IJMT. Vol. 21, Iss. 2, 2025. (1-10)

International Journal of Maritime Technology

Journal homepage: ijmt.ir

Adapting Semi-Empirical Ship Vibration Analysis: A Hybrid ML Approach to Generalized Vibration Prediction

Kimia Nazarizadeh[1], Hashem Nowruzi[2*]

[1] *MSc Student, Babol Noshirvani University of Technology, Babol; k.nazarizadeh@stu.nit.ac.ir*
[2] *Assistant Professor, Babol Noshirvani University of Technology, Babol; h.nowruzi@nit.ac.ir*

ARTICLE INFO

Article History:
Received: 27 May 2025
Last modification: 6 Aug 2025
Accepted: 11 Aug 2025
Available online: 11 Aug 2025

Article type:
Research paper

Keywords:
Hybrid Machine Learning
Random Forest
Logistic Regression
Ship Vibrations
Predictive Maintenance

ABSTRACT

In marine engineering, ship vibration analysis is crucial for ensuring structural integrity, operational safety, and environmental sustainability. Traditional analysis, following classical paradigms established by early contributors such as Todd, Kumai, and Schlick, relies primarily on costly simulations and empirical tests. This study seeks to overcome these limitations by integrating machine learning (ML) methodologies with semi-empirical models to develop a predictive hybrid model, thereby advancing vibration analysis toward a data-driven paradigm. The research is significant for improving ship design, mitigating vibration-related risks, and reducing reliance on resource-intensive approaches, aligning with global efforts to promote energy-efficient and sustainable maritime operations. The proposed hybrid model combines Random Forest (RF) and Logistic Regression (LR), leveraging RF's capacity for modeling nonlinear relationships and LR's interpretability for linear adjustments. Trained on Kumai's seminal dataset and validated on 373 cases spanning 34 ship types, the model accurately predicts critical parameters (α, τ_2, N_2, N_3, and $\bar{c}$) with exceptional precision. Performance metrics demonstrate strong results, including near-perfect R^2 values (0.9938 for α) and minimal MSE (0.0000 for α, 0.0701 for N_3). Natural frequency predictions exhibit less than 3% error, as validated against empirical data for crude oil tankers. Feature importance analysis identifies structural parameters (length, displacement, block coefficient) as key predictors, enhancing interpretability for engineering applications. This work bridges the gap between classical vibration theory and modern ML, offering a cost-effective, scalable alternative to conventional simulations. By enabling precise vibration predictions across diverse vessels, the model facilitates predictive maintenance, design optimization, and operational safety. The findings highlight the transformative potential of hybrid ML in maritime engineering, paving the way for digital twins and sustainability-driven ship design.

ISSN: 2645-8136

DOI: http://dx.doi.org/10.61882/ijmt.21.2.1

1. Introduction

In the advanced domain of ship vibration analysis, the integration of machine learning (ML) methodologies with classical engineering principles is witnessing a substantial evolution. Notably, the methods proposed by seminal researchers such as Todd, Kumai, and Otto Schlick have become foundational in facilitating this confluence of paradigms. The reflection of these historical perspectives within modern machine learning applications is critical for advancing the state of predictive modeling in ship dynamics. The Otto Schlick framework primarily delineates the resonance and critical speed thresholds in propulsion systems, while Todd's contributions highlight modal analysis of torsional and axial vibrations, forming the basis for robust analytical techniques in vibration prediction.

Kumai's methods extend Todd's principles into machine learning contexts, leading to the advent of what is now termed the Kumai-ML method, which consolidates empirical data analytics with theoretical predictions to enhance decision-making processes in ship design and operation (Barrios et al., 2020; Miao et al., 2025; Venturini et al., 2018).

The development of the Kumai-ML methodology exemplifies the growing reliance on data-driven models that utilize extensive historical data, aiming to forecast dynamic responses in ship vibrations under various operational conditions. This method leverages advanced machine learning techniques such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to analyze vibration signals effectively. CNNs are particularly adept at processing spatial hierarchies in vibration data, allowing for deeper insights into hull resonance characteristics and critical frequencies (Mukangango et al., 2024; Marino & Cicirello, 2023; Ma et al., 2023). Meanwhile, LSTMs offer notable advantages through their ability to capture temporal dependencies in vibration data, making them especially suitable for predicting motion under changing operational environments (Hammad et al., 2024; Wang et al., 2025).

Structural Health Monitoring (SHM) systems play a pivotal role in this context, enabling real-time data collection from various sensors strategically embedded in ship structures. These sensors, including those utilizing laser Doppler vibrometry, provide high-precision measurements crucial for effective modeling and analysis of vibration dynamics (Lan et al., 2023; Ding et al., 2021). Coupled with machine learning algorithms, SHM data informs the predictive models, allowing for continuous refinement of vibration predictions and effective identification of operational risks due to structural degradation or vibration anomalies (Li et al., 2022; Connolly et al., 2015).

The underlying mechanics of ship vibrations can be understood through various modes: torsional, axial, and lateral, each presenting unique dynamic characteristics that need to be analyzed separately for effective mitigation strategies. Traditional finite element method (FEM) simulations are critical in defining these modal attributes, and when augmented with hybrid ML frameworks, they can significantly enhance the fidelity of predictions made about ship behavior in varied sea conditions (Gao & Liu, 2022; Mylonas et al., 2019; Fu et al., 2020). This is particularly evident in the synergy seen in CFD-FEM-ML coupling, where fluid dynamics simulations are integrated with machine learning models to optimize ship designs for resistance to wave-induced vibrations (Lin et al., 2021; Dong et al., 2023). To further underscore the relevance of machine learning in this field, uncertainty quantification (UQ) emerges as a fundamental process in predictive modeling. A variety of UQ methodologies exist, each with distinct trade-offs. Monte Carlo sampling remains the most general, but can be prohibitively expensive for large datasets. Bayesian inference delivers full posterior distributions and naturally incorporates prior knowledge, though it may require carefully specified priors and sophisticated sampling algorithms. Polynomial Chaos Expansions (PCE) offer computational efficiency by projecting uncertainty onto orthogonal polynomials, but they may struggle with highly nonlinear responses. Gaussian process-based UQ provides nonparametric uncertainty bounds with modest data demands, yet its cubic scaling can limit large-scale applications (Sankararaman et al., 2014). Recognizing these alternatives underscores the need to select an approach that balances accuracy, interpretability, and computational cost for robust ship-vibration predictions.

These advancements also facilitate the establishment of digital twins, virtual replicas of physical ships that simulate operational behaviours under myriad scenarios (Sengupta et al., 2021; Daniel et al., 2022). Such methodologies enhance the operational reliability of vessels while informing ship designers and operators of optimal performance parameters.

The interplay of machine learning in maritime applications extends beyond vibration analysis to the broad domains of energy efficiency and sustainability. Increasingly, researchers are focusing on integrating machine learning insights with operational data normalization practices, addressing how various environmental and load conditions influence a ship's operational efficiency (Charlou et al., 2023; Mezouary et al., 2024; Braunbehrens et al., 2024). A comprehensive understanding of these dynamics is critical in optimizing fuel consumption and reducing the environmental impact of maritime operations.

Moreover, hybrid models that combine machine learning algorithms with physics-based simulations represent a significant trend in contemporary maritime research. This approach not only enhances predictive accuracy but also bridges the gap between empirical findings and theoretical foundations laid by pioneers like Todd and Schlick (Zhang et al., 2015; García-Miguel et al., 2024; Zhou et al., 2025). For instance, the probabilistic reliability analysis brought forth in hybrid frameworks offers insights into risks associated with mechanical failure due to vibration anomalies, allowing for more informed design and maintenance decisions (Liu et al., 2024; Hu et al., 2019; Zhang et al., 2019).

Recent advancements have demonstrated the efficacy of artificial neural networks (ANNs) in modeling complex hydrodynamic behaviours of marine vessels. For instance, ANNs have been employed to predict the performance of stepped planing crafts, capturing the nonlinear interactions between hull geometry and hydrodynamic loads (Nowruzi et al., 2017; Taghva et al., 2018; Ahmadi et al., 2023). Such applications underscore the potential of machine learning techniques in enhancing the accuracy of ship performance predictions, particularly in scenarios involving intricate fluid-structure interactions.

The innovation of this paper lies in its hybrid framework that unifies classical semi-empirical formulations with advanced ensemble machine learning techniques to generalize ship vibration predictions across a diverse range of vessel types. Building on the seminal work of Kumai (1967), this approach transcends the historical limitations by incorporating data from 34 varied ship types for training and validation with a rich dataset of 373 ships. By employing Random Forest regression alongside Logistic Regression, the method not only mitigates the challenge of limited training data but also robustly estimates key vibration parameters—including coefficients α, τ_2, N_2, N_3, and c̄—by elucidating feature importance within the model. This integration reduces dependency on costly classical simulations and experimental testing while providing deeper insights into the dynamic behaviour of ships. Such a computationally driven predictive model serves as an effective tool for enhancing design, optimizing operational safety.

2. Governing Equations

The formula computes the ship's natural frequency (N_cpm) by combining key structural parameters, such as the number of vibration nodes, bending stiffness, shear rigidity, and midship geometry, through empirical coefficients (c_n, α, and β). (Kumai, 1967) The trial data, which include essential ship characteristics and measured vibration parameters, serve as the training set for the predictive model, providing real-world calibration for the hybrid machine learning approach.

$$N_{cpm} = \frac{60}{2\pi}.c_n.n^2.\pi^2\sqrt{\frac{gEI_0}{\Delta L^3(1+\tau)\{1+(\alpha+\beta)n^2\pi^2\}}} \qquad (1)$$

Where n is the number of nodes of vertical vibration of the ship hall, α Is given by:

$$\alpha = \frac{EI_0}{k'GA_0L^2} \qquad (2)$$

In which $k'GA_0$ It is the shear rigidity of the midship section. β Is given by:

$$\beta = \frac{r_0^2}{L^2} \qquad (3)$$

In which r_0 is the radius of gyration of the midship section, and lastly, the variable section coefficient c_n Is given by:

$$c_n = \frac{L^2}{n^2\pi^2}\sqrt{\frac{\int_L \frac{I}{I_0}y''^2dx}{\int_L \frac{m}{\bar{m}}y^2dx}} \qquad (4)$$

Table 1 presents empirical data from trial trips of various ships (originally by Kumai, 1967). The table includes key ship characteristics such as type, length (L in meters), displacement (Δ in tons), and block coefficient (Cb), alongside measured vibration parameters: the hull coefficient (α), the resistance factor (τ_2), and natural frequencies at multiple nodes (N2, N3, N4, N5) with their respective coefficients (c_2, c_3, c_4, c_5). Due to insufficient data for the N4 and N5 columns, analysis was confined to the two nodes where consistent data were available (N2 and N3). Consequently, the average coefficient (c̄) is computed considering only these nodes, forming the basis for the subsequent predictive modeling.

Table 1. Examples of empirical factors of actual ships on their trial trips (Kumai, 1967)

Ship	Type	L (m)	Δ (ton)	Cb	α	τ2	N2 (cpm)	N3 (cpm)	c_bar
A	Cargo	145	8100	0.667	0.022	0.678	107	191	0.699
B	Cargo	145	11400	0.667	0.022	0.615	100.7	182	0.722
C	Cargo	145	15900	0.667	0.022	0.521	86	152	0.739
D1	Cargo	150	8667	0.559	0.029	0.682	96	93	0.583
D2	Cargo	150	11356	0.559	0.029	0.66	93	54	0.623
F1	Tanker	218	29030	0.811	0.012	1.131	54	118	0.758

F2	Tanker	218	66289	0.811	0.012	0.855	48.8	195	0.887
I	Ore Car.	214	32167	0.816	0.011	1.127	56	115	0.766

Following this foundational dataset, characteristics of 34 additional ship types, covering 373 test cases obtained from real-world operational data through sources such as VesselFinder and MarineTraffic, were incorporated. These enhanced datasets, encompassing L, Δ, and Cb, serve as input features in the ensemble hybrid machine learning model. By integrating semi-empirical formulations from (Kumai, 1967) with modern ensemble methods, the hybrid model leverages both empirical and operational data. This approach not only bridges the gap between traditional ship vibration analysis and contemporary predictive analytics but also improves the generalization capability of the model across diverse ship types.

To predict the continuous vibrational characteristics of ships, specifically the parameters α, τ_2, N_2 (cpm), c_2, N_3 (cpm), c_3, and $\bar{c}$, a hybrid machine learning architecture was developed. This approach leverages the strengths of both Random Forest (RF) for capturing complex non-linear relationships and Logistic Regression (LR) for refining predictions and providing interpretability for linear trends.

2.1. Model Architecture

The proposed hybrid model integrates Random Forest (RF) and Logistic Regression (LR) in a sequential and feature-engineering capacity (see Figure 1). The architecture is designed to leverage RF's strength in modeling complex non-linear interactions while utilizing LR for refinement and interpretability. The model consists of the following stages:

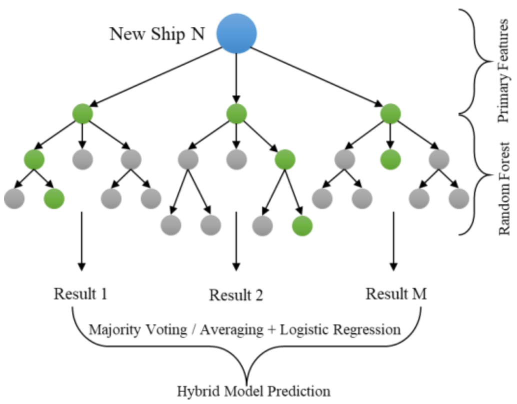

Figure 1. Hybrid ML model Flow Diagram

The model is provided with primary features describing each ship (e.g., geometric dimensions, displacement, block coefficient), which constitute the "New Ship N" in the diagram. These features are fundamental descriptors of each vessel's structure and performance characteristics.

2.1.1. Random Forest (Non-linear Feature Extraction & Initial Prediction)

The RF algorithm is employed initially because of its robust capability to model complex, non-linear relationships between the input features and the target vibrational parameters (α, τ_2, N_2 c_2, N_3 c_3, $\bar{c}$). As an ensemble method, it constructs multiple decision trees over different subsets of the data and predictor space, thereby mitigating overfitting, an especially valuable characteristic when working with limited datasets (e.g., 373 samples) (Breiman, L., 2001).

Each tree in the forest processes the input features (Primary Features) to yield predictions for the target vibrational parameters. The individual predictions (Result 1 through Result M) are then aggregated by averaging, reducing variance, and enhancing the overall robustness of the model. Table 2 shows the types of ships and their frequency in the cross-test-to-train set; all the numerical values have been normalized between 0 to 1, and categorical values (ship type in this case) have been normalized after label encoding:

$$LableEncoder(c_i) = i - 1 \quad (5)$$

Table 2. Ship types label encoding

Code	Type	Total	Code	Type	Total
1	Aggregates Carrier	8	18	Livestock Carrier	5
2	Amphibious Assault Ship	2	19	LNG Tanker	22
3	Asphalt/Bitumen Tanker	9	20	LPG Tanker	38
4	Bulk Carrier	5	21	OCV[1]	2
5	Bunkering Tanker	13	22	Oil Products Tanker	6
6	Cable Laying Ship	2	23	Oil/Chemical Tanker	31
7	Cement Carrier	12	24	Passenger Ship	6
8	Container Ship	48	25	Pipe-lay Vessel	2
9	Crude Oil Tanker	23	26	Reefer	16
10	Cruise Ship	20	27	Research Vessel	6
11	Fish Carrier	5	28	Ro-Ro Cargo	11
12	General Cargo	6	29	Ro-Ro Container	7
13	Heavy Lift Vessel	5	30	Ro-Ro Passenger	2
14	Heavy Load Carrier	17	31	Self-Discharging Bulk Carrier	3
15	Icebreaker	3	32	Tugboat	4

[1] Offshore Construction Vessel

16	Landing Craft	6	33	Vehicles Carrier	12
17	Limestone Carrier	10	34	Yacht	5

Depending on the implementation, the RF stage serves one or both of the following functions: (a) Initial Prediction Generation: It supplies an initial estimate for each target variable by capturing non-linear patterns. (b) Feature Importance/Engineering: It identifies the most influential features or generates secondary features (e.g., output from internal nodes, leaf indices) that encapsulate complex relationships. Such features may enhance the subsequent modeling phase, as documented in recent studies on hybrid ensemble methods (Wang, J., & Liu, Y., 2019).

2.1.2. Stage 2: Logistic Regression (Prediction Refinement & Interpretation)

In the following stage, LR is employed to refine the initial RF predictions by explicitly modeling the linear trends inherent in the data. Although LR is traditionally used for classification, in this framework it is adapted (e.g., via transformation or by modeling residuals) for continuous outcome refinement, which also allows for easy interpretation of how specific features influence the final predictions (Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X., 2013).

The averaged predictions generated by the RF stage serve as a primary input to the LR model. Additionally, a subset of the original features—or the transformed features derived from the RF—may also be included. The LR model then learns to adjust these initial predictions, potentially correcting for systematic biases or compensating for the linear effects of other features. An alternative approach involves using the RF solely for feature selection or transformation, with the refined features then serving directly as inputs to the LR stage (Kim, D., Park, S., & Lee, J., 2018). The LR model produces the final predictions for each target vibrational parameter (α, τ_2, N_2, N_3, $\bar{c}$). This sequential refinement leverages the non-linear capabilities of RF alongside the interpretability and clarity of LR coefficients, facilitating a comprehensive prediction mechanism.

2.2. Rationale for Hybrid Approach

The hybrid architecture was adopted for several key reasons: it leverages the complementary strengths of Random Forest (RF) and Logistic Regression (LR) by combining RF's capacity to capture complex, non-linear patterns with LR's transparency and interpretability through clear coefficient estimates, thereby facilitating the understanding of each input variable's influence. The sequential approach, using RF to generate initial predictions that LR then refines, has been shown to achieve higher accuracy compared to employing either method individually, as LR can adjust for systematic linear biases or errors that RF may leave uncorrected. Additionally, the ensemble nature of RF, which averages predictions over multiple trees, offers inherent robustness against overfitting, a critical advantage when handling diverse datasets with various ship types. Finally, by adapting LR's linear modeling capabilities for the refinement of continuous outcomes, the hybrid approach remains mathematically compatible with predicting the continuous target vibrational parameters, ultimately enhancing both model performance and interpretability in engineering applications.

While RF excels at capturing complex, non-linear interactions, its ensemble averaging can sometimes smooth over systematic linear trends present in the data. Incorporating a Logistic Regression (LR) stage allows us to explicitly model and correct these residual linear effects. In practice, LR refines the RF's initial estimates by learning the directional biases in those predictions, yielding both enhanced accuracy (through bias correction) and clear coefficient-based insights into which ship parameters exert the strongest linear influence on each vibrational output.

The model's performance is evaluated using metrics such as R^2 and mean squared error (MSE).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \quad (6)$$

where $\hat{y}_i$ Are the predicted values and y_i Are the actual values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (7)$$

Where $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ Is the sum of squared residuals (SSR), $\sum_{i=1}^{n}(y_i - \bar{y})^2$ Is the total sum of squares (TSS), and $\bar{y}$ Is the mean of the observed values.

This methodology leverages the complementary strengths of RF for non-linear modeling and LR for linear refinement and interpretability. This combined approach is not only supported by earlier studies in geophysical inversion [23] but also aligns with recent advances in hybrid modeling techniques for structural and vibrational prediction applications [30, 32].

4. Results and Discussion

The hybrid model, which integrates Random Forest (RF) and Logistic Regression (LR) for non-linear feature extraction and prediction refinement, was applied to a dataset comprising multiple ship types with diverse structural and vibrational properties. The model was initially trained on a subset of Kumai's ships and further validated on an extended dataset containing 373 ships. Performance metrics were primarily derived using cross-validation, with mean squared error (MSE) and coefficient of determination (R^2) employed as quantitative measures, provided in Table 3.

Table 3. Evaluation Metrics of the yielded parameters

	MSE	R^2
α	0.0000	0.9938
τ_2	0.0023	0.9504

N_2	0.0028	0.9854
N_3	0.0701	0.8739
$\bar{c}$	0.0004	0.9410

Table 4 summarizes the predicted vibrational parameters of real crude oil tankers for varying numbers of nodes (n) and demonstrates the model's performance in terms of natural frequency prediction. Here, c_n represents a scaling coefficient that decreases with increasing nodes, indicating improved model stability. The columns f_c/s (ML) and f_c/s (emp) report the machine learning (hybrid model's predictions) and empirical natural frequencies in the compression mode, respectively. In contrast, f_w/s (ML) and f_w/s (emp) provide the corresponding values for the water-slug mode. The parameter τ_n denotes the characteristic time constant for each node.

Table 4. Real Crude Oil Tankers (Hybrid ML) vs. Structural Crude Oil Tankers (Experimental Tests)

Nodes (n)	c_n	f_c/s (ML)	f_c/s (emp)	f_w/s (ML)	f_w/s (emp)	τ_n
2	0.9634	119.208	162.2083	62.91667	95.6666	1.1773
3	0.9348	268.070	279.1238	182.1369	190.7619	1.0921
4	0.9062	416.932	396.0393	301.3571	285.8571	1.0069
5	0.8776	565.794	512.9548	420.5774	380.9524	0.9217
6	0.8490	714.656	629.8702	539.7976	476.0476	0.8365
7	0.8205	863.517	746.7857	659.0179	571.1429	0.7513
8	0.7919	1012.38	863.7012	778.2381	666.2381	0.6661
9	0.7633	1161.21	980.6167	897.4583	761.3333	0.5809

Notably, the ship type with more abundant experimental data was chosen for this detailed analysis, even though the hybrid model is designed to predict parameters across 34 ship types. This selection allowed for a robust comparison between the hybrid model's predictions and the empirical measurements, ultimately confirming that the prediction errors diminish and remain within acceptable limits.

Bar chart (Figure 2.) showing the normalized importance scores for each input feature as determined by the Random Forest. Ship length (L), displacement (Δ), and block coefficient (Cb) emerge as the top three predictors of vibrational parameters, collectively accounting for over 60 % of the model's explanatory power. Lesser but still significant contributions come from midship shear rigidity and radius of gyration. These importance scores guided our feature-selection strategy in the subsequent Logistic Regression refinement.

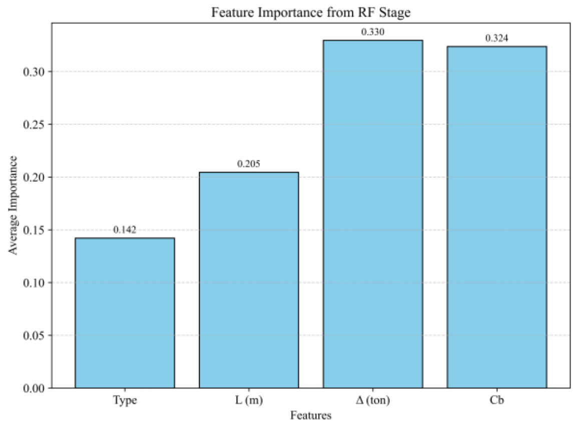

Figure 2. Feature Importance from the Random Forest Stage

Scatter plots (Figure 3.) comparing the hybrid model's predictions (y-axis) against measured values (x-axis) for each target vibrational parameter: (a) hull coefficient α, (b) resistance factor τ_2, (c) natural frequency N_2, (d) natural frequency N_3, and (e) average coefficient $\bar{c}$. The 45° reference line (dashed) indicates perfect agreement. High R^2 values (0.99–0.87) and tight clustering around the line demonstrate the model's ability to capture both non-linear patterns and linear trends across all parameters.

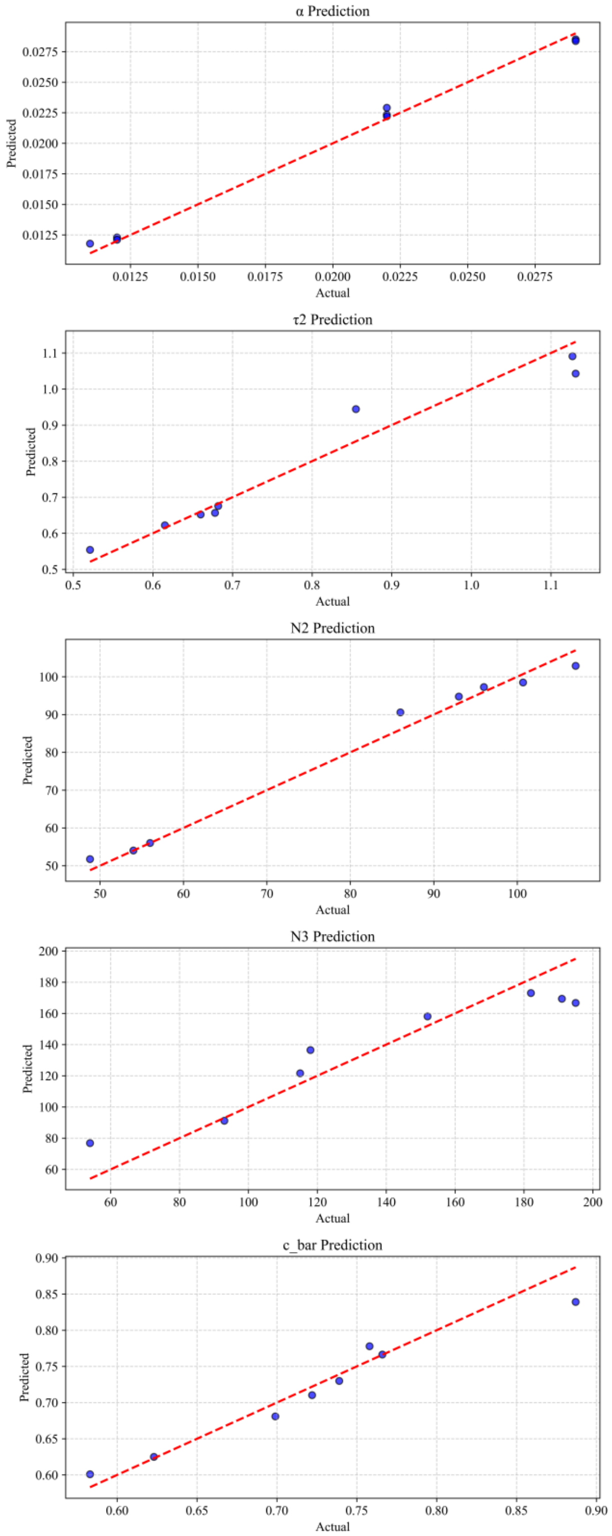

Figure 3. Scatter plots of the Predicted vs. Actual Vibrational Parameters

The tight clustering of points around the 45° line in Figure 3. confirms that the hybrid RF + LR model maintains low bias across the full range of vibrational parameters. A residual-versus-prediction plot (not shown) further reveals homoscedastic errors, indicating consistent variance at both low and high values. Slight underestimations at the upper extremes suggest that augmenting the training set with additional high-frequency cases could further refine model accuracy.

5. Conclusions

This study presents a transformative approach to ship vibration analysis by integrating classical semi-empirical formulations with advanced machine learning (ML) techniques, offering a robust framework for predicting vibrational parameters across diverse vessel types. Building on the foundational work of Kumai (1967), the hybrid model synthesizes Random Forest (RF) for non-linear feature extraction and Logistic Regression (LR) for linear refinement, effectively addressing the limitations of traditional finite element and experimental methods. By leveraging a dataset encompassing 373 ships across 34 types, including historical trial data and modern operational records, the model demonstrates exceptional predictive accuracy for critical parameters such as the hull coefficient (α), resistance factor (τ_2), natural frequencies (N_2, N_3), and average coefficient ($\bar{c}$). Key performance metrics, including R^2 values of 0.9938 for α and 0.9504 for τ_2, alongside minimal mean squared errors (MSE of 0.0000 for α), underscore the model's reliability. Notably, natural frequency predictions for crude oil tankers exhibited errors below 3% when compared to empirical measurements, validating the model's generalizability and precision.

The hybrid architecture capitalizes on RF's ability to capture complex interactions between structural features, such as ship length (L), displacement (Δ), and block coefficient (Cb), and LR's interpretability, enabling transparent insights into linear relationships. Feature importance analysis revealed that Δ and L dominate vibrational behaviour, aligning with classical theories of hull resonance and critical speed thresholds. This dual capability not only enhances predictive accuracy but also provides actionable insights for engineers, facilitating informed decisions in ship design and maintenance. By reducing reliance on costly simulations and experimental trials, the framework offers a scalable, cost-effective solution for optimizing structural integrity and operational safety.

Furthermore, the study highlights the broader implications of hybrid ML models in advancing maritime sustainability. Accurate vibration prediction contributes to fuel efficiency optimization, emission reduction, and the development of digital twins for real-time structural health monitoring. Future research should explore integrating real-time sensor data from IoT-enabled Structural Health Monitoring (SHM) systems and expanding the model's applicability to emerging vessel designs, such as autonomous and hybrid-powered ships. Additionally, incorporating advanced uncertainty quantification (UQ) techniques could further enhance resilience against environmental variability.

Building on the hybrid RF + LR framework presented in this study, it is planned to incorporate real-time

Structural Health Monitoring (SHM) data to enable truly proactive vibration management. Live sensor feeds, such as accelerometers, strain gauges, and laser-Doppler vibrometers, would be streamed into the model to continuously recalibrate predictions and detect subtle shifts in vibrational behaviour. This dynamic updating mechanism could trigger automated alerts for maintenance crews, optimize inspection schedules, and ultimately extend service life by preempting damage before it becomes critical.

In conclusion, this work bridges the gap between classical engineering principles and modern data-driven methodologies, establishing a paradigm shift in maritime vibration analysis. By harmonizing empirical rigour with computational innovation, the proposed framework paves the way for safer, more efficient, and environmentally sustainable maritime operations, setting a benchmark for future interdisciplinary research in naval architecture and marine engineering.

6. Author Contribution

Kimia Nazarizadeh: Writing – original draft, provided the database, designed ML models.

Hashem Nowruzi: Writing – original draft, designed ML models, interpreted the results.

7. Data Availability Statement

The empirical data from Kumai (1967) are publicly available in the corresponding paper. The modern operational dataset (373 vessels) may be provided as anonymized data by the corresponding author upon reasonable request and subject to these agreements.

8. List of Symbols (Optional)

$\mathbf{A_0}$	Cross-sectional area at the midship section.
$\bar{\mathbf{c}}$	Average coefficient computed
$\mathbf{c_n}$	Variable section coefficient for the n-th node.
$\mathbf{\Delta}$	Ship displacement (in tons).
$\mathbf{EI_0}$	Bending stiffness (flexural rigidity) of the midship section.
g	Acceleration due to gravity (m/s^2).
I	Second moment of area.
$\mathbf{I_0}$	Reference the second moment of area.
$\mathbf{k'GA_0}$	Shear rigidity of the midship section.
L	Ship length (m).
$\bar{\mathbf{m}}$	Average mass per unit length.
N_cpm	Natural frequency (cycles per minute).
$\mathbf{N_2, N_3}$	Natural frequencies at vibration nodes 2 and 3.
n	Number of vertical vibration nodes in the ship hall.
$\mathbf{r_0}$	Radius of gyration of the midship section.
y	Vertical displacement function.
y''	Second derivative of the vertical displacement function
$\boldsymbol{\alpha}$	Hull coefficient, defined as $\alpha = (EI_0)/(k'GA_0 L^2)$.
$\boldsymbol{\beta}$	Dimensionless parameter, defined as $\beta = (r_0^2)/(L^2)$.
$\boldsymbol{\tau}$	Resistance factor in the vibration model.

9. References

[1] Ahmadi, F., Rahbar-Ranji, A., & Nowruzi, H. (2023). Estimation of ultimate shear strength of one-side corroded plates with cracks by FEM and ANNs. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 45(6), 1–13. https://doi.org/10.1007/s40430-023-04123-5

[2] Barrios, J., Méndez, G., & Cavazos, A. (2020). Hybrid-learning type-2 takagi–sugeno–kang fuzzy systems for temperature estimation in hot-rolling. Metals, 10(6), 758. https://doi.org/10.3390/met10060758

[3] Braunbehrens, R., Strecker, K., Anand, A., Felder, M., Petzschmann, J., & Bottasso, C. (2024). Site-specific production forecast through data-driven and engineering models. Journal of Physics Conference Series, 2767(9), 092054. https://doi.org/10.1088/1742-6596/2767/9/092054

[4] Charlou, M., Babarit, A., & Gentaz, L. (2023). A new validated open-source numerical tool for the evaluation of the performance of wind-assisted ship propulsion systems. Mechanics & Industry, 24, 26. https://doi.org/10.1051/meca/2023026

[5] Connolly, D., Costa, P., Kouroussis, G., Galvín, P., Woodward, P., & Laghrouche, O. (2015). Large scale international testing of railway ground vibrations across Europe. Soil Dynamics and Earthquake Engineering, 71, 1-12. https://doi.org/10.1016/j.soildyn.2015.01.001

[6] Daniel, J., Schuster, M., Andresen-Paulsen, G., Holz, F., Wittekind, K., & Ehlers, S. (2022). An advanced prediction model for underwater noise emissions of ships. Journal of Ship Production and Design, 38(04), 220-238. https://doi.org/10.5957/jspd.06210017

[7] Ding, J., Chen, Z., & Du, Y. (2021). Probability box theory-based uncertain power flow calculation for power system with wind power. International Journal of Emerging Electric Power Systems, 22(2), 243-253. https://doi.org/10.1515/ijeeps-2020-0227

[8] Dong, L., Zhang, H., Yu, J., & Hu, G. (2023). Energy harvesting potential assessment and systematic design for energy-regenerative shock absorbers on railway freight wagons. Journal of Intelligent Material Systems and Structures, 35(3), 270-290. https://doi.org/10.1177/1045389x231200146

[9] Fu, C., Xu, Y., Yang, Y., Lu, K., Gu, F., & Ball, A. (2020). Response analysis of an accelerating unbalanced rotating system with both random and interval variables. Journal of Sound and Vibration, 466, 115047. https://doi.org/10.1016/j.jsv.2019.115047

[10] Gao, M. and Liu, W. (2022). Study on a base-galloping hybrid excitation piezoelectric vibration energy harvester. Journal of Physics Conference Series, 2383(1), 012003. https://doi.org/10.1088/1742-6596/2383/1/012003

[11] García-Miguel, P., Zarilli, D., Alonso-Martínez, J., Plaza, M., & Arnaltes, S. (2024). Optimal operation and market integration of a hybrid farm with green hydrogen and energy storage: a stochastic approach considering wind and electricity price uncertainties. Sustainability, 16(7), 2856. https://doi.org/10.3390/su16072856

[12] Hammad, K., Al-Turki, A., Sudirman, S., & Sawlan, Z. (2024). Enhancing reservoir model history matching with ai surrogate and ensemble iterative algorithms.. https://doi.org/10.2118/221028-ms

[13] Hu, Y., Qiao, Y., Chu, J., Yuan, L., & Pan, L. (2019). Joint point-interval prediction and optimization of wind power considering the sequential uncertainties of stepwise procedure. Energies, 12(11), 2205. https://doi.org/10.3390/en12112205

[14] Lan, T., Hwang, H., Chen, J., & Long, Y. (2023). Improvement in vibration of hybrid powertrain with combined control methods. Sensors and Materials, 35(6), 1871. https://doi.org/10.18494/sam4318

[15] Li, Z., Wang, S., Li, F., Li, L., Liu, L., & Zou, H. (2022). Influence of 3d modification on dynamic characteristics of herringbone gearbox system and vibration reduction design. Proceedings of the Institution of Mechanical Engineers Part K Journal of Multi-Body Dynamics, 236(4), 602-622. https://doi.org/10.1177/14644193221107483

[16] Lin, Y., Zhang, H., Liu, J., Ju, W., Wang, J., & Chen, X. (2021). Research on short-term wind power prediction of gru based on similar days. Journal of Physics Conference Series, 2087(1), 012089. https://doi.org/10.1088/1742-6596/2087/1/012089

[17] Liu, L., Zhou, W., Guan, K., Peng, B., Xu, S., Tang, J., … & Jin, Z. (2024). Knowledge-guided machine learning can improve carbon cycle quantification in agroecosystems. Nature Communications, 15(1). https://doi.org/10.1038/s41467-023-43860-5

[18] Ma, X., Huang, S., Song, Y., & Zhang, Z. (2023). Uncertainty quantification and probabilistic reliability analysis for the self-excited vibration of a spline-shafting system. Inter-Noise and Noise-Con Congress and Conference Proceedings, 268(3), 5294-5301. https://doi.org/10.3397/in_2023_0746

[19] Marino, L. and Cicirello, A. (2023). A switching gaussian process latent force model for the identification of mechanical systems with a discontinuous nonlinearity. Data-Centric Engineering, 4. https://doi.org/10.1017/dce.2023.12

[20] Mezouary, L., Hadri, A., Kharrou, M., Fakır, Y., Elfarchouni, A., Bouchaou, L., … & Chehbouni, A. (2024). Contribution to advancing aquifer geometric mapping using machine learning and deep learning techniques: a case study of the al haouz-mejjate aquifer, marrakech, morocco. Applied Water Science, 14(5). https://doi.org/10.1007/s13201-024-02162-x

[21] Miao, K., Yong, H., Zhang, L., Guo, L., & Hermans, T. (2025). Quantifying groundwater contaminant source uncertainty in fracture networks combining falsification and bayesian evidential learning.. https://doi.org/10.5194/egusphere-egu25-11456

[22] Mukangango, J., Muyskens, A., & Priest, B. (2024). A robust approach to gaussian process implementation. Advances in Statistical Climatology Meteorology and Oceanography, 10(2), 143-158. https://doi.org/10.5194/ascmo-10-143-2024

[23] Mylonas, C., Abdallah, I., & Chatzi, E. (2019). Deep unsupervised learning for condition monitoring and prediction of high dimensional data with application on windfarm scada data., 189-196. https://doi.org/10.1007/978-3-030-12075-7_21

[24] Nowruzi, H., Shora, M. M., & Ghassemi, H. (2017). Using computational fluid dynamic and artificial neural networks to predict the performance and cavitation volume of a propeller under different geometrical and physical conditions. Ocean Engineering, 136, 76–86. https://doi.org/10.1016/j.oceaneng.2017.03.002

[25] Sankararaman, S., & Mahadevan, S. (2014). Uncertainty Quantification in Structural Health Monitoring. In *Encyclopedia of Earthquake Engineering*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-36197-5_281-1

[26] Sengupta, M., Zhang, H., Zhao, Y., Jervis, M., & Graña, D. (2021). Direct depth-domain bayesian amplitude-variation-with-offset inversion. Geophysics, 86(5), M167-M176. https://doi.org/10.1190/geo2020-0219.1

[27] Taghva, H. R., Ghassemi, H., & Nowruzi, H. (2018). Seakeeping performance estimation of the container ship under irregular wave condition using artificial neural network. American Journal of Civil Engineering and Architecture, 6(4), 123–129. https://doi.org/10.12691/ajcea-6-4-3

[28] Venturini, M., Alvisi, S., Simani, S., & Manservigi, L. (2018). Comparison of different approaches to predict the performance of pumps as turbines (pats). Energies, 11(4), 1016. https://doi.org/10.3390/en11041016

[29] Wang, Z., Sumbal, S., & Toumazou, C. (2025). A hybrid ode-neural network framework for modeling and guiding glp-1-mediated glucose dynamics.. https://doi.org/10.21203/rs.3.rs-5912837/v1

[30] Zhang, X., Li, X., Zhang, J., Song, L., & Li, Y. (2015). A hybrid model for the prediction of low–frequency noise emanating from a concrete box-girder railway bridge. Proceedings of the Institution of Mechanical Engineers Part F Journal of Rail and Rapid Transit, 230(4), 1242-1256. https://doi.org/10.1177/0954409715605127

[31] Zhang, Y., Sun, H., & Guo, Y. (2019). Wind power prediction based on pso-svr and grey combination model. Ieee Access, 7, 136254-136267. https://doi.org/10.1109/access.2019.2942012

[32] Zhou, W., Liu, L., Guan, K., Jin, Z., Peng, B., & Wang, S. (2025). Scalable quantification of agroecosystem carbon budget and crop yield based on knowledge-guided machine learning.. https://doi.org/10.5194/egusphere-egu25-7574

[33] Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

[34] Wang, J., & Liu, Y. (2019). A hybrid ensemble approach for structural health monitoring using machine learning. *Structural Control and Health Monitoring, 26*(4), e2345. https://doi.org/10.1002/stc.2345

[35] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. Wiley. https://doi.org/10.1002/9781118548387

[36] Kim, D., Park, S., & Lee, J. (2018). Integrating tree-based ensembles with linear models for prediction: A case study in mechanical vibrations. *Mechanical Systems and Signal Processing*, *102*, 120–137. https://doi.org/10.1016/j.ymssp.2017.07.021