are high-level optimisation techniques in which the aim is to achieve effective searching of complex, nonlinear, and multimodal search spaces that classical gradient-based methods cannot search effectively, or which cause them to become stuck in a local minimum. A balance between exploration and exploitation is typically achieved using stochastic operators, population-based or trajectory-based search, and flexible neighbourhood structures. Metaheuristics are therefore suitable for the simultaneous tuning of machine-learning models and engine operating parameters with several constraints.

Simulated annealing

SA is a trajectory-based metaheuristic based on the annealing of solids, in which a material is initially heated and then allowed to cool gradually to achieve a low-energy crystalline form. In the optimisation algorithm, a candidate solution is perturbed repeatedly to create neighbours whose acceptance is based on the change in objective value and a temperature parameter [48,49]. The Metropolis criterion can be used to accept poor solutions, which enables the algorithm to overcome local minima at increased temperatures. The search becomes more exploitative as the temperature decreases, following a certain cooling schedule, and focuses on areas with the lowest objective value. Only objective-function assessments are needed in SA, which is why it shows promise for issues where the problem has a complex and black-box representation, such as the prediction of engine emissions with ensemble learning [50,51]. Its straightforward implementation, small parameterisation, and capacity to deal with continuous, discrete, and constrained variables have made it very popular for use in engineering design, optimisation of energy systems, and hyperparameter optimisation of machine-learning models. A schematic representation of SA is shown in Fig. 2(a).

Artificial bee colony

ABC is a population-based metaheuristic approach that emulates the foraging process of honeybee swarm colonies seeking nectar sources. A schematic representation of the ABC algorithm is shown in Fig. 2(b). The algorithm consists of employed bees, onlooker bees, and scout bees, each of which has a different role in terms of the exploration and exploitation of the search space [52,53]. Employed bees have affiliated food sources (candidate solutions), and their location is changed locally to find improved quantities of nectar, comparable to a neighbourhood search of current solutions. The bystander bees can observe a hive waggle-dance-like probabilistic system and choose the most convenient food sources according to their fitness, a process that tends to concentrate the search in high-quality areas [54,55]. Scout bees carry out random surveying of the territory when a food source has been depleted, an activity that introduces diversity and avoids responses to a sub-optimal minimum. ABC is especially suitable where the optimisation of continuous engineering is needed, as it has a limited number of control parameters, can be hybridised with machine learning surrogates, and supports noisy or multi-modal objective functions [56,57]. Hence, ABC has been extensively used for combustion, energy-system, and emission control optimisation problems.

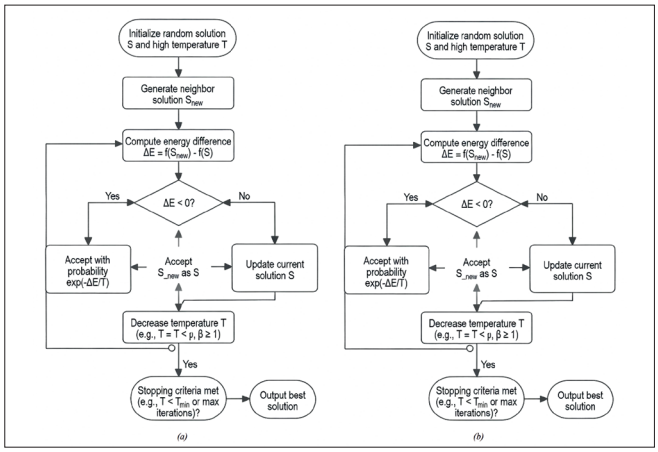

Fig. 2. Workflow diagrams for (a) simulated annealing; (b) artificial bee colony

MACHINE LEARNING

Machine learning offers data-driven models that are capable of capturing the nonlinear and interacting effects between engine operating parameters and fuel properties, and are frequently more effective than the traditional empirical correlations. Trained regression algorithms can be used to predict the emissions of NOx, CO, and HC with a high level of accuracy and speed. Such models also work well with multi-objective engine calibration methods that are based on metaheuristic optimisers.

Support vector regression

SVR is a variation of the support vector machine, which approximates continuous values through the use of the epsilon-insensitive loss. It works by building a regression tube around the data but only penalising deviations greater than the value of epsilon, and the complexity of the model is controlled through the regularisation parameter C [58,59]. Kernel functions (e.g. the radial basis function, RBF) are used to transform inputs to high-dimensional feature spaces, in which nonlinear relationships between engine parameters and emissions become almost linear [60,61]. Since a solution can be obtained based

on a relatively small number of samples (support vectors), SVR can work with relatively large datasets, and is resistant to overfitting in a high-dimensional feature space [62]. These attributes render the use of SVR an important tool for emissions prediction, where data are limited but strong interactions and nonlinearities are present, and where experimental campaigns are costly.

Extreme gradient boosting

Extreme Gradient Boosting (XGBoost) is an efficient and regularised version of gradient-boosted decision trees that yields high predictive accuracy and high scalability. The algorithm constructs a chain of shallow trees, with each new tree applied to the residual of the preceding trees, to minimise a differentiable loss (such as squared error) in a regression. XGBoost applies shrinkage (learning rate), column and row subsampling, and explicit L1/L2 regularisation to regulate the complexity of the model and decrease overfitting, an important aspect when modelling engine emission data that are noisy [63,64]. Its tree-based representation inherently indicates nonlinearities, threshold effects, and higher-order interactions between the operating conditions and fuel properties without the need for explicit feature engineering. XGBoost has also given excellent results in engine-emission experiments, providing high coefficients of determination and stable extrapolation within the experimental space, and has been shown to be efficient enough to be used in the optimisation loop (e.g. a metaheuristic search) with other methods [65,66].

RESULTS AND DISCUSSION

Several Python libraries were used for data management, modelling, and visualisation in the Jupyter Notebook environment. Pandas and Numpy were used for data preparation and tabular manipulation, as they allowed for the creation of derived features and the effective management of experimental engine emission datasets. The XGBoost models employed the XGBRegressor class from the xgboost package via its scikit-learn–compatible API, while the SVR models were constructed using scikit-learn (sklearn.svm.SVR). Pure Python was used to code the metaheuristic optimisation techniques (i.e. ABC and SA) with numpy to support randomisation and numerical operations. Line graphs, scatter plots, and boxplots comparing actual and expected emissions for the training and testing datasets were created using matplotlib and seaborn. The experimentation process and interpretation of results were made much easier by Jupyter Notebook's interactive workflow, which allows the user to run cells iteratively, examine interim results, adjust hyperparameters, and record code, figures, and discussion in a single repeatable file.

METAHEURISTIC OPTIMISATION

Simulated annealing-based optimisation

Fig. 3 illustrates the exploration of the search space by the SA algorithm and how it reaches an optimum operating condition for a DME-fuelled engine. Fig. 3(a) shows the temperature schedule, in which the temperature is gradually reduced, without interruption, to almost zero after approximately 500 iterations. This schedule implies that initially, the algorithm will often accept even worse solutions, allowing for a broad search and avoidance of a sub-optimal local minimum. The lower the temperature, the less often there is acceptance of worse moves, and the search process is slowly concentrated on exploitation close to promising areas of the design space. The path of candidate solutions on the brake mean effective pressure (BMEP) blend (in this case, LCV) plane can be seen in Fig. 3(b). The blue line is initially dispersed, and thus occupies a wide area, but over successive steps the range is narrowed, and this is how SA is steered towards a steady optimum. The scatter in the objective values with increasing iteration number is shown in Fig. 3(c). The initial iterations show a high degree of variation with the relatively high objective value, but later iterations tend to be concentrated around low values, implying better solutions as the algorithm cools. The best-so-far object plotted in Fig. 3(d) shows a sharp decrease in the first few tens of iterations and a long straight line near the end, thus verifying that the global or close-to-global minimum is obtained in the initial iterations and then followed. The present objective history can be seen in Fig. 3(e). Although fluctuations persist over the run, as some uphill movements are still tolerated, the overall trend is downward, and there is convergence towards the best-so-far curve. This is a positive activity of exploration and exploitation, which is a sign of successfully tuned SA. At convergence, an optimum operating point with values of BMEP = 0.5544 MPa and LCV = 40.4246 MJ/kg is achieved (see Table 1), which implies a minimised composite objective value of 0.9108. In such a state, the projected values of the emissions are NOx = 1253.4 ppm, CO = 0.0267, and HC = 34.24 ppm, representing a good compromise for the competing emission targets involved in the multi-objective formulation.

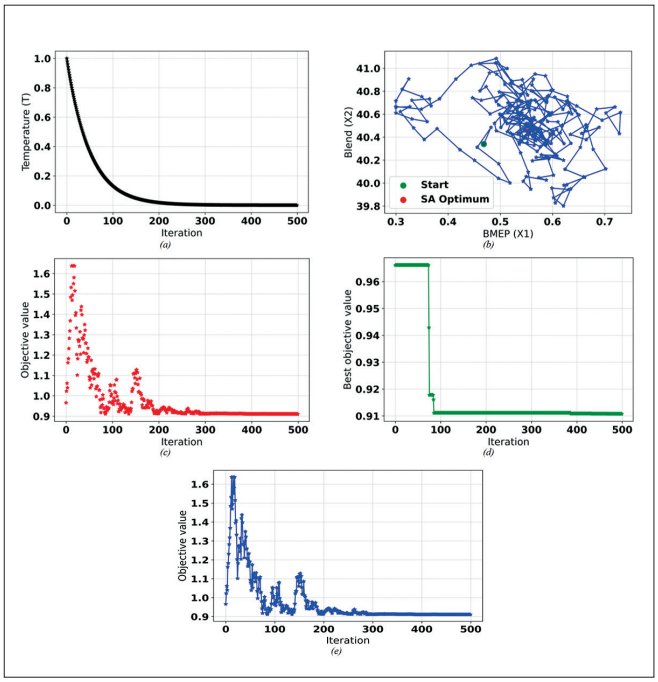

Fig. 3. Results for SA-based optimisation: (a) temperature schedule; (b) search trajectory; (c) scatter of objective; (d) best so far objective; (e) current objective

Artificial bee colony

The ABC algorithm examines the decision space and converges to an ideal operating condition for the engine powered by DME, as shown in Fig. 4. The food sources in Figs. 4(a–c) show potential solutions in the BMEP–LCV plane throughout the various cycles. The food sources in Cycle 1 are widely dispersed, suggesting that the search space was initially explored almost at random. The swarm progressively shrinks around a small area as the iterations advance to Cycles 150 and 299, and the suggested ABC optimum stabilises at the final solution. The employed, observer, and scout bee phases work together to steer the population from a variety of beginning positions in the direction of a high-quality area, as this graphic demonstrates. The distribution of the best objective values across several runs is displayed in Fig. 4(d). The algorithm appears to be quite consistent, as evidenced by the very sharp peak at 0.911, and continually converges to almost the same optimum with little variation in the quality of the final answer. The evolution of the global best objective as a function of the cycle number is shown in Fig. 4(e, f). There is a dramatic initial drop in the first few cycles followed by a long plateau, highlighting the early convergence of the ABC algorithm and showing that only a few cycles are needed to achieve almost ideal performance. Strong exploitation of the prospective food sources chosen probabilistically by spectator bees is the cause of this quick improvement. An ideal operating state with BMEP = 0.5558 MPa and LCV = 40.4368 MJ/kg, or a minimised composite objective value of 0.9107, is found by the ABC algorithm at convergence. The generated operating point represents a strong compromise between the conflicting emission objectives within the specified search domain, as can be seen from the related anticipated emissions of NOx = 1255.59 ppm, CO = 0.0268%, and HC = 34.13 ppm (shown in Table 1).

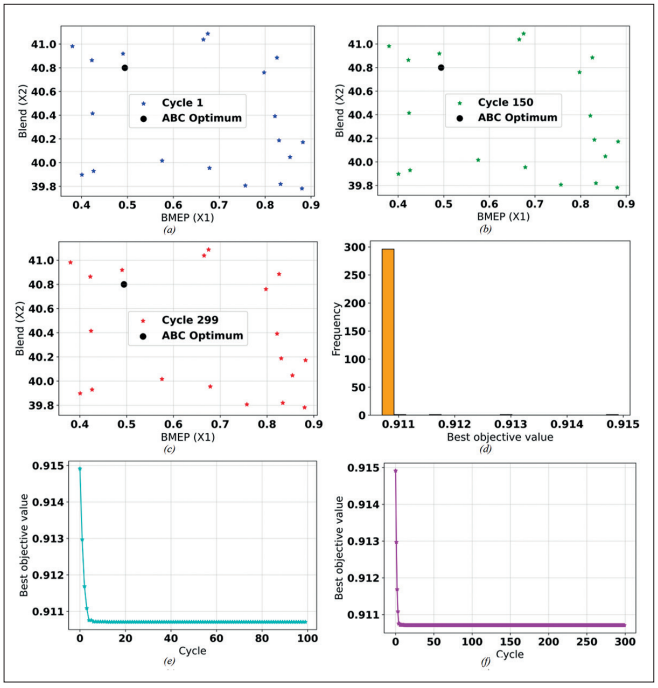

Fig. 4. Results for ABC optimisation: (a) food sources in the decision space for the first cycle: (b) food sources in the decision space for the 150th cycle; (c) food sources in the decision space for the 299th cycle; (d) distribution of best values; (e) early convergence; (f) best objective per cycle

Tab. 1. Optimisation results

Algorithm	Simulated annealing	Artificial bee colony
BMEP, MPa	0.5544	0.5558
LCV, MJ/kg	40.4246	40.4368
Objective	0.9108	0.9107
NOx, ppm	1253.4	1255.59
CO, %	0.0267	0.0268
HC, ppm	34.24	34.13

PROGNOSTIC MODELLING

Correlational analysis

The correlation matrix in Fig. 5(a) shows how the variables of the BMEP and the percentage of diesel and diesel blend in the diesel blend varied in relation to the main exhaust gas pollutants of nitrogen oxide (NOx), carbon monoxide (CO), and hydrocarbons (HC) during the experimental engine test. There is a strong positive correlation (0.92) between the BMEP and NOx emissions, indicating that the greater the engine load, the greater the formation of NOx as a result of the increased combustion temperatures. NOx has moderate negative correlations with both CO (−0.42) and HC (−0.58), representing the common trade-off in which situations that support full combustion give rise to increased NOx and decreased CO and HC. The blend ratio has a weak negative correlation with NOx (−0.03) and a significant positive correlation with CO (0.48), meaning that an increase in the DME content will increase CO by a few points and decrease NOx by a few points. HC emissions are found to have a moderate negative correlation with BMEP (−0.44), blend ratio (−0.45), and NOx (−0.58), suggesting that the combustion efficiency is enhanced under high loads and that the higher the DME content, the lower the unburned hydrocarbons. HC is positively related to CO, but the correlation is weak (0.19). Overall, the matrix indicates prevailing load-based NOx growth, a desirable decrease in HC with the DME mix, and conventional emission trade-offs that are of interest in regard to machine learning modelling and metaheuristic optimisation.

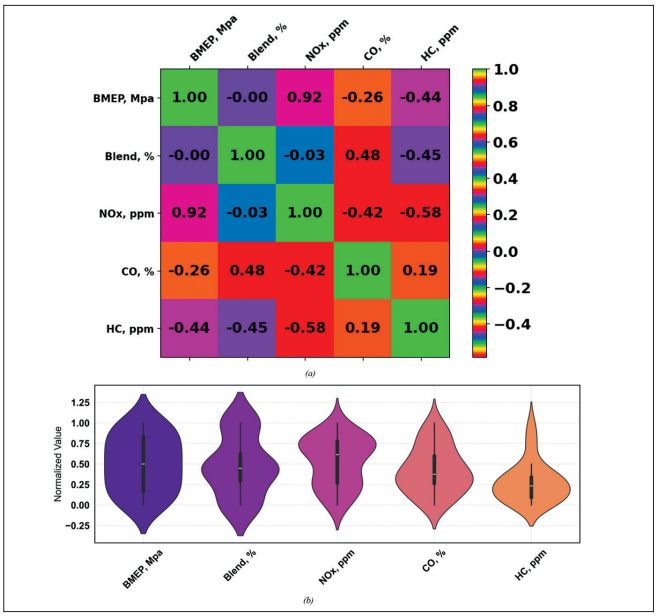

Fig. 5. (a) Data correlation heatmap; (b) violin plots

The violin plots in Fig. 5(b) illustrate how the normalised variables of the DME/diesel engine experimental data, such as the BMEP, the blend percentage, and the emissions of NOx, CO, and HC, are distributed. Each of the violin plots represents a kernel density estimate (with probability density) with an embedded boxplot (with median, interquartile range, and whiskers). The distribution of BMEP and percentage blend is relatively symmetric, with values concentrated around 0.5, indicating an equal distribution of the operating loads and DME blending ratios. The distribution of NOx is mediocre, as it has a peak in the middle and represents the normal range of variations that are found in various combustion scenarios. The CO distribution is a little skewed, with a lower median and more frequent readings in the low-CO range, whereas HC is the most skewed towards the right, and the central density is smaller, indicating occasional high readings of unburned hydrocarbons. On the whole, there is a reasonable data distribution with no extreme outliers following normalisation, a finding that supports strong machine learning modelling and optimisation in the research.

NOx model

The results of the SVR-based model for NOx emission are depicted in Fig. 6, and show high predictive performance. Hyperparameter tuning was done carefully, with the RBF kernel being the most successful over the linear kernel, and the best parameters were identified as C = 150, epsilon = 0.08 and gamma = 0.05 within the ranges considered for C (1–500), epsilon (0.01–0.3), and gamma (scale to 0.5). The hyperparameter optimisation results are listed in Table 2. The model attained robust values for R2 (0.982), mean squared error (MSE) (2246.68), mean absolute percentage error (MAPE) (0.038), and Kling–Gupta efficiency (KGE) (0.970) on the training set, implying that the model fits the observed values of NOx almost perfectly. On the unseen test set, the performance was still satisfactory, with an R2 of 0.820 and increases in MSE to 38382.7, MAPE to 0.175, and KGE to 0.674. The results indicate a minor level of overfitting due to the constraints of the dataset sizes and natural variations in measurements of engine emissions testing, as listed in Table 3. These quantitative measures were further supported by visual measures: the line plot of actual and predicted values (Fig. 6(a)) indicated that there was close agreement with the training points and reasonable tracking of test samples, whereas the scatter plot (Fig. 6(b)) indicated a strong linear correspondence of the training data along the ideal 45° line and acceptable dispersion of the test data along the ideal regression line. The comparisons between real and model distributions in Fig. 6(c) show similar spreads and medians for the actual and model distributions on both datasets, indicating that the model is reliable even when tested on unknown data. All of these findings confirm that SVR is a successful method of predicting NOx emissions with high accuracy, and thus provides a strong basis for proceeding to the next stage of optimisation based on a multi-objective model that can reduce harmful exhaust emissions while maintaining the efficiency of the engine in the applications using DME fuel.

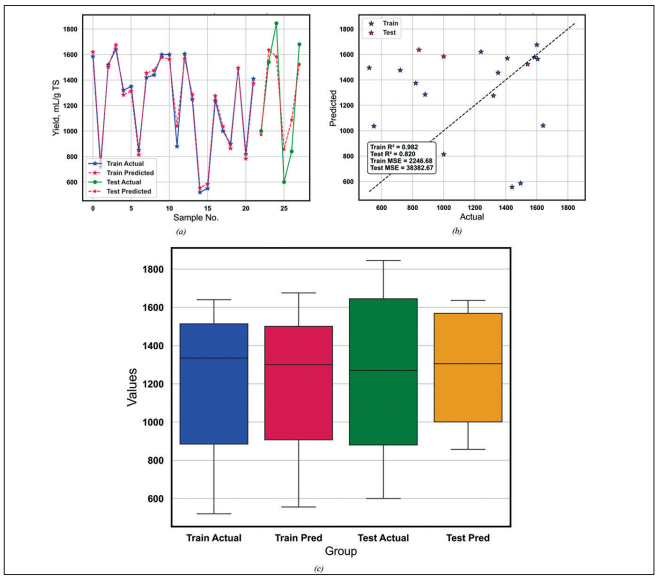

Fig. 6. Results for the SVR-based NOx emission model: (a) actual vs predicted values; (b) scatter plot; (c) box plots

Training of the XGBoost model to predict the NOx emissions from a DME-diesel-fuelled compression-ignition engine was based on experiments, which provided inputs in the form of BMEP, blend ratio, and other corresponding operating conditions. After a systematic trial, the best configuration identified through hyperparameter optimisation was as follows: number of estimators = 350 (in the range 100–600), learning rate = 0.06 (in the range 0.01–0.2), maximum tree depth = 5 (in the range 3–8), subsample ratio = 0.9 (in the range 0.6–1.0), and column sampling ratio = 0.8 (in the range 0.6–1.0), as listed in Table 2. These environments represent a trade-off between model complexity and generalisation, meaning that XGBoost closely reproduces the intricate and interactive influences on NOx formation during combustion. The performance metrics listed in Table 3 were found to be robust for both the training and work datasets. On the training set, XGBoost achieved an R^2 of 0.963, which implies that the amount of variance in the NOx emissions accounted for is 96.3%. The MSE was low at 5782.01, the MAPE was 0.054 (5.4%), and the Kling–Gupta efficiency (KGE) was high at 0.969. These values are close to ideal in terms of fitting to the training data. When applied to the unseen test set, the model performed relatively well, with an R^2 of 0.928 (explaining 92.8% of the variance), an MSE of 6437.37, an MAPE of just 0.060 (6.0%), and a KGE of 0.806. This small change in MSE and the slight change in KGE between training and testing indicate a low level of overfitting, and hence better generalisation by XGBoost than most other regressions, especially since engine emission datasets are small and noisy.

The visual diagnostics in Fig. 7 also testify to the strength of the model. The line plot of actual versus predicted results (train + test combined) in Fig. 7(a) shows close tracking between the actual (blue/green) and predicted (red/orange) values at most sample points, with the low yield regions at the extreme points showing only slight deviations. As can be seen from the scatter plot (Fig. 7(b)), the training points (blue stars) are well aligned along the ideal 45° dashed line, whereas the testing points (red stars) are also close to linear, with acceptable scatter, a finding that agrees with the high R^2 for the test. A comparison of the boxplots

(Fig. 7(c)) for both distributions shows that the median and interquartile ranges for the actual and predicted values are very similar for both the training and testing samples, except for one mild outlier in the test actual set, suggesting that the prediction is very reliable over the range of operation. Together, these findings identify XGBoost as an effective and consistent predictor of NOx emissions in engines that are fuelled with DME. These high values of R^2, low values for the error, and good spatial agreement between the true and estimated values confirm that this algorithm can be incorporated into a multi-objective optimisation system based on simulated annealing, to determine the engine parameter settings that are needed to reduce NOx under performance and for the management of other pollutants. The results demonstrate the importance of gradient boosting ensembles in modelling the complex and nonlinear relations of highly developed alternative-fuel combustion systems.

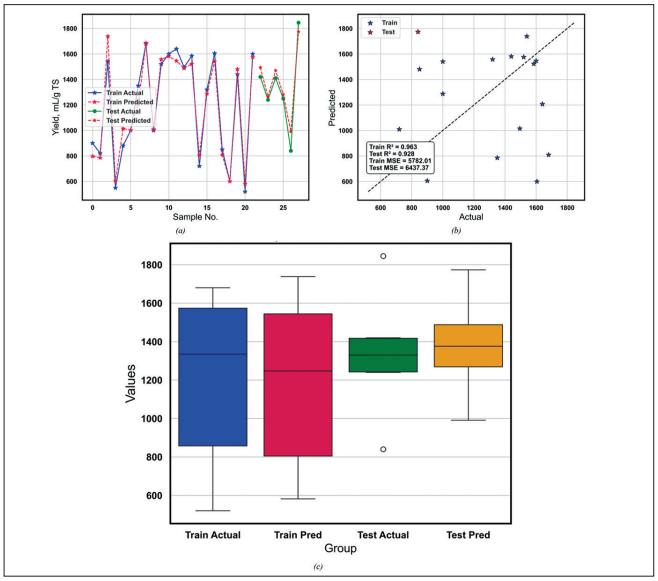

Fig. 7. Results for the XGBoost-based NOx emission model: (a) actual vs predicted values; (b) scatter plot; (c) box plots

CO emission model

The SVR model was effective in modelling the CO emissions of a DME-diesel fuelled compression-ignition engine. The hyperparameters were optimised to ensure maximum performance, as listed in Table 2, with the optimum values identified as C = 120 (in the range 1–500), epsilon = 0.05 (in the range 0.01–0.3) and gamma = scale (out of scale to 0.5). These parameters allowed the model to successfully represent the nonlinearity of the relationships that govern the formation of CO in combustion. The strong predictive capability shown in Table 3 was demonstrated by the performance evaluation. On the training dataset, SVR had an R^2 of 0.966, meaning that 96.6 % of the variance in CO emissions was accounted for, whereas the MSE was almost zero, the MAPE was 0.074 (7.4%), and the KGE had a high value of 0.970, as listed in Table 3. These statistics indicate that the training data were well fitted. When applied to the unseen test set, the model again showed good generalisation, with an R^2 of 0.854 (explaining 85.4% of the variance), an MSE of zero, an MAPE of 0.141, and a KGE of 0.751. The small generalisation to train and test values increases of MAPE and a small decrease of KGE are indicative of reasonable generalisation, where relatively little degradation is experienced in spite of the small size of the dataset and the inherent variability in the emissions measurements. A statistical evaluation of the model is shown in Table 3.

These results are confirmed by the graphs in in Fig. 8. The line plot of the actual and predicted values in Fig. 8(a) (train and test together) shows that the actual values (blue and green lines) and the predicted values (red and orange lines) are close to each other at most of the sample points, with slight differences at the lowest CO concentrations. The scatter plot in Fig. 8(b) indicates that the training data (blue stars) are tightly clustered around the ideal line, whereas the test data (red stars) show an acceptable scatter with an evident linear relationship, in accordance with the value of R2 for the test. A comparison of distributions across groups using boxplots (Fig. 8(c)) implies that the median and interquartile ranges for the values of prediction are similar to those of the actual values for the training and test sets; the dispersion is also similar, and no strong systematic bias is found. In summary, the SVR model is a very successful prediction model for CO emissions from DME-powered engines, as it provides high accuracy on the training data and good performance on the test data. These findings suggest the applicability of this algorithm for multi-objective optimisation models, including SA, in which the aim is to determine the optimal engine operating conditions in terms of reducing CO and other pollutants at the same time, with reasonable efficiency and power output. Since the values of MSE and R^2 are always low and high, respectively, these results confirm that SVR is a good predictive model for use in emission control research involving alternative-fuel combustion engines.

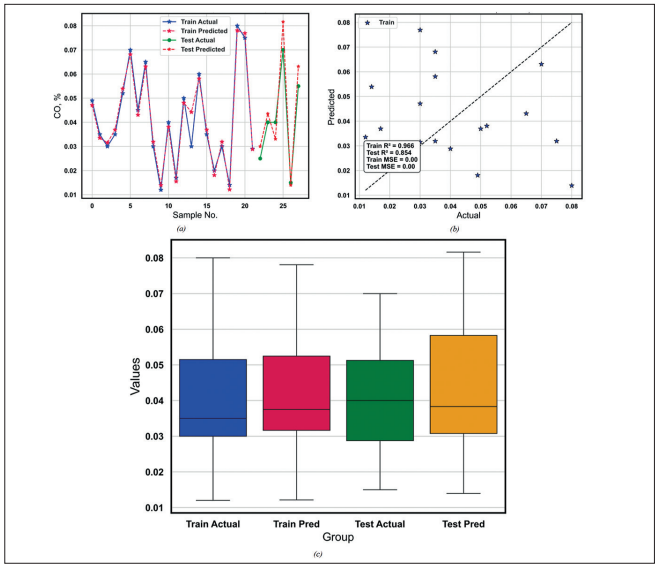

Fig. 8. Results for the SVR-based CO emission model: (a) actual vs. predicted values; (b) scatter plot; (c) box plots

The results for the XGBoost model in terms of predicting CO emissions from the DME-diesel-fuelled compression-ignition engine are shown in Fig. 9. Hyperparameter optimisation was carried out to produce balanced performance, which resulted in the following optimal model: the number of estimators was 280 (from the range 100–600), the learning rate was 0.04 (from the range 0.01–0.2), the maximum tree depth was four (3–8), the subsample ratio was 0.85 (0.6–1.0), and the column sampling

ratio was 0.9 (0.6–1.0). The hyperparameter optimisation results are shown in Table 2. These environments offered a level of model complexity that was adequate to describe the nonlinear and interactive effects on the formation of CO, with no overfitting. The predictive performance was good in quantitative terms, especially on the training data. This model achieved a value for R^2 of 0.855 on the training set, explaining 85.5% of the variance in CO emissions, with a zero MSE, an MAPE of 15.7%, and a KGE of 0.883. On the unseen test set, the performance was also reasonable, with an R^2 of 0.731 (explaining 73.1% of the variance), an MSE that was still close to zero, an MAPE of 0.48, and a KGE of 0.746. The significant rise in the MAPE on the test set indicates that there are certain difficulties in extrapolating to unknown samples, which may be explained by the small size of the dataset, the noise in the measurements of low-level CO, and the overall distribution of the CO values across the operating conditions, which is rather flat. These findings are justified in Fig. 9. The plot of actual versus predicted values in Fig. 9(a) indicates that the actual and predicted values are reasonably close at most of the points in the sample, but at the lowest and highest CO concentrations, the deviation is more significant. The scatter plot in Fig. 9(b) indicates that the training points (blue stars) have moderately good degree of fit to the ideal line, whereas the test points (red stars) are more dispersed, a finding that is consistent with the lower value of R^2 in the test. A comparison of the distributions in the form of boxplots (Fig. 9(c)) shows that both the median and interquartile ranges for the predictive values are similar to those of the actual values in both training and test groups, the general dispersion is similar, and no significant systematic bias is present. Overall, XGBoost has a high CO emission prediction, high training accuracy, and good test generalisation. The findings validate the capability of this model in regard to the modelling of CO behaviour in engines fuelled with DME, although it is limited in the extreme low-concentration variability. These results confirm its suitability for inclusion in multi-objective optimisation algorithms, such as simulated annealing, for finding engine parameter sets that are effective in minimising CO and other pollutants.

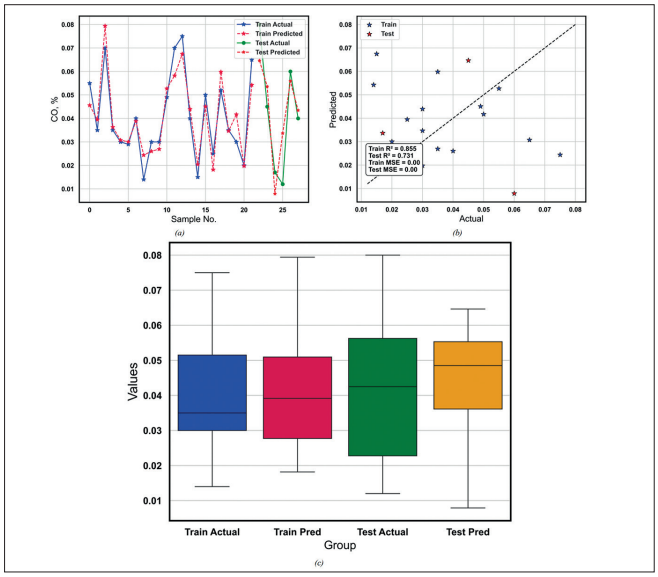

Fig. 9. Results for the XGBoost-based CO emission model: (a) actual vs. predicted values; (b) scatter plot; (c) box plots

HC emission model

The SVR model was applied to predict unburned HC emissions from the DME-diesel-fuelled compression-ignition engine, based on experimental data on the BMEP, blend ratio, and operating parameters. Through hyperparameter optimisation, the RBF kernel was identified as the most effective of the available options (RBF, linear, and polynomial), with optimal values of C = 200 (in the range 1–500), epsilon = 0.1 (in the range 0.01–0.3), and gamma = 0.03 (on a scale to 0.5). These settings enabled the model to accurately capture the complex nonlinear patterns associated with HC formation and oxidation during combustion. The values of the performance metrics indicated outstanding predictive accuracy. On the training dataset, SVR achieved an exceptional value for R^2 of 0.992, explaining 99.2% of the variance in HC emissions, with a very low MSE of 0.46, an MAPE of only 0.018, and a high KGE of 0.967. These results indicate near-perfect fitting to the training samples. On the unseen test set, the model maintained strong generalisation, with an R^2 of 0.854 (explaining 85.4% of the variance), an MSE of 10.89, an MAPE of 0.056 (5.6%), and a KGE of 0.770. The modest increase in MSE and slight reduction in KGE reflect an acceptable level of performance degradation, which is primarily due to the limited dataset size and the inherent variability in low-concentration HC measurements.

The quantitative results are also shown in Fig. 10. The line plot of the actual and predicted values in Fig. 10(a) indicates a very good fit between the actual measurements (blue and green lines) and the predicted values (red and orange lines) at almost all the points in the sample, and only slight deviations are observed at the extreme ends. The scatter plot in Fig. 10(b) shows a close cluster of training points (blue stars) on the optimal line, whereas the test points (red stars) have a reasonable linear relationship with a low level of scatter, a finding that is in line with the high value of R^2 for the test.

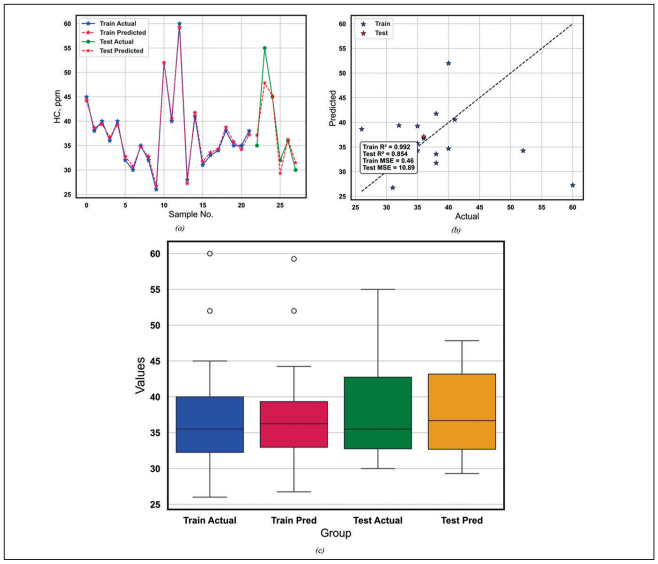

Fig. 10. Results for the SVR-based HC emission model: (a) actual vs. predicted values; (b) scatter plot; (c) box plots

From the boxplots in Fig. 10(c), it can be seen that the interquartile ranges and medians for the predicted values

are very close to those of the real values on both the training and test sets, and the overall distributions are similar, with no major systematic bias. On the whole, it is possible to say that the SVR model is extremely efficient and consistent in the prediction of HC emissions in DME-powered engines. The combination of near-perfect training fit, good test performance, and high visual agreement means that it can be integrated into multi-objective optimisation systems, such as SA, to find the optimal engine parameters to reduce HC, NOx, and CO.

The XGBoost model was built to predict unburned hydrocarbon (HC) emissions from a DME-diesel-fuelled compression-ignition engine based on experimental input data. Hyperparameter tuning yielded an effective configuration consisting of 320 estimators (within the range 100–600), a learning rate value of 0.05 (from the range 0.01–0.2), a maximum tree depth of six (from the range 3–8), a subsample ratio of 0.8 (from the range 0.6–1.0), and a column sampling ratio of 0.75 (from the range 0.6–1.0), as listed in Table 2. These parameters allowed this model to account for important nonlinear interactions that affect the formation of HC while keeping the complexity reasonably low. The values of the performance measures represented a moderate predictive ability. On the training data, the R^2 score for XGBoost was 0.754, representing 75.4% variance in HC emissions, while the MSE was 15.01, the MAPE was 0.085 (8.5%), and KGE was 0.815. On the unseen test set, the model yielded values of $R^2 = 0.686$ (68.6% variance explained), MSE = 15.58, MAPE = 0.098 (9.8%), and KGE = 0.609. The relatively low increase in MSE and small reduction in KGE from the training to the test data indicate acceptable generalisation, although the overall reduction in R^2 indicates difficulties in fully accounting for the variability of low concentration HC emissions, which is likely to be a result of dataset and measurement noise.

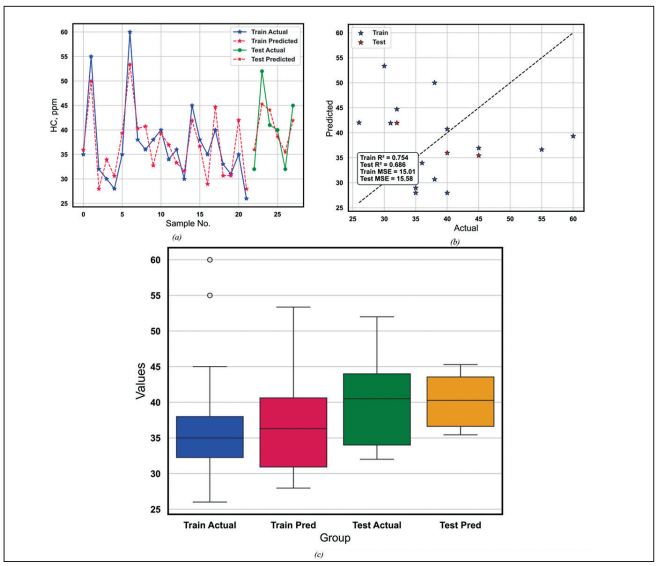

Fig. 11. Results for the XGBoost-based HC emission model: (a) actual vs. predicted values; (b) scatter plot; (c) box plots

A comprehensive visual evaluation of the results is given in Fig. 11. From Fig. 11(a) (line plot of actual vs. predicted results for the training and test datasets), it can be seen that there is good tracking between the actual (blue and green lines) and predicted (red and orange lines) values for the majority of sample points, with some notable deviations mainly at the ends of the HC range. In Fig. 11(b) (scatter plot of actual vs. predicted values), the training points (blue stars) have a moderate level of fit to the ideal 45° dashed line, and the test points (red stars) have a greater amount of scatter, as would also be aligned with the R^2 test.

Fig. 11(c) (boxplot of the actual vs. predicted values) shows that the median and interquartile ranges for the predicted values are similar to those of the actual values on both the training and test sets, with similar overall distributions, despite the presence of a few outliers in the training set. Overall, acceptable performance of the XGBoost model was observed in terms of the predictability of HC emissions, with moderate performance during training, and moderate test generalisation.

Tab. 2. Optimised values of the training hyperparameters

Target	Model	Hyperparameter	Tested Range	Best Value
NOx	SVR	kernel	RBF, linear	RBF
		C	1–500	150
		epsilon	0.01–0.3	0.08
		gamma	scale, 0.01–0.5	0.05
	XGBoost	n_estimators	100–600	350
		learning_rate	0.01–0.2	0.06
		max_depth	3–8	5
		subsample	0.6–1.0	0.9
		colsample_bytree	0.6–1.0	0.8
CO	SVR	kernel	RBF, linear	RBF
		C	1–500	120
		epsilon	0.01–0.3	0.05
		gamma	scale, 0.01–0.5	scale
	XGBoost	n_estimators	100–600	280
		learning_rate	0.01–0.2	0.04
		max_depth	3–8	4
		subsample	0.6–1.0	0.85
		colsample_bytree	0.6–1.0	0.9
HC	SVR	kernel	RBF, linear, poly	RBF
		C	1–500	200
		epsilon	0.01–0.3	0.1
		gamma	scale, 0.01–0.5	0.03
	XGBoost	n_estimators	100–600	320
		learning_rate	0.01–0.2	0.05
		max_depth	3–8	6
		subsample	0.6–1.0	0.8
		colsample_bytree	0.6–1.0	0.75

Tab. 3. Results of a statistical evaluation of the models

Target	Model	Metric	Train	Test
NOx	SVR	R^2	0.982	0.820
		MSE	2246.68	38382.7
		MAPE	0.038	0.175
		KGE	0.970	0.674
	XGBoost	R^2	0.963	0.928
		MSE	5782.01	6437.37
		MAPE	0.054	0.060
		KGE	0.969	0.806
CO	SVR	R^2	0.966	0.854
		MSE	0.00	0.00
		MAPE	0.074	0.141
		KGE	0.970	0.751
	XGBoost	R^2	0.855	0.731
		MSE	0.00	0.00
		MAPE	0.157	0.480
		KGE	0.883	0.746
HC	SVR	R^2	0.992	0.854
		MSE	0.46	10.89
		MAPE	0.018	0.056
		KGE	0.967	0.770
	XGBoost	R^2	0.754	0.686
		MSE	15.01	15.58
		MAPE	0.085	0.098
		KGE	0.815	0.609

CONCLUSION

This research involved the development and validation of a computational framework that integrated advanced ensemble machine learning with SA optimisation. An efficient model prediction/optimisation framework was proposed for the minimisation of emission from DME-diesel-fuelled compression-ignition engines. Comprehensive experimental data from a commercial direct injection diesel engine study were used for modelling and optimisation.

Both the SVR and XGBoost models showed good predictive performance for all types of targeted emissions, but SVR performed especially well for NOx and HC due to its capacity to handle nonlinear relationships well. Correlation and distribution analyses also confirmed load-dominated NOx formation and the positive effect of DME on HC suppression. The use of SA with the trained predictive models allowed the framework to handle multiple optimisation objectives, involving the identification of sets of engine parameters leading to optimal NOx, CO, and HC emissions, while maintaining acceptable thermal efficiency and power output.

This work fills an important gap in the existing literature through the synergistic coupling of high-fidelity emissions prediction with global stochastic optimisation for DME combustion systems. The methodology described here is scalable, computationally efficient and adaptable to other alternative fuels. Future efforts should be directed towards real-time implementation, validation with various engine architectures, and combinations of renewable DME production pathways to further improve the lifecycle sustainability. Ultimately, the proposed approach offers a clean, efficient transportation method that is consistent with global decarbonisation and emissions reduction objectives.

ACKNOWLEDGMENT

The authors wish to express their appreciation to the Ministry of Education and Training Vietnam for supporting this research under the project B2026-DNA-09, entitled "Investigation and Development of an Extended-Range Electric Motorcycle (EREM) Powered by a Hydrogen carrier fuel Engine".

REFERENCES

1. Kowalski J, Tarelko W. NOx emission from a two-stroke ship engine. Part 1: Modeling aspect. Appl Therm Eng 2009, vol. 29, pp. 2153–9. https://doi.org/10.1016/j.applthermaleng.2008.06.032

2. Pham TN, Nguyen HP, Nguyen VN, Sirichokpokin P, Hadiyanto H, Rudzki K, et al. Application of hybrid fuzzy-TOPSIS approach in enhancing green logistics towards sustainable maritime. Polish Marit Res 2025, vol. 32, pp. 156–74. https://doi.org/10.2478/pomr-2025-0030

3. Cao DN, Johnson AJT. A simulation study on a premixed-charge compression ignition mode-based engine using a blend of biodiesel/diesel fuel under a split injection strategy. Int J Adv Sci Eng Inf Technol 2024, vol. 14, pp. 451–71. https://doi.org/10.18517/ijaseit.14.2.20007

4. Hoang AT, Nayak SK, Vujanović M, Guerrero-Pérez MO, Rodríguez-Castellón E, López-Escalante MC, et al. Towards sustainable development goals: Application of hydrogen-enriched mahua biodiesel/diesel blend to dual-fuel diesel engine. Glob Challenges 2025, vol. 9. https://doi.org/10.1002/gch2.202500260

5. Wei C, Jiang G, Wu G, Zhou Y, Liu Y. Effects on of blended biodiesel and heavy oil on engine combustion and black carbon emissions of a low-speed two-stroke engine. Polish Marit Res 2024, vol. 31, pp. 94–101. https://doi.org/10.2478/pomr-2024-0010

6. Othman MF, Adam A, Najafi G, Mamat R. Green fuel as alternative fuel for diesel engine: A review. Renew Sustain Energy Rev 2017, vol. 80, pp. 694–709. https://doi.org/10.1016/j.rser.2017.05.140

7. Zeńczak W, Łuszczynski D. Biomass as an alternative fuel—Selected aspects. Polish Marit Res 2025, vol. 32, pp. 147–53. https://doi.org/10.2478/pomr-2025-0042

8. Nguyen VN, Sharma P, Kumar A, Pham MT, Le HC, Truong TH, et al. Optimization of biodiesel production from Nahar oil using Box-Behnken design, ANOVA and grey wolf optimizer. Int J Renew Energy Dev 2023, vol. 12, pp. 711–9. https://doi.org/10.14710/ijred.2023.54941

9. Nguyen H-B, Nguyen V-T-A, Ly V-D, Bui T-A. Biodiesel produced from pangasius oil operating a diesel engine: Case Study in Vietnam. Int J Adv Sci Eng Inf Technol 2022, vol. 12, pp. 477–82. https://doi.org/10.18517/ijaseit.12.2.16159

10. Cao W, Geng P, Xu X, Tarasiuk T. Energy management strategy considering energy storage system degradation for hydrogen fuel cell ship. Polish Marit Res 2023, vol. 30, pp. 95–104. https://doi.org/10.2478/pomr-2023-0025

11. Hoang AT, Bora BJ, Huynh DNL, Nguyen DKP, Le VV. Dual-fuel diesel engine features powered by ammonia under various pilot-fuel injection timing: Comprehensive analysis and response surface methodology-based optimization. Int J Engine Res 2025. https://doi.org/10.1177/14680874251341036

12. Hoang AT, Chen W-H, López-Escalante MC, Guerrero-Pérez MO, Rodríguez-Castellón E, Kowalski J, et al. Methanol for decarbonization of the maritime sector: From ideological strategy to practical solutions. Renew Sustain Energy Rev 2026, vol. 229, p. 116638. https://doi.org/10.1016/j.rser.2025.116638

13. Zastempowski M, Kaszkowiak J, Jablonicky J, Hujo L, Gromnicki M. Effect of using gasoline with bioethanol on power, torque and selected exhaust gas components: Case study of a small spark-ignition outboard engine. Polish Marit Res 2025, vol. 32, pp. 95–102. https://doi.org/10.2478/pomr-2025-0009

14. Wei C, Chen X-X, Chen L, Li P, Liu M. Effects of pilot oil injection timing on combustion, covariance and knocking of a natural gas–diesel duel-fuel low-speed engine. Polish Marit Res 2024, vol. 31, pp. 69–75. https://doi.org/10.2478/pomr-2024-0051

15. Bui VG, Bui TMT, Nguyen MT, Bui VH, Do PN, Tran NAH, et al. Enhancing the performance of syngas-diesel dual-fuel engines by optimizing injection regimes: From comparative analysis to control strategy proposal. Process Saf Environ Prot 2024, vol. 186, pp. 1034–52. https://doi.org/10.1016/j.psep.2024.04.042

16. Bui VG, Tu Bui TM, Ong HC, Nižetić S, Bui VH, Xuan Nguyen TT, et al. Optimizing operation parameters of a spark-ignition engine fueled with biogas-hydrogen blend integrated into biomass-solar hybrid renewable energy system. Energy 2022, vol. 252, p. 124052. https://doi.org/10.1016/j.energy.2022.124052

17. Nguyen TA, Pham TY, Le HC, Nguyen VG, Nguyen LH. Exploring the feasibility of dimethyl ether (DME) and LPG fuel blend for small diesel engine: A simulation perspective. Int J Renew Energy Dev 2024, vol. 13, pp. 559–71. https://doi.org/10.61435/ijred.2024.60250

18. Doan QB, Nguyen XP, Pham VV, Dong TMH, Pham MT, Le TS. Performance and emission characteristics of diesel engine using ether additives: A review. Int J Renew Energy Dev 2022, vol. 11, pp. 255–74. https://doi.org/10.14710/ijred.2022.42522

19. Anggarani R, Aisyah L, Wibowo CS, Nugroho YS, Dhiputra IMK. Experimental comparison of working region, flame stability, and flame height of LPG, DME, and DME-mixed LPG in an atmospheric diffusion cylindrical burner. Int J Technol 2020, vol. 11, p. 400. https://doi.org/10.14716/ijtech.v11i2.3914

20. Sun C, Liu Y, Qiao X, Ju D, Tang Q, Fang X, et al. Experimental study of effects of exhaust gas recirculation on combustion, performance, and emissions of DME-biodiesel fueled engine. Energy 2020, vol. 197, p. 117233. https://doi.org/10.1016/j.energy.2020.117233

21. Le TT, Balasubramanian D, Le AT, Cao DN, Chen W-H, Venugopal IP, et al. Partially-charged advanced low-temperature combustion in diesel engine: Progress and prospects. Alexandria Eng J 2025, vol. 129, pp. 373–440. https://doi.org/10.1016/j.aej.2025.06.021

22. Arcoumanis C, Bae C, Crookes R, Kinoshita E. The potential of di-methyl ether (DME) as an alternative fuel for compression-ignition engines: A review. Fuel 2008, vol. 87, pp. 1014–30. https://doi.org/10.1016/j.fuel.2007.06.007

23. Jie L, Shenghua L, Yi L, Yanju W, Guangle L, Zan Z. Regulated and nonregulated emissions from a dimethyl ether powered compression ignition engine. Energy & Fuels 2010, vol. 24, pp. 2465–9. https://doi.org/10.1021/ef9016043

24. Park SH, Kim HJ, Lee CS. Effects of dimethyl-ether (DME) spray behavior in the cylinder on the combustion and exhaust emissions characteristics of a high speed diesel engine. Fuel Process Technol 2010, vol. 91, pp. 504–13.

25. Yoon SH, Cha JP, Lee CS. An investigation of the effects of spray angle and injection strategy on dimethyl ether (DME) combustion and exhaust emission characteristics in a common-rail diesel engine. Fuel Process Technol 2010, vol. 91, pp. 1364–72. https://doi.org/10.1016/j.fuproc.2010.04.017

26. Namasivayam AM, Korakianitis T, Crookes RJ, Bob-Manuel KDH, Olsen J. Biodiesel, emulsified biodiesel and dimethyl ether as pilot fuels for natural gas fuelled engines. Appl Energy 2010, vol. 87, pp. 769–78. https://doi.org/10.1016/j.apenergy.2009.09.014

27. Soltic P, Hilfiker T, Wright Y, Hardy G, Fröhlich B, Klein D. The potential of dimethyl ether (DME) to meet current and future emissions standards in heavy-duty compression-ignition engines. Fuel 2024, vol. 355, p. 129357. https://doi.org/10.1016/j.fuel.2023.129357

28. Preuß J, Munch K, Denbratt I. Performance and emissions of renewable blends with OME3-5 and HVO in heavy duty and light duty compression ignition engines. Fuel 2021, vol. 303, p. 121275. https://doi.org/10.1016/j.fuel.2021.121275

29. Gelner AD, Pang GA, Weber M, Haisch C, Beck HA, Pastoetter C, et al. Gaseous emissions of a heavy-duty engine fueled with polyoxymethylene dimethyl ethers (OME) in transient cold-start operation and methods for after-treatment system heating. Environ Sci Adv 2022, vol. 1, pp. 470–82. https://doi.org/10.1039/D2VA00080F

30. Liu H, Wang Z, Zhang J, Wang J, Shuai S. Study on combustion and emission characteristics of polyoxymethylene dimethyl ethers/diesel blends in light-duty and heavy-duty diesel engines. Appl Energy 2017, vol. 185, pp. 1393–402. https://doi.org/10.1016/j.apenergy.2015.10.183

31. Shojae K, Mahdavian M. Modification of piston bowl geometry and injection strategy, and investigation of EGR composition for a DME-burning direct injection engine. Adv Environ Technol 2017, vol. 3, pp. 1–10. https://doi.org/10.22104/aet.2017.503

32. Benajes J, Novella R, Pastor JM, Hernández-López A, Kokjohn S. Computational optimization of a combustion system for a stoichiometric DME fueled compression ignition engine. Fuel 2018, vol. 223, pp. 20–31. https://doi.org/10.1016/j.fuel.2018.03.022

33. Jamei M, Sharma P, Ali M, Bora BJ, Malik A, Paramasivam P, et al. Application of an explainable glass-box machine learning approach for prognostic analysis of a biogas-powered small agriculture engine. Energy 2024, vol. 288, p. 129862. https://doi.org/10.1016/j.energy.2023.129862

34. Sunil Kumar K, Surakasi R, Patro SGK, Govil N, Ramis MK, Razak A, et al. Performance, combustion, and emission analysis of diesel engine fuelled with pyrolysis oil blends and n-propyl alcohol-RSM optimization and ML modelling. J Clean Prod 2024, vol. 434, p. 140354. https://doi.org/10.1016/j.jclepro.2023.140354

35. Tuan Hoang A, Nižetić S, Chyuan Ong H, Tarelko W, Viet Pham V, Hieu Le T, et al. A review on application of artificial neural network (ANN) for performance and emission characteristics of diesel engine fueled with biodiesel-based fuels. Sustain Energy Technol Assessments 2021, vol. 47, p. 101416. https://doi.org/10.1016/j.seta.2021.101416

36. Tarelko W, Rudzki K. Applying artificial neural networks for modelling ship speed and fuel consumption. Neural Comput Appl 2020, p. 1–17. https://doi.org/10.1007/s00521-020-05111-2

37. Zhang E, Zhou F, Xi H, Duan X, Liu J. Predicting cycle-to-cycle variations in liquid methane engines using CTGAN-augmented machine learning. J Mar Sci Eng 2025, vol. 13, p. 1513. https://doi.org/10.3390/jmse13081513

38. Mohammedali AAM, Albadwi A, Omara AAM, Ali K, Ali MIH. Machine learning prediction and evolutionary multi-objective optimization on performance and emissions of a syngas-diesel dual-fuel engine. Int J Hydrogen Energy 2025, vol. 162, p. 150781. https://doi.org/10.1016/j.ijhydene.2025.150781

39. Nachippan M, Pathmanabhan P, Nagappan B, Upadhye VJ, Kaliappan N, Balaji V, et al. Machine learning predictions for enhancing engine performance and emission using aluminum oxide nano additives in castor biodiesel. Sci Rep 2025, vol. 15, p. 36514. https://doi.org/10.1038/s41598-025-02388-y

40. Godwin DJ, Varuvel EG, Martin MLJ. Prediction of combustion, performance, and emission parameters of ethanol powered spark ignition engine using ensemble least squares boosting machine learning algorithms. J Clean Prod 2023, vol. 421, p. 138401. https://doi.org/10.1016/j.jclepro.2023.138401

41. Sharma P, Sahoo BB. 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 2022, vol. 321, p. 124131. https://doi.org/10.1016/j.fuel.2022.124131

42. Sharma P, Sahoo BB, Said Z, Hadiyanto H, Nguyen XP, Nižetić S, et al. Application of machine learning and Box-Behnken design in optimizing engine characteristics operated with a dual-fuel mode of algal biodiesel and waste-derived biogas. Int J Hydrogen Energy 2023, vol. 48, pp. 6738–60. https://doi.org/10.1016/j.ijhydene.2022.04.152

43. Bhowmik S, Panua R, Debroy D, Paul A. Artificial neural network prediction of diesel engine performance and emission fueled with diesel–kerosene–ethanol blends: A fuzzy-based optimization. J Energy Resour Technol 2017, vol. 139. https://doi.org/10.1115/1.4035886

44. Hoang AT, Chen W-H, Olga Guerrero-Pérez M, Rodríguez-Castellón E, López-Escalante MC, Huynh DNL, et al. Machine learning and simulated annealing/moth-flame based prediction-optimization for improving ultrasonication-assisted co-anaerobic digestion. Energy 2026, vol. 344, p. 139831. https://doi.org/10.1016/j.energy.2025.139831

45. Asadi J, Jalali Farahani F. Optimization of dimethyl ether production process based on sustainability criteria using

a homotopy continuation method. Comput Chem Eng 2018, vol. 115, pp. 161–78. https://doi.org/10.1016/j.compchemeng.2018.03.014

46. Bashiri S, Yasari E, Tayyebi S. Comparison of different sampling and surrogate modelling approaches for a multi-objective optimization problem of direct dimethyl ether synthesis in the fixed-bed reactor. Chemom Intell Lab Syst 2022, vol. 230, p. 104683. https://doi.org/10.1016/J.CHEMOLAB.2022.104683

47. Ying W, Genbao L, Wei Z, Longbao Z. Study on the application of DME/diesel blends in a diesel engine. Fuel Process Technol 2008, vol. 89, pp. 1272–80. https://doi.org/https://dx.doi.org/10.1016/j.fuproc.2008.05.023

48. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. Science (80-) 1983, vol. 220, pp. 671–80. https://doi.org/10.1126/science.220.4598.671

49. Sylejmani K, Gashi E, Ymeri A. Simulated annealing with penalization for university course timetabling. J Sched 2023, vol. 26, pp. 497–517. https://doi.org/10.1007/s10951-022-00747-5

50. Dubey K, Sharma SC, Nasr AA. A simulated annealing based energy-efficient VM placement policy in cloud computing. 2020 Int. Conf. Emerg. Trends Inf. Technol. Eng., IEEE; 2020, pp. 1–5. https://doi.org/10.1109/ic-ETITE47903.2020.119

51. Liu C, Zhang F, Zhang H, Shi Z, Zhu H. Optimization of assembly sequence of building components based on simulated annealing genetic algorithm. Alexandria Eng J 2023, vol. 62, pp. 257–68. https://doi.org/10.1016/j.aej.2022.07.025

52. Karaboga D, Akay B. A comparative study of artificial bee colony algorithm. Appl Math Comput 2009, vol. 214, pp. 108–32. https://doi.org/10.1016/j.amc.2009.03.090

53. Karaboga D, Gorkemli B, Ozturk C, Karaboga N. A comprehensive survey: Artificial bee colony (ABC) algorithm and applications. Artif Intell Rev 2014, vol. 42, pp. 21–57. https://doi.org/10.1007/s10462-012-9328-0

54. Luo J, Liu Q, Yang Y, Li X, Chen M, Cao W. An artificial bee colony algorithm for multi-objective optimisation. Appl Soft Comput 2017, vol. 50, pp. 235–51. https://doi.org/10.1016/j.asoc.2016.11.014

55. Karaboga D, Basturk B. Artificial bee colony (ABC) Optimization algorithm for solving constrained optimization problems. Found Fuzzy Log Soft Comput, Berlin, Heidelberg: Springer Berlin Heidelberg; n.d., p. 789–98. https://doi.org/10.1007/978-3-540-72950-1_77

56. Sonmez M, Akgüngör AP, Bektaş S. Estimating transportation energy demand in Turkey using the artificial bee colony algorithm. Energy 2017, vol. 122, pp. 301–10. https://doi.org/10.1016/j.energy.2017.01.074

57. Shirneshan A, Samani BH, Ghobadian B. Optimization of biodiesel percentage in fuel mixture and engine operating conditions for diesel engine performance and emission characteristics by artificial bees colony algorithm. Fuel 2016, vol. 184, pp. 518–26. https://doi.org/10.1016/j.fuel.2016.06.117

58. Sabzekar M, Hasheminejad SMH. Robust regression using support vector regressions. Chaos, Solitons & Fractals 2021, vol. 144, p. 110738. https://doi.org/10.1016/j.chaos.2021.110738

59. Dong H, Jin X, Lou Y, Wang C. Lithium-ion battery state of health monitoring and remaining useful life prediction based on support vector regression-particle filter. J Power Sources 2014, vol. 271, pp. 114–23. https://doi.org/10.1016/j.jpowsour.2014.07.176

60. Ahmad MW, Reynolds J, Rezgui Y. Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. J Clean Prod 2018, vol. 203, pp. 810–21. https://doi.org/10.1016/j.jclepro.2018.08.207

61. Huang H, Wei X, Zhou Y. An overview on twin support vector regression. Neurocomputing 2022, vol. 490, pp. 80–92. https://doi.org/10.1016/j.neucom.2021.10.125

62. Ban B, Yang J, Chen P, Xiong J, Wang Q. Ship track regression based on support vector machine. IEEE Access 2017. https://doi.org/10.1109/ACCESS.2017.2749260

63. Kumar K P, Alruqi M, Hanafi HA, Sharma P, Wanatasanappan VV. Effect of particle size on second law of thermodynamics analysis of Al2O3 nanofluid: Application of XGBoost and gradient boosting regression for prognostic analysis. Int J Therm Sci 2024, vol. 197, p. 108825. https://doi.org/10.1016/j.ijthermalsci.2023.108825

64. Liu Y, Dong Y, Jiang Z, Li Y, Zha W, Du Y, et al. XGBoost-based model for predicting hydrogen content in electroslag remelting. J Iron Steel Res Int 2023, vol. 30, pp. 887–96. https://doi.org/10.1007/s42243-023-00962-0

65. Bae D-J, Kwon B-S, Song K-B. XGBoost-based day-ahead load forecasting algorithm considering behind-the-meter solar PV generation. Energies 2021, vol. 15, p. 128. https://doi.org/10.3390/en15010128

66. Han P, Liu Z, Sun Z, Yan C. A novel prediction model for ship fuel consumption considering shipping data privacy: An XGBoost-IGWO-LSTM-based personalized federated learning approach. Ocean Eng 2024, vol. 302, p. 117668. https://doi.org/10.1016/j.oceaneng.2024.117668