International Journal of Marine Engineering and Applications
Volume 3, Number 1, March 2026
e-ISSN: 3063-640X

Date of Received: December 21, 2025
Date of Accepted: March 4, 2026
Date of Published: March 31, 2026
DOI: doi.org/10.30649/ijmea.v3i1.399

DEVELOPING A DATA-DRIVEN METHOD FOR YACHT DIMENSION PREDICTION

Md. Tawhid Hassan[1]*, Md. Riad Khan [2]

[1] Department of Naval Architecture and Offshore Engineering, Bangladesh Maritime University, Dhaka-1216, Bangladesh
[2]Lecturer, Department of Naval Architecture and Offshore Engineering, Bangladesh Maritime University, Dhaka-1216, Bangladesh

*Corresponding Author: naoe22005.tawhid@bmu.edu.bd

ABSTRACT

A yacht's principal dimensions can be found in many ways. However, due to the massive technological advancements in the yacht industry, using these ancient methods is just a waste of time. A new statistical method is necessary to determine yacht dimensions in an easy and effective way. In this paper, 122 modern yacht data have been used to investigate the relationship between length, breadth, draught, speed, gross tonnage, and power, and to perform regression analysis to develop a new method for estimating yacht dimensions. This study developed two predictive models: Model 1 utilizes empirical ratios and trendline equations, while Model 2 employs sequential stepwise multiple regression. The models effectively estimate breadth, draft, gross tonnage (GT), speed, and power from a specified length, with geometric parameters (e.g., GT prediction R^2=0.97 in Model 1) showing higher reliability than performance parameters. The minimum and maximum ratio of Length to Breadth, draught to breadth, for a different range of ship length is also determined. This research is conducted in such a way that the owner's requirement for a new yacht is the length, and other particulars are determined accordingly.

Keywords: Yacht, regression, polynomial trendline, linear regression, power trendline

Introduction

A yacht should have modern technologies, facilities, as well as economically beneficial. In particular, the length of the yacht represents the main discriminating factor with regard to the technical and commercial typologies of the vessels, which have given rise to the category's 'superyachts', 'mega yachts', 'giga yachts', and 'dream yachts' [1]. However, the exact definition of these categories in terms of length is, to a certain degree, subjective and not clearly defined, and the only objective classification is that which divides the fleet into vessels below 24 m in length ('small yachts') or over 24m ('superyachts') [1]. A yacht has a linear relation between its length, breadth, and draught. Analyzing modern yachts, a new method for determining dimensions can be found.

There is a clear need for simpler, data-driven tools that reflect contemporary yacht practice. A regression-based approach can leverage databases of recent yacht designs to predict one dimension from others. In particular, if an owner specifies a desired length, it should be possible to estimate beam, draft, volume, speed, and power requirements using statistical correlations. To our knowledge, such methods are scarce in the published literature. For example, [2] used regression on 20-40 meter yachts to determine LOA, beam, draft, and related parameters in early design, and [3] developed formulas for light displacement from length, breadth, and draft on 240 yachts. However, these studies focus on displacement or tonnage prediction. In contrast, the present work aims to directly predict hull dimensions and performance metrics from length.

Although regression has been used in previous studies on the design of yachts [3], the studies tend to address the specific output of the regression, such as displacement. The uniqueness of this work is in presenting a detailed, coherent combination of basic empirical equations obtained solely out of contemporary yacht data, which are able to forecast the entire array of major dimensions and essential performance indicators directly and consecutively out of a single owner-keyboard input, the length overall (LOA). The strategy is to bridge the gap with a fast, data-driven preliminary design tool.

The main aim of the research is to work out and prove a functional and regression-based methodological framework for the quick initial dimensioning of contemporary motor yachts. The framework is designed to offer a dependable tool to the naval architects that, given one input parameter from the owner of Length Overall (LOA), gives correct first-estimates of the main dimensions (Breadth and Draught), volumetric measurement (Gross Tonnage), and important performance measures (Speed and Power). This study aims to simplify the design stage to produce explicit equations, which are based on current data, and not the conventional, time-consuming, displacement-based computation.

First, it makes use of a filtered set of 122 yachts (since 2005), so that the predictive models are up to date with the current tendencies in designing the yachts and their technological features, as opposed to using the results of the old-fashioned empirical data. Second, it presents a two-model solution: Model 1 will provide flexibility by allowing empirical ratios and trendline equations to provide independent parameter estimates, whereas Model 2 will provide a sequential and integrated method of calculation, with each of the predicted parameters serving as the input to the model, with only LOA as the input. Third, it presents a generalized set of straightforward, explicit equations of an expanded variety of outputs (B, T, GT, V, P) directly based on length, which fills a literature gap that frequently discusses predictive displacement or tonnage separately but does not provide a holistic, sequential sizing instrument to be used in the preliminary design discovery.

The initial design of the yacht used in the past has depended on rules of thumb, experience, and extrapolation of the existing vessels, and on the tradition practiced by commercial ship designers. Nonetheless, the current design of a yacht has changed hugely owing to the high demands of an owner, the swift change in technology, and a broad diversity in hull shapes and performance objectives. Research on yacht design also highlights that length overall (LOA) is the main parameter that is placed at the concept phase, and other key dimensions are calculated later, which explains the necessity of systematic and data-driven estimation tools that are specific to yachts [1].

Recent studies have shown that the principal dimensions of motor yachts, length, breadth, and draught, show high interdependences and strong scaling properties over the variations in size. The statistical results of modern databases of yachts indicate that these geometric relations can be appropriately utilized in preliminary design in an attempt to predict dimensions and volumetric features [2]. These findings suggest that geometrical parameters tend to be more predictable than performance-based parameters, which are affected by other design and operational aspects.

Computational intelligence and regression-based applicant techniques have been successively applied to the design of yachts and ships. Cepowski demonstrated that regression and neural-network-based methods are capable of reliably predicting light displacement of motor yachts based on simple design parameters, and regression models with similar efficiency are less complex and more transparent [3]. This justifies the appropriateness of regression methods in estimating the dimensions of a yacht at the initial stages, especially when very little information is found.

The increased use of artificial intelligence and machine learning in the field of maritime engineering also contributes to the utilization of data-driven models in design-associated work. Vessel identification, resistance prediction, and performance estimation have been successfully performed using machine learning methods and have proven that meaningful patterns can be identified through large datasets [4] [5]. Nevertheless, these papers always indicate higher uncertainty in the forecasting of dynamic performance parameters like speed, resistance, and power than in geometric dimensions, and this indicates inherent variability in performance forecasts.

In terms of design methodology, rational and risk-based ship design frameworks note the significance of sound early-stage estimation tools since preliminary design choices have a potent impact on safety, cost, and performance issues in the lifecycle of the vessel [6]. It has been a well-established fact that simplified analytical and statistical models are needed to reduce the design space prior to the process of more detailed analyses [7]. The latest numerical research on the added resistance and ship performance further proves that geometry-based predictions are stronger compared to those that consider operational conditions and environmental impact [8].

Extensive analyses of machine learning and modelling in ship design indicate that, although highly advanced AI methods are emerging rapidly, regression-based and statistical models are still of great value in preliminary design since they are interpretable, inexpensive to compute, and simple to implement [9]. However, the available body of literature is predominantly concerned with the prediction of single parameters - i.e., displacement, resistance, or classification - but not a combined, length-based framework that would be able to estimate all the main dimensions and the main performance indicators of the modern yachts.

Methodology

The study was intended to offer contemporary, fast, and reliable predictive tools to the modern naval architect, particularly, the complexity and time-intensive nature of the older, displacement-driven empirical method of determination. The methodology was based on fitting trendlines and sequential stepwise multiple linear regression, which means that the formulated equations would be representative of the contemporary superyacht design practice.

a. Data Source and Sampling

The corpus of raw data (Table 1) was collected from the SuperYacht Times database, which is a reliable source of industry information specializing in documenting and tracking the large modern pleasure boats and providing relevant information on the contemporary design tendencies, and making the statistical models applicable to a wide range of classifications [10].

To place this study under a valid foundation, the dataset used contained more modern vessel specifications so that the resulting predictive equations applied would more accurately represent the current state of practices in naval architecture and technological development, and as such, the foundation was more reliable than the dependence on the old-fashioned empirical methods.

a.1 Sample Characteristics and Variables

A total sample size of 122 contemporary yacht data points was collected to be analysed.1 The length overall of data sampled vessels ranges extensively (between a minimum of 24.0 m and a maximum of 162.5 m). This extensive range is essential to the creation of statistically significant models, since it will enable the equations to mirror effects of scale between vessels between the standard superyachts and giga-yachts, avoiding extrapolation error when predicting the dimensions of large vessels.

Table 1. Yacht raw data description

Parameter	Unit	Min. Value	Max. Value
Length overall (L)	m	24	162.5
Breadth (B)	m	5.65	25.7
Draught (T)	m	1.46	6.15
Speed (V)	knots	8	29
Power (P)	kw	306	8200
Gross Tonnage (GT)	-	70	15917

b. Analytical Framework and Validation Criteria

The basic analysis method was to fit mathematical models to the observed correlation in the obtained data (e.g., L vs. B and L vs. T and the volumetric proxy L*B*T vs. GT).

b.1 Model Selection Criterion: Coefficient of Determination (R^2)

In all of the generated trendline equations (Model 1) and multiple regression equations (Model 2), the Coefficient of Determination (R^2) was used as the model validation and selection measure. The study required a very strict standard; the equations with greater values of R^2 were accepted only. This was used to make sure that the final selected equations made the maximum percentage of the overall variance in the dependent variable that could be statistically explained by the independent variables, creating a level of great confidence in the empirical data fitting.

One important point made by the analytical framework is the predictability order amongst the parameters. The model results showed that the geometric parameters (L, B, T, GT) had a consistently high R^2 value (more than 0.8), indicating that there was a strong, predictable linear or non-linear relationship between scales depending on the sample data. On the other hand, the dynamic performance parameters, which include speed (V) and power (P) calculated in Model 2, showed a significantly lower statistical fidelity (R^2 was 0.1182 and 0.5839, respectively). This difference confirms the complexity and variability inherent in the prediction of dynamic parameters in the specific cases of various hull forms and technological applications in comparison with the more stable relationships between fixed geometrical dimensions.

c. Model 1: Empirical Ratio and Trendline Analysis

Model 1 was aimed at giving a flexible initial dimension forecast by using either the required length of the yacht or the set hull form ratios (e.g., L/B and B/T). Linear, power, and polynomial trendline fitting are the methods that were used to come up with this model.

c.1 Initial Data Transformation and Ratio Analysis

The first step was to compute three key empirical ratios of all 122 sampled yachts L/ B, B/T ratio, and the product of major dimensions (L*B*T), which served as a proxy of volume. To provide context to preliminary design and hull shape assumption, the dataset was grouped categorically by length, which enabled the determination of the current ratio ranges of L/B and B/T in particular classes of designs, which serves as a guidepost to the designer when choosing the hull form parameters prior to creating the particular dimensions shown in Figure 1.

c.2 Trendline Generation and Modelling Dependencies

Many types of trendlines have been tested and implemented to determine the deterministic relationships among the major dimensions; only the ones having the highest value of the R2 were accepted.

- Breadth Prediction (Linear Trendline): The width of the yacht was adequately modelled as a linear expression of the length.
- Draught Prediction (Polynomial Trendline): Draught was predicted as a second-order polynomial function of breadth.
- Alternative Draught Prediction (Linear Trendline): This is the simplest alternative model that was used to forecast draught as a linear regression of length.
- Gross Tonnage Prediction (Power Trendline): Gross Tonnage is a measure of internal volume directly and was represented as a power defect of the volumetric proxy (L x B x T).

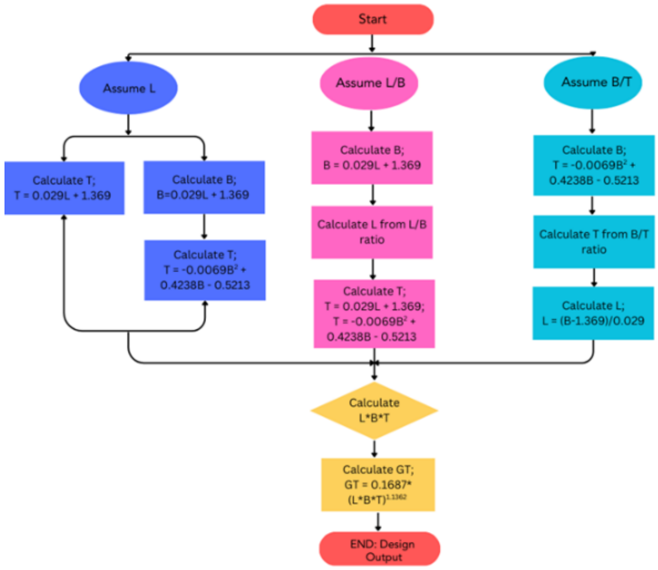

Figure 1. Flowchart of Model 1

d. Model 2: Sequential Stepwise Multiple Regression Analysis

This structure led to Model 2, which set out to design a fully integrated and simplified preliminary design tool using only the owner-specified yacht length as the input variable. The regression process was performed in five different steps, where L is the initial parameter, whose value is known at the outset of the series.

- Breadth Prediction: This step defined the basic transverse dimension using a univariate linear regression. Here, yacht length is the input variable.
- Draught Prediction: Draught was calculated with multiple linear regression, including the recalculated and predicted breadth to enhance the statistical accuracy, as the

draught is strongly dependent on length and beam.

- Speed Prediction: The four-dimensional dynamic parameter of performance, speed, needed all three initial hull sizes 9length, breadth, draught) to be predicted using multiple linear regression.
- Power Prediction: Multiple linear regression was used to determine required engine power, including the calculated dimensions (length, draught) and the target speed calculated.
- Prediction of Gross Tonnage: Gross Tonnage, which is the height of the vessel in terms of internal volume, was modelled by taking into consideration only the three physical dimensions (L, B, T). The particular decision to exclude V and P will keep the volumetric estimation unchanged and reduce the possibility of the volatile performance parameters affecting it.

The regression analysis follows the given procedure below (Figure 2):

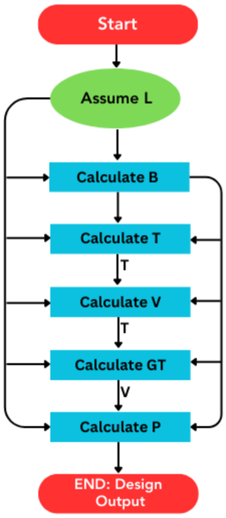

Figure 2. Flowchart of Model 2

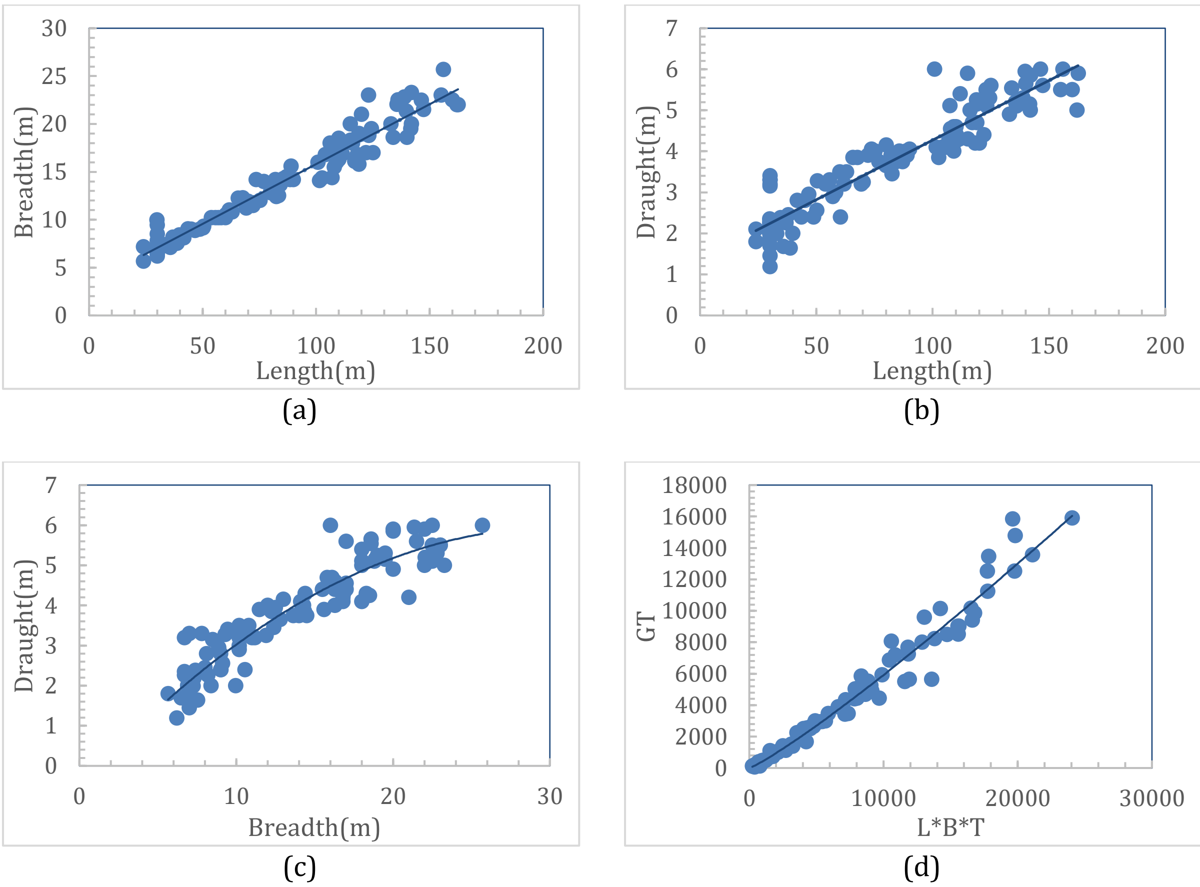

Figure 3. Relation between a) Breadth & Length, b) Draught & Length, c) Breadth & Draught, d) L*B*T & GT

Result and Discussion

a. Model 1

Figure 3 illustrates the foundational relationships within the dataset, which informed Model 1: (a) Breadth vs. Length shows a strong linear correlation; (b) Draught vs. Length displays more scatter by linear correlation (c) Draught vs. Breadth is better captured by a polynomial trend; and (d) Gross Tonnage scales predictably with the volumetric proxy (LBT) via a power-law relationship.

Generated Equations-

For Yacht length 20<L<50:
L/B = 3.00-5.42, B/T = 2.01-5.21
For Yacht length 50<L<100:
L/B = 5.18-6.68, B/T = 2.84-4.40
For Yacht length L>100:
L/B = 5.35-7.52, B/T = 2.67-5.00

$$B = 0.029L + 1.369 \quad (1)$$
$$T = -0.0069B^2 + 0.4238B - 0.5213 \quad (2)$$
$$T = 0.029L + 1.369 \quad (3)$$
$$GT = 0.1687*(L*B*T)^{1.1362} \quad (4)$$

Proposed Method to Estimate Main Dimensions of a New Yacht:

- For equations 1, 2, and 3, assume any yacht length as per client demand and put it in equation 1 to determine yacht breadth. Then put the breadth value in equation 2 to determine the yacht's draught or use the length in equation 3. Using the term L*B*T, determine the GT value from equation 4.
- If the owner's requirement is the L/B & B/T ratio, then determine the L value from the L/B ratio and put it in equation 1 to get breadth. Then, from the B/T ratio, put the B value in equation 2 or use L in equation 3 to determine draught. Finally, using the L*B*T term, GT can be found.

b. Model 2

The following equations were found through regression analysis:

$$B = 3.64 + 0.12*L \quad (5)$$
$$T = 1.07 + 0.02*L + 0.08*B \quad (6)$$
$$V = 19.83 + 0.05*L - 0.04*B - 2.11*T \quad (7)$$
$$P = -1429.88+9.46*L+277.28*B-474.82*T+80.80*V \quad (8)$$
$$GT = -4246.96 + 60.05*L + 279.46*B - 205.01*T \quad (9)$$

Proposed Method to Estimate Main Dimensions of a New Yacht-

- Take any Length value as the owner wants and put it in equation 5 to get the breadth value.
- Put length and breadth values in equation 6 to get the draft value.
- Put length, breadth, draught value at equations 7 & 9 to determine the Velocity & GT value.
- Then determine the power value from equation 8 using length, breadth, draft, and velocity.

This paper has given the minimum and maximum ratio of L/B & B/T with principal particulars dimension prediction. This ratio will help to assume the hull shape. This paper also categorized yachts in different ranges as 20-50 m, 50-100 m, and >100 m in length. It helps to accurately predict the L/B & B/T ratio. Model 1 (Table 2) was designed to predict breadth, draught, and GT from yacht length or the L/B, B/T ratio. Whereas model 2 (Table 3) was designed to predict breadth, draught, velocity, power, and GT from yacht length. The findings of this paper can be used for all sizes of yachts, as data from all sizes of yachts were used in this paper. All types of trendlines and interpolation have been used in the paper, and only equations with higher R^2 values were accepted.

Table 2. Model 1 coefficient of determination values

Equations	R^2
$B = 0.029L + 1.369$	0.9546
$T = -0.0069B^2 + 0.4238B - 0.5213$	0.8876
$T = 0.029L + 1.369$	0.8724
$GT = 0.1687*(L*B*T)^{1.1362}$	0.9725

The regression models showed that yacht dimensions were strongly interrelated. The beam grows nearly linearly with length: both Model 1 and Model 2 yielded linear fits (e.g., $B = 0.029L + 1.369$, $B = 3.642677426+ 0.122562498L$) with $R^2 \approx 0.92$–0.95. This implied a relatively uniform length-to-beam ratio across the sample, reflecting that larger yachts were generally slender (typical L/B ratios are on the order of 8–10). The draft depended nonlinearly on beam: a quadratic fit, $T = -0.0069B^2 + 0.4238B - 0.5213$ ($R^2 \approx 0.89$) captured the variation well. This quadratic form makes sense because hull volume (and thus draft) scales roughly with cross-sectional area.

Gross tonnage (GT) was also well-predicted by geometry: using the volume-based power law, $GT = 0.1687*(L*B*T)^{1.1362}$ gave $R^2 \approx 0.97$, indicating that

GT is essentially proportional to displaced volume for our yachts. A multiple linear regression in L, B, and T also achieved R^2 ≈0.86–0.97. These high R^2 values agree with the expectation that superyacht GT is primarily a function of hull volume and shape.

Table 3. Model 1 coefficient of determination values

Equations	R^2
B = 3.64 + 0.12*L	0.9205
T = 1.07 + 0.02*L + 0.08*B	0.8229
V = 19.83 + 0.05*L – 0.04*B – 2.11*T	0.1182
P = -1429.88+9.46*L+277.28*B-474.82*T+80.80*V	0.5839
GT = -4246.96 + 60.05*L + 279.46*B – 205.01*	0.8628

By contrast, predictions for speed and power were less accurate. In Model 2, the cruise speed regression (V vs. L, B, T) had a very low R^2 (~0.12), meaning that length, beam, and draft alone do not strongly determine yacht speed in our dataset. This reflects reality: yachts of the same size can have widely different design speeds depending on engine choice and intended use. Similarly, the power regression (P vs. L, B, T, V) yielded $R^2 \approx 0.58$. This moderate fit suggests that factors beyond hull dimensions—such as propulsion efficiency, hull roughness, and load conditions—contribute significantly to required power. In other words, while larger yachts need more power on average, there is considerable scatter. Designers should therefore use the speed and power estimates as rough guides; detailed resistance calculations would be needed for final specifications.

Overall, the models were most reliable for estimating geometric dimensions and volume (beam, draft, tonnage) and less so for performance (speed, power). This hierarchy of accuracy was expected: the relationships between L, B, and T were constrained by geometry and stability requirements, whereas V and P also depend on non-dimensional coefficients and operating conditions. The strength of the regression fitted for B and T ($R^2 \geq 0.87$) means that, given a length, a designer can predict beam and draft with confidence.

These findings provide actionable insight. In early design, one can use the equations to quickly size a hull: assume an LOA, compute B and T, then check that the implied L/B and B/T ratios fall within acceptable ranges (e.g., typical yachts have L/B around 8–11 and B/T around 3–4). The GT estimate (from volume) indicates the interior volume and regulatory tonnage, useful for cost and classification considerations. If a particular speed or power is required, one could use the inverse form of our equations to adjust the other variables, though such tuning should be validated by further analysis.

This research paper considered length, breadth, draught, velocity, power, and gross tonnage. But net tonnage and displacement were not considered. In the future, equations for these two parameters can be found. There are 2 models in this paper. More models can be found using more yacht data in the future.

Conclusion

The research was able to create and test an empirical-based, regression-oriented methodology framework for the quick preliminary determination of motor yacht base dimensions and the main performance indicators. With the help of a filtered set of 122 contemporary yachts, two unique predictive models were developed: Model 1, using accumulated ratios and trendline analysis, and Model 2, a stepwise model of multiple regression that only needs Length Overall (LOA) as an input.

The analysis established the predictability of geometric parameters: breadth (B), draught (T), and gross tonnage (GT) with high statistical accuracy based on length, which is indicated by R^2 values that are always greater than 0.87 and even near 0.97 when predicting the gross tonnage (GT) in Model 1. These high correlations are demonstrations of the underlying geometric and volumetric principles of the modern yacht design. Moreover, the research developed useful bands of critical hull form ratios (L/B and B/T) in various size groups (20-50 m, 50-100m, and over 100m), which is an important guide in the early assumptions of hull form.

Predictions of parameters of performance, speed (V), and power (P), in contrast, exhibited substantially less statistical fidelity (R^2 of 0.12 and 0.58, respectively). The outcome highlights the increased complexity and contributing variables to these metrics, such as hull form efficiency, propulsion technology, and operational profile, that are not completely represented in principal dimensions only.

One weakness of this study is that it dwells on the major dimensions and GT, with the omission of other significant parameters like displacement and net tonnage. Future studies need to consider these variables and increase the number of variables in order to improve the robustness of the models. Also, it might be better to investigate non-linear regression algorithms or machine learning algorithms to enhance the performance parameter prediction. Finally, the models suggested in this paper will provide a consistent empirical base for the preliminary sizing of contemporary motor yachts to balance the needs of the owner and the comprehensive design of the naval architect.

Data Availability Statement

The raw yacht data were obtained from the publicly accessible SuperYacht Times website (https://www.superyachttimes.com/). The authors processed and organized the data for analysis. The processed dataset is available from the corresponding author upon reasonable request.

References

[1] D. Boote et al., "ISSC 2012 Committee V.8 YACHT DESIGN," Oct. 2015, doi: https://doi.org/10.13140/rg.2.1.3985.7366

[2] F. Mauro et al., "The effect of main dimensions on the preliminary design of motor yachts," International Marine Design Conference, May 2024, doi: https://doi.org/10.59490/imdc.2024.904

[3] T. Cepowski, "An estimation of motor yacht light displacement based on design parameters using computational intelligence techniques," Ocean Engineering, vol. 231, p. 109086, Jul. 2021, doi: https://doi.org/10.1016/j.oceaneng.2021.109086

[4] H. Karna, M. Braović, A. Gudelj, and K. Buličić, "Artificial intelligence-based prediction model for maritime vessel type identification," Information, vol. 16, no. 5, Art. no. 367, Apr. 2025, doi: 10.3390/info16050367.

[5] Y. Yang et al., "Research on ship resistance prediction using machine learning with different samples," *Journal of Marine Science and Engineering*, vol. 12, no. 4, Art. no. 556, Mar. 2024, doi: 10.3390/jmse12040556.

[6] A. D. Papanikolaou, *Risk-Based Ship Design: Methods, Tools and Applications*. Berlin, Germany: Springer, 2009.

[7] O. F. Hughes, F. Mistree, and V. Žanić, "A practical method for the rational design of ship structures," *Journal of Ship Research*, vol. 24, no. 2, pp. 101–113, Jun. 1980, doi: 10.5957/jsr.1980.24.2.101.

[8] M. Mittendorf, *Data-driven prediction of added resistance on ships in waves*, Ph.D. dissertation, Dept. of Civil and Mechanical Engineering, Technical University of Denmark, 2023.

[9] P. D. Kaklis, K. Kostas, and S. Khan, "Machine learning and modeling for ship design," *Journal of Marine Science and Engineering*, vol. 13, no. 12, Art. no. 2304, Dec. 2025, doi: 10.3390/jmse13122304.

[10] "SuperYacht Times," *SuperYacht Times*, Nov. 28, 2019. [Online]. Available: https://www.superyachttimes.com/. [Accessed: Jan. 09, 2026].