SNAME Maritime Convention 2020
29 September- 2 October, Houston TX

www.sname.org

A Review on Applications of Machine Learning in Shipping Sustainability

Blanca Pena[1, 2] (SM), **Luofeng Huang**[1, *] (SM), **Fredrik Ahlgren**[2,3] (V)

1. Department of Mechanical Engineering, University College London, United Kingdom
2. Department of Mechanical Engineering, University of British Columbia, Canada
3. Department of Computer Science, Linnaeus University, Kalmar, Sweden

The shipping industry faces a large challenge as it needs to significantly lower the amounts of Green House Gas emissions at the same time as it is expected to meet the rising demand. Traditionally, optimizing the fuel consumption for ships is done during the ship design stage and through operating it in a better way, for example, with more energy-efficient machinery or optimizing the speed or route. During the last decade, the area of machine learning has evolved significantly, and these methods are applicable in many more fields than before. The field of ship efficiency improvement is by using Machine Learning methods is significantly progressing due to the available big volumes of data from online measuring, experiments and computations. This amount of data has made machine learning a powerful tool that has been successfully used to extract information and complex patterns that can be translated into attractive ship energy savings. This article, therefore, presents an overview of past history, current developments, and emerging opportunities of Machine Learning for ship efficiency. This article covers the fundamentals of Machine Learning and discusses the methodologies available for ship efficiency optimization. Besides, this article reveals the potentials of this promising technology and future challenges.

KEY WORDS: Machine learning; ship efficiency; Optimization; big data.

INTRODUCTION

About 70% of the Earth's surface is covered by water, and approximately 90% of all transports are waterborne. In the long term, maritime transport will still be the most common option for goods, as marine fuels are much cheaper compared to other main transport modes, and the amount of cargo can be carried on a ship is comparable to 2000 trucks, or 2500 airplanes, or 225 trains (Stamatopoulou and Psaraftis 2013). However, for the year 2012, global shipping emissions were approximately 938 million tonnes CO_2 and 961 million tonnes CO_{2e} for GHGs combining CO_2, CH_4 and N_2O; This signifies around 2.2% of global anthropogenic Greenhouse Gases (GHG) (Smith et al. 2014). By 2050, the maritime transport segment needs to reduce its total annual GHG emissions by 50% compared to 2008 to be in line with the global GHG reduction target to limit the global temperature rise to no more than 2 ° C above the pre-industrial level (Cames et al. 2015). With the trend of global warming, the dominate role of waterborne transport means great importance to optimise maritime efficiency, thus achieve green shipping.

Optimising maritime transport has a long history and been an ongoing task. Since hundreds of years ago, naval architects have started to aspire better hull forms so the ships would feel less resistance when operating in water. Although those approaches are mainly empirical and based on simplified classic physics, they did establish the fundamental theories of naval architecture, significantly improved hull design and brought up several centuries of maritime blossom. This is then accompanied by the optimisation of marine engines after the industrial revolution; by improving the engine efficiency, less fuel would be required.

More recently, with the development of computer technique, ship design becomes viable using the Computational Fluid Dynamics (CFD) method to produce highly realistic sailing simulations (Jasak 2017). At the same time, enhancement in satellite observation has allowed ships to plan their voyages based on weather conditions, which has been improving maritime sustainability and safety by always choosing an optimised route.

During the last decade, many data management frameworks supporting distributed storage have been developed. Following this big data trend, Machine Learning (ML) has stepped into the shipping industry and is transforming it in a way that has never been seen before. Big data in this field has been established based on Geospatial Data Systems such as Copernicus Marine Service (Schuckmann et al. 2018). Those systems integrate historical weather data and provide future projections to support voyage planning. On top of that, ship fuel consumptions corresponding to specific weather conditions can also be recorded. The amount of data may no longer be handled together by traditional manual methods; instead, ML can help to ascertain the rules within and more importantly it gives the chance to give integrated operations. Subsequently, ML analyses can point out ways to

*Corresponding author: ucemlhu@ucl.ac.uk (L. Huang)

improve shipping efficiency and reduce emissions. Moreover, this process can be automated along with the update of data in real-time.

Fig. 1 : An conceptional illustration of big-data-oriented shipping (CIAOTECH Srl 2019).

Despite the great potential of ML in the shipping field, it may sound fearing for people not from a Computer Science background to conduct relevant research. In such a context, this paper reviews how ML has been applied in this field and has facilitated a green shipping industry. As demonstrated in Fig. 1. the concept of machine learning in the shipping industry relies on a data stream from and to the ships, which is analysed onshore (CIAOTECH Srl 2019). It covers relevant applications in naval architecture, marine engine design and route planning. Since it is a cross-discipline work between Ocean Engineering and Computer Science, the present work provides an overview for scholars from different fields to better understand the mechanism and foresee future research opportunities. As a result, this work aims to promote the advancement of ML in waterborne, thus towards achieving a zero-emission future.

MACHINE LEARNING FUNDAMENTALS

ML has demonstrated to be a feasible alternative to solve conventional engineering problems when development cost and time are constraints. In addition, it has demonstrated to be effective in solving extremely complex engineering problems that the current analysis methods cannot solve.
According to (Simeone 2017), Machine Learning can be implemented when the following applies:

- *'The task involves a function that maps well-defined inputs to well-defined outputs;*
- *Large data sets exist or can be created containing input-output pairs;*
- *The task provides clear feedback with clearly definable goals and metrics;*
- *The task does not involve long chains of logic or reasoning that depend on diverse background knowledge or common sense;*
- *The task does not require detailed explanations for how the decision was made;*
- *The task has a tolerance for error and no need for provably correct or optimal solutions;*
- *No specialised dexterity, physical skills, or mobility is required'.*

Depending on how the learning task is achieved, machine learning algorithms can be classified into *Supervised Learning, Unsupervised Learning, Semi-supervised Learning and Reinforcement Learning.* In ML a feature is an input to the model, that is a variable which is used for training and for feeding the trained model. A label is a true value, what is used for the training of the model. (Raschka and Mirjalili 2017).

The term supervised learning represents a tool to classify and process data in a relatively simple way. A supervised learning algorithm relies on a set of input dataset whose characteristics, output and relationship are known and also requires experience from the ML-engineer. A learning algorithm then trains a model to generate a prediction for the response to new data or the test dataset. The most known techniques are linear regression and classification techniques. Linear regression is typically used to predict relationships between quantitative data. A very common example that is used to illustrate its capability is the linear relationship between a radiation therapy and a tumour size. The classification techniques, on the other hand, predict a relationship by analysing data and distinguishing patterns. This technique is typically used to predict whether a credit card transaction is fraudulent or not.

There are two classes of models in supervised learning, a parametric model is when the model has a fixed number of parameters and a non-parametric model the parameters grows with the amount of training data. Parametric algorithms include the linear regression, logistic regression, least shrinkage and selection operator regression (LASSO) and linear discriminant analysis (LDA), where non-parametric models include gaussian process (GP) and Support Vector Machines (SVM). Parametric models are faster, but with less flexibility in comparison with non-parametric. Linear regression is the linear relationship with the model features, it is common to use the ordinary least squares method for fitting the model, where each data point residual (distance between model fit and training data) is squared. Logistic regression is a model that uses a logistic function instead of a linear function. The LASSO regression uses shrinkage methods to minimize the needed inputs (features) for the model response (prediction), and features that have their regression coefficient shrunken to zero are excluded. An LDA calculates the mean and variance for each dimension, and is often used in pre-processing, feature extraction and dimensionality reduction. A GP is a non-parametric approach using Bayesian inference over many possible functions, which means no prior assumptions are made on the functions for training. The SVM works by separating the decision boundaries by hyperplanes (support vectors), and the most simple SVM is when a two-class data is separated by a linear hyperplane (Murphy 2012; Gareth et al. 2013)

*Corresponding author: ucemlhu@ucl.ac.uk (L. Huang)

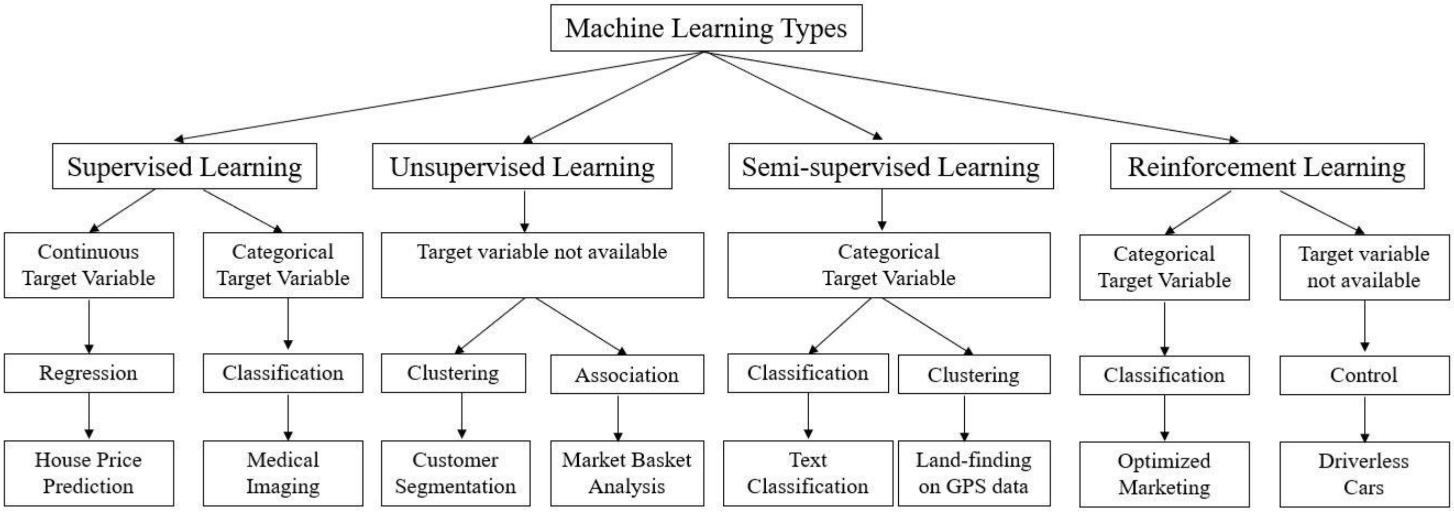

Fig. 2: A classification map of various ML techniques and their application examples (reproduce based on information provided on proft.me).

There are a vast number of different algorithms for machine learning, as an example, the popular Python ML-library Scikit-learn comprises 17 different linear regression (Pedregosa et al. 2011). Depending on how large the data are, and how strong is the connection between the data and the output, a model can be chosen out of the experience of the data scientist.

As demonstrated in Fig. 2, the process of ML requires several steps depending on the task and type of data. If the amount of data is too small it is often not feasible with machine learning, but instead it is better with physical modelling. Also, if the model is going to be used for classification, clustering or regression, it limits the choice of model. In essence, all algorithms are trained by minimising the error of the predicted value with the training data.

Generally, if the amounts of training data are growing, the applicability of artificial neural networks (ANN) demonstrate better accuracy. An ANN is inspired of the neurons in a biological brain. ANN's have proven to be successful in many areas, such as natural language processing and image classification. It consists of units called perceptions that act as thresholds to an input, the perceptron can receive multiple inputs, and these are multiplied by a weight and passes an activation function. An ANN is trained by optimisation of the weights of the perceptrons. The network can consist of several layers, which is how deep the network is. If the network is larger it generally needs more data for training. Generally, a Deep neural network (DNN) is where there are several hidden layers between input and output layers (Raschka and Mirjalili 2017).

The term unsupervised learning, however, finds structures in data which for multiple reasons has not been labelled before. This fact makes unsupervised learning attractive in applications where there is a large amount of data or where data labels are simply not available. The main techniques used in unsupervised learning are principal component (used to group datasets with shared attributes to extrapolate other data relationships) and cluster analysis (which analyses and identifies relationships in the input data which are then extrapolated to a new dataset). Principal component analysis (PCA) is a commonly used technique for dimensionality reduction, in unsupervised learning and in exploratory data analysis. It can help identify the correlation between features, and works by finding the maximum variance in high dimension data and projecting this to fewer dimensions. When having a large number of correlated variables a PCA can explain these with fewer dimensions. The K-nearest neighbour (KNN) is a non-parametric model that can be used as a classifier for clustering data, it looks at the points which are closest to the nearest centroid (Gareth et al. 2013).

Semi-supervised Learning, on the other hand, can be considered a hybrid between supervised and unsupervised learning which combines a small amount of labelled data with a large amount of unlabelled data which is typically used during the algorithm training. The labelled data improves the learning accuracy without the necessity of producing a large amount of data that are required in a supervised learning algorithm.

As there are numerous different algorithms and frameworks to use in the field of ML, the concept of automated Machine Learning (AutoML) is growing, these are tools that optimise on both the pre-processing, the model selection as well as the tuning of hyper-parameters. There are several opensource AutoML tools, tree-based pipeline optimisation tool, Autosk-learn, AutoWeka and more (Cortes et al. 2013; Olson et al. 2016; Kotthoff et al. 2017).

Another approach, called Reinforcement Learning enables an algorithm to learn by using a continuous trial and error approach. A reinforcement learning algorithm can consist of several elements, reward signal, a policy value function and a sometimes a model of the environment. The policy is the rules defining the actions of the model, the reward signal is what defines the goal of

*Corresponding author: ucemlhu@ucl.ac.uk (L. Huang)

the learning problem and the value function is the total reward over time. Q-learning is one of several reinforcement learning algorithms that optimise the learning outcome by random values for the policy, that means actions are off-policy (Sutton and Barto 2018). This technique is particularly important in the area such as active flow control and ship design, which could be used for improving the efficiency of ships.

APPLICABILITY

Naval Architecture Design

In terms of naval architecture, the design of a vessel constitutes an essential task to achieve superior hydrodynamic performance to minimise fuel consumption. Designing a ship relies on sophisticated experimental and computational techniques for hydrodynamic performance evaluation of multiple hull sizes and shapes; which at the same time require significant time as well as expertise setting up the complex problem physics. Traditionally, naval architects have obtained a ship hydrodynamic performance guidance from existing hull-forms. Linear regression analyses have been typically used at the early stages of the design to select the hull-form geometric coefficients based on the performance of existing vessels. This approach, however, may lead to multiple disappointments at later stages of the design since a regression approach does not consider the non-linearity relationship between performance and geometry. During a later optimisation process, which is a critical step in improving the performance of vessels, ship designers rely on their personal experience assisted with the direct simulations and experiments results. However, this traditional approach largely depends on the designer's skills and could make it hard to find the most optimal configuration without spending a significant amount of time testing multiple geometry combinations during the optimisation phase.

A strong impetus has aimed at turning a tedious ship design into a much simpler process. These attempts have been facilitated the fast development of artificial intelligence together with the availability of High-Performance Computers (HPC); so now a semi-automatic ship design process has been made a reality. The first applications in the area of machine learning could be considered the Holtrop and Mennen's empirical algorithms which present a statistical method to determine the ship resistance based on the results of multiple model basin tests. On the other hand, pioneers in the area of assisted ship design were Ray and Sha (1994). They incorporated accepted naval architectural estimation methods, a decision system handler and a non-linear optimisation tool with a decision system which identifies the weights corresponding to different objectives based on the relative importance of the objectives using multi-attribute decision-making methods. This approach was used during the containership design. However, accurately modelling of non-linear hydrodynamic phenomena for the purposes of ship design is a highly sophisticated task which requires looking into several design variants to account for the hydrodynamic performance fully. Neural networks, therefore, are applicable for the purpose of modelling a phenomenon in which mathematical nature cannot be determined, or in which the model is too complicated such as in ship hydrodynamics. However, the development of an accurate mathematical model requires a sufficient number of observation results for a given process so the determined relations can be used as a component or a block of computational models. Neural networks may also serve for the simplification of some earlier developed models in such a way that the variables left in the model are only those that are important for a given phase or a formulation of a design problem.

Cui et al. (2012) proposed a Q-learning reinforcement learning optimisation approach which is based on the human learning process and that successfully introduced as a useful tool during the ship optimisation phase to improve the searching ability. The authors successfully used their reinforcement learning-based approach to improving the structural optimisation process of a bulk-carrier ship with two objectives of weight and fatigue, which was successfully integrated with JAVA and ABAQUS. Their algorithm proved to shows great potential to minimise a ship's structure weight (which could be used to minimise ship's fuel consumption). Cepowski (2020) investigated an ANN to estimate added resistance in regular head waves with the training data obtained through model test experiments. The study showed that added wave resistance values predicted by the neural network soundly correlated with measured data and had good generalisation ability during the first stages of the design to minimise ship hydrodynamic resistance in rough seas. However, it is essential to remember that the predictions of resistance are given in model scale, and therefore, full-scale data is still subject to scaling issues.

Going a bit further, Yu and Wang (2018) revolutionised the ship design process by creating a set of complex hull forms by using a Principal Component Analysis (PCA) approach which generate a large number of derived hull forms, which were evaluated computationally for their hydrodynamic performance. The results from the process were then used to train a Deep Neural Network (DNN) to accurately establish the relation between different hull forms and their associated performances. Then, based on the fast, parallel DNN-based hull-form evaluation, the large-scale search for optimal hull forms is performed. By using this approach, the authors showed a novel application of machine learning which allows first to create an extensive database as well as get fast results. Additionally, Yu et al. (2019) designed an algorithm by predicting ship dynamics to assist the achievement of a thruster & mooring balanced system.

Propulsion Control

Propulsion efficiency is crucial as it governs how much fuel consumption can be actually converted into the ship movement, while this efficiency is not static, usually related in what condition the ship is operating. Upon such data are collected, ML can come in handy here to derive the relationships behind. Petersen et al. (2012) demonstrated the usage of ANN and GP for this purpose. And yet the total energy consumption of a ship is not only dependent on the propulsion, but also the different support systems that produce electricity, heating, ventilation and other auxiliary demands. Similarly, those relationships can be

*Corresponding author: ucemlhu@ucl.ac.uk (L. Huang)

established; For example, Yang et al. (2018) created neural networks for predicting waste heat recovery performance. By analysing how the energy efficiency changes with environmental variables, different components can be designed according to different kinds of operation that vessels are expected to conduct. Another example was given by Raptodimos and Lazakis (2018), in which they apply ML to link monitoring data with situations where machinery failure could happen, thus enabling diagnostic purposes.

Perera and Mo (2016) designed an ML-based automation system consisting of a power management architecture for engine and propulsion control systems with respect to various engine room operations. It achieved a coupling control of different engines' power, ship speed, shaft speed and corresponding fuel consumptions. Meanwhile, a marine engine centred data flow chart has been established to handle large-scale data sets. Thereby, they forged a big data solution that can automatically improve the quality of engine strategies and advise bridge crew on decisions such as speed selection. Nonetheless, a gap here is that different ML approaches can provide notably different accuracies in engine performance prediction. In such a context, Yuan and Wei (2018) compared the outcomes of ANN and GP in this procedure and found out GP provides more accurate data; however, as Petersen et al. (2012) indicated, there still lack benchmarking cases that can be used to verify different methods, thus the conclusion of Yuan and Wei can be one case but cannot generally mean GP is best option to optimise shipping energy. Ongoing work within this area will focus on improving these models, considering the possibility to combine them so that different variables can all be dealt with their suitable ML methods.

On the other hand, marine diesel engines operating with heavy fuel oil or marine diesel oil are not a viable powering solution for the shipping industry in terms of the required reduction in GHG and pollutants. There have been trends to develop green and renewable energies to alternatively power ships. Wu and Backnall (2020) designed a hybrid fuel cell and Lithium-ion battery propulsion system for vessels. This system achieves complementation between the two powering methods: since fuel cell has the shortcoming of slow response, Lithium-ion battery can cover the ac/deceleration processes; whereas Lithium-ion battery is very slow to refill, the fuel cell can be used as the primary energy source. They provided simulations to repeat previous voyages and proved a minimum 65% GHG emission reduction can be achieved by utilising the hybrid system. Subsequently, the next obvious question is to determine when to swap between the methods in a certain operation scenario. To address this, they applied Reinforcement Learning (Wu et al. 2020) to allocate the optimal strategies for different scenarios. Nevertheless, the novel system is just applicable to coastal vessels committing short-distance voyages, as it is limited by the total amount of fuel carriage. For global cargo shipping, marine diesel will still be the dominant energy resource in a couple of decades, thus using ML methods to optimise the traditional engine efficiency and reduce wastes will still be a worthwhile research area.

Maritime Operations

In recent years, with the benefits of reducing marine incidents as well as optimising energy efficiencies, there have been increasing deployments of automated route planning which is currently supported by weather routing systems and radar systems. In this approach, environmental factors such as the wave height, direction, wind and currents as well as densities and temperatures of air and water are considered, while radars are normally used to identify other vessels and obstacles to secure safety.

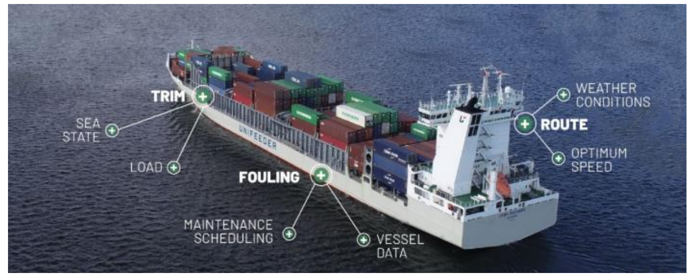

Fig. 3: Demonstration of variables for ML optimisation (GreenSteam 2019)

Voyage Planning Tools (VPT) based on weather systems are often achieved by Ship Performance Models (SPM), where respond surface of ships is built for various input conditions. Such respond surface can be built upon empirical equations using extensive data from experiments or simulations, traditionally by regression methods, and now compatible with ML.

As shown in Fig. 3, apart from hull design and engine performance, the efficiency of a ship is also related to its current trim and fouling, Also, choosing an optimised route is essential for time and fuel savings (GreenSteam 2019). Tillig (2020) proposed an SPM, which is a generic ship energy systems model to predict the fuel consumption under operational conditions with limited required input of the ship's characteristics. The model can be divided into two main parts: (i) a static part for calm water power prediction based on empirical methods and standard propeller and hull series as well as the estimation of all required ship dimensions and properties using empirical formulas, and (ii) a dynamic part for the analysis of the required power under realistic operational conditions, including effects from wind, waves, current, temperature differences, biofouling and shallow water. Based on such an SPM, the VPT can map out the fuel consumption of all potential routes and choose the best one: like "Google Map" on oceans.

The use of ML for predicting fuel consumption has been demonstrated in several studies. Wang et al. (2018) proposed a Least Absolute Shrinkage and Selector Operator (LASSO) regression predicting the fuel consumption for several container ships, with features on ship and weather data extracted from a fleet management system. The LASSO regression was able to produce better results compared to Support Vector Machines (SVR), ANN and GP. Meng et al. (2016) used operational data, based on 24 h daily snapshots, from two sister container ships and using regression modelling predicting the daily fuel consumption. Bal et al. (2016) demonstrated an ANN with training data based

*Corresponding author: ucemlhu@ucl.ac.uk (L. Huang)

on noon-reports from an oil tanker, which was used as an input a forecasting model to a decision support system. Ahlgren et. al (2019) demonstrated a method predicting the fuel consumption of a cruise ship, with an automated mechanism using noon report data together with logged machinery data. Gkerekos et al. (2019) did a comparative study of different ML methods for predicting fuel consumption of two different ships, and demonstrating results of R^2 scores of ~90 % for noon report data.

Liu and Bucknall (2015) designed an algorithm for planning routes of Unmanned Surface Vehicles (USV) that can achieve avoiding obstacles. They applied the Fast Marching (FM) method that can identify corresponding safe shipping area and forbidden area in real-time, to ensure the planned trajectory to not encounter any obstacle. The method works in both static environments (with natural obstacles, offshore structures etc.) and dynamic environments (with other moving vessels). Chen et al. (2019) demonstrated the usage of Reinforcement Learning to train USVs, in which the ships can be rewarded based on how rational the decisions are, and the route optimisation can be done by choosing the best reward value; however, their work only considered a static environment thus still need to incorporate a dynamic environment as Liu and Bucknall did. Similar examples can also be found using Deep Learning (see (Perera 2020)).

A combination of both energy saving and obstacle avoidance has been done by the VPT of Li et al. (2020a) In their application, an SPM has been linked with ice conditions to guide ship navigation in the Arctic. On one hand, the VPT can choose a route with the least fuel consumption; on the other hand, it receives ice conditions from satellite to avoid encountering significant ice conditions such as icebergs and ice ridges. It can reroute automatically considering the drifting direction of icebergs. Following validation, the fuel consumption predicted by their model has agreed well with full-scale measurement data (Li et al. 2020b). The work of Li et al. has demonstrated the excellent potential to apply AI technique in this area to handle the non-negligible ice data and risks, which is motivated particularly by the opening of Arctic shipping routes in recent years (Huang et al. 2020a; 2020b). Another example of applying ML technique to predict ship speed in ice fields has been given by Milaković et al. (2019). As ship-ice interactions contain very complex physics, using ML in this field has revealed an advantage by using derived relationships rather than to run an advance simulation for every input condition (Huang et al. 2020c).

DISCUSSION

This review has presented the applicability of Machine Learning fundamentals and algorithms in optimizing shipping efficiency. It highlighted successful implementations of ML in three main fields, which are naval architecture, propulsion control and maritime operations:

- In naval architecture, ML algorithms based on statistical regression have been traditionally used as part of the design process. New applications which facilitate a semi-automatic ship design process from a hydrodynamic or structural perspective have been made possible. This includes data-driven optimisation and applied regression techniques that are well suited for non-linear problems which are typically encountered during the design of a vessel.

- Propulsion control relies on machine learning algorithms to establish relationships of fuel consumption with engine powers, ship speed, shaft speed, energy wastes and weather data, by which an optimal propulsion setup can be advised in a given navigation condition. Moreover, green engine options such as fuel cells and batteries have been developed as alternatives to traditional marine diesel – they are optimized by ML.

- From a maritime operations point of view, there have been increasing deployments of automated route planning which consider factors such as weather forecasts as well as route obstacles that can be encountered by ships, being governed an objective function of fuel consumption and sailing risks. These ML techniques have demonstrated to achieve considerable fuel savings.

There are several other areas in shipping that can be facilitated by ML but not explicit in the three areas classified in the present paper, such as underwater vehicles, wave energy converters, condition monitoring and maintenance. Ports can leverage on ML for real time data from cargo containers, and also minimising the manual work of paperwork. Also, not only the fuel consumption but also the air emissions and underwater noise can be reduced by combining data from an engine performance and routes.

Despite ML algorithms have demonstrated their capabilities in ship efficiency, traditional knowledge still dominates the maritime industry. One of the reasons could be that the algorithms firstly rely on big amounts of data while requiring high computational costs. The marine industry has been traditionally conservative and still reluctant to share data which does not support the data training and validation process of ML. Still, creating, optimising and maintaining relevant algorithms will require expertise and extensive human inputs, given the fact that errors in ship design calculations and marine operations could carry catastrophic consequences. Therefore, human validation and verification will be needed in the foreseeable future. It is also expected that the development of HPC resources, as well as the availability of the 5th generation mobile network (5G), will facilitate the process, particularly during route-planning that requires constant data transmission. In addition, ML techniques could be used to link this data with ship structural issues and safety incidents. The key to such applications will require advanced measurement network and sensors/monitors on ships.

With large databases available to the scientific community, linear approaches and old-fashioned empirical functions will be slowly substituted by Machine Learning algorithms that consider the usually neglected physics non-linearity and complexity. This is the case of the current-in-use guideline formulae such as the

*Corresponding author: ucemlhu@ucl.ac.uk (L. Huang)

ITTC-57 and ITTC-78 (ITTC 2008; 2014) which calculate friction coefficients based on empirical results from model scale experiments and for old ship geometries. In addition, large databases could replace the simplest computations such as CFD resistance or self-propulsion studies which are constantly used as part of the evaluation of ship performance. Improvements in the calculation of ship hydrodynamic performance characteristics when using ML algorithms could be achieved by the implementation of the latest approaches for evaluation of ship flows. For example, Yu and Wang (2018) method could see further accuracy improvement when combined with the most advanced turbulence modelling approach for detailed and complex analyses of the flow around ships (Pena et al. 2019) which has shown exceptionally accuracy and reliability when compared to conventional RANS methods (Pena et al. 2020). Yu and Wang's DNN-based method could therefore achieve higher accuracy if trained with enhanced CFD simulation results.

Also, ship design, as still a long and complicated process, is expected to be slowly transitioned into an automated process which will just require a set of inputs to find the most efficient solution with little human intervention and replacing large teams. The industry should be aware that such a scenario would cause corresponding job cuts. According to a study by World Maritime University, the new technologies will likely result in a shift in the workforce, rather than a labour reduction (Schröder-Hinrichs et al. 2017). On the whole, ML is expected to substantially support shipping sustainability.

Based on the above progress, there has arisen an excited scientific community and learners who believe that Machine Learning has the capability to resolve any kind of practical problem. However, this group of people should still be very careful, since the data-based solution is easy to go a deviated way from classic mathematical and physical principles, which should still be prerequisite. ML efficiency and efficacy are totally dependant on properly selecting the algorithm and the training that is used as part of the process. Factors such as the quality and quantity of data, the desired outputs and the target function must be reasonably selected but those are not guaranteed, and all ML users should avoid the laziness of selecting an algorithm without sufficient data training. Demonstrating the validity of an algorithm should go through a comprehensive validation process in multiple problems and geometries to ensure that overfitting or underfitting is not an issue. And yet, to date there has not formed conclusive options on which ML algorithm is the best for a given application. The development of many algorithms are still in an ongoing base and need extensive calibrations. A very possible scenario is, as Cheng et al. (2020) demonstrated, that a specifically optimal ML technique is a combination of multiple basic methods.

REFERENCES

Ahlgren, Fredrik, Maria E. Mondejar, and Marcus Thern. 2019. 'Predicting Dynamic Fuel Oil Consumption on Ships with Automated Machine Learning'. *Energy Procedia* 158: 6126–6131.

Beşikçi, E. Bal, O. Arslan, O. Turan, and A. I. Ölçer. 2016. 'An Artificial Neural Network Based Decision Support System for Energy Efficient Ship Operations'. *Computers & Operations Research* 66: 393–401.

Cames, Martin, Jakob Graichen, Anne Siemons, and Vanessa Cook. 2015. 'Emission Reduction Targets for International Aviation and Shipping'. *Directorate General for Internal Policies; European Parliament—Policy Department A; Economic and Scientific Policy: Bruxelles, Belgium*.

Cepowski, Tomasz. 2020. 'The Prediction of Ship Added Resistance at the Preliminary Design Stage by the Use of an Artificial Neural Network'. *Ocean Engineering* 195 (January): 106657. https://doi.org/10.1016/j.oceaneng.2019.106657.

Chen, Chen, Xian-Qiao Chen, Feng Ma, Xiao-Jun Zeng, and Jin Wang. 2019. 'A Knowledge-Free Path Planning Approach for Smart Ships Based on Reinforcement Learning'. *Ocean Engineering* 189: 106299.

Cheng, Chen, Peng-Fei Xu, Hongxia Cheng, Yanxu Ding, Jinhai Zheng, Tong Ge, Dianhong Sun, and Jin Xu. 2020. 'Ensemble Learning Approach Based on Stacking for Unmanned Surface Vehicle's Dynamics'. *Ocean Engineering* 207: 107388.

CIAOTECH Srl. 2019. 'Autonomous Shipping Initiative for European Waters'. *EU Horizon 2020*, 2019.

Cortes, Corinna, M. Kloft, and M. Mohri. 2013. *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc.

Cui, Hao, Osman Turan, and Philip Sayer. 2012. 'Learning-Based Ship Design Optimization Approach'. *Computer-Aided Design* 44 (3): 186–95. https://doi.org/10.1016/j.cad.2011.06.011.

Gareth, James, Witten Daniela, Hastie Trevor, and Tibshirani Robert. 2013. *An Introduction to Statistical Learning: With Applications in R*. Spinger.

Gkerekos, Christos, Iraklis Lazakis, and Gerasimos Theotokatos. 2019. 'Machine Learning Models for Predicting Ship Main Engine Fuel Oil Consumption: A Comparative Study'. *Ocean Engineering* 188: 106282.

*Corresponding author: ucemlhu@ucl.ac.uk (L. Huang)

GreenSteam. 2019. 'Innovative Partnership Paves The Way For Sustainable Shipping With Machine Learning'. *International Shipping News*, 2019.

Huang, Luofeng, Minghao Li, Tuomas Romu, Azam Dolatshah, and Giles Thomas. 2020. 'Simulation of a Ship Operating in an Open-Water Ice Channel'. *Ships and Offshore Structures* 0 (0): 1–10. https://doi.org/10.1080/17445302.2020.1729595.

Huang, Luofeng, Zhiyuan Li, Christopher Ryan, Minghao Li, Jonas Ringsberg, Bojan Igrec, GLR Andrea, Dimitris Stagonas, and Giles Thomas. 2020. 'Ship Resistance When Operating in Floating Ice Floes: A Derivation of Empirical Equations'. In *ASME 2020 39th International Conference on Ocean, Offshore and Arctic Engineering (OMAE)*.

Huang, Luofeng, Jukka Tuhkuri, Bojan Igrec, Minghao Li, Dimitris Stagonas, Alessandro Toffoli, Philip Cardiff, and Giles Thomas. 2020. 'Ship Resistance When Operating in Floating Ice Floes: A Combined CFD&DEM Approach'. *Marine Structures*. 74, p.102817.

ITTC. 2008. '1978 ITTC Performance Prediction Method'. *Recommended Procedures and Guidelines*.

ITTC. 2014. 'Example for Uncertainty Analysis of Resistance Tests in Towing Tanks'. *Recommended Procedures and Guidelines*.

Jasak, Hrvoje. 2017. 'CFD Analysis in Subsea and Marine Technology'. *IOP Conference Series: Materials Science and Engineering* 276 (1): 012009. https://doi.org/10.1088/1757-899X/276/1/012009.

Kotthoff, Lars, Chris Thornton, Holger H. Hoos, Frank Hutter, and Kevin Leyton-Brown. 2017. 'Auto-WEKA 2.0: Automatic Model Selection and Hyperparameter Optimization in WEKA'. *The Journal of Machine Learning Research* 18 (1): 826–830.

Li, Zhiyuan, Jonas W. Ringsberg, and Francisco Rita. 2020. 'A Voyage Planning Tool for Ships Sailing between Europe and Asia via the Arctic'. *Ships and Offshore Structures*, 1–10.

Li, Zhiyuan, Christopher Ryan, Luofeng Huang, Li Ding, J. W. Ringsberg, and Giles Thomas. 2020. 'A Comparison of Two Ship Performance Models against Full-Scale Measurements on a Cargo Ship on the Northern Sea Route'. In *The 5th International Conference on Ships and Offshore Structures*.

Liu, Yuanchang, and Richard Bucknall. 2015. 'Path Planning Algorithm for Unmanned Surface Vehicle Formations in a Practical Maritime Environment'. *Ocean Engineering* 97: 126–144.

Meng, Qiang, Yuquan Du, and Yadong Wang. 2016. 'Shipping Log Data Based Container Ship Fuel Efficiency Modeling'. *Transportation Research Part B: Methodological* 83: 207–229.

Milaković, Aleksandar-Saša, Fang Li, Mohamed Marouf, and Sören Ehlers. 2019. 'A Machine Learning-Based Method for Simulation of Ship Speed Profile in a Complex Ice Field'. *Ships and Offshore Structures*, 1–7.

Murphy, Kevin P. 2012. *Machine Learning: A Probabilistic Perspective*. MIT press.

Olson, Randal S., Ryan J. Urbanowicz, Peter C. Andrews, Nicole A. Lavender, and Jason H. Moore. 2016. 'Automating Biomedical Data Science through Tree-Based Pipeline Optimization'. In *European Conference on the Applications of Evolutionary Computation*, 123–137. Springer.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, and Vincent Dubourg. 2011. 'Scikit-Learn: Machine Learning in Python'. *The Journal of Machine Learning Research* 12: 2825–2830.

Pena, Blanca, E. Muk-Pavic, Giles Thomas, and Patrick Fitzsimmons. 2020. 'An Approach for the Accurate Investigation of Full-Scale Ship Boundary Layers and Wakes.' *Ocean Engineering*, 214, p.107854.

Pena, Blanca, Ema Muk-Pavic, and Dmitriy Ponkratov. 2019. 'Achieving a High Accuracy Numerical Simulations of the Flow around a Full Scale Ship'. In *International Conference on Offshore Mechanics and Arctic Engineering*, 58844:V07AT06A058. American Society of Mechanical Engineers.

Perera, Lokukaluge P. 2020. 'Deep Learning Toward Autonomous Ship Navigation and Possible COLREGs Failures'. *Journal of Offshore Mechanics and Arctic Engineering* 142 (3).

Perera, Lokukaluge P., B. Mo, and Guedes Soares. 2016. 'Machine Intelligence for Energy Efficient Ships: A Big Data Solution'. *Maritime Engineering and Technology III, Guedes Soares & Santos (Eds.)* 1: 143–150.

Petersen, Jóan Petur, Daniel J. Jacobsen, and Ole Winther. 2012. 'Statistical Modelling for Ship Propulsion Efficiency'. *Journal of Marine Science and Technology* 17 (1): 30–39.

Raptodimos, Yiannis, and Iraklis Lazakis. 2018. 'Using Artificial Neural Network-Self-Organising Map

*Corresponding author: ucemlhu@ucl.ac.uk (L. Huang)

for Data Clustering of Marine Engine Condition Monitoring Applications'. *Ships and Offshore Structures* 13 (6): 649–656.
Raschka, Sebastian, and Vahid Mirjalili. 2017. *Python Machine Learning*. Packt Publishing Ltd.
Ray, Tapabrata, and Om Prakash Sha. 1994. 'Multicriteria Optimization Model for a Containership Design'. *Marine Technology and SNAME News* 31 (04): 258–268.
Schröder-Hinrichs, J. U., D. W. Song, T. Fonseca, K. Lagdami, X. Shi, and K. Loer. 2017. 'Transport 2040: Automation, Technology, Employment-The Future of Work'. *World Maritime University, Malmö, Report*.
Schuckmann, Karina von, Pierre-Yves Le Traon, Neville Smith, Ananda Pascual, Pierre Brasseur, Katja Fennel, Samy Djavidnia, Signe Aaboe, Enrique Alvarez Fanjul, and Emmanuelle Autret. 2018. 'Copernicus Marine Service Ocean State Report'. *Journal of Operational Oceanography* 11 (sup1): S1–S142.
Simeone, Osvaldo. 2017. 'A Brief Introduction to Machine Learning for Engineers'. *ArXiv Preprint ArXiv:1709.02840*.
Smith, T. W. P., J. P. Jalkanen, B. A. Anderson, J. J. Corbett, J. Faber, S. Hanayama, E. O'keeffe, S. Parker, L. Johansson, and L. Aldous. 2014. 'Third IMO Greenhouse Gas Study 2014'. *Int. Marit. Organ* 327.
Stamatopoulou, Eirini V., and Harilaos N. Psaraftis. 2013. 'Northern Sea Route: Effect on Modal Shift and Modal Choice'. In *International Association of Maritime Economists Conference (IAME 2013):"Managing Complexity in Shipping and Port Markets"*.
Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction*. MIT press.
Tillig, Fabian. 2020. 'Simulation Model of a Ship's Energy Performance and Transportation Costs'. PhD Thesis, Chalmers University of Technology.
Wang, Shengzheng, Baoxian Ji, Jiansen Zhao, Wei Liu, and Tie Xu. 2018. 'Predicting Ship Fuel Consumption Based on LASSO Regression'. *Transportation Research Part D: Transport and Environment* 65: 817–824.
Wu, Peng, Julius Partridge, and Richard Bucknall. 2020. 'Cost-Effective Reinforcement Learning Energy Management for Plug-in Hybrid Fuel Cell and Battery Ships'. *Applied Energy*, 275, p.115258.
Wu, Peng, and Richard Bucknall. 2020. 'Hybrid Fuel Cell and Battery Propulsion System Modelling and Multi-Objective Optimisation for a Coastal Ferry'. *International Journal of Hydrogen Energy* 45 (4): 3193–3208.
Yang, Fubin, Heejin Cho, Hongguang Zhang, Jian Zhang, and Yuting Wu. 2018. 'Artificial Neural Network (ANN) Based Prediction and Optimization of an Organic Rankine Cycle (ORC) for Diesel Engine Waste Heat Recovery'. *Energy Conversion and Management* 164: 15–26.
Yu, Dongchi, and Lu Wang. 2018. 'Hull Form Optimization with Principal Component Analysis and Deep Neural Network'. *ArXiv Preprint ArXiv:1810.11701*.
Yu, Shangyu, Lei Wang, Bo Li, and Huacheng He. 2019. 'Optimal Setpoint Learning of a Thruster-Assisted Position Mooring System Using a Deep Deterministic Policy Gradient Approach'. *Journal of Marine Science and Technology*, 1–12.
Yuan, Jun, and Shumei Wei. 2018. 'Comparison of Using Artificial Neural Network and Gaussian Process in Ship Energy Consumption Evaluation'. *DEStech Transactions on Environment, Energy and Earth Sciences*, no. iceee.

*Corresponding author: ucemlhu@ucl.ac.uk (L. Huang)