УДК 656.6:004.8 doi.org/10.33298/2226-8553.2026.1.45.16

© Kochubei P., Appazov E.

FUELSIGHT: MODEL FOR PREDICTING SHIP FUEL CONSUMPTION AND PROVIDING EXPLANATIONS

The ability to predict ship fuel consumption is fundamental to improving the efficiency and sustainability of maritime operations. Existing approaches predominantly focus on predictive accuracy while overlooking interpretability, thereby limiting their practical applicability for operational decision-making. This study proposes FuelSight, a unified model architecture that jointly performs multi-horizon fuel consumption forecasting and generates structured, human-readable insights reflecting seafarers' best practices. The proposed framework leverages multivariate time-series data derived from onboard telemetry and environmental conditions to capture complex operational dynamics. A large language model backbone based on GPT-2 is employed to process sequential inputs and enable both numerical prediction and explanation generation within a single architecture, providing an efficient and coherent modeling approach. Empirical evaluation on the FuelCast benchmark demonstrates that the proposed method achieves competitive performance across multiple vessels and forecasting horizons. In addition to numerical accuracy, the model produces interpretable outputs, which achieve good results when evaluated as a structured classification task. The results indicate that integrating forecasting and explanation within an LLM-based framework offers a promising direction for developing interpretable decision-support systems in maritime applications.

Keywords: *machine learning, large language models, XAI, interpretability, ship fuel consumption forecasting, decision support system, human factor, automatic control system*

Introduction. Maritime shipping is the underpinning of global commerce and accounts for over 90% of international trade volume [1]. However, the industry's heavy reliance on fossil fuels presents a growing environmental and economic challenge. The International Maritime Organization (IMO) estimates that global shipping accounted for 1,056 billion tons of CO_2 emissions in 2018, representing 2.89% of global greenhouse gas emissions. By 2025, emissions could reach 90% or even 130% of 2008 levels [2]. In response, the IMO has introduced regulatory instruments, including the Energy Efficiency Existing Ship Index (EEXI) and the Carbon Intensity Indicator (CII), which require operators to maintain detailed awareness of their vessels' energy consumption profiles. At the same time, fuel costs represent a significant and rising proportion of fleet operating expenses, making fuel optimization both an environmental and commercial imperative [3].

Accurate prediction of ship fuel consumption is necessary to address those challenges. Assessment of future consumption enables route optimization and voyage planning that could reduce both emissions and operational costs [4]. However, the prediction task is inherently complex: fuel consumption depends on the interplay among vessel speed, hull condition, engine characteristics, and environmental factors such as wind, waves, currents, and water depth [5].

Despite significant progress in data-driven fuel prediction – from statistical methods to classical machine learning to deep neural networks – two gaps remain. First, there's a lack of standardized and open benchmarks for comparing models. Second, almost all research focuses on prediction accuracy and overlooks interpretability, thereby leaving unexplored the "black-box" predictive value that navigators can use for operational decision-making.

In this work, we propose FuelSight, a dual-head architecture that addresses both gaps. Built on a frozen GPT-2 backbone with low-rank adaptation, the model jointly produces multi-horizon fuel consumption forecasts and human-readable prescriptive advisories. We evaluate on the FuelCast benchmark [6], a recently published open dataset. Our main contributions are as follows. First, we demonstrate that a single LLM-based architecture achieves state-of-the-art forecasting accuracy, outperforming both traditional machine learning baselines and zero-shot foundation models on four of six ship–horizon configurations. Second, we introduce a prescriptive explanation head that generates structured operational advisories — including regime classification, fuel trend analysis, and condition summaries — at zero additional parameter cost, through weight tying with the language model's embedding matrix.

Related Work. Various approaches exist for predicting ship fuel consumption, ranging from physics-based models and statistical methods [7] to simple and sophisticated machine learning methods such as Gaussian Process Regression and Back-Propagation Neural Networks [8], Artificial Neural Networks [9], and gradient-boosted ensembles including XGBoost [10]. Recently, deep learning architectures have dominated the field – LSTMs [11], Bi-LSTMs with multi-head self-attention [12], genetic algorithm-optimized LSTMs [13], hybrid TCN-GRU architectures with attention mechanisms [14], and CNN-BiGRU networks with metaheuristic optimization [15].

Most published studies rely on proprietary datasets from individual vessels – a single tanker [16], one container ship [17], a Kamsarmax bulk carrier [18], or short-duration records of two to three months [19]. This heterogeneity makes it impossible to compare reported accuracies across studies. Most recently, researchers [6] introduced the FuelCast benchmark dataset, which is easily accessible to researchers worldwide and provides extensive data on ship telemetry and weather conditions. Such a benchmark is a significant step forward in energy consumption, enabling standardized, iterative, and measurable improvements in model performance. We are grateful for this contribution and build our work upon their foundation.

Another important gap is the field's singular focus on architectural improvements while disregarding interpretability and insights for the end user – ship operators. A singular numeric forecast of 0.45 kg/s offers no actionable guidance on whether adjusting speed or course would meaningfully impact consumption. The classical approach is to use explanation frameworks such as SHAP [20] or LIME [21]. However, the recent advent of Large Language Models, whose textual nature implicitly provides a foundation for the model's contextual output, motivates inquiry into their application in this regard. Additionally, the emergence of reprogramming approaches such as Time-LLM [22], cross-modal fine-tuning frameworks such as CALF [23], and autoregressive adaptation methods such as AutoTimes [24] has demonstrated that pre-trained language models can be effectively adapted for time-series prediction tasks. These approaches demonstrated strong performance on standardized time-series benchmarks such as ETTh, Weather, and Electricity, and showed that LLMs can be effectively used for time-series predictions with minimal adjustments.

In the maritime domain, AIS-LLM [] has been proposed as an LLM-based architecture that describes AIS data, provides intelligent insights into trajectory prediction, anomaly detection, and collision risk, and serves as a descriptive author. We draw inspiration from their work and propose a system that not only describes but also provides actionable insights to the navigator.

Given the predictive performance of LLMs and inherent natural language capabilities, we propose an architecture that, in addition to providing high-accuracy forecasts, generates operational advisories for maritime leadership, bridging the gap between predictive performance and decision support.

Data. We adopt the FuelCast benchmark dataset, a publicly available maritime fuel-consumption dataset designed to enable standardized comparison of machine learning models. The dataset contains operational and environmental measurements from three vessels (CPS Poseidon, CPS Triton, and OSS Ceto). See Table 1 for details.

Table 1 – Overview of the FuelCast ships and corresponding data

	Type	Gross Tonnage	# Samples	Missing Values
CPS Triton	Cruise Passenger Ship	11,000	25,351	0.04 %
CPS Poseidon	Cruise Passenger Ship	70,000	105,422	3.2 %
OSS Ceto	Offshore Supply Ship	24,000	43,213	0.96 %

Each ship's dataset contains operational data, such as engine RPM and environmental variables, including wind direction and wave height. To ensure comparability and reproducibility, we follow the feature selection approach in the original paper, with plans to extend it to more operational variables in the future. See Table 2 for an overview of the variables.

Table 2 – Overview of FuelCast variable/features used for modeling

Variable	Description	Source	Unit
Total.MomentaryFuel	Total momentary fuel consumption of all consumers on the vessel	Flowmeter	kg/s
SpeedOverGround	Speed over ground of the vessel	GPS	m/s
SeaFloorDepth	Sea floor depth below sea level (bathymetry)	Copernicus Marine	m
WindDirection10M	Wind direction at 10 meters above ground	Open-Meteo	°
WindSpeed10M	Wind speed at 10 meters above ground	Open-Meteo	m/s
OceanCurrentDirection	Ocean current direction considering all components	Open-Meteo	°
OceanCurrentVelocity	Ocean current velocity considering all components	Open-Meteo	m/s
WaveDirection	Mean direction of significant waves	Open-Meteo	°
WaveHeight	Significant mean wave height	Open-Meteo	m
WavePeriod	Period between significant waves	Open-Meteo	s
Temperature2M	Air temperature 2 meters above ground	Open-Meteo	°C

Preprocessing. For each vessel, we retain the selected variables and temporal index. Rows with missing temporal indices are removed, the index column is cast to integer format, and the rows are sorted chronologically.

Transformations. We process the directional variables – wind direction, ocean current direction, and wave direction –relative to the ship bearing using Eq. 1. Each relative angle is encoded as its sine and cosine components. After this step, the final input representation consists of 13 operational and environmental features.

$$\theta_{rel} = \left(\theta_{env} - \theta_{bearing}\right) \, mod \; 360° \tag{1}$$

Cross-validation. Each dataset is divided into five contiguous intervals of approximately equal size. For each fold, one interval is used as the test set, and the remaining four form the training set. The validation subset is constructed from the end of the ordered training anchors subset, ensuring that validation data always occur after the corresponding training samples. An anchor is considered valid only if the full lookback window and the full prediction horizon lie entirely within the same interval.

Missing values. The missing values are imputed with per-channel mean values computed from the rows covered by the training windows.

Methodology. Our model architecture is depicted in Figure 1 and is designed to predict the vessel's fuel consumption K-steps. We have selected prediction intervals of 15 and 30 min as the most operationally viable. In addition to predicting the scalar value of fuel consumption, our model outputs human-readable explanations intended to provide a framework for informed decisions based on planned actions and expected conditions.

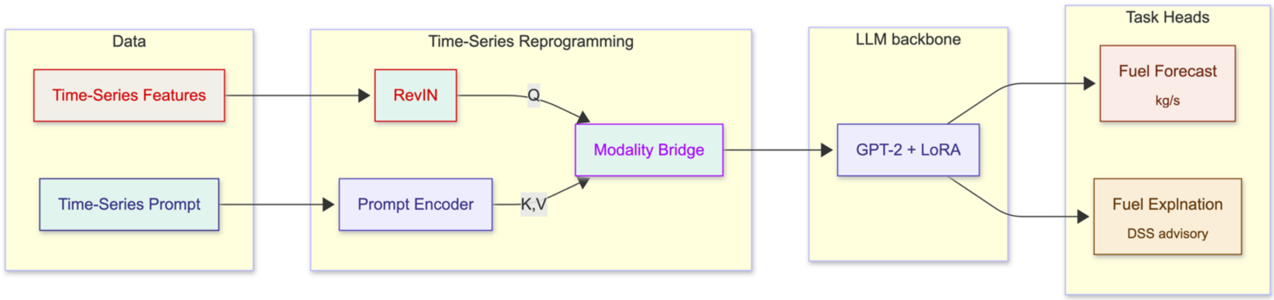

Figure 1 – Overview of LLM-based fuel prediction

Time-series Features. Time-series data collected onboard vessels exhibit non-stationary distributions that shift across voyages, seasons, and operational regimes. To address this, we apply Reversible Instance Normalization (RevIN) [26], which standardizes each sample independently and retains sufficient statistics to restore the original scale at inference:

$$\hat{X} = \gamma \odot \frac{X-\mu}{\sigma} + \beta \quad (2)$$

The output preserves the original feature dimensionality for downstream cross-modal alignment.

Time-series Prompts. We construct natural-language prompts that describe each lookback window using domain-specific maritime terminology — sea state, speed regime, and fuel trend. Those prompts are meant to serve as maritime best practices, basically summarizing the situation from a human perspective. For example, we have included basic speed recommendations based on wave height [27, 28]. Still, the prompts can also contain statistical distributions or feature attribution from surrogate models that use SHAP or LIME explanations.

The prompts are embedded using the frozen GPT-2 [29] token embedding layer and refined through a single self-attention layer:

$$H^P = \mathrm{SelfAttention}(\mathrm{wte}(p)) \quad (3)$$

This refinement captures inter-token dependencies within the prompt before alignment with the time-series branch.

The following labels are designed using simple heuristics to illustrate maritime best practices. For sea state, we're using a simplified Douglas scale with three classes: CALM, MODERATE, and ROUGH. The speed regime is classified into LOW, NORMAL, and HIGH, and the recommended action into REDUCE_SPEED, INCREASE_SPEED, and MAINTAIN_SPEED.

Modality Bridge. The temporal encoder and prompt encoder produce representations in different embedding spaces — the time-series branch operates in the original feature dimension C. In contrast, the prompt branch operates in the LLM dimension d. Modality bridge reconciles these two modalities through a cross-attention mechanism where time-series tokens attend to the prompt embeddings:

$$H^A = \mathrm{CrossAttention}(H^{TS}W_q,\ H^P,\ H^P) + H^{TS}W_q \quad (4)$$

This allows each timestep in the time series to selectively attend to the most relevant parts of the textual context. For instance, a timestep with high wave impact can attend more strongly to wave-related prompt tokens. The projection $\in \mathbb{R}^{C\times d}$is the sole dimensionality-reduction bridge between the two modalities throughout the entire architecture.

LLM Backbone. Rather than training a forecasting model from scratch, we leverage the sequential reasoning capabilities of a pre-trained GPT-2 model, whose self-attention mechanism is inherently suited to processing ordered sequences. The backbone is kept frozen to preserve the general-purpose representations acquired during pre-training, while task-specific adaptation is achieved through LoRA [30] — low-rank matrices injected into the attention layers:

$$h = W_0 x + BAx \tag{5}$$

where W_0 are the frozen weights B, A are the trainable low-rank factors. This strategy adapts the model to the maritime fuel domain while keeping the trainable parameter count around 1-2% of the total, significantly reducing the risk of overfitting on smaller ship datasets. A single set of adapters serves both the forecasting and explanation tasks, as both rely on the same contextual understanding of the input sequence.

Forecast Head. Direct prediction of absolute fuel consumption values is challenging because the target distribution varies substantially across operational regimes. Instead, we adopt a residual prediction approach [31] in which the model predicts normalized deviations from the last observed fuel value, reducing the learning problem to small corrections around a strong baseline:

$$\hat{y} = y_T + \hat{\delta} \odot \sigma_\delta + \mu_\delta \tag{6}$$

where y_T is the last observed fuel consumption, $\hat{\boldsymbol{\delta}}$ is the predicted normalized delta, and $\boldsymbol{\mu}_\delta$ $\boldsymbol{\sigma}_\delta$ are per-horizon statistics computed from the training set. This formulation stabilizes training and improves accuracy at longer horizons, where cumulative error in absolute predictions would otherwise compound. A separate head is maintained for each forecast horizon.

Explanation head. Numerical forecasts alone provide limited operational value without context for why fuel consumption is expected to change. To address this, the model generates prescriptive advisories through a causal language modeling head whose weights are tied to the GPT-2 embedding matrix, introducing zero additional parameters. During training, ground-truth advisory strings — containing regime classification, trend analysis, and condition summaries — are appended to the input sequence. The causal attention mask in GPT-2 ensures that appended text tokens cannot influence the time-series position hidden states. Hence, the forecast head produces identical outputs regardless of whether explanation tokens are present. This property enables both tasks to be served by a single forward pass without interference:

$$H = \mathrm{GPT2}([H^P | H^A | E^{exp}]) \tag{7}$$

In inference, advisories are generated autoregressively from the shared hidden states.

Training objective. Both heads are jointly trained via shared LoRA adapters. The total loss combines a Smooth L1 loss for the forecast task with a weighted cross-entropy loss for the explanation task:

$$\mathcal{L} = \mathcal{L}_{SmoothL1}(\hat{\delta}, \delta^*) + \lambda \cdot \mathcal{L}_{CE} \tag{8}$$

The explanation task, being a low-entropy template completion, converges within the first few epochs and thereafter contributes negligible gradients — acting as an implicit regularizer for the shared adapter weights during early training without degrading forecast accuracy at convergence.

Experimental setup. We evaluate FuelSight against four baseline models spanning three distinct paradigms. XGBoost is a gradient-boosted tree method that has consistently demonstrated strong performance on structured tabular data in the maritime domain [10]. XLinear [32] serves as a lightweight linear baseline that isolates the contribution of non-linear modeling. For the foundation model paradigm, we include TimesFM [33], a decoder-based time-series foundation model pre-trained on a large corpus of public time-series data, and Chronos-2 [34], a probabilistic foundation model that tokenizes time-series values and generates forecasts autoregressively. Both foundation models are evaluated in a zero-shot setting with covariates — no task-specific fine-tuning is performed. We acknowledge the asymmetry between these zero-shot baselines and our fine-tuned model. All these models demonstrate strong performance in predicting fuel consumption; however, our focus is to show that our model performs on par with large models while also providing meaningful insights into its performance.

Evaluation metrics. Forecasting performance is assessed using three complementary metrics computed per horizon and averaged across 5 folds: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). MAE serves as the primary metric for early stopping and model selection. All metrics are reported on the original scale (kg/s) after delta denormalization, ensuring comparability across models.

Results. Table 4 reports the mean MAE and R^2 across 5-fold temporal cross-validation for all models and ship–horizon combinations. FuelSight achieves the lowest MAE on four of the six configurations – both horizons for Ceto and Poseidon — while remaining competitive on Triton, the smallest dataset. On Ceto at the 15-minute horizon, FuelSight attains a MAE of 0.0086 kg/s, improving over the next best baseline (TimesFM, 0.0093 kg/s) by approximately 8%. On Poseidon, FuelSight achieves $R^2 = 0.978$ at the 15-minute horizon, the highest across all models and configurations.

On the smaller Triton dataset (~25,000 rows), XGBoost outperforms all deep learning and foundation model approaches, achieving a MAE of 0.0138 kg/s at the 15-minute horizon compared to 0.0151 kg/s for FuelSight. This result is consistent with the well-documented advantage of gradient-boosted tree methods in low-data regimes, where deep architectures risk overfitting. Notably, FuelSight still outperforms both zero-shot foundation models (TimesFM and Chronos-2) on Triton at the 15-minute horizon, suggesting that domain-specific fine-tuning provides measurable benefit even on limited data.

Table 3 – Results comparison across evaluated models

Ship	***Horizon***	***Metric***	***XGBoost***	***XLinear***	***TimesFM***	***Chronus-2***	***FuelSight***
Ceto	3	MAE	0.0173	0.0099	0.0093	0.0098	**0.0084**
Ceto	3	R^2	0.879	0.938	0.934	0.929	**0.949**
Ceto	6	MAE	0.0223	0.0118	0.0117	0.0125	**0.0109**
Ceto	6	R^2	0.821	**0.911**	0.894	0.880	0.916
Poseidon	3	MAE	0.0341	0.0334	0.0316	0.0339	**0.0301**
Poseidon	3	R^2	0.976	0.974	0.971	0.972	**0.978**
Poseidon	6	MAE	0.0432	0.0390	0.0395	0.0423	**0.0368**
Poseidon	6	R^2	0.961	0.962	0.950	0.952	**0.965**
Triton	3	MAE	**0.0138**	0.0154	0.0175	0.0167	0.0148
Triton	3	R^2	**0.874**	0.820	0.727	0.752	0.815
Triton	6	MAE	**0.0193**	0.0211	0.0239	0.0231	0.0217
Triton	6	R^2	**0.785**	0.686	0.541	0.577	0.675

The zero-shot foundation models (TimesFM and Chronos-2) achieve strong performance without any task-specific training, particularly on Ceto, where TimesFM approaches FuelSight's accuracy. However, neither foundation model produces operational insights beyond numerical forecasts.

Evaluating explanation head. The results are presented in Table 4. We structure the evaluation of explanation code as a structured classification problem rather than relying on ROUGE or BLEU, which are commonly accepted in NLP circles [36]. In our case, we have decided to use classification metrics, as they are better suited to the contrived example of prompt and label outputs we're using in the following paper.

For each field, we report the average **F1 score**. Including the F1 score helps shed light on the imbalanced classes in the data we're evaluating. All metrics are computed independently for each fold within the same five-fold temporal cross-validation protocol used for forecasting. Additionally, we report the **Joint Exact Match (EM)** score, which measures strict multi-label correctness (9).

$$\text{Joint EM} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(\hat{y}_i = y_i) \tag{9}$$

This metric is meant to capture compositional reasoning accuracy across multiple decision dimensions.

Furthermore, we report **Consistency**, which evaluates whether the predicted action aligns with the direction of the predicted fuel consumption trend.

$$\text{Consistency} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\left(a_i = 0 \wedge \Delta\hat{y}_i = 0 \ \ \vee \ \ \text{sign}(a_i) = \text{sign}(\Delta\hat{y}_i)\right) \qquad (10)$$

We have included this metric to reflect the alignment between explanations and numerical forecasts analogous to faithfulness in explainable literature [37].

Table 4. Evaluation results of the explanation head

Ship	*Horizon*	*SEA STATE (F1)*	*FUEL RISK (F1)*	*SPEED ACTION (F1)*	*Joint EM*	*Consistency*
Ceto	3	0.865	0.491	0.660	0.503	0.732
Ceto	6	0.864	0.473	0.664	0.514	0.662
Poseidon	3	1.000	0.685	0.990	0.715	0.830
Poseidon	6	0.999	0.716	0.983	0.745	0.815
Triton	3	0.990	0.496	0.836	0.523	0.929
Triton	6	0.977	0.462	0.802	0.499	0.954

Conclusion. In our work, we proposed FuelSight, an integrated architecture for producing a k-step fuel forecast and generating structured, human-readable insights. The empirical evaluation of the FuelCast benchmark demonstrates competitive performance across multiple ship horizons. This further supports the claim that LLMs, when appropriately adapted, can be effective across multiple domains for time-series forecasting.

Additionally, our model produces syntactically valid, structured output that provides a good foundation for Explainable AI in the maritime industry. However, we acknowledge that further refinement and investigation are needed in this area. The following finding aligns with prior research that explanation methods often struggle to achieve faithfulness and stability, even when predictive performance is high [38].

Future work will focus on improving explanation fidelity and robustness. Specifically, we plan to adopt gradient-based attribution methods to ground explanations in model sensitivity and to explore methods for evaluating human-readable explanations produced by LLMs. Furthermore, we plan to extend the feature list used for modeling and include more ship-specific metrics, such as engine RPM, to provide more actionable insights.

Overall, the results suggest that LLM-based architectures provide a necessary foundation for combined prediction-explanation systems in the maritime industry.

REFERENCES

1. Pallotta G, Vespe M, Bryan K. Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction. Entropy. 2013; 15(6):2218-2245. https://doi.org/10.3390/e15062218.
2. IMO. (2020). Fourth Greenhouse Gas Study 2020. International Maritime Organization. https://www.imo.org/en/ourwork/environment/pages/fourth-imo-greenhouse-gas-study-2020.aspx
3. Yan, R., Wang, S., & Psaraftis, H. N. (2021). Data analytics for fuel consumption management in maritime transportation: Status and perspectives. Transportation Research Part E: Logistics and Transportation Review, 155, 102489.
4. Zhou, Y., Pazouki, K., Murphy, A. J., Uriondo, Z., Granado, I., Quincoces, I., & Fernandes-Salvador, J. A. (2023). Predicting ship fuel consumption using a combination of metocean and on-board data. Ocean Engineering, 285, 115509.
5. Wang, K., Wang, J., Huang, L., Yuan, Y., Wu, G., Xing, H., ... & Jiang, X. (2022). A comprehensive review on the prediction of ship energy consumption and pollution gas emissions. Ocean Engineering, 266, 112826.

6. Viga, J., Mueck, P., Löser, A., & Weis, T. (2025, September). Fuelcast: Benchmarking tabular and temporal models for ship fuel consumption. In International Workshop on Advanced Analytics and Learning on Temporal Data (pp. 54-69). Cham: Springer Nature Switzerland.

7. Petersen, J.P., Jacobsen, D.J. & Winther, O. Statistical modelling for ship propulsion efficiency. J Mar Sci Technol 17, 30–39 (2012). https://doi.org/10.1007/s00773-011-0151-0

8. Hu, Z., Jin, Y., Hu, Q., Sen, S., Zhou, T., & Osman, M. T. (2019). Prediction of fuel consumption for enroute ship based on machine learning. Ieee Access, 7, 119497-119505.

9. Nguyen, V. G., Sakthivel, R., Rudzik, K., Kozak, J., Sharma, P., Pham, N. D. K., ... & Nguyen, X. P. (2023). Using artificial neural networks for predicting ship fuel consumption. Polish Maritime Research

10. Agand, P., Kennedy, A., Harris, T., Bae, C., Chen, M., & Park, E. J. (2023). Fuel consumption prediction for a passenger ferry using machine learning and in-service data: A comparative study. Ocean Engineering, 284, 115271.

11. Lei, L., Wen, Z., & Peng, Z. (2021, September). Prediction of main engine speed and fuel consumption of inland ships based on deep learning. In Journal of Physics: Conference Series (Vol. 2025, No. 1, p. 012012). IOP Publishing.

12. Ilias, L., Kapsalis, P., Mouzakitis, S., & Askounis, D. (2023). A multitask learning framework for predicting ship fuel oil consumption. Ieee Access, 11, 132576-132589.

13. Wang, K., Hua, Y., Huang, L., Guo, X., Liu, X., Ma, Z., ... & Jiang, X. (2023). A novel GA-LSTM-based prediction method of ship energy usage based on the characteristics analysis of operational data. Energy, 282, 128910.

14. Liu, Y., Wang, K., Lu, Y., Zhang, Y., Li, Z., Ma, R., & Huang, L. (2024). A ship energy consumption prediction method based on TGMA model and feature selection. Journal of Marine Science and Engineering, 12(7), 1098.

15. Wang, Z., Wang, K., Li, Z., Liang, H., Yin, S., Ma, Q., ... & Xiong, W. (2026). A Novel ROA-Optimized CNN-BiGRU Hybrid Network with an Attention Mechanism for Ship Fuel Consumption Prediction. Journal of Marine Science and Engineering, 14(4), 324.

16. Pedersen, B. P., & Larsen, J. (2009, May). Prediction of full-scale propulsion power using artificial neural networks. In Proceedings of the 8th international conference on computer and IT applications in the maritime industries (COMPIT'09), Budapest, Hungary May (pp. 10-12).

17. Hu, Z., Jin, Y., Hu, Q., Sen, S., Zhou, T., & Osman, M. T. (2019). Prediction of fuel consumption for enroute ship based on machine learning. Ieee Access, 7, 119497-119505.

18. Zhang, M., Tsoulakos, N., Kujala, P., & Hirdaris, S. (2024). A deep learning method for the prediction of ship fuel consumption in real operational conditions. Engineering Applications of Artificial Intelligence, 130, 107425.

19. Petersen, J. P., Winther, O., & Jacobsen, D. J. (2012). A machine-learning approach to predict main energy consumption under realistic operational conditions. Ship Technology Research, 59(1), 64-72.

20. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

21. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

22. Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J. Y., Shi, X., ... & Wen, Q. (2023). Time-llm: Time series forecasting by reprogramming large language models. arXiv preprint arXiv:2310.01728.

23. Liu, P., Guo, H., Dai, T., Li, N., Bao, J., Ren, X., ... & Xia, S. T. (2025, April). Calf: Aligning llms for time series forecasting via cross-modal fine-tuning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 18, pp. 18915-18923).

24. Liu, Y., Qin, G., Huang, X., Wang, J., & Long, M. (2024). Autotimes: Autoregressive time series forecasters via large language models. Advances in Neural Information Processing Systems, 37, 122154-122184.

25. Park, H., Jung, J., Seo, M., Choi, H., Cho, D., Park, S., & Choi, D. G. (2025). AIS-LLM: A Unified Framework for Maritime Trajectory Prediction, Anomaly Detection, and Collision Risk Assessment with Explainable Forecasting. arXiv preprint arXiv:2508.07668.

26. Kim, T., Kim, J., Tae, Y., Park, C., Choi, J. H., & Choo, J. (2021, May). Reversible instance normalization for accurate time-series forecasting against distribution shift. In International conference on learning representations.

27. Degiuli, N., Ćatipović, I., Martić, I., Werner, A., & Čorić, V. (2017). Increase of ship fuel consumption due to the added resistance in waves. Journal of sustainable development of energy, water and environment systems, 5(1), 1-14.

28. Farkas, A., Degiuli, N., Martić, I., & Mikulić, A. (2023). Benefits of slow steaming in realistic sailing conditions along different sailing routes. Ocean engineering, 275, 114143.
29. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.
30. Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., & Chen, W. (2022). Lora:Low-rank adaptation of large language models. Iclr, 1(2), 3.
31. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
32. Chen, X., Jin, H., Huang, Y., & Feng, Z. (2026). XLinear: A Lightweight and Accurate MLP-Based Model for Long-Term Time Series Forecasting with Exogenous Inputs. arXiv preprint arXiv:2601.09237.
33. Das, A., Kong, W., Sen, R., & Zhou, Y. (2024, July). A decoder-only foundation model for time-series forecasting. In Forty-first international conference on machine learning.
34. Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., ... & Wang, Y. (2024). Chronos: Learning the language of time series. arXiv preprint arXiv:2403.07815.
35. Kim, T., Kim, J., Tae, Y., Park, C., Choi, J. H., & Choo, J. (2021, May). Reversible instance normalization for accurate time-series forecasting against distribution shift. In International conference on learning representations.
36. Abdul Samad, S., Sushma, R., Bharathi Mohan, G., Samuji, P., Repakula, S., & Kothamasu, S. R. (2024, March). Advancing abstractive summarization: Evaluating GPT-2, BART, T5-small, and pegasus models with baseline in ROUGE and BLEU metrics. In International Conference on Innovations in Cybersecurity and Data Science Proceedings of ICICDS (pp. 119-131). Singapore: Springer Nature Singapore.
37. Aksu, T., Liu, C., Saha, A., Tan, S., Xiong, C., & Sahoo, D. (2024). Xforecast: Evaluating natural language explanations for time series forecasting. arXiv preprint arXiv:2410.14180.
38. Alvarez-Melis, D., & Jaakkola, T. S. (2018). On the robustness of interpretability methods. arXiv preprint arXiv:1806.08049

Стаття прийнята 10.02.2026