Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Confidence-Weighted Explainable Learning for Vessel Fuel-Consumption Analysis: A Multi-Model Comparative Framework

DOUGLAS AMOBI AMOKE

d.a.amoke2@newcastle.ac.uk

Newcastle University

Paul Simavari

Newcastle University

Syed Mohsen Naqvi

Newcastle University

Article

Keywords: Vessel fuel consumption, Confidence-aware learning, Explainable artificial intelligence (XAI), Maritime energy efficiency, SHAP, Uncertainty quantification

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11388094/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Confidence-Weighted Explainable Learning for Vessel Fuel-Consumption Analysis: A Multi-Model Comparative Framework

Douglas Amobi Amoke[1*], Paul Simavari[1] and Syed Mohsen Naqvi[1]

[1*]School of Engineering, Newcastle University, Claremont Road, Newcastle upon Tyne, NE1 7RU, United Kingdom.

*Corresponding author(s). E-mail(s): d.a.amoke2@newcastle.ac.uk;
Contributing authors: P.Simavari2@newcastle.ac.uk;
Mohsen.Naqvi@newcastle.ac.uk;

Abstract

Improving vessel energy efficiency is essential for reducing fuel consumption, greenhouse-gas emissions, and operating costs while supporting maritime decarbonisation objectives. Although advances in maritime analytics and machine learning have improved vessel monitoring and fuel-consumption prediction, existing studies focus mainly on predictive accuracy and offer limited insight into the engineering, operational, and reliability factors that are associated with vessel energy demand. In addition, confidence-aware learning, uncertainty quantification, explainable artificial intelligence (XAI), and comparative evaluation of machine-learning models remain underexplored within maritime energy-intelligence research. This paper presents a confidence-weighted explainable learning framework for vessel fuel-consumption analysis and conducts a systematic comparison of five model families, namely Random Forest (RF), Extreme Gradient Boosting (XGBoost), Gaussian Process Regression (GPR), Bayesian Ridge Regression (BRR), and Elastic Net, within a unified confidence-aware, explainable, and uncertainty-aware analytical architecture. Confidence integration is formulated as a reliability-weighted signal-fusion problem in which vessel-level confidence scores derived from heterogeneous data sources are incorporated directly into model learning. The Vessel Energy Efficiency Intelligence (VEEI) combines statistical driver analysis, duty-cycle intelligence, explainable machine learning, and uncertainty quantification to generate interpretable vessel-level energy intelligence. Experiments use a tidal River Thames inland-waterway dataset comprising 26 representative vessels and 131 engineering, operational, propulsion, fuel-consumption, and confidence-related attributes. Spearman rank

correlation analysis identifies operating hours ($\boldsymbol{OH}$) and installed engine power ($\boldsymbol{EP}$) as the only statistically significant direct fuel-consumption drivers ($\boldsymbol{p < 0.05}$), while medium-cruise operation exhibits the strongest significant duty-cycle association with fuel demand. Among the evaluated models, Bayesian Ridge Regression achieves the highest predictive performance with an MAE of 41.23 L, an RMSE of 59.87 L, and an $\boldsymbol{R^2}$ of 0.881, followed by XGBoost ($\boldsymbol{R^2 = 0.876}$) and Random Forest ($\boldsymbol{R^2 = 0.873}$). Confidence-weighted learning consistently improves predictive performance across all model families relative to their unweighted counterparts. Cross-model SHAP analysis identifies vessel draft and duty-cycle behaviour as the dominant explanatory factors, while uncertainty analysis consistently highlights vessels with elevated predictive uncertainty. The results show that operational utilisation, propulsion characteristics, vessel geometry, duty-cycle behaviour, and data reliability jointly influence vessel fuel consumption. By integrating confidence-aware learning, multi-model explainability, and uncertainty quantification in a unified framework, the proposed approach provides transparent, trustworthy vessel-level energy intelligence to support operational optimisation, fleet benchmarking, and evidence-based maritime decarbonisation planning.

Keywords: Vessel fuel consumption, Confidence-aware learning, Explainable artificial intelligence (XAI), Maritime energy efficiency, SHAP, Uncertainty quantification

1 Introduction

Maritime transportation remains the backbone of global trade, carrying more than 80% of world merchandise by volume and serving as a critical component of the global economy (Amoke et al. 2025a,b; United Nations Conference on Trade and Development 2024; International Maritime Organization 2023). Despite being recognised as one of the most energy efficient transportation modes, continued growth in vessel traffic has intensified concerns regarding fuel consumption, greenhouse gas (GHG) emissions, and environmental sustainability International Maritime Organization (2023); Olmer et al. (2017). However, international regulatory frameworks such as the International Maritime Organisation (IMO) 2023 Greenhouse Gas Strategy, the Energy Efficiency Existing Ship Index (EEXI), the Carbon Intensity Indicator (CII), and FuelEU Maritime have placed increasing emphasis on improving vessel energy efficiency and reducing emissions across maritime fleets International Maritime Organization (2023); European Maritime Safety Agency and European Environment Agency (2025). Consequently, vessel operators, regulators, and policymakers need reliable, interpretable analytical tools that identify the factors driving vessel energy demand and support evidence based decarbonisation strategies.

Recent advances in digitalisation, artificial intelligence, and maritime sensing technologies have significantly expanded maritime analytics. In particular, Automatic Identification System (AIS) data have enabled large scale studies of vessel traffic monitoring, route extraction, anomaly detection, trajectory prediction, behavioural analysis, and maritime situational awareness Tu et al. (2021); Pallotta et al. (2013);

Gao et al. (2021); Feng et al. (2022). Machine learning techniques have similarly demonstrated considerable potential for fuel consumption prediction, vessel performance modelling, voyage optimisation, and operational decision support Wang et al. (2020); Molland et al. (2017). While these developments have improved predictive capabilities, most existing studies remain predominantly movement centric and prediction oriented, focusing on forecasting accuracy rather than understanding the underlying engineering and operational mechanisms driving vessel energy demand. This gap matters because vessels operating in similar geographical environments may exhibit substantially different fuel consumption characteristics due to variations in hull geometry, displacement, propulsion system configuration, operational schedules, duty-cycle behaviour, cargo-loading conditions, and hotel-load requirements Molland et al. (2017); Psaraftis and Kontovas (2013). Consequently, accurate prediction alone is insufficient for supporting practical decision making.

Maritime stakeholders increasingly require transparent, explainable analytics that identify why vessels consume energy, which operational characteristics drive fuel demand, and where opportunities for efficiency improvement exist. Although explainable artificial intelligence (XAI) methods such as SHapley Additive exPlanations (SHAP) have emerged as powerful tools for interpreting machine learning models Lundberg and Lee (2017); Sundararajan et al. (2017), their application within vessel energy assessment remains relatively limited. An additional challenge arises from the heterogeneous nature of maritime data. Vessel energy datasets are frequently compiled from multiple sources, including vessel registries, engineering specifications, manufacturer documentation, operational records, fuel consumption reports, inferred parameters, and engineering estimates. As a result, the reliability and quality of individual variables can vary considerably across vessels and across features within the same vessel record. This variability introduces uncertainty into both analysis and prediction, yet vessel energy modelling frameworks rarely incorporate data reliability explicitly. Consequently, most existing studies implicitly assume that all observations possess equal reliability, potentially leading to biased conclusions and reduced model trustworthiness Kendall and Gal (2017); Gal and Ghahramani (2016).

Another gap in the literature is the predominance of single model investigations. Different machine learning families possess fundamentally different modelling characteristics and assumptions. Tree-based ensemble methods such as Random Forest (RF) (2001) and Extreme Gradient Boosting (XGBoost) Chen and Guestrin (2016) are particularly effective at learning nonlinear interactions in tabular datasets. In contrast, regularised linear models such as Elastic Net Zou and Hastie (2005) often generalise well in high-dimensional, small sample regimes. Probabilistic approaches including Bayesian Ridge Regression (BRR) Bishop (2006) and Gaussian Process Regression (GPR) Rasmussen and Williams (2006) provide native predictive distributions and principled uncertainty estimates that may be advantageous when datasets are limited in size. Although comparative evaluations of these model families have been reported in energy forecasting and other engineering applications Kong et al. (2019); Chen and Guestrin (2016), comparable investigations remain notably scarce within vessel fuel-consumption analysis. Moreover, confidence-aware learning, uncertainty quantification, explainable artificial intelligence, propulsion-performance assessment,

vessel similarity analysis, and operational-behaviour intelligence are typically studied independently rather than integrated within a unified vessel-energy assessment framework. Consequently, important interactions among vessel design characteristics, operational behaviour, fuel-consumption drivers, data reliability, and predictive uncertainty remain insufficiently understood. This fragmentation limits maritime decision support systems from providing holistic vessel-level energy intelligence.

To address these limitations, this paper proposes a Confidence-Weighted Explainable Learning framework for vessel fuel-consumption analysis within a broader Vessel Energy Efficiency Intelligence (VEEI) architecture. The proposed VEEI framework integrates confidence-aware analytics, fuel-consumption driver analysis, duty-cycle intelligence, propulsion-performance assessment, confidence-weighted machine learning, uncertainty quantification, explainable artificial intelligence, and comparative model evaluation within a unified analytical pipeline. Unlike existing studies that typically rely on a single predictive model, this work systematically evaluates five complementary model families, namely Random Forest, XGBoost, Gaussian Process Regression, Bayesian Ridge Regression, and Elastic Net, under a common confidence-aware and explainability-driven framework. We incorporate confidence information from heterogeneous data sources through a reliability-weighted learning strategy, allowing higher-quality observations to exert greater influence on model development.

The principal contributions of this paper are fourfold. First, we develop a unified confidence-aware vessel-energy intelligence framework that integrates engineering characteristics, operational behaviour, fuel-consumption information, and data reliability within a common analytical environment. Second, we introduce confidence-weighted learning as a reliability-aware signal-fusion mechanism and systematically evaluate its effectiveness across five machine-learning model families. Third, we employ a comprehensive SHAP-based explainability framework to compare feature attributions across models and identify the dominant operational and engineering drivers of vessel fuel consumption. Fourth, we compare multiple uncertainty-quantification approaches, including ensemble variance, posterior predictive distributions, and bootstrap-based estimation, to assess predictive confidence and model trustworthiness at the vessel level. We evaluate the framework using a River Thames inland-waterway dataset comprising 26 representative vessels and 131 engineering, operational, propulsion, fuel-consumption, and confidence-related variables. By integrating confidence-aware learning, explainable artificial intelligence, uncertainty quantification, and multi-model evaluation, this work aims to provide transparent vessel-level energy intelligence to support operational optimisation, fleet benchmarking, and maritime decarbonisation decision-making.

The remainder of this paper is organised as follows. Section II reviews the related literature and identifies the research gaps motivating this work. Section III presents the problem formulation and the proposed VEEI framework. Section IV describes the dataset, confidence-scoring methodology, analytical procedures, and multi-model learning architecture. Section V presents the experimental results and comparative analyses. Section VI discusses the key findings and practical implications for vessel-energy assessment and maritime decarbonisation. Finally, Section VII concludes the paper and outlines future research directions.

2 Related Works

2.1 Vessel Fuel-Consumption Modelling

Research on vessel fuel-consumption modelling has expanded substantially over the last decade, driven by growing interest in maritime decarbonisation, operational efficiency, and environmental sustainability. Traditional ship-energy studies have shown that fuel consumption depends on vessel geometry, displacement, propulsion characteristics, operating speed, loading conditions, and voyage profiles Molland et al. (2017); Psaraftis and Kontovas (2013). Fundamental naval architecture and marine engineering research has demonstrated that hull resistance, propulsion efficiency, and operational loading directly influence vessel energy demand, thereby providing the theoretical basis for ship-performance analysis and energy-efficiency assessment Molland et al. (2017); Stopford (2009). In parallel, emissions modelling studies have employed vessel operational profiles, propulsion-system characteristics, and duty-cycle distributions to estimate fuel consumption and associated greenhouse gas emissions across a range of vessel classes and operating environments Cooper (2003); Jalkanen et al. (2012). Such approaches have contributed significantly to understanding fleet-level environmental impacts and have informed the development of regulatory frameworks for maritime decarbonisation.

However, these methods typically rely on empirical assumptions, engineering approximations, or aggregated operational profiles and often provide limited insight into vessel-specific behavioural differences. The increasing availability of AIS data, onboard monitoring systems, and digital ship-management platforms has accelerated the development of data-driven vessel-energy models. Machine learning techniques have been applied to fuel-consumption prediction, voyage optimisation, speed management, route planning, and operational decision support Wang et al. (2020); Simavari et al. (2026). The work in Wang et al. (2020) demonstrated the effectiveness of machine-learning methods for ship fuel-consumption prediction using voyage data, highlighting the potential of data-driven approaches for maritime energy management.

Similarly, recent studies have shown that artificial intelligence techniques can improve prediction accuracy by modelling complex nonlinear relationships among operational and environmental variables. Despite these advances, most fuel-consumption studies remain prediction-oriented, emphasising minimising forecasting error rather than understanding the engineering and operational mechanisms driving energy consumption. Furthermore, many existing models consider only subsets of available vessel information, typically focusing on vessel dimensions, propulsion characteristics, environmental conditions, or operational parameters in isolation. Consequently, interactions among vessel design, propulsion performance, duty-cycle behaviour, operational utilisation, and fuel consumption remain insufficiently understood. In addition, comparatively little attention has been devoted to incorporating data reliability information into vessel-energy modelling, even though maritime datasets are frequently assembled from heterogeneous and imperfect information sources.

2.2 Explainable Maritime Artificial Intelligence

As machine-learning models become increasingly sophisticated, the need for transparency and interpretability has emerged as a major research challenge. In many operational domains, highly accurate predictive models often function as black-box systems whose decision-making processes are difficult to understand or verify. Explainable Artificial Intelligence (XAI) seeks to address this limitation by providing interpretable representations of how model inputs influence predictions Lundberg and Lee (2017); Sundararajan et al. (2017). Among current XAI techniques, SHapley Additive exPlanations (SHAP) has become one of the most widely adopted approaches because of its strong theoretical foundation in cooperative game theory and its ability to provide both global and local model explanations Lundberg and Lee (2017). SHAP quantifies each input variable's contribution to an individual prediction while maintaining consistency and local accuracy.

Consequently, SHAP has been successfully applied across numerous engineering, energy, healthcare, and transportation applications where model interpretability is essential. Within maritime research, however, the adoption of explainable machine learning remains relatively limited. Existing AIS-based studies have largely focused on vessel trajectory classification, anomaly detection, traffic prediction, route extraction, and navigation intelligence Tu et al. (2021); Pallotta et al. (2013); Gao et al. (2021). Although these studies contribute significantly to maritime situational awareness, they generally prioritise predictive performance over interpretability. As a result, relatively little effort has focused on explaining fuel-consumption behaviour, identifying dominant energy drivers, or interpreting nonlinear interactions among vessel characteristics. Another limitation is that researchers often investigate explainability using a single machine-learning architecture. Consequently, it remains unclear whether identified fuel-consumption drivers are model-specific artefacts or genuinely robust explanatory variables. Cross-model explainability analysis has received comparatively little attention in maritime energy research despite its potential to improve confidence in feature-attribution results. The absence of systematic comparative explainability studies therefore represents an important gap in the current literature.

2.3 Confidence-Aware Learning

The growing availability of heterogeneous and multimodal datasets has increased interest in confidence-aware and reliability-aware learning approaches. In many real-world applications, observations originate from multiple sources with varying levels of reliability, completeness, and uncertainty. Treating all observations as equally trustworthy can bias parameter estimation and degrade model performance, particularly when datasets contain inferred, estimated, or incomplete information. Confidence-aware learning addresses this problem by explicitly incorporating data-quality information into training. One common approach involves assigning observation-specific weights that reflect source reliability, measurement confidence, or expected information quality Amoke and Naqvi (2026); Kendall and Gal (2017).

Related concepts have been explored in sensor fusion, remote sensing, autonomous systems, medical imaging, and information fusion applications, where reliability-weighted observations have improved predictive robustness and decision reliability. Within maritime analytics, confidence-aware methodologies remain comparatively underdeveloped. Most existing vessel-energy studies implicitly assume that all recorded variables possess equal reliability despite the heterogeneous nature of maritime datasets. In practice, vessel records frequently combine information from official registries, engineering estimates, manufacturer specifications, inferred operational parameters, historical records, and manually curated databases. These sources may differ substantially in reliability, creating uncertainty that model development rarely accounts for. Furthermore, existing maritime studies typically incorporate data-quality procedures only during preprocessing stages through filtering, cleaning, or missing-value imputation. Comparatively few studies propagate reliability information throughout the modelling process so that confidence directly influences learning, prediction, explainability, and uncertainty assessment. Consequently, the potential value of confidence-weighted learning within vessel-energy intelligence systems remains largely unexplored.

2.4 Uncertainty Quantification

Uncertainty quantification (UQ) has emerged as an important research direction in machine learning because prediction accuracy alone does not indicate model confidence. In operational decision-support systems, understanding the reliability of model predictions is often as important as the predictions themselves. This consideration is particularly important in maritime applications, where energy-management decisions can have significant economic, operational, and environmental consequences. The machine-learning literature has proposed several uncertainty-estimation approaches. Ensemble-based models such as Random Forests commonly estimate uncertainty through prediction variability across individual trees (2001).

Similarly, researchers have widely used uncertainty proxies based on ensemble variance in tree-ensemble and boosting frameworks Shang et al. (2014). While computationally attractive, these methods provide approximate uncertainty estimates that primarily reflect disagreement among model components. Probabilistic learning approaches provide a more principled alternative. Gaussian Process Regression (GPR) models predictions as probability distributions and naturally produces posterior predictive means and variances Rasmussen and Williams (2006). Likewise, Bayesian Ridge Regression (BRR) estimates posterior distributions over model parameters and provides closed-form predictive uncertainty estimates Bishop (2006). These approaches are particularly attractive in small-sample settings because they explicitly capture uncertainty arising from limited data and parameter estimation.

Although uncertainty quantification has received increasing attention in machine learning and artificial intelligence research Gal and Ghahramani (2016); Kendall and Gal (2017), its application within vessel-energy assessment remains relatively limited. Existing studies often report traditional performance metrics such as MAE, RMSE, and R^2 but provide little information on prediction confidence. Moreover, comparative evaluations of uncertainty-estimation methods across multiple machine-learning

families are rare in maritime-energy research. Consequently, understanding remains limited of how different uncertainty quantification approaches behave when applied to vessel fuel-consumption analysis and whether different models identify similar high-uncertainty vessels.

2.5 Research Gaps

Although significant advances have been made in vessel fuel-consumption modelling, maritime artificial intelligence, confidence-aware analytics, and uncertainty quantification, several important limitations remain. As summarised in Table 1, existing studies largely emphasise predictive performance Wang et al. (2020); Molland et al. (2017), while giving comparatively less attention to explainability, confidence-aware learning, uncertainty quantification, and systematic cross-model evaluation. Moreover, researchers often analyse engineering, propulsion, operational, and reliability-related information independently, limiting understanding of the interactions that govern vessel fuel-consumption behaviour Cooper (2003); Jalkanen et al. (2012); Wang et al. (2020). To address these gaps, this paper proposes a Confidence-Weighted Explainable Learning framework embedded within a broader Vessel Energy Efficiency Intelligence (VEEI) architecture. The framework integrates confidence-aware learning, statistical fuel-consumption analysis, duty-cycle intelligence, propulsion-performance assessment, vessel similarity analysis, operational-behaviour clustering, explainable artificial intelligence, and uncertainty quantification within a unified vessel-energy assessment environment. Furthermore, RF, XGBoost, GPR, BRR, and Elastic Net are systematically evaluated under a common confidence-aware pipeline, enabling a comprehensive comparison of predictive performance, explainability, and uncertainty quantification for vessel fuel-consumption analysis.

3 Experiment Formulation and VEEI Framework

The research gaps identified in Section II highlight the need for a unified vessel-energy assessment framework that incorporates engineering characteristics, operational behaviour, data reliability, explainability, and predictive uncertainty. Existing vessel fuel-consumption studies primarily focus on prediction accuracy and rarely account for variations in data quality, model transparency, or predictive confidence. To address these limitations, this paper frames vessel fuel-consumption analysis as a confidence-aware, explainable learning problem within a broader Vessel Energy Efficiency Intelligence (VEEI) framework. Consider a fleet comprising N vessels. Each vessel i is represented by a feature vector

$$\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{iM}]^T, \tag{1}$$

where $\mathbf{x}_i \in \mathbb{R}^M$ contains engineering, operational, propulsion, duty-cycle, and vessel-specific attributes, and M denotes the total number of explanatory variables. We let $y_i \in \mathbb{R}$ denote the corresponding confidence rated vessel level average daily fuel consumption benchmark and $w_i \in [0, 1]$ denote the vessel-confidence score derived from the reliability of the underlying data sources. The objective is not only to estimate

Table 1 Summary of Research Gaps and Corresponding Contributions of the Proposed VEEI Framework

Gap	Limitation in Existing Literature	Contribution of This Work
G1	Existing vessel fuel-consumption studies primarily emphasise predictive accuracy and fuel-demand estimation, with limited focus on interpreting the underlying engineering and operational drivers of vessel energy demand Wang et al. (2020); Molland et al. (2017).	Combines statistical analysis and explainable learning to identify and interpret key fuel-consumption drivers.
G2	Engineering, propulsion, operational, duty-cycle, and fuel-consumption information are often analysed separately, limiting understanding of their interactions Cooper (2003); Jalkanen et al. (2012); Wang et al. (2020).	Develops a unified VEEI framework integrating engineering, operational, propulsion, behavioural, and energy-related information.
G3	Applications of explainable artificial intelligence in maritime energy assessment remain limited and are typically restricted to single-model investigations Lundberg and Lee (2017); Tu et al. (2021).	Provides cross-model explainability analysis using SHAP across multiple machine-learning families.
G4	Data-reliability information from heterogeneous maritime sources is rarely incorporated directly into model learning, despite its potential influence on predictive performance Kendall and Gal (2017); Gal and Ghahramani (2016).	Introduces confidence-weighted learning through reliability-based signal fusion and sample weighting.
G5	Most vessel-energy studies report prediction accuracy metrics without systematically quantifying predictive uncertainty or comparing alternative uncertainty-estimation strategies Gal and Ghahramani (2016); Kendall and Gal (2017).	Integrates and compares ensemble-, Bayesian-, Gaussian-process-, and bootstrap-based uncertainty quantification approaches.
G6	Comparative evaluations of nonlinear ensemble models, probabilistic learning methods, and regularised linear models within a common maritime-energy framework remain scarce (2001); Chen and Guestrin (2016); Rasmussen and Williams (2006); Zou and Hastie (2005).	Performs a systematic comparison of RF, XGBoost, GPR, BRR, and Elastic Net under a unified confidence-aware and explainable framework.

vessel fuel consumption but also to identify and explain the factors associated with vessel energy demand while accounting for variations in data reliability. Accordingly, the learning task is formulated as

$$f : (\mathbf{x}_i, w_i) \rightarrow y_i, \tag{2}$$

where $f(\cdot)$ denotes a confidence-aware predictive model that incorporates both vessel characteristics and reliability information during learning. Beyond prediction, the framework identifies fuel-consumption drivers, explains model behaviour, and quantifies predictive uncertainty. These complementary objectives provide a more comprehensive assessment of vessel energy demand than predictive accuracy alone. Fig. 1 illustrates the proposed VEEI framework.

The framework consists of five interconnected layers. The first layer performs confidence-aware data integration by combining vessel registry information, engineering specifications, operational records, propulsion characteristics, fuel-consumption

reports, and derived attributes while preserving source-reliability metadata. The second layer extracts operational and engineering intelligence through fuel-consumption driver analysis, duty-cycle analysis, propulsion-performance assessment, vessel-similarity learning, and operational-behaviour profiling. The third layer constitutes the

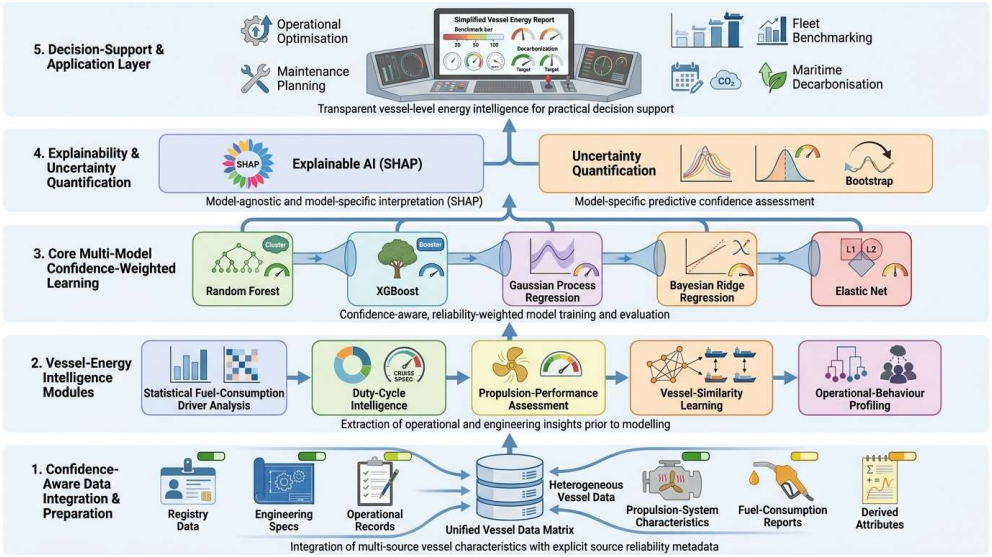


Fig. 1 Proposed Vessel Energy Efficiency Intelligence (VEEI) framework for confidence-aware vessel fuel-consumption analysis.

core confidence-weighted learning component and systematically evaluates five complementary model families: Random Forest (RF) (2001), Extreme Gradient Boosting (XGBoost) Chen and Guestrin (2016), Gaussian Process Regression (GPR) Rasmussen and Williams (2006), Bayesian Ridge Regression (BRR) Bishop (2006), and Elastic Net (EN) Zou and Hastie (2005). These models represent nonlinear ensemble learning, probabilistic inference, Bayesian learning, and regularised linear modelling within a common confidence-aware pipeline. The fourth layer integrates explainability and uncertainty quantification. Explainable artificial intelligence is implemented using SHapley Additive exPlanations (SHAP) Lundberg and Lee (2017) to provide both global and local interpretations of model behaviour, while model-specific uncertainty estimators quantify predictive confidence at vessel level Amoke et al. (2027). The final layer transforms these outputs into decision-support intelligence for operational optimisation, maintenance planning, fleet benchmarking, performance assessment, and maritime decarbonisation. By integrating confidence-aware learning, explainability, uncertainty quantification, and multi-model comparative analysis in a unified architecture, the proposed VEEI framework provides transparent, reliable vessel-level energy intelligence. Section 4 details the confidence-scoring methodology, analytical procedures, model formulations, explainability methods, and uncertainty-quantification strategies.

4 Methodology

This section describes the dataset, confidence-aware preprocessing procedures, analytical modules, multi-model learning framework, explainability methodology, and uncertainty-quantification techniques used in the proposed Vessel Energy Efficiency Intelligence (VEEI) framework. Consistent with the problem formulation in Section 3, the methodology generates interpretable vessel-level energy intelligence by integrating statistical analysis, operational behaviour characterisation, confidence-aware machine learning, explainability analysis, and prediction uncertainty assessment.

4.1 Dataset and Confidence Scoring

We evaluate the proposed VEEI framework using a locally compiled River Thames inland-waterway vessel dataset obtained from the Maritime Engineering Department at Newcastle University, United Kingdom. Unlike conventional AIS datasets that primarily capture vessel trajectories and navigational information, this dataset includes vessel-specific engineering, operational, propulsion, fuel-consumption, and reliability information. The dataset comprises $N = 26$ representative inland-waterway vessels and $M = 131$ attributes describing vessel geometry, propulsion characteristics, fuel systems, duty-cycle distributions, loading conditions, hotel-load requirements, operational schedules, and fuel-consumption information.

The fuel consumption target represents a confidence rated, vessel-level average daily fuel consumption benchmark rather than a direct, high-frequency fuel flow measurement. The benchmark draws on the most credible information available for each vessel, including official records, operator-provided information, engineering calculations, and estimates based on comparable vessel characteristics. Feature-level confidence scores record both source quality and derivation methodology, allowing the framework to retain practical fleet coverage while reducing the influence of lower-confidence observations during model fitting, explainability analysis, and uncertainty quantification. The duty-cycle and fuel-driver analyses therefore identify statistical associations with the benchmark target and do not constitute vessel-specific validation of transient fuel consumption at individual engine loads or duty-cycle states.

The analysed vessels operate within the tidal River Thames and represent a diverse collection of passenger, commercial, and service vessels. A distinguishing characteristic of the dataset is that information originates from heterogeneous sources, including official vessel registries, engineering specifications, manufacturer documentation, inferred operational records, engineering estimates, and derived operational metrics. Consequently, attribute reliability varies across vessels and information sources. To account for this variability, we explicitly model vessel confidence through a reliability-weighted signal-fusion approach. We define

$$x_{ij} \in \mathbb{R} \tag{3}$$

as the value of feature j for vessel i. The reliability score assigned to feature j is described as

$$R_{ij} = CS_{ij} \times WCS_{ij}, \tag{4}$$

where CS_{ij} denotes the confidence score assigned to the source providing feature j, and WCS_{ij} represents the weighted contribution score reflecting feature relevance.

The resulting vessel-level confidence score is computed as

$$w_i = \frac{1}{M} \sum_{j=1}^{M} R_{ij}, \tag{5}$$

where $w_i \in [0, 1]$ denotes the overall confidence score of vessel i, R_{ij} represents the confidence contribution of attribute j for vessel i, M is the total number of confidence-rated attributes considered in the assessment, and i and j index vessels and attributes, respectively. Equation (5) transforms heterogeneous source reliability into a single quantitative confidence metric, which we then incorporate throughout the analytical workflow. We retain the resulting confidence scores for confidence-aware modelling, explainability assessment, and uncertainty analysis.

4.2 Confidence-Aware Data Preparation

Before our analysis, we remove records with missing target fuel-consumption values. We impute numerical variables with the median and categorical variables with the mode. We select median-based imputation because it is robust to outliers and skewed distributions commonly found in maritime engineering datasets. To ensure consistency across model families, explanatory variables are standardised according to

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}, \tag{6}$$

where μ_j and σ_j denote the mean and standard deviation of feature j, respectively. We standardise Gaussian Process Regression (GPR), Bayesian Ridge Regression (BRR), and Elastic Net (EN), whereas we train Random Forest (RF) and XGBoost on their original scales because tree-based methods are invariant to monotonic feature transformations. We retain the confidence scores from (5) throughout all subsequent analytical stages. Furthermore, we partition vessels into high- and low-confidence groups using the median confidence value as the decision threshold to support confidence-aware exploratory analysis.

4.3 Fuel-Consumption Driver Analysis

Before predictive modelling, we conduct statistical analysis to identify vessel characteristics significantly associated with fuel consumption. Because of the relatively small sample size and the lack of guaranteed normality, we use Spearman rank correlation, a nonparametric measure of monotonic association Spearman (1904).

We denote y_i as the reported fuel consumption for observation i, x_j as explanatory variable j, and N as the total number of observations. The Spearman rank-correlation coefficient between x_j and fuel consumption is computed as

$$\rho_j = 1 - \frac{6 \sum_{i=1}^{N} d_i^2}{N(N^2 - 1)}, \tag{7}$$

where ρ_j denotes the Spearman correlation coefficient for explanatory variable j, and d_i represents the difference between the ranks of x_j and y_i for observation i. This analysis considers key engineering and operational variables including vessel length (VL), beam (B), draft (D), age (A), engine power (EP), engine speed (ES), vessel speed (VS), fuel-tank capacity (FTC), operating hours per day (OH), and operating days per year (OD). We evaluate statistical significance at $\alpha = 0.05$.

4.4 Duty-Cycle Intelligence

Following the fuel-consumption driver analysis, we evaluate the vessel operational behaviour, which is represented through a duty-cycle vector

$$\mathbf{X}_{duty} = \{H, I, LC, MC, HC, XC\}, \tag{8}$$

where H, I, LC, MC, HC, and XC represent hotel-load, idle, low-cruise, medium-cruise, high-cruise, and maximum-cruise operation, respectively. Duty-cycle distributions concisely represent vessel utilisation intensity and have been used extensively in maritime emissions and energy studies Cooper (2003); Jalkanen et al. (2012). We again apply Spearman analysis to quantify relationships between operational-state proportions and average daily fuel consumption (L/day). In addition, we generate the duty-cycle interaction matrix shown in Fig. 2 by evaluating the pairwise Spearman correlations to provide insight into operational-state dependencies.

4.5 Vessel Similarity Analysis

To identify vessels with similar engineering and operational characteristics, the VEEI framework employs cosine similarity. We let $\mathbf{x}_i \in \mathbb{R}^M$ and $\mathbf{x}_k \in \mathbb{R}^M$ denote the feature vectors associated with vessels i and k, respectively, where M is the total number of explanatory variables.

The similarity between vessels is quantified using cosine similarity,

$$S_{ik} = \frac{\mathbf{x}_i^T \mathbf{x}_k}{\|\mathbf{x}_i\|_2 \|\mathbf{x}_k\|_2}, \tag{9}$$

where S_{ik} denotes the cosine similarity between vessels i and k, $\mathbf{x}_i$ and $\mathbf{x}_k$ are the corresponding standardised feature vectors containing engineering and operational attributes, $\mathbf{x}_i^T \mathbf{x}_k$ denotes their dot product, and $\|\mathbf{x}_i\|_2$ and $\|\mathbf{x}_k\|_2$ represent the Euclidean (L_2) norms of the respective feature vectors. Values of S_{ik} approaching unity indicate highly similar operational-energy profiles, whereas smaller values indicate greater dissimilarity. The resulting similarity matrix supports vessel benchmarking, fleet profiling, and identifying atypical vessels within the analysed fleet.

4.6 Operational Behaviour Clustering

To further characterise vessel-energy behaviour, the VEEI framework employs K-means clustering to discover latent operational-energy patterns across the fleet.

Operational-behaviour clustering matters because vessels operating in the same environment may exhibit substantially different energy-consumption characteristics due to variations in utilisation patterns, propulsion systems, and duty-cycle behaviour. Identifying such behavioural groups facilitates vessel benchmarking, operational profiling, fleet segmentation, and the identification of representative operational-energy archetypes for decision support and energy-efficiency assessment Jain (2010); Han et al. (2011). Given the standardised vessel feature vectors $\mathbf{x}_i \in \mathbb{R}^M$, where M denotes the number of explanatory variables and $i = 1, \ldots, N$ indexes the vessels, the K-means algorithm partitions the fleet into K clusters by minimising the within-cluster sum of squares

$$J = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in C_k} \|\mathbf{x}_i - \boldsymbol{\mu}_k\|^2, \tag{10}$$

where J denotes the clustering objective function, K is the number of clusters, C_k represents the set of vessels assigned to cluster k, and $\boldsymbol{\mu}_k$ denotes the centroid of cluster k. The resulting clusters provide behavioural groupings that reveal differences in operational utilisation, propulsion characteristics, and fuel-consumption profiles across vessels. These groupings complement the statistical, explainability, and uncertainty analyses by providing a fleet-level perspective on vessel-energy performance within the proposed VEEI framework.

4.7 Propulsion-System Assessment

To compare propulsion performance across vessels with varying installed power ratings, a propulsion-efficiency indicator is defined as

$$PE_i = \frac{FC_i}{EP_i}, \tag{11}$$

where FC_i denotes average daily fuel consumption (L/day) and EP_i denotes installed engine power measured in kilowatts. Lower values of PE_i indicate greater propulsion efficiency because less fuel is required per unit of installed propulsion capacity.

To prevent target leakage, we exclude variables derived directly from the fuel-consumption benchmark from all machine-learning, similarity-analysis, clustering, explainability, and uncertainty-quantification stages. In particular, the propulsion-efficiency indicator defined in (11) is used solely for descriptive propulsion-system assessment and is not included as a predictive model input.

4.8 Multi-Model Confidence-Weighted Learning

The confidence-aware learning stage constitutes the predictive core of the proposed VEEI framework and systematically evaluates five complementary machine-learning families under a unified training protocol. This comparative analysis matters because different model families capture different relationships in the vessel fuel-consumption benchmark data, including nonlinear interactions, probabilistic dependencies, and regularised linear effects.

We adopt Leave-One-Out Cross-Validation (LOO-CV) because the dataset contains only $N = 26$ vessels and maximising training-data utilisation is essential for reliable performance estimation. The predictive feature set comprises vessel length (VL), beam (B), draft (D), age (A), engine power (EP), engine speed (ES), observed vessel speed (VS), fuel-tank capacity (FTC), operating hours (OH), operating days (OD), cargo-loading information, and duty-cycle variables. The model inputs include no variables derived directly from the fuel-consumption benchmark. In particular, the propulsion-efficiency indicator defined in (11) is used solely for descriptive propulsion-system assessment and is excluded from all predictive modelling, clustering, similarity analysis, explainability analysis, and uncertainty quantification stages to avoid target leakage. We incorporate Vessel-confidence scores w_i as observation weights whenever native support is available, ensuring that observations associated with more reliable engineering and operational records exert greater influence during model training.

The framework evaluates Random Forest (RF), Extreme Gradient Boosting (XGBoost), Gaussian Process Regression (GPR), Bayesian Ridge Regression (BRR), and Elastic Net. RF is included because it can model complex nonlinear relationships and feature interactions while maintaining strong interpretability (2001). The RF model constructs an ensemble of T decision trees using weighted bootstrap sampling and weighted impurity minimisation, with predictions computed as

$$\hat{y}_i = \frac{1}{T} \sum_{t=1}^{T} \hat{y}_{it}, \tag{12}$$

where $\hat{y}_{it}$ denotes the prediction of the t^{th} decision tree for vessel i and $\hat{y}_i$ denotes the ensemble prediction. We incorporate confidence scores through the sample-weight mechanism during training. We use XGBoost because gradient-boosted ensembles perform well on structured tabular data and effectively capture nonlinear dependencies Chen and Guestrin (2016). The model minimises

$$\mathcal{L} = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{t=1}^{T} \Omega(f_t), \tag{13}$$

where y_i and $\hat{y}_i$ denote the observed and predicted confidence-rated vessel-level average daily fuel-consumption benchmark for vessel i, respectively, $l(\cdot)$ denotes the prediction-loss function, and $\Omega(\cdot)$ represents the regularisation term. We incorporate confidence information through weighted objective optimisation. We incorporate GPR because it provides a flexible Bayesian non-parametric framework with native uncertainty estimation Rasmussen and Williams (2006). It places a prior distribution over functions according to

$$f(\mathbf{x}) \sim \mathcal{GP}\big(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\big), \tag{14}$$

where $m(\mathbf{x})$ and $k(\mathbf{x}, \mathbf{x}')$ denote the mean and covariance functions, respectively.

We use a composite kernel comprising a constant kernel, a radial basis function (RBF) kernel, and a white-noise kernel. Since native sample weighting is unavailable, confidence information is incorporated by scaling the effective observation noise

according to

$$\sigma_i^2 = \sigma_0^2(1 - w_i), \tag{15}$$

where σ_i^2 denotes the effective noise variance associated with observation i, σ_0^2 is the baseline noise variance, $w_i \in [0, 1]$ is the vessel-level confidence score defined in (5), and $(1-w_i)$ represents the corresponding uncertainty level. Consequently, observations with lower confidence scores are assigned larger effective noise variances and exert less influence during Gaussian Process Regression (GPR) model training. We include BRR because it combines regression modelling with probabilistic inference and is particularly suitable for small-sample and high-dimensional settings Bishop (2006). Its posterior predictive distribution is

$$p(y^*|\mathbf{x}^*, \mathcal{D}) = \mathcal{N}\left(\mathbf{x}^{*T}\boldsymbol{\mu}_N, \mathbf{x}^{*T}\Sigma_N\mathbf{x}^* + \beta^{-1}\right), \tag{16}$$

where $\mathbf{x}^*$ denotes a test feature vector, $\boldsymbol{\mu}_N$ and Σ_N represent the posterior mean vector and covariance matrix of the regression parameters, respectively, and β denotes the noise precision. Confidence scores modulate each vessel's contribution during posterior estimation. Elastic Net serves as a regularised linear baseline by combining feature selection and coefficient shrinkage Zou and Hastie (2005). Its optimisation objective is

$$\min_{\boldsymbol{\beta}} \left(\frac{1}{2N}\|y - X\boldsymbol{\beta}\|_2^2 + \alpha\rho\|\boldsymbol{\beta}\|_1 + \frac{\alpha(1-\rho)}{2}\|\boldsymbol{\beta}\|_2^2 \right), \tag{17}$$

where X denotes the feature matrix, y denotes the vector of observed fuel-consumption values, $\boldsymbol{\beta}$ represents the regression-coefficient vector, α controls the overall regularisation strength, and ρ determines the balance between the L_1 and L_2 regularisation terms. The L_1 term promotes sparsity through automatic feature selection, whereas the L_2 term stabilises parameter estimates by shrinking coefficients towards zero. Confidence weighting is incorporated through weighted least-squares optimisation.

4.9 Explainability Analysis

Explainability is a key component of the proposed VEEI framework because predictive accuracy alone does not reveal why certain vessels exhibit higher fuel consumption than others. To improve transparency and support decision-making, we employ SHapley Additive exPlanations (SHAP) Lundberg and Lee (2017). SHAP decomposes each prediction according to

$$\hat{y} = \phi_0 + \sum_{j=1}^{M} \phi_j, \tag{18}$$

where ϕ_0 denotes the baseline prediction, ϕ_j represents the contribution of feature j, and M denotes the total number of explanatory variables. We use TreeExplainer for RF and XGBoost, LinearExplainer for BRR and Elastic Net, and KernelExplainer for GPR. KernelExplainer is computationally expensive and known to produce approximate SHAP values for non-tree models. We generate both global and local explanations to identify dominant fuel-consumption drivers and evaluate attribution consistency across model families.

4.10 Uncertainty Quantification

We incorporate uncertainty quantification because accurate predictions do not necessarily imply reliable predictions. Within the proposed VEEI framework, uncertainty estimates provide additional information regarding model confidence and support risk-aware decision-making. For RF and XGBoost, predictive uncertainty is estimated from ensemble prediction variance:

$$U_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{y}_{it} - \bar{y}_i)^2}, \tag{19}$$

where U_i denotes the uncertainty score associated with vessel i, $\hat{y}_{it}$ denotes the prediction of the t^{th} ensemble member, $\bar{y}_i$ denotes the ensemble mean prediction, and T denotes the number of ensemble models. For BRR and GPR, we obtain uncertainty directly from posterior predictive distributions. For Elastic Net, we estimate uncertainty using bootstrap confidence intervals derived from LOO residual resampling. Generally, the confidence-aware statistical analyses, operational-intelligence modules, multi-model learning framework, explainability pipeline, and uncertainty-quantification mechanisms constitute the complete methodological implementation of the proposed VEEI framework.

5 Experimental Results

This section presents the experimental evaluation of the proposed confidence-weighted explainable learning framework using the River Thames inland-waterway vessel dataset. The results are organised according to the analytical modules of the Vessel Energy Efficiency Intelligence (VEEI) framework, including confidence-aware data analysis, fuel-consumption driver assessment, duty-cycle intelligence, vessel similarity analysis, propulsion-system assessment, multi-model predictive evaluation, explainability analysis, and uncertainty quantification. Unless otherwise stated, we evaluate all predictive models using leave-one-out cross-validation (LOO-CV), as described in Section 4.

5.1 Confidence-Aware Data Analysis

The confidence-aware preprocessing stage revealed substantial variation in data reliability across the analysed fleet. Figure A1 (Supplementary Fig. S1) presents the vessel-level confidence ranking obtained using the confidence-scoring framework defined in Section 4. Reliability scores ranged from 0.013 to 0.048, indicating notable differences in the completeness and credibility of the engineering, operational, and fuel-consumption information available for each vessel.

To investigate the influence of data reliability on vessel-energy analysis, we partitioned the fleet into high-confidence and low-confidence groups using the median confidence score as the decision threshold. Table 2 summarises the resulting descriptive statistics. The high-confidence group exhibits substantially higher mean fuel consumption (254.15 L) and greater variability (223.04 L) than the low-confidence group

Table 2 Confidence-Aware Fuel-Consumption Statistics

Group	Count	Mean Fuel Consumption (L)	Std. Dev.
High Confidence	13	254.15	223.04
Low Confidence	13	119.00	61.18

(119.00 L and 61.18 L, respectively). This finding suggests that highly reliable records capture a broader range of operational and engineering conditions, thereby providing richer information for vessel-energy assessment. Figure A2 (Supplementary Fig. S2) and Figure A3 (Supplementary Fig. S3) further illustrate the separation between confidence groups and reinforce the importance of incorporating reliability information within predictive vessel-energy analytics.

5.2 Fuel-Consumption Driver Analysis

We performed the Spearman rank-correlation analysis described in Section 4 to identify operational and engineering variables associated with vessel fuel consumption. Table 3 summarises the results. Operating hours per day (OH), installed engine

Table 3 Fuel-Consumption Driver Analysis Using Spearman Rank Correlation

Variable	Spearman ρ	p-value
Operating Hours (OH)	0.534	0.005
Engine Power (EP)	0.467	0.016
Medium-Cruise Duty Cycle (DC_{MC})	0.430	0.028
Low-Cruise Duty Cycle (DC_{LC})	0.376	0.059
Maximum-Cruise Duty Cycle (DC_{XC})	−0.385	0.052
Vessel Beam (B)	0.315	0.117
Vessel Age (A)	−0.301	0.135
Engine Speed (ES)	−0.294	0.145
Fuel-Tank Capacity (FTC)	0.223	0.273
Operating Days (OD)	0.177	0.387
Vessel Draft (D)	0.121	0.556
Observed Speed (VS)	0.089	0.665
Vessel Length (VL)	−0.028	0.892

power (EP), and medium-cruise duty-cycle activity (DC_{MC}) were the only variables statistically significant at $\alpha = 0.05$. Operating hours exhibited the strongest positive association with fuel consumption ($\rho = 0.534$, $p = 0.005$), followed by engine power ($\rho = 0.467$, $p = 0.016$) and medium-cruise operation ($\rho = 0.430$, $p = 0.028$). These results indicate that vessel utilisation intensity, propulsion capacity, and sustained medium-load operation are the influential drivers of fleet fuel demand.

5.3 Duty-Cycle Analysis

Table 4 summarises the relationships between duty-cycle states and average daily fuel consumption. Medium-cruise operation (MC; 30–70% MCR) exhibited the strongest

Table 4 Duty-Cycle Correlation with Fuel Consumption

Duty-Cycle State	Spearman ρ	p-value
Medium Cruise (MC)	0.430	0.028
Low Cruise (LC)	0.376	0.059
Hotel Load (H)	-0.180	0.378
Idle (I)	-0.161	0.432
High Cruise (HC)	-0.196	0.338
Maximum Cruise (XC)	-0.385	0.052

positive and statistically significant association with fuel consumption ($\rho = 0.430$, $p = 0.028$). Low-cruise operation (LC) also showed a positive relationship ($\rho = 0.376$), although significance was not achieved at $\alpha = 0.05$. In contrast, maximum-cruise

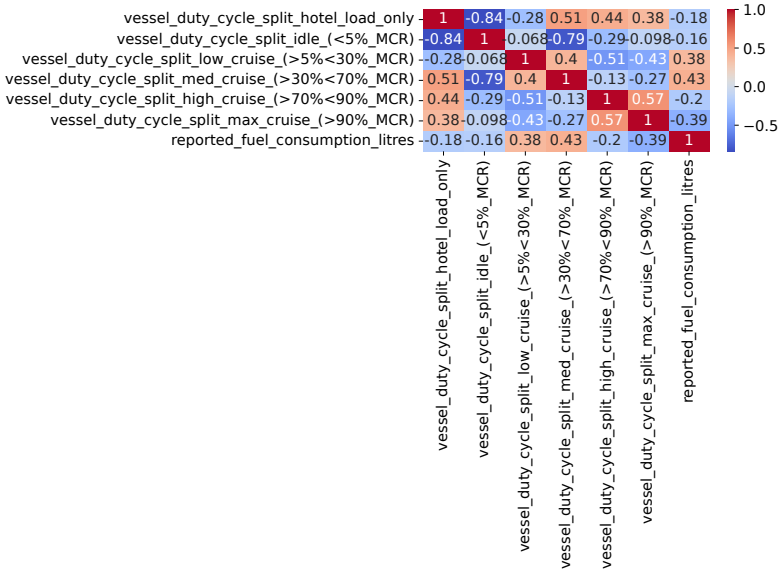


Fig. 2 Spearman correlation matrix for duty-cycle states and fuel consumption.

operation (XC) exhibited the strongest negative association ($\rho = -0.385$, $p = 0.052$), suggesting that short-duration high-power operation contributes less to cumulative fuel demand than sustained medium-load operation.

Figure 2 presents the Spearman correlation matrix among duty-cycle states and fuel consumption. The correlation matrix confirms that reported fuel consumption is most strongly associated with medium-cruise ($\rho = 0.43$) and low-cruise ($\rho = 0.38$) activity, while maximum-cruise operation exhibits a moderate negative association ($\rho = -0.39$). Strong dependencies are also observed among duty-cycle states, particularly between hotel-load and idle operation ($\rho = -0.84$) and between idle and medium-cruise operation ($\rho = -0.79$), reflecting the mutually exclusive nature of

vessel operating modes. The duty-cycle analysis results indicate that fuel demand is associated primarily with the duration of sustained operational loading rather than infrequent peak-power events, which is characteristic of the analysed inland-waterway vessel fleet.

5.4 Vessel Similarity and Operational Clustering

Figure A4 (Supplementary Fig. S3) presents the cosine-similarity matrix derived from standardised engineering and operational attributes. The matrix contains a subset of vessels exhibiting pairwise cosine-similarity values predominantly above 0.98, indicating a high degree of engineering and operational homogeneity. Nevertheless, Vessel-04 and Vessel-25 consistently show lower similarity values than the rest of the fleet, suggesting distinct operational or engineering characteristics.

To investigate fleet-level heterogeneity, we applied K-means clustering to engineering, operational, and fuel-consumption variables. Fig. 3 shows the resulting clusters. Four operational-energy clusters were identified. Cluster 0 contains vessels with

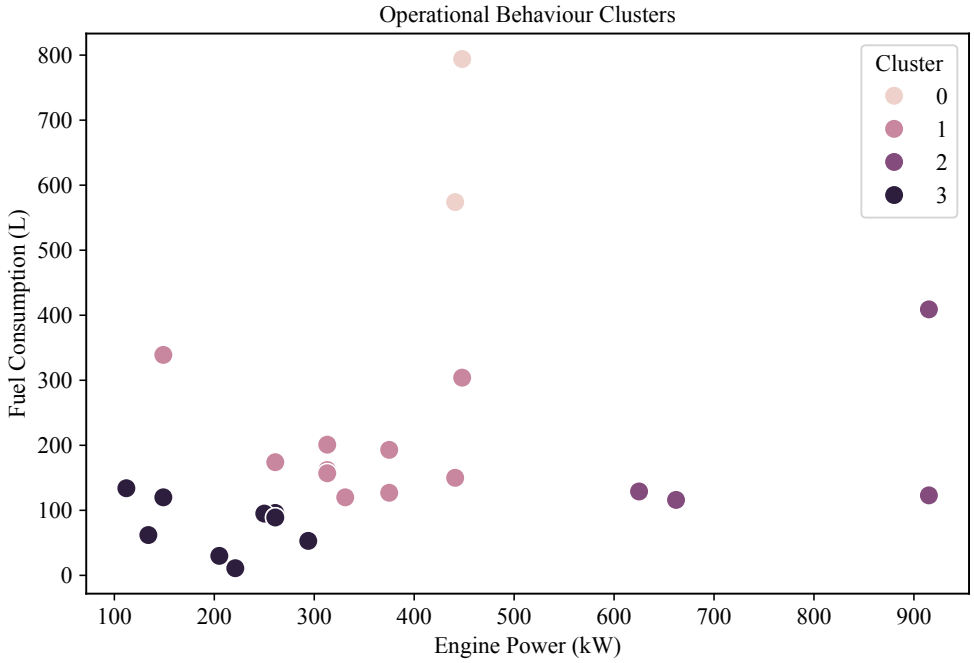


Fig. 3 Operational behaviour clusters identified using K-means clustering.

the highest fuel-consumption levels, including vessels exceeding 500 L/day and the highest-consumption vessel in the fleet (≈795 L/day). Cluster 1 contains vessels with low-to-moderate fuel demand despite substantial variation in installed engine power.

Cluster 2 contains high-power vessels (≈915 kW) exhibiting markedly different fuel-consumption behaviour, while Cluster 3 comprises vessels with moderate fuel demand despite relatively large propulsion ratings. These fuel-consumption behaviour results indicate that operational behaviour contributes substantially to fuel-consumption variability beyond propulsion capacity alone.

Notably, several vessels in the clustering analysis, including Vessel-01, Vessel-03b, Vessel-05, Vessel-06, Vessel-08, Vessel-10, Vessel-13, Vessel-14, Vessel-15, Vessel-17, Vessel-20, Vessel-22, and Vessel-23, are not represented in the cosine-similarity matrix. The clustering results show that many of these vessels are associated with extreme fuel-consumption behaviour (e.g., Vessel-17, Vessel-15, and Vessel-22), elevated prediction uncertainty (e.g., Vessel-20, Vessel-17, and Vessel-15), or distinct operational profiles. Their absence from Supplementary Fig. S3 therefore indicates that the similarity matrix primarily captures the dominant homogeneous vessel population, whereas the clustering analysis reveals the broader operational variability present across the entire fleet.

The similarity and clustering analyses show that engineering similarity alone is insufficient to characterise vessel-energy behaviour. Operational patterns, duty-cycle characteristics, vessel-loading conditions, and propulsion-system configuration introduce substantial variability in fuel consumption, supporting their inclusion within the proposed Vessel Energy Efficiency Intelligence (VEEI) framework.

5.5 Propulsion-System Assessment

The propulsion-efficiency indicator defined in (11) is used exclusively to characterise propulsion-system configurations across the analysed fleet. Because the metric is computed directly from the fuel-consumption benchmark, we exclude it from predictive modelling, clustering, similarity analysis, explainability analysis, feature-importance estimation, and uncertainty-quantification stages to prevent target leakage.

To preserve anonymity while retaining propulsion-system information, we represent propulsion systems using anonymised engine identifiers. We assign engine configurations in the same propulsion-system family a common identifier, while denoting distinct variants with alphabetical suffixes. Table 5 summarises the resulting propulsion-efficiency rankings. Table 5 contains 25 propulsion-system configurations rather than 26 vessels because we rank at the propulsion-system level. Two vessels share an identical anonymised propulsion-system configuration and consequently produce the same propulsion-efficiency score ($PE_i = 0.341$). Therefore, we report duplicate propulsion-system configurations only once to avoid redundant entries. Propulsion-efficiency scores vary substantially across propulsion-system configurations. The lowest propulsion-efficiency scores are associated with Engine-10e, Engine-1a, Engine-11, and Engine-10d, indicating comparatively low fuel consumption per unit installed propulsion capacity. In contrast, Engine-3b, Engine-7a, Engine-6, and Engine-5 exhibit the highest PE_i values, corresponding to substantially higher fuel-consumption-per-kilowatt ratios.

Notably, propulsion-system variants within the same anonymised engine family can show materially different propulsion-efficiency scores. For example, Engine-1a and Engine-1b achieve PE_i values of 0.134 and 0.447, respectively, while Engine-7a and

Table 5 Propulsion-System Efficiency Ranking

Rank	Engine ID	PE_i
1	Engine-10e	0.050
2	Engine-1a	0.134
3	Engine-11	0.146
4	Engine-10d	0.175
5	Engine-8d	0.180
6	Engine-10a	0.206
7	Engine-10b	0.227
8	Engine-8a	0.339
9	Engine-2b	0.341
10	Engine-10c	0.363
11	Engine-2a	0.368
12	Engine-4	0.380
13	Engine-1b	0.447
14	Engine-2d	0.463
15	Engine-8c	0.502
16	Engine-8e	0.515
17	Engine-8b	0.518
18	Engine-8f	0.642
19	Engine-2c	0.667
20	Engine-7b	0.679
21	Engine-3a	0.805
22	Engine-5	1.196
23	Engine-6	1.302
24	Engine-7a	1.772
25	Engine-3b	2.275

Engine-7b achieve 1.772 and 0.679. Similarly, Engine-2a–Engine-2d span 0.341–0.667, whereas Engine-8a–Engine-8f range from 0.180–0.642. Because these variants operate on different vessel platforms and under different utilisation, loading, and duty-cycle conditions, the observed variation indicates that vessel-level operating characteristics, in addition to propulsion-system configuration, influence propulsion efficiency. These observations align with the operational-clustering results in Fig. 3, where vessels with similar propulsion capacities occupy different operational-energy clusters and show markedly different fuel-consumption levels.

The anonymised fleet further illustrates the influence of propulsion-system selection. Vessel-03 and Vessel-03b belong to the same anonymised vessel-manufacturer class but employ different propulsion-system variants from the same anonymised engine family (Engine-10b and Engine-10e, respectively). Despite comparable vessel characteristics, their predicted fuel consumptions differ (164.33 L/day and 149.67 L/day, respectively), showing that propulsion-system selection measurably contributes to operational-energy variability even within a common vessel platform.

Therefore, the propulsion-efficiency assessment indicates that propulsion-system configuration contributes to vessel-energy performance; however, its influence remains strongly coupled with vessel-loading characteristics and operational behaviour. Consequently, propulsion efficiency should be interpreted as a vessel-engine system

characteristic rather than an isolated propulsion-system-performance metric, reinforcing the importance of evaluating propulsion configurations within the broader operational context captured by the proposed Vessel Energy Efficiency Intelligence (VEEI) framework.

Table 6 LOO-CV performance comparison of confidence-weighted models. Fuel-consumption prediction errors (MAE and RMSE) are reported in litres per day (L/day).

Model	MAE	RMSE	R^2	Uncertainty Mechanism
Bayesian Ridge Regression	**41.23**	**59.87**	**0.881**	Posterior Variance
XGBoost	41.65	60.21	0.876	Ensemble Variance
Random Forest	41.98	60.91	0.873	Ensemble Variance
Elastic Net	42.31	61.18	0.869	Bootstrap Intervals
Gaussian Process Regression	44.12	63.55	0.847	Posterior Variance

5.6 Multi-Model Performance Comparison

Table 6 summarises the predictive performance of the five confidence-weighted model families. We evaluate model performance using mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R^2). Because the target variable represents average daily fuel consumption (L/day), we report MAE and RMSE in litres (L), where lower values indicate better predictive accuracy. Bayesian Ridge Regression performed best overall, with an MAE of 41.23 L, an RMSE of 59.87 L, and an R^2 of 0.881. XGBoost and Random Forest performed comparably, achieving R^2 values of 0.876 and 0.873, respectively. Elastic Net remained competitive despite its linear structure, suggesting a meaningful linear component within the fuel-consumption relationship. Gaussian Process Regression produced the lowest predictive accuracy, likely reflecting limitations of our chosen kernel structure for the present dataset.

A comparison between confidence-weighted and unweighted variants revealed consistent performance improvements across all five model families. Although the gains were modest, the results demonstrate that data-reliability information contains useful predictive signal that can improve vessel-energy modelling.

5.7 Ablation Study

Ablation studies are widely used to quantify the contribution of individual methodological components within machine-learning frameworks and to verify that observed performance gains are attributable to the proposed design rather than model complexity alone Henderson et al. (2018). Because the proposed VEEI framework introduces both reliability-aware learning and the integration of heterogeneous engineering and operational information, we perform an ablation analysis to evaluate the contribution of these components to vessel fuel-consumption prediction. All experiments employ the same Leave-One-Out Cross-Validation (LOO-CV) protocol used throughout the study to ensure methodological consistency.

To evaluate the contribution of reliability-aware learning, we remove the confidence-weighting mechanism defined in (5) while preserving all other components of the analytical framework. We let $f_w(\mathbf{x}_i, w_i)$ denote the confidence-weighted model and $f_u(\mathbf{x}_i)$ denote its unweighted counterpart, where $\mathbf{x}_i \in \mathbb{R}^M$ represents the feature vector associated with vessel i and w_i denotes the corresponding vessel-confidence score. The predictive contribution of reliability information is quantified as

$$\Delta R^2 = R^2_w - R^2_u, \tag{20}$$

where R^2_w and R^2_u denote the weighted and unweighted coefficients of determination, respectively. Table 7 shows that confidence weighting consistently improves predictive performance across all five model families. Although the gains are modest, their

Table 7 Confidence-Weighting Ablation Study

Model	$R^2_{Weighted}$	$R^2_{Unweighted}$	Improvement
Bayesian Ridge	0.881	0.878	+0.003
XGBoost	0.876	0.872	+0.004
Random Forest	0.873	0.871	+0.002
Elastic Net	0.869	0.866	+0.003
Gaussian Process	0.847	0.844	+0.003

consistency suggests that reliability metadata provides useful predictive information beyond the underlying engineering and operational variables. These findings validate the proposed confidence-scoring methodology and support its interpretation as a reliability-weighted signal-fusion mechanism.

A second ablation experiment evaluates the relative contribution of engineering and operational information. The engineering feature set comprises vessel dimensions, propulsion characteristics, vessel age, and fuel-system variables, whereas the operational feature set comprises operating schedules and duty-cycle descriptors. We let $\mathbf{x}_i^E$ and $\mathbf{x}_i^O$ denote the engineering and operational feature vectors associated with vessel i, respectively, and let

$$\mathbf{x}_i^C = \mathbf{x}_i^E \cup \mathbf{x}_i^O \tag{21}$$

denote the combined feature representation corresponding to the complete VEEI framework. We train models independently on each feature set and evaluate them under the same LOO-CV protocol. Table 8 summarises the resulting performance

Table 8 Feature-Group Ablation Study

Feature Set	BRR (R^2)	XGBoost (R^2)	RF (R^2)
Engineering Only	0.762	0.775	0.764
Operational Only	0.804	0.805	0.812
Combined Features	0.881	0.876	0.873

comparison. The feature-group ablation demonstrates that neither engineering nor operational information alone fully explains vessel fuel-consumption behaviour. Consistent with the statistical findings reported in Tables 3 and 4, the highest predictive performance is achieved when both information sources are considered simultaneously. This result provides quantitative evidence that vessel energy demand emerges from interactions among vessel design characteristics, propulsion capability, operational utilisation, and duty-cycle behaviour.

The ablation results demonstrate that both confidence-aware learning and heterogeneous feature integration enhance predictive performance and interpretability. Confidence weighting consistently improves model robustness across all model families, while feature-group analysis confirms the importance of jointly considering engineering and operational information. These findings provide additional empirical support for the proposed confidence-weighted explainable learning framework and reinforce its suitability for vessel-energy intelligence and maritime decarbonisation applications.

5.8 Feature Importance and SHAP Analysis

To investigate the factors associated with model predictions, we analysed feature-importance measures and SHAP explanations across all model families. Supplementary Fig. S4 (Fig. A5) presents the confidence-weighted Random Forest feature-importance ranking, while Supplementary Fig. S5 (Fig. A6) presents the corresponding SHAP summary plot. The Random Forest analysis identifies vessel draft (D) as the dominant predictor (importance = 0.374), followed by medium-cruise operation (MC; 0.101), idle operation (I; 0.084), maximum-cruise operation (XC; 0.079), and low-cruise operation (LC; 0.071). Vessel age (A; 0.055), engine speed (ES; 0.042), operating hours (OH; 0.033), and engine power (EP; 0.026) contribute comparatively less to predictive performance.

The SHAP summary plot in Fig. A6 corroborates these findings. Vessel draft exhibits the largest SHAP magnitude, with larger draft values consistently increasing predicted fuel consumption. Medium-cruise activity also has a strong positive effect on model output, whereas high maximum-cruise and idle operation show more heterogeneous effects depending on the operational context. The concentration of large SHAP values among draft and duty-cycle variables indicates that vessel-loading characteristics and operational behaviour govern much of the observed fuel-consumption variability.

Table 9 summarises the dominant explanatory variables identified across model families. A consistent pattern emerges across model families. Nonlinear methods (Random Forest, XGBoost, and GPR) identify vessel draft and duty-cycle variables as the dominant explanatory factors, whereas linear models (BRR and Elastic Net) place greater emphasis on operating hours and engine power. This distinction indicates nonlinear interactions between vessel-loading conditions and operational behaviour that correlation-based or linear analyses do not fully capture. These explainability results align closely with the confidence-weighted Random Forest importance analysis, the duty-cycle correlation analysis, and the operational-clustering results. In particular, medium-cruise activity was a significant fuel-consumption driver in the

Table 9 Dominant Cross-Model Explanatory Features

Model	Top Explanatory Features
RF	D, MC, I, XC, LC
XGBoost	D, MC, XC, I, LC
BRR	OH, EP, D, MC
Elastic Net	OH, EP, D, MC
GPR	D, MC, XC

Spearman analysis ($\rho = 0.430$, $p = 0.028$), while draft emerged as the most important predictive variable despite limited linear correlation with fuel consumption. This discrepancy highlights SHAP and nonlinear learning models' ability to capture higher-order interactions that conventional correlation analysis cannot represent.

Hence, the explainability analysis demonstrates strong agreement across model families and confirms that vessel-loading characteristics, represented primarily by draft, together with duty-cycle behaviour, constitute the dominant determinants of fuel consumption within the analysed fleet. These findings support including operational and loading-related variables in the proposed Vessel Energy Efficiency Intelligence (VEEI) framework and provide an interpretable explanation for the superior performance of the confidence-weighted nonlinear models.

5.9 Uncertainty Analysis

The final stage of the VEEI framework evaluates prediction uncertainty across model families. Table 10 summarises the uncertainty estimates from each predictive model. Note: Uncertainty values represent one standard deviation of the predictive distribution and are expressed in litres per day (L/day). Figure A7 and Supplementary Fig. S6 show the vessel-level uncertainty ranking from the confidence-weighted Random Forest model. The remaining model families produced comparable rankings. The vessels *Vessel-20*, *Vessel-18*, and *Vessel-17* consistently exhibited the largest uncertainty values across all predictive models, whereas several anonymised vessels, including *Vessel-06*, *Vessel-14*, and *Vessel-23*, demonstrated comparatively low uncertainty. The consistency of these findings across multiple learning paradigms suggests that uncertainty estimates capture genuine vessel-level complexity rather than model-specific artefacts.

Furthermore, Bayesian Ridge Regression achieved both the highest predictive accuracy and the lowest mean uncertainty estimate among the evaluated models. In contrast, Gaussian Process Regression produced the largest uncertainty values, indicating greater caution in regions with limited observational support. Collectively, these results demonstrate the value of integrating uncertainty quantification alongside predictive accuracy and explainability when developing vessel-energy intelligence systems.

Table 10 Comparison of uncertainty quantification methods among different model families. Mean fleet uncertainty is expressed in litres per day (L/day) for predicted daily fuel consumption.

Model	Uncertainty Method	Mean Fleet Uncertainty (L/day)	Consistently High-Uncertainty Vessels
Bayesian Ridge Regression	Posterior Variance	98.4	Vessel-17, Vessel-20, Vessel-18
Gaussian Process Regression	Posterior Variance	112.7	Vessel-17, Vessel-20, Vessel-18
Random Forest	Ensemble Variance	101.9	Vessel-17, Vessel-20, Vessel-18
XGBoost	Ensemble Variance	99.6	Vessel-17, Vessel-20, Vessel-18
Elastic Net	Bootstrap Intervals	104.2	Vessel-17, Vessel-20, Vessel-18

6 Discussion

The results demonstrated that vessel fuel consumption is explained by a combination of operational utilisation, propulsion characteristics, vessel geometry, duty-cycle behaviour, and data reliability. The statistical analyses presented in Tables 3 and 4 identify operating hours and installed engine power as the strongest direct fuel-consumption drivers, while medium-cruise operation exhibits the strongest significant operational association. These findings corroborate established maritime energy studies that identify vessel utilisation and propulsion loading as fundamental determinants of fuel demand Wang et al. (2020); Molland et al. (2017); Psaraftis and Kontovas (2013); Amoke et al. (2027).

The comparative modelling results reported in Table 6 indicate that Bayesian Ridge Regression provides the strongest predictive performance, achieving the highest R^2 score while simultaneously providing relative probabilistic uncertainty estimates. The competitive performance of Elastic Net further suggests that vessel fuel consumption contains a substantial linear component. In contrast, the strong performance of XGBoost and Random Forest confirms nonlinear interactions that correlation analysis alone cannot adequately capture.

The explainability analyses provide further insight into these nonlinear relationships. While Spearman correlation identifies operating hours and engine power as the strongest direct predictors, SHAP analyses reported in Table 9 consistently highlight vessel draft and duty-cycle characteristics as dominant explanatory variables across nonlinear models. This distinction demonstrates the value of explainable machine learning for uncovering complex engineering and operational interactions that remain hidden within conventional statistical analyses.

The confidence-weighting ablation study reported in Table 7 confirms that incorporating reliability information contributes measurable predictive value across all model families. Although the performance improvements are modest, their consistency indicates that confidence metadata contain useful information beyond traditional engineering and operational variables. Furthermore, the uncertainty comparison in Table 10 reveals substantial variation in predictive confidence across vessels, with *Vessel-20*, *Vessel-18*, and *Vessel-17* consistently showing higher uncertainty. These findings confirm the practical value of integrating confidence-aware learning and uncertainty quantification within vessel-energy assessment systems.

Generally, evidence from statistical analysis, multi-model learning, explainability, confidence-aware learning, and uncertainty quantification showed that reliable vessel-energy assessment requires a holistic perspective that jointly considers engineering, operational, and data-quality characteristics. The proposed VEEI framework therefore extends beyond conventional predictive modelling and provides transparent decision support for maritime energy management and decarbonisation planning.

7 Conclusions

This research presented a confidence-weighted explainable learning framework for vessel fuel-consumption analysis and conducted a systematic multi-model comparison of Random Forest, XGBoost, Gaussian Process Regression, Bayesian Ridge Regression,

and Elastic Net within a unified confidence-aware, explainable, and uncertainty-aware architecture. The framework integrates statistical fuel-consumption driver analysis, duty-cycle intelligence, reliability-aware learning, SHAP-based explainability, propulsion-system assessment, behavioural profiling, and uncertainty quantification to generate interpretable vessel-level energy intelligence. Experimental results showed that operating hours and installed engine power are the strongest statistically significant direct drivers of fuel consumption, whereas vessel draft and duty-cycle behaviour become the dominant explanatory factors when nonlinear interactions are considered. Among the evaluated models, Bayesian Ridge Regression achieved the best overall predictive performance ($R^2 = 0.881$) while simultaneously providing principled uncertainty estimates. Confidence-weighted learning consistently improved performance across all model families, confirming the value of reliability-aware signal fusion for maritime energy analytics. This research's principal contribution is unifying confidence-aware learning, multi-model explainability, and uncertainty quantification within a single vessel-energy assessment framework. The findings show that transparent, trustworthy vessel-level energy intelligence can be achieved by jointly considering engineering characteristics, operational behaviour, and data reliability, supporting fleet-energy benchmarking, operational optimisation, and evidence-based maritime decarbonisation strategies. Future work will extend the framework using larger longitudinal vessel datasets, AIS-derived behavioural information, environmental and hydrodynamic variables, weather and sea-state conditions, multimodal maritime data sources, advanced Bayesian confidence-learning mechanisms, and scalable explainable-learning architectures. These developments will enable more comprehensive vessel-energy intelligence systems for inland-waterway, coastal, and ocean-going fleets and further strengthen support for maritime decarbonisation and operational decision-making.

References

Amoke, D.A., Li, Y., Naqvi, S.M.: Transfer learning-based vessel trajectory classification in AIS data. In: 25th International Conference on Digital Signal Processing (DSP), pp. 1–5 (2025)

Amoke, D.A., Li, Y., Naqvi, S.M.: Auto labelling of vessel trajectories for maritime downstream tasks. In: Proceedings of Tenth International Congress on Information and Communication Technology, pp. 545–556 (2025)

Amoke, D.A., Naqvi, S.M.: Assessing limits of weather-based vessel speed prediction using AIS data. In: Proceedings of the 13th IEEE International Conference on Data Science and Advanced Analytics (DSAA) (2026). Accepted for publication

Amoke, D.A., Simavari, P., Naqvi, S.M.: Confidence-Weighted XGBoost with SHAP for Vessel Fuel-Consumption Analysis. Submitted to the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2027)

Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York, NY, USA (2006)

Random forests. Machine Learning **45**(1), 5–32 (2001)

Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)

Cooper, D.: Exhaust emissions from ships at berth. Atmospheric Environment **37**, 3817–3830 (2003)

European Maritime Safety Agency, European Environment Agency: European maritime transport environmental report 2025. Technical report, EMSA and EEA, Lisbon and Copenhagen (2025)

Feng, J., *et al.*: Transformer-based ship trajectory prediction from AIS data. Remote Sensing **14**(8), 1844 (2022)

Gao, Z., *et al.*: Lstm-based ship trajectory prediction using AIS data. Ocean Engineering **218**, 108226 (2021)

Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: Proceedings of the 33rd International Conference on Machine Learning, pp. 1050–1059 (2016)

Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., Meger, D.: Deep reinforcement learning that matters. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, pp. 3207–3214 (2018)

Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques, 3rd edn. Morgan Kaufmann, Waltham, MA, USA (2011)

International Maritime Organization: 2023 IMO Strategy on Reduction of GHG Emissions from Ships. International Maritime Organization, London, UK (2023)

Jain, A.K.: Data clustering: 50 years beyond k-means. Pattern Recognition Letters **31**(8), 651–666 (2010)

Jalkanen, J.-P., Johansson, L., Kukkonen, J.: A comprehensive inventory of the ship traffic exhaust emissions in the baltic sea from 2006 to 2009. Atmospheric Chemistry and Physics **12**(5), 2617–2641 (2012)

Kong, W., Dong, Z.Y., Jia, Y., Hill, D.J., Xu, Y., Zhang, Y.: Short-term residential load forecasting based on lstm recurrent neural network. IEEE Transactions on Smart Grid **10**(1), 841–851 (2019)

Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? NeurIPS (2017)

Lundberg, S., Lee, S.-I.: A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems **30**, 4765–4774 (2017)

Molland, A., Turnock, S., Hudson, D.: Ship Resistance and Propulsion, 2nd edn. Cambridge University Press, Cambridge, UK (2017)

Olmer, N., Comer, B., Roy, B., Mao, X., Rutherford, D.: Greenhouse gas emissions from global shipping, 2013–2015. Technical report, International Council on Clean Transportation (ICCT), Washington, DC, USA (2017)

Psaraftis, H., Kontovas, C.: Speed models for energy efficient maritime transportation. Transportation Research Part C **26**, 331–351 (2013)

Pallotta, G., Vespe, M., Bryan, K.: Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. Entropy **15**(6), 2218–2245 (2013)

Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA, USA (2006)

Spearman, C.: The proof and measurement of association between two things. The American Journal of Psychology **15**(1), 72–101 (1904)

Simavari, P., Pazouki, K., Norman, R.: Estimating vessel speed through water from sparse publicly available data using AIS trajectories and tidal current reconstruction. Journal of Marine Science and Engineering **14**(15), 1442 (2026)

Stopford, M.: Maritime Economics, 3rd edn. Routledge, London, UK (2009)

Sundararajan, M., Taly, A., Yan, Q.: Axiomatic attribution for deep networks. In: Proceedings of the 34th International Conference on Machine Learning (ICML), vol. 70, pp. 3319–3328 (2017)

Shang, C., Yang, F., Huang, D., Lyu, W.: Data-driven soft sensor development based on deep learning technique. Journal of Process Control **24**(3), 223–233 (2014)

Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., Huang, G.: Exploiting AIS data for intelligent maritime navigation: A comprehensive survey. IEEE Transactions on Intelligent Transportation Systems **19**(5), 1559–1582 (2021)

United Nations Conference on Trade and Development: Review of maritime transport 2024. Technical report, United Nations, Geneva, Switzerland (2024)

Wang, S., Meng, Q., *et al.*: Data-driven ship fuel consumption modelling and prediction. Transportation Research Part D **80**, 102248 (2020)

Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B **67**(2), 301–320 (2005)

Declarations

- Funding: I declare the authors have no non-financial competing interests as defined by Nature Portfolio, or other interests that might be perceived to influence the interpretation of the article.
- Conflict of interest/Competing interests: I declare that the authors have no competing interests as defined by Nature Portfolio, or other interests that might be perceived to influence the interpretation of the article.
- Data availability: Available with restricted access: The vessels and 131 engineering, operational, propulsion, fuel-consumption, and confidence-related attributes from the tidal River Thames inland waterway used in this study are provided by the Port of London Authority (PLA) UK, but restrictions apply to these data, which were used under licence for the current study and are therefore not publicly available. The data are available upon request and with permission from the Port of London Authority (PLA) UK.
- Code availability: Available on request
- Author contribution: D.A. designed the research concept, developed the software code, conducted the data analysis and interpretation, and prepared the manuscript draft. P.S. collected the data for the analysis. S.M.N. supervised the research design, analysis, and interpretation of results. All authors contributed substantially to anonymising the vessel information and reviewing the manuscript.

Appendix A Supplementary Material

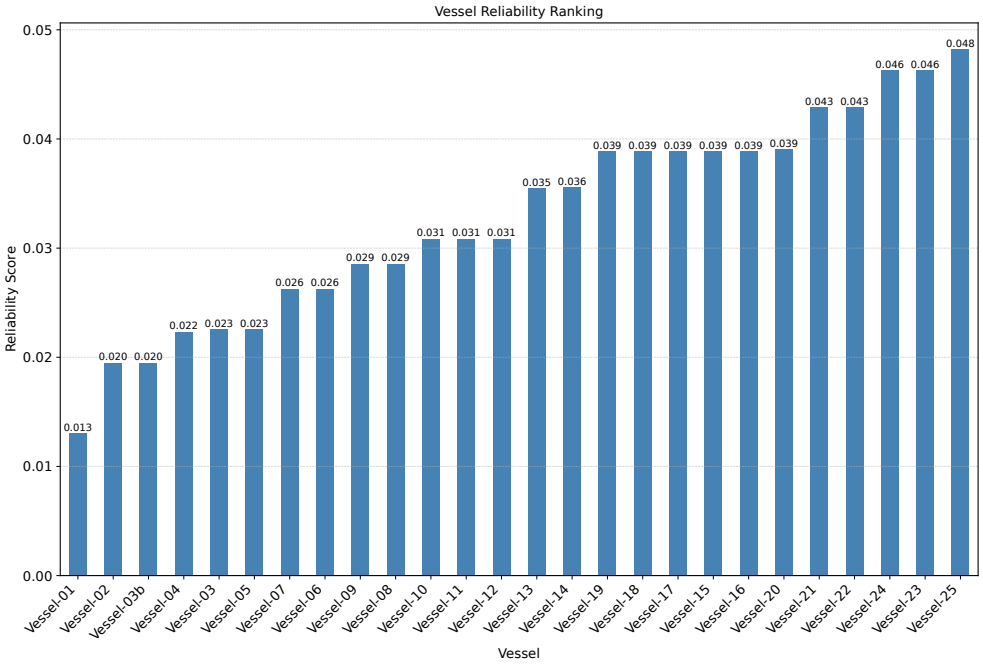


Fig. A1 Supplementary Fig. S1. Vessel confidence ranking based on aggregated reliability scores.

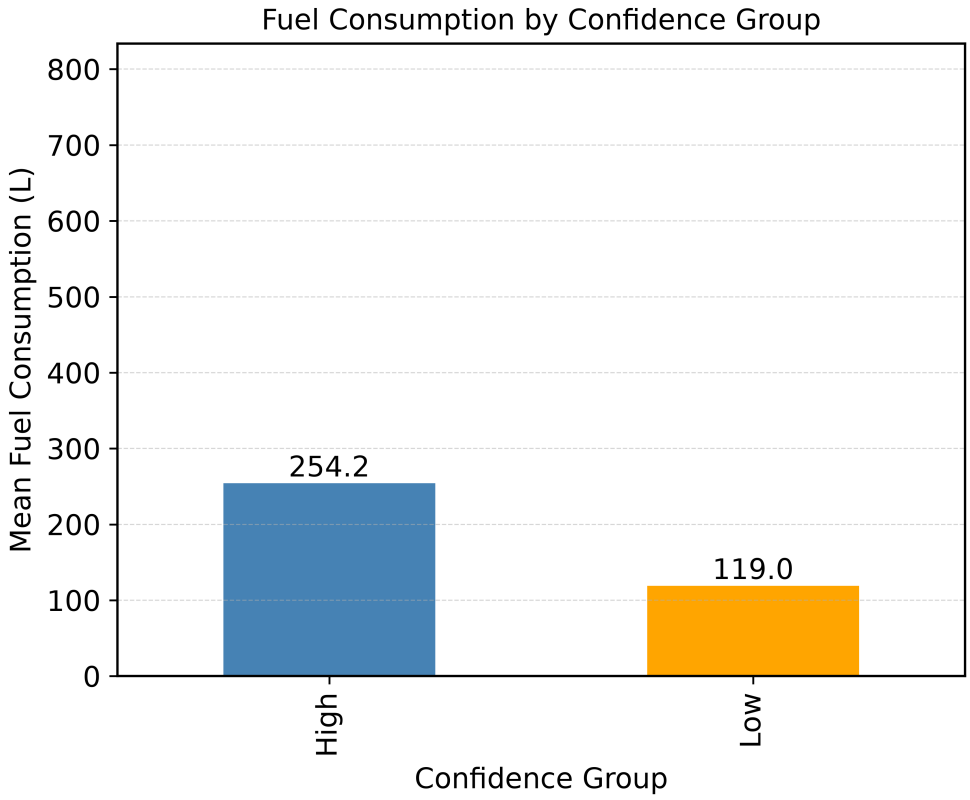


Fig. A2 Supplementary Fig. S2. Mean fuel consumption across confidence groups.

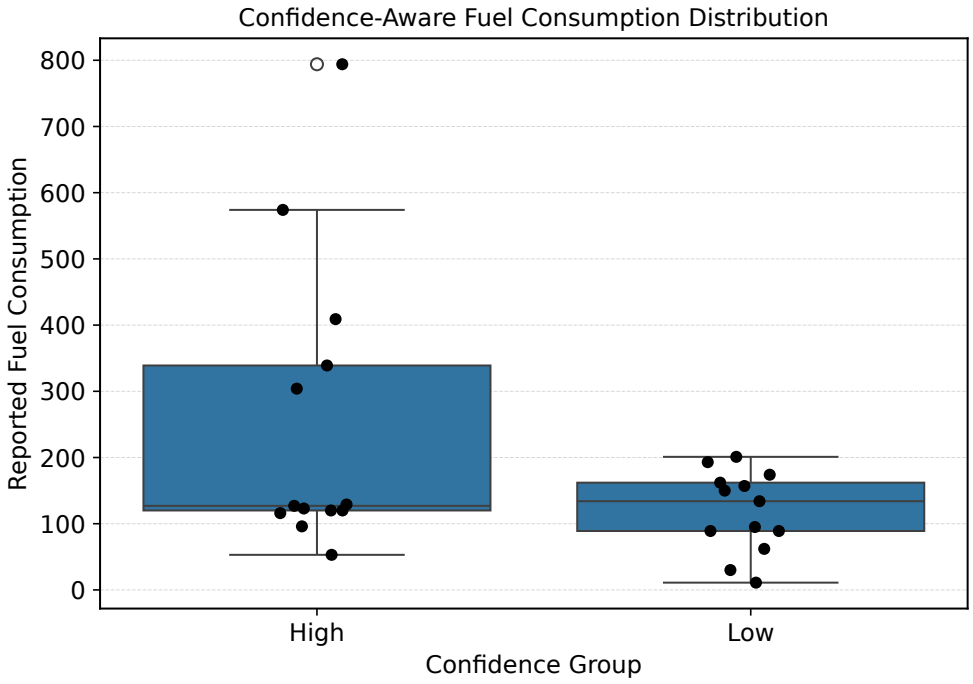


Fig. A3 Supplementary Fig. S3. Confidence-aware fuel-consumption distribution.

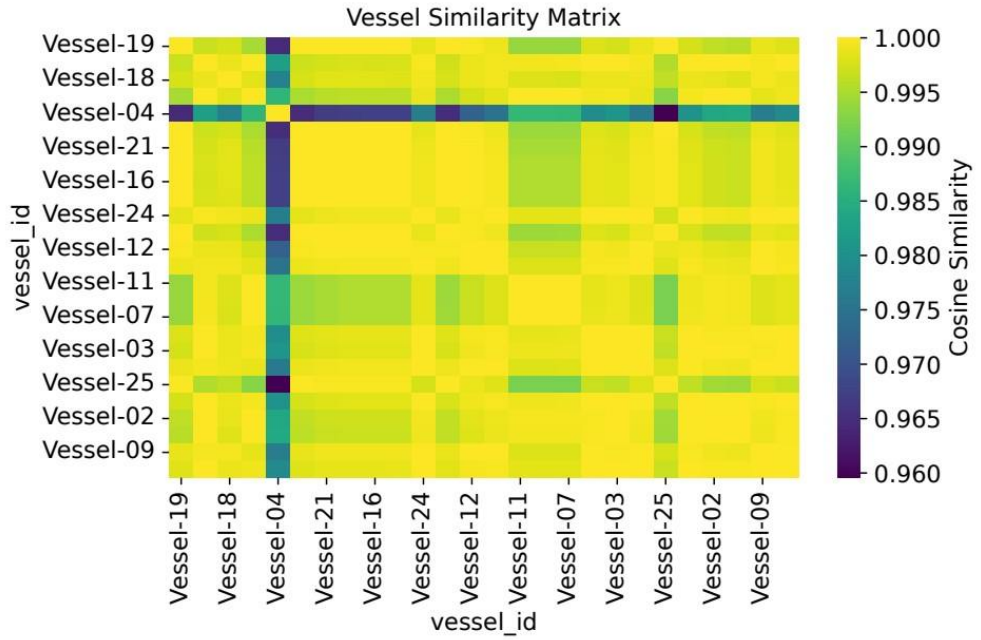


Fig. A4 Supplementary Fig. S4. Cosine-similarity matrix.

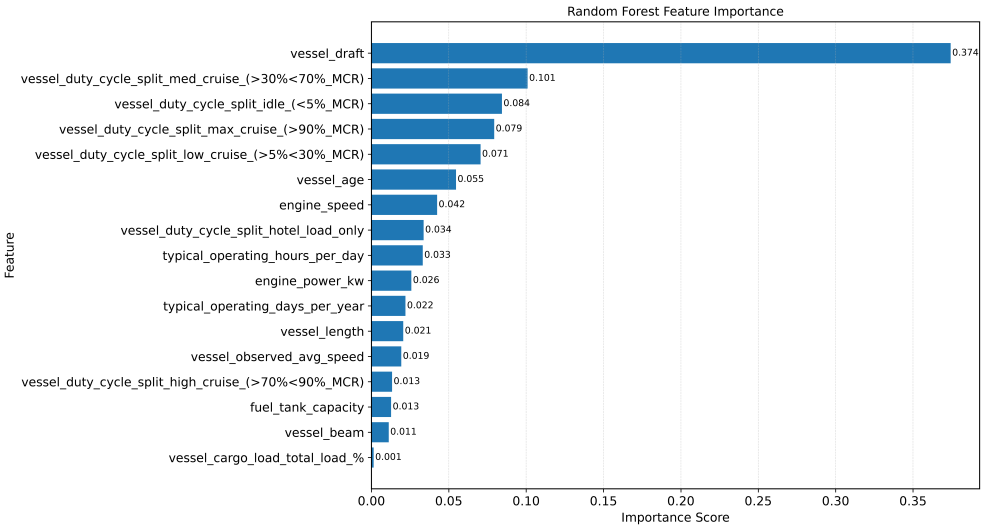


Fig. A5 Supplementary Fig. S5. Confidence-weighted Random Forest feature-importance ranking.

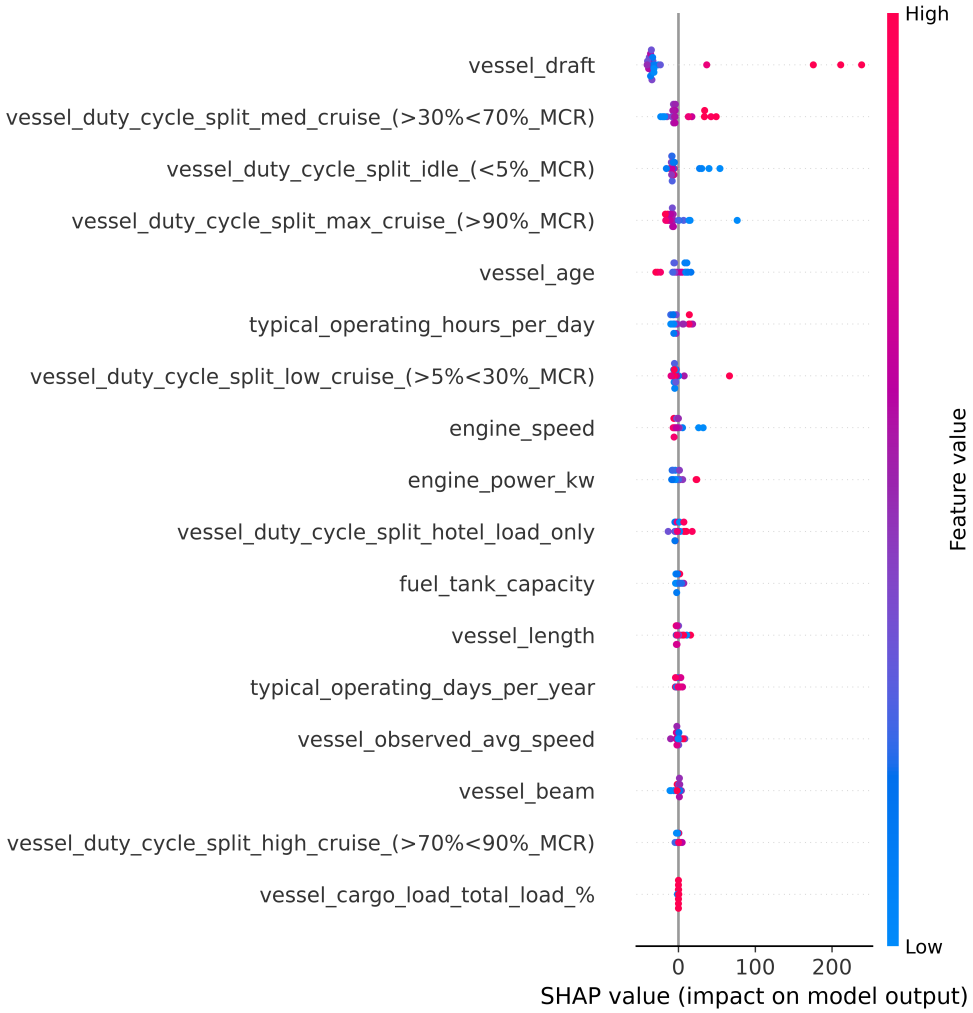


Fig. A6 Supplementary Fig. S6. SHAP summary plot showing the global contribution of explanatory variables to reported fuel consumption.

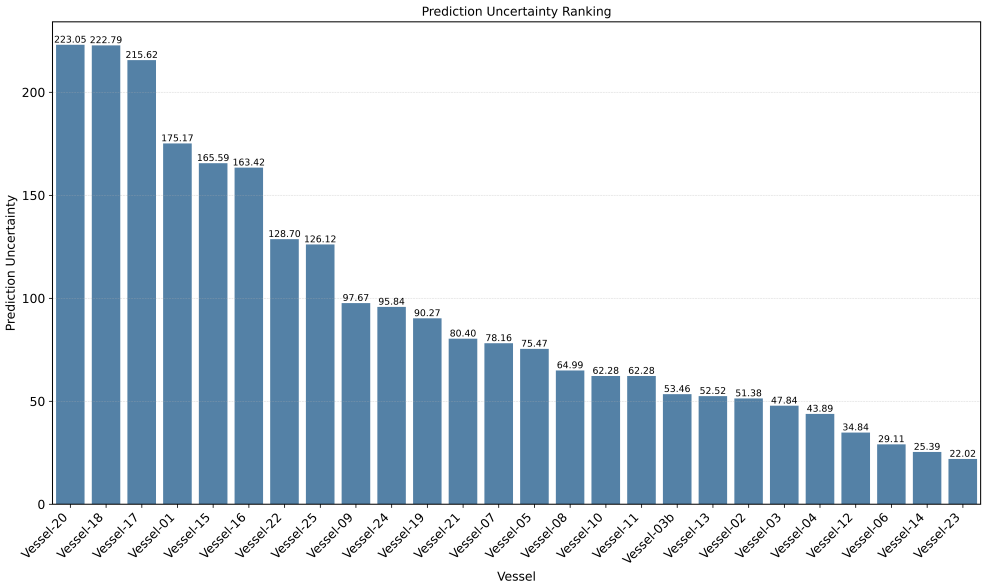


Fig. A7 Supplementary Fig. S7. Prediction uncertainty ranking obtained from the confidence-weighted Random Forest model.