Scientific Reports

https://doi.org/10.1038/s41598-026-64316-y

Article in Press

Machine learning based prediction of TOPSIS scores for internal combustion engine performance and emissions

Received: 19 March 2026

Accepted: 24 July 2026

Published online: 07 August 2026

Cite this article as: Karaoğlu A., Demirtaş G. & Söyler H. Machine learning based prediction of TOPSIS scores for internal combustion engine performance and emissions. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-64316-y

Ahmet Karaoğlu, Güven Demirtaş & Hüseyin Söyler

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Machine Learning Based Prediction of TOPSIS Scores for Internal Combustion Engine Performance and Emissions

Ahmet Karaoğlu [a], Güven Demirtaş [b*], Hüseyin Söyler [c]

[a] *Department of Computer Engineering, Sinop University, 57000, Sinop, Turkey*

akaraoglu@sinop.edu.tr

[b] *Akdağmadeni Vocational School, Yozgat Bozok University, 66000, Yozgat, Turkey*

guven.demirtas@bozok.edu.tr

[c] *Gerze Vocational School, Sinop University, 57000, Sinop, Turkey*

hsoyler@sinop.edu.tr

Abstract

Internal combustion engines operating on alternative fuels face an inherent challenge: optimizing power output and fuel efficiency while simultaneously minimizing exhaust emissions requires the simultaneous management of multiple, often conflicting objectives. Although multi-criteria decision-making (MCDM) and machine learning (ML) methods have each been applied extensively in engine research, they are typically used in isolation, and the direct prediction of a composite MCDM decision score from engine control parameters has not previously been demonstrated. This study addresses that gap by proposing a hybrid TOPSIS-Stacking framework in which a TOPSIS-based composite score, aggregating eight performance and emission indicators into a single figure of merit, is itself treated as a continuous regression target. Experimental data from a single-cylinder Otto-cycle engine operating across a range of ethanol-water blend ratios (0 to 50% water by volume), two compression ratios (7.44:1 and 9.44:1), and five rotational speeds (2000 to 4000 rpm) were used. Entropy-based weighting was adopted to assign objective criterion importance, and three regression models (RSM, XGBoost, and a stacking ensemble combining both) were trained to predict the TOPSIS score from four engine control parameters. A 5-fold Group Cross-Validation strategy was implemented to prevent repetition leakage between experimental replicates and ensure conservative generalization estimates. The stacking ensemble achieved the highest predictive accuracy (R^2 = 0.814, RMSE = 0.061, MAPE = 6.36%), outperforming both individual models. Feature importance analysis identified water fraction as the

ARTICLE IN PRESS

dominant predictor, followed by compression ratio, consistent with their thermodynamic influence on NO_x formation and thermal efficiency. These results demonstrate that the proposed framework provides a reliable and computationally efficient tool for virtual engine calibration and scenario-based performance evaluation, significantly reducing the need for exhaustive multi-criteria experimental campaigns.

Keywords: Internal Combustion Engine, TOPSIS, Ensemble Learning, Artificial intelligence Engine Performance Modeling, Multi-Criteria Decision Making, Artificial intelligence (Industry innovation and infrastructure)

1. Introduction

Internal combustion engines (ICEs) remain crucial in transportation, with 99.8% of global transport powered by ICEs as of 2020. Despite alternatives like battery electric vehicles, ICEs are expected to dominate through 2040. Improving ICE efficiency and reducing emissions is vital for environmental impact mitigation. Advanced technologies can potentially reduce fuel consumption by 50% in light-duty vehicles [1]. Emissions from ICEs include NO_x, CO, CO_2, hydrocarbons, and particulate matter [2]. These emissions result from complex interactions between fuel properties and engine parameters [3].

The chemical properties of the fuel, combustion quality, and engine design parameters have a direct impact on engine performance and emission levels [4]. Therefore, improving both performance and reducing emissions is essential to ensure engines operate efficiently while minimizing their environmental footprint. At the same time, reducing fuel consumption is important not only for lowering operational costs but also for decreasing dependence on fossil fuels and mitigating environmental impact [5,6]. Achieving this balance under real-world operating conditions has become a critical goal in modern engineering. Especially today, with increasing environmental pressures and the growing use of alternative fuels, integrated and data-driven approaches are needed to meet these challenges effectively [7].

In today's world, growing environmental concerns, energy security challenges, and the finite nature of fossil fuel resources have intensified the demand for cleaner and more sustainable alternative fuels in ICEs. In this context, renewable fuels such as biodiesel and alcohol-based fuels have gained significant attention [8].

ARTICLE IN PRESS

Bioalcohols are an important group of renewable fuels that provide sustainable energy in ICEs. Short-chain alcohols such as methanol, ethanol, propanol, butanol, and octanol are promising alternatives to fossil fuels for internal combustion engines due to their high octane numbers and clean combustion potential [9–11]. These bioalcohols can be produced from biomass through various thermochemical and biochemical processes. When blended with gasoline, they generally improve engine performance, increasing brake thermal efficiency and reducing emissions [9,10]. Isobutanol, in particular, offers advantages over ethanol, including higher energy content and lower hygroscopicity [12]. However, each alcohol has unique properties that affect its suitability for different engine types. For example, ethanol is more suitable for spark-ignition engines, while octanol is better for compression-ignition engines. Recent experimental studies demonstrate that while ethanol blends effectively reduce soot and NO_x emissions, they can simultaneously increase carbon monoxide (CO) and unburned hydrocarbon (THC) emissions depending on the combustion mode [13]. This proves that the application of ethanol in internal combustion engines creates unavoidable trade-offs between performance and emission targets. Further research is needed to address challenges in large-scale production, purification, and commercial viability of these biofuels [14].

ICEs convert the chemical energy of fuels into mechanical power through controlled combustion processes. Key performance parameters of ICEs include brake power, torque, brake thermal efficiency (BTE), and brake specific fuel consumption (BSFC). These parameters are closely tied to the efficiency of combustion and are influenced by both the physical properties of the fuel and engine design characteristics.

Understanding the complex relationships among these variables is crucial for the development of modern, environmentally friendly, and high-efficiency engine technologies. In this context, both experimental and computational methods are employed to analyze fuel–engine interactions and to optimize engine behavior. Recently, the integration of artificial intelligence (AI) and machine learning (ML) techniques into combustion studies has enabled more effective multivariate modeling, allowing for faster and more accurate predictions of engine performance under various fuel compositions and operating conditions [15].

Recent research highlights the limitations of traditional ICE modelling approaches and explores ML techniques as promising alternatives [16]. Aliramezani et al. review various ML methods for

ARTICLE IN PRESS

ICE modeling, optimization, and control, proposing grey-box approaches that combine physics-based and ML models [17]. Salam et al. discuss analytical, regression, and artificial neural network methods for engine research, recommending the integration of advanced statistical and ML algorithms [18]. Huang et al. demonstrate the effectiveness of artificial neural networks (ANN) in modelling engine performance and emissions, accurately capturing complex relationships between operating variables and engine responses [19]. Soltanalizadeh et al. propose hybrid deep learning approaches, combining Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) networks to model engine emissions under both warm and cold operating conditions, demonstrating improved predictive accuracy compared to conventional methods. These studies emphasize the potential of ML techniques to address the complexities of ICE modelling and optimization, offering more flexible and accurate solutions than traditional empirical or thermodynamic approaches [19].

Recent advancements in ML have significantly improved engine modelling and analysis. ML techniques, including artificial neural networks and deep learning, have demonstrated superior performance in predicting engine behaviour and in-cylinder pressure [20,21]. These approaches offer advantages over traditional methods, particularly when large experimental datasets are available [17]. Hybrid models combining analytical functions with ML techniques have shown enhanced accuracy and generalization across dynamic engine operating conditions [20]. ML applications in engine research span various areas, including combustion analysis, emissions modelling, and fault diagnostics [17,22]. The integration of ML with physics-based models, known as grey-box approaches, has been proposed as a promising solution for robust and high-fidelity engine modelling and control [22]. These advancements in ML-based engine analysis are accelerating optimization processes and improving our understanding of complex in-cylinder phenomena [21,22]. Advanced data-driven techniques are particularly valuable for interpreting these phenomena, such as the high-frequency cylinder pressure fluctuations and knock limits associated with novel prechamber combustion systems [23].

Furthermore, the evaluation of engine performance must not rely solely on predictive accuracy; it should also incorporate trade-offs among multiple criteria such as power output, fuel efficiency, and emissions. For instance, recent research has demonstrated that advanced optimization schedulers, such as multi-objective game theory (MOGT) and genetic algorithms, can effectively

ARTICLE IN PRESS

manage these complex trade-offs by simultaneously maximizing work output and minimizing NO emissions and entropy generation in alternative dual-fuel engines [24]. Similarly, hybrid optimization strategies integrating Design of Experiment (DoE) methodologies with evolutionary algorithms, such as NSGA-II, have proven highly effective in accelerating the convergence toward optimal trade-offs between performance maximization and NO_x emission reduction [25]. In this regard, multi-criteria decision-making (MCDM) techniques, such as the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), offer a systematic framework for balancing conflicting objectives. For instance, recent research has successfully applied MCDM methods such as the Analytic Hierarchy Process (AHP) to identify optimal injection strategies in complex tri-fuel engine configurations by systematically evaluating the critical trade-offs between thermal efficiency and emissions [26]. By integrating ML-based prediction with MCDM-based evaluation, a hybrid data-driven strategy can be formulated capable of not only forecasting engine parameters with high precision but also supporting optimal decision-making across diverse operating scenarios [27].

Although MCDM techniques and ML models have each individually matured within engine research, their genuine integration, in which an ML model is trained to learn and reproduce a composite MCDM-based decision score directly from control parameters, remains a largely unexplored research direction. The present study is positioned to close this specific gap. Three converging but distinct strands of recent literature illustrate this point. First, several studies apply AI exclusively to the multi-objective prediction or optimization of raw performance and emission targets without aggregating them into a single, learnable decision metric. For instance, Zhu et al. [28] proposed an artificial intelligence based multi-objective optimization strategy for an Otto-cycle engine operating under ethanol/water blends, the same experimental envelope investigated in the present work, yet their framework outputs separate optimized operating points for individual objectives rather than a unified score that can itself be regressed from control parameters. Second, a substantial body of work couples response-surface or AI-based surrogate models with evolutionary optimization and subsequently applies TOPSIS only as a final selection step over a small set of Pareto-optimal solutions. A representative example is the RSM-NSGA-III-TOPSIS framework of Gao et al. [29] for diesel/PODE3/n-butanol blends and related plateau-environment diesel studies, in which TOPSIS ranks fewer than a dozen candidate solutions rather than being learned as a continuous function of the operating envelope. Third, outside the combustion engine

ARTICLE IN PRESS

domain, a small number of recent studies have begun coupling stacking ensembles with TOPSIS-based MCDM. Hannah Grace [30], for example, applied TOPSIS together with MEREC-based weighting to rank the performance of stacked classification models for heart-disease prediction. In that and similar studies, however, TOPSIS evaluates competing ML models rather than serving as the prediction target of an ML model itself. In contrast to all three of these strands, the present study computes a single TOPSIS-based decision score across the full set of 360 experimental alternatives and then trains a stacking ensemble combining RSM and XGBoost to directly predict that composite score from four engine control parameters under a leakage-safe Group Cross-Validation protocol. To the best of the authors' knowledge, this is the first demonstration that a unified MCDM decision score for an internal combustion engine can itself be treated as a continuous, learnable regression target, enabling virtual estimation of overall engine performance and emission trade-offs for unseen operating conditions without re-running the full multi-criteria evaluation.

It is important to note that the novelty of this study does not reside in the invention of a new algorithm, a new MCDM method, or a new engine model. Rather, it resides in a new research formulation: demonstrating for the first time that a composite MCDM decision score, derived from eight conflicting engine performance and emission criteria, can itself be treated as a continuous, learnable regression target and directly predicted from engine control parameters. This formulation transforms a one-time, data-dependent ranking exercise into a transferable predictive function, and it has not been demonstrated in the engine domain or in any MCDM context involving a stacking ensemble as the primary prediction instrument.

In summary, the primary novelty and contributions of this study are:

- Integrating classical TOPSIS with a stacking ensemble (RSM + XGBoost) to predict a unified engine performance score directly from control parameters.
- Treating an MCDM decision score as a continuous, learnable regression target for the first time in internal combustion engine optimization.
- Providing a data-driven virtual calibration tool that significantly reduces the reliance on exhaustive multi-criteria experimental evaluations for unseen operating conditions.

2. Materials and methods

ARTICLE IN PRESS

2.1.Data Set

The dataset used in this study was obtained from an experimental investigation conducted by Torres et al. [31]. It contains engine performance and emission data collected from a single-cylinder Otto cycle internal combustion engine operating under various conditions, including different ethanol/water blend ratios (from 0% to 50% water by volume) and two compression ratios (7.44:1 and 9.44:1). The experiments were performed at five engine speeds ranging from 2000 to 4000 rpm, and each operating condition was repeated six times, resulting in a total of 360 data points.

The dataset is publicly available through the Mendeley Data repository, DOI 10.17632/yvb7khhbrj.1, published 23 January 2024, under a Creative Commons Attribution 4.0 license, with Darlan Jorge Gil Torres, André de Souza Mendes, and Cyro Albuquerque listed as the original data contributors and owners [31]. The present authors hold no ownership claim over the raw experimental data and used it strictly as a secondary, openly licensed data source for the development and validation of the proposed TOPSIS based decision support and machine learning framework. All original measurements were collected independently of, and prior to, the present study, and proper attribution to the original data source is maintained throughout this manuscript via citation [31], in accordance with the terms of the license.

ARTICLE IN PRESS

The dataset includes both directly measured and computed engine parameters. The complete list of features used in this study is as follows:

- Performance metrics: Torque (Nm), Power (kW), Specific Fuel Consumption (g/kWh), Efficiency (%)
- Thermal parameters: Fuel Temperature (°C), Exhaust Temperature (°C), Oil Temperature (°C)
- Emissions: CO (%), CO_2 (%), O_2 (%), HC (ppm), NO_x (ppm)
- Experimental setup: Water fraction (0–0.5), Rotational speed (rpm), Compression ratio, Ignition advance angle (BTDC)

The descriptive statistics of the main engine performance and emission parameters are presented in Table 1. As observed, the mean power output is approximately 3.38 kW with a standard deviation of 0.97 kW, while torque averages at 10.72 Nm. The wide standard deviation observed in specific fuel consumption (302.44 g/kWh) indicates substantial variability across different

ARTICLE IN PRESS

engine operating conditions and water blending ratios. Likewise, NO_x emissions exhibit a large range, with values spanning from 0 to 2733.2 ppm, reflecting the sensitivity of NO_x formation to combustion temperature and oxygen availability. The relatively high standard deviation in hydrocarbon emissions (HC) also suggests unstable combustion events under certain conditions. These values provide a preliminary understanding of the variability and central tendencies of the dataset and support subsequent modelling and decision-making processes based on this experimental data.

Table 1. Descriptive statistics of the main engine performance and emission parameters under varying ethanol water blend ratios and compression ratios.

Parameter	**Mean**	**Std. Dev.**	**Min**	**Max**	**Outlier Count**
Power (kW)	3.38	0.97	1.76	5.09	0
Torque (Nm)	10.72	1.34	5.70	13.47	11
Fuel Consumption (g/s)	0.55	0.23	0.21	1.12	1
BSFC (g/kWh)	614.83	302.44	265.00	1687.00	11
Efficiency (%)	40.03	5.52	21.31	47.71	5
CO (%)	0.60	0.13	0.16	0.98	1
CO_2 (%)	10.04	1.40	7.30	12.60	0
O_2 (%)	14.93	1.39	12.20	18.00	0
HC (ppm)	288.78	429.03	8.00	1520.00	10
NO_x (ppm)	580.01	689.21	0.00	2733.20	11

The distribution of outliers detected in each parameter using the Z-score method (|Z-score| > 3) indicates that specific fuel consumption, torque, and NO_x exhibit the highest number of extreme values, each with 11 outliers, reflecting their high variability and sensitivity to experimental conditions. The HC and efficiency parameters follow with 10 and 5 outliers, respectively, suggesting a notable influence on the overall evaluation, while the remaining variables show negligible or no outlier presence, indicating more stable statistical behaviour. Importantly, these outliers particularly in emission-related parameters such as NO_x and HC were deliberately retained in the dataset. In internal combustion engine research, such extreme values generally correspond to physically meaningful operating conditions rather than measurement errors, and their exclusion would limit the model's ability to represent the full operational envelope of the engine. Consequently, accounting for outlier behaviour is essential when selecting weighting strategies in multi-criteria decision-making analyses, as it can substantially influence correlation structures, statistical dispersion, and the robustness of the resulting decision scores.

ARTICLE IN PRESS

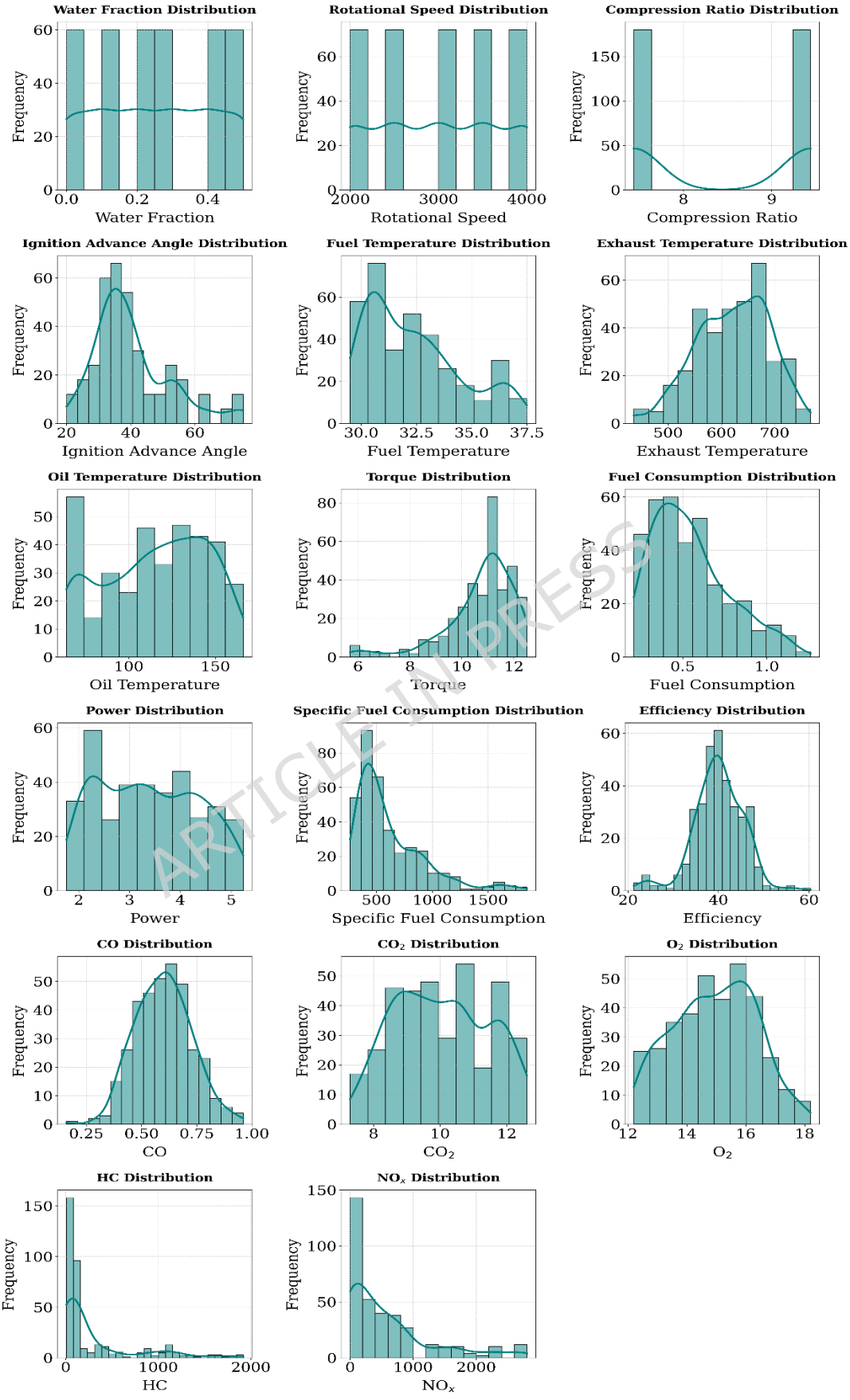
Table 2. Comparison of predictive performance (R^2) using StandardScaler versus RobustScaler on outlier-prone parameters.

Target Parameter	StandardScaler (R^2)	RobustScaler (R^2)	Difference
NO_x	0.7783	0.7784	+0.0001
HC	0.6946	0.6948	+0.0002
CO	0.1381	0.1383	+0.0002

In the context of the modelling strategy, the effect of outliers on prediction accuracy was investigated by comparing StandardScaler (mean-variance based) and RobustScaler (median-quartile based). The comparative performance results obtained from this analysis are summarized in Table 2. As can be seen from the table, the choice of scaling method yielded negligible differences in predictive performance (identical R^2 values). Furthermore, tree-based ensemble models like XGBoost are inherently robust to outliers as they rely on hierarchical splitting rather than distance metrics. Therefore, standard scaling was retained to preserve the original magnitude information.

Figure 1 presents the frequency distributions and kernel density estimations (KDE) for all experimental parameters measured in this study. The control parameters (water fraction, rotational speed, compression ratio, and ignition advance angle) exhibit distinct distribution patterns: water fraction and rotational speed show uniform distributions indicating balanced experimental design, compression ratio displays a bimodal distribution with two distinct levels (7.5 and 9.4), while ignition advance angle follows a unimodal distribution centered around 35-40 degrees. The measured parameters demonstrate varied distributions reflecting the complex engine response characteristics. Fuel and oil temperatures maintain relatively narrow ranges (30-37°C and 75-150°C respectively), confirming stable operating conditions.

ARTICLE IN PRESS



ARTICLE IN PRESS

ARTICLE IN PRESS

Figure 1. Frequency distribution plots of selected experimental parameters.

Performance parameters (power, torque, efficiency) and emission parameters (CO, CO_2, NO_x, HC) show characteristic distributions with HC emissions notably exhibiting a highly skewed distribution with most values concentrated below 500 ppm but extending to 2000 ppm in extreme cases. The specific fuel consumption shows a right-skewed distribution centered around 400-600 g/kWh. These distribution patterns validate the comprehensive coverage of the experimental design space and highlight the presence of genuine operational extremes rather than measurement errors, which were deliberately retained to capture the full operational envelope of the engine under various ethanol-water blend conditions.

2.2.Multi-Criteria Decision Making (TOPSIS)

To evaluate the engine's operating performance under conflicting objectives, the TOPSIS method was employed. Addressing the necessity for scale invariance, this study strictly utilizes Vector Normalization to transform different scales and units into a comparable dimensionless environment. This step is crucial to prevent parameters with large magnitudes (e.g., NO_x) from dominating those with smaller values (e.g., efficiency).

The criteria direction is critical for constructing the decision matrix. Performance metrics such as power and torque are treated as benefit-type criteria and are therefore maximized, whereas fuel consumption and emission-related indicators are considered cost-type criteria and are minimized. The classification of all criteria is presented in Table 3.

Table 3. Classification of engine parameters as Benefit or Cost criteria

Parameter	**Unit**	**Type**	**Goal**
Power	kW	Benefit	Maximize (↑)
Torque	Nm	Benefit	Maximize (↑)
Efficiency	%	Benefit	Maximize (↑)
Specific Fuel Cons. (SFC)	g/kWh	Cost	Minimize (↓)
CO Emission	% vol	Cost	Minimize (↓)
CO_2 Emission	% vol	Cost	Minimize (↓)
NO_x Emission	ppm	Cost	Minimize (↓)
HC Emission	ppm	Cost	Minimize (↓)

The TOPSIS procedure follows five standard steps: vector normalization of the decision matrix, multiplication by entropy-derived criterion weights, identification of positive and negative ideal

ARTICLE IN PRESS

solutions, computation of Euclidean separation measures, and calculation of the relative closeness coefficient for each alternative. The full mathematical formulation is provided in Appendix A.

2.3.Determination of Criteria Weights and Correlation-Based Method Selection

MCDM processes, the assignment of appropriate weights to evaluation parameters plays a crucial role in ensuring the objectivity and accuracy of the final ranking. This section outlines the application of three objective weighting methods, Standard Deviation, Entropy, and CRITIC for determining the relative importance of performance and emission-related parameters. Each method leverages different statistical or information-theoretic principles to quantify parameter significance without relying on expert judgment. Additionally, a correlation matrix is computed to assess the interrelationships among variables, and the average correlation coefficient is used to guide the selection of the most suitable weighting strategy. This integrated approach enhances the reliability of the evaluation framework by aligning methodological choice with data structure characteristics.

In the standard deviation method, the variability of each criterion is interpreted as a measure of its discriminative capability, such that criteria with higher dispersion exert greater influence in the decision-making process. The entropy method, on the other hand, evaluates the degree of uncertainty associated with each parameter and assigns higher weights to criteria exhibiting more informative and diverse data distributions. Building upon these principles, the CRITIC method simultaneously accounts for both the intrinsic variability of the criteria and the level of conflict among them by reducing the influence of highly correlated parameters to mitigate redundancy. To support the selection of an appropriate weighting strategy, inter-parameter correlations were examined, and the overall correlation structure was used as a guiding indicator, favouring entropy-based weighting under low to moderate dependency conditions and CRITIC-based weighting when strong interdependencies were observed.

Table 4. Comparison of parameter importance values calculated by Standard Deviation, Entropy, and CRITIC methods.

Parameter	**Std Deviation (%)**	**Entropy (%)**	**CRITIC (%)**
Power	0.07	2.59	0.07
Torque	0.09	0.53	0.13
Efficiency	0.39	0.61	0.39
Specific fuel consumption	21.15	6.42	26.83
CO	0.01	1.42	0.00
CO_2	0.10	0.60	0.09

ARTICLE IN PRESS

NO_x	48.20	38.66	40.41
HC	30.00	49.17	32.07

The results presented in Table 4 provide a comparative overview of parameter weight distributions obtained using three objective weighting methods. Across all approaches, NO_x and HC emerge as the most influential criteria, indicating their strong variability and high informational contribution to the dataset. The entropy-based weighting approach assigns the greatest importance to HC, whereas the CRITIC method emphasizes both NO_x and brake specific fuel consumption as dominant parameters. These variations illustrate how each weighting strategy captures different characteristics of the data by prioritizing information content, statistical dispersion, or inter-criterion independence. The consistent prominence of specific parameters across all methods further confirms their critical role in multi-criteria evaluation frameworks such as TOPSIS.

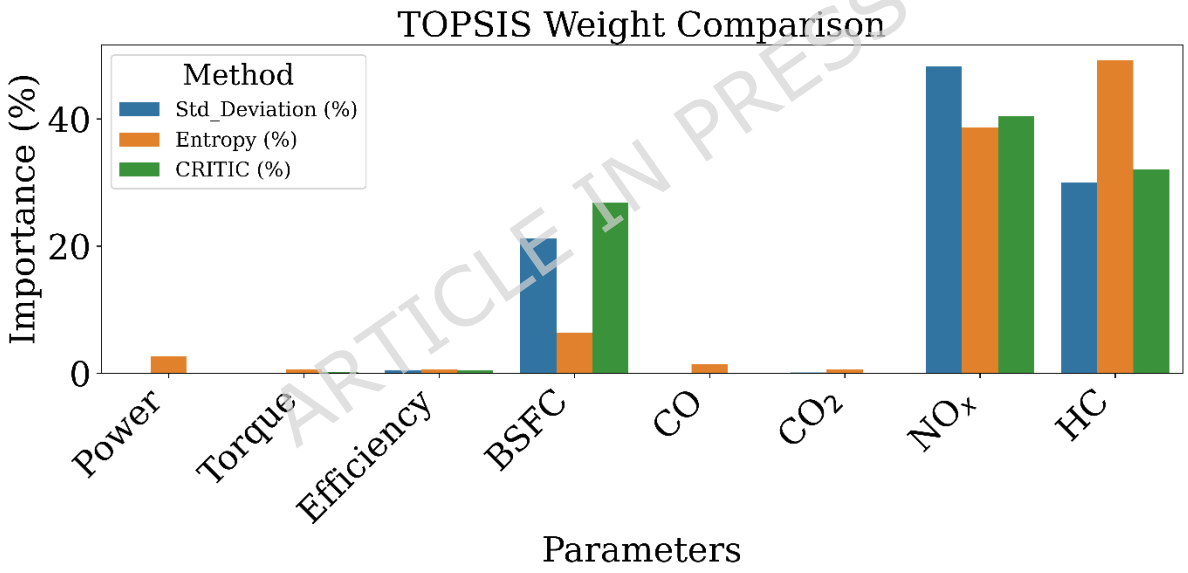


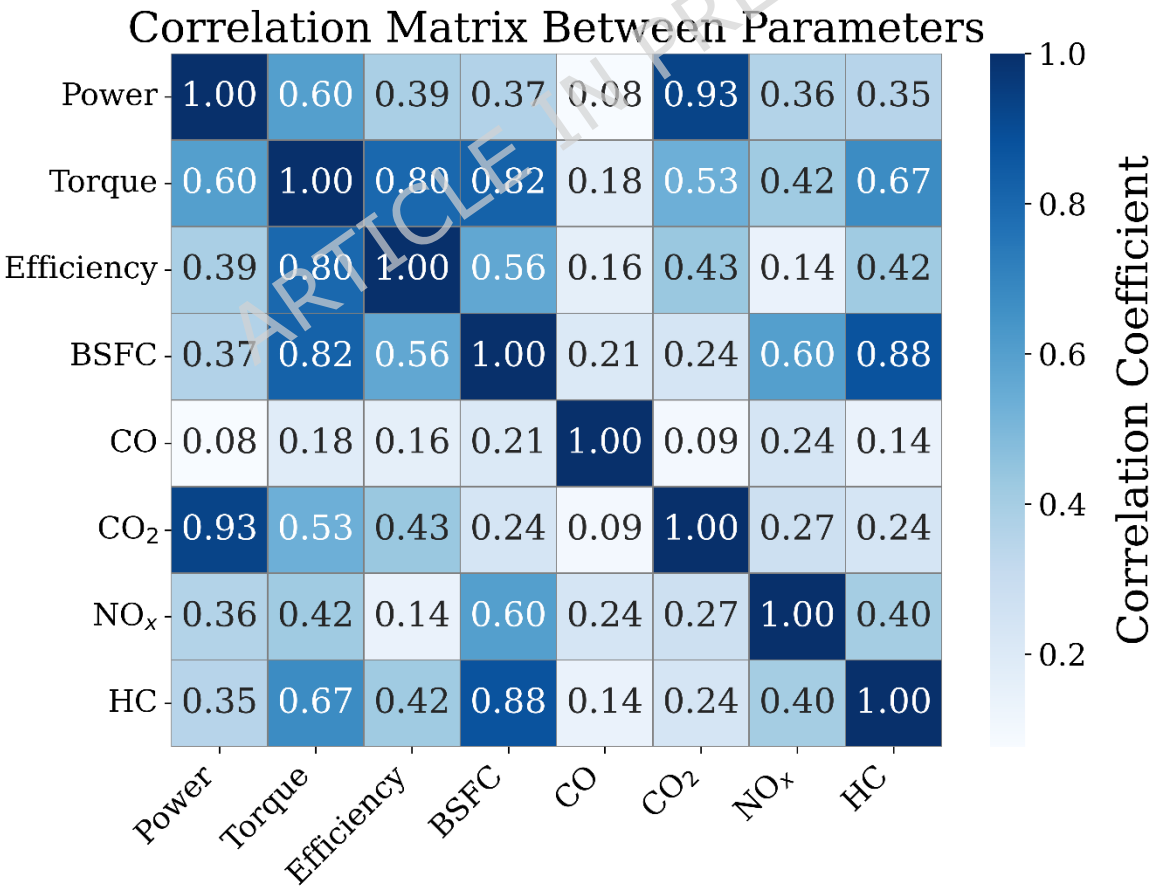
Figure 2. Parameter importance values computed using three different objective weighting methods: Standard Deviation, Entropy, and CRITIC.

The bar chart in Figure 2 presents a comparative analysis of the weights assigned to eight performance and emission parameters using Standard Deviation, Entropy, and CRITIC methods. It is evident that NO_x and HC are consistently identified as the most influential parameters across all three methods, reflecting their substantial variability and contribution to decision-making. While the Standard Deviation method emphasizes the influence of BSFC and NO_x, the Entropy method assigns the highest weight to HC, indicating a high degree of information content in its

ARTICLE IN PRESS

distribution. The CRITIC method, which accounts for both variability and inter-parameter correlation, offers a more balanced weight distribution, particularly elevating the role of BSFC. These differences highlight the methodological sensitivity of each approach and underline the importance of aligning weighting strategies with the underlying data structure in multi-criteria evaluations.

Figure 3 visualizes the Pearson correlation coefficients calculated between eight key performance and emission parameters. The results reveal strong positive correlations, particularly between power and CO_2 (r = 0.93), torque and BSFC (r = 0.82), and torque and efficiency (r = 0.80), indicating that these variables tend to increase together under similar conditions. In contrast, parameters such as co exhibit generally weak correlations with others, suggesting more independent behaviour. These variations in interdependence underscore the relevance of correlation-based weighting techniques such as the CRITIC method in multi-criteria analysis, as they account for redundancy and potential overlap in the decision-making framework.



ARTICLE IN PRESS

Figure 3. Heatmap illustrating the Pearson correlation coefficients among selected performance and emission parameters.

The average absolute correlation was computed using Equation A.6 in Appendix A, yielding 41.20%.

Based on this result, the Entropy method was recommended for assigning criteria weights, as it performs best under low-correlation conditions by prioritizing parameters with higher information entropy and minimal redundancy. This data-driven selection enhances the objectivity and appropriateness of the weighting strategy employed in the subsequent TOPSIS analysis. This selection criterion is grounded in the underlying mechanics of the two methods. The CRITIC method, originally proposed by Diakoulaki et al. [32], extends pure variance-based weighting by explicitly penalizing criteria that are strongly correlated with others, on the rationale that highly correlated criteria carry redundant information and should not be allowed to dominate the aggregate score through repeated emphasis. This correction mechanism delivers its greatest benefit precisely when inter-criteria correlations are strong and pervasive, commonly considered to begin around an absolute correlation coefficient of 0.7 to 0.8 [33]. In the present dataset, the average absolute correlation of 41.20% falls short of this threshold, even under conservative conventions for classifying correlation magnitude such as Cohen's medium effect benchmark of approximately 0.30 [34], indicating that while some pairwise relationships are present, for example between power and CO_2, or torque and BSFC, they do not represent the pervasive redundancy that CRITIC's correlation correction term is designed to address. Under such conditions, CRITIC's weighting tends to converge toward that of the variance based methods, while introducing additional estimation variance through the correlation matrix itself, particularly given the documented presence of outliers in NO_x, HC, and BSFC discussed in Section 2.1. The entropy method, by contrast, requires no explicit correlation correction and instead directly rewards criteria with high information content and dispersion across alternatives [33], which is consistent with the moderate to low correlation structure observed here. This reasoning supports its selection over CRITIC for the present dataset.

The entropy weighting procedure evaluates information uncertainty in each parameter through column normalization, entropy computation, utility calculation, and weight normalization. The formulation is provided in Appendix A.

ARTICLE IN PRESS

The results of the analysis are summarized in Table. This dual representation allows for both numerical precision and graphical interpretation of the parameter importance values.

2.4.Evaluation of TOPSIS Results and Correlation Analysis

The outcomes of the TOPSIS analysis, conducted across a total of 360 distinct experimental alternatives, are presented in Table 5, which displays the top five ranked results. To enhance physical interpretability in real-world contexts, the parameters for these ranked alternatives are presented in their original units rather than normalized values. This approach enables a more comprehensive understanding of how specific engine operating conditions influence the final decision rankings. The TOPSIS analysis results for the total of 360 alternatives show that the highest-ranked alternative achieves a score of 0.97104, highlighting a superior balance between performance and emissions. The decreasing trend in scores among the top-ranked results reflects the diminishing optimality of subsequent alternatives. Here, an 'alternative' represents a unique engine operating condition evaluated within the TOPSIS framework, and all reported parameter values are presented in their original physical units (kW, Nm, ppm, etc.) to ensure consistency with the descriptive statistics in Table 1. All 360 experimental data points were treated as individual alternatives in the TOPSIS decision matrix.

Table 5 Performance and emission parameters, TOPSIS scores, and final rankings of the top 5 alternatives.

Rank	TOPSIS Score	Power	Torque	Efficiency	Bsfc	CO	CO_2	NO_x	HC
1	0.97104	2.07	9.9	40.01	486.10	0.49	8.4	124	23
2	0.97095	2.03	9.7	37.85	513.90	0.50	8.4	121	24
3	0.97066	2.09	10.0	37.72	515.70	0.50	8.3	125	21
4	0.97042	2.09	10.0	40.42	481.30	0.56	8.3	125	27
5	0.97031	2.07	9.9	37.34	520.90	0.53	8.4	123	26

ARTICLE IN PRESS

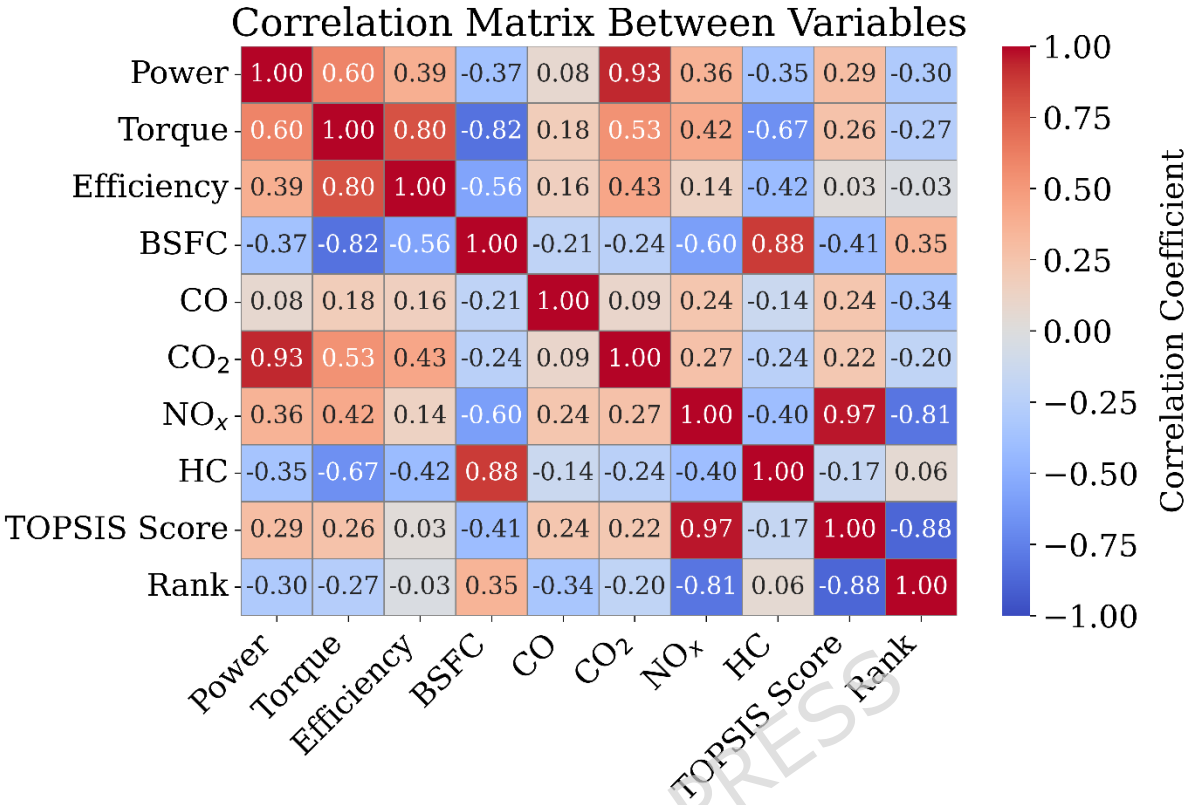


Figure 4. Correlation heatmap of performance and emission parameters with TOPSIS Score and Rank.

In addition to the ranking results, a correlation heatmap was generated to examine the relationships between the input variables and the TOPSIS outcome. The visual representation is shown in Figure 4.

The figure visualizes pairwise correlations among the decision variables, highlighting strong associations such as between NO_x and TOPSIS Score ($r = 0.97$), and their inverse relationship with the final ranking ($r = –0.88$). The heatmap reveals that NO_x is the most influential parameter in determining the final TOPSIS scores, aligning with its highest Entropy weight. In contrast, variables such as efficiency and CO exhibit weak correlations with the TOPSIS score, suggesting minimal influence under the applied weighting strategy.

These patterns reinforce the effectiveness of the Entropy-based weight assignment, which emphasizes parameters with high information variance and low redundancy. By combining score comparison and correlation analysis, this section substantiates the objectivity and consistency of the decision-making process applied in this study.

ARTICLE IN PRESS

2.5.Machine Learning-Based Prediction of Classical TOPSIS Scores

In this study, three distinct regression models were implemented to predict the relative scores obtained from the classical TOPSIS method: the Response Surface Methodology (RSM), the tree-based gradient boosting algorithm (XGBoost), and a hybrid Stacking Ensemble model that combines the two. This predictive modelling

aimed to evaluate whether the final TOPSIS-based decision scores could be reliably approximated using the original technical and emission parameters.

This choice of prediction target, the aggregated TOPSIS score itself rather than the eight individual performance and emission parameters, reflects a deliberate methodological decision with both statistical and practical motivations, and is conceptually aligned with the predict-then-optimize literature in operations research and machine learning [35]. An alternative two-stage strategy would be to train eight separate regression models, one for each performance and emission parameter, and subsequently compute the TOPSIS score from these predicted values. However, because the TOPSIS procedure itself functions as a decision rule that maps the eight criteria onto a single ranking, this two-stage strategy mirrors the classical predict-then-optimize pipeline, in which a model is trained to minimize prediction error on intermediate parameters without accounting for how those errors propagate through the downstream decision step. Because the TOPSIS procedure involves normalization, weighting, and Euclidean distance based ranking across all criteria simultaneously, prediction errors from each of the eight independent models would propagate and interact nonlinearly through these steps, compounding into a final decision score with substantially greater and harder to quantify uncertainty than the individual model errors alone. Directly regressing the TOPSIS score avoids this error compounding cascade, consistent with the rationale that motivates decision-focused or end-to-end learning over indirect, two-stage approaches in other prediction-and-decision settings [35]. From a practical standpoint, a unified score offers direct operational value for engine calibration. Rather than requiring engineers to simultaneously weigh multiple, often conflicting trade-offs, such as higher power against higher NO_x, or lower BSFC against higher HC, for every candidate operating condition, the predicted TOPSIS score condenses these trade-offs into a single, physically interpretable figure of merit that can be used for rapid virtual screening of the operating envelope, design space exploration, or, as discussed in Section

ARTICLE IN PRESS

7, onboard estimation of overall engine performance where the eight underlying quantities cannot all be measured simultaneously.

2.5.1. Machine Learning Framework

To ensure the reliability of the predictive models and prevent data leakage, a rigorous validation strategy was implemented. The machine learning models were designed to predict the TOPSIS score (output variable, y) from four engine control parameters: water fraction, rotational speed, compression ratio, and ignition advance angle (input features, X). Importantly, the TOPSIS score itself is calculated from eight performance and emission metrics (power, torque, efficiency, specific fuel consumption, CO, CO_2, NO_x, and HC), which are not used as model inputs but rather aggregated into the single target variable through the TOPSIS methodology.

The dataset consists of 6 repetitions for each experimental condition. A standard random split could result in 'repetition leakage,' where different repetitions of the same condition appear in both training and test sets, leading to artificially inflated performance metrics. To address this, a 5-Fold Group Cross-Validation (GroupKFold) strategy was adopted. The data points were grouped based on unique combinations of operational parameters (water fraction, speed, compression ratio, ignition timing). This ensures that all repetitions of a specific condition are kept strictly within the same fold (either training or validation), assessing the model's ability to generalize to unseen operating conditions. Furthermore, for the Stacking Ensemble model, Out-of-Fold (OOF) predictions were generated. The meta-learner was trained using these OOF outputs from the base learners, ensuring that the input to the second-level model was not seen during the initial training phase.

2.5.2. Modelling Approaches

The first model was based on Response Surface Methodology (RSM), which utilizes second-order polynomial terms in a multivariate regression setting. The second model employed the XGBoost algorithm, a gradient boosting approach optimized for computational efficiency and prediction accuracy. Finally, both RSM and XGBoost were combined in a Stacking Ensemble architecture, with a linear regressor serving as the meta-learner. Early stopping was not applied in the XGBoost training procedure, since the model evaluation was performed under a strict 5-fold GroupKFold cross-validation protocol.

ARTICLE IN PRESS

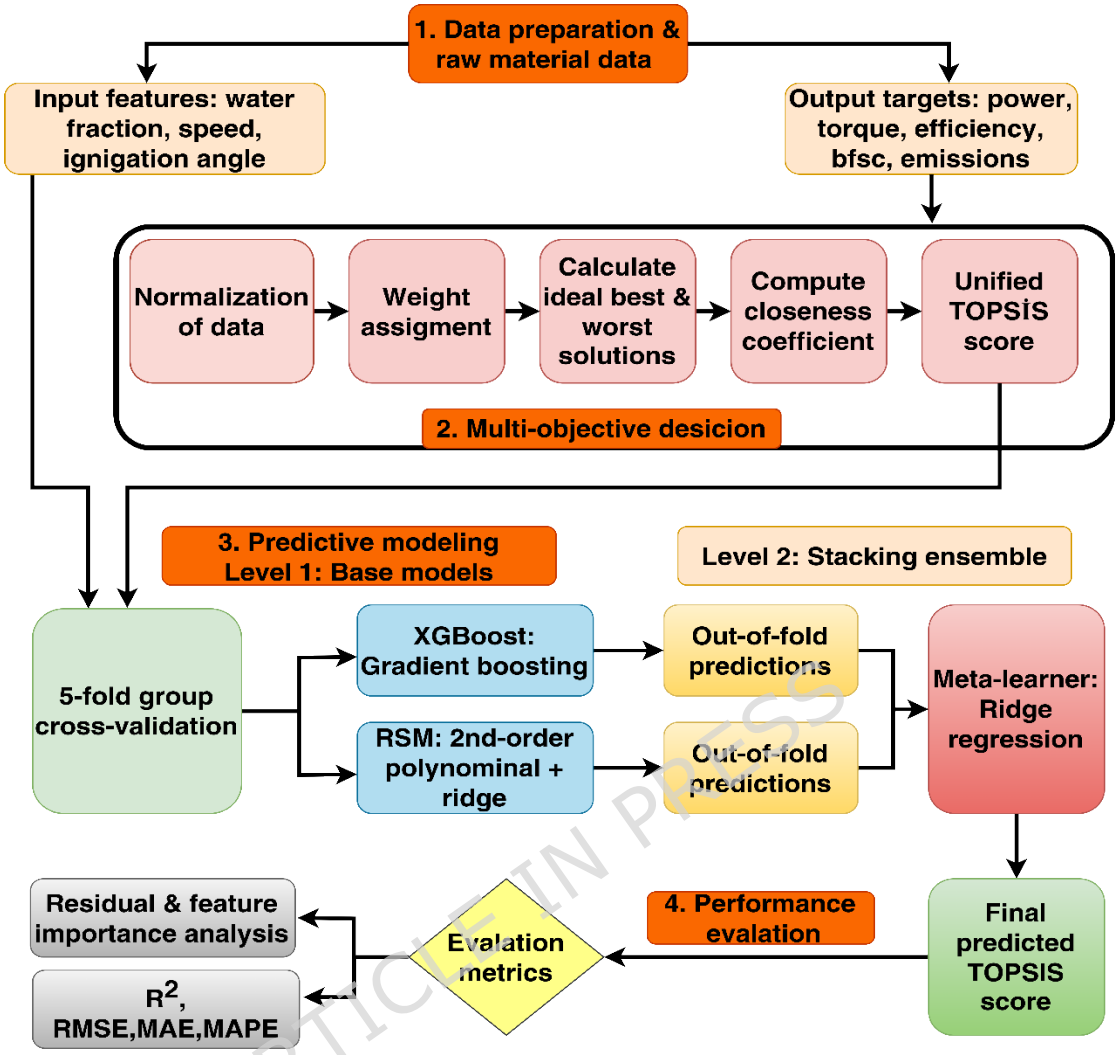


Figure 5. Workflow of the machine learning models used to predict classical TOPSIS scores, including data preparation, model training, and performance evaluation.

The hyperparameters were selected using a manual heuristic tuning strategy based on stable generalization performance under GroupKFold validation, rather than an exhaustive grid search or Bayesian optimization. Each model's prediction performance was evaluated using four standard metrics: coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). These metrics were used to assess and compare model accuracy under consistent conditions. A schematic representation of the entire machine learning pipeline is illustrated in Figure 5, which includes preprocessing, input/target structure, model configurations, and performance evaluation stages.

ARTICLE IN PRESS

This structured pipeline, as illustrated in Figure 5, ensures that the classical TOPSIS decision scores can be effectively modelled through various regression techniques. By systematically organizing the preprocessing steps, feature-target selection, training setup, and model-specific configurations, the workflow supports reproducibility, transparency, and robust performance evaluation. Such a framework facilitates not only the interpretability of prediction outcomes but also provides a solid foundation for comparative analysis of model efficiency in multicriteria decision-making contexts. The detailed configuration parameters and hyperparameter settings for the implemented RSM, XGBoost, and Stacking models are provided in Table 6.

Table 6 Configuration and hyperparameters of the implemented regression models.

Model Component	**Parameter / Setting**	**Value / Description**
Data Preprocessing	**Train/Test Split**	80% Training / 20% Testing
	Cross-Validation	5-fold Group Cross-Validation strategy
	Input Features (4)	Water Fraction, Rotational Speed, Compression Ratio, Ignition Advance Angle
Model 1: RSM	**Method**	Response Surface Methodology (Ridge Regression)
	Polynomial Degree	2 (Second-order terms with interactions)
	Regularization	Ridge (L2) Alpha: 1.0
Model 2: XGBoost	**Algorithm**	Tree-based Gradient Boosting
	Hyperparameters	n_estimators: 300, Learning Rate: 0.05, Max Depth: 6
	Regularization	Subsample: 0.8, Colsample_bytree: 0.8
	Random Seed	42
Model 3: Stacking	**Architecture**	Heterogeneous Ensemble
	Base Learners	RSM (Polynomial), XGBoost
	Meta-Learner	Ridge Linear Regressor (L2 regularized, Alpha: 1.0)

All model training and validation procedures were performed on a workstation equipped with an Intel Xeon W7-2495X CPU, 128 GB DDR5 RAM, 2× NVIDIA RTX A5500 GPUs (24 GB each; total 48 GB VRAM), and a 2 TB NVMe SSD, running Windows 11 Professional.

2.6.Sensitivity Analysis and Robustness Check

ARTICLE IN PRESS

To address the concern that the Entropy method's assignment of high weights to specific parameters (specifically HC at 49.17% and NO_x at 38.66%) might bias the decision-making process towards a "single-criteria" optimization, a comprehensive robustness check was conducted. The TOPSIS rankings were re-calculated using three alternative objective weighting strategies: CRITIC (Criteria Importance Through Inter-criteria Correlation), Standard Deviation (SD), and Equal Weighting. First, the weight distributions were analyzed to highlight the methodological differences. As shown in Figure 6, the Entropy method exhibits a highly skewed distribution driven by data variance, whereas the Equal Weighting method assigns a constant importance (12.5%) to all eight parameters. The CRITIC method, which accounts for contrast and conflict, assigns a more moderate weight distribution, notably increasing the importance of Specific Fuel Consumption compared to Entropy.

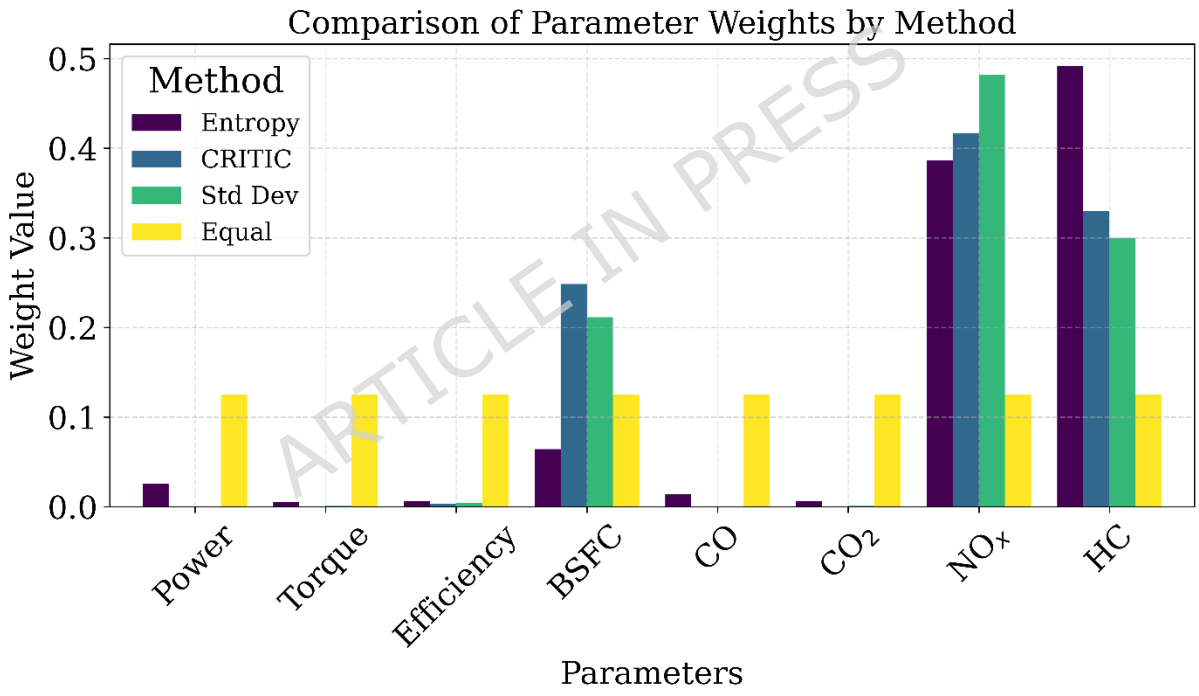


Figure 6. Comparison of parameter importance weights derived from Entropy, CRITIC, Standard Deviation, and Equal Weighting methods.

Despite these significant variations in weight assignments, the core objective is to determine if the identification of the optimal engine operating conditions remains stable. To quantify the similarity between the rankings produced by the proposed Entropy method and the alternative scenarios, Spearman's rank correlation coefficients were calculated. The resulting correlation matrix is presented in Figure 7.

ARTICLE IN PRESS

The analysis reveals a strong positive correlation between the Entropy-based rankings and those obtained from other methods. Specifically, the correlation coefficient between the Entropy method and the CRITIC method is 0.97, indicating that accounting for inter-parameter conflicts yields nearly identical decision outcomes.

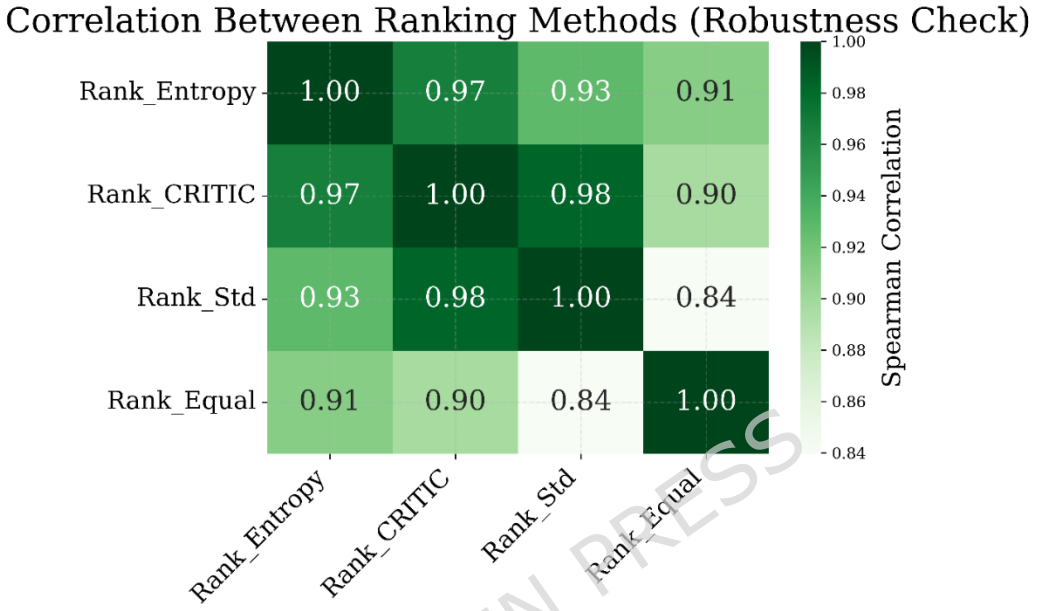


Figure 7. Spearman rank correlation heatmap illustrating the high degree of consistency between rankings obtained from Entropy, CRITIC, SD, and Equal Weighting methods.

Most importantly, the correlation between the Entropy method and the Equal Weighting method is 0.91. This is a critical finding; it demonstrates that even when the dominance of HC and NO_x is completely removed (by assigning equal weights to all criteria), the ranking of the alternatives remains 91% consistent. This empirical evidence refutes the hypothesis that the proposed framework relies solely on emission parameters. Instead, it confirms that the top-ranked alternatives represent "global optima" within the experimental design space, satisfying both performance and emission trade-offs regardless of the specific weighting logic applied.

ARTICLE IN PRESS

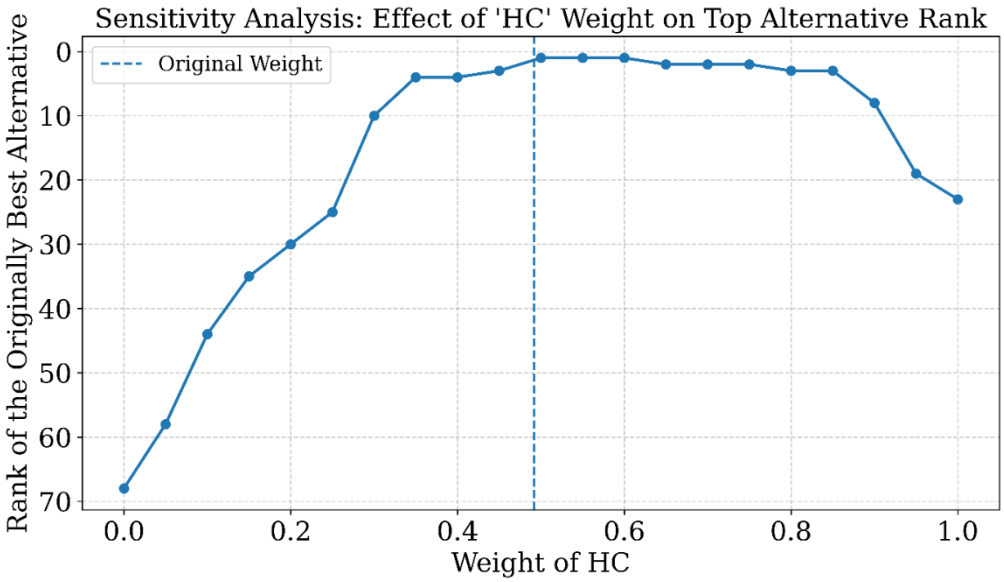


Figure 8. Sensitivity analysis showing the rank stability of the top alternative under varying weights of the dominant HC parameter.

Finally, a single-parameter sensitivity analysis was performed on the most dominant criterion (HC) to observe rank stability. As illustrated in Figure 8, the top-ranked alternative maintains its leading position (Rank 1-5) even when the weight of HC varies significantly around its original value. However, as expected, if the weight of HC is reduced towards zero, the ranking of the environmentally optimal alternative drops, which confirms that HC is indeed a distinguishing factor for sustainability but does not invalidate the multi-criteria nature of the solution.

3. Results

In this study, the predictive capabilities of three distinct regression approaches (RSM, XGBoost, and Stacking Ensemble) were evaluated for predicting the single TOPSIS decision score, which aggregates eight performance and emission indicators. To ensure the reliability of the validation process and prevent data leakage between experimental repetitions, a 5-Fold Group Cross-Validation strategy was implemented. The comprehensive performance metrics obtained from this analysis are summarized in Table 7.

The comparative performance of the RSM, XGBoost, and Stacking Ensemble models is summarized quantitatively in Table 7 and visually illustrated in Figure 9(a–d). Figure 9(a) shows the coefficient of determination (R^2) values obtained using 5-fold Group Cross-Validation, while

ARTICLE IN PRESS

Figures 9(b), 9(c), and 9(d) present the corresponding RMSE, MAE, and MAPE metrics, respectively.

Table 7 Performance metrics (R^2 and RMSE) of RSM, XGBoost, and Stacking models obtained using 5-Fold Group Cross-Validation.

Model	**R²**	**RMSE**	**MAE**	**MAPE**
RSM	0.774334	0.067627	0.050688	7.124762
XGBoost	0.780295	0.066728	0.046091	6.542203
Stacking Ensemble	0.814314	0.061345	0.044456	6.359036

Overall, the stacking ensemble achieved the best performance for predicting the TOPSIS score among the compared models; however, we report this conclusion strictly based on the aggregated evaluation metrics in Table 7. As shown in Figure 9(a), the RSM model achieved an R^2value of 0.7743, indicating a moderate predictive capability. However, this performance is accompanied by comparatively higher prediction errors, as evidenced by the RMSE (0.0676), MAE (0.0507), and MAPE (7.12%) values illustrated in Figures 9(b–d).

ARTICLE IN PRESS

ARTICLE IN PRESS

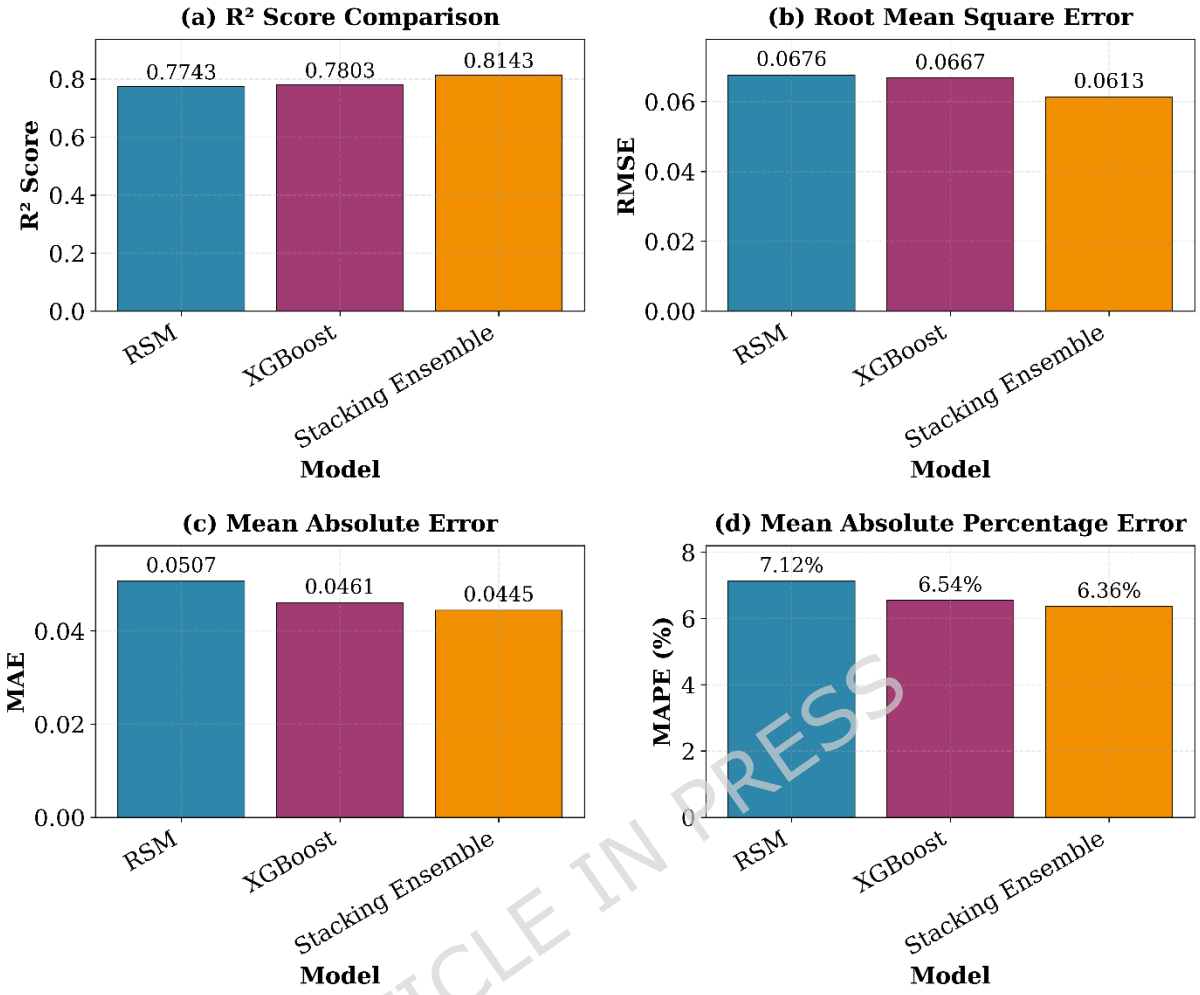


Figure 9. Model performance metrics from 5-fold cross-validation: (a) R², (b) RMSE, (c) MAE, (d) MAPE for RSM, XGBoost, and Stacking Ensemble.

These results suggest that the polynomial-based RSM approach is limited in capturing the complex nonlinear relationships governing the TOPSIS score. The XGBoost model demonstrates an improvement over RSM across all performance indicators. As observed in Figure 10(a), the R^2 value increases to 0.7803, while the corresponding error metrics decrease to RMSE = 0.0667, MAE = 0.0461, and MAPE = 6.54% (Figures 10(b–d)). This improvement reflects the ability of gradient boosting algorithms to model nonlinear interactions and variable dependencies more effectively. The proposed Stacking Ensemble model consistently outperforms both individual models, as clearly illustrated in Figure 10. It achieves the highest R^2 value of 0.8143 (Figure 10(a)) and the lowest error metrics, with RMSE = 0.0613, MAE = 0.0445, and MAPE = 6.36% (Figures 10(b–d)). The simultaneous improvement in goodness-of-fit and error reduction indicates that the stacking strategy successfully integrates the complementary strengths of RSM and XGBoost,

ARTICLE IN PRESS

leading to enhanced generalization performance. Overall, the visual trends observed in Figure 10 corroborate the numerical results reported in Table 8 and confirm the robustness of the Stacking Ensemble model under group-based cross-validation conditions.

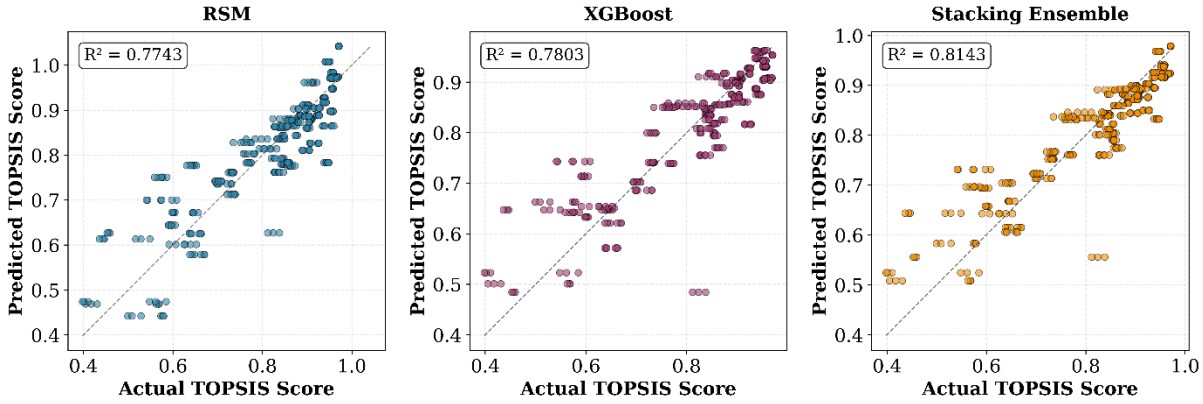


Figure 10. Actual versus predicted TOPSIS scores for (a) RSM, (b) XGBoost, and (c) Stacking Ensemble models.

Figure 10 illustrates the relationship between the experimentally obtained TOPSIS scores and the model-predicted values for the RSM, XGBoost, and Stacking Ensemble models. As observed in Figure 10(a), the RSM model exhibits noticeable dispersion around the ideal diagonal line, indicating limited predictive accuracy, particularly at lower TOPSIS score ranges.

The XGBoost model (Figure 10(b)) demonstrates an improved alignment with the diagonal, reflecting enhanced predictive capability compared to RSM, although some deviations remain. In contrast, the Stacking Ensemble model (Figure 10(c)) shows the closest clustering of data points around the ideal line, confirming its superior prediction accuracy and consistency. These visual observations are consistent with the quantitative performance metrics reported in Table 7.

The compression ratio is identified as the second most significant factor (0.262), followed by the ignition advance angle (0.167). In contrast, rotational speed exhibits a comparatively minor contribution, with the lowest importance score (0.045). These results suggest that fuel composition and combustion-related parameters have a stronger impact on the overall TOPSIS-based performance evaluation than operating speed under the investigated conditions.

Figure 11 presents a comprehensive residual diagnostic analysis for the Stacking Ensemble model, serving to validate the underlying regression assumptions and overall model reliability.

ARTICLE IN PRESS

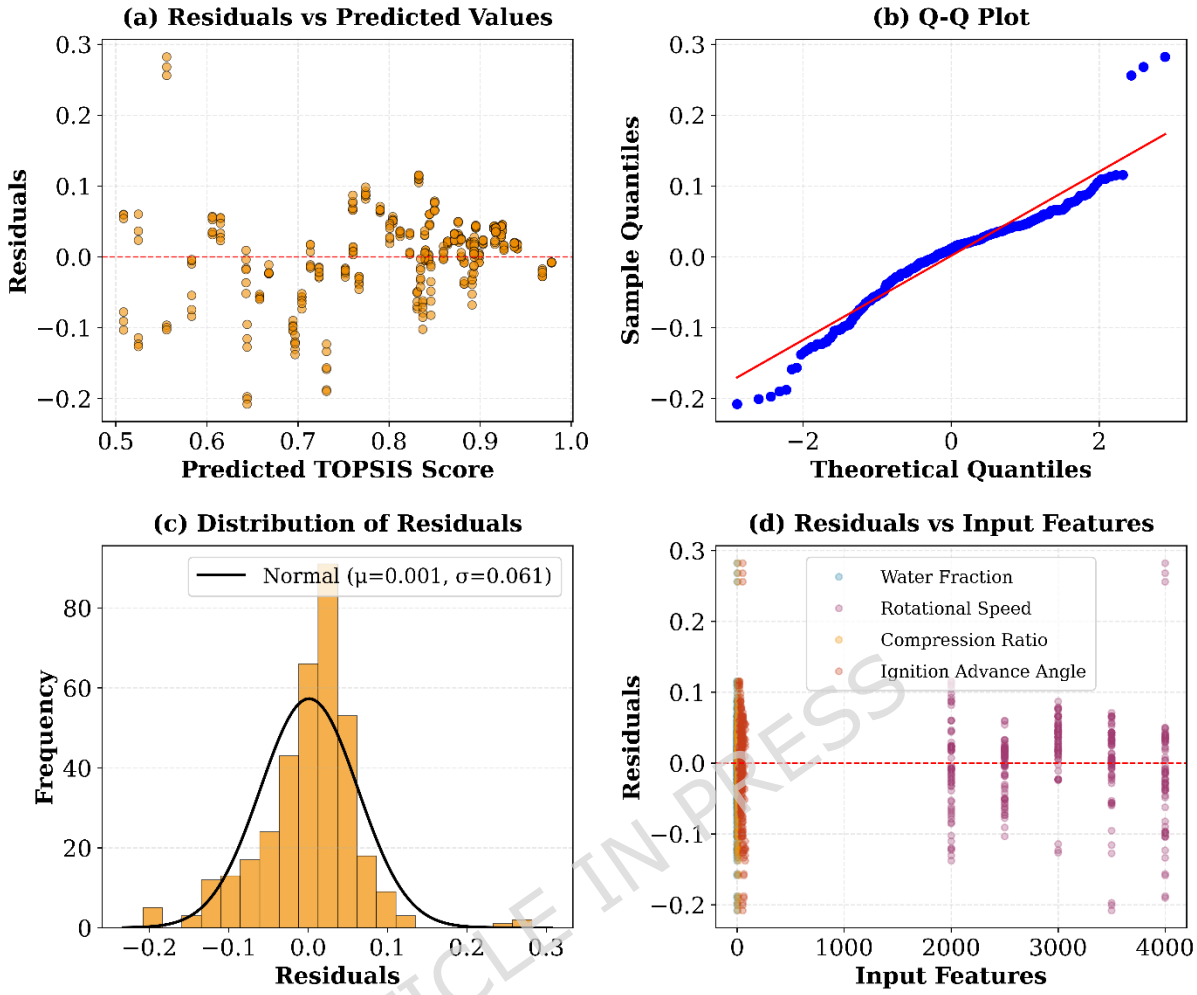


Figure 11. Residual diagnostics for the Stacking Ensemble model: (a) residuals vs predicted values, (b) Q-Q plot, (c) residual distribution with normal curve overlay, and (d) residuals vs input features showing model adequacy and assumption validation.

The plot of residuals against predicted values demonstrates a largely homoscedastic pattern, with residuals evenly distributed around the zero-reference line across the full prediction range from 0.5 to 1.0, although a limited number of extreme points exhibit residual magnitudes exceedingly plus or minus 0.2. The quantile–quantile analysis further confirms the approximate normality of the residual distribution, as most data points closely follow the theoretical diagonal reference line, with only minor deviations observed at the distribution tails, suggesting slightly heavier tail behaviour than that of a perfect normal distribution. The residual distribution histogram (Figure 11c) exhibits a near-perfect normal shape with mean μ=0.001 and standard deviation σ=0.061, confirming unbiased predictions and consistency with the reported RMSE of 0.0613.

ARTICLE IN PRESS

The residuals versus input features plot (Figure 11d) reveals no systematic patterns across water fraction, rotational speed, compression ratio, or ignition advance angle, with uniform scatter indicating the model captures the relationships without systematic bias. Overall, this residual analysis validates the fundamental regression assumptions of linearity, independence, homoscedasticity, and normality, supporting the reliability of the Stacking Ensemble model's R^2 value of 0.814 and confirming its robustness for TOPSIS score prediction in engine performance evaluation applications.

Figure 12 presents the distribution characteristics of TOPSIS scores across the experimental dataset and their variation with rotational speed.

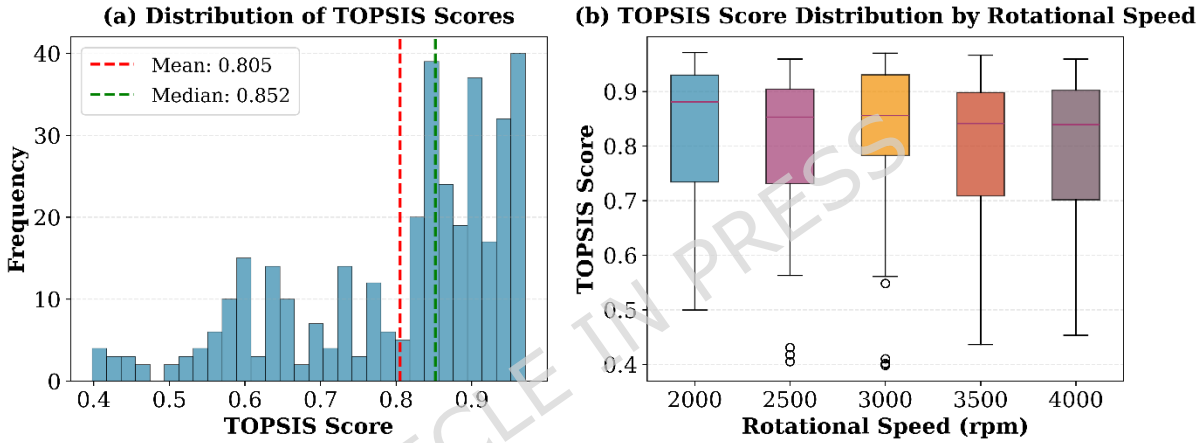


Figure 12. Distribution of TOPSIS scores: (a) Overall frequency histogram with statistical markers, (b) Score variation by rotational speed showing median consistency but varying dispersion.

The histogram in Figure 12(a) reveals a bimodal distribution with a distinct gap around 0.75-0.80, where the lower mode (centered around 0.60-0.65) represents suboptimal operating conditions while the upper mode (0.85-0.95) indicates superior performance configurations. The mean TOPSIS score of 0.805 and median of 0.852 suggest a slight negative skew, with more data points concentrated in the higher performance region.

This bimodal pattern indicates that the engine operating conditions naturally segregate into two distinct performance clusters rather than forming a continuous spectrum. Figure 12(b) demonstrates the impact of rotational speed on TOPSIS score variability through boxplots, showing relatively consistent median values across all speed ranges (2000-4000 rpm) but with notable differences in dispersion. The 2000 rpm condition exhibits the widest interquartile range,

ARTICLE IN PRESS

suggesting greater sensitivity to other operating parameters at lower speeds, while higher speeds (3500-4000 rpm) show more compressed distributions with fewer outliers. The presence of outliers, particularly at 2500 and 3000 rpm, corresponds to specific combinations of water fraction and ignition timing that produce exceptionally poor or exceptional performance, confirming the complex interaction effects between operational parameters in determining overall engine performance.

4. Discussion

Although the prediction of classical TOPSIS scores through machine learning (ML) has recently attracted attention, the integration of regression-based models with Multi-Criteria Decision-Making (MCDM) techniques in technical fields such as engine performance and emission evaluation remains largely underexplored. This study introduces a novel hybrid framework that combines Response Surface Methodology (RSM) and XGBoost within a stacking ensemble architecture to predict classical TOPSIS scores derived from real-world engine performance and emission data. To the best of our knowledge, this represents the first implementation of such a hybrid stacking model in the context of engine-focused classical decision-making, offering a data-driven alternative to conventional MCDM-based evaluation.

4.1. Comparison with Existing ML-MCDM Hybrid Approaches and Methodological Positioning

Existing studies that combine ML and MCDM methods in engineering contexts can be grouped into three broad categories, each of which differs fundamentally from the approach proposed in this study. The first category comprises studies that apply ML or AI exclusively to the prediction or optimization of individual, directly measurable engine performance and emission parameters without constructing a unified decision metric. Zhu et al. [28], working on the same Otto-cycle ethanol/water experimental dataset used in the present study, developed an AI-based multi-objective optimization strategy that outputs separate optimized operating points for each individual objective rather than a single composite score. While this approach achieves high predictive accuracy for individual parameters, it transfers the multi-criteria reconciliation burden back to the engineer, who must manually weigh competing outputs for each candidate operating condition. The second category encompasses studies that couple response-surface or surrogate models with evolutionary optimization algorithms and subsequently apply TOPSIS or another MCDM method

ARTICLE IN PRESS

only as a post-hoc selection step over a small set of Pareto-optimal solutions. The RSM-NSGA-III-TOPSIS framework of Gao et al. [29] is representative of this category: TOPSIS is applied to rank fewer than a dozen candidate solutions identified by the optimizer rather than being used as a continuous, learnable function across the full operating envelope. In such frameworks, TOPSIS serves as a selection filter rather than a prediction target, meaning the composite decision logic cannot be queried for arbitrary, unseen operating conditions without re-running the full evolutionary optimization. The third category includes studies outside the combustion engine domain that couple stacking ensembles with TOPSIS-based evaluation, such as Hannah et al. [30], who applied TOPSIS with MEREC-based weighting to rank competing classification models in a healthcare context. In these studies, TOPSIS evaluates the quality of ML models rather than engine operating conditions, and the composite score is never treated as a regression target.

The present study departs from all three categories in a single, defining methodological step: the composite TOPSIS score, computed across the full set of 360 experimental alternatives using entropy-weighted criteria, is itself treated as a continuous regression target and directly learned by a stacking ensemble combining RSM and XGBoost. This formulation transforms the MCDM evaluation from a one-time, data-dependent ranking exercise into a transferable predictive function that can estimate overall engine performance for any unseen combination of the four control parameters without re-running the multi-criteria analysis. The methodological advancement over the first category is that the multi-criteria reconciliation is embedded into the prediction target itself rather than deferred to the user. The advancement over the second category is that the full operating envelope is covered continuously rather than discretely through a handful of Pareto-optimal points. The advancement over the third category is that TOPSIS defines what is being predicted rather than how competing predictors are ranked. These distinctions are summarized in Table 8 and collectively constitute the primary novelty of the proposed framework.

Several studies have addressed the convergence of ML and MCDM in technical systems, yet none has integrated stacking ensemble models to predict classical decision scores in engine-related applications. For instance, Hoxha et al. [36] utilized stacking for forecasting transportation energy demand in Türkiye, but without incorporating MCDM methods. Wang et al. [37] applied XGBoost within a decision framework that lacked classical inference, while Wu et al. [38] used XGBoost for hydroelectric power prediction, an unrelated domain with no MCDM integration. Das et al. [39], on

ARTICLE IN PRESS

the other hand, employed classical TOPSIS in green logistics but did not implement ensemble learning techniques.

Table 8 Summary of related literature and identified methodological gaps.

Study	**Focus**	**Gap Highlighted**
(Hoxha et al., 2023) [36]	Stacking ensemble for energy demand	No MCDM integration
(Wang et al., 2025) [37]	XGBoost in decision framework	No classical score estimation
(Wu et al., 2024) [38]	XGBoost in hydroelectricity prediction	Not related to MCDM
(Das et al., 2025) [39]	classical TOPSIS in green logistics	No ensemble modeling
(Zhu et al., 2025) [28]	AI based multi-objective optimization on the same ethanol/water Otto-cycle dataset	Optimizes individual objectives separately; no unified MCDM score, no ML regression of a decision metric
(Gao et al., 2024) [29]	RSM-NSGA-III-TOPSIS for diesel/PODE3/n-butanol blends	TOPSIS used only as a post-hoc Pareto-front selector over a few candidate solutions, not as an ML regression target across the full operating envelope
(Hannah et al., 2025) [30]	Stacked ensemble with TOPSIS-MEREC ranking of classification models (healthcare)	TOPSIS ranks competing ML models rather than engine operating conditions; different domain
Current Study	**Stacking (RSM + XGBoost) for engine-based** classical **scores**	**First integration in this context**

More recently, and closer to the engine domain examined here, Zhu et al. [28] optimized the same Otto-cycle ethanol/water dataset using an AI-based multi-objective strategy, but treated each performance and emission target separately rather than aggregating them into a single MCDM score; Gao et al. [29], together with a wider family of RSM/ANN/RF–NSGA-III–TOPSIS studies on diesel blends, applied TOPSIS only as a final selection step over a handful of Pareto-optimal points generated by evolutionary optimization, rather than training a regression model to reproduce the TOPSIS score itself across the full experimental envelope; and Hannah Grace [30] combined a stacking ensemble with TOPSIS in a healthcare application, but used TOPSIS to rank competing classifiers rather than as the prediction target of an ML model. These gaps are summarized in Table 8, which clearly illustrates that none of the reviewed studies applied a stacking ensemble to estimate classical decision scores in an engine performance context, nor used a composite MCDM score itself as a continuous, learnable regression target. The higher R^2 values reported in the comparison studies reflect not only stronger model fit but also the intrinsically lower complexity

ARTICLE IN PRESS

of single-parameter prediction targets; predicting a composite MCDM score that simultaneously encodes trade-offs across eight conflicting criteria represents a fundamentally more demanding regression task, and the R^2 of 0.814 achieved here should be interpreted in that context.

A key methodological question that merits explicit discussion is why predicting the composite TOPSIS score is preferable to predicting the eight individual engine performance and emission parameters directly. From a statistical perspective, an alternative two-stage approach would require training eight separate regression models and subsequently computing the TOPSIS score from their combined outputs. However, the TOPSIS procedure involves sequential normalization, entropy weighting, and Euclidean distance-based ranking across all criteria simultaneously, meaning that prediction errors from the eight independent models would propagate and interact nonlinearly through these steps, compounding into a final decision score with substantially greater uncertainty than the individual model errors suggest in isolation. Directly regressing the TOPSIS score eliminates this error compounding cascade entirely [35]. From a practical perspective, the advantages are equally significant. First, a single composite score provides an immediately actionable figure of merit for engine calibration: rather than requiring engineers to manually reconcile eight separate, often conflicting outputs for each candidate operating point, the predicted TOPSIS score condenses all trade-offs into one number that directly reflects holistic optimality. Second, not all eight underlying parameters can be measured simultaneously in a real engine during operation; sensors for NO_x, HC, and torque may not all be available on a given platform, whereas the TOPSIS score can still be estimated from the four control parameters alone. Third, the composite score is more robust to individual sensor noise or measurement uncertainty, since errors in a single parameter are diluted across the multi-criteria aggregation rather than propagating unchecked to the final decision. Taken together, these statistical and practical advantages justify treating the TOPSIS score as the primary prediction target rather than using it as a post-processing step applied to individually predicted parameters.

The integration of ML with MCDM approaches represents a transformative step in engineering optimization; however, the direct prediction of classical decision scores using ensemble learning has remained a largely underexplored domain. This study addresses this gap by establishing a novel hybrid framework that not only evaluates internal combustion engine performance using TOPSIS but also predicts these decision scores with high precision using a Stacking Ensemble

ARTICLE IN PRESS

model. One of the most critical contributions of this research lies in its rigorous methodological validation. Unlike many prior studies that rely on simple random data splitting which often leads to "data leakage" due to repeated experimental measurements, this study implemented a 5-Fold Group Cross-Validation strategy. As demonstrated in the results, this approach ensured that all repetitions of a specific experimental condition were kept strictly within the same fold. Consequently, the high R^2 scores and low RMSE values obtained are not artifacts of memorization but clear evidence of the model's ability to generalize to unseen engine operating conditions. This methodological rigor sets a new benchmark for data-driven engine performance studies, confirming that the proposed framework is robust enough for real-world calibration tasks.

The physical and engineering significance of the results warrants deeper interpretation. The dominant influence of NO_x in the TOPSIS scoring, reflected by its highest entropy weight of 38.66% and its near-perfect Pearson correlation with the final TOPSIS score (r = 0.97), is consistent with the known thermodynamic sensitivity of NO_x formation to peak combustion temperature. According to the extended Zeldovich mechanism, NO_x production rises exponentially with in-cylinder temperature; consequently, operating conditions that simultaneously suppress peak temperatures while preserving power output tend to dominate the upper range of the composite score. In the ethanol/water blend context investigated here, increasing water content reduces peak combustion temperature through charge cooling and the high latent heat of vaporization of water, thereby suppressing NO_x formation, a mechanism confirmed by experimental water injection studies reporting NO_x reductions of 20% to 48% at water-to-fuel ratios of 0.1 to 0.5 [40,41]. This physical mechanism directly explains why the top-ranked TOPSIS alternatives consistently exhibit low NO_x values (121 to 125 ppm) despite modest power outputs, rather than being dominated by peak power conditions. The bimodal distribution of TOPSIS scores, with a lower cluster centered around 0.60 to 0.65 and an upper cluster around 0.85 to 0.95, also has a clear thermodynamic interpretation: the lower cluster corresponds to operating conditions at higher engine speeds and lower water fractions, where combustion temperatures are elevated and NO_x formation is intensive, while the upper cluster corresponds to moderate speeds with sufficient water content to suppress NO_x without critically impairing combustion completeness. The feature importance analysis further reinforces this interpretation. Water fraction emerges as the most important predictor of the TOPSIS score because it simultaneously governs combustion temperature, NO_x formation, and HC emissions through charge dilution, making it the

ARTICLE IN PRESS

single control parameter with the broadest thermodynamic reach across the decision criteria. Compression ratio ranks second (0.262), consistent with the well-established thermodynamic principle that higher compression ratios improve thermal efficiency through increased expansion work, but simultaneously raise peak temperatures and NO_x emissions, creating a fundamental performance-emission trade-off that is precisely the kind of multi-criteria tension that TOPSIS is designed to resolve [42]. Rotational speed, by contrast, exerts a comparatively minor influence on the composite score (0.045), reflecting that under steady-state conditions, speed primarily affects volumetric efficiency rather than the emission-dominated criteria that carry the heaviest TOPSIS weights in this dataset.

Table 9 Comparison of the proposed model with recent ML-based engine prediction studies.

Study	**Prediction Target**	**Models Compared**	**Best R^2**	**Dataset**
Khayum et al. [43] (2026)	Individual emission parameters (HC, CO, NO_x)	ML algorithms, dual-fuel diesel	0.936-0.998	Diesel/biodiesel dual-fuel engine
Pandey et al. [44](2024)	Individual performance/emission parameters	Multiple ML algorithms, LHR diesel engine	0.95-0.98 (RF)	Ternary blend diesel engine
Gad et al. [45](2025)	Individual performance/emission parameters	XGBoost-RF-MLP stacking	0.979-0.999	WCO biodiesel CI engine
Nachippan et al. [46](2025)	BSFC, individual emission parameters	XGBoost, Random Forest	0.91-0.96	Castor biodiesel engine
Zhu et al. [28](2025)	Individual performance/emission parameters	ANN, surrogate models	>0.95	Same ethanol/water Otto-cycle dataset
Our study	**Composite TOPSIS score (8 criteria aggregated)**	**RSM, XGBoost, Stacking**	**0.814 (Stacking)**	**Ethanol/water Otto-cycle engine**

Table 9 places the predictive performance of the proposed framework in the context of recent ML-based engine studies. It is important to note that all comparison studies report R^2 values for individual, directly measurable engine parameters such as BSFC, NO_x, or HC, where values typically range from 0.91 to 0.999. The present study, by contrast, predicts a composite TOPSIS score that nonlinearly aggregates eight performance and emission criteria simultaneously. The inherently greater complexity of this target, which combines correlated and conflicting criteria through normalization, entropy weighting, and Euclidean distance-based ranking, makes direct

ARTICLE IN PRESS

numerical comparison inappropriate; however, the obtained R^2 of 0.814 for the stacking ensemble is consistent with the expected reduction in predictive accuracy when transitioning from single-parameter to multi-criteria composite targets. This observation is supported by Karaoğlu and Söyler [15], who report an R^2 of 0.81 for a stacking ensemble predicting an analogous composite efficiency metric in a biodiesel conversion context. Furthermore, the computational efficiency of the proposed framework is noteworthy: with training times on the order of a few seconds per fold and inference times well below one millisecond per sample, the stacking ensemble is suitable for both offline virtual calibration and potential onboard deployment on modern Engine Control Units.

A more systematic examination of how each control parameter shapes the TOPSIS score distribution reveals distinct and physically interpretable trends. Regarding water fraction, the relationship with TOPSIS score is non-monotonic: at low water fractions, combustion temperatures are elevated and NO_x formation is intensive, placing operating conditions in the lower TOPSIS cluster; as water fraction increases toward 0.3 to 0.4 by volume, the charge-cooling and thermal dilution effects progressively suppress NO_x and HC emissions, pushing conditions into the upper performance cluster; however, further increases beyond this range begin to impair combustion stability, raising HC emissions and counteracting the NO_x benefit, which explains why the top-ranked alternatives in Table 5 consistently correspond to intermediate rather than maximum water fractions. Regarding compression ratio, the two discrete levels investigated (7.44:1 and 9.44:1) create a fundamental trade-off: the higher compression ratio improves thermal efficiency and reduces BSFC through enhanced expansion work, which benefits the three performance-oriented criteria in the TOPSIS evaluation; however, it simultaneously raises peak in-cylinder temperatures, intensifying NO_x formation and imposing a substantial penalty on the emission-dominated side of the composite score, where NO_x alone carries an entropy weight of 38.66%. The net effect of this trade-off explains why compression ratio ranks as the second most important predictor (importance score 0.262) rather than the dominant one, since its influence on the TOPSIS score is partially self-cancelling across the two competing criterion groups. Regarding ignition advance angle, advancing ignition timing beyond the optimum for a given operating condition improves combustion phasing and thermodynamic efficiency up to a point, reducing BSFC and increasing power output; however, over-advanced timing raises peak cylinder pressure and temperature, increasing NO_x emissions and partially offsetting the performance gains in the composite score. This non-linear, bounded influence explains its intermediate importance ranking

ARTICLE IN PRESS

(0.167) and the absence of a simple monotonic trend in the residuals versus ignition advance angle shown in Figure 11d. Regarding rotational speed, the comparatively minor influence on the TOPSIS score (importance score 0.045) reflects the fact that, under the steady-state experimental conditions investigated, speed primarily governs volumetric efficiency and friction losses rather than the combustion temperature-sensitive emission criteria that carry the heaviest TOPSIS weights. As shown in Figure 12b, median TOPSIS scores remain broadly consistent across the 2000 to 4000 rpm range, while the widest dispersion at 2000 rpm indicates greater sensitivity to the other three parameters at lower speeds, where combustion duration is longest and the effects of water fraction and ignition timing on temperature and mixture quality are most pronounced.

In terms of predictive model comparison, the stacking ensemble's R^2 of 0.814 on a composite decision score is broadly consistent with findings in related engine ML literature, while acknowledging that direct numerical comparison is complicated by differences in prediction targets. For individual engine performance and emission parameters, recent studies report R^2 values of 0.91 to 0.99 for XGBoost and Random Forest on BSFC and emission outputs in biodiesel-blended engines [45,46]. The comparatively lower R^2 obtained in the present study reflects the additional complexity of predicting an aggregated composite score that nonlinearly integrates eight parameters simultaneously, rather than any single measurable output. This is consistent with the observation of Karaoğlu and Söyler [15], who report that a linear regression-based stacking ensemble combining Random Forest, XGBoost, and a deep neural network achieved an R^2 of 0.81 on a similarly aggregated efficiency metric in a biodiesel conversion efficiency context. The superiority of XGBoost over RSM in predicting NO_x and HC in the present study also aligns with findings reported by Gad et al. [45], who demonstrate that tree-based ensemble architectures outperform polynomial surface models for emission parameters exhibiting strong nonlinear sensitivity to operating conditions. Taken together, these comparisons confirm that the predictive performance achieved by the proposed stacking framework is both physically interpretable and methodologically competitive with the broader engine ML literature.

In terms of predictive performance, the comparative analysis revealed distinct characteristics among the regression models. The Response Surface Methodology (RSM), while computationally efficient, showed limitations in capturing the highly non-linear behavior of emission parameters such as CO and NO_x. This aligns with the known limitations of polynomial-based approaches in

ARTICLE IN PRESS

modelling complex combustion chemistry. In contrast, the XGBoost model effectively handled these non-linearities, significantly reducing prediction errors. The success of tree-based algorithms in modelling engine performance is also in parallel with other recent studies in the literature, which report that Random Forest regression offers higher R^2 values (0.95–0.98) compared to Polynomial regression for complex fuel blends [44].

This new study proves that tree-based algorithms are also superior to Polynomial Regression (PR) in predicting emission parameters (HC, CO) (R^2 = 0.936-0.998 vs. R^2 = 0.78-0.891). In this study, we also found that XGBoost outperformed RSM (polynomial-based) in predicting emission parameters (CO, NO_x). Therefore, this new study directly and powerfully corroborates your findings [43].

However, the Stacking Ensemble emerged as the superior approach by effectively integrating the global trend-capturing ability of RSM with the local pattern-recognition power of XGBoost. The Stacking model maintained consistently low error rates across all parameters, regardless of their magnitude, demonstrating superior stability compared to single-model approaches. Furthermore, the reliability of the target variable (the TOPSIS score) was validated through a comprehensive sensitivity analysis. The fact that the decision rankings remained stable across different weighting methods (Entropy vs. Equal Weighting) confirms that the ML model is learning a robust and objective definition of "optimal performance." This dual validation, both MCDM inputs and the ML outputs, strengthens the study's conclusion that hybrid AI-MCDM frameworks can reliably substitute for expensive and time-consuming experimental campaigns. Despite these promising results, it is worth noting that the performance gain of Stacking over XGBoost was marginal for some highly linear parameters (e.g., specific fuel consumption). This suggests that for simpler tasks, advanced boosting algorithms alone may suffice. However, for holistic optimization where emission trade-offs are critical, the Stacking architecture provides an essential layer of reliability. Future research could expand this framework by incorporating deep learning-based meta-learners or testing the model's adaptability on different engine types and alternative fuel blends to further generalize its applicability.

The increasing structural complexity of modern internal combustion engines, together with the variability in fuel properties introduced by alternative fuels, has significantly intensified the challenges associated with engine calibration. For instance, recent studies demonstrate that

ARTICLE IN PRESS

exploiting the full potential of advanced renewable fuels, such as hydrotreated vegetable oil (HVO), to achieve high-efficiency and ultra-low emission targets necessitates highly specific and optimized combustion control strategies [47]. Traditional Design of Experiments approaches used to populate static, map-based control strategies require an extensive number of dynamometer tests as the calibration search space expands, resulting in substantial time and cost burdens during the development process. To reduce this experimental burden, previous studies have successfully coupled 1D engine cycle simulations with heuristic algorithms, such as Multi-Objective Particle Swarm Optimization (MOPSO), to virtually calibrate operating parameters for simultaneously boosting in-cylinder pressure and controlling NO_x emissions [48]. Building upon such virtual optimization concepts, the TOPSIS–Stacking hybrid framework proposed in this study offers a data-driven and holistic solution by enabling an integrated evaluation of engine performance under limited experimental conditions.

During the development phase, the proposed model is trained on a sparse experimental dataset to virtually generate high-fidelity performance maps that represent the full engine operating range. This virtual calibration capability allows the trade-offs between power output, efficiency, and emissions to be analyzed without the need for exhaustive physical testing, thereby enabling the identification of optimal engine control parameters in a virtual environment. As a direct consequence, dynamometer operating time and fuel consumption during the calibration process can be significantly reduced, leading to a more efficient engine development cycle.

Furthermore, the computational efficiency of the Stacking Ensemble architecture facilitates the extension of this framework to on-board, real-time applications. With the increasing processing capabilities of modern Engine Control Units, the proposed model can be deployed on the vehicle to estimate unified performance indicators, such as the TOPSIS score, that cannot be directly measured during operation. This capability provides a strong foundation for transitioning from static, pre-calibrated maps toward adaptive, model-based control strategies. Consequently, in internal combustion engines, ignition timing and combustion-related control parameters can be dynamically adjusted in response to fuel properties and environmental conditions, allowing the engine to operate consistently close to its holistic optimal operating point under all conditions.

Despite these contributions, several limitations of the proposed framework should be acknowledged to properly contextualize its generalizability. The model was trained and validated

ARTICLE IN PRESS

exclusively on data collected from a single-cylinder Otto-cycle engine operating under a specific set of ethanol/water blend ratios, two compression ratios, and five engine speed levels. Consequently, its predictive validity cannot be assumed to extend directly to other engine architectures such as multi-cylinder, compression-ignition, or turbocharged configurations, nor to other fuel families such as biodiesel blends, gasoline, or hydrogen mixtures. The TOPSIS score computed in this study is also inherently tied to the specific set of eight criteria selected and the entropy-based weighting applied; a different criteria set or weighting strategy would yield a different composite score, and the trained model would not transfer to that revised formulation without retraining. Additionally, although the 5-Fold Group Cross-Validation strategy employed here ensures that no experimental repetition leakage occurs between folds and provides a conservative estimate of generalization performance, it remains an internal validation procedure applied to a single dataset. External validation on an independent engine dataset was not performed in this study, which represents a limitation that future work should address. The robustness of the TOPSIS rankings to weighting method choice was demonstrated through Spearman rank correlation analysis across four weighting strategies, yielding correlation coefficients between 0.91 and 0.97, which provides indirect evidence that the learned composite score reflects stable underlying performance trade-offs rather than an artifact of the specific weighting scheme. However, this analysis does not substitute for cross-engine or cross-fuel validation. Future research should therefore focus on applying the proposed framework to additional engine types and fuel blends, incorporating physics-informed constraints into the machine learning architecture to improve transferability, and testing the model's performance under transient operating conditions beyond the steady-state measurements available in the present dataset.

ARTICLE IN PRESS

5. Conclusions

- This study introduces a novel hybrid framework that evaluates internal combustion engine performance using TOPSIS and predicts these decision scores with high accuracy through a Stacking Ensemble (RSM + XGBoost) model.
- A 5-Fold Group Cross-Validation strategy prevented data leakage arising from repeated experimental measurements, confirming that the model generalizes reliably to unseen engine operating conditions.

ARTICLE IN PRESS

- The Stacking Ensemble outperformed the individual RSM and XGBoost models by combining global trend-capturing with local pattern recognition, maintaining low and stable error rates across all parameters; stable decision rankings across different weighting methods further confirm the robustness of the predicted TOPSIS scores.
- The proposed framework enables virtual calibration that reduces dynamometer testing time and cost, with clear potential for extension to real-time, on-board engine control applications.

Abbreviation	Description
AHP	Analytic Hierarchy Process
AI	Artificial Intelligence
ANN	Artificial Neural Network
BTDC	Before Top Dead Center
BTE	Brake Thermal Efficiency
BSFC	Brake Specific Fuel Consumption
CI	Compression Ignition
CNN	Convolutional Neural Network
CO	Carbon Monoxide
CO_2	Carbon Dioxide
CR	Compression Ratio
CRITIC	Criteria Importance Through Inter-criteria Correlation
DoE	Design of Experiment
HC	Hydrocarbon
HVO	Hydrotreated Vegetable Oil
ICE	Internal Combustion Engine
LSTM	Long Short-Therm Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MOGT	Multi-Objective Game Theory
MOPSO	Multi-Objective Particle Swarm Optimization
MCDM	Multi-Criteria Decision-Making
ML	Machine Learning
MLP	Multi Layer Perceptron
MSE	Mean Squared Error
NO_x	Nitrogen Oxides
NSGA	Non-dominated Sorting Genetic Algorithm
OOF	Out-of-Fold
O_2	Oxygen
MEREC	Method based on the Removal Effects of Criteria
PR	Polynomial Regression
R^2	Coefficient of Determination
RF	Random Forest
RMSE	Root Mean Squared Error
RSM	Response Surface Methodology
SD	Standard Deviation

ARTICLE IN PRESS

SFC	Specific Fuel Consumption
THC	Total/Unburned Hydrocarbons
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
XGBoost	Extreme Gradient Boosting
GPU	Graphics Processing Unit
KDE	Kernel Density Estimation
Z-score	Standardized score (standard deviation units)
WCO	Waste Cooking Oil
CRediT	Contributor Roles Taxonomy

Data Availability Statement

The experimental dataset analyzed in this study is openly available in the Mendeley Data repository at https://doi.org/10.17632/yvb7khhbrj.1, originally published by Torres, Mendes, and Albuquerque [31] under a CC BY 4.0 license. The custom Python code developed in this study for entropy-weighted TOPSIS computation, RSM and XGBoost model training, stacking ensemble construction, and 5-fold Group Cross-Validation is publicly available in the OSF repository at https://doi.org/10.17605/OSF.IO/CERKV.

ARTICLE IN PRESS

Funding Statement

This research received no external funding.

Conflict of Interest Statement

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.

CRediT Author Statement

Ahmet Karaoğlu: Methodology, Software, Formal analysis, Data curation, Validation, Visualization, Writing - original draft; Hüseyin Söyler: Conceptualization, Methodology, Supervision, Writing - review & editing, Project administration.

References

1. Leach, F., Kalghatgi, G., Stone, R. & Miles, P. The scope for improving the efficiency and environmental impact of internal combustion engines. *Transp. Eng.* **1**, 100005 (2020).

ARTICLE IN PRESS

2. Singh, D. *et al.* A review on emissions reduction techniques used in internal combustion engines. *Int. J. Environ. Sustain. Dev.* **20**, 232 (2021).

3. Wallington, T. J., Kaiser, E. W. & Farrell, J. T. Automotive fuels and internal combustion engines: a chemical perspective. *Chem. Soc. Rev.* **35**, 335 (2006).

4. Söyler, H. Boost pressure influence on combustion, emission characteristics, and performance of diesel engines with various fuel types. *Eng. Sci. Technol. Int. J.* **63**, 101983 (2025).

5. Söyler, H., Balki, M. K. & Sayin, C. The impact of injection timing and pressure on a CRDI engine's combustion characteristics. *J. Eng. Res.* S2307187724000439 (2024) doi:10.1016/j.jer.2024.02.015.

6. Demirtaş, G. & Sayın, C. MWCNT katkılı biyodizel, n-oktanol karışımının tek silindirli CRDI bir dizel motorun performans, yanma ve gürültü düzeyine etkisi. *Gazi Üniversitesi Mühendis. Mimar. Fakültesi Derg.* **40**, 899–910 (2025).

7. Verhaegh, J., Kupper, F. & Willems, F. Data-Driven Air-Fuel Path Control Design for Robust RCCI Engine Operation. *Energies* **15**, 2018 (2022).

8. Temür, M. & Sayın, C. Metanol-Dizel Çift Yakıtlı Reaktivite Kontrollü Sıkıştırma ile Ateşlemeli (RCCI) Bir Motorda Farklı Yüklerin Yanma Karakteristiklerine Etkisinin Deneysel İncelenmesi. *Int. J. Adv. Eng. Pure Sci.* **37**, 389–409 (2025).

9. Masum, B. M., Kalam, M. A., Masjuki, H. H., Palash, S. M. & Fattah, I. M. R. Performance and emission analysis of a multi cylinder gasoline engine operating at different alcohol–gasoline blends. *RSC Adv* **4**, 27898–27904 (2014).

10. Abrar, I., Arora, T. & Khandelwal, R. Bioalcohols as an alternative fuel for transportation: Cradle to grave analysis. *Fuel Process. Technol.* **242**, 107646 (2023).

11. Demirtaş, G., Balki, M. K. & Sayin, C. Effects of a CRDI Engine Running on Biodiesel, n-Octanol and Nanoparticle Blended Nanofuel on Performance, Emissions and Combustion. *Arab. J. Sci. Eng.* **49**, 2565–2580 (2024).

12. Olson, A. L., Tunér, M. & Verhelst, S. A Review of Isobutanol as a Fuel for Internal Combustion Engines. *Energies* **16**, 7470 (2023).

ARTICLE IN PRESS

ARTICLE IN PRESS

13. Belgiorno, G., Ianniello, R. & Di Blasio, G. Ethanol Fuel as Enabler for High-Efficiency and Low-Soot Combustion in Dual-Fuel and Blend Modes. *SAE Int. J. Engines* **18**, 521–539 (2025).

14. Kremer, F., Blank, L. M., Jones, P. R. & Akhtar, M. K. A Comparison of the Microbial Production and Combustion Characteristics of Three Alcohol Biofuels: Ethanol, 1-Butanol, and 1-Octanol. *Front. Bioeng. Biotechnol.* **3**, (2015).

15. Karaoğlu, A. & Söyler, H. A New Approach for Improving Biodiesel Conversion Efficiency: A Stacking Ensemble Model Based on Linear Regression Approach with GAN-Enhanced. *Arab. J. Sci. Eng.* https://doi.org/10.1007/s13369-025-10227-5 (2025) doi:10.1007/s13369-025-10227-5.

16. Dam, Q. T. Adaptive PID Controller Design for Velocity Control of a Hydrogen Internal Combustion Engines using RBF Neural Networks. *Eng. Perspect.* **1**, 41–48 (2025).

17. Aliramezani, M., Koch, C. R. & Shahbakhti, M. Modeling, diagnostics, optimization, and control of internal combustion engines via modern machine learning techniques: A review and future directions. *Prog. Energy Combust. Sci.* **88**, 100967 (2022).

18. Salam, S., Choudhary, T., Pugazhendhi, A., Verma, T. N. & Sharma, A. A review on recent progress in computational and empirical studies of compression ignition internal combustion engine. *Fuel* **279**, 118469 (2020).

19. Huang, Q., Liu, J., Ulishney, C. & Dumitrescu, C. E. On the use of artificial neural networks to model the performance and emissions of a heavy-duty natural gas spark ignition engine. *Int. J. Engine Res.* **23**, 1879–1898 (2022).

20. Brusa, A. *et al.* The Enhancement of Machine Learning-Based Engine Models Through the Integration of Analytical Functions. *Energies* **17**, 5398 (2024).

21. Ristow Hadlich, R., Loprete, J. & Assanis, D. A Deep Learning Approach to Predict In-Cylinder Pressure of a Compression Ignition Engine. *J. Eng. Gas Turbines Power* 1–12 (2024) doi:10.1115/1.4064480.

22. Zhao, F. & Hung, D. L. S. Applications of machine learning to the analysis of engine in-cylinder flow and thermal process: A review and outlook. *Appl. Therm. Eng.* **220**, 119633 (2023).

ARTICLE IN PRESS

ARTICLE IN PRESS

23. Corrigan, D. J. *et al.* Engine Knock Detection Methods for Spark Ignition and Prechamber Combustion Systems in a High-Performance Gasoline Direct Injection Engine. *SAE Int. J. Engines* **15**, 883–897 (2022).

24. Taghavifar, H., Roy, S. & Roskilly, A. P. Comparison of game theory and genetic algorithm optimisation schedulers for diesel-hydrogen powered system reconfiguration. *Int. J. Hydrog. Energy* **105**, 1435–1450 (2025).

25. Derikvand, H., Shafiey Dehaj, M. & Taghavifar, H. The effect of different sampling method integrated in NSGA II optimization on performance and emission of diesel/hydrogen dual-fuel CI engine. *Appl. Soft Comput.* **128**, 109434 (2022).

26. Pandey, K. K., Deb, P., Singh, P. & Kumar, S. Experimental and AHP-based assessment of ammonia and hydrogen enrichment in tri-fuel CI engine operation. *Fuel* **427**, 139962 (2027).

27. A hybrid machine learning-enhanced MCDM model for transport safety engineering | Scientific Reports. https://www.nature.com/articles/s41598-025-21297-8.

28. Zhu, Z., Wang, J., Deng, T. & Dai, H. An artificial intelligence-based strategy for multi-objective optimization of internal combustion engine performance and emissions. *Expert Syst. Appl.* **270**, 126472 (2025).

29. Gao, S. *et al.* Improvements of performance and emission characteristics of a diesel engine fueled with diesel/PODE3/*n*-butanol blended fuels by RSM-NSGA III in plateau environment. *Process Saf. Environ. Prot.* **185**, 184–210 (2024).

30. Ashika, T. & Hannah Grace, G. Enhancing heart disease prediction with stacked ensemble and MCDM-based ranking: an optimized RST-ML approach. *Front. Digit. Health* **7**, 1609308 (2025).

31. Torres, D. J. G., Mendes, A. D. S. & Albuquerque, C. Performance and emissions data of an internal combustion engine operating with different ethanol/water mixtures and compression ratios. *Data Brief* **54**, 110390 (2024).

32. Diakoulaki, D., Mavrotas, G. & Papayannakis, L. Determining objective weights in multiple criteria problems: The critic method. *Comput. Oper. Res.* **22**, 763–770 (1995).

ARTICLE IN PRESS

ARTICLE IN PRESS

33. Mukhametzyanov, I. Specific character of objective methods for determining weights of criteria in MCDM problems: Entropy, CRITIC and SD. *Decis. Mak. Appl. Manag. Eng.* **4**, 76–105 (2021).

34. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. (Psychology Press, New York, NY, 2009).

35. Elmachtoub, A. N. & Grigas, P. Smart "Predict, then Optimize". *Manag. Sci.* **68**, 9–26 (2022).

36. Hoxha, J., Çodur, M. Y., Mustafaraj, E., Kanj, H. & El Masri, A. Prediction of transportation energy demand in Türkiye using stacking ensemble models: Methodology and comparative analysis. *Appl. Energy* **350**, 121765 (2023).

37. Wang, X., Zhang, B., Xu, Z., Li, M. & Skare, M. A multi-dimensional decision framework based on the XGBoost algorithm and the constrained parametric approach. *Sci. Rep.* **15**, 4315 (2025).

38. Wu, Y., Xie, Y., Xu, F., Zhu, X. & Liu, S. A runoff-based hydroelectricity prediction method based on meteorological similar days and XGBoost model. *Front. Energy Res.* **11**, 1273805 (2024).

39. Das, S., Tamang, S. K. & Kalita, P. An integrated ANFIS-TOPSIS approach for enhanced performance and emissions characteristics in syngas-diesel powered dual-fuel engine. *Int. J. Hydrog. Energy* **139**, 1116–1132 (2025).

40. Experimental Study of Port Water Injection on GDI Engine Fuel Economics and Emissions | ACS Omega. https://pubs.acs.org/doi/10.1021/acsomega.3c06859.

41. Voris, A. R., Lundberg, M. & Puzinauskas, P. V. Effect of ethanol content, water injection, and compression ratio on combustion and detailed emissions in an SI-GDI engine. *Fuel* **381**, 133200 (2025).

42. Sakthivel, P., Subramanian, K. A. & Mathai, R. Effects of different compression ratios and spark timings on performance and emissions of a two-wheeler with 30% ethanol-gasoline blend (E30). *Fuel* **277**, 118113 (2020).

43. Khayum, N. *et al.* Harnessing machine learning: prediction of emission patterns in a dual fuel diesel engine. *Int. J. Ambient Energy* **47**, 2678557 (2026).

ARTICLE IN PRESS

ARTICLE IN PRESS

44. Pandey, K. K., Khayum, N. & Shaik, J. H. Comparison of machine learning algorithms on a low heat rejection diesel engine running on ternary blends. *J. Renew. Sustain. Energy* **16**, 053101 (2024).

45. Gad, M. S., Soliman, M. S. & Helal, E. B. Hybrid XGBoost-RF-MLP model and PSO optimization for performance and emissions of CI engine using waste cooking biodiesel blends. *Sci. Rep.* **15**, 43832 (2025).

46. Nachippan, M. *et al.* Machine learning predictions for enhancing engine performance and emission using aluminum oxide nano additives in castor biodiesel. *Sci. Rep.* **15**, 36514 (2025).

47. Di Blasio, G., Ianniello, R. & Beatrice, C. Hydrotreated vegetable oil as enabler for high-efficient and ultra-low emission vehicles in the view of 2030 targets. *Fuel* **310**, 122206 (2022).

48. Taghavifar, H. & Mazari, F. 1D diesel engine cycle modeling integrated with MOPSO optimization for improved NOx control and pressure boost. *Energy* **247**, 123517 (2022).

ARTICLE IN PRESS

ARTICLE IN PRESS

Appendix A. Mathematical Formulation of TOPSIS and Entropy Weighting

The TOPSIS procedure is implemented in the following steps:

Step 1: To achieve comparability across diverse criteria, the decision matrix is subjected to a vector normalization process to eliminate scale-related differences and unit inconsistencies, as defined in Equation (1):

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \tag{1}$$

Step 2: In the subsequent stage, the relative importance of each criterion is integrated into the analysis by multiplying the normalized values with their respective weights derived from the Entropy method, yielding the weighted normalized decision matrix as expressed in Equation (2):

$$v_{ij} = w_j \times r_{ij} \tag{2}$$

Step 3: Following the construction of the weighted matrix, the positive ideal solution (A^+) and the negative ideal solution (A^-) are identified to establish the benchmark limits for each criterion, where the selection of maximum or minimum values is determined by the specific nature of the benefit or cost-oriented criteria, as formulated in Equation (3):

$$\begin{aligned} A^+ &= \{v_1^+,\ldots,v_n^+\} \quad \text{where} \quad v_j^+ = \begin{cases} \max(v_{ij}) if\ j \in Benefit \\ \min(v_{ij}) if\ j \in Cost \end{cases} \\ A^- &= \{v_1^-,\ldots,v_n^-\} \quad \text{where} \quad v_j^- = \begin{cases} \min(v_{ij}) if\ j \in Benefit \\ \max(v_{ij}) if\ j \in Cost \end{cases} \end{aligned} \tag{3}$$

Step 4: To quantify the performance of each candidate, the separation measures are determined by calculating the Euclidean distance of each alternative from both the positive ideal and negative ideal solutions, providing a numerical basis for their relative proximity to the optimal benchmarks as shown in Equation (4):

ARTICLE IN PRESS

$$S_i^+ = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^+\right)^2},\ S_i^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^-\right)^2} \tag{4}$$

Step 5: The final relative closeness coefficient for each alternative is determined by calculating the ratio of its distance from the negative ideal solution to the sum of its distances from both ideal benchmarks, where a result approaching unity signifies a superior performance and higher preference ranking, as defined in Equation (5):

$$C_i = \frac{S_i^-}{S_i^+ + S_i^-} \tag{5}$$

To determine the most suitable objective weighting method based on inter-parameter relationships, the average of the absolute Pearson correlation coefficients was calculated using Equation (6). This calculation excludes the diagonal elements of the correlation matrix to prevent self-correlation from inflating the result. The formula used is:

$$\bar{r} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} |r_{ij}| \tag{6}$$

where $\bar{r}$ denotes the average absolute correlation, and r_{ij} represents the correlation coefficient between parameters i and j. Applying this equation to the selected parameter matrix yielded an average correlation value of 41.20%, indicating a moderate to low level of interdependence among the variables.

The entropy weighting method used in this study evaluates the degree of disorder or uncertainty in each parameter to assign objective importance scores. The procedure begins with normalizing the data matrix by column, followed by the computation of entropy values for each parameter. Then, the information utility is calculated by subtracting each entropy score from 1. Finally, the normalized importance weights are obtained by dividing each utility value by the sum of all utilities. The final weight w_j of criterion j is calculated using Equation (7):

$$w_j = \frac{1 - E_j}{\sum_{j=1}^{n} (1 - E_j)} \tag{7}$$