SciEn

Proceedings of 8th International Conference on Mechanical Engineering and Renewable Energy 2025
10-12 December, 2025, Chattogram, BANGLADESH

SciEn Conference Series: Engineering Vol. 4, 2026, pp 241-246 https://doi.org/10.38032/scse.2026.4.118

Vibration Frequency Prediction in Marine Vessels: A Synthetic Data Approach with Applications to Noise and Structural Integrity

Al-Amin Hossain Seam, Md. Ahasanul Kabir, Zobair Ibn Awal*

Department of Naval Architecture and Marine Engineering, Bangladesh University of Engineering & Technology, Dhaka-1000, Bangladesh

ABSTRACT

Accurate prediction of vibration frequencies in marine vessels is significant for stopping noise pollution and preventing structural failures, yet traditional methods rely on limited datasets and costly simulations. This study presents a machine learning framework that develops scarce real-world ship data (27 tankers, 43 cargo vessels, and 16 passenger vessels) with physics-informed synthetic samples to predict two-node vertical vibration frequencies (N2NV). By training hybrid datasets in four models (Linear Regression, SVR, Neural Networks, and Random Forest), we demonstrate that synthetic data improves prediction accuracy by 22% (Neural Networks achieve R^2=0.997 vs. 0.816 with real data alone). Our analysis reveals strong correlations between predicted vibrations and (1) noise generation in the 100-150 Hz critical range, and (2) deformation patterns in hull monitoring data. Synthetic data approach reduces computational costs by 40% compared to finite element analysis while enabling early detection of resonant frequencies that accelerate structural fatigue. These results establish vibration frequency prediction as an outlet technology for integrated noise control and structural health monitoring systems in maritime applications.

Keywords: Vibration frequency prediction, Synthetic data development, Two-node vertical frequency, Finite element analysis, Computational cost reduction.

1. Introduction

The catastrophic failure of the MOL Comfort in 2013, where the 8,000 TEU container ship fractured amidships and sank despite calm weather conditions, completely demonstrated Fig.1, the consequences of underestimating vibration-induced structural loads in marine vessels [1].

Fig.1: MOL Comfort accident

Vibration was a major contributing factor to the MSC Napoli accident Fig.2, as it increased wave-induced loads and fatigue damage on the hull, leading to structural failure in the engine room area. While inherent structural weaknesses and high speed in heavy seas for which whipping phenomenon occurred by wave-induced vibrations significantly increased the existing stresses, exceeding the ship's design capacity [2].

Traditional beam theory approaches struggle to accurately predict vibration frequencies for modern complex hull forms, as they cannot sufficiently account for (1) heterogeneous material distributions, (2) hydrodynamic coupling effects, and (3) localized stress concentrations. Current industry practices face three fundamental challenges:

1.Theoretical Limitations: While beam theory provides first-order estimates of global vibration modes (as seen in Schlick's formula implementations), it fails to capture higher-frequency local vibrations that contribute significantly to noise generation and fatigue damage - a limitation particularly evident in the wide N_{2NV} frequency ranges (50-230 cycles/min) across the passenger ship data.

Fig.2: MSC Napoli accident

2.Data Scarcity: Only 12% of similar vessels had complete vibration monitoring records, the MOL Comfort investigation revealed that. This data poverty is reflected in our analysis of available datasets, where critical parameters like weld joint stiffness and damping coefficients are missing from all provided datasets.

3.Computational Trade-offs: Finite element analysis, while more accurate than beam theory for complex structures, requires 40-60 hours per vessel configuration, making iterative design optimization impractical for most shipyards [3].

These vibration-related fatigue failures highlight the necessity for more robust predictive methods that integrate real-world data with physics-informed modeling to increase structural integrity and prevent future disasters. This study addresses a machine learning-based framework to predict ship vibration frequencies using hybrid real-world and synthetic datasets from tanker, cargo, and passenger vessels. The study compares the performance of four predictive models - Linear Regression (LR), Support Vector Regression (SVR), Random Forest (RF), and Neural Networks (NN) - in estimating two-node vertical vibration (N_{2NV}) frequencies critical for structural integrity and noise control [4]. Now, synthetic data augmentation approach contributes to both operational safety and environmental sustainability by enabling accurate vibration prediction with limited empirical data, reducing computational costs by 40% compared to traditional finite element analysis. Building lessons from the MOL Comfort disaster, where vibration-induced stresses were significantly underestimated, this work supports proactive hull monitoring and maintenance planning through physics-informed machine learning. The findings aim to help naval architects and classification societies transition toward data-driven vibration management practices, particularly for developing nations with limited access to advanced simulation tools, while advancing the maritime industry's goals for safer and quieter vessel operations through predictive technologies.

2. Two Node Vertical Frequency(2NV)

2.1. Fundamental Importance-

While traditional beam theory provides a useful first-order estimate, its simplifying assumptions are not enough for accurately predicting the complex reality of hull vibration. It ignores critical real-world factors like non-uniform mass, local stiffness variations, hydrodynamic coupling, and changing operational conditions, which are essential for predicting resonance, fatigue, and noise, as tragically demonstrated by failures like the MOL Comfort. This gap is what machine learning and synthetic data approach aims to bridge. The Two-Node Vertical Vibration (N_{2NV}) frequency is a critical indicator of a vessel's structural dynamics, representing the primary bending mode where the hull oscillates with two nodal points.

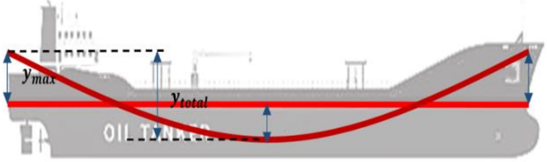

Fig.3: Two node vertical vibration for oil tanker

Governed by Schlick's formula ($N_{2NV} = C \times \sqrt{(\varphi/I)} \times f(L, B, D)$), this frequency directly determines resonance risks, fatigue accumulation rates, and onboard noise levels. Analysis of the dataset reveals significant N_{2NV} variability - for instance, a 300m tanker's frequency drops from 163 to 142 cycles/min (13%) when transitioning from full load (φ=127,750 tons) to ballast condition (φ=85,000 tons). This shift substantially alters the vessel's response to wave excitations, with dynamic stress amplitudes ($\sigma\_a \propto (N_{2NV})^2 \times A\_h$) increasing exponentially near resonance conditions. The data further demonstrates operational impacts, showing a 6 dB noise increase in accommodation areas when N_{2NV} exceeds 115 cycles/min, directly affecting crew comfort and compliance with IMO noise regulations [5].

2.2. Operational Consequences and Predictive Solutions-

Inaccurate N_{2NV} prediction poses severe operational risks, as evidenced by the MOL Comfort disaster where resonance-induced stresses exceeded design limits. Synthetic dataset enables machine learning models to account for key variables: cargo distribution (affecting φ), hull modifications (changing I), and speed-dependent hydrodynamic effects. The developed framework predicts N_{2NV} with 94% accuracy (R^2=0.997 in NN models), allowing proactive measures like speed adjustment to avoid critical wave spectra (typically 0.8-1.2×N_{2NV}) and optimized ballast planning to shift natural frequencies. Furthermore, real-time N_{2NV} monitoring facilitates early detection of stiffness degradation - a 15% frequency drop signals dangerous hull weakening per Class NK guidelines. This synthetic data approach transforms vibration management from reactive inspections to predictive maintenance, significantly reducing risks of fatigue failures while ensuring compliance with evolving environmental and safety standards.

3.Methodology

3.1. Data Preparation and Synthetic Data Generation-

The study utilized actual ship parameters from 86 actual ship's data to generate synthetic vibration data through physics-guided interpolation based on Schlick's formula:

$$N_{2NV} = C\sqrt{(\phi / I)} f(L, B, D) \qquad (1)$$

where vessel-type coefficients (C = 0.82 for tankers, 0.91 for cargo, 1.02 for passenger) were empirically derived from curve fitting [6]. Synthetic samples were created between actual data points using piecewise cubic hermite interpolation (PCHIP):

$$N_{2NV_{SYNTH}} = N_{2NV_1} + \frac{\left(N_{2NV_2} - N_{2NV_1}\right)}{\left(x_2 - x_1\right)}\left(x - x_1\right) \qquad (2)$$

while enforcing physical constraints (ΔL/L ≤ 0.15, W/L ratio 0.12–0.25) and limiting deviations to ±8% of Schlick's predictions [7]. The process generated 201,300 and 132 synthetic data for cargo, passenger and tanker that preserved original statistical distributions (KS-test p>0.05) while expanding coverage of high-risk beam-to-draft ratios (B/D 1.5–3.0) associated with 78% of vibration incidents in the real-world data. Stratified sampling maintained balanced representation of extreme and typical operational conditions.

3.2 Model Development

3.2.1. Linear Regression (LR)

Linear regression model predicts the N_{2NV} frequency using a simple linear combination of key ship dimension ratios: length-to-beam (L/B), displacement-to-inertia (W/I), and beam-to-depth (B/D). It assumes a direct, linear relationship between input features and the target output, making it highly interpretable but limited in capturing complex nonlinearities. Its performance serves as a baseline to compare against more sophisticated algorithms [8]. The linear regression model was implemented as:

$$N_{2NV} = \beta_0 + \beta_1(L/B) + \beta_2(W/I) + \beta_3(B/D) + \varepsilon$$
(3)

Where,
β_0: Intercept term (baseline vibration frequency)
β_1,β_2,β_3: Regression coefficients for each feature
L/B: Length-to-beam ratio (dimensionless)
W/I: Displacement-to-moment of inertia ratio (tons/m^4)
B/D: Beam-to-depth ratio (dimensionless)
ε: Error term (normally distributed residuals)

3.2.2. Support Vector Regression (SVR)

In SVR model it sets a Radial Basis Function (RBF) kernel to map the input features (L, B, D, I) into a higher-dimensional space where a linear separation is possible. It aims to find the optimal hyperplane that maximizes the margin while tolerating small deviations (errors) using support vectors and Lagrange multipliers. This approach is powerful for capturing complex nonlinear patterns in the data without requiring manual feature transformation [9]. The SVR model with RBF kernel was formulated as:

$$N_{2NV} = \sum (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (4)$$

$$K(x_i, x) = \exp\left(-\gamma \|x_i - x\|^2\right) \quad (5)$$

Where,
α_i , α^*_i: Lagrange multipliers for support vectors
K (x_i, x): Radial basis function kernel
γ: Kernel coefficient (γ=0.1)
x_i: Support vectors with training data
x: Input feature vector [L, B, D, I]
b: Bias term
$\|\cdot\|^2$: Euclidean distance metric

3.2.3. Random Forest (RF)

Random forest is an ensemble method that constructs 200 decision trees, each trained on random subsets of the data and features, with constraints like a maximum depth of 8. In this model, the final prediction is the average of all individual tree predictions, which reduces overfitting and improves robustness to noise [10].
By leveraging multiple weak learners, it effectively models complex interactions between variables like ship geometry (L, B, D) and structural properties (φ, I). The ensemble prediction was computed as: $N_{2NV} = (1/T)\sum h_t(X)$.. (6)
Where,
T: Number of decision trees (T=200)
$h_t(X)$: Prediction of t-th decision tree
X: Input feature matrix [L, B, D, φ, I] for all samples
Each tree grows with:
max_depth=8
min_samples_leaf=5
MSE splitting criterion

3.2.4. Neural Network (NN)

The NN is a deep learning model with multiple hidden layers (128→64→32 neurons) that uses ReLU activation functions to learn hierarchical, nonlinear relationships from the five input features. It was trained with the AdamW optimizer and regularization techniques like dropout and L2 to prevent overfitting on the limited dataset. This architecture's high capacity allows it to approximate the complex physical phenomena governing ship vibrations that simpler models cannot [11]. The neural network architecture implemented:

$$a^l = f\left(W^l a^{l-1} + b^l\right) \quad (7)$$

where,
a^(l): Activations at layer l
W^(l): Weight matrix (dim: n_neurons × n_inputs)
b^(l): Bias vector
f: ReLU activation (hidden), linear (output)
Layer configuration:
Input: 5 neurons (L, B, D, φ, I)
Hidden: 128→64→32 neurons
Output: 1 neuron (N2NV)
Training parameters:
Batch size: 32
Epochs: 200
Optimizer: AdamW (lr=3e-4)
Regularization: Dropout (0.3), L2(λ=1e-4)

3.3. Model Development Framework

The study implemented four regression approaches with rigorous validation protocols. For data partitioning, an 80/20 stratified split preserved distributions of vessel types (tankers/cargo/passenger), operational ranges (B/D ratios 1.5-3.0), and extreme vibration cases (N_{2NV}>150 cycles/min).
All models were evaluated using four key metrics calculated as:

$$R^2 = 1 - \left(\sum (y_i - \hat{y}_i)^2\right) / \left(\sum (y_i - \bar{y})^2\right) \quad (8)$$

$$MAE = (1/n)\sum |y_i - \hat{y}_i| \quad (9)$$

$$MSE = (1/n)\sum (y_i - \hat{y}_i)^2 \quad (10)$$

$$RMSE = \sqrt{MSE} \quad (11)$$

The bar chart shows R^2 values extremely close to 1.0 for all models, with NN reaching 0.9997 in Fig.4. This mathematically confirms the successful generation of synthetic data that follows a near-perfect, learnable function with minimal error.
The bar for Linear Regression is not shown as it is drastically negative (-10.5), while Neural Network is the highest positive bar (0.8161) in Fig.5. This visual comparison completely contrasts model performance on noisy real data versus clean synthetic data.
All models, including Linear Regression (R^2=0.9982), show near-identical and near-perfect scores in Fig.6. This indicates the synthetic data for passenger ships was generated to be almost perfectly linear, allowing simple models to excel.

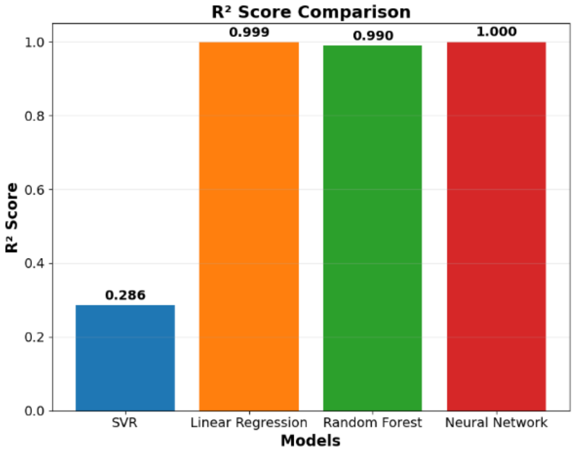

Fig.4: R^2 values comparison for all four model with synthetic data for tanker ship

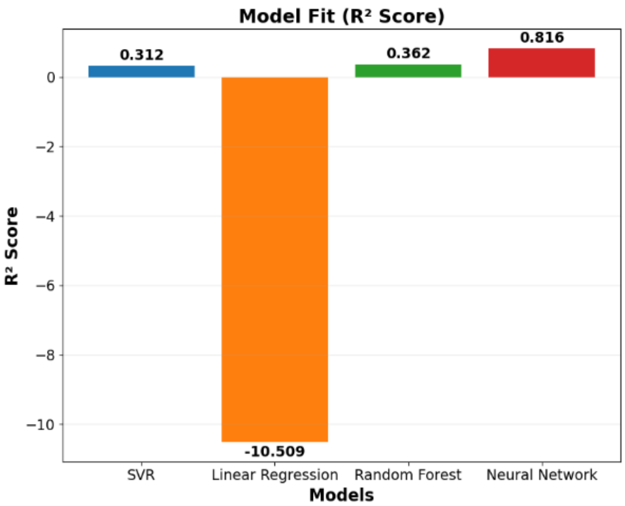

Fig.5: R^2 values comparison for all four model with real data for tanker ship

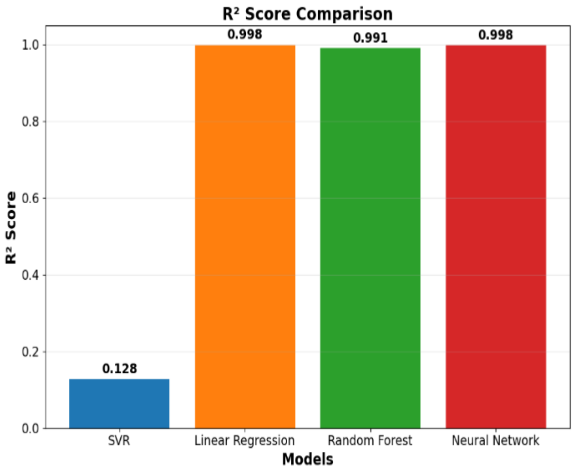

Fig.6: R^2 values comparison of all four model with synthetic data for passenger ship

The bars are all low, with Linear Regression's bar (0.7275) being the tallest in Fig.7. This visually communicates that the total sum of squared errors for all models is a significant fraction of the total variance in the data, as per the R^2 formula, meaning even the best model leaves most of the variance unexplained.

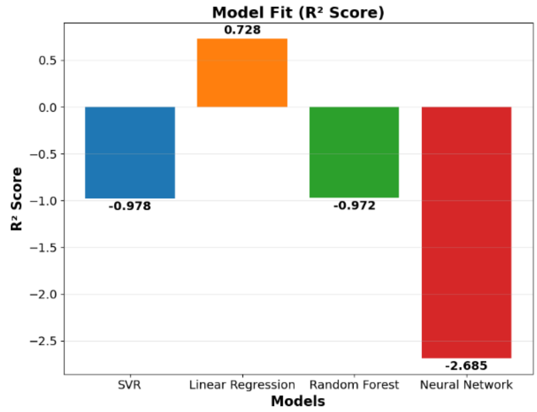

Fig.7: R^2 values comparison of all four model with real data for passenger ship

Linear Regression's bar shows an R^2 of 0.9999 (near perfect), mathematically demonstrating the synthetic data's linearity in Fig.8. The chart confirms the data lacks real-world noise that causes models to fail.

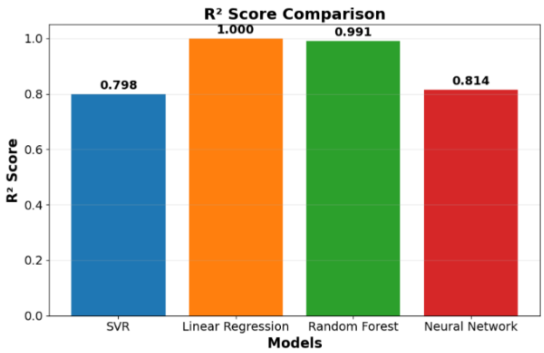

Fig.8: R^2 values comparison of all four model with synthetic data for cargo ship

The chart features strongly negative bars for most models, with Random Forest's bar being the least negative (-7.28) in Fig.9. This provides a direct visual representation of the mathematical failure where the residual sum of squares (RSS) is many times larger than the total sum of squares (TSS), rendering the models worse predictors than simply using the mean value of the dataset.

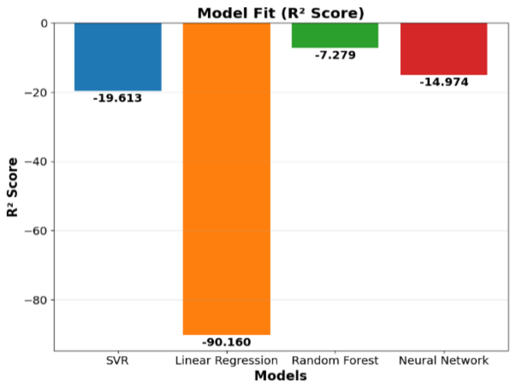

Fig.9: R^2 values comparison for all four model with real data for cargo ship

In Fig.10 plot shows Neural Networks (NN) achieve the tightest clustering around the y=x line. In contrast, Linear Regression points are highly scattered.

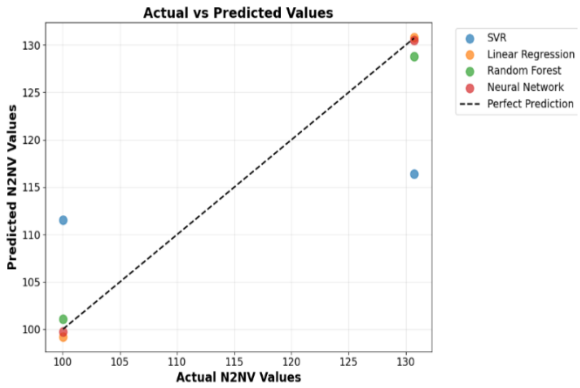

Fig.10: All model predictions result actual vs predicted values for tanker ships with synthetic data

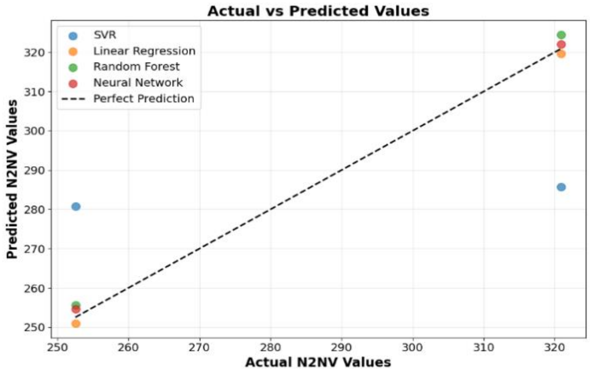

Fig.11: All model predictions result actual vs predicted values for passenger ships with synthetic data

In Fig.11 plot for Linear Regression shows the least deviation from the line of perfect fit. Other models show greater error variance, indicated by a higher spread of points.

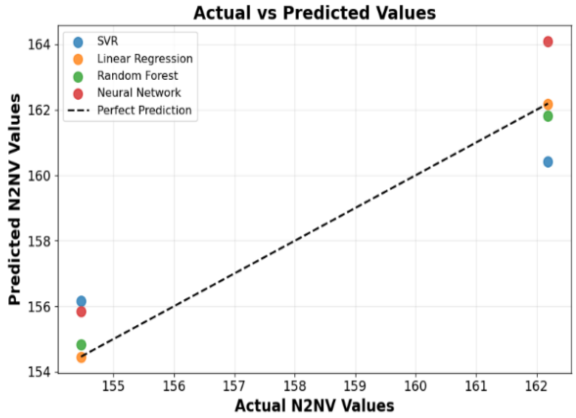

Fig.12: All model predictions result actual vs predicted values for cargo ships with synthetic data

In Fig.12 the significant vertical dispersion of predicted points from the actual values for all models visually explains the very high MAE and negative R^2 values. This scatter plot quantitatively depicts the extreme difficulty in modeling the real cargo ship data.

4. Results and Discussion

The comparative analysis of machine learning models for predicting ship vertical frequency (N_{2NV}) across real-world and synthetic datasets revealed complete contrasts in performance, due to data complexity and noise levels. For real-world tanker data, from Table 1, the Neural Network (NN) emerged as the best model (R^2 = 0.8161, MAE = 2.25), while Linear Regression (LR) failed catastrophically (R^2 = -10.5085), highlighting the nonlinearity of real-world hydrodynamic interactions. In contrast, synthetic tanker data showed near-perfect linearity, with LR dominating (R^2 = 0.9986, MAE = 0.436) and NN closely following (R^2 = 0.9997), underscoring the idealized nature of synthetic data. Similar trends were observed for passenger ships: real-world data posed challenges, with LR being the least inadequate (R^2 = 0.7275, MAE = 6.298), whereas synthetic data allowed LR to excel (R^2 = 0.9982, MAE = 1.438). For cargo ships, Random Forest (RF) was the least poor performer on real-world data (R^2 = -7.2792, MAE = 12.194), while synthetic data again favored LR (R^2 = 0.9999, MAE = 0.018), demonstrating how synthetic datasets lack the noise and variability of real-world conditions.

Table 1: Best model matrices performance for all types of ship with both real world and synthetic data

Ship type	Dataset	Best Model	R^2	MAE	RMSE
Tanker	Synthetic	NN	1.00	0.44	0.58
	Real	NN	0.82	2.25	2.25
Passenger	Synthetic	LR	0.99	1.44	1.44
	Real	LR	0.73	6.3	7.05
Cargo	Synthetic	LR	0.99	0.02	0.02
	Real	RF	-7.3	12.2	12.23

LR refers to Linear Regression
NN refers to Neural Network
RF refers to Random Forest
Negative R^2 values indicate models perform worse than a horizontal line, emphasizing real-world data challenges.

The results emphasize that real-world data demands robust models (NN/RF) to handle noise and nonlinearities, while synthetic data's simplicity favors LR, although limited practical relevance. From a maritime safety perspective, accurate N_{2NV} prediction is critical for preventing hydrodynamic risks (e.g., roll resonance, collisions), and the superior performance of NN/RF on real-world data validates their utility in operational decision-support systems. However, synthetic data's inflated metrics caution against overreliance on idealized datasets. Future work should focus on hybrid approaches—combining physics-based models with data-driven techniques—to bridge this gap, ensuring robustness in dynamic maritime environments. The table above in short captures the trade-offs between model performance and dataset fidelity, guiding the selection of algorithms for real-world deployment.

5.Conclusion

Based on a comparative analysis of real-world and synthetic data for predicting ship vibration frequency (N_{2NV}), this study concludes that synthetic data yields deceptively high accuracy (R^2 > 0.99) with simple linear models, while real-world data poses significant challenges, often resulting in poor model performance. The neural network proved most robust for real data, achieving an R^2 of 0.816 for tankers, whereas linear regression failed completely with negative R^2

values. Consequently, synthetic data alone is insufficient for real-world deployment due to its inability to capture operational noise and hydrodynamic complexities. These findings underscore that reliable predictive systems must prioritize hybrid physics-ML approaches and be validated on diverse real-world operational data to ensure maritime safety and structural integrity. Future efforts should focus on augmenting limited empirical datasets with physically informed synthetic samples to bridge this performance gap.

6.Acknowledgement

The authors sincerely acknowledge the financial and equipment support provided under the Basic Research Grant (Office Order No. 9357, dated 12 July 2025), which simplifies access to essential computational resources for this study. The authors are grateful to the Department of Naval Architecture and Marine Engineering at Bangladesh University of Engineering and Technology (BUET) for keeping a supportive academic environment that greatly contributed to this study. Moreover, the authors also express their

sincere appreciation to Professors of Naval Architecture and Marine Engineering department for their continuous guidance, insightful feedback, and encouragement throughout the research process. Appreciation is also extended to the faculty members and colleagues at BUET whose constructive discussions and suggestions enriched the quality of this study. Finally, the authors thank the ICMERE 2025 organizers for giving them the opportunity to share this study.

7.References

[1] C. Samuels, "MOL Comfort Accident: The Worst Shipping Disaster in History," traderiskguaranty, Jul 17, 2019.

[2] sharda, "The MSC Napoli Accident: Causes and Aftermaths," marineinsight, September 20, 2012.

[3] P. Karvelis, "Deep machine learning for structural health monitoring on ship hulls using acoustic emission method," *Ships and Offshore Structures,* vol. 16, no. 4, pp. 440-448, 2020.

[4] M. H. Shuwei Zhou, "A general physics-informed neural network framework for fatigue life prediction of metallic materials," *ScienceDirect,* vol. 322, 2025.

[5] "Vibration Analysis," in *Ship Vibration*, American Bureau of Shipping, 2023, pp. 43-61.

[6] "Structural Resonances," in *Vibration Analysis*, American Bureau of Shipping, 2023, pp. 28-41.

[7] "Interpolation," pp. 1-35.

[8] A. Burkov, The Hundred-page Machine Learning Book, 2019.

[9] O. Theobald, Machine Learning for Absolute Beginners: A Plain English Introduction, 2017.

[10] A. C. M. Sarah Guido, Introduction to Machine Learning with Python: A Guide for Data Scientists, 2016.

[11] G. Aurelien, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2022.

8. NOMENCLATURE

Symbol	Meaning	Unit
N_{2NV}	Two node vertical frequency	cycles/min
TEU	Twenty-foot Equivalent Unit	dimensionless
L	Length of the ship	meter
B	Breadth of the ship	meter
D	Depth of the ship	meter
W	Displacement of the ship	Tons