IJMT. Vol. 21, Iss. 1, 2025. (71-79)

International Journal of Maritime Technology

Journal homepage: ijmt.ir

AI-Driven Ship Resistance Prediction Using Three Key Hydrodynamic Parameters

Poorya Khorsandi [1], Ahmad Hajivand [2*]

[1] *BSc, Khorramshahr University of Marine Science and Technology; poorya.khorsandy.official@gmail.com*
[2] *Faculty of Marine Engineering Department, Khorramshahr University of Marine Science and Technology;hajivand@kmsu.ac.ir*

ARTICLE INFO

Article History:
Received: 13 May 2025
Last modification: 9 Jul 2025
Accepted: 11 Jul 2025
Available online: 11 Jul 2025

Article type:
Research paper

Keywords:
Ship Resistance
Hydrodynamic Parameters
Artificial Intelligence
Machine Learning
Correlation Analysis
Ensemble Methods
Design Optimization

ABSTRACT

This paper introduces an innovative, AI-driven methodology for predicting ship resistance using only three fundamental input parameters: Length Between Waterlines (LWL), Beam at Waterline (BWL), and Draft (T). Traditional resistance prediction techniques such as empirical methods, towing tank experiments, and computational fluid dynamics (CFD) simulations are highly accurate but involve significant time, cost, and complexity. Our approach leverages machine learning algorithms, including XGBoost, CatBoost, and Gradient Boosting, to derive a comprehensive suite of hydrodynamic characteristics from a robust dataset comprising 308 full-scale experiments across 22 different hull shapes. The methodology begins with meticulous data preprocessing and feature engineering, including normalization, outlier analysis, and correlation assessment, to ensure reliability and minimize error propagation. By transforming raw hydrodynamic data into dimensionless groups, our models effectively capture both linear and non-linear relationships among critical parameters such as displacement, wetted surface area, midship section area, waterplane area, and the longitudinal center of buoyancy (LCB). Simple linear regression techniques were successfully used to derive parameters with perfect correlations, while more complex non-linear interactions were accurately predicted using advanced ensemble methods. The integration of these AI models into a Django-based web application further enhances the utility of our approach, providing naval architects and marine engineers with a user-friendly, real-time tool for design optimization and performance evaluation. Comparative analysis indicates that our streamlined model delivers predictions of residual and frictional resistance with accuracy comparable to traditional methods, while offering significant improvements in computational efficiency and cost-effectiveness. Overall, this research bridges the gap between classical hydrodynamic theory and modern artificial intelligence techniques, offering a rapid, reliable, and scalable solution for ship resistance prediction that has the potential to significantly enhance early-stage design processes in naval architecture.

ISSN: 2645-8136

DOI: http://dx.doi.org/10.61882/ijmt.21.1.71

1. Introduction

Ship resistance is a fundamental aspect of naval architecture that directly impacts a vessel's performance, fuel consumption, and overall operational efficiency. Accurate prediction of ship resistance has traditionally relied on empirical methods, such as the Holtrop-Mennen formulation [1], towing tank experiments adhering to standardized protocols [2], and computational fluid dynamics (CFD) simulations [3][4]. While these methods offer high accuracy, they are often time-consuming, costly, and limited by scale effects inherent in experimental and numerical setups [5]. In recent years, the advent of artificial intelligence (AI) and machine learning (ML) techniques has presented a promising alternative to conventional methods, enabling rapid and cost-effective resistance prediction without the need for extensive experimental setups [6]. Recent advancements in AI have shown that with sufficient training data, machine learning models can accurately predict complex hydrodynamic behaviors and resistance components from minimal input parameters. However, most existing studies focus on specific aspects of ship performance and require large, multidimensional datasets.

In this work, we propose an innovative approach that streamlines the prediction process by utilizing only three fundamental input parameters Length Between Waterlines (LWL), Beam at Waterline (BWL), and Draft (T) to derive a comprehensive set of hydrodynamic characteristics, including Maximum Beam, displacement, wetted surface area, midship section area, waterplane area, and the longitudinal center of buoyancy (LCB). By integrating AI-driven models with established hydrodynamic theory, our methodology not only simplifies the input requirements but also maintains high predictive accuracy for both residual and frictional resistance. The methodology presented in this paper leverages a robust dataset from Delft University [7], comprising 308 full-scale experiments across 22 different hull shapes. This dataset, combined with advanced preprocessing techniques, correlation analysis, and sophisticated machine learning algorithms (such as XGBoost [8], CatBoost [9], and Adaptive Boosting [10]), forms the backbone of our predictive framework. Additionally, the developed model is integrated into a Django-based [11] web application, providing naval architects and marine engineers with a user-friendly tool for real-time hydrodynamic analysis and design optimization. In summary, this study aims to bridge the gap between traditional hydrodynamic analysis and modern AI methodologies, offering a rapid, reliable, and cost-effective solution for ship resistance prediction. The proposed approach not only enhances design efficiency but also paves the way for further research into AI applications in naval architecture.

2. Methodology

2.1. Data Acquisition and Preprocessing

In this study, we employ the Delft University dataset, which comprises 308 full-scale experiments conducted on 22 different hull shapes. The dataset encompasses essential hydrodynamic parameters such as Maximum beam (Bmax), Waterline beam (BWL), Draft (T), Depth (D), Waterline length (LWL), displacement, wetted surface area (S), midship section area (Ax), prismatic coeffects (Cp), waterplane area (Aw), and the longitudinal center of buoyancy (LCB). These parameters are pivotal for analysing ship resistance, ensuring a comprehensive representation of the hydrodynamic behaviour across diverse vessel designs. For data cleaning, we utilized Python libraries including NumPy [12], Pandas [13], and Scikit-learn [14]. Although the original dataset contained several hydrostatic features such as metacenter height and buoyancy keel distance that are generally valuable in hydrodynamic studies, these parameters were considered non-essential for our specific focus on ship resistance. By constructing a correlation matrix and leveraging our hydrodynamic knowledge, we identified and removed these extraneous features.

Furthermore, an examination of the dataset using Pandas confirmed that there were no missing (Nan) values (Figure 1), thereby ensuring data completeness and reliability [15]. To ensure the robustness of our dataset, we conducted an outlier analysis using box plots (Figure 2) and standard deviation charts (Figure 3) through Tukey's interquartile range method:

$$\boldsymbol{Q^1 = 25th\ percentile\ Q^3 = 75th\ percentile}$$

$$\begin{aligned} &\boldsymbol{IQR = Q_3 - Q_1} \\ &\boldsymbol{Lower\ bound = Q^1 - 1.5 \times IQR} \\ &\boldsymbol{Upper\ bound = Q^3 + 1.5 \times IQR} \end{aligned} \tag{1}$$

where Q_1 and Q_3 represent the 25th and 75th percentiles respectively, and $IQR = Q_1 - Q_3$ is the interquartile range. This method identifies outliers as observations falling outside this range, with the 1.5 multiplier providing a conservative threshold that preserves physically plausible extreme values while filtering statistical anomalies.

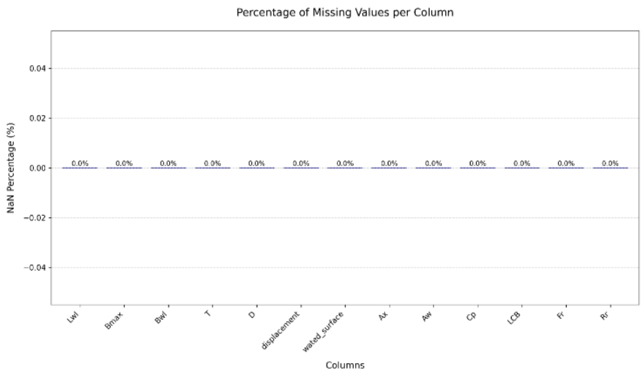

Figure 1. Percentage of missing values across hydrodynamic parameters.

These statistical tools revealed that while some extreme values exist in the dataset (visible in Figure 3's deviation analysis), they are limited in number and fall within reasonable physical bounds for ship hydrodynamic parameters. The standard deviation visualization (Figure 3) further confirms that nearly all data points lie within ±3σ of the mean. Given their limited quantity and potential validity as rare but authentic measurements, we retained these outliers to preserve the dataset's integrity and avoid introducing bias through excessive filtering.

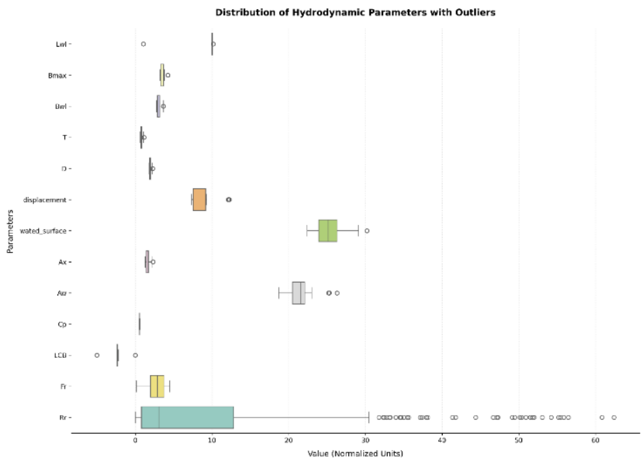

Figure 2. Distribution of parameters with outlier detection using box plots.

Given that the dataset comprises parameters measured in varying units and scales, feature standardization was a critical preprocessing step. We applied normalization and standardization [16] techniques using Scikit-learn preprocessing tools to transform the data such that each feature has a mean of zero and a standard deviation of one. This standardization minimizes error propagation and ensures that all features contribute equally during the machine learning model training, ultimately enhancing the stability and performance of our predictive analyses. The robust nature of our chosen machine learning approaches further mitigates potential outlier effects while maintaining sensitivity to genuine hydrodynamic phenomena.

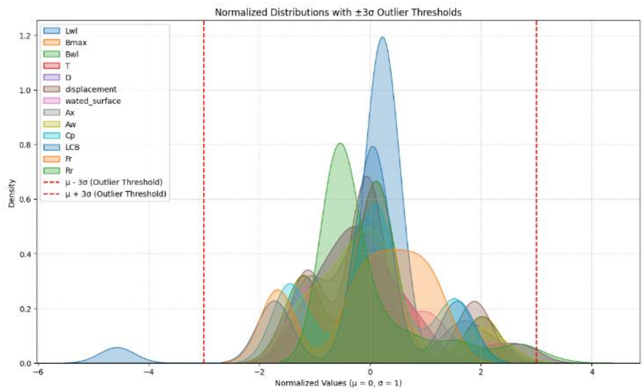

Figure 3. Normalized parameter distributions with ±3σ outlier thresholds.

2.2 Correlation Analysis and Feature Engineering

In the next phase of our methodology, we conducted a comprehensive correlation analysis [17] to elucidate the relationships between the hydrodynamic parameters in our dataset. By constructing a correlation matrix (Figure 4), we were able to identify several strong dependencies among the variables. Notably, perfect correlations were observed between Bmax and BWL, as well as between D and T. These findings allowed us to derive Bmax directly from BWL and D from T using simple linear regression models, each achieving 100% accuracy. The correlation matrix was constructed using the Pearson correlation coefficient, which quantifies the linear relationship between two variables, as defined by Equation 2:

$$r_{xY} = \frac{\left(\sum(x_i - \bar{x})(y_i - \bar{y})\right)}{\sqrt{\sum(x_i - \bar{x})^2 \times \sum(y_i - \bar{y})^2}} \qquad (2)$$

For parameters with more complex, non-linear interdependencies, such as the ship's displacement, we combined LWL, BWL, and T using a CatBoost Regressor. This approach effectively captured non-linear relationships, providing highly accurate predictions for displacement. Building upon these derived parameters, we further estimated other critical hydrodynamic quantities including the wetted surface area (S), midship section area (Ax), waterplane area (Aw), and longitudinal center of buoyancy (LCB) by employing appropriate regression models guided by the strengths of the observed correlations. To remove scale effects and ensure universal applicability of our model, all raw parameters were subsequently converted into dimensionless groups (e.g., L/B, L/T, Cp, L/Displacement^ (1/3), LCB, and various Froude numbers). This conversion not only normalized the data but also enhanced the robustness and generalizability of our predictive modelling framework.

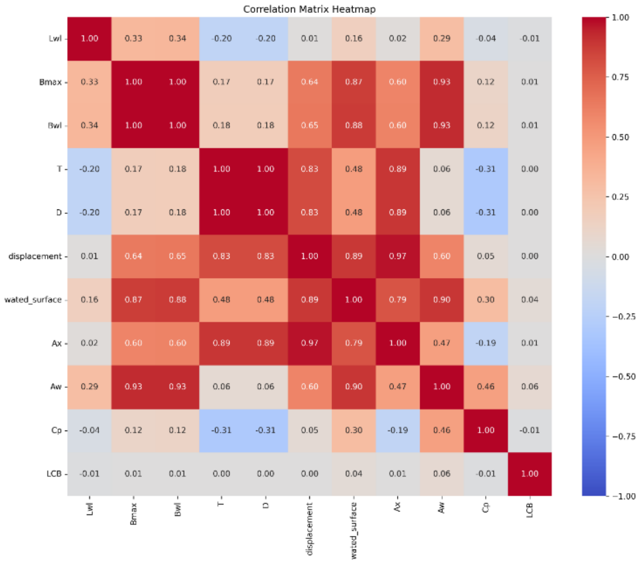

Figure 4. Correlation matrix of hydrodynamic parameters highlighting linear dependencies.

2.3 Residual and Frictional Resistance Estimation

In this phase, we focus on accurately quantifying the two major components of ship resistance residual and frictional resistance and combining them to compute the total resistance. To begin with residual resistance, our strategy involves converting the raw hydrodynamic parameters into dimensionless groups, such as L/B, L/T, Cp, L/Displacement^ (1/3), LCB, and various Froude numbers. This transformation is critical to eliminate scale effects and to embed the essential physical characteristics of the ship into our modelling process, thereby facilitating universal comparisons across different vessel sizes and designs.

We then evaluated several machine learning algorithms including XGBoost, LightGBM [18], CatBoost, Random Forest [19], Adaptive Boosting, and Support Vector Machine (SVM) [20] using these dimensionless parameters as inputs. The training process employed a 10-fold cross-validation strategy [21], ensuring that the models were robust and generalizable. Among the various models tested, XGBoost emerged as the best candidate, demonstrating an accuracy of approximately 99.8% in predicting residual resistance.

A key aspect of optimizing the XGBoost model involved using grid search for hyperparameter tuning to systematically evaluate combinations of learning rate, maximum tree depth, number of estimators, subsample ratios, and regularization parameters. Model performance was assessed through four key metrics: Mean Absolute Error (MAE) quantifying average prediction deviations (Eq. 3), Mean Squared Error (MSE) emphasizing larger errors through squaring (Eq. 4), Root Mean Square Error (RMSE) maintaining dimensional consistency (Eq. 5), and R-squared (R^2) measuring explained variance (Eq. 6). The optimal configuration was selected through 10-fold cross-validation by simultaneously minimizing MAE, MSE, and RMSE while maximizing R^2, ensuring robust capture of non-linear interactions for residual resistance prediction. This rigorous process yielded the following evaluation metrics:

$$\boldsymbol{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (3)$$

$$\boldsymbol{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (4)$$

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (5)$$

$$\boldsymbol{R^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (6)$$

For frictional resistance calculation, we adopted the well-established ITTC-1957 method (Eq. 7), which computes the friction coefficient (C*F*) based on the Reynolds number (Re). This formulation is widely accepted in naval architecture for estimating viscous drag:

where Re=VL/*ν* is the Reynolds number, with V representing ship velocity, *L* the waterline length, and *ν* the kinematic viscosity of water. The ITTC-1957 method provided a reliable measure of the frictional resistance component, particularly at lower speeds where viscous effects dominate.

$$\boldsymbol{Re} = \frac{(V \times L)}{\nu}$$

$$\boldsymbol{C_F} = \mathbf{0}.\frac{075}{(\log 10\,(Re) - 2)^2} \qquad (7)$$

Finally, the total resistance of the ship is determined by summing the residual resistance (predicted via the optimized XGBoost model) and the frictional resistance (calculated using the ITTC57 method). This integrated approach, which combines data-driven machine learning with classical hydrodynamic theory, offers a robust and accurate estimation of the forces opposing the ship's movement. Such a comprehensive methodology not only streamlines the resistance prediction process but also enhances the precision of performance evaluations and design optimizations in naval architecture.

The complete code is available at https://github.com/pooryakhorsand/ml_yacht_resistance_3parameters/blob/main/Untitled5.ipynb.

3. Results and Discussion

In this study, our primary objective was to develop an AI-driven model capable of predicting all essential hydrodynamic properties and the total resistance of sailing yachts using only three fundamental input parameters: Length Between Waterlines (LWL), Beam at Waterline (BWL), and Draft (T). This innovative approach not only streamlines the data input process but also integrates multiple predictive steps to accurately derive additional parameters such as Bmax, D, displacement, wetted surface area, Ax, Ay, Cp, and LCB culminating in the calculation of both frictional and residual resistance, and ultimately the total resistance.

This section is organized to systematically present the predictive performance of each step in our model. We begin with the derivation of key hydrodynamic parameters through correlation analysis and regression techniques, followed by a detailed evaluation of our AI models, including the comparison of several machine learning algorithms used for predicting residual resistance. Furthermore, the section discusses how our AI-driven methodology compares with traditional resistance estimation techniques such as towing tank experiments and CFD simulations, highlighting both the advantages and potential limitations of our approach.

3.1 Correlation Analysis and Parameter Derivation

Our initial analysis involved constructing a comprehensive correlation matrix to understand the relationships between all hydrodynamic parameters. As illustrated in the accompanying heatmap (Figure 4), two standout findings emerged: a perfect 100% correlation between Bmax and BWL, and likewise between D and T. These perfect correlations imply that Bmax can be directly derived from BWL and D from T without loss of information. Consequently, these findings significantly simplify our input space, allowing us to expand the parameter set while maintaining the integrity of the predictions.

To leverage these perfect correlations, we implemented simple Linear Regression models. For Bmax, using BWL as the sole predictor yielded a regression model with 100% accuracy. A similar approach was applied to predict D from T, resulting in equally flawless performance. The resulting regression equations are presented in Eq. 8. These outcomes not only validate the reliability of our dataset but also ensure that the derived parameters enhance the overall model without introducing errors.

$$\boldsymbol{Bmax} = -\mathbf{0.05} + (\mathbf{1.18}) * \boldsymbol{Bwl}$$
$$\boldsymbol{D} = \mathbf{1.15} + (\mathbf{1.00}) * \boldsymbol{T} \tag{8}$$

Unlike the linear relationships observed for Bmax and D, the displacement of the ship exhibited a non-linear correlation with the primary input parameters (BWL, T, and LWL). As shown in the correlation heatmap (Figure 4), displacement has a strong correlation with D (0.83), a moderate correlation with BWL (0.65), and a weak correlation with LWL (0.1). The low correlation between displacement and LWL can be attributed to the limited variation in LWL across the dataset. Our dataset includes 22 ship bodies, with LWL values ranging only from 10 to 10.15. This narrow range means that LWL has minimal influence on other parameters and overall resistance, leading to its weak correlation with displacement. To effectively model this complexity, we utilized the CatBoost Regressor, a powerful gradient-boosting algorithm known for efficiently handling non-linear relationships. The model demonstrated 100% accuracy in predicting displacement, as confirmed by our comparative performance metrics and scatter plots (Figure 5). These results affirm the suitability of the CatBoost approach in capturing intricate hydrodynamic dependencies.

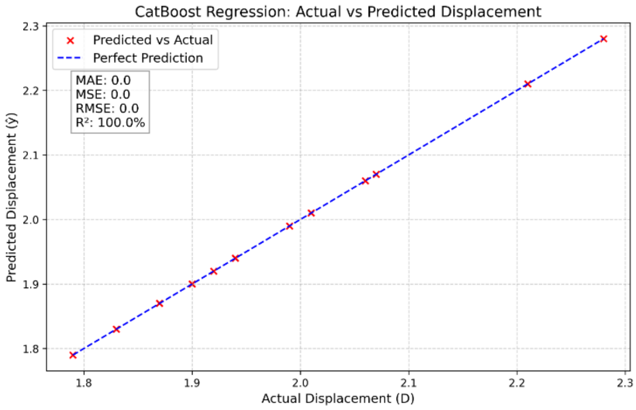

Figure 5. Scatter plot of predicted vs. actual ship displacement, demonstrating CatBoost Regressor accuracy.

For the prediction of wetted surface area (S), we employed the CatBoost Regressor to capture the non-linear relationships among the input parameters and S. Our model used LWL, BWL, T, and displacement as inputs, with observed correlations of 0.16, 0.88, 0.48, and 0.89 with S, respectively. The CatBoost model achieved near 100% accuracy in predicting S, as demonstrated by the performance plot in Figure 6, resulting in highly precise predictions with an error margin below 1%.

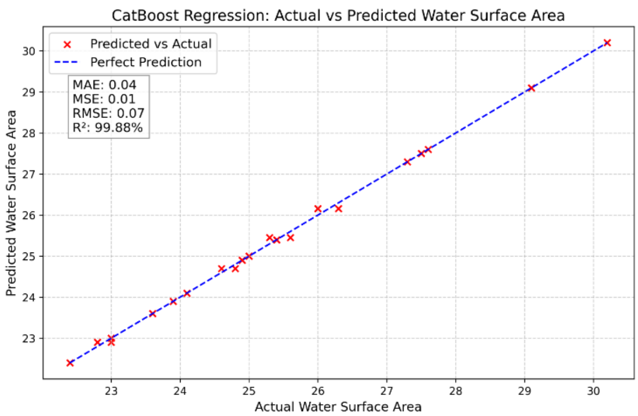

Figure 6. Scatter plot of predicted vs. actual water surface, demonstrating CatBoost Regressor accuracy.

For the prediction of the midship section area (Ax), we utilized a Linear Regression model, leveraging T, BWL, and displacement as the primary input parameters. These variables demonstrated strong correlations with Ax, including 0.97 with displacement, 0.89 with T, and 0.6 with BWL, validating the choice of the Linear Regression approach. The model provided highly accurate predictions, confirming its effectiveness in capturing the relationships between the input parameters and Ax, as represented in Equation 9.

$$\boldsymbol{Ax} = -\mathbf{12511111033481.34} + (\mathbf{0.52}) * \boldsymbol{Bwl} + (-\mathbf{10879226985632.47}) * \boldsymbol{T} + (\mathbf{10879226985634.52}) * \boldsymbol{Displacement} \quad (9)$$

The prediction of the waterplane area (Aw) was similarly based on S, Ax, BWL, and displacement, which demonstrated correlations of 0.9, 0.47, 0.93, and 0.6, respectively, with Aw. A Linear Regression model was employed, and the formula for Aw is provided in Equation 10.

$$\boldsymbol{Aw} = \mathbf{11.42} + (\mathbf{0.98}) * \boldsymbol{wated_{surface}} + (\mathbf{0.65}) * \boldsymbol{Ax} + (\mathbf{0.65}) * \boldsymbol{Bwl} + (-\mathbf{9.00}) * \boldsymbol{Displacement} \quad (10)$$

For the block coefficient (Cb), we used its well-known formula, which defines Cb as the ratio of displacement volume to the product of LWL, BWL, and T. Since this parameter is calculated directly rather than predicted via regression, no additional model was necessary. Similarly, for predicting the prismatic coefficient (Cp), we used its well-known formula, which defines Cp as the ratio of displacement to the product of the midship section area (Ax) and the waterline length (LWL). This formula allows us to directly calculate Cp without requiring a regression model. The formulas for Cb and Cp are shown in Equation 11.

$$\boldsymbol{C_B} = \frac{\boldsymbol{Displacement}}{\boldsymbol{LWL} * \boldsymbol{BWL} * \boldsymbol{T}}$$

$$\boldsymbol{C_p} = \frac{\boldsymbol{Displacement}}{\boldsymbol{Ax} * \boldsymbol{LWL}} \quad (11)$$

The prediction of the longitudinal center of buoyancy (LCB) proved challenging due to its weak correlation with the primary input parameters. To improve the model's predictive capability, we included all available parameters, enabling the model to better capture the complex, non-linear relationships influencing LCB. As a result, we employed a Gradient Boosting model, which efficiently handled these intricate dependencies. The corresponding scatter plot shown in Figure 7 illustrates the model’s satisfactory accuracy, despite the inherent challenges in prediction.

Table. 1. summarizes the prediction models, input features, and performance metrics for each hydrodynamic parameter.

Feature	Model	MAE	MSE	RMSE	R^2
Bmax	Linear Regression (BWL)	0.00	0.00	0.00	100%
D	Linear Regression (T)	0.00	0.00	0.00	100%
Displacement	CatBoost (BWL, T, LWL)	0.00	0.00	0.00	100%
Wetted Surface Area (S)	CatBoost (LWL, BWL, T, Displacement)	0.04	0.01	0.07	99.88%
Midship Section Area (Ax)	Linear Regression (T, BWL, Displacement)	0.01	0.00	0.02	99.68%
Waterplane Area (Aw)	Linear Regression (S, Ax, BWL, Displacement)	0.12	0.02	0.16	99.36%

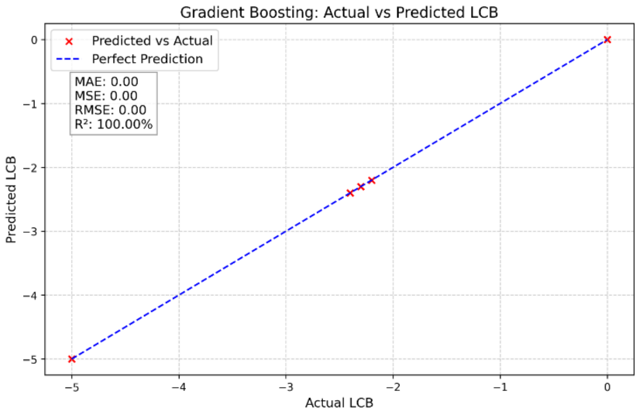

Figure 7. Scatter plot of predicted vs. actual water surface, demonstrating CatBoost Regressor accuracy.

In addition, Table 1 summarizes the performance of our predictive models across all hydrodynamic features. For each predicted parameter, the table details the selected input features, the machine learning model used (e.g., Linear Regression, CatBoost, Gradient Boosting [23]), and key performance metrics such as R^2, MSE, MAE, RMSE, and MAPE. This comprehensive overview confirms the high accuracy and reliability of our approach, highlighting the suitability of simple regression for directly correlated parameters (Bmax and D), CatBoost for capturing non-linear relationships (displacement and wetted surface area), and Gradient Boosting for modelling complex dependencies (LCB).

3.2 Residual, Frictional, and Total Resistance Estimation

Residual resistance, which includes wave-making and viscous pressure resistance, was predicted using machine learning algorithms. Given our dataset of 308 rows and a limited number of columns, traditional deep learning approaches resulted in poor accuracy due to overfitting [24]. Instead, we tested six machine learning models XGBoost, LightGBM, CatBoost, Random Forest, Adaptive Boosting, and SVM.

To enhance model performance, we transformed key hydrodynamic parameters into dimensionless groups such as L/B, L/T, Cp, L/Displacement^ (1/3), LCB, and various Froude numbers. These transformations remove scale effects and enable universal comparisons across different vessel sizes. The inclusion of multiple Froude numbers further improved model accuracy by capturing residual resistance across varying speed ranges.

Among the tested models, XGBoost achieved the highest accuracy (99.8%), demonstrating superior handling of non-linear dependencies in predicting residual resistance. Detailed performance metrics MAE, RMSE, and R^2 are presented in Figure 8, and the corresponding scatter plots for each algorithm are shown in Figure 9. Additionally, Table 2 consolidates these metrics (MAE, MSE, RMSE, and R^2) for all six algorithms, providing a comprehensive comparison of their predictive performance.

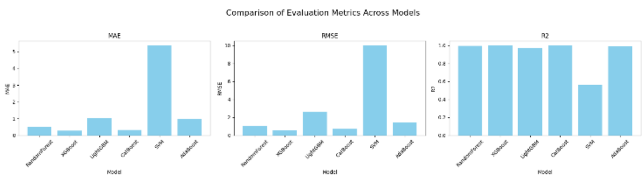

Figure 8. MAE, RMSE, and R² for each machine learning model.

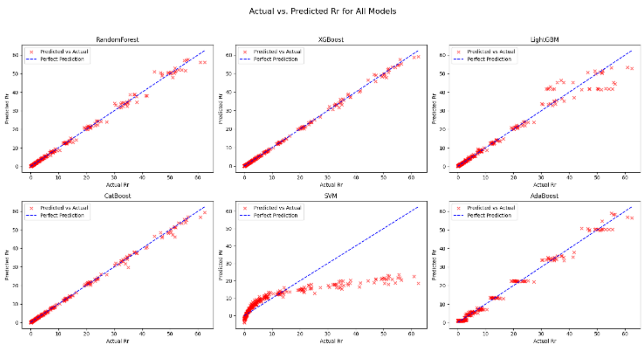

Figure 9. Scatter plots of predicted versus actual residual resistance.

Table. 2. presents the performance metrics of six machine learning models for predicting residual resistance.

Model	MAE	MSE	RMSE	R²
XGBoost	0.270534	0.297569	0.545498	99.87%
CatBoost	0.321662	0.503819	0.709802	99.78%
Random Forest	0.487186	1.031691	1.015722	99.55%
AdaBoost	0.992492	2.024278	1.422771	99.12%
LightGBM	1.030206	6.928767	2.632255	96.98%
SVM	5.365370	99.734449	9.986714	56.47%

Frictional resistance, representing the viscous drag due to the interaction between the ship's hull and water, was computed using the ITTC57 method, a widely accepted standard in naval architecture. This method calculates the friction coefficient based on the Reynolds number, as formulated in Eq. 7, yielding a reliable estimate of frictional drag, especially at lower speeds where viscous effects are most significant.

The total resistance of the ship is determined by summing the residual resistance (predicted via our machine learning model) and the frictional resistance (computed using the ITTC57 method). This integrated approach, which combines data-driven insights with classical hydrodynamic theory, provides a comprehensive and accurate estimation of the forces opposing the ship's motion, offering a robust tool for optimizing the design and performance of sailing yachts.

3.3 Comparison with Traditional Methods

Traditional methods, such as towing tank experiments and Computational Fluid Dynamics (CFD) simulations, have long been the backbone of hydrodynamic analysis in naval architecture. Towing tank experiments are known for their high accuracy and direct measurement of resistance forces, but they are expensive, time-consuming, and limited by scale effects. Similarly, CFD simulations provide detailed insights into fluid dynamics around the hull, yet they require substantial computational resources and specialized expertise. In contrast, our AI-driven approach using a streamlined three-parameter input model (length, width, and draft) offers rapid predictions at a fraction of the cost. By converting parameters into dimensionless groups and employing machine learning algorithms like XGBoost, we achieve near-instantaneous estimates of both residual and frictional resistance. Although this method may not capture every minute physical nuance, it strikes an excellent balance between computational efficiency and acceptable experimental accuracy.

One of the primary advantages of our approach is its simplicity: using only three input parameters within a pre-defined range (e.g., width values between 2 and 3, as determined by our dataset) significantly reduces data collection and processing complexity. This minimalistic input model allows for quick iterations during the early design phases, which is crucial for optimizing performance and fuel consumption. However, there are certain limitations to this approach. The primary constraint is that users can only modify the length, width, and draft within the tested range of the dataset. This ensures accurate predictions but restricts flexibility in exploring designs beyond the dataset's predefined limits. Additionally, while the model effectively estimates resistance parameters, any errors in the original dataset may influence the predicted values. Despite these constraints, this method remains a powerful tool for initial yacht design, allowing for rapid evaluation of hydrodynamic performance within the validated parameter space.

Our AI-driven methodology has been successfully integrated into a Django-based application, where users can input the three key parameters length, width, and draft within specified ranges. Once the data is entered, the application immediately displays the remaining hydrodynamic parameters and resistance values through intuitive graphs and tables. This real-time visualization not only aids designers in quickly assessing design viability but also facilitates early-stage optimization of yacht models, enhancing both fuel consumption estimates and overall operational efficiency. According to Figure 10, the interface presents a clear and structured layout, allowing users to efficiently analyse resistance components and make informed design decisions.

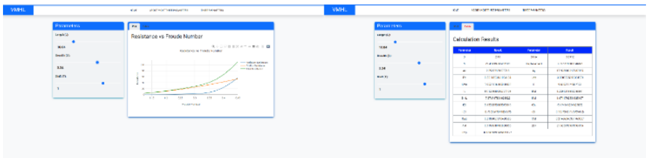

Figure 10. Django app interface showing resistance plots (left) and chart view (right) for user inputs.

The implementation of this AI-driven approach offers significant benefits for the design process. By streamlining iterations and reducing dependency on time-consuming experiments or computationally expensive simulations, it enables rapid prototyping and cost-effective design optimization. Designers can experiment with different configurations and immediately see the hydrodynamic implications, leading to faster and more informed decision-making. Looking forward, there is potential for further refinement by incorporating additional parameters or expanding the dataset's range, which could enhance the model's accuracy and extend its applicability to a wider variety of hull forms and operating conditions.

3.4 Discussion of Error Sources and Future Work

Our experiments indicate that when the length, width, and draft of the ship remain within the limits defined by our experimental dataset, the prediction chain is highly reliable with minimal error propagation. This is largely because the derived parameters are calculated from accurately constrained inputs. However, potential sources of error do exist. For example, assumptions inherent in the regression models and the quality of the dataset could contribute to discrepancies if the inputs fall outside the tested range. Any deviation from these limits may lead to error propagation through the derived parameters, affecting the overall resistance predictions.

To enhance the robustness and generalizability of our AI-driven model, future work should focus on expanding the dataset. Creating a dataset with more than 300 results that covers a much larger range of dimensions would allow users to experiment with a broader spectrum of designs. This expansion would facilitate try-and-error processes over an extended parameter space, ultimately improving model accuracy and versatility. Additionally, integrating real-time simulation tools and incorporating feedback from field data could further refine the model. Extending the methodology to accommodate other vessel types and operational conditions is also a promising avenue for future research.

In summary, our work demonstrates the feasibility of deriving multiple hydrodynamic parameters from just three primary inputs length, width, and draft without compromising on accuracy, provided the inputs remain within the experimental limits. The AI model, particularly when leveraging XGBoost for residual resistance, has proven to be a reliable predictor of ship resistance.

4. Conclusions

In conclusion, this study successfully demonstrates the viability of an AI-driven approach to ship resistance prediction using a minimal set of three primary inputs Length Between Waterlines (LWL), Beam at Waterline (BWL), and Draft (T). By integrating classical hydrodynamic theory with advanced machine learning algorithms, our methodology not only simplifies the design process but also maintains a high degree of predictive accuracy. The robust dataset from Delft University, encompassing 308 full-scale experiments across 22 hull types, provided a solid foundation for developing models that accurately derive multiple hydrodynamic parameters such as displacement, wetted surface area, midship section area, waterplane area, and longitudinal center of buoyancy (LCB). The research highlights the strengths of employing both linear regression for parameters with perfect correlations and sophisticated ensemble methods like XGBoost, CatBoost, and Gradient Boosting for capturing more complex, non-linear interactions. These techniques effectively bridge the gap between traditional empirical methods and modern computational practices, offering a streamlined process that is both cost-effective and computationally efficient. The integration of these models into a Django-based web application further enhances their practical applicability, enabling real-time analysis and design optimization for naval architects and marine engineers. Moreover, while the proposed approach significantly reduces the dependency on time-consuming experimental setups and computationally expensive simulations, it remains robust within the validated range of input parameters. The findings pave the way for future research aimed at expanding the dataset and refining model accuracy, thereby extending the approach's applicability to a broader range of vessel designs and operating conditions. Overall, this work not only contributes to the evolving landscape of AI applications in naval architecture but also offers a promising tool for enhancing early-stage design processes and operational performance in maritime engineering.

5. References

1- Holtrop, J., and Mennen, G. G. J., (1982), An approximate power prediction method, International Shipbuilding Progress, 29(335), p.166-170.

2- ITTC, (2017), Practical guidelines for ship resistance tests, International Towing Tank Conference.
3- Molland, A. F., Turning, S. and Forbes, P., (2010), Principles of Naval Architecture, Society of Naval Architects and Marine Engineers.
4- Blevins, R. D., (2014), Applied Fluid Dynamics Handbook, Krieger Publishing Company.
5- Larsson, L., Stern, F. and Visonneau, M. (Eds.), (2014), Numerical ship hydrodynamics: An assessment of the Gothenburg 2010 workshop, Springer.
6- Panda, J. P., (2021), Machine Learning for Naval Architecture, Ocean and Marine Engineering, arXiv:2109.05574 (CC BY 4.0).
7- Gerritsma, J., Onnmk, R. and Versluis, A., (1981), Geometry, Resistance and Stability of the Delft Systematic Yacht Hull Series, Delft University of Technology.
8- Chen, T., and Guestrin, C., (2016), XGBoost: A scalable tree boosting system, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, p.785-794.
9- Dorogush, A. V., Gulin, A., Kazeev, V. and Prokhorenkova, L., (2018), CatBoost: gradient boosting with categorical features support, arXiv preprint.
10- Freund, Y., and Schapire, R. E., (1997), A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Sciences, 55(1), p.119-139.
11- Django Software Foundation, (2023), Django: A high-level Python web framework.
12- Harris, C. R., et al., (2020), Array programming with NumPy, Nature, 585(7825), p.357-362.
13- McKinney, W., (2010), Data structures for statistical computing in Python, Proceedings of the 9th Python in Science Conference, p.51-56.
14- Pedregosa, F., et al., (2011), Scikit-learn: Machine learning in Python, Journal of Machine Learning Research, 12, p.2825-2830.
15- Tukey, J. W., (1977), Exploratory Data Analysis, Addison-Wesley.
16- Witten, I. H., Frank, E., Hall, M. A. and Pal, C. J., (2016), Data Mining: Practical Machine Learning Tools and Techniques (4th ed.), Morgan Kaufmann.
17- Pearson, K., (1895), Notes on regression and inheritance in the case of two parents, Proceedings of the Royal Society of London, 58, p.240-242.
18- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... and Liu, T. Y., (2017), LightGBM: A highly efficient gradient boosting decision tree, Advances in Neural Information Processing Systems, 30, p.3146-3154.
19- Breiman, L., (2001), Random forests, Machine Learning, 45(1), p.5-32.
20- Cortes, C., and Vapnik, V., (1995), Support-vector networks, Machine Learning, 20(3), p.273-297.
21- Kohavi, R., (1995), A study of cross-validation and bootstrap for accuracy estimation and model selection, Proceedings of the 14th International Joint Conference on Artificial Intelligence, 2(12), p.1137-1143.
22- Bergstra, J., and Bengio, Y., (2012), Random search for hyper-parameter optimization, Journal of Machine Learning Research, 13(Feb), p.281-305.
23- Friedman, J. H., (2001), Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics, 29(5), p.1189-1232.
24- Géron, A., (2019), Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.), O'Reilly Media. ISBN: 978-1492032649.