Proceedings of 15th International Marine Design Conference (IMDC-2024)
June 2-6, 2024
Amsterdam, the Netherlands

Prediction of main engine power of oil tankers using artificial intelligence algorithms

Darin Majnarić[1,*], Nikola Anđelić[2], Sandi Baressi Šegota[2], and Jerolim Andrić[1]

ABSTRACT

In the preliminary ship design, the accurate determination of a vessel's main engine power is one of the most critical aspects next to service speed, main particulars, and cargo capacity. However, this task can be quite intricate due to its reliance on an extremely great number of influencing factors. In the research that is presented in this paper dataset of 357 oil tankers was gathered and developed to research the idea in which genetic programming is applied to the mentioned dataset to obtain mathematical equations (MEs) that can estimate the ship's main engine power with high accuracy. The highest estimation accuracy of MEs is achieved by tuning the GP hyperparameter values through the random hyperparameter search (RHS) method. The initial dataset was divided into train and test datasets in a 70:30 ratio. The train dataset was used to train GP in a 5-fold cross-validation process and after the process was done the obtained MEs were evaluated on the test dataset. To evaluate the GP training testing process several evaluation metrics were used i.e., coefficients of determination (R2), mean absolute error (MAE), root mean square error (RMSE), and length of obtained MEs. The conducted investigation showed that GP generated MEs that can estimate ship main engine power with high accuracy.

KEY WORDS

Preliminary ship design; Genetic programming; Main engine power; Random hyperparameter search; 5-fold cross-validation

INTRODUCTION

In the initial design phase of ship design, it is necessary to select key values that define the general characteristics of the ship. These initial values can be specified as requirements by the client or determined by the naval architect. Main particulars, cargo capacity, service speed, and main engine power are among the values that form the foundation of the ship design process.

Upon defining the primary project inputs, the naval architect can roughly estimate the main engine power required to propel the ship and consequently determine the type of main engine needed. When establishing these initial values, which serve as the starting point for the entire design process, the designer must consider limiting factors such as port and channel specifications. The sizes of channels significantly constrain ship dimensions, as evident from the names of standard oil tanker categories such as Panamax and Suezmax, which are designed to navigate through specific channels. Therefore, initial ship design phase takes into consideration the mentioned requirements but also the limits of the environment in which the ship will be operating. It's evident that the entire design process relies heavily on accurately establishing initial values. Improved estimation of these values will likely result in fewer iterations later in the process, thereby shortening the design time. This correlation is evident in the widely adopted ship design spiral, where naval architects encounter fewer cycles. Ship design spiral was introduced by Evans (7) and is set up in a way in which as mentioned designer starts with rough design values and is through iterations arriving at final ship values and final design. There were many modifications made to the entire process

[1] Affiliation (Department of Naval Architecture and Offshore Engineering, Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Zagreb, Croatia)

[2] Affiliation (Department of Automation and Electronics, Faculty of Engineering, University of Rijeka, Rijeka, Croatia);
* Corresponding Author: darin.majnaric@gmail.com

Submitted: 23 February 2024, Revised: 1 May 2024, Accepted: 3 May 2024, Published: 23 May 2024

Conference paper, DOI: https://doi.org/10.59490/imdc.2024.887 e-ISSN: 3050-486

with the goals of shortening it, reducing required effort, automating certain aspects, and enhancing the overall quality of the process. For example, the design spiral itself was modified by Andrews (1), later on, Watson developed two design spirals of which one was for warships and the other for merchant ships (18). Further Rawson and Tupper have introduced more steps in their design spiral with which main design parameters were separated into the components (15). To reduce the number of iterations in the design spiral and to ”prepare” the process, so that it can accommodate different types of vessels Papanikolaou (12) has presented a ship design procedure in the form of a straightened spiral. As presented there was constant effort to improve the process itself and through time new methods are being developed and proposed. Besides improvements on the process itself as already mentioned it was recognized that initial values have a significant impact on the rest of the design process and also the final results. Therefore efforts were also placed into the improvement of the initial values with which the design process is starting. Multiple approaches were developed and utilized to predict initial values. Most efforts were directed toward predicting the main particulars. Among the first to develop and apply statistical regression equations was Piko who used the nonlinear approximation method (14). Later on, Papanikolaou used the power regression model for initial values prediction as deadweight as an input (13). Linear 2nd-degree polynomial together with a power regression model was used by Kristensen to predict initial values which were in his research focused on container ships (8).

In recent years, new approaches for predicting initial ship values have been developed and used. Among them most significant is the utilization of artificial intelligence algorithms. Specifically, a lot of research has been conducted on the possible usability of artificial neural networks (ANN). For example, in his research, Clausen has (4) used Bayesian network and regression analysis at the same time as ANN for the estimation of ship main particulars. Ekinci (6) was predicting the main parameters of oil tankers as well by using ANN and the method that has shown as most successful was the Model Trees (M5P) method. When considering ship types most of the research was been conducted on container ships, Majnaric et al. (10) developed a database of 250 container ships which was analyzed with ANN to get high accurate estimation of initial ship values. Additionally, research was also conducted on the utilization of synthetic data to get to the amount of data that can be usable for ANN (9). The research focused on the estimation of ship length between perpendiculars was conducted by Cepowski et al. (2) and by using ANN they have produced two equations for preliminary design purposes. Part of the overall research focused on the implementation of artificial intelligence algorithms was also focused on the prediction of main engine power. In the research (3) authors have used a set of ANNs to update the design equations for the estimation of main engine power and fuel consumption on top of which it was proposed how to estimate CO2 emissions. Similar research was carried out by Ozsari (11) at which ANN was used to predict main engine power and pollutant emission. Develop model was presented and proposed for future studies on fuel consumption and energy efficiency.

From the previous literature overview, it can be noticed that the majority of used AI algorithms were ANNs. The main problem with ANNs is that they can’t be easily transformed into simple mathematical equations due to a large number of interconnected neurons. The other problem is that they require considerable computational resources for storing trained models (memory) and for processing new data samples (CPU power). So, in this paper, the idea is to apply genetic programming symbolic regressor (GPSR) to obtain symbolic expressions (SEs) that could estimate the ship’s main engine power with high estimation accuracy. The benefit of using GPSR is that the trained AI model does not require string after it was trained since it after each execution generates SEs i.e. mathematical equation that connects input variables with the target (output) variable). In order to determine the optimal combination of GPSR hyperparameter values for generating SE with high estimation accuracy, the random hyperparameter value search (RHVS) method was developed and applied. To create a robust system of SEs the GPSR will be trained using a 5-fold cross-validation method. A database of Oil tankers was developed so that genetic programming could be applied to obtain mathematical equations that can be used for accurate estimation of main engine power. However, the dataset created for this research had to be preprocessed (outliers removed, synthetically oversampled). Based on the detailed literature overview and the idea in this paper the following questions arise:

- Is it possible to obtain SEs using GPSR that could estimate main engine power with high estimation accuracy?
- Do data preprocessing techniques (outliers removal) and synthetic oversampling have any effect on the estimation accuracy of the trained GPSR model i.e. obtained SEs?
- Can the RHVS method be utilized to find the optimal combination of GPSR hyperparameters, leading to SEs with the highest estimation accuracy achieved by GPSR?
- Is it possible to obtain a robust system of SEs using a 5-fold Cross-Validation process for the estimation of electrical power output?

The outline of this paper consists of the following sections: materials and methods, results and discussion, and conclusion. In materials and methods, the dataset is described as well as the statistical analysis performed. Besides that, in the materials and methods section, the GPSR is described as well as, the evaluation metrics used and the training-testing procedure. The results section contains the description of the best set of SEs obtained with the optimal combination of GPSR hyperparameter values, and evaluation metric values. The conclusion section contains the conclusions drawn from this research ed based on the results and discussion given in the previous section.

MATERIALS AND METHODS

This section of the paper will begin by detailing the initial creation of the database, emphasizing its quality assurance measures. Subsequently, an analysis of the database will be presented. Finally, dataset statistics will be provided and analyzed at the conclusion of this section.

Database

For this study, a dataset of 357 oil tankers was developed. Information for each ship was collected, including net tonnage (NT), service speed (v), and main engine power.

Additionally, to the mentioned data deadweight tonnage (DWT) and length overall (LOA) were collected as well but were not used in this study other than for the representation purposes of main data.

The decision was made to collect data directly from classification society databases, as the quality of results and analysis depends on the input data. Three maritime classification societies had open and maintained databases available to the public (Register; Veritas; DNV) and were used to collect the data. The databases of these classification societies exhibited some inconsistencies in data presentation, and at times, information was missing. Inconsistencies were observed in the form of misplaced commas in values, leading to physically impossible scenarios, as well as incorrect numbers resulting in significantly inflated or deflated values. For example, incorrect values, when located at the first digit of the ship's service speed, introduced significant oscillation from the usual speed expected for a ship of a certain length and main machine power. Therefore, all of these questionable values were identified, checked, and either corrected or obsoleted. For Genetic programming, it is preferable to have larger databases, so that training and validation sets can be of sufficient size, ensuring that the database size does not limit the analysis. Because of this, the intention was to gather as much data as possible. In cases where data was missing or incorrect, best efforts were made to fill in the gaps. Missing or incorrect values were addressed only if the information was available from shipowners' websites or by directly contacting shipping companies. The important quality check performed on the developed database aimed to identify sister ships and recurring vessels, which were subsequently removed to prevent any distortion of the final results. In order to keep the database relevant it was filled in only by the ships built from the year 2000 and onwards. Distribution of the ships through the years is visible at the figure 1.

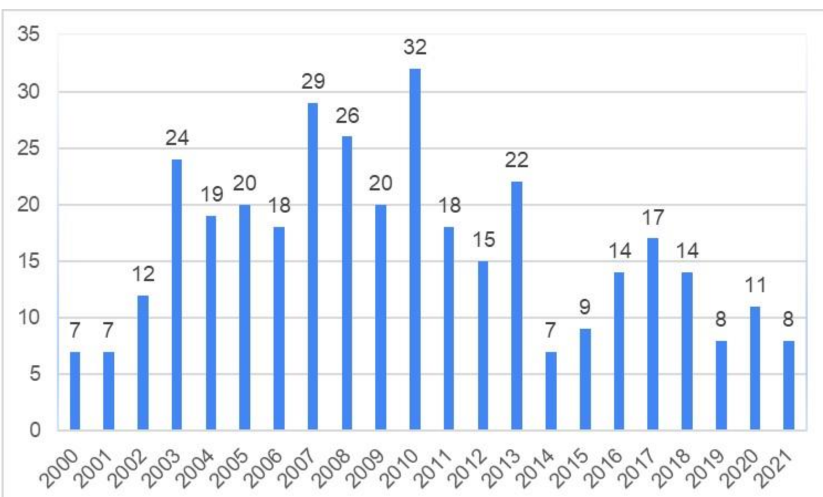


Figure 1. Distribution of ships in the database through the built years

The mentioned deadweight tonnage that was as already specified collected additionally, was used in this research only to present the ship categorization. In 1 ship categorization is presented based on typically used ship categorization in the industry.

As expected, most of the collected ships belong to the smallest Panamax category, while there are only two records of ultralarge crude carriers (ULCCs). Although more ULCC records were initially present in the database, some were obsoleted. For example, in one instance, there were four sister ships, but only one record was retained. Including records of sister ships in the database would introduce clusters that could distort the dataset and compromise the reliability and realism of the analysis.

Table 1: Distribution of ships through classes

Sub-types	DWT range		Number of ships
Panamax	-	79,999	197
Aframax	80,000	119,999	68
Suezmax	120,000	199,999	47
Very large crude carrier (VLCC)	200,000	319,999	43
Ultra large crude carrier (ULCC)	320,000	-	2

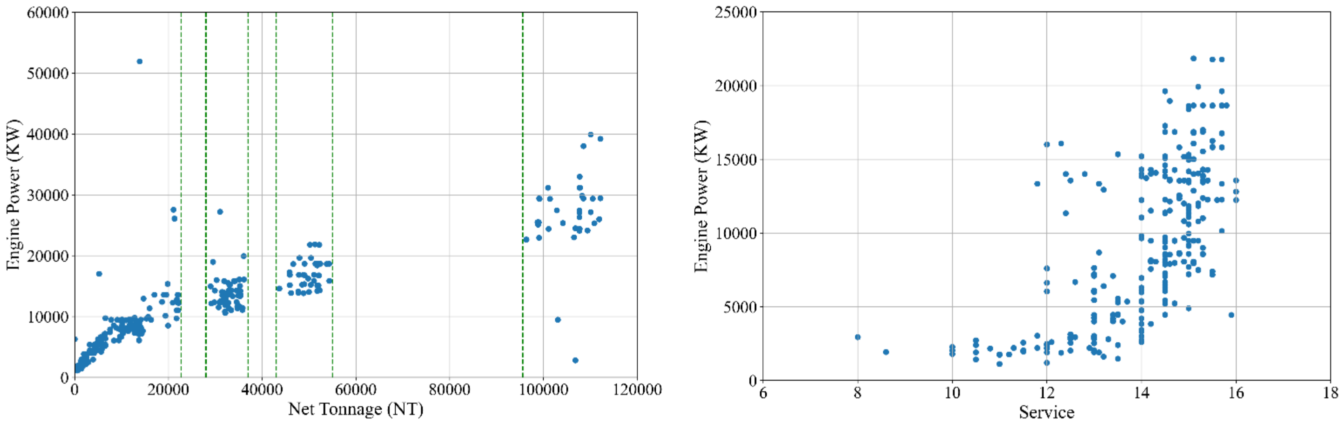
Dataset statistics

The dataset consists of 357 samples and three variables i.e. net tonnage, service speed, and main engine power. The initial statistical analysis of the dataset i.e. the number of samples (count), mean, standard deviation (std), minimum (min), maximum (max), and GPSR variable representation is listed for each dataset variable in Table 2.

Table 2: Initial statistical analysis of the dataset

	Net Tonnage (NT)	Service	Engine Power (KW)
count	357		
mean	30213.52	14.03221	11890.5
std	32342.13	1.360539	8332.385
min	5.49	8	1118
max	112192	16.8	51913
GPSR Variable Represntation	X_0	X_1	y

From Table 2 it can be seen that all three dataset variables have the same number of samples (357) which means that there are no missing values in the dataset. The mean, std, min, and max values show that net tonnage and engine power have a larger value range than the service variable. This variation in the value range of the dataset variable would indicate that the scaling/normalization technique is required. However, the idea in this paper was to perform the investigation with the original range of dataset variable values. The GPSR variable representation shows how the dataset variables will be represented in the symbolic expressions obtained with this method. The input variables Net Tonnage (NT) and Service (input variables) will be represented in symbolic expressions as $X0$ and $X1$, respectively. The plots Engine Power versus Net Tonnage (NT) and Engine Power versus Service are shown in Figure 2.



(a) Engine versus Net Tonnage (NT) plot (The green lines represent the gaps)

(b) Engine versus Service plot

Figure 2: Initial plots of dataset input variables versus the Engine power

From Figure 2 it can be noticed that data is scattered. However, some general trends can be noticed for lower values of net tonnage (NT) (0-60000). There are also gaps where no samples consist for example net tonnage (NT) first gap (22700 - 28000), second gap (37000 - 43000), and third gap (55000 - 95600). Besides the gaps in the general data trend, there are also a number of outliers i.e. samples that greatly deviate from the general data trend and should be somehow handled. Regarding the Engine power versus service plot, the general trend could be noticed for lower values of the engine power however the majority of data is scattered in 12 to 16 range Service values. The next step in dataset analysis is to perform Pearson's correlation analysis to determine the correlation between input and output variables. Pearson's correlation between two variables can be in the -1 to 1 range. The -1 value indicates perfect negative correlation between two variables and indicates that if the value of one variable increases the value of the other variable will decrease and vice versa. The value of 1 indicates a perfect positive correlation where an increase of one variable will increase the value of the other variable and vice versa. The worst possible correlation value is 0 between two variables which indicates that the increase/decrease in one variable value will not affect the other variable. The results of the Pearson's correlation analysis in the form of the heatmap are shown in Figure 3.

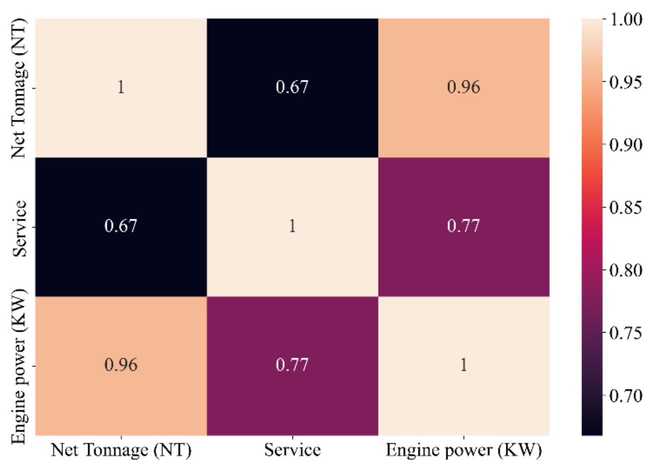


Figure 3: Results of Pearson's correlation analysis in the form of the heatmap

From Figure 3 it can be noticed that the Net Tonnage (NT) has a high positive correlation value (0.86) with the main engine power while the Service has a low positive correlation (0.29) value with the main engine power. Regarding the correlation between the net tonnage (NT) and the service speed variable, the correlation value is 0.18 which is the lowest correlation between the two variables in this dataset. This value indicates that the change in Net Tonnage (NT) will have a very small effect on the Service variable, and vice versa. The value of 1 on the diagonal line in the heatmap indicates that the variable is correlated with itself. Outliers and boxplots play crucial roles in data analysis. Outliers, which are data points significantly different from others, can distort interpretations and analyses, making their identification important. Boxplots offer a concise visual representation of data distribution, displaying key summary statistics such as median, quartiles, and outliers. They facilitate a quick understanding of data spread, central tendency, and variability, aiding in comparison across different groups or categories within a dataset. Moreover, boxplots are robust to skewness and outliers, providing reliable insights even with non-normally distributed data. Their effectiveness in communicating findings makes them valuable tools for analysts to convey insights to stakeholders succinctly. Overall, outliers and boxplots are essential in detecting anomalies, summarizing distributions, comparing groups, and communicating data characteristics effectively. The results of outlier detection are presented in the form of the boxplot and shown in Figure 4.

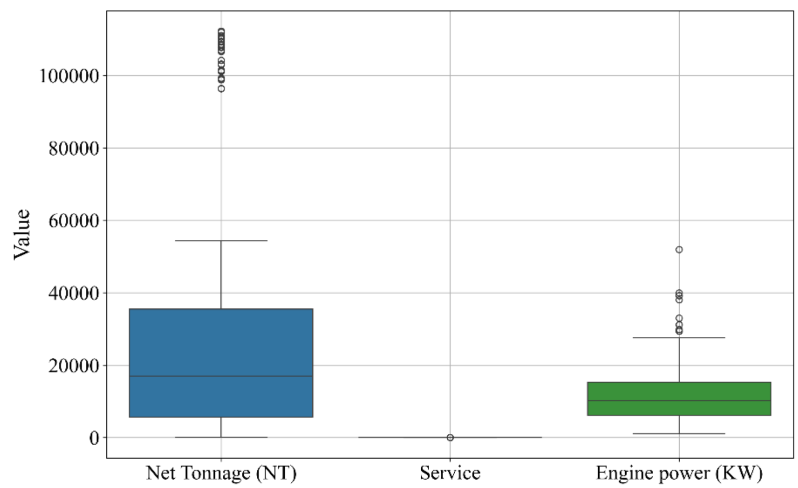


Figure 4: Distribution of dataset variables.

As observed in Figure 4, all dataset variables exhibit outliers to some extent. To enhance estimation accuracy with the GPSR, these outliers need to be addressed. The capping method known as median absolute deviation (MAD) was employed for outlier removal. MAD is a robust statistical measure of dispersion, offering reduced sensitivity to outliers compared to measures such as standard deviation. To apply the MAD method for capping outliers, the following steps were followed: Firstly, the median of the dataset was calculated to establish the central tendency. Next, the absolute deviations of each data point from the median were computed to capture the variability. Subsequently, the median of these absolute deviations (MAD) was determined. Finally, MAD was multiplied by a constant factor, typically 2 or 3 (in this case, 3), to establish the threshold for identifying outliers.

$$LT = \text{median} - (3 \cdot \text{MAD}), \quad (1)$$
$$UT = \text{median} + (3 \cdot \text{MAD}), \quad (2)$$

Here, LT and UT denote the lower and upper thresholds, respectively, beyond which any data points are regarded as outliers and may be either capped or removed. Any data points outside these thresholds can be considered outliers and in this paper, these outliers were removed. Generally, this is a very useful method when dealing with skewed or non-normally distributed data where traditional methods like standard deviation may not be appropriate due to sensitivity to outliers. With the application of this outlier removal method, the outliers were successfully removed from the dataset as seen from Figure 5.

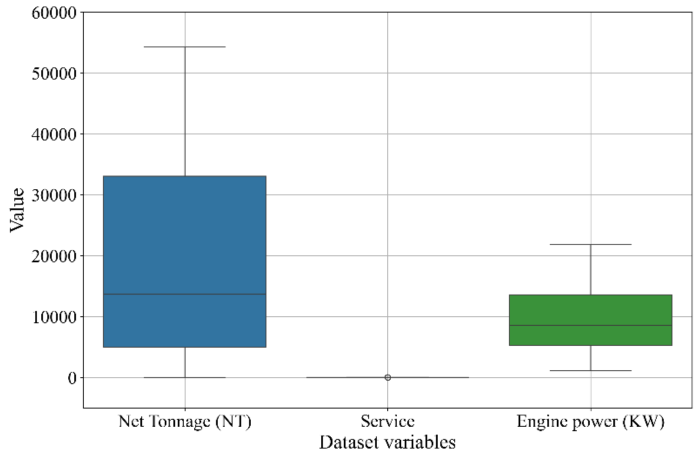
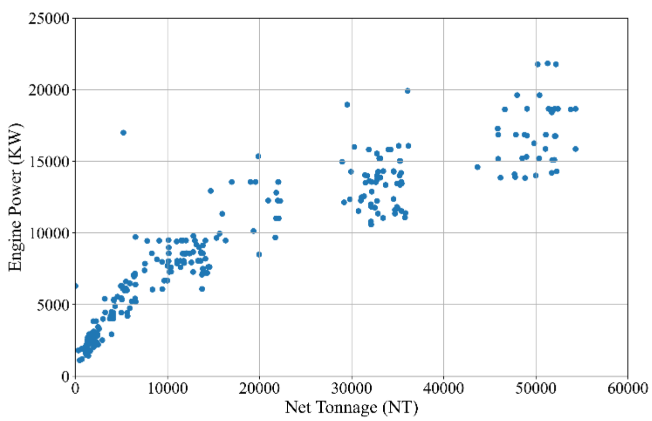
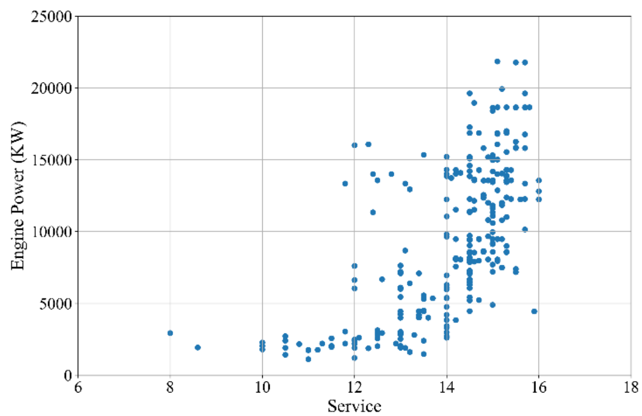


Figure 5: The distribution of dataset variables after application of MAD outlier capping method.

When Figure 4 and 5 are compared it can be noticed that the outliers (black dots) were removed from dataset variables (Net Tonnage (NT) and Engine Power (KW)), however, some small number of outliers remains in the Service dataset variable. With the application of the MAD capping the majority of outliers were successfully removed however the number of dataset samples was reduced from the original 357 to 307 samples. After the outliers were removed from the dataset using the MAD method the plots Engine versus Net Tonnage (NT) and Engine versus Service are shown in Figure 6.



(a) Engine versus Net Tonnage (NT) plot after outliers were removed using MAD technique



(b) Engine versus Service plot after outliers were removed using MAD technique

Figure 6: The Engine power versus net tonnage and service after outliers were removed using MAD technique

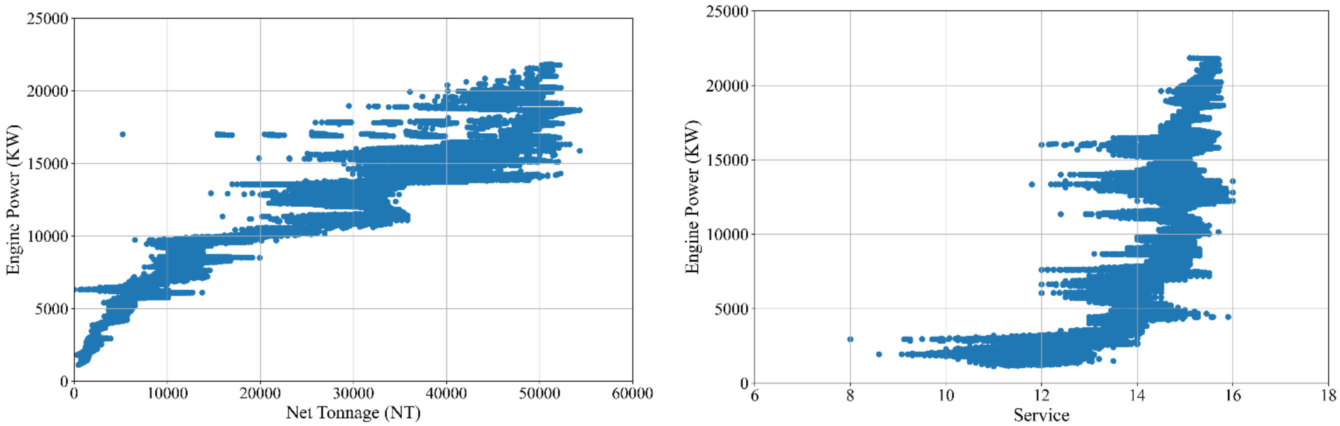
As seen form Figure 6 it can be noticed that the MAD technique has removed outliers from the dataset. This is especially valid when Figure 6a is compared to Figure 2a. Besides outliers that greatly deviate from the general data trend a large number of samples where removed in which Net Tonnage (NT) exceeded 90000. This is due to the large gap in data since the majority of samples in terms of Net Tonnage (NT) are concentrated in the 0 to 60000 range.

The other problem with the obtained dataset is the small number of samples and the idea is to see if somehow the number of samples could be synthetically generated. In this paper the averaging method is considered that consists of the following steps:

1. sort the dataset samples from minimum to maximum value of the target variable (main engine power),
2. for dataset samples in range from 1 to the maximum number of dataset samples:
    - create an average dataset sample between the current sample and the sample + 1, sample +2, sample +3, sample +4, sample +5, sample + 6, and sample +7 for all dataset variables,

3. concatenate the original dataset with newly created samples,
4. repeated the process an arbitrary number of times.

In this paper, the process of creating the average values was repeated 3 times. By doing so the number of samples was enlarged from 307 up to 76460.



(a) Engine versus Net Tonnage (NT) plot after average oversampling was performed 3 times

(b) Engine versus Service plot after average oversampling was performed 3 times

Figure 7: The Engine power versus net tonnage and service after average oversampling was performed

As seen from Figure 7 visually the number of samples was greatly increased. The advantage of the average oversampling is that the gaps (in the case of Net Tonnage) are filled but the problem is the small number of outliers generated (samples that deviate from the general data trend). The correlation plot of the final synthetically oversampled dataset is shown in Figure 8.

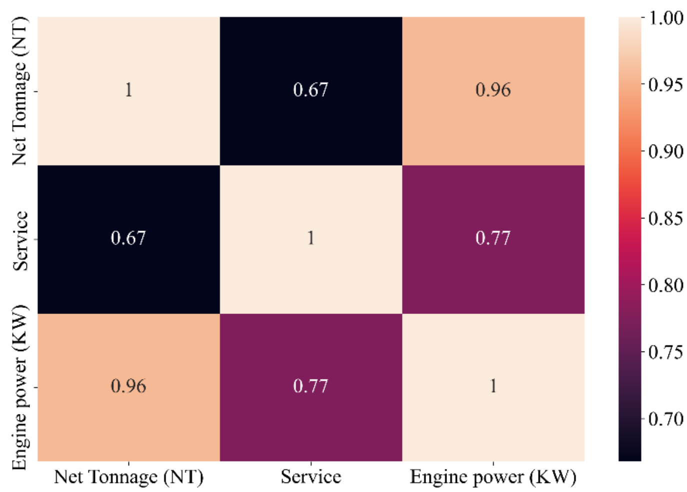


Figure 8: The correlation heatmap of the synthetically enlarged dataset.

When Figures 3 and 8 are compared it can be noticed that with the application of outlier removal and the synthetic oversampling, the correlation between Net Tonnage (NT) and the main engine Power was increased from 0.86 to 0.96, between Service and main engine power (KW) the correlation was increased from 0.29 to 0.77. It can be also noticed that the correlation between Net Tonnage and Service speed was increased from 0.18 to 0.67. This final version of the dataset obtained using outlier removal and synthetic oversampling will be used in genetic programming symbolic regressor to obtain symbolic expressions for estimation of main engine power.

Genetic programming symbolic regressor

The genetic programming symbolic regressor begins its execution by randomly creating the population of naïve symbolic expression by random selection of input variables, constant values, and mathematical functions to estimate the target value. The estimation accuracy of the initial population is generally very low. However, with the application of genetic operations such as crossover and mutation through a consecutive number of generations in the end the symbolic expression is obtained that estimates the target variable with certain accuracy. Since the GPSR has a large number of hyperparameters the random hyperparameter values search method was developed and applied to find the optimal combination of hyperparameter values using GPSR will produce the SE with high estimation accuracy. The process of developing the RHVS includes the following steps:

1. Definition of the initial value range for each GPSR hyperparameter value,
2. Testing the lower and upper boundaries of each GPSR hyperparameter to see if the GPSR can successfully be executed i.e. produce the SE,
3. If the GPSR fails in its execution due to a specific hyperparameter boundary value adjust the hyperparameter value
range.

In the RHVS method, the optimal values of the following hyperparameters were searched:

- Population size – the size of the population (SEs) that will be evolved during its execution
- Number of generations – The maximum number of generations for which GPSR will executed. This is also one of the termination criteria which means that GPSR will terminate the execution after the predefined number of generations is reached.
- Tournament size – the tournament size is the size of a randomly selected population in each generation that will compete to become the parents of the next generation. In every tournament selection, there is only one winner and on this winner, genetic operations such as crossover or mutation are performed.
- Initial depth size – inside the GPSR the population members are represented as tree structures which means that the size of the population member is presented in terms of depth measured from the root node up to the deepest leaf in the tree structure. It should be noted that the initial depth size is defined as a value hyperparameter indicating the range of the population member's depth size. For example (3,12) indicates that the initially created population the depth of population members will be in 3 to 12 depth.
- Crossover probability value – probability value of crossover operation performed in each generation. The crossover is performed using two tournament selection winners. On both tournament selection winners, the random sub-tree is selected and the sub-tree from the second tournament winner replaces the sub-tree of the first winner to create offspring for the next generation.
- Subtree mutation probability value – the probability value of the subtree mutation. This operation and other mutations used in this research require only one tournament selection winner on which the random sub-tree is selected and replaced with a randomly created sub-tree by randomly picking constant values (from a predefined range), input variables, and mathematical functions.
- Point mutation probability value – is the probability value of the point mutation genetic operation. In point mutation, the nodes are randomly selected on the tournament winner and replaced. In other words, the randomly selected constant value node is replaced with a constant value, the input variable with other input variables, and mathematical functions with randomly picked mathematical functions. However, when mathematical functions are replaced with other randomly selected mathematical functions the number of arguments in the original mathematical function must be the same in the mathematical function that will replace the original one.
- Hoist mutation probability value – In hoist mutation, the random subtree is selected on the tournament winner, and on that tree a random node is selected. Then the randomly selected node is "hoisted" i.e. it replaces the entire randomly selected subtree.
- Range of constant values – the range of constant values that will be used in GPSR to develop the initial population, to perform the different mutation operations.
- Stopping criteria – is the predefined minimum value of the fitness function. In the case of GPSR, the fitness function is the Mean Absolute error defined with the equation:

$$MAE = \frac{\sum_i^n (y_{ti} - y_{pi})}{n}, \tag{3}$$

Where y_{ti}, y_{pi}, and n are the true target value, predicted target value and the number of dataset samples. So, in the case of stopping criteria if the fitness function value of only one population member falls below the predefined stopping criteria value the GPSR execution will terminate and the GPSR will give as the output the best SE. In other words, the stopping criteria are the second GPSR termination criteria, alongside the number of generations.

- Max Samples – the fraction of the samples that will be drawn from the training dataset and used to evaluate each SE. The max samples are a useful tool to see how the SE performs on the unseen data during the execution. The output value in this case will be out of bag or raw fitness value. For example, if you use a training dataset you can specify a very small portion of that dataset that will not be used during training i.e. it will be unseen by SEs during training, and in each generation, the majority of the training dataset will be used to calculate the fitness value while this small portion will be used to calculate the raw fitness value. Generally, the raw fitness value should be close to the real fitness value in order to at the end obtain SE with high estimation accuracy.
- Parsimony coefficient – is the coefficient used in the parsimony pressure method to stop the bloat phenomenon. This phenomenon occurs when the size of the population members rapidly grows from generation to generation without any indication of lowering the fitness value. In extreme cases, this bloat phenomenon can cause the GPSR to quickly terminate due to memory overflow or it can execute after a long time with an extremely large size (1000 or more elements in SE). The coefficient is the most sensitive one which means that small values can lead to extremely large population members (SEs) while large values can lead to small SEs with low estimation performance.

It should be noted that the sum of all genetic operations (probabilities) must be equal to 1 or slightly lower than 1 for example 0.99. If the value is smaller than 1 then some tournament selection winners enter the next generation unchanged i.e. the genetic operations were not applied to some tournament selection winners. In all GPSR investigations, the following mathematical functions were used addition, subtraction, multiplication, division, square root, absolute value, sine, cosine, tangent, natural logarithm, and logarithm with bases 2 and 10. However, division, square root, natural logarithm, and logarithm with bases 2 and 10 had to be modified to avoid imaginary or not number values which could lead to failed GPSR executions.

The division function was modified in the following way:

$$y_{\mathrm{DIV}}(x_1, x_2) = \begin{cases} \frac{x_1}{x_2} & \text{if } |x_2| > 0.001 \\ 1 & \text{if } |x_2| \leq 0.001 \end{cases} \tag{4}$$

The square root function was modified in the following way:

$$y_{SQRT} = \sqrt{|x|} \tag{5}$$

The natural logarithm and logarithm with bases 2 and 10 are modified in the following way:

$$y_{LOG}(x) = \begin{cases} \log_i(|x|) & |x| > 0.001 \\ 0 & |x| < 0.001 \end{cases} \tag{6}$$

Where i represents the base of the logarithm (e, 2, and 10). It should be noted that x_1, x_2, and x in previous equations do not have any connection with the input variables used in this research. They are only used here to describe the modifications made to the aforementioned mathematical functions.

Table 1: The RHVS ranges for previously described hyperparameters are listed

Hyperparameter	Range
Population size	1000-2000
Number of generations	25 - 50
Tournament size	100-500
Initial depth size	3-18
Crossover	0.001-1
Subtree mutation	0.001-1
Point mutation	0.001-1
Hoist mutation	0.001-1
Range of constant values	-10000 - 10000
Stopping criteria	$1 \times 10^{-9} - 1 \times 10^{-3}$
Max samples	0.99-1
Parsimony coefficient	1×10^{-6} - 1×10^{-1}

Evaluation metrics

The evaluation metrics used in this research were coefficient of determination (R^2), mean absolute error (MAE), and Root mean square error (RMSE). The R^2 can be defined as:

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \overline{y}_i)^2}, \tag{7}$$

Where y_i is the real value (target variable), f_i is the predicted value and the y_i is the mean value of the real data. The MAE was previously defined in the description of GPSR as it was used as a fitness function for the evaluation of population members. RMSE stands for Root Mean Square Error. It is a commonly used metric to evaluate the performance of a regression model. RMSE measures the average magnitude of the errors between predicted values and actual values. Mathematically, RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i}^{n}(y_i - f_i)^2} \tag{8}$$

Where n is a total number of samples.

Training-Testing Procedure

The training-testing procedure is graphically shown in Figure 9.

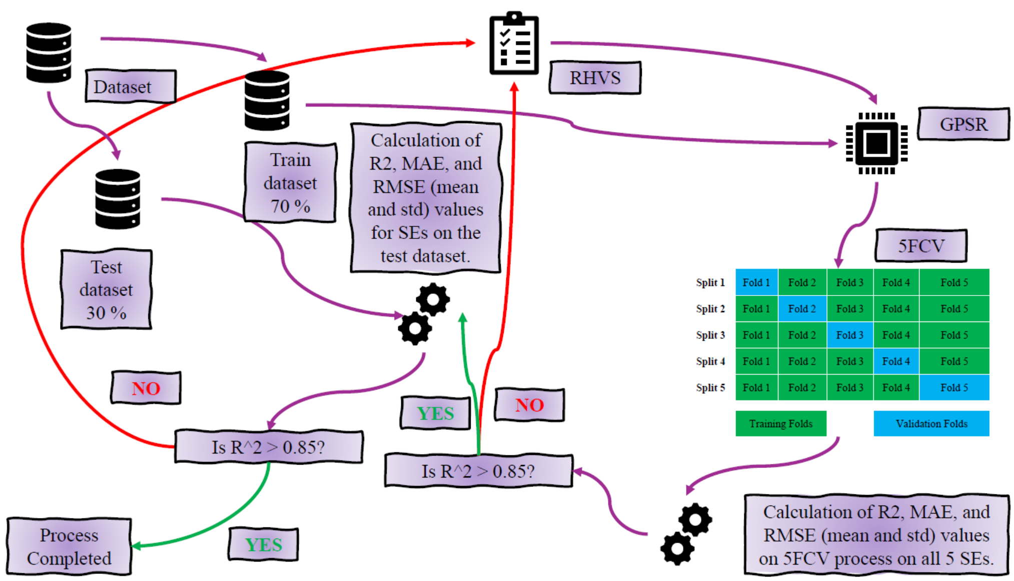


Figure 9: The scheme of training GPSR using the 5FCV process and RHVS method for GPSR random hyperparameter selection.

The training-testing procedure shown in Figure 9 consists of the following steps:

1. The dataset obtained after preprocessing (elimination of outliers and oversampling) is divided on training and testing
   datasets in 70:30 ratio. The training dataset will be used in GPSC and will be trained using 5FCV while the test dataset will be used from the evaluation of SEs obtained with GPSC.
2. The random selection of GPSC hyperparameter values using the RHVS method and training of GPSC using 5FCV. It
   should be noted that after each split in 5FCV (5 splits in total) the SE is obtained so this means that after 5FCV a total of 5 SEs will be obtained using GPSC.
3. After training was done evaluate the obtained SEs using R2, MAE, and RMSE evaluation metric methods. Since after 5FCV, the GPSC generated 5 SEs the evaluation metric values have to be obtained for each SEs on train and validation sets. The mean and std values of the aforementioned metric methods are obtained and if the mean value of R2 is higher than 0.85 the process continues to the testing phase. However, if the R2 is lower than 0.85 the process continues from the beginning i.e. by selecting the random GPSC hyperparameter values using the RHVS method.
4. In the testing phase the test dataset (30% of the initial) dataset is provided to the 5 SEs and the output is generated. This output is compared to the original output and the R2, MAE, and RMSE mean and std values are obtained. Finally, if the mean R2 value is higher than 0.85 the process is completed. Otherwise, the process starts from the beginning i.e. from the RHVS method.

RESULTS AND DISCUSSION

The best set of SEs obtained in this research was achieved using the following combination of hyperparameters listed in Table 4.

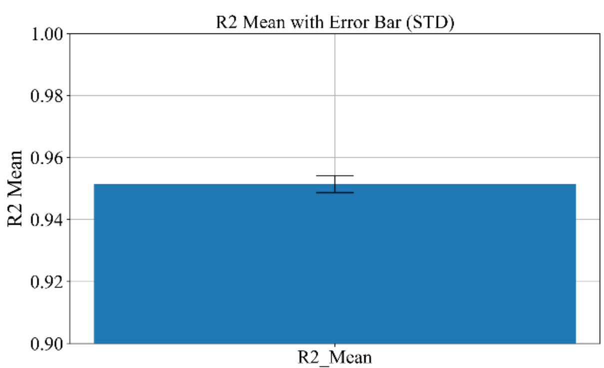
Table 4: Optimal combination of GPSR hyperparameter values

	Population size	Number of generations	Tournament size,	Initial depth size	Crossover	Subtree mutation
Case 1	1126	25	350	(6, 11)	0.0035	0.9667
	Point mutation	Hoist mutation	Range of Constant Values	Stopping Criteria	Max Samples	Parsimony Coefficient
Case 1	0.019	0.0102	(-8861.28, 1162.32)	6.91×10^{-7}	0.995	0.0098

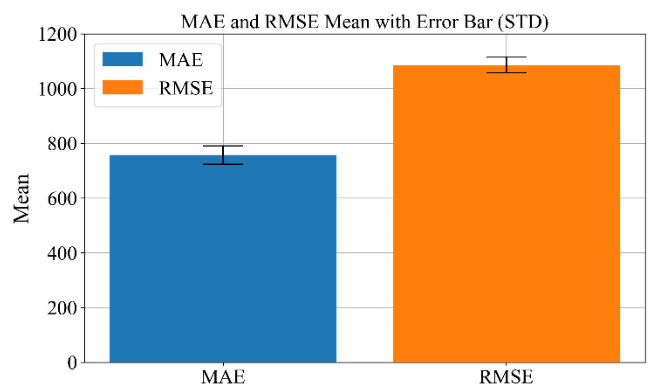
As seen from Table 4 the best set of SEs was obtained with GPSR in which the population consisted of 1126 population members that were evolved for 25 generations. These two hyperparameters are near the lower boundary used in the RHVS method (Table 3). The dominating genetic operation was subtree mutation (0.9667) while other genetic operations had values in the 0.0035 to 0.019 range. The stopping criteria value was defined so low any population member never reached it during the execution of GPSR which means that the dominating termination criteria in all these investigations was the number of generations. The parsimony coefficient was near the upper boundary (Table 3) which prevented the occurrence of bloat phenomenon and generated smaller SEs. The example of SE is shown in Eq.(9).

$$\begin{aligned} y \;=\; & 67.06\sqrt{|X_0|}\Big(\tan(\log(\max(\min(-3953.5, 1.443\log(0.43|\log(X_0)|)), \\ & \min(0.93\log(X_1), 1.2\sqrt{\log(\frac{X_0}{X_1}+\sqrt{X_0})} - \sin(\log(\min(|X_1|, \\ & 1.4\log(\min(-7556.63, X_0)))))))))\Big)^{\frac{1}{2}} - \sqrt{|\log(0.00031\log(X_0))(X_0+\sqrt[9]{X_0}-7073.28)|}+10.31 \end{aligned} \tag{9}$$

As seen the Eq.(9) is not so large however it requires both input variables. It should be noted that using the 5FCV training process the GPSR generated five different SEs to estimate the target variable. The evaluation metric values (R2, MAE, and RMSE) were obtained on both training and testing datasets. The mean and standard deviation values are shown in Figure 10.



(a) Mean R^2 value with standard deviation obtained with GPSR generated SEs on the synthetically oversampled dataset.



(b) Mean *MAE* and *RMSE* values with standard deviation obtained with GPSR generated SEs on the synthetically oversampled dataset.

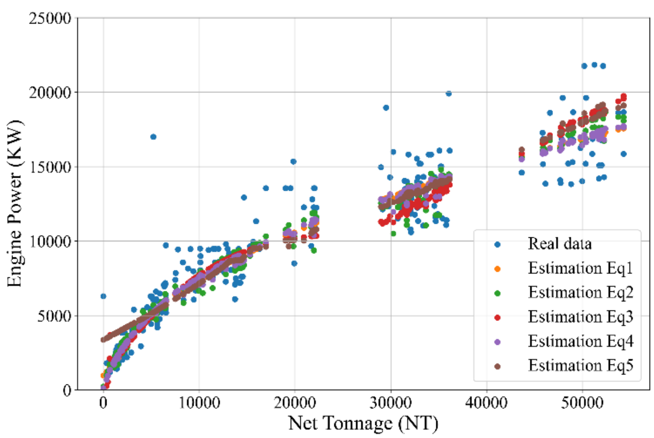
Figure 10: Evaluation metric values of SEs obtained with GPSR trained on the synthetically oversampled dataset.

As seen from Figure 10 the mean value of R^2 is pretty high with a small standard deviation value. The MAE and $RMSE$ values are low when these values are compared with the range of the target variable (Table 5). After the 5 SEs obtained on the synthetically oversampled dataset are evaluated on the initial dataset (307 samples) i.e. the dataset obtained after outlier removal using MAD capping the following evaluation metric values are obtained and listed in Table 5.

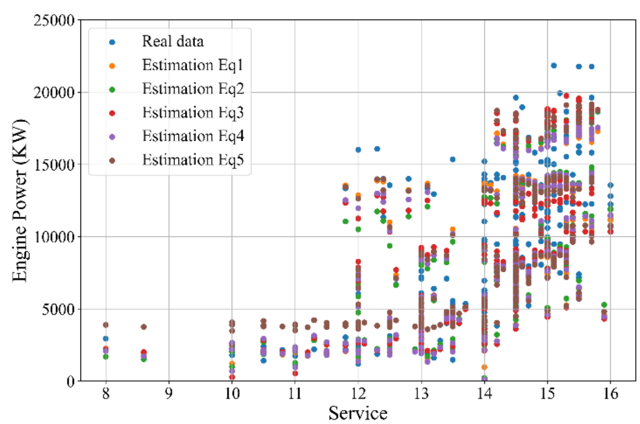
Table 5: Summary of Data

Metric	Value
R_2 Mean	0.883
R_2 STD	0.0113
MAE	1205.92
$\sigma(MAE)$	105.094
$RMSE$	1736.75
$\sigma(RMSE)$	83.67

The results from Table 5 have shown that the obtained SEs using GPSR on the synthetically oversampled dataset have slightly lower estimation accuracy when applied on the initial dataset (outliers removed). This is because the obtained SEs follow the general trend of the data however there are some samples that deviate from the general trend. The graphical representation of the Engine versus Net Tonnage (NT) and Engine versus Service is shown in Figure 11.



(a) Engine versus Net Tonnage (NT) comparison of real data and estimation made by all five best equations obtained with GPSR.



(b) Engine versus Service comparison of real data and estimation made by all five best equations obtained with GPSR.

Figure 11: Comparison of estimation made by five best equations obtained with GPSR and the real data (dataset with removed outliers).

As seen from Figure 11 the best SEs captured the trend in the data. This can be seen from Figure 11a where Engine Power (KW) versus Net Tonnage (NT) is shown.

CONCLUSIONS

As presented in this paper, a dataset was developed that consisted of 357 oil tankers for which data was collected as net tonnage (NT), service speed (v), and main engine power. The original dataset variables contained several outliers which had to be removed since they greatly influenced the estimation accuracy of the trained AI model. After the removal of the majority of outlier dataset samples the dataset was synthetically oversampled by the introduction of averaging samples. This modified dataset was used in GPSR with the RHVS method that was trained using 5FCV. After the application of this procedure, the 5SEs were obtained. The best set of SEs achieved great estimation accuracy of main engine power on synthetically oversampled datasets and the initial dataset without outliers. Based on the conducted investigation the following conclusions can be drawn:

- The SEs with high estimation accuracy of main engine power can be obtained using GPSR however the accuracy is dependent on two factors i.e. dataset quality, a combination of hyperparameters.
- Dataset preprocessing methods (outliers removal) and synthetic oversampling greatly contributed to the GPSR-generated SEs since the GPSR produced the SEs with high estimation performance.
- The RHVS method proved to be useful in finding the optimal combination of GPSR hyperparameters using which the SEs obtained with GPSR have achieved high estimation performance
- The 5FCV proved to be a useful method for training the GPSR since using this approach a robust set of 5 SEs was obtained.
- Approach can be used in the preliminary phase of ship design to generate a value of main engine power.
- Based on the results initial main engine power value will be an accurate estimation and therefore close to the value that will be reached at the end of the design. With that fewer iterations will be needed during the design development process.

  The equations developed through this approach have the potential to be utilized during the initial design phase of ship development. The practicality of this approach lies in its speed and accuracy, making it a valuable complement to existing methods for estimating initial engine main power. Furthermore, based on this research equations produced by this method can eventually even be used as a sole source of initial values.

The proposed approach in this paper has its pros and cons. The pros are:

- The outlier removal and synthetic oversampling are great approaches in dataset preparation which will be used for train of AI algorithm,
- The GPSR in combination with RHVS and 5FCV is a great method to obtain a robust set of SEs with high classification Performance.

The cons of the proposed research are:

- Removing outliers removes the number of original dataset samples, and in general, the trained AI model on the dataset without outliers cannot predict the outliers. The other problem with outliers removal using this approach is that the majority of outliers i.e. values of Net Tonnage (NT) higher than 50000 were removed from the original dataset. So this approach could be used for Net Tonnage in the 0 to 55000 range.
- The GPSR with RHVS and 5FCV can take some time to obtain a robust set of 5SEs with high estimation performance because the optimal combination of GPSR hyperparameters cannot be found instantly and each combination of GPSR hyperparameter values had to be tested in GPSR which is trained using 5FCV. This means that GPSR will be trained 5 times on different splits of the training sets i.e. generating 5 SEs in the process.
- This approach may not be suitable for designing oil tankers that deviate from typical designs.

Based on the pros and cons of the proposed research methodology future work will be focused on exploring alternative artificial intelligence algorithms (Neural networks) to see if better estimation accuracy of engine power could be obtained.
Besides the alternative artificial intelligence algorithms in the future work different input variables will be considered since the correlation between Net Tonnage (NT) and the Service with the Engine Power (KW) is initially too low.

DATA ACCESS STATEMENT

The dataset used in this study is available on request.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN WRTITING

Not applicable.

CONTRIBUTION STATEMENT

Author 1: Conceptualization; Investigation; Data curation; Writing original.
Author 2: Conceptualization; methodology; Software; Validation; Formal analysis; Writing original draft; visualization.
Author 3: Conceptualization; methodology; writing – original draft.
Author 4: Supervision; writing - review and editing; project administration; funding acquisition.

ACKNOWLEDGEMENTS

This research has been (partly) supported by the CEEPUS network CIII-HR-0108, the Erasmus+ project WICT, under the grant 2021-1-HR01-KA220-HED-000031177; the University of Rijeka, under scientific grants uniri-tehnic-18-275-1447 and uniri-mladi-technic-22-61.

REFERENCES

[1] Andrews, D. (1986). An integrated approach to ship synthesis. *Trans. RINA*, 128:73–102.

[2] Cepowski, T. and Chorab, P. (2021a). Determination of design formulas for container ships at the preliminary design stage using artificial neural network and multiple nonlinear regression. *Ocean Engineering*, 238:109727.

[3] Cepowski, T. and Chorab, P. (2021b). The use of artificial neural networks to determine the engine power and fuel consumption of modern bulk carriers, tankers and container ships. *Energies*, 14(16):4827.

[4] Clausen, H. B., Lutzen, M., Friis-Hansen, A., and Bjørneboe, N. (2001). Bayesian and neural networks for preliminary ship design. *Marine technology and SNAME news*, 38(04):268–277.

[DNV] DNV. Dnv gl vessel register. https://vesselregister.dnvgl.com/vesselregister/vesselregister. html. Accessed: 14.2.2023.

[6] Ekinci, S., Çelebi, U. B., Bal, M., Amasyali, M. F., and Boyaci, U. K. (2011). Predictions of oil/chemical tanker main design parameters using computational intelligence techniques. *Applied Soft Computing*, 11(2):2356–2366.

[7] Evans, J. H. (1959). Basic design concepts. *Journal of the American Society for Naval Engineers*, 71(4):671–678.

[8] Kristensen, H. O. (2013). Statistical analysis and determination of regression formulas for main dimensions of container ships based on ihs fairplay data. *University of Southern Denmark: Odense, Denmark*.

[9] Majnarić, D., Baressi Šegota, S., Anđelić, N., and Andrić, J. (2024). Improvement of machine learning-based modelling of container ship's main particulars with synthetic data. *Journal ofMarine Science and Engineering*, 12(2):273.

[10] Majnarić, D., Šegota, S. B., Lorencin, I., and Car, Z. (2022). Prediction of main particulars of container ships using artificial intelligence algorithms. *Ocean Engineering*, 265:112571.

[11] Ozsari, I. (2023). Predicting main engine power and emissions for container, cargo, and tanker ships with artificial neural network analysis. *Brodogradnja: Teorija i praksa brodogradnje i pomorske tehnike*, 74(2):77–94.

[12] Papanikolaou, A. (2009). Andersen p, kristensen ho, levander k, riska k, singer d, vassalos d. state of the art design for x. In *Proc. 10th Int. Marine Design Conference-IMDC09, Trondheim*.

[13] Papanikolaou, A. (2014). *Ship design: methodologies ofpreliminary design*. Springer.

[14] Piko, G. (1980). *Regression analysis ofship characteristics*. Australian Government Publishing Service.

[15] Rawson, K. J. and Tupper, E. C. (2001). *Basic Ship Theory: Hydrostatics and Strength*, volume 1. Butterworth-Heinemann.

[Register] Register, L. Lloyd's ships in class. https://www.lr.org/en/about-us/who-we-are/ lr-ships-in-class/. Accessed: 14.2.2023.

[Veritas] Veritas, B. Bv fleet. https://marine-offshore.bureauveritas.com/bv-fleet/#/bv-fleet/. Accessed: 14.2.2023.

[18] Watson, D. G. (1998). *Practical ship design*, volume 1. Elsevier.