ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb129

Ship Fuel Consumption Prediction and Optimization Under Limited Logbook Data: A Bunker Vessel Case Study

Ye Si Thu Aung[1]; Le Van Diem[2]; Tran Hong Ha[3]

[1]Faculty of Marine Engineering, Vietnam Maritime University, Hai Phong, Vietnam
[2]Faculty of Marine Engineering, Vietnam Maritime University, Hai Phong, Vietnam
[3]Faculty of Marine Engineering, Vietnam Maritime University, Hai Phong, Vietnam

Publication Date: 2026/02/09

Abstract: Ship fuel consumption prediction and optimization is challenging due to limited operational data even often collected for many years for a ship, which is barely enough for training models, especially when limiting collecting only from ship logbook data. In this study, only 76 usable records were able to collected from 2 years of voyages from a bunker ship and it was trained with radient Boosting, Random Forest, and XGBoost, showed weak predictive performance, with R^2 values remaining below 0.64. To overcome this limitation, the training data were expanded using distribution-preserving augmentation, increasing the sample size to 1,000 while keeping the original statistical characteristics. After augmentation, prediction accuracy improved markedly, reaching an MAE of 0.027, an RMSE of 0.050, and an R^2 of 0.995. The improved Random Forest model was then used for fuel optimization. Four different optimization methods, Genetic Algorithm, Real-Coded Genetic Algorithm, NSGA-II, and Particle Swarm Optimization were applied to adjust controllable variables such as vessel speed and trim under fixed voyage conditions. All four methods led to nearly the same optimal operating point and resulted in fuel savings of about 20.6 percent compared with the baseline voyage. This shows that once prediction stability is achieved, optimization results become consistent and reliable even when only logbook-scale data are available.

Keywords: *Bunker Ship; Fuel Consumption Prediction; Limited Data; Machine Learning; Optimization.*

How to Cite: Ye Si Thu Aung; Le Van Diem; Tran Hong Ha (2026) Ship Fuel Consumption Prediction and Optimization Under Limited Logbook Data: A Bunker Vessel Case Study. *International Journal of Innovative Science and Research Technology*, 11(2), 100-107. https://doi.org/10.38124/ijisrt/26feb129

I. INTRODUCTION

Fuel consumption remains the dominant operational cost and emissions driver in commercial shipping. Regulatory instruments such as EEXI and CII convert this operational variable into a compliance constraint, not merely an efficiency target [1], [2]. In practice, however, fuel consumption is rarely observed under controlled conditions. It is produced by coupled effects of speed, loading, trim, and environment, recorded inconsistently across voyages, and documented primarily through logbooks rather than continuous sensors on many vessels [3]. The predictive problem is therefore not one of model expressiveness alone, but of inference under sparse, irregular, and partially structured data.

Conventional physics-based approaches estimate fuel consumption by resolving resistance and propulsion components under assumed operating states. These methods provide interpretable baselines but degrade when confronted with operational variability that is not explicitly parameterized, including fluctuating weather exposure, trim adjustments, and human decision-making [4]. Their applicability is further limited when required inputs are unavailable or coarsely recorded, as is common in logbook-derived datasets. Under such conditions, prediction error reflects missing structure rather than hydrodynamic uncertainty.

Data-driven models address this limitation by learning empirical relationships directly from operational records. Machine learning methods, particularly tree-based ensembles, have shown the capacity to capture nonlinear dependencies among speed, displacement proxies, and environmental factors without explicit physical formulation [5], [6]. Their effectiveness, however, is strongly conditioned on data volume and diversity. Most reported successes rely on large, sensor-rich datasets that are unavailable for a significant portion of the active fleet. When sample size collapses to tens of observations, model variance dominates, and apparent underperformance becomes indistinguishable from data insufficiency [7].

This constraint defines the core problem addressed in this study. The objective is not to introduce a new predictive

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb129

algorithm, but to examine how predictive structure can be recovered when only small-scale logbook data are available. In such regimes, the central methodological risk is not bias from model choice, but instability arising from undersampled operating space. Any practical solution must therefore act on the data regime itself while preserving physical plausibility and statistical structure.

Data augmentation offers one possible intervention, but naïve resampling or noise injection risks distorting operational distributions and violating domain constraints. For fuel consumption, even small distributional shifts can yield physically implausible predictions when extrapolated [8]. Augmentation must therefore be constrained, distribution-aware, and explicitly bounded by operational limits. The effectiveness of such augmentation should be evaluated not by in-sample fit alone, but by behavior on genuinely unseen voyages.

This study investigates fuel consumption prediction for a bunker vessel using two years of operational logbook data under these constraints. Baseline ensemble regressors are first evaluated on the original dataset to characterize performance limits imposed by sample size. A distribution-preserving augmentation procedure is then introduced to expand the training set while maintaining covariance structure and physical bounds. Models retrained on the augmented data are evaluated against both internal test splits and independent holdout voyages. The analysis focuses on whether augmentation recovers stable predictive structure without introducing artefactual accuracy.

Fuel optimization approaches have been applied to speed optimization, trim optimization, and energy-efficient routing, but their applicability is limited when training with real operational data. A common feature of most existing optimization studies is the reliance on large datasets obtained from high-frequency onboard sensors or continuous monitoring systems, with those data, data-driven fuel consumption models have been incorporated into optimization frameworks. Several studies have coupled machine learning–based fuel predictors with optimization techniques to minimize fuel consumption while satisfying voyage constraints such as schedule adherence or safety margins [9]. As a result, robustness and reliability under data-scarce conditions, such as those encountered when only logbook records are available, remain insufficiently addressed. This gap motivates this study to further investigation into optimization frameworks that can operate effectively when predictive models themselves must be learned from limited and irregular data.

The contribution is methodological rather than algorithmic. The results clarify how, under logbook-scale data conditions, constrained augmentation can shift model behavior from variance-dominated fitting toward operationally reliable prediction and optimization. This has direct relevance for fleets where sensor coverage is limited and regulatory pressure demands quantitative fuel estimation and optimization without the infrastructure assumed in most data-driven studies

II. RESEARCH METHODOLOGY

➢ *Problem Formulation*

Let the operational dataset be defined as a finite set $D = \{(x_i, y_i)\}_{i=1}^{N}$; where each record corresponds to a voyage segment extracted from ship logbooks. The feature vector $x_i \in \mathbb{R}^6$; contains operational variables: distance run, vessel speed, cargo load, wave height, wind speed, and trim. The target variable $y_i \in \mathbb{R}$; denotes the corresponding fuel consumption rate expressed in tons per day.

The prediction task is formulated as supervised regression: $\hat{y} = f(x)$; where $f(\cdot)$ is a learned mapping that estimates fuel consumption under a given operational state. No temporal ordering is assumed. Records are treated as conditionally independent samples due to irregular logging intervals and incomplete voyage continuity.

The central constraint is dataset size of 76 rows and 7 columns as shown ing Table 1, the operating space is sparsely sampled, and conventional asymptotic assumptions underlying statistical learning do not apply. Model performance is therefore governed primarily by variance and sensitivity to sampling noise rather than representational capacity.

➢ *Data Validation and Cleaning*

Prior to modeling, the raw dataset was subjected to automated validation using a Python-based inspection pipeline. The process targeted structural and semantic consistency rather than exploratory inference. Column names were standardized, non-numeric tokens removed, and mixed-format fields resolved. Range constraints were enforced for physically bounded variables, including wind direction, wind force, and fuel consumption, with violations logged for inspection.

Duplicate records were removed, missing values imputed using median statistics for numeric variables, and outliers adjusted using an interquartile-range-based winsorization procedure. This approach suppresses extreme leverage points while retaining rank structure within each variable. After cleaning, all variables were numeric, non-null, and bounded within operationally plausible ranges. The resulting dataset preserves empirical variability while eliminating artifacts introduced by logging inconsistencies.

➢ *Dataset Partitioning*

To evaluate generalization under limited data, the cleaned dataset was partitioned into training and testing subsets using random sampling with a fixed seed to ensure reproducibility. Eighty percent of the records were allocated to training, with the remaining twenty percent reserved for testing. The split was performed once and held fixed across all baseline experiments to prevent evaluation drift.

Given the small sample size, no cross-validation was applied. Repeated resampling would have reduced the effective training set and amplified variance effects. Instead, emphasis was placed on consistency between training behavior and performance on genuinely unseen records

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb129

Table 1 Sample Operational Logbook Records

Distance Run (NM)	Vessel Speed	Cargo Load	Wave Height (m)	Wind Speed (Knots)	Trim	Fuel Consumption
226	9.40	9417	3.25	18.4	-0.02	10.50
257	10.70	9417	3.25	24.3	-0.02	10.48
219	9.10	9417	3.25	24.3	-0.02	10.50
236	9.70	2340	1.88	18.4	-2.50	9.79
248	10.00	2340	1.88	18.4	-2.50	9.80
224	9.33	7000	0.88	12.7	-0.94	9.96
246	8.90	7000	0.88	7.8	-0.94	10.21
232	10.00	6600	0.30	7.8	-0.40	9.58
186	10.30	2500	0.05	4.9	-2.10	7.35
258	10.70	2100	0.88	12.7	-2.40	9.80

➢ *Baseline Models*

Three ensemble-based regression models were selected as baselines: Gradient Boosting Regressor, Random Forest Regressor, and Extreme Gradient Boosting. These models were chosen not for novelty, but for their known robustness under nonlinear, low-dimensional settings. All models were trained using identical feature sets and a uniform preprocessing pipeline consisting of median imputation. No hyperparameter tuning was performed at this stage in order to isolate the effect of data regime rather than optimization effort as shwon in Fig 1.

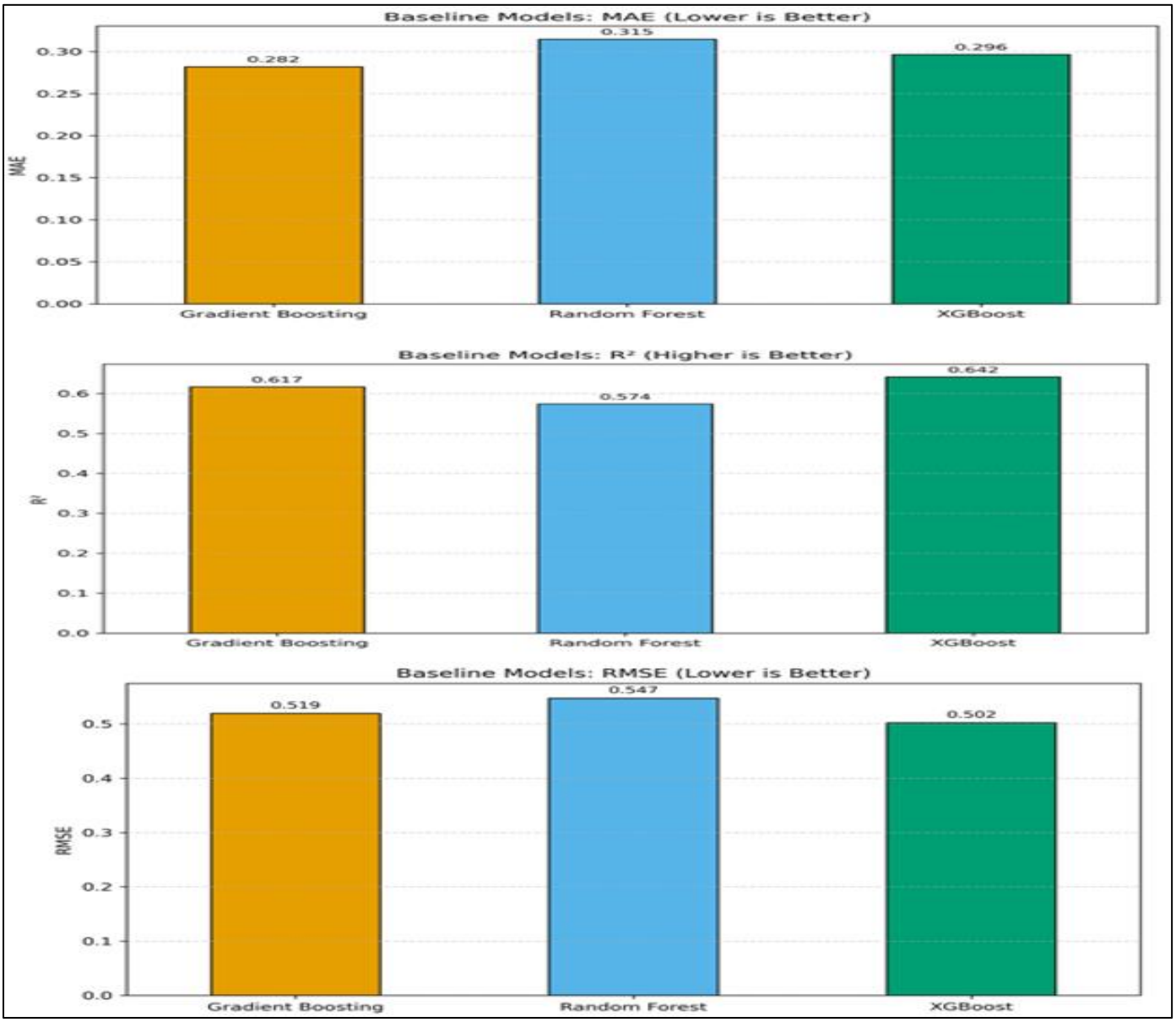

Fig 1 The MAE, R2, EMSE of Three Models Trained with Original Dataset

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb129

In table 2, performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R2). These metrics jointly characterize absolute deviation, sensitivity to large errors, and explained variance, which is necessary under small-sample conditions where single metrics can be misleading.

Table 2 The MAE, R2, EMSE of Three Models Trained with Original Dataset

Model	MAE	RMSE	R^2
XGBoost	0.2963	0.5017	**0.6416**
Gradient Boosting	0.2818	0.5189	0.6166
Random Forest	0.3147	0.5473	0.5736

➢ *Distribution-Preserving Data Augmentation*

To address instability arising from limited sample size, a constrained data augmentation procedure was introduced. The training subset was expanded via bootstrap resampling with replacement, followed by the injection of Gaussian perturbations scaled to a fixed fraction of each feature's empirical standard deviation. Perturbation magnitude was set conservatively to avoid altering marginal distributions or introducing implausible operating states.

Post-perturbation, all features were clipped at physically meaningful bounds to enforce non-negativity and operational feasibility. The augmented samples were concatenated with the original training data to produce an expanded dataset of 1,000 records. This procedure preserves first-order statistics and approximate covariance structure while increasing coverage of the observed operating space.

➢ *Training on Augmented Data*

The same baseline models were retrained on the augmented dataset using the identical preprocessing and evaluation protocol. Train–test splitting was repeated with the same random seed to ensure comparability with non-augmented results. Model artifacts, predictions, and performance metrics were stored for subsequent comparison.

The methodological focus is not on achieving maximal in-sample accuracy, but on evaluating whether augmentation stabilizes model behavior and improves predictive performance on held-out data without introducing distributional distortion.

➢ *Holdout Validation and Comparative Evaluation*

Final evaluation was conducted using independent holdout records not involved in augmentation or model fitting. Predictions from models trained on original and augmented datasets were compared against observed fuel consumption values. Error distributions, percentage deviations, and global metrics were computed to assess calibration and generalization.

Comparative analysis emphasizes relative behavior between models trained under different data regimes. Improvements are interpreted as evidence of recovered predictive structure rather than absolute performance gains in Fig 2. This framing is necessary to distinguish genuine generalization from artefacts induced by synthetic data inflation.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb129

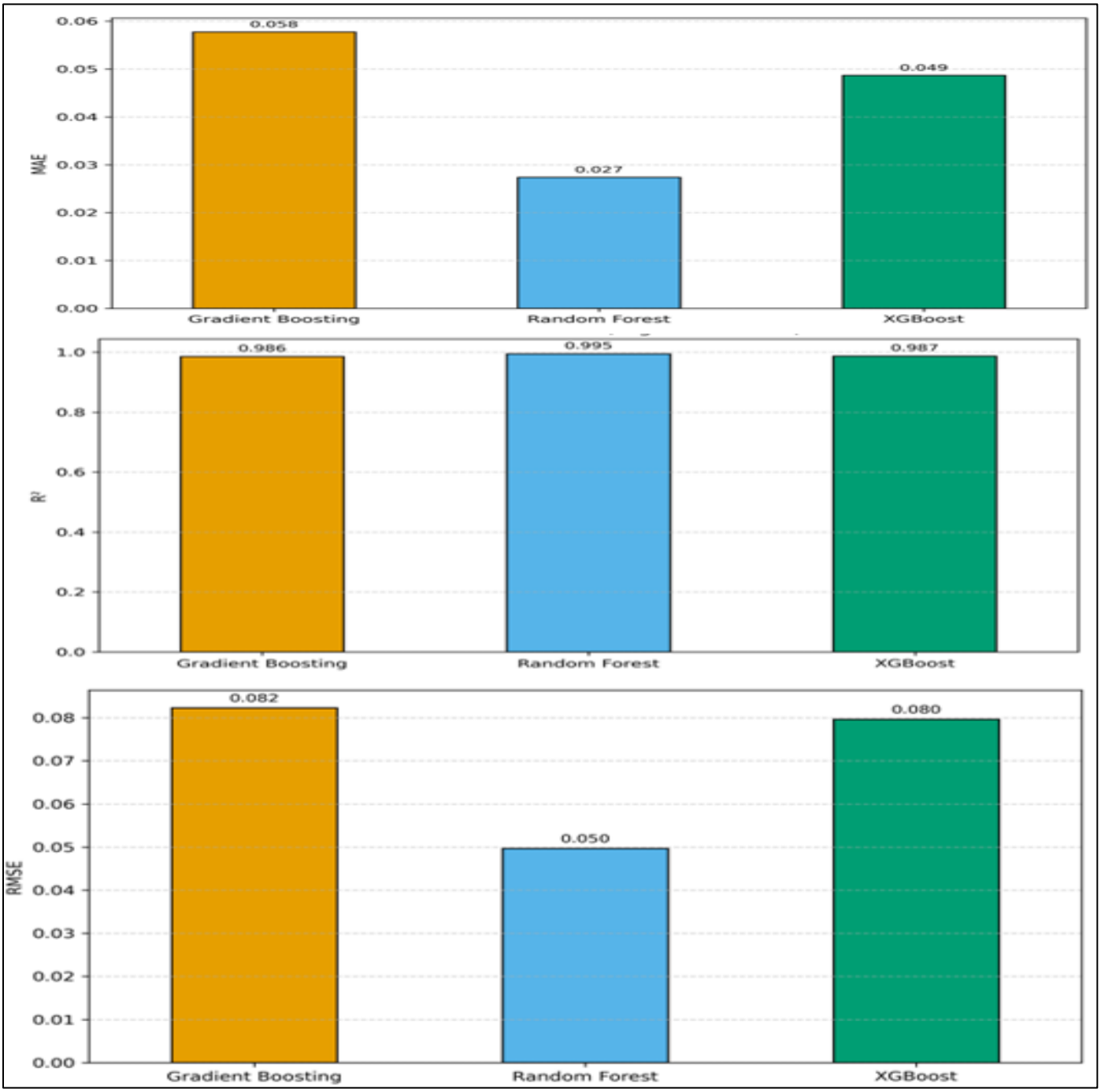

Fig 2 The MAE, R2, EMSE of New Models Trained with 1000 Augmented Data

III. RESULTS AND DISCUSSION

➢ *Baseline Performance on the Original Dataset*

Baseline models trained on the original 76-record dataset exhibited constrained but interpretable performance. Among the three ensemble regressors, XGBoost achieved the highest explanatory power, with R2=0.642, followed by Gradient Boosting (R2=0.617) and Random Forest (R2=0.574). Absolute error levels remained moderate across models, with MAE values clustered between 0.28 and 0.32 tons/day.

These results should not be interpreted as algorithmic inadequacy. With six input variables and sparse coverage of the operating space, the effective degrees of freedom are dominated by sampling variance. The observed ceiling in R2 reflects incomplete representation of operational regimes rather than systematic bias. Model rankings were stable across metrics, indicating that performance differences arise from variance reduction capacity rather than feature sensitivity.

Tree-based ensemble methods outperformed linear baselines under this regime, consistent with their ability to capture threshold effects and interaction structure without assuming functional form. However, none of the models demonstrated strong generalization margins at this scale, underscoring the limits imposed by data volume.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb129

➢ *Effect of Distribution-Preserving Augmentation*

Augmentation expanded the training set from 61 to 1,000 records while maintaining empirical distributions and physical bounds. Retraining on the augmented dataset produced a pronounced shift in model behavior. All three models exhibited substantial reductions in MAE and RMSE, accompanied by near-complete recovery of explained variance.

The Random Forest model achieved the strongest test performance, with MAE = 0.027 tons/day, RMSE = 0.050 tons/day, and R2=0.995. XGBoost and Gradient Boosting followed closely, both exceeding R2=0.985 as shown in table 3. Error magnitudes decreased by an order of magnitude relative to the non-augmented baseline, indicating that variance suppression rather than marginal fitting drove the improvement.

Table 3 Performance Comparison of Machine Learning Models

Model	**MAE**	**RMSE**	**R^2**
XGBoost	0.049	0.080	0.987
Gradient Boosting	0.058	0.082	0.986
Random Forest	0.027	0.050	0.995

Importantly, these gains did not arise from distributional distortion. Summary statistics and feature ranges of the augmented dataset remained aligned with the original data, and predictions remained within operationally plausible bounds. The improvement therefore reflects stabilization of the learned mapping rather than artefactual overfitting.

Percentage-based errors followed the same pattern. Mean absolute percentage error decreased from 2.57% to 1.80%, and symmetric MAPE showed a comparable reduction as in table 4. These differences are operationally non-trivial given the narrow absolute range of daily fuel consumption observed in the dataset.

➢ *Validation on Independent Holdout Voyages*

Generalization was evaluated using held-out voyage records not involved in training or augmentation. When applied to this dataset, the baseline XGBoost model trained on the original data achieved R2=0.816, with MAE = 0.240 tons/day. The augmented Random Forest model improved performance across all metrics, reaching R2=0.856 and MAE = 0.172 tons/day.

Direct comparison between the two prediction sets showed a mean symmetric percent difference of approximately 1.37%, indicating close calibration between models as shown in figure 3. The augmented model's advantage arises from reduced dispersion rather than systematic shift, suggesting improved robustness rather than altered bias.

Table 4 Comparison of Error Percentage Between Baseline and Augmented Detection

Model	MAE	RMSE	R^2	MAPE (%)	sMAPE (%)
Base	0.240	0.300	0.816	2.57	2.54
1000	0.172	0.265	0.856	1.80	1.79

➢ *Error Structure and Operational Consistency*

Error distributions for both models were compact and centered near zero, with the augmented model exhibiting tighter concentration and fewer extreme deviations. Visual inspection of prediction-versus-actual plots confirmed close alignment with the identity line across the full operating range. Localized discrepancies persisted at the edges of the observed space, particularly under high wind and extreme trim conditions, but their magnitude was reduced after augmentation.

Predicted fuel consumption values responded monotonically to distance run and vessel speed, consistent with propulsion fundamentals. Secondary effects associated with trim and wind speed were preserved without amplification, indicating that augmentation did not introduce spurious sensitivities. This consistency is critical for operational use, where implausible response patterns undermine trust regardless of numerical accuracy.

➢ *Interpretation Under Small-Data Constraints*

The results clarify the role of augmentation under logbook-scale data conditions. Without augmentation, ensemble models capture partial structure but remain variance-limited. With constrained augmentation, the same models recover stable mappings that generalize to unseen voyages without violating physical plausibility.

The improvement does not imply that synthetic data substitute for real observations. Rather, augmentation densifies the observed operating manifold, allowing ensemble methods to average over noise that would otherwise dominate fitting. The effectiveness of the approach depends on preserving covariance structure and enforcing domain constraints; unconstrained augmentation would likely inflate apparent accuracy while degrading operational validity.

From a methodological perspective, the findings indicate that data regime manipulation can be as consequential as model selection when observational capacity is limited. Under such conditions, predictive reliability is governed less by algorithmic sophistication than by how effectively the available data represent the operating space.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb129

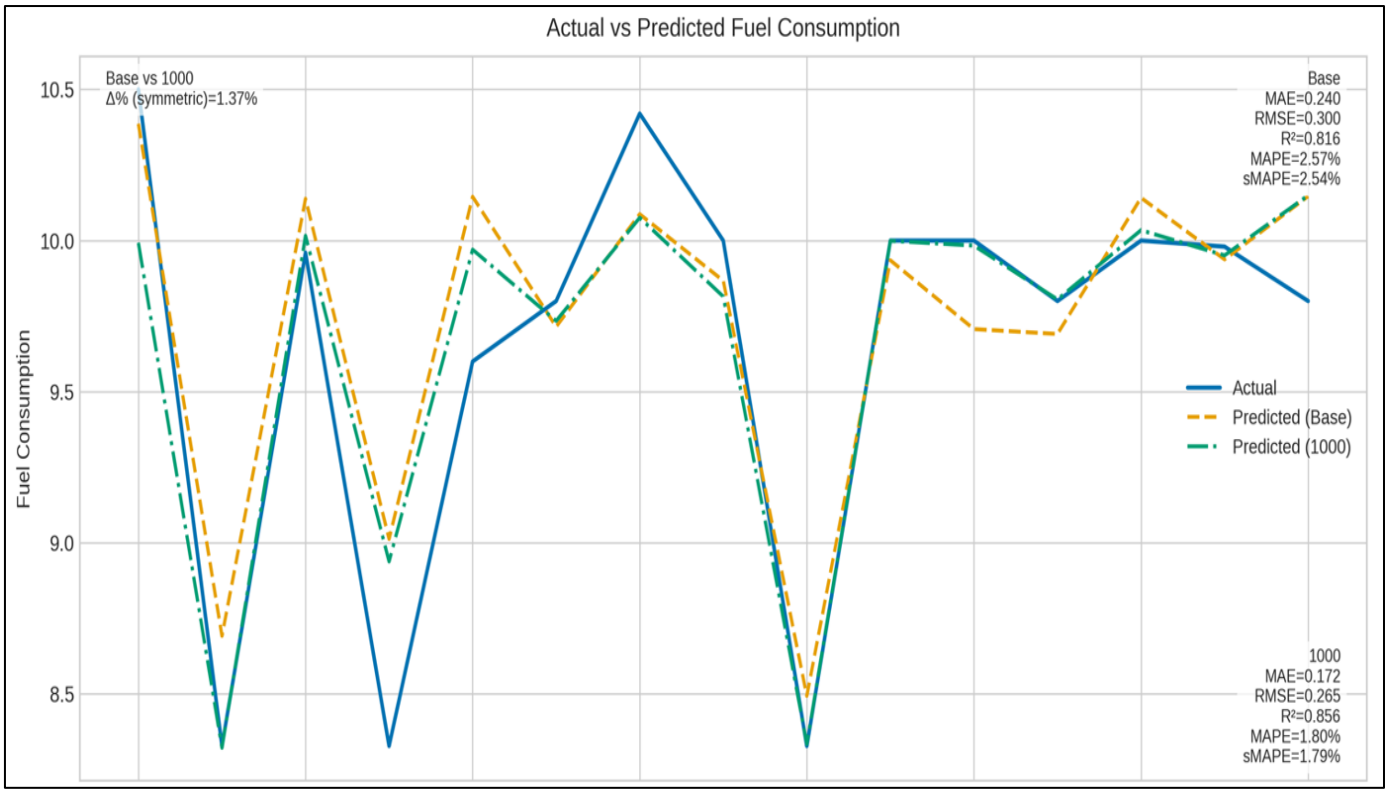

Fig 3 Actual vs Predicted Fuel Consumptions with Unseen Data

IV. OPTIMIZATION OF FUEL CONSUMPTION

After improvement of the prediction model, an optimization phase was carried out to determine fuel-efficient operating conditions for a bunker vessel. To ensure consistency and reliability, all optimization experiments used the same pre-trained Random Forest model (Random.pkl), which was previously identified as the best-performing and most stable predictor after distribution-preserving data augmentation.

Total voyage fuel consumption was defined as the product of the Random Forest–predicted fuel rate and the estimated time of arrival (ETA). Vessel speed and trim were treated as decision variables, while distance run, cargo load, wave height, and wind speed were fixed to represent a typical operating scenario. Practical bounds were imposed on speed (7–12 kn) and trim (−2.5 to 0.0 m) to maintain operational feasibility. Four population-based optimization techniques, Genetic Algorithm (GA), Real-Coded Genetic Algorithm (RCGA), NSGA-II, and Particle Swarm Optimization (PSO), were applied using the same Random Forest surrogate model and identical objective formulation. This design ensures that any differences in outcomes arise from the optimization strategy itself rather than from variations in the predictive model.

Across all optimizers, convergence was achieved at nearly identical operating points. The optimal solution consistently corresponded to a vessel speed close to the upper bound (approximately 12 kn) and a negative trim between −2.07 m and −2.16 m. Under these conditions, total voyage fuel consumption decreased from 8.8082 t in the baseline case to 6.9941 t, yielding a fuel saving of 1.8141 t, or 20.6% as shown in table 5.

Table 5 Comparison of Optimization Results Using the Pre-Trained Random Forest Mode

Optimizer	Objective	Optimal Speed (kn)	Optimal Trim (M)	ETA(h)	Total fuel (t)	Fuel saving vs baseline
Baseline	Observed Operation	9.80	-1.20	22.45	8.8082	-
GA	Minimize total (rate* ETA)	12.00	-2.156	18.33	6.9941	-1.8141t (-20.6%)
RCGA	Minimize (total fuel, ETA)	12.00	-2.125	18.33	6.9941	-1.8141t (-20.6%)
NSGA-II	Minimize (total fuel, ETA)	12.00	-2.125	18.33	6.9941	-1.8141t (-20.6%)
PSO	Minimize (rate* ETA)	12.00	-2.071	18.33	6.9941	-1.8141t (-20.6%)

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb129

The agreement among GA, RCGA, NSGA-II, and PSO indicates that the optimized solution is governed primarily by the structure learned by the Random Forest fuel model, rather than by optimizer-specific behavior. In the examined operating range, the reduction in ETA with increasing speed was not offset by a proportional increase in the predicted fuel rate, resulting in minimum total fuel consumption at the upper speed bound.

A local sensitivity analysis around the optimum showed only minor variations in total fuel consumption, confirming that the identified solution lies within a stable region of the operating space. These results demonstrate that, when a robust Random Forest model is used as a surrogate, metaheuristic optimization can provide consistent and operationally meaningful fuel-saving recommendations even under limited logbook data conditions.

V. CONCLUSION

This study examined ship fuel consumption prediction under conditions where only small-scale operational logbook data are available. Using two years of records from a bunker vessel, baseline ensemble models exhibited moderate predictive capability, with performance bounded by sample size rather than model structure. Moreover, fuel consumption optimzation task was performed. These results reflect a common operational reality: limited data coverage constrains inference more strongly than algorithm choice.

Introducing a distribution-preserving augmentation procedure altered this regime. By expanding the training set through constrained bootstrap resampling with covariance-scaled perturbations, model variance was substantially reduced without distorting empirical distributions or violating physical bounds. Ensemble models retrained on the augmented data achieved stable and accurate predictions, and with the prediction model based, optimization results showed very good. The findings carry practical implications for fuel prediction and optimization in fleets without using sensor data. Although the prediction results are quite satisfied and reliable, the optimization results are still needed to make more robust and reliable, which leads us to the further studies for this research work.

REFERENCES

[1]. "2023 IMO Strategy on Reduction of GHG Emissions from Ships." Accessed: Nov. 14, 2025. [Online]. Available: https://www.imo.org/en/ourwork/environment/pages/2023-imo-strategy-on-reduction-of-ghg-emissions-from-ships.aspx

[2]. United Nations Conference on Trade and Development, Review of Maritime Transport 2024. 2024. Accessed: Oct. 16, 2025. [Online]. Available: https://unctad.org/system/files/official-document/rmt2024_en.pdf

[3]. A. Fan, J. Yang, L. Yang, D. Wu, and N. Vladimir, "A review of ship fuel consumption models," Ocean Engineering, vol. 264, p. 112405, Nov. 2022, doi: 10.1016/j.oceaneng.2022.112405.

[4]. R. Campbell, M. Terziev, T. Tezdogan, and A. Incecik, "Computational fluid dynamics predictions of draught and trim variations on ship resistance in confined waters," Applied Ocean Research, vol. 126, p. 103301, Sep. 2022, doi: 10.1016/j.apor.2022.103301.

[5]. T. Uyanık, Ç. Karatuğ, and Y. Arslanoğlu, "Machine learning approach to ship fuel consumption: A case of container vessel," Transportation Research Part D: Transport and Environment, vol. 84, p. 102389, Jul. 2020, doi: 10.1016/j.trd.2020.102389.

[6]. C. Papandreou and A. Ziakopoulos, "Predicting VLCC fuel consumption with machine learning using operationally available sensor data," Ocean Engineering, vol. 243, p. 110321, Jan. 2022, doi: 10.1016/j.oceaneng.2021.110321.

[7]. S. Wang, B. Ji, J. Zhao, W. Liu, and T. Xu, "Predicting ship fuel consumption based on LASSO regression," Transportation Research Part D: Transport and Environment, vol. 65, pp. 817–824, Dec. 2018, doi: 10.1016/j.trd.2017.09.014.

[8]. IN HEAVY VEHICLES," International Journal of Innovative Research in Technology, vol. 9, no. 12, pp. 1230–1239, May 2023.

[9]. Y. Gu, Y. Wang, and J. Zhang, "Fleet deployment and speed optimization of container ships considering bunker fuel consumption heterogeneity," MSE, vol. 1, no. 1, p. 3, Oct. 2022, doi: 10.1007/s44176-022-00003-2.