Comparison of Deep Learning and Ensemble Learning as Surrogate Models in the Early Stage of Ship Design

Nanda Yustina, Ade Ratih Ispandiari, Annisa Roschyntawati, Zulfa Qonita, Nurul Shabrina, Nanda Itohasi Gutami, Noor Muhammad Ridha Fuadi, Waluyo, Siti Sadiah, Abdul Kadir, Zaid Cahya, Mochamad Adityo Rachmadi

***Abstract*—The growing competitiveness and digitalization-driven demand for faster logistics promote rapid and optimized ship design processes. To overcome this need, stakeholders are focusing on the development of safer and environmentally friendly vessels with shorter lead times and reduced Energy Efficiency Design Index (EEDI). Conventional iterative design methods, like the Spiral Design approach, are often time-consuming and prone to local optima. This study evaluates surrogate modelling approaches to simplify early-stage ship design by predicting power and steel weight requirements—key metrics for sustainability and cost-effectiveness. Three approaches are compared: individual surrogate models (Polynomial Regression [PR], Kriging, K-Nearest Neighbours [KNN]), ensemble models (weighted aggregation and stacking regressors integrating PR, Kriging, and KNN), and deep learning architectures. PR achieved the best accuracy with a Mean Absolute Error (MAE) of 106.75, Root Mean Square Error (RMSE) of 320.31, and near-perfect R^2 (0.999), outperforming all other methods. Ensemble models ranked second, leveraging the combined strengths of PR, Kriging, and KNN to deliver robust predictions and higher hypervolume in optimization tasks. While following PR and ensembles, deep learning models surpassed standalone Kriging and KNN, demonstrating strong nonlinear fitting capabilities, with optimization results improving 58% lower power requirements and 20% reduced steel weight compared to the original design. This study concludes that no universal surrogate model suits all ship design challenges.**

***Index Terms*— Deep learning, Ensemble learning, Surrogate model, Ship design, Optimization**

Manuscript received December 20, 2023; revised May 7, 2025. This work is supported by the National Research and Innovation Agency, which has provided financial support for our research through the Electronics and Informatics Research Organisation programme house.

Nanda Yustina is a researcher at the Transport Technology Research Centre, a national innovation research agency in Puspiptek Serpong, Banten, 15314, Indonesia. (corresponding author to provide phone: +62 851-6188-5587 ; (e-mail: nand013@brin.go.id)

Ade Ratih Ispandiari is a researcher at the Transport Technology Research Centre, a national innovation research agency in Puspiptek Serpong, Banten, 15314, Indonesia. (e-mail: ader006@brin.go.id)

Annisa Roschyntawati is a researcher at the Transport Technology Research Centre, a national innovation research agency in Puspiptek Serpong, Banten, 15314, Indonesia. (e-mail: anni006@brin.go.id)

Zulfa Qonita is a researcher at Research Center for Geological Disaster, National Research and Innovation Agency in Sangkuriang Dago Street, Kecamatan Coblong, Kota Coblong, Bandung, West Java, 40135, Indonesia. (e-mail: zulf007@brin.go.id)

Nurul Shabrina is a researcher at the Transport Technology Research Centre, a national innovation research agency in Puspiptek Serpong, Banten, 15314, Indonesia. (e-mail: nuru025@brin.go.id)

Nanda Itohasi Gutami is a researcher at the Transport Technology Research Centre, a national innovation research agency in Puspiptek Serpong, Banten, 15314, Indonesia. (e-mail: nand016@brin.go.id).

Noor Muhammad Ridha Fuadi is a researcher at the Transport Technology Research Centre, a national innovation research agency in Puspiptek Serpong, Banten, 15314, Indonesia. (e-mail: noor014@brin.go.id)

Waluyo is a researcher at the Transport Technology Research Centre, a national innovation research agency in Puspiptek Serpong, Banten, 15314, Indonesia. (e-mail: walu001@brin.go.id)

Siti Sadiah is a researcher at the Transport Technology Research Centre, a national innovation research agency in Puspiptek Serpong, Banten, 15314, Indonesia. (e-mail: siti007@brin.go.id)

Abdul Kadir is a researcher at the Transport Technology Research Centre, a national innovation research agency in Puspiptek Serpong, Banten, 15314, Indonesia. (e-mail: abdu014@brin.go.id)

Zaid Cahya is a researcher at the Electronics Research Centre, a national innovation research agency in Puspiptek Serpong, Banten, 15314, Indonesia. (e-mail: zaid002@brin.go.id)

Mochamad Adityo Rachmadi is a researcher at the Electronics Research Centre, a national innovation research agency in Puspiptek Serpong, Banten, 15314, Indonesia. (e-mail: moch045@brin.go.id)

I. Introduction

Nowadays, maritime transport is forced to undergo structural changes that include low emissions and clean energy (GHG), more robust supply chains and logistics, digitization, and data-driven business models. It must also adapt to changing demand and consumption patterns and increasingly fragmented, localized, or regionalized operating and trading settings. [1]. The COVID-19 pandemic accelerated shifts in consumer attitudes and behaviors, leading to a notable rise in online purchases of consumer goods, many of which are shipped in containers. Global e-commerce accounted for 15% of all retail sales in 2019; by 2021, that percentage had risen to 21% [2]. This trend is expected to continue as the disruption persists, and the growth of e-commerce demands near real-time delivery [3][4]. The increasing competitiveness of the shipping industry, in line with the need for fast logistics due to digitalization, is driving rapid changes in the field of ship design. One is the demand for safer and more environmentally friendly ships with shorter lead times and a lower Energy Efficiency Design Index (EEDI). Ship designers strive to reduce the ship's fit observed power requirements as much as possible to obtain a low EEDI value and reduce the ship's carbon emissions. through weight efficiency while minimizing resistance [5]. Meanwhile, calculating the Total Cost of Ownership (TCO) is also required to calculate economic feasibility when designing a ship. TCO on ships generally consists of

building costs (CAPEX) and operating costs (OPEX). Material costs heavily influence building costs and can be represented in the weight of steel required. Meanwhile, operating costs are also related to the speed of the vessel and the capacity of the cargo being moved, which determines the vessel's power requirements and subsequently affects fuel consumption. Therefore, minimizing the steel weight and the vessel's power requirement can reduce the vessel's TCO [6].

Several studies have been conducted to reduce the power and steel weight requirement. One study successfully determined that the optimized hull form, particularly one incorporating a bulbous design, proves more efficient power and offers superior sea-keeping performance. This finding underscores the significance of minimizing total resistance for fuel efficiency and onboard comfort [7]. In addition, research aimed at reducing ship power has utilized various machine learning methods and modifications to geometric algorithms. A study by [8] states that using machine learning (ML) and principal component analysis (PCA) for hull vane (HV) shape optimization offers efficiency gains. PCA addresses dimensionality challenges, streamlining design space exploration. ML models, like artificial neural networks (ANN), accurately predict resistance even with a small dataset, which is crucial for HV shape optimization. To reduce ship emissions, in parallel with optimizing the hull form to minimize ship resistance, as mentioned earlier, another factor that can be fought for is the steel weight reduction of the ship itself. For the steel weight reduction of the vessel itself.

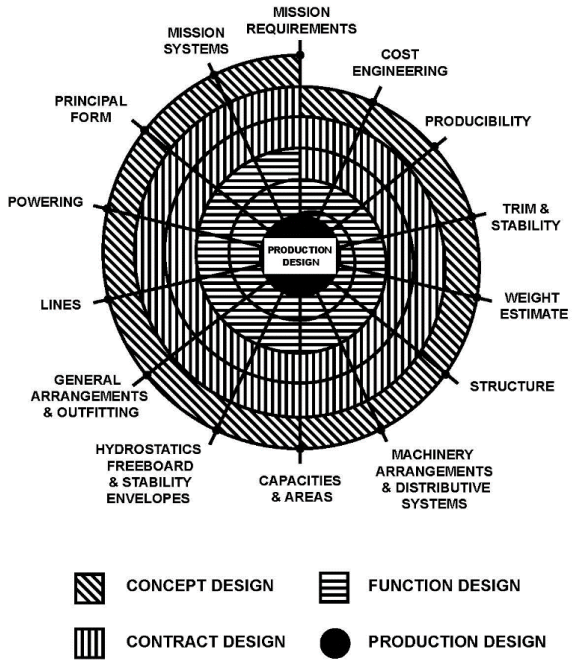

Fig. 1. Spiral concept of ship design.

Regarding methodology, ship design has evolved from an iterative sequential process in the Spiral Design concept [9] to the more recent holistic design approach [10]. In the Spiral Design concept in Fig. 1, ship design components are optimized sequentially in an iterative manner in the design stages. Besides being considered time- and resource-consuming, this approach is often trapped in local optima [11]. Previous studies on design optimization in several engineering design fields have used surrogate models (commonly called surrogate models or metamodels) to replace high-fidelity models [12], [13], [14]. Surrogate models are approximations to describe the relationship between optimization targets, design objectives, constraints, and design variables [15]. The higher the accuracy of the surrogate model, the closer the predicted values for the objective and constraints are to the actual conditions. The surrogate model must be trained to model multiple objective and constraint functions before using the optimization algorithm.

In ship design problems, previous research used Kriging [6], [15], [16] as a surrogate model for the objective function and Radial Basis Function (RBF) [6] as a surrogate model for the constraint function. Some previous research compares several surrogate models individually as surrogate models of objective functions in ship design. Namely, the comparison of Kriging with Radial Basis Function Neural Network (RBFNN) and Support Vector Regression (SVR) [17]. In another study [18], Kriging was compared with Backpropagation – Particle Swarm Optimizer (BPNN-PSO) and Multi-layer Perceptron (MLP). These studies are similar because they use a single machine-learning method as the surrogate model.

Besides surrogate modelling for optimization, machine learning has also been widely utilized to predict specific design parameters in ship design. For instance, Majnarić [19] applied MLP and Gradient Boosting to predict the principal dimensions of container ships, demonstrating the potential of machine learning in early-stage design predictions. Similarly, Cepowski [20] predicted ship-added resistance during the early design phase, emphasizing the role of machine learning in estimating performance metrics.

However, recent research shows that using several surrogate models (ensemble), Polynomial Regression (PR), RBF, and Kriging [21] get better result accuracy than using one surrogate model. This is because the ensemble surrogate model can reduce the generalization error of the prediction [22]. Ensemble surrogate models are used to obtain the advantages of several surrogate models. In this way, the optimization algorithm can use the most suitable surrogate model for different problems, and the adaptability of the surrogate model to the problem can also be improved [23]. Despite these advancements, further analysis is necessary to investigate the effectiveness of surrogate models when comparing ensemble methods with deep-learning approaches. This study focuses primarily on comparing the performance of Deep Learning and ensemble surrogate models to predict power and steel weight requirements in the early stage of ship design. Ensemble methods such as Extra Trees, Gradient Boosting, Weighted Ensemble, and Stacking Regressor are considered, as they are robust for small datasets and can handle imbalanced distributions effectively [24]. Among these, Extra Trees and Gradient Boosting are particularly notable for their high variance reduction and ability to capture non-linear relationships [25]. Stacking Regressor combines the strengths of multiple base models, offering flexibility and improved predictive accuracy.

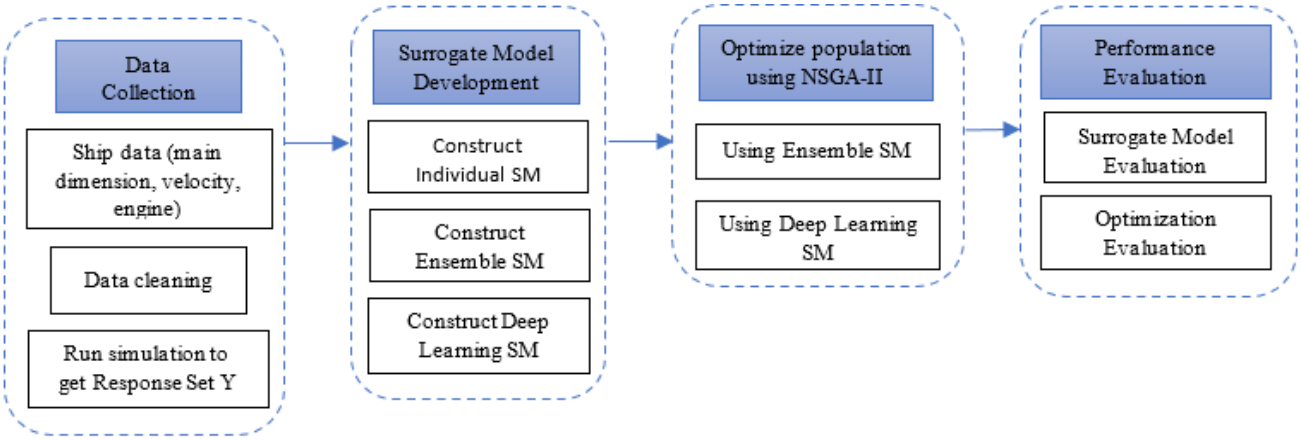

Fig. 2. Research Flow

Machine learning methods such as Kriging, Polynomial Regression (PR), and K-nearest neighbour (KNN) are also used individually as surrogate models. Kriging is widely known for its ability to model spatially correlated data with high precision, while PR is chosen for its simplicity, interpretability, and low computational cost [21], [26]. KNN, on the other hand, effectively captures local patterns within the dataset [27].

Additionally, deep learning models such as Multi-Layer Perceptron (MLP) and Autoencoders are explored as surrogate models. MLP is chosen for its simplicity and ability to model complex relationships in small datasets, while Autoencoders are used to uncover latent representations that improve prediction accuracy [28], [29].

By employing a diverse set of machine learning and deep learning techniques, this study comprehensively compares surrogate modelling approaches for early-stage ship design optimization. Its primary contribution lies in identifying the most effective surrogate models to support and simplify the iterative design process, as framed by the Spiral Design concept, through a holistic optimization strategy. Notably, the study utilizes actual ship performance data—rather than relying solely on design of experiments (DoE) data—to train the surrogate models, enhancing their practical relevance. Through this data-driven evaluation and cross-method comparison, the research offers valuable insights into the strengths, limitations, and suitability of different surrogate modelling techniques for complex engineering design tasks.

II. Research Methods

This research aims to develop the most suitable surrogate model for application in the early stages of ship design. The surrogate model is trained to identify relationships, similarities, and patterns across a large dataset of ship designs. Ship designers define the dependent and independent variables, and the surrogate model learns the functional relationship between them during training. The primary objective is to accurately predict key performance metrics—such as power requirements and steel weight—while satisfying various design constraints. As illustrated in Fig. 2, comparing deep learning and ensemble learning approaches in the context of hull shape optimization involves four key steps.

The research begins with collecting ship-related data, including main dimensions, velocity, and engine specifications. The collected data undergoes a data-cleaning process to ensure quality and consistency. Afterwards, simulations are conducted to generate the response set Y, which serves as the output variables for surrogate modelling. Next is surrogate model development. In this step, three types of surrogate models are constructed:

- Individual Surrogate Models (SM): Machine learning methods such as Kriging, Polynomial Regression, and K-Nearest Neighbours (KNN) are developed independently as surrogate models.
- Ensemble Surrogate Models: Extra Trees, Gradient Boosting, Stacking Regressor and Weighted Ensemble. Stacking and Weighted methods combine multiple surrogate modelling techniques, such as Polynomial Regression, Kriging, and KNN, using weighted aggregation methods to enhance prediction accuracy.
- Deep Learning Surrogate Models: Deep learning architectures like Multi-Layer Perceptrons (MLP) and Autoencoders are developed as alternative surrogate models to predict ship performance metrics.

After the surrogate model is constructed, it is integrated into the optimization process using the Non-dominated Sorting Genetic Algorithm II (NSGA-II). Two optimization approaches are employed. Using Ensemble Surrogate Models to approximate the objective functions. We also use Deep Learning Surrogate Models for the same purpose, allowing a comparative analysis of their optimization performance. The final stage evaluates the performance of the surrogate models and the optimization results. This includes Surrogate Model Evaluation, which assesses prediction accuracy, computational efficiency, and adaptability of each surrogate model. As for Optimization Evaluation, we compare the effectiveness of ensemble and deep learning surrogate models in achieving optimal ship design objectives.

A. Collection of Data

The 456-hull data analyzed in this study were obtained from the International Association of Classification Societies (IACS). The International Association of Classification Societies (IACS) is a technical-based non-

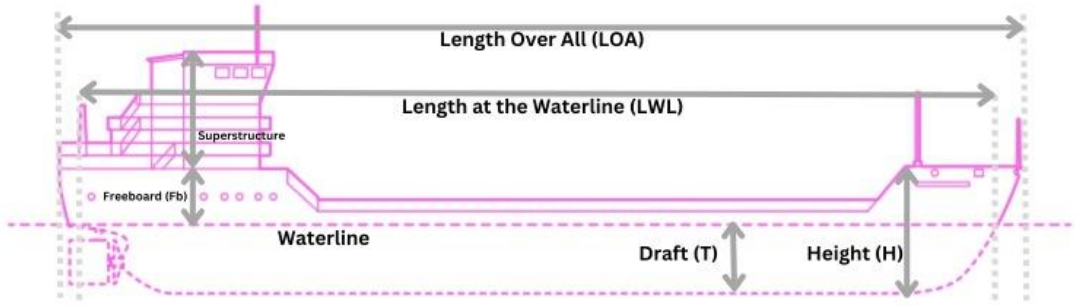

Fig. 4. Main dimensions of the vessel

governmental organization that regulates the strength of ship construction and machinery, quality assurance of marine materials, and supervises the construction, maintenance, and overhaul of ships by classification regulations [30]. The actual ship data collected includes five types of ships: bulk carriers, container ships, tankers, ore carriers, and RoRo ships. The data distribution per ship type is shown in Fig. 3.

The actual data analyzed in this research is taken from 12 IACS members, including the American Bureau of Shipping, Bureau Veritas, Croatian Register of Shipping, China Classification Society, DNV, Indian Register of Shipping, Lloyd's Register, Korean Register of Shipping, Nippon Kaiji Kyokai, Polish Register of Shipping, Registro Italiano Navale, and Türk Loydu [31].

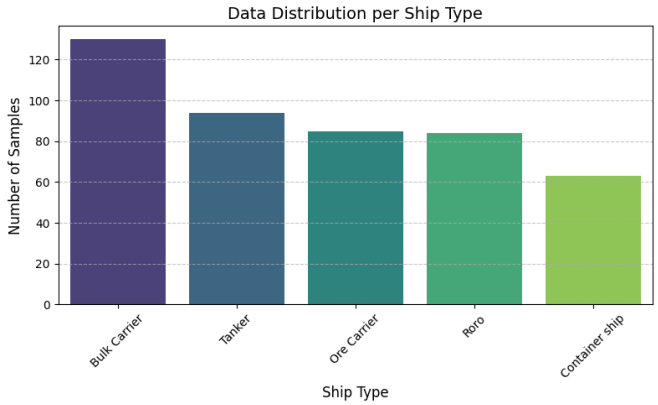

Fig. 3. Data Distribution per Ship Type

The data collected includes the type of vessel, the International Maritime Organisation (IMO) number of the vessel [32], the year the ship was built, LOA (Length Overall), LWL (Length at the Waterline), BOA (Breadth Over All), D (Depth), T (draft), DWT (Dead Weight), GT (Gross Tonnage), V (speed), V_max (maximum speed), P (Power). The actual ship data focuses on vessels constructed after the year 2000. Fig. 4 illustrates the main dimensions of a ship commonly used during the preliminary design phase. These dimensions include the LOA, which measures the vessel's entire length from bow to stern, and the LWL, which specifies the ship's length at the waterline when afloat. The picture also shows the T, which is the vertical distance between the waterline and the lowest point of the hull, indicating how deep the vessel is in the water. Height (H) is the vertical distance from the waterline to the ship's tallest point, sometimes including the superstructure.

B. *Problem Formulation*

Engineering design optimization problems typically have decision variables, objectives, and constraint functions. When optimization has more than one objective, it can be stated as a Multi-objective Problem (MOP), as in Equation 1.

$$\min_{\vec{x}} \vec{f}(\vec{x}) \quad \text{where} : \vec{f}(\vec{x}) = [f_1(\vec{x}), \dots, f_k(\vec{x})] \tag{1}$$

where $\vec{f}$ is a vector function, with k design objectives, and $\vec{x}$ is a decision variable vector. A decision variable vector $\vec{x}$ consists of n variables, $\vec{x} = [x_1, \dots, x_n]$, and is part of the feasible solution $\vec{x} \in \Omega$.

Decision variables in ship design problems are numerical values that can be adjusted during optimization [6]. These variables are usually denoted as x_j, where $j = 1, \dots, n$, where x_j represents one decision variable. The vector $\vec{x}$ is then represented by:

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \tag{2}$$

In this study, the decision variables used are variables related to the main dimensions of the ship: Length (L), Breadth (B), Height (H), Draught (T), and Velocity (V).
Constraints are frequently viewed as hard goal functions that must be met before minimizing the remaining soft objective functions. The feasible region Ω contains all solutions that satisfy some given constraints, including the bounds of variables and inequality constraints [33]:

$$x_i^{lower} \le x_i \le x_i^{upper} \qquad i = \{1, \dots, n\} \tag{3}$$

$$g_i(\vec{x}) \le c_i \ \vee \ g_i(\vec{x}) \ge c_i \quad i = \{1, \dots, m\} \tag{4}$$

The search space Ω represents all potential combinations between a preset lower and upper bound of $\vec{x}$. These upper and lower bounds were obtained from the ship dataset, as in Table 1. The table also defines the constraints that limit the boundaries of the ship design related to the main dimensions of the ship used in this study [34]. Due to the constraints, not all combinations will result in a feasible solution.

The objectives in the ship design optimization problem are usually contradictory. Consequently, there is hardly a perfect solution, but rather a group of alternatives known as non-dominated solutions. This ship design problem has two objectives: minimizing the effective power requirement

(EP) and the weight of the ship's steel (Ws).

TABLE I
CONSTRAINTS USED IN THIS STUDY

Constraint	Range
Variable Limits	$23 \le LOA \le 400$
	$22 \le LOA \le 399$
	$6 \le B \le 65$
	$3 \le D \le 62$
	$2 \le T \le 25$
	$9 \le V \le 28$
Froude Number	$Fn \le 0.32$
L/B	$3.5 \le L/B \le 10$
B/T	$1.8 \le B/T \le 5$
L/T	$10 \le L/T \le 30$
L/D	$L/D \le 15$
DWT	$30 \le DWT \le 403{,}000$
Type of ship	*Bulk Carrier, Container Ship, Tanker, Ore Carrier, and RoRo Ship*

C. Surrogate Model

Before we train the surrogate model, we acquire a response set Y, for each data point using manual simulations. A manual simulation is a thorough computation carried out by a ship designer manually. As previously mentioned in the spiral design concept, this manual simulation typically takes a long time to complete. It takes a naval architect days to gather information and run manual simulations for preliminary computations. The collected data (decision variables and associated response set) is then trained independently using five distinct surrogate models: Deep Learning, PR, Kriging, KNN, and an ensemble of PR-Kriging-KNN. three distinct types of surrogate models:

- Individual Surrogate Models (SM): Kriging, Polynomial Regression, and K-Nearest Neighbors (KNN).
- Ensemble Surrogate Models: Extra Trees, Gradient Boosting, Weighted Ensemble Method, and Stacking Regressor combine 3 surrogate models: Polynomial Regression, Kriging, and KNN.
- Deep Learning Surrogate Models: Deep learning architectures like Multi-Layer Perceptrons (MLP) and Auto-Encoders

Polynomial Regression (PR)

Polynomial Regression - Polynomial regression is a technique that can model the relationship between multiple independent variables (X and Y) to the dependent variable (Z) through a non-linear relationship Click or tap here to enter text.[35]. The following is equation 1 of polynomial regression:

$$y = c_0 + c_1 x + c_1 x^2 + \cdots + c_n x^n + + e \qquad (5)$$

where c is the coefficient set, n is the polynomial degree, and e is the unobserved random error.

Kriging

A widely recognized surrogate model employed to approximate computationally intensive functions is the Kriging model. The Kriging approach was initially introduced by Daniel G. Krige [36]. Later, the Kriging methodology was also called the stochastic process model for designing and analyzing computer experiments (DACE). Kriging - is a supervised learning method that combines polynomial model x i.e. $f(x)$ and localized deviation x i.e. $Z(x)$ [37] as in equation 2.

$$y(x) = f(x) + Z(x) \qquad (6)$$

where $Z(x)$ is a normally distributed Gaussian random process with zero mean and non-zero covariance. In the above equation, $f(x)$ is the polynomial function of the RSM, which gives the 'global' model and is fixed.

K-nearest neighbour (KNN)

The fundamental concept underlying nearest neighbour methods is approaching the training set as the model and predicting new points according to their proximity to those in the training set [38]. Choosing the K value, distance metric, and decision rule are three fundamental hyperparameters of the KNN method, a basic classification and regression technique [39].

In the given training set of $q_1, q_2, \ldots, q_n$, and the corresponding target values $y_1, y_2, \ldots, y_n$ (where "y^{i}" represents any feature in yi), the prediction in a new point q^p can be achieved by initially locating a set of K nearest neighbors in the training set. Subsequently, that $\widehat{y^p}$ is calculated as a weighted average of K nearest neighbours [40].

$$\widehat{y^p} = \sum_{i=1}^{k} w_i y^i \qquad (7)$$

Where w_i represents the weight of the $i-th$ neighbour, it should be noted that the greater the weight assigned to a point, the closer it is to the predicted point. w_i is defined as

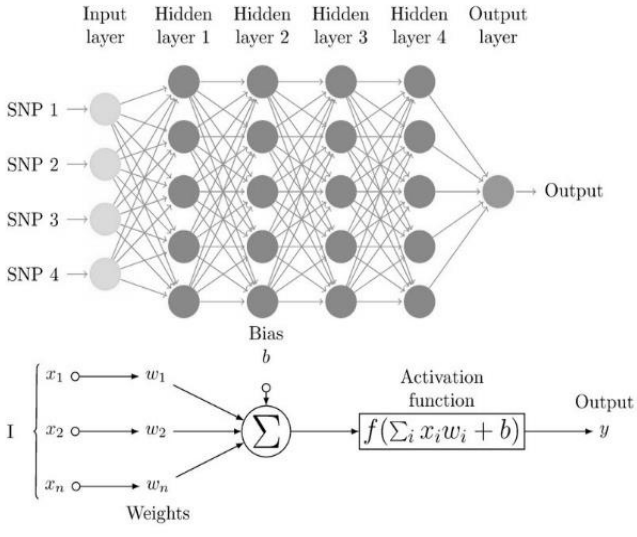

Fig. 5. Multi-layer perceptron architecture

$$w_i = \frac{1/L(q^i,q^j)}{\sum_{j=1}^{k} 1/L(q^j,q^p)} \quad (8)$$

where L represents a distance function, denoted as $L1$ or $L2$.

$$L1(q^i,q^j) = \sum_{k=1}^{4}[q^j - q^j] \quad (9)$$

$$or\ \ L2(q^i,q^j) = \sqrt{\sum_{k=1}^{4}(q^j - q^j)^2} \quad (10)$$

Multi-Layer Perceptron (MLP)

Deep learning used in this study is fully connected neural networks or multi-layer perceptron (MLP) and Autoencoders. MLP is frequently employed for applications involving regression and classification. Except for the possibility of several hidden layers, the architecture of MLP is comparable to conventional neural networks, as in Fig. 5 [41]. Every node links input (x) to outputs (y); it then adds up the inputs, applies the weighting factors ($w_{i,j}$), and modifies the data. The network's weights are then updated via the backpropagation technique [18]. The final weights are defined as an approximate output following training. We employ five hidden layers for MLP in this work, with two output layers and 64 hidden nodes for each hidden layer.

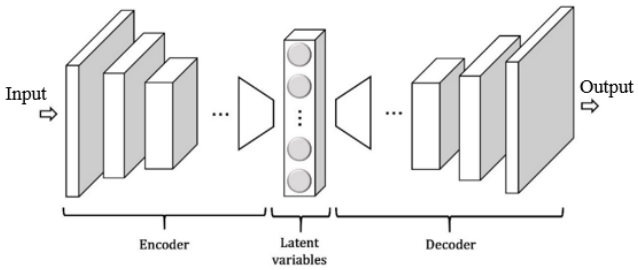

Fig. 6. Auto-encoder architecture

Auto-Encoders

The concept of Auto-Encoders (AEs) was first introduced by Rumelhart et al. [42] as a neural network designed to learn from unlabeled datasets unsupervised. An AE consists of two main components: an encoder and a decoder, as shown in Fig. 6 [29]. The encoder compresses the input data into a reduced representation, known as the encoding, while the decoder reconstructs the original input from this encoded representation with minimal error. In between, there are latent variables, a compact bottleneck representation of the input data containing its critical information. The primary goal of an AE is to achieve this reconstruction with the lowest possible loss, typically measured as the mean squared error between the input data x_i and its reconstructed output $\tilde{x}_i$.

Extra Trees

The Extra Tree Regression (ETR) algorithm, introduced by Geurts et al.,[43] is an ensemble method derived from the Random Forest (RF) model. Like RF, it constructs multiple decision or regression trees but introduces greater randomization. Unlike RF, which searches for optimal split thresholds, ETR selects split points randomly [44]. Additionally, ETR typically uses the entire dataset to train each tree instead of relying on bootstrapping. For regression tasks, ETR predicts by averaging the outputs of all trees, making it computationally efficient and robust to overfitting [45].

Gradient Boosting

Gradient Boosting (GB) is an ensemble learning method that builds a strong predictor by combining multiple weak learners, typically decision trees. It optimizes a cost function by iteratively fitting new models to the residual errors of previous ones, effectively minimizing prediction errors using functional gradient descent [19], [24]. Due to its robustness and interpretability, Gradient Boosting is widely used in both regression and classification tasks, making it a powerful tool for various machine-learning applications [46].

Stacking Regressor

Stacking regression is an ensemble learning method combining multiple base models (referred to as "level 0" learners) to enhance predictive performance. Base learners can include algorithms such as linear regression, support vector machines, gradient boosting, and random forests [47]. While each base model may focus on different features or patterns in the data, the stacking framework combines their predictions into a more accurate and robust final output. However, the effectiveness of stacking depends on the diversity and quality of the base learners, as redundant or overly noisy models can harm overall accuracy rather than improve it [48].

Weighted Ensemble

Ensemble methods are often developed to overcome the shortcomings of individual surrogate models. Several ensemble surrogate model methods have been developed, often based on error correlation or prediction variance [21]. For example, [49] used errors from cross-validation (CV) to calculate ensemble weights and proposed an ensemble surrogate model approach for single objective optimization, also using Root Mean Square Error (RMSE) to calculate ensemble weights. The methods used are PR, KNN, and Kriging. The final output of the ensemble surrogate model is obtained using:

$$\hat{f}_{ens}(x) = w_1\hat{f}_1 + w_2\hat{f}_2 + w_3\hat{f}_3 \quad (6)$$

is where $\hat{f}_i(x)$ is i-th output from each member, and w_i is a weight from $i-th$ output which defined as:

$$w_i = 0.5 - \frac{e_i}{2(e_1+e_2+e_3)} \quad (7)$$

Where e_1 is *Root Mean Square Error* (RMSE) from i-th model, and later calculated by weighted aggregation method.

D. Optimization

Python library PyMOO [41]is used for the optimization algorithm (NSGA-II). Hyperparameters used include: population size N = 100, number of generations τ = 300, SBX crossover probability pc = 0.95, polynomial mutation probability pm = 1/n (where n is the number of variables), For Kriging, the parameter used is Kernel, which defines the covariance function that determines the shape of the

Gaussian process before and after. This study used ConstantKernel from sklearn with a value = 1.0 and bounds = (1e-3, 1e3). For polynomial regression, a degree = 3 was used to determine the degree of power.

E. Performance Evaluation

This section will evaluate the efficacy of the several surrogate models discussed in this research. First, the relationship between the input parameters and the response was examined by analyzing the local sensitivity curves. Local sensitivity analysis is computationally more efficient since it concentrates on a particular area of parameter space or set of variables. This localized method addresses specific questions that necessitate thoroughly comprehending the model's behavior in a given area. It makes it possible to do focused research on how minor adjustments to input parameters impact model outcomes in that area [50]. Additionally, local sensitivity analysis helps in factor prioritization by confirming or measuring the impact of each component. This strategy narrows the scope of the investigation and saves time and resources, especially when computational power is limited.

Then, two prevalent model error metrics are employed to gauge the accuracy of the models: Maximum Absolute Error (MAE) and Root Mean Square Error (RMSE) [51]. R-squared (R^2), mean absolute error (MAE), and root mean squared error (RMSE) are quantitative performance metrics utilized in regression modelling. In predicting future instances of the output, the model demonstrates greater precision as the MAE decreases. The standard deviation of the response variable constitutes the RMSE. R^2 values vary between 0 and 1, with 1 indicating an optimal fit and 0 indicating that there is no advantage to employing the model compared to solely utilizing fixed background response rates. The one with the highest R^2 and the lowest RMSE and MAE is favoured when comparing models [52]. MAE model criteria in equation 8, RMSE criteria in equation 9, and R^2 criteria in equation 10 [51], [52], [53].

$$\frac{1}{n} = \sum_{i=1}^{n}[yi - \hat{y}i] \tag{8}$$

$$\sqrt{\sum_{i=1}^{n}[yi - \hat{y}i]^2 / n} \tag{9}$$

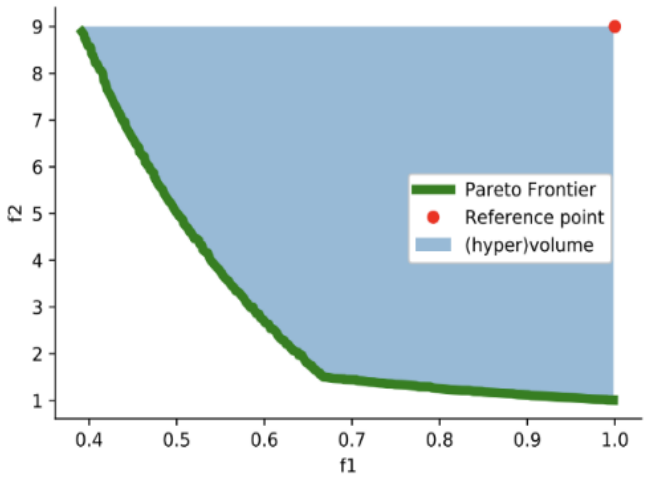

Fig. 7. Hypervolume ilustration

$$1 - \frac{\sum_{i=1}^{n}[yi - \hat{y}i]^2}{\sum_{i=1}^{n}[yi - \bar{y}i]^2} \tag{10}$$

Where MAE represents mean absolute error; RMSE denotes mean square error; n represents the number of datasets in the dataset; yi signifies the actual value of the output variable; $\hat{y}i$ signifies the predicted value of the output variable; and $\hat{y}i$ represents that age value of all output variables in the dataset.

After obtaining the ensemble model and DNN, data generation and evolution were conducted using NSGA-II as

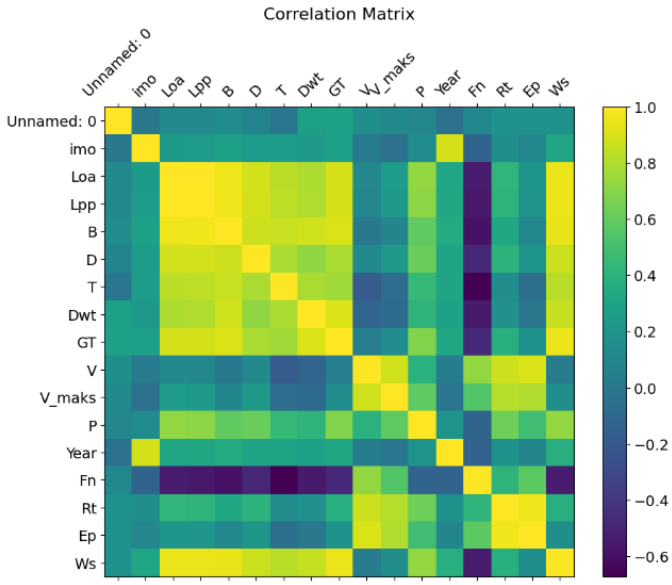

Fig. 8. Correlation matrix

the optimization algorithm. This optimization was performed by comparing each of the two surrogate models. Subsequently, the hypervolume values were calculated. The Hypervolume (HV) metric is a tool for evaluating optimization outcomes. It represents the volume enclosed between a fixed reference point and the Pareto front, offering a measure of solution quality in multi-objective problems where solutions are not dominated by others (Pareto fronts), as depicted in Fig. 7. A Reference Point is utilized, derived from a fixed point that dominates all points in the effectiveness set [6][20]. Generally, a higher HV suggests superior algorithm performance for the given problem.

In addition to HV, the Pareto front's dispersion is considered. This dispersion refers to how points are distributed along Pareto front approximations. A shorter distance between points in these approximations suggests a more evenly distributed set of solutions [54].

III. Results and Discussion

This section discusses the experiment results, from the sensitivity analysis and surrogate model accuracy.

A. Sensitivity Analysis

A correlation matrix was built to see the reaction and how it is related to other parameters. Stated differently, it is necessary to ascertain the factors that exhibit both strong and weak correlations with the design response. For this objective, all data were employed to filter along the correlation value using Pearson's rank correlation [46]. The

viewer may observe which factors have the strongest association with the response variable by looking at the correlation matrix in Fig. 8. It visually depicts the correlations. The factors that should be prioritized for optimization or additional research can be determined using the information provided. Each cell represents the correlation coefficient between this matrix's corresponding row and column variables. The outcome was a colour-coded matrix that showed how strongly each component and the response variable correlated. The colour-coded matrix resulting from this analysis indicated the strength of the correlation between each factor and the response variable. A stronger relationship is expected when the value is closer to the absolute value of 1.

According to the matrix, LOA, LPP, B, D, T, DWT, and GT most influence Ws with a correlation value above 0.8. These same parameters also significantly influence EP. The year the ship was built shows varying degrees of correlation with other variables. This could reflect changes in ship design standards and technology advancements over time. The Froude number (Fn) has an interesting pattern of correlations. It shows a strong negative correlation with some dimensions, such as D and B. This is important for hull design, as it influences the hydrodynamic efficiency of the vessel. Rt (total resistance) and EP have moderate correlations with design parameters, reflecting the impact of design choices on the ship's operational efficiency.

Strong positive correlations exist among the ship's main dimensions, such as LOA, LPP, and B. This indicates that as one of the dimensions increases, the others also tend to increase. There is a noticeable positive correlation between V and EP. This is intuitive as higher speeds generally require more power. Furthermore, the correlation between the Froude number and other design parameters highlights the importance of hydrodynamic considerations in achieving efficient ship designs. By minimizing resistance through optimized hull shapes, designers can significantly enhance the ship's overall performance and reduce environmental impact

B. Surrogate Model Accuracy

The dataset, which has 456 data, is divided 70:30 between training and test data. Table II presents the performance comparison of various surrogate models evaluated based on three metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). The models include Polynomial Regression (PR), Kriging, K-nearest neighbours (KNN), Extra Trees, Gradient Boosting, Stacking Regressor, Weighted Ensemble, Multi-layer Perceptron (MLP), and Autoencoders (AE). Among the models, PR performs the best, achieving the lowest MAE (102.19) and RMSE (248.34) along with the highest R^2 value of 0.99, indicating its superior accuracy and reliability in predictions.

The Stacking Regressor and Weighted Ensemble also perform well, with R^2 values of 0.989 and 0.97, respectively, suggesting the effectiveness of combining multiple models to enhance prediction accuracy. On the other hand, KNN shows the weakest performance, with the highest MAE (702.24) and RMSE (1236.11), along with an R^2 of 0.90, indicating that it may not be suitable for this particular

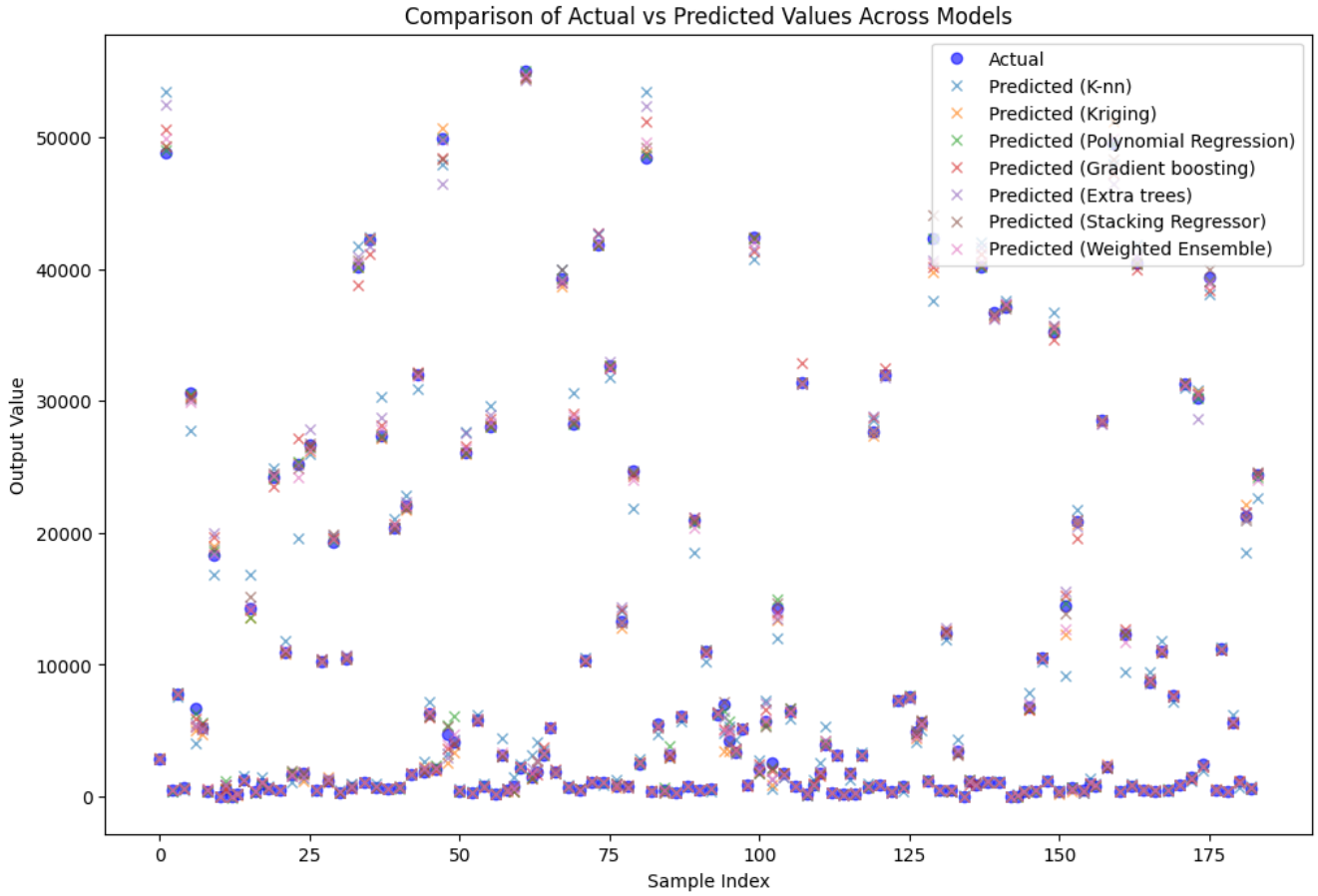

Fig. 9. Comparison Actual vs Predicted Values for Individual and Ensemble Surrogate Model

problem. Other models, such as Gradient Boosting, Extra Trees, and Kriging, achieve moderate performance, with R^2 values ranging from 0.91 to 0.976, showing that they can capture underlying patterns but with some limitations in accuracy.

TABLE II
SURROGATE MODEL PERFORMANCE RESULTS

Surrogate Model	Test Score		
	MAE	RMSE	R^2
PR	**102.19**	**248.34**	**0.99**
Kriging	199.75	496.65	0.91
KNN	702.24	1236.11	0.90
Extra Trees	301.06	686.11	0.97
Gradient Boosting	272.19	544.10	0.976
Stacking Regressor	145.85	297.68	0.989
Weighted Ensemble	181.36	362.62	0.97
MLP	570.05	1108.74	0.986
AE	315.07	622.94	0.984

Interestingly, the MLP and AE models achieve competitive R^2 values of 0.986 and 0.984, respectively, highlighting their effectiveness in handling complex relationships in the data. However, their RMSE and MAE values indicate they are slightly less accurate than PR and the ensemble methods. Overall, the results demonstrate that ensemble models and Polynomial Regression are more suitable for this task, offering a balance between accuracy and reliability. These findings align with prior research indicating the robustness of ensemble approaches and polynomial methods in surrogate modelling tasks.

Higher R^2 often denotes better accuracy and Quality of Fit. As noted in Table II, the PR Quality of Fit (e.g., R^2 = 0.999) is stronger, so the PR model is preferable. Simply put, the Quality of Fit measures how well the observation points are fitted statistically based on the surrogate model coefficients. This technique is called optimization using local surrogate models. In the meantime, cross-validation and Root Mean Square Error (RMSE) can be used to evaluate the interpolation model to provide a more precise approximation over the design points. It also addressed this argument, and the technique used was called global surrogate model-based optimization. The quality of fit is higher when the RMSE value is closer to zero. Since the error is caused by model inadequacies rather than noise, the present error estimates were derived using bias error rather than noise.

The performance results from Table II align well with the visualization in the comparison plot of actual versus predicted values across different surrogate models as shown in Fig 9. The plot illustrates the predictive accuracy of each model by showing how closely its predictions align with the actual data points.

Polynomial Regression (PR), which achieved the best scores in Table II, shows the plot's closest alignment between predicted and actual values. This consistency reinforces its superior ability to capture the underlying patterns in the dataset accurately. Similarly, the Stacking Regressor and Weighted Ensemble models, which also performed well in the tabular metrics (R^2: 0.989 and 0.97, respectively), demonstrate tightly clustered predictions around the actual values, indicating their robustness.

In contrast, KNN, which showed the weakest performance in Table II (MAE: 702.24, RMSE: 1236.11, R^2: 0.90), is evident in the plot as having larger deviations between predicted and actual values. This discrepancy highlights its limited suitability for the problem.

Models such as Gradient Boosting, Extra Trees, MLP, and Autoencoders exhibit moderate performance in both the table and the plot. While their predictions are reasonably close to the actual values, they display slightly higher variance compared to PR and the ensemble methods. Combining quantitative metrics and visual comparisons provides a comprehensive understanding of model performance. It confirms that models like PR and ensemble methods achieve low error metrics and produce predictions that are highly consistent with the actual data.

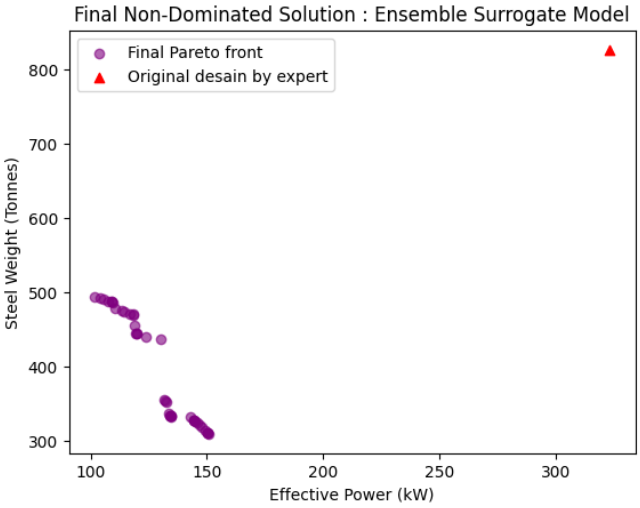

Fig. 10. Optimization using Ensemble Surrogate Model

D. Optimization

After evaluating the surrogate model, it will be used in design optimization. The challenge is that a surrogate model cannot produce reliable predictions without an adequate supply of high-quality data. Error-free data is crucial because the surrogate model will pick up the incorrect trend without it and produce inaccurate predictions. So, the use of actual ship data is preferred for the problems in this research. Table II presents the optimization results,

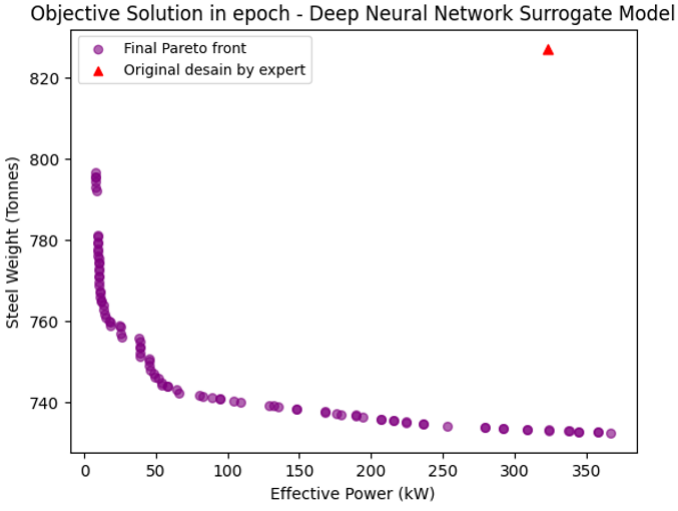

Fig. 11. Optimization using MLP Surrogate Model

comparing two surrogate models utilizing NSGA-II. These findings suggest that the ensemble yields the highest hypervolume value and shorter computational time compared to MLP.

TABLE III
OPTIMIZATION PERFORMANCE RESULTS

Surrogate Model	Score	
	Hypervolume	Time (seconds)
MLP	568.54	180.45
Ensemble	625.12	57.20

Figs. 10 and 11 illustrate the objective solution with the Pareto fronts visualized for the ensemble and MLP surrogate model, respectively. In Fig. 11, utilizing the MLP surrogate model with NSGA-II results in closer distances between points in the Pareto front than the ensemble surrogate model. On the other hand, the objective values $f1$ and $f2$ in these figures show a contradictory relationship. Specifically, when the solution achieves the minimum value for the objective function $f1$, the value of the objective function $f2$ is observed to be higher and conversely. Furthermore, upon comparing the objective values with the expert's original design, both optimization outcome reveals the most notable reductions in objectives within their respective Pareto front design variations, highlighted in red in Fig. 10.

Consequently, the objective values for ensemble surrogate models utilizing NSGA-II lead to enhancements in design. The MLP with NSGA-II results in 58% lower power requirements and 20% less steel weight compared to the original design. Meanwhile, ensemble SM with NSGA-II results in approximately 52% lower power requirements and 40% less steel weight. Despite the ensemble SM resulting in higher hypervolume values, it has a different dispersion of the Pareto front and lower model accuracy performance (MAE) compared to MLP. This divergence may be attributed to model errors influencing the attainment of the lowest objective value, consequently impacting the hypervolume metric. Thus, hypervolume alone may not be the sole performance metric for optimizing the surrogate model approach

IV. CONCLUSION

This study evaluated three categories of surrogate modelling techniques—individual surrogate models, ensemble surrogate models, and deep learning surrogate models—to determine their effectiveness in optimizing early-stage ship design. The goal was to identify the most suitable approach for establishing functional relationships between design variables and performance responses.

Key findings revealed distinct performance hierarchies among the methods. Polynomial Regression (PR) emerged as the top-performing individual surrogate model, achieving the lowest prediction errors (MAE: 106.75; RMSE: 320.31) and the highest accuracy (R^2: 0.999). Its simplicity and interpretability make it particularly effective for this application.

Ensemble surrogate models, combining PR, Kriging, and KNN via weighted aggregation and stacking regressors, secured second place. These models demonstrated strong predictive capabilities by leveraging the complementary strengths of their constituent methods. Notably, ensemble approaches achieved a higher hypervolume in optimization tasks than deep learning models, yielding practical design improvements such as reduced power requirements and lower steel weight compared to the original design.

Deep learning models ranked third, including MLP and autoencoders (AE). While less accurate than PR and ensemble models, they outperformed standalone Kriging and KNN methods and proved robust in capturing complex nonlinear relationships between variables.

The results underscore that no single surrogate modelling method is universally optimal for ship design optimization. Instead, the choice depends on the problem's characteristics and the design space. PR and ensemble models explore response surfaces and efficiently maximize performance metrics. Meanwhile, despite their computational complexity, deep learning models are better suited for scenarios requiring precise approximation of intricate functional relationships.

This study emphasizes the importance of tailoring the surrogate modelling approach to the specific demands of the optimization task. A thorough evaluation of model strengths, computational costs, and problem complexity is critical for identifying the most effective strategy. Future work could explore hybrid frameworks that integrate the interpretability of PR, the flexibility of ensemble methods, and the nonlinear modelling power of deep learning to further advance ship design optimization.

ACKNOWLEDGMENT

We would like to express our deepest gratitude to Dr. Ir. Iskendar, MS., for his help in problem identification as naval experts, and their assistance in data acquisition.

REFERENCES

[1] United Nations, *Review of Maritime Transport 2023*. in United Nations Conference on Trade and Development, Geneva, no. 2023. Geneva: United Nations, 2023.

[2] IMO, "Marine Environmental Protection Committee." 2022. Accessed: Dec. 18, 2023. [Online]. Available: https://wwwcdn.imo.org/localresources/en/OurWork/Environment/Documents/Air%20pollution/MEPC.352(78).pdf

[3] ILO, "Trends 2022: ILO Flagship Report - World Employment and Social Outlook." Accessed: Dec. 08, 2023. [Online]. Available: https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@dgreports/@dcomm/@publ/documents/publication/wcms_834081.pdf

[4] UNCTAD, "Fast Tracking Implementation of eTrade Readiness Assessments - Second Edition." Accessed: Dec. 06, 2023. [Online]. Available: https://unctad.org/system/files/official-document/dtlstict2022d5_en.pdf

[5] S. Chakraborty, "How The Power Requirement Of A Ship Is Estimated?," 2019. Accessed: Feb. 05, 2021. [Online]. Available: https://www.marineinsight.com/naval-architecture/power-requirement-ship-estimated/

[6] R. de Winter, B. van Stein, M. Dijkman, and T. Bäck, "Designing Ships Using Constrained Multi-Objective Efficient Global Optimization," Springer International Publishing, Cham, 2019. doi: 10.1007/978-3-030-13709-0_16.

[7] M.-P. B. James, A. J. Chuku, and C. U. Orji, "HULL FORM OPTIMISATION FOR IMPROVED HYDRODYNAMIC PERFORMANCE," *International Research Journal of Engineering and Technology (IRJET)*, vol. 08, no. 09, 2021.

[8] C. Çelik, D. B. Danışman, S. Khan, and P. Kaklis, "A reduced order data-driven method for resistance prediction and shape optimization

of hull vane," *Ocean Engineering*, vol. 235, p. 109406, Sep. 2021, doi: 10.1016/j.oceaneng.2021.109406.

[9] J. H. Evans, "Basic Design Concepts," *Journal of the American Society for Naval Engineers*, vol. 71, no. 4, pp. 671–678, Nov. 1959, doi: 10.1111/j.1559-3584.1959.tb01836.x.

[10] A. D. Papanikolaou, "Holistic Approach to Ship Design," *JMSE*, vol. 10, no. 11, p. 1717, Nov. 2022, doi: 10.3390/jmse10111717.

[11] R. de Winter, J. Furustam, T. Bäck, and T. Muller, "Optimizing Ships Using the Holistic Accelerated Concept Design Methodology," in *Lecture Notes in Civil Engineering*, Springer Science and Business Media Deutschland GmbH, 2021, pp. 38–50. doi: 10.1007/978-981-15-4680-8_3.

[12] W. Qin, J. Dong, M. Wang, Y. Li, and S. Wang, "Fast Antenna Design Using Multi-Objective Evolutionary Algorithms and Artificial Neural Networks," in *2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE)*, Hangzhou, China: IEEE, Dec. 2018, pp. 1–3. doi: 10.1109/ISAPE.2018.8634075.

[13] A. Whyte and G. Parks, "Surrogate Model Optimization of a 'Micro Core' PWR Fuel Assembly Arrangement Using Deep Learning Models," *PHYSOR2020 – International Conference on Physics of Reactors: Transition to a Scalable Nuclear Future*, vol. 247, p. 8, 2020, doi: 10.1051/epjconf/202124712003.

[14] S. de Lucas, A. Velazquez, and J. M. Vega, "An Optimization Method for an Aircraft Rear-end Conceptual Design Based on Surrogate Models," Lecture Notes in Engineering and Computer Science: *Proceedings of the World Congress on Engineering 2011*, WCE 2011, July 6 - 8, 2011, London, U.K [Online]. Available: https://www.iaeng.org/publication/WCE2011/WCE2011_pp2610-2615.pdf

[15] A. Nazemian and P. Ghadimi, "CFD-based optimization of a displacement trimaran hull for improving its calm water and wavy condition resistance," *Applied Ocean Research*, vol. 113, p. 102729, Aug. 2021, doi: 10.1016/j.apor.2021.102729.

[16] A. Priftis, E. Boulougouris, O. Turan, and G. Atzampos, "Multi-objective robust early stage ship design optimisation under uncertainty utilising surrogate models," *Ocean Engineering*, vol. 197, p. 106850, Feb. 2020, doi: 10.1016/j.oceaneng.2019.106850.

[17] P. Wang, Z. Chen, and Y. Feng, "Many-objective optimization for a deep-sea aquaculture vessel based on an improved RBF neural network surrogate model," *J Mar Sci Technol*, vol. 26, no. 2, pp. 582–605, Jun. 2021, doi: 10.1007/s00773-020-00756-z.

[18] N. Yustina and A. Saptawijaya, "Surrogate Model-based Multi-Objective Optimization in Early Stages of Ship Design," *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 6, no. 5, pp. 782–789, Oct. 2022, doi: 10.29207/resti.v6i5.4248.

[19] D. Majnarić, S. Baressi Šegota, I. Lorencin, and Z. Car, "Prediction of main particulars of container ships using artificial intelligence algorithms," *Ocean Engineering*, vol. 265, p. 112571, Dec. 2022, doi: 10.1016/j.oceaneng.2022.112571.

[20] T. Cepowski, "The prediction of ship added resistance at the preliminary design stage by the use of an artificial neural network," *Ocean Engineering 195 (2020) 106657*.

[21] H. Wang, Y. Jin, and J. Doherty, "Committee-Based Active Learning for Surrogate-Assisted Particle Swarm Optimization of Expensive Problems," *IEEE TRANSACTIONS ON CYBERNETICS*, vol. 47, no. 9, pp. 2664–2677, 2017, doi: doi:10.1109/tcyb.2017.2710978.

[22] V. Kotu and B. Deshpande, *Data Science : Concept and Practice*, 2nd ed. Elsevier, 2019. [Online]. Available: https://doi.org/10.1016/C2017-0-02113-4

[23] Y. Sun, J. Wang, and Z. Lu, "Asynchronous Parallel Surrogate Optimization Algorithm Based on Ensemble Surrogating Model and Stochastic Response Surface Method," in *2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, IEEE, May 2019, pp. 74–84. doi: 10.1109/BigDataSecurity-HPSC-IDS.2019.00024.

[24] W. Li, W. Wang, and W. Huo, "RegBoost: a gradient boosted multivariate regression algorithm," *IJCS*, vol. 4, no. 1, pp. 60–72, Jan. 2020, doi: 10.1108/IJCS-10-2019-0029.

[25] D. A. Otchere, T. O. A. Ganat, J. O. Ojero, B. N. Tackie-Otoo, and M. Y. Taki, "Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions," *Journal of Petroleum Science and Engineering*, vol. 208, p. 109244, Jan. 2022, doi: 10.1016/j.petrol.2021.109244.

[26] M. Shi, Q. Geng, L. Lv, and S. L. Zhang, "A decision tree-assisted polynomial regression model with application in the cutting force analysis of cutters of a tunnel boring machine," *Engineering Optimization*, vol. 55, no. 5, pp. 823–840, 2023, doi: 10.1080/0305215X.2022.2039131.

[27] R. Goyal, P. Chandra, and Y. Singh, "Suitability of KNN Regression in the Development of Interaction based Software Fault Prediction Models," *IERI Procedia*, vol. 6, pp. 15–21, 2014, doi: 10.1016/j.ieri.2014.03.004.

[28] S. Nikolopoulos, I. Kalogeris, and V. Papadopoulos, "Non-intrusive surrogate modeling for parametrized time-dependent partial differential equations using convolutional autoencoders," *Engineering Applications of Artificial Intelligence*, vol. 109, p. 104652, Mar. 2022, doi: 10.1016/j.engappai.2021.104652.

[29] N. Wang, H. Chang, and D. Zhang, "Theory-guided Auto-Encoder for surrogate construction and inverse modeling," *Computer Methods in Applied Mechanics and Engineering*, vol. 385, p. 114037, Nov. 2021, doi: 10.1016/j.cma.2021.114037.

[30] IACS, "What are classification societies?," 2004.

[31] IACS, "IACS Members," *The International Association of Classification Societies*. Accessed: Dec. 17, 2023. [Online]. Available: https://iacs.org.uk/membership/iacs-members

[32] IMO, "IMO identification number schemes." Accessed: Dec. 17, 2023. [Online]. Available: https://www.imo.org/en/ourwork/msas/pages/imo-identification-number-scheme.aspx

[33] T. Peter, "Using Deep Learning as a surrogate model in Multi-objective Evolutionary Algorithms," Otto-von-Guericke-Universität, Magdeburg, 2018.

[34] A. Charchalis, "Determination of Main Dimensions and Estimation of Propulsion Power of A Ship," *Journal of KONES Powertrain and Transport*, vol. 21, no. 2, 2014, doi: 10.5604/12314005.1133863.

[35] N. K. Lankton, D. H. McKnight, R. T. Wright, and J. B. Thatcher, "Research Note—Using Expectation Disconfirmation Theory and Polynomial Modeling to Understand Trust in Technology," *Information Systems Research*, vol. 27, no. 1, pp. 197–213, Mar. 2016, doi: 10.1287/isre.2015.0611.

[36] D. G. Krige, "A statistical approach to some basic mine valuation problems on the Witwatersrand," *Journal of The Chemical, Metallurgical, and Mining Society of South Africa*, vol. 52 No.6, 1951, [Online]. Available: https://hdl.handle.net/10520/AJA0038223X_4792

[37] L. Zhang, Z. Lu, and P. Wang, "Efficient structural reliability analysis method based on advanced Kriging model," *Applied Mathematical Modelling*, vol. 39, no. 2, pp. 781–793, Jan. 2015, doi: 10.1016/j.apm.2014.07.008.

[38] L. Cai, L. Ren, Y. Wang, W. Xie, G. Zhu, and H. Gao, "Surrogate models based on machine learning methods for parameter estimation of left ventricular myocardium," *R. Soc. open sci.*, vol. 8, no. 1, p. 201121, Jan. 2021, doi: 10.1098/rsos.201121.

[39] Y. Liang, Y. Pan, X. Yuan, W. Jia, and Z. Huang, "Surrogate modeling for long-term and high-resolution prediction of building thermal load with a metric-optimized KNN algorithm," *Energy and Built Environment*, vol. 4, no. 6, pp. 709–724, Dec. 2023, doi: 10.1016/j.enbenv.2022.06.008.

[40] H. Tong, C. Huang, L. L. Minku, and X. Yao, "Surrogate models in evolutionary single-objective optimization: A new taxonomy and experimental study," *Information Sciences*, vol. 562, pp. 414–437, Jul. 2021, doi: 10.1016/j.ins.2021.03.002.

[41] M. Pérez-Enciso and L. M. Zingaretti, "A Guide on Deep Learning for Complex Trait Genomic Prediction," *Genes*, vol. 10, no. 7, p. 553, Jul. 2019, doi: 10.3390/genes10070553.

[42] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015, doi: 10.1016/j.neunet.2014.09.003.

[43] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach Learn*, vol. 63, no. 1, pp. 3–42, Apr. 2006, doi: 10.1007/s10994-006-6226-1.

[44] M. M. Hameed, M. K. AlOmar, F. Khaleel, and N. Al-Ansari, "An Extra Tree Regression Model for Discharge Coefficient Prediction: Novel, Practical Applications in the Hydraulic Sector and Future Research Directions," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–19, Sep. 2021, doi: 10.1155/2021/7001710.

[45] T. Sudhamathi and K. Perumal, "Ensemble regression based Extra Tree Regressor for hybrid crop yield prediction system," *Measurement: Sensors*, vol. 35, p. 101277, Oct. 2024, doi: 10.1016/j.measen.2024.101277.

[46] F.-K. Wang and T. Mamo, "Gradient boosted regression model for the degradation analysis of prismatic cells," *Computers & Industrial Engineering*, vol. 144, p. 106494, Jun. 2020, doi: 10.1016/j.cie.2020.106494.

[47] A. Ahrens, E. Ersoy, V. Iakovlev, H. Li, and M. E. Schaffer, "An Introduction to Stacking Regression for Economists," in *Credible Asset Allocation, Optimal Transport Methods, and Related Topics*, vol. 429, S. Sriboonchitta, V. Kreinovich, and W. Yamaka, Eds., in Studies in Systems, Decision and Control, vol. 429. , Cham: Springer International Publishing, 2022, pp. 7–29. doi: 10.1007/978-3-030-97273-8_2.

[48] Y. Cai *et al.*, "An Adaptive Stacking Regressor With a Self-Iterative Optimization Module for Improving Fractional Woody Cover Mapping," *IEEE Geosci. Remote Sensing Lett.*, vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2023.3281646.

[49] F. A. C. Viana, G. Venter, and V. Balabanov, "An algorithm for fast optimal Latin hypercube design of experiments: AN ALGORITHM FOR FAST OPTIMAL LHD," *Int. J. Numer. Meth. Engng.*, vol. 82, no. 2, pp. 135–156, Apr. 2010, doi: 10.1002/nme.2750.

[50] J. Lu, Y. Fang, and W. Han, "A novel adaptive-weight ensemble surrogate model base on distance and mixture error," *PLoS ONE*, vol. 18, no. 10, p. e0293318, Oct. 2023, doi: 10.1371/journal.pone.0293318.

[51] P. Vurtur Badarinath, M. Chierichetti, and F. Davoudi Kakhki, "A Machine Learning Approach as a Surrogate for a Finite Element Analysis: Status of Research and Application to One Dimensional Systems," *Sensors*, vol. 21, no. 5, p. 1654, Feb. 2021, doi: 10.3390/s21051654.

[52] S. Koziel and A. Pietrenko-Dabrowska, "Fundamentals of Data-Driven Surrogate Modeling," in *Surrogate Modeling for High-Frequency Design*, WORLD SCIENTIFIC (EUROPE), 2022, pp. 1–37. doi: 10.1142/9781800610750_0001.

[53] M. Al-Mukhtar and F. Al-Yaseen, "Modeling Water Quality Parameters Using Data-Driven Models, a Case Study Abu-Ziriq Marsh in South of Iraq," *Hydrology*, vol. 6, no. 1, p. 24, Mar. 2019, doi: 10.3390/hydrology6010024.

[54] R. de Winter, J. Furustam, T. Bäck, and T. Muller, "Optimizing Ships Using the Holistic Accelerated Concept Design Methodology," in *Practical Design of Ships and Other Floating Structures*, T. Okada, K. Suzuki, and Y. Kawamura, Eds., in Lecture Notes in Civil Engineering, vol. 65. Singapore: Springer Singapore, 2021, pp. 38–50. doi: 10.1007/978-981-15-4680-8_3.