



A survey of the opportunities and challenges of supervised machine learning in maritime risk analysis

Andrew Rawson & Mario Brito

To cite this article: Andrew Rawson & Mario Brito (2022): A survey of the opportunities and challenges of supervised machine learning in maritime risk analysis, Transport Reviews, DOI: [10.1080/01441647.2022.2036864](https://doi.org/10.1080/01441647.2022.2036864)

To link to this article: <https://doi.org/10.1080/01441647.2022.2036864>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 07 Feb 2022.



Submit your article to this journal [↗](#)



Article views: 219



View related articles [↗](#)



View Crossmark data [↗](#)

RESEARCH ARTICLE



A survey of the opportunities and challenges of supervised machine learning in maritime risk analysis

Andrew Rawson ^a and Mario Brito ^b

^aElectronics and Computer Science, University of Southampton, Southampton, UK; ^bCentre for Risk Research, Southampton Business School, University of Southampton, Southampton, UK

ABSTRACT

Identifying and assessing the likelihood and consequences of maritime accidents has been a key focus of research within the maritime industry. However, conventional methods utilised for maritime risk assessment have been dominated by a few methodologies each of which have recognised weaknesses. Given the growing attention that supervised machine learning and big data applications for safety assessments have been receiving in other disciplines, a comprehensive review of the academic literature on this topic in the maritime domain has been conducted. The review encapsulates the prediction of accident occurrence, accident severity, ship detentions and ship collision risk. In particular, the purpose, methods, datasets and features of such studies are compared to better understand how such an approach can be applied in practice and its relative merits. Several key challenges within these themes are also identified, such as the availability and representativeness of the datasets and methodological challenges associated with transparency, model development and results evaluation. Whilst focused within the maritime domain, many of these findings are equally relevant to other transportation topics. This work, therefore, highlights both novel applications for applying these techniques to maritime safety and key challenges that warrant further research in order to strengthen this methodological approach.

ARTICLE HISTORY

Received 21 June 2021
Accepted 25 January 2022

KEYWORDS

Machine learning; navigation safety; AIS data; maritime; risk assessment; accidents

1. Introduction

The movement of vessels across seas and between continents involves the potential for significant accidents. Vessels might run aground, collide, be lost in storms or any number of other hazards could result in loss of life and environmental disaster. Yet, these risks have changed significantly over the previous 500 years. Estimates for the sixteenth through early nineteenth century are that voyages between Europe and Asia might involve the loss of a vessel around 5% of trips (Bowen, 2020). By 1910, the loss rate per year had reduced to 1%, and in recent years is in the order of 0.1% (Allianz, 2012).

CONTACT Andrew Rawson  a.rawson@soton.ac.uk  Electronics and Computer Science, University of Southampton, Highfield, Southampton SO17 1BJ, UK

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Similarly, the number of oil spills has reduced from 78.8 per year in the 1970s to 6.2 per year in the 2010s, and the total amount spilt has reduced by 95% during that same period (ITOPF, 2020). Huge technical, procedural and technological changes have enabled this reduction in accidents at sea. Developments in Radar, ship design, Global Navigation Satellite Systems (GNSS), improved training, the Automatic Identification System (AIS) and many other technological breakthroughs have contributed to the safety of navigation (Allianz, 2012; Yang, Wu, Wang, Jia, & Li, 2019).

Yet, major accidents continue to occur. In the eight years 2011–2018, the European Maritime Safety Agency (EMSA) reported more than 3200 incidents per year which resulted in almost 8000 injuries, 700 fatalities, 65 incidents of pollution and 230 vessels lost (EMSA, 2019). Finding new methods to identify, monitor, predict and then mitigate these risks would contribute to reducing the frequency of these accidents still further. As such, a growing body of academic literature has sought to develop quantitative methods to predict accident likelihood and consequence.

Comprehensive reviews of maritime risk analysis techniques are contained in Li, Meng, and Qu (2012), Lim, Cho, Bora, Biobaku, and Parsaei (2018), Chen, Huang, Mou, and van Gelder (2019), Kulkarni, Goerlandt, Li, Banda, and Kujala (2020) amongst many other works, but it can be summarised that only a few approaches dominate the literature on maritime risk analysis. These include statistical analysis of accident data, expert judgement in the form of Bayesian Networks and the development of geometric route models or time-domain simulations. Evaluations of these methods (OpenRisk, 2018; EMSA, 2018) have identified limitations that make their practical implementation for maritime risk assessment both challenging and potentially flawed.

A growing area of interest in transportation research is the application of machine learning methods (Wen, Xie, Jiang, Pu, & Ge, 2021). Machine learning can be described as a subset of artificial intelligence whereby computers make predictions or decisions without being explicitly programmed to perform that task. These models can be supervised, whereby, the model is constructed on data containing both input and outputs, or unsupervised, whereby structure is sought on unlabelled data. Furthermore, machine learning tasks may involve regression, predicting continuous numeric values, or classification, determining classes or types for data points. There are a multitude of applications of supervised machine learning in the maritime domain. These include anomaly detection of vessel transits (Riveiro, Pallotta, & Vespe, 2018), visual identification of vessels (Chen et al., 2018) from imagery or sensors, prediction of fuel consumption and ship efficiency (Uyanik, Karatug, & Arslanoglu, 2020) or path planning of autonomous vessels (Chen, Chen, Ma, Zeng, & Wang, 2019) amongst others. Many of these applications have a view to a possible future of autonomous shipping, supporting the technological requirements necessary for this concept to develop.

A less discussed, but growing area of research is the use of machine learning algorithms to enhance maritime safety (Tang, Tang, Zhang, Du, & Wang, 2019; Jin, Shi, Yuen, Xiao, & Li, 2019). By combining vessel traffic data, incident data and other datasets, machine learning models may be able to differentiate the likelihood or consequences of marine accidents better than conventional methods. Machine learning for risk analysis has seen success in disciplines ranging from landslide susceptibility (Merghadi et al., 2020), flooding (Mojaddadi, Pradhan, Nampak, Ahmad, & Ghazali, 2017) to road traffic collisions (Yuan, Zhou, Yang, Tamerius, & Mantilla, 2017; Wen et al., 2021). However, few

studies have sought to apply such techniques to maritime risk assessment (Tang et al., 2019; Jin, Shi, Yuen, et al., 2019), specifically predicting the likelihood or consequence of vessel accidents. Such an approach could be argued to represent a strong realist view of maritime risk assessment (Goerlandt & Montewka, 2015). In such a view, risk exists objectively as part of the maritime system, and the algorithm seeks to learn this attribute through discovering the relationship between accident events and various input features.

Given the relative sparsity of research on applying machine learning methods for maritime risk assessment, this study contributes a comprehensive review of this embryonic discipline in order to highlight future research opportunities for both practitioners and researchers. Bibliometric analysis is conducted of relevant literature to provide insights into the approaches, datasets, features and methods which have been applied. From this review, a discussion highlights several key themes and collective challenges evident in such works, that should serve as future research directions. Whilst this review focuses on the maritime discipline, many of these findings are equally applicable to challenges faced in other transportation domains.

The remainder of this article is organised as follows. Section 2 outlines the literature review method, through which a selection of the articles are extracted and the criteria against which they are evaluated. Section 3 provides a summary of the outputs of this review identifying the key applications and purposes of these studies. In Section 4, the principal datasets and key features used to develop the models are identified and their strengths and weaknesses evaluated. Section 5 considers methodological considerations such as the types of models employed and how issues of class imbalance are addressed. Finally, Section 6 draws some concluding statements and identifies several research gaps that warrant future work.

2. Literature search and selection

2.1. Data collection

The scope of the review is to identify relevant works where machine learning algorithms have been used to assess maritime risk. The methodology implemented to identify these articles is comparative to other review articles (Lim et al., 2018; Chen, Huang, et al., 2019; Kulkarni et al., 2020) and involves searching for selected keywords in bibliographic databases and manually extracting relevant literature based on the information provided in the titles and abstracts. The search was conducted in the databases of “Web of Knowledge”, “Scopus” and “ScienceDirect” during April 2021. Given the potentially significant number of articles that could be related to either machine learning or maritime risk, a combination (AND) of keywords was used to identify the most relevant articles, namely:

- Keyword 1: Maritime OR Navigation.
- Keyword 2: Risk OR Safety OR Accident.
- Keyword 3: Machine Learning OR Deep Learning OR Logistic Regression OR Support Vector OR Random Forest OR Artificial Neural Networks.

The first set of keywords was used to limit the search to articles that relate to the marine discipline, principally the navigation of vessels. Secondly, an additional keyword sought to limit the search to articles related to the frequency or severity of accidents. Finally, as opposed to alternative methods used for maritime risk assessment, the use of keywords related to specific popular algorithms was used. In total more than 1000 articles were identified.

Based on the search criteria, the resulting works' titles and abstracts were then reviewed by the authors to ensure relevance. Firstly, the articles were checked to ensure that the principal focus of the research was for maritime transportation. Secondly, machine learning has been utilised in the maritime domain for purposes other than risk analysis, such as route prediction, anomaly detection or computer vision. Unless these were specifically developed for accident prediction and evaluation, they were not included. The resulting dataset contained 76 papers.

2.2. Literature summary

In order to review the 76 retained articles, bibliometric analysis is conducted to categorise and evaluate aspects of each work. Numerous methods are available for conducting bibliometric analysis, such as trend analysis, clustering, and network analysis (Li, Goerlandt, & Reniers, 2021). Firstly, the yearly distribution of publications is used to determine the temporal trend of this topic. Figure 1 shows the yearly distribution of the 76 papers. There has been a marked increase in the number of published articles between 2014 and 2021, with 80% of the total within the last eight years and more than a third published between January 2020 and April 2021. The earliest example is Hashemi, Blanc, Rucks, and Shearry (1995) who utilised a simple neural network architecture to differentiate different accident types based on waterway conditions, exceeding the performance of logistic regression. It is notable that this significant increase in publications recently is similar to the trends for the wider maritime risk analysis discipline identified by Kulkarni et al. (2020), Lim et al. (2018) and Yang et al. (2019).

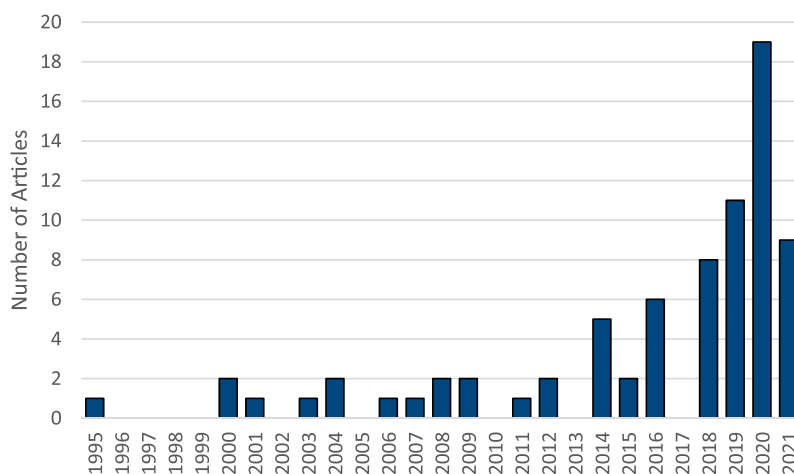


Figure 1. Article distribution by publication date (2021 to April).

Table 1. Top 10 journals.

N	Journal name	ISSN	No. of publications
1	<i>Ocean Engineering</i>	0029-8018	8
2	<i>Reliability Engineering and System Safety</i>	0951-8320	6
3	<i>Marine Policy</i>	0308-597X	4
4	<i>Accident Analysis and Prevention</i>	0001-4575	3
5	<i>Journal of Marine Science and Engineering</i>	2077-1312	3
6	<i>Maritime Policy and Management</i>	0308-8839	3
7	<i>Transportation Research Part A</i>	0965-8564	3
8	<i>Expert Systems with Applications</i>	0957-4174	2
9	<i>Journal of Risk and Reliability</i>	1748-006X	2
10	<i>Risk Analysis</i>	0272-4332	2

Secondly, the publications by journal are used to identify the key communities undertaking this work. Table 1 identifies the top 10 journals within which this work is published, principally within Ocean Engineering, Reliability Engineering and System Safety, and Marine Policy. A total of 47 different journals or conference proceedings published work on this topic, albeit only 13 accounted for more than one publication.

2.3. Review criteria

To provide a more detailed understanding of the research questions and methodological approaches in the literature, thematic analysis of the purpose, datasets and methods of each work is presented in the following sections. Firstly, the works are categorised by their purpose or application, in order to better understand the scope of machine learning in maritime risk assessment. Several papers addressed multiple applications, such as accident factors and probability analysis, and therefore, were dual categorised. A list of four principal applications were identified, namely:

- (1) **Analysis of accident factors** – understanding the causal factors associated with accident occurrence.
- (2) **Analysis of accident probability/likelihood** – predicting the likelihood or frequency of accident occurrence.
- (3) **Analysis of accident severity** – predicting the consequence or outcomes of accident occurrence.
- (4) **Analysis of accident description** – interpreting textual data to better understand accident types or causal factors.

Secondly, the capabilities of machine learning algorithms are to some extent dependent upon the availability of relevant datasets that describe the likelihood or consequences of accidents. Eight key data sources were identified:

- (1) **Accident Databases** – use of tabular data consisting of incident details such as type, time, location, vessels and causes.
- (2) **Vessel Movement Data** – positions of vessels from various sources such as AIS, Radar or artificially generated data.
- (3) **Inspection Data** – details of the outcomes of Port or Flag State Control inspections that include details of deficiencies or detentions of that vessel.

- (4) **List of Vessels/Designs** – details of a portion of the world shipping fleet, such as ship characteristics or designs.
- (5) **Expert Judgement** – the use of information obtained from experienced navigators through questionnaires or interviews.
- (6) **Textual Data** – the use of unstructured textual data, such as NAVTEX navigational and meteorological data or accident descriptions.
- (7) **Simulator Data** – the use of outputs from a full bridge simulator used to test specific scenarios.
- (8) **In-situ measurements** – data obtained from in-situ ship-board sensors, such as accelerometers or pressure sensors.

Thirdly, from the available datasets identified above, analysis is undertaken on which features have been included as independent or dependent variables in the various studies. 32 different features were identified that are grouped into five categories:

- (1) **Vessel details** – including vessel size, age and type.
- (2) **Vessel management** – including flag state, classification society, nationality, company and builder.
- (3) **Voyage details** – including location, date and time of day, depth of water, distance from shore, traffic level, vessel behaviour and speed.
- (4) **MetOcean Conditions** – including season, wind, wave, ice, precipitation, visibility, pressure, temperature and current.
- (5) **Incident History** – including inspection history, deficiencies, detentions, insurance claims, accident consequences and accident details.

Finally, articles are classified by their implemented machine learning algorithms and evaluated across methodological considerations. This includes the types of algorithms used and how the results are evaluated to assess the relative merits of different approaches.

3. Machine learning approaches to maritime risk assessment

Four principal applications of supervised machine learning for maritime risk assessment are shown in [Figure 2](#). The most frequent research question investigated by the reviewed studies concerns predicting the likelihood of accident occurrence, accounting for almost two-thirds of articles. For example, Jin, Shi, Yuen, et al. (2019) demonstrate the capabilities of several algorithms in order to classify which of a list of 1934 tankers, 633 accidents have occurred, based on vessel descriptive features. Others have used the outcomes of inspections to predict either ship detentions (Xiao, Wang, Lin, Qi, & Li, 2020) or accident occurrences (Knapp & Franses, 2008; Heij & Knapp, 2018; Fan, Zhang, Yin, & Wang, 2019; Knapp & Heij, 2020). In Heij and Knapp (2018), the authors found that those vessels with relatively high number of deficiencies had a risk of total loss that was more than six times those with few deficiencies. Such an approach can be expanded by including details of a vessel's voyage, to predict the likelihood of an accident based on the conditions a vessel has experienced (Adland, Jia, Lode, & Skontorp, 2021) or between different waterways

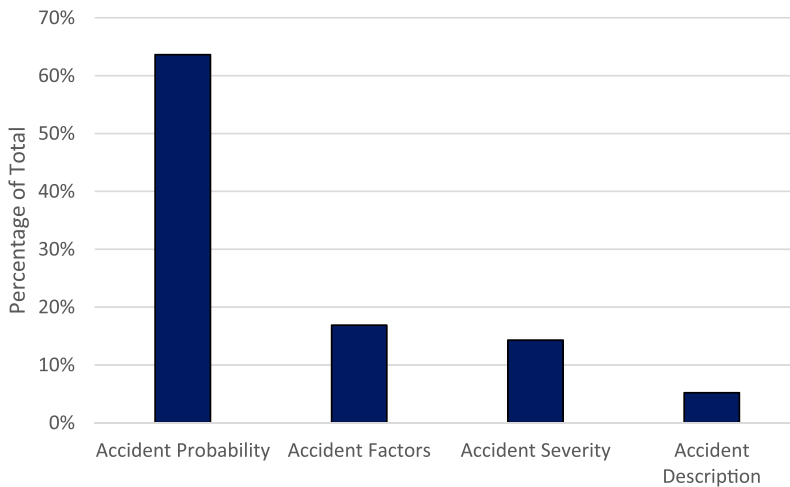


Figure 2. Proportion of articles by purpose/application.

(Hoorn & Knapp, 2015). In some cases, the risk of accidents is defined by expert labelling based on their experience (Zhang, Feng, Goerlandt, & Liu, 2020). Similarly, modelling near misses building upon the concept of a ship domain can provide an alternative dataset (Pietrzykowski, 2001; Zheng, Chen, Jiang, & Qiao, 2020; Rawson & Brito, 2021a, 2021b). Each of these applications has the potential to provide decision makers and coastguard authorities with actionable intelligence to identify and reduce accident occurrence.

With regards to accident factors analysis, these studies sought to determine the relative importance of different descriptive factors in accident occurrence (Bye & Aalberg, 2018; Fan, Yang, Blanco-Davis, Zhang, & Yan, 2020; Qiao, Liu, Ma, & Liu, 2020). For example, Fan et al. (2020) developed a Naïve Bayesian Network to identify the relationship between accident type and risk factors based on accident investigation reports. Bye and Aalberg (2018) seek to understand the causal factors behind different accident types in Norwegian waters. This can be used to inform policy decisions on the application of risk control measures.

Alternatively, some studies consider the consequences of hazards, whereby the task is to predict the severity of an accident, given an accident occurrence and the presence of different factors (Chang & Park, 2019; Tang et al., 2019; Wang, Wang, Jiang, Wang, & Yang, 2020; Wang, Liu, Wang, Graham, & Wang, 2021). For example, Rezaee, Pelot, and Finnis (2016) utilise a dataset of fishing vessel accidents and develop a model to predict damage and fatalities in accidents based on numerous input factors, such as weather conditions, time of incident and vessel type. The severity might also be dependent upon the outcomes of search and rescue efforts, which can be predicted in a similar way (Norrington, Quigley, Russell, & Meer, 2008).

Four of the papers used machine learning for accident description and identification through text classification and natural language processing. For example, this might involve determining accident types in unstructured textual datasets, such as navigational warnings or news articles (Razavi, Inkpen, Falcon, & Abielmona, 2014). Recent work by Dorsey, Wang, Grabowski, Merrick, and Harrauld (2020) sought

to identify and correct errors in fields of a large accident database, through use of a Naïve Bayes classifier.

4. Availability of datasets and choice of features

In order to construct a supervised machine learning algorithm, two principal types of training data are required: namely, a positive accident class and a negative non-accident class. The success of any derived algorithm is dependent upon the availability and representativeness of these datasets (Guikema, 2020). Availability can be defined as whether there is sufficient accurately labelled training data for algorithms to learn the relationship between the two classes. By representativeness, it is meant that future or unseen data is similar to data on which the models have been trained. The reviewed studies include a plethora of different datasets and features to construct their models and challenges associated with the availability and representativeness of data are identified.

4.1. Deriving class labels

In order to develop the class labels required for supervised machine learning, a variety of datasets have been identified and processed. Figure 3 shows the proportion of studies which utilise specific datasets, with historical accident data the dominant source of a positive class label. Accident data in isolation cannot assess risk (Bye & Almklov, 2019), but by combining multiple datasets a far greater understanding can be achieved (Kulkarni et al., 2020). Some authors have used a list of candidate vessels details and designs (Li, Yin, Bang, Yang, & Wang, 2012; Jin, Shi, Yuen, et al., 2019; Jin, Shi, Lin, & Li, 2019), port state control inspections (Li, Yin, & Fan, 2014) or the use of higher resolution vessel traffic data such as AIS (Zhang et al., 2020). Other spatial and temporal datasets can be combined within accident data, including the use of AIS data (Adland et al., 2021), shipboard weather reporting data (Knapp, Kumar, Sakurada, & Shen, 2011) or fishing vessel reporting data (Wu, Pelot, & Hilliard, 2009; Rezaee, Pelot, & Ghasemi, 2016). Each of these datasets are described and evaluated in turn.

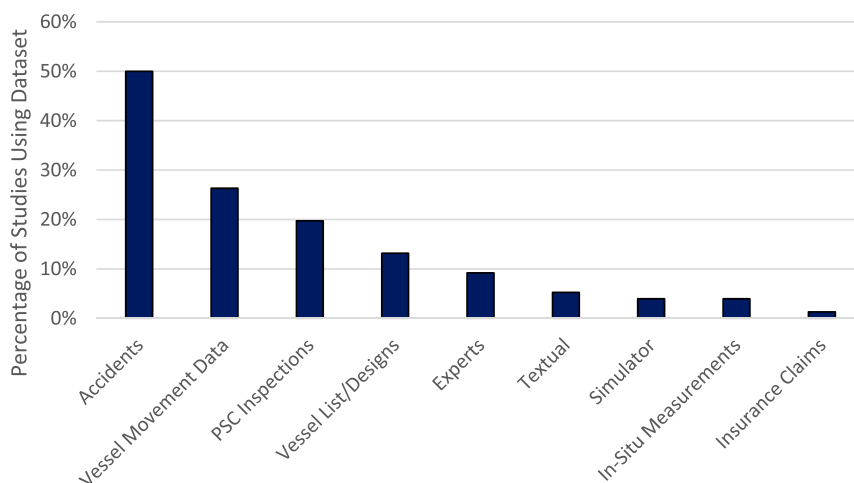


Figure 3. Principal datasets used in studies.

4.1.1. Accident data

Accident data is the most common input into these studies, accounting for half of the total. Several sources of accident data are available for researchers, including from national administrations such as the Chinese Maritime Safety Administration (Likun & Zaili, 2018), Hong Kong Marine Department (Wang, Wang, et al., 2020), Canadian Coastguard (Rezaee, Pelot, & Finnis, 2016; Fan et al., 2020), US Coastguard (Chang & Park, 2019), the UK's Marine Accident Investigation Branch (Fan et al., 2020) or a combination (Wang, Liu, et al., 2021). Other international datasets are also available, such as the IMO's GISIS (Jin, Shi, Yuen, et al., 2019; Fan, Wang, & Yin, 2019; Wang, Huang, Shi, & Zhang, 2021) or commercial datasets such as IHS-Maritime or Lloyds Maritime Intelligence Unit (Hoorn & Knapp, 2015). Some studies have used non-navigational accidents, such as modelling piracy attacks (Jin, Shi, Lin, et al., 2019).

Whilst accident data is a definitive metric of risk, the use of accident data in maritime risk analysis has been subject to criticism in several works. Firstly, incident data is often unreliable or under-reported which undermines statistical analysis (Hassel, Asbjornslett, & Hole, 2011), particularly minor accidents. As a result, trained models will have an inherent bias towards predicting that accidents are less frequent and more severe than would likely be the case. In both Fan, Wang, et al. (2019) and Jin, Shi, Yuen, et al. (2019), the IMO's GISIS accident database is utilised which is known to have underreporting and quality issues (Zhang, Sun, Chen, & Cheng, 2021). Similarly, accident databases have quality issues, such as missing fields (Zhang et al., 2021) or incorrect positioning (Mazaheri, Montewka, Kotilainen, Sormunen, & Kujala, 2014). To prevent undermining the predictions of developed models, these limitations need to be addressed but little evidence is reported in the reviewed works as to how this has been achieved. An innovative solution of quality assurance is proposed by Dorsey et al. (2020) through using natural language processing of the free text fields, but not all databases have this information available.

A second key limitation of accident data is its inherent infrequency, that both constrains the training capability of many algorithms but undermines the representativeness of the training data. For example, in some novel environments, such as the Arctic, or where shipping routes have changed over time, the learnt function between the historical training data and future predictions is invalidated. Given such conditions, some have argued that expert developed Bayesian Networks are more appropriate than using historical accident data (Baksh, Abbassi, Garaniya, & Khan, 2018). Similarly, there is a general bias in the literature to the Baltic Sea (Kulkarni et al., 2020) and as a result it is difficult to state whether models trained in this environment are representative of conditions elsewhere (Zhang et al., 2020).

Furthermore, due to the infrequency of accident occurrence, collecting a sufficient sample might require multiple years of accident data to be collated. During that period, the conditions, technologies and behaviours of vessels which led to these accidents changes, and therefore, the representativeness of the training data to future predictions becomes less strong (Guikema, 2020). For example, new technologies such as AIS, changes in Bridge Resource Management and training, or implementation of new risk control measures such as pilotage or traffic lanes would change the risk profile of that waterway. Care must be taken by authors to

gauge the degree to which such effects might influence the predictive capabilities of their models.

4.1.2. Ship inspection data

The outcomes of ship inspections, particularly information on ship detentions or deficiencies, could arguably indicate which vessels are most at risk (Kretschmann, 2020). Port state control (PSC) inspections are used by national administrations to evaluate the safety of a vessel, identify any deficiencies that warrant correction and if necessary, detain the vessel until they are rectified. To maximise transparency and improve the collective safety of navigation, inspection data is shared between administrations through various Memorandums of Understanding (MoUs). For this reason, data on hundreds of thousands of inspections are widely accessible online and have been used by many researchers, such as use of the Tokyo MoU (Fan, Wang, et al., 2019; Wang, Wang, et al., 2020; Xiao et al., 2020), Paris MoU (Yang, Yang, & Yin, 2018), Asia-Pacific MoU (Wu et al., 2021) or a combination (Knapp & Franses, 2008). Accident data and inspection data are often combined, to understand whether models can be developed to target inspections to those vessels most likely to have an accident (Knapp & Franses, 2008).

However, ship inspections are likely to predict only a subset of potential hazards as they provide a single snapshot when a vessel is in port. Navigational accidents such as collisions and groundings have spatial temporal causal factors that cannot be assessed from inspection data alone. Others have noted that inspection data might not be consistent between MoU's, and therefore, the models may have little generalisation between locations and datasets (Xiao et al., 2020).

4.1.3. Qualitative data

Expert labelling can provide an alternative method of assessing risk. Experts are provided with different situations and tasked with labelling the relative risk of an incident, becoming the target label for model development (Ozturk, Birbil, & Kadir, 2019; Zhang et al., 2020). Pietrzykowski, Wielgosz, and Breitsprecher (2020) use simulated close quarters situations, labelled by experts in terms of risk-taking behaviour, before training a decision tree to predict these labels. Ozturk et al. (2019) study the risk of ship manoeuvring by combining expert judgement with semi-supervised SVMs and Random Forest to determine relative risk levels. Whilst still assessing collision risk, Zhang et al. (2020) develop an innovative approach by utilising a Convolutional Neural Network (CNN) and a traffic image to classify ship-ship collision risk in encounter scenarios. Such views might be elicited through the use of questionnaires (Karakasnaki, Vlachopoulos, Pantouvakis, & Bouranta, 2018). However, such an approach is unlikely to provide calibrated, quantitative metrics of risk, instead providing a qualitative, high, medium, low scaling. Manual labelling is also time expensive, and therefore, it may not be possible to gather sufficient labelled data from maritime professionals. Furthermore, expert judgement is subjective, and may be subject to heuristics and biases that might affect the predictive accuracy (Kahneman, Slovic, & Tversky, 1982).

4.1.4. Vessel traffic data

The former datasets are limited as they do not contain information on the activities or behaviour of a vessel, specifically the external factors and influences on the navigation of the vessel (Xiao et al., 2020). For example, one vessel might spend twice as long at

sea as another, or operate in particularly challenging environments, influencing the risk profile (Bye & Almklov, 2019). Key vessel traffic datasets are available to researchers such as the Automatic Identification System (AIS), ship movement data from ports and Radar data amongst others. Adland et al. (2021) were able to demonstrate that adding AIS derived features increased their accident prediction model performance by 18%.

A number of studies apply machine learning to AIS data to predict the risk of vessel transits and particularly collision situations. Gang, Wang, Sun, Zhou, and Zhang (2016) propose the use of SVMs to predict collision risk in meeting vessel situations. Liu, Wang, Cai, Liu, and Liu (2020) propose the use of Density-Based clustering and a Recurrent Neural Network (RNN) to determine high risk meeting situations in AIS data. AIS data is also used to predict vessel speed and heading in a port approach channel (Tsou, 2019). High resolution quantitative data can also be sourced from in-situ measurements from a vessel, such as ice load on a ship's hull or movement in adverse weather conditions (Liu, Duan, Huang, Duan, & Ma, 2020).

Furthermore, AIS data can be used to model non-accident events such as near misses might reasonable be used in lieu of historical accidents (Du, Goerlandt, & Kujala, 2020). These might be developed through, for example, ship domain analysis (Rawson & Brito, 2021a) and would greatly increase the number of positive class samples. However, several studies have challenged both the validity of such methods (Goerlandt & Kujala, 2014) and the predictive relationship between ship encounters and collisions (Rawson & Brito, 2021b).

Adding dynamic movement data such as AIS has computational and representation challenges, particularly given class imbalance. Given the low frequency of accident occurrence, several years of data would be required to obtain enough positive samples, necessitating massive AIS datasets which are costly to obtain, challenging to handle and time consuming to train. For example, Adland et al. (2021) analyse 42,000 Pacific ship transits consisting of 600 million ship positions which required processing and combination with heterogeneous datasets such as weather for use in model development. With this volume and variety of datasets, model development becomes unmanageable, and any downsampling or aggregation could result in potentially valuable information being lost.

Even though AIS offers several advantages in providing far greater resolution and details of vessel movements, significant limitations have been highlighted by many authors (IALA, 2002; Harati-Mokhtari, Brooks, Wall, & Wang, 2007; Iphar, Ray, & Napoli, 2020; Yang et al., 2019). These include limitations of reception range and coverage between vessels and terrestrial receiving stations or message collision in busy areas such as the North Sea. Furthermore, transmitted data can be false, with one study finding that 8% of AIS transmissions contained some form of error, including vessel identity, length, position, type or draught (Harati-Mokhtari et al., 2007). Therefore, in order to properly utilise these datasets, without introducing errors, some form of pre-processing and error checking is warranted. Without this, the input training data for a learning algorithm may result in errors propagated through the training and testing set, leading to sub-optimal or misleading results (Zhang et al., 2020).

4.2. Feature analysis

A principal advantage of the use of machine learning methods is the ability to integrate numerous datasets, thereby improving predictive accuracy. Based on the review of each

article, key features used as inputs are shown in [Figure 4](#) as a proportion of each study. Given their frequent use as the class label ([Figure 3](#)), accident details are the most commonly utilised feature in the reviewed studies. These include the types of vessels involved, location, time and causes of each individual accident.

The second most common group relate to the details of a vessel, such as their type, size, tonnage, age and flag state. It would be anticipated that certain vessel types are more prone to accidents than others. For example, some research determined that larger vessels are more prone to incidents (Li, Yin, et al., 2012; Li et al., 2014; Fan, Zhang, et al., 2019; Fan, Wang, et al., 2019) and others that they are relatively safer (Bye & Aalberg, 2018; Jin, Shi, Yuen, et al., 2019; Jin, Shi, Lin, et al., 2019). To some extent this contradiction might reflect data collection issues with incidents involving larger vessels likely more readily recorded, or different routes taken by different vessel sizes. Similarly, it has been concluded that older vessels are less safe than newer ones (Bye & Aalberg, 2018; Jin, Shi, Yuen, et al., 2019) and vice versa (Li, Yin, et al., 2012, 2014). Contrary arguments to explain this are presented, such as older vessels being proven more seaworthy or robust due to their continued use (Li et al., 2014) whilst newer vessels have greater technological safety systems (Fan, Zhang, et al., 2019).

Similarly, the use of an IACS classification society or a reputable flag state should indicate greater ship safety (Jin, Shi, Lin, et al., 2019; Li, Yin, et al., 2012, 2014), but these results are not supported by Fan, Wang, et al. (2019), suggesting either some bias in their data or that safety management might be more relaxed on perceived safer vessels. Details on inspections, deficiencies and detentions are used for both dependent and independent variables throughout these works. It is logical that vessels with safety deficiencies identified during inspections are those most likely to have safety related incidents (Knapp & Franses, 2008; Heij & Knapp, 2018; Fan, Zhang, et al., 2019). Yet, Fan, Wang, et al. (2019) found contradictory results, suggesting that identified defects during inspections result in a greater effort by the operators to repair and improve the vessel, enhancing safety. These contradictions are often justified by the authors based on their own interpretation as to why accidents occur.

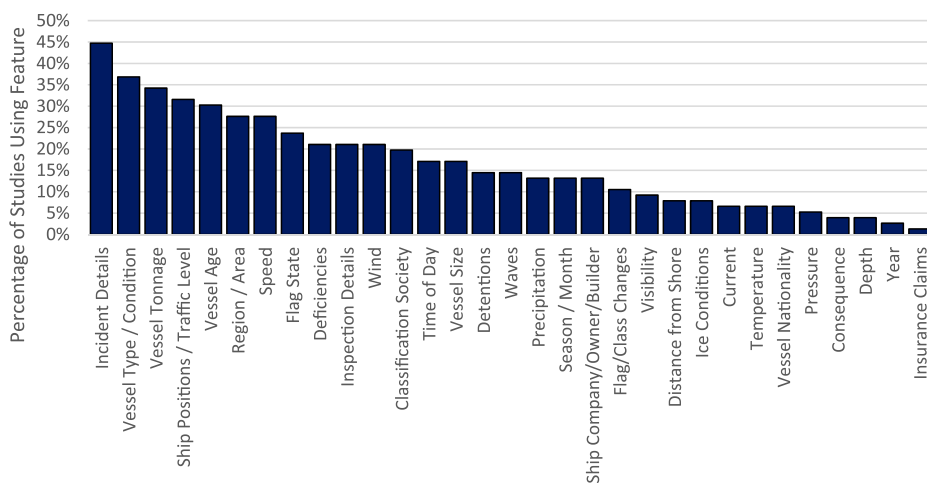


Figure 4. Key features used in studies.

Other factors such as weather are more difficult to integrate as they require spatial temporal joining between different datasets, but this has been achieved in several works to demonstrate the influence of wind and waves on incident probability or severity (Rezaee, Pelot, & Finnis, 2016; Rezaee, Pelot, & Ghasemi, 2016; Knapp et al., 2011; Adland et al., 2021). Some conditions such as ice are rarely encountered and can only be included when the analysis is limited to specific regions where the effect on risk is significant (Wu et al., 2009; Rezaee, Pelot, & Finnis, 2016; Rezaee, Pelot, & Ghasemi, 2016). In some cases, season is used as a proximate measure of weather conditions assuming that winter months have more extreme conditions than summer months (Blanc, Hashemi, & Rucks, 2001), but this relationship is not as significant (Li et al., 2014; Bye & Aalberg, 2018).

Many factors have been identified in both the academic and industry literature as to why accidents might occur (Mazaheri et al., 2014), yet the features identified in Figure 4 are only a small proportion due to challenges in quantitatively representing certain factors as features. Human error constitutes the most significant causal factor for maritime accidents (MCA, 2010), yet factors such as the alertness of the bridge team cannot be easily encoded, without direct observations of the bridge environment which is not feasible at the scale required. As a result, Bayesian Networks with probabilistic inputs from expert elicitation are widely used to model specific aspects of human error such as situational awareness, safety culture or bridge resource management (Mazaheri, Montewka, & Kujala, 2016). Other factors have been used as indicating the safety culture of a vessel. For example, whether a vessel is registered to a flag of convenience is believed to indicate that the safety regulations on board might be less (Jin, Shi, Lin, et al., 2019; Li, Yin, et al., 2012, 2014).

In addition to human factors, there are other spatial and temporal factors which logically should be included but are surprisingly underexplored. For example, predicting ship groundings would require inclusion of features related to the depths of water, but few studies have incorporated this. The presence of risk controls such as pilotage or shipping lanes would have a significant effect on the risk profile of any individual movements. Furthermore, visibility (fog) is a likely contributor to accident occurrence but few studies have been able to source such a dataset (Likun & Zaili, 2018). There are significant opportunities to expand the number of features to include spatiotemporal risk factors.

These challenges can lead to an asymmetry of information between accident datasets and non-accident datasets, which has been recognised in other disciplines (Mehdizadeh et al., 2020). High quality vessel traffic data might be available for small durations or areas but limits the number of accident samples. Furthermore, Lensu and Goerlandt (2019) characterise that most maritime research compromises either the extent of the data or the methodological complexity. Extending the accident data to larger areas and longer durations makes obtaining non-accident data more costly and challenging, and therefore, features such as speed and vessel behaviour may be lost (Hoorn & Knapp, 2015). These more detailed features that relate to navigational behaviour may be crucial in improving accident prediction, but few of the reviewed studies offered methodologies that have these capabilities. Recent focus in several works has been on integrating heterogeneous maritime datasets using big-data technologies, which might overcome these challenges (Lensu & Goerlandt, 2019; Kulkarni et al., 2020). Machine learning methods offer a scalable method to construct complex models using massive heterogeneous datasets which might further improve the predictive capability of accident models.

5. Methodological considerations

Table 2 lists the most frequent machine learning algorithms and methods utilised within the reviewed articles. Several articles compare or utilise multiple methods (such as Jin, Shi, Yuen, et al., 2019), and therefore, the total exceeds the total number of articles. A multi-methodological approach might reflect that there is no reason a priori why one algorithm would work better than another on a particular problem.

5.1. Model choice and development

The most widely applied is Logistic Regression, a type of generalised linear model that utilises a sigmoid function to make predictions that are representative of probability. Their widespread use reflects both their simplicity and effectiveness, and are often used to derive baseline models (Merghadi et al., 2020). Examples of this include assessing tanker accidents (Jin, Shi, Yuen, et al., 2019), fishing vessel accidents (Rezaee, Pelot, & Finnis, 2016; Rezaee, Pelot, & Ghasemi, 2016), accidents based on previous detentions (Heij & Knapp, 2018) or weather-related incidents (Knapp et al., 2011). One of the key advantages of Logistic Regression is the ability to easily evaluate the impact of each individual variable on the dependent variable and produce a probabilistic likelihood of an accident occurring, improving interpretability. For example, Jin, Shi, Yuen, et al. (2019) note that Logistic Regression had the lowest accuracy of the models tested, but they dedicated significant discussion to the coefficients associated with different exploratory factors. Computationally it requires little tuning compared to some of the following models, but as a result is unable to solve complex non-linear relationships as the decision surface is linear.

Artificial Neural Networks (ANNs) are biologically inspired computational networks, originally proposed by McCulloch and Pitts (1943), but popularised following the development of backpropagation as an optimised means of model training. ANNs consist of input layers (independent variables), hidden layers and an output layer (dependent variable) and are trained through gradient descent as a means to minimise the cost function. Part of the strength of this method is its ability to learn non-linear and complex relationships, which arguably are inherent in risk related studies (Hedge & Rokseth, 2020), achieving high performance in the reviewed studies (Lisowski, 2016; Wang et al., 2004; Wang, Wang, et al., 2020). CNNs are a special form of ANNs used for complex classification problems, particularly involving images. One study reviewed sought to utilise this approach to assess collision risk from traffic pictures, previously assigned a risk score by a group of

Table 2. Top 10 methods/algorithms.

N	Method	No. of studies	Percentage
1	Logistic Regression	29	28%
2	Artificial Neural Network (ANN)	13	13%
3	Support Vector Machine (SVM)	11	11%
4	Random Forest	10	10%
5	Bayesian Networks	8	8%
6	Naïve Bayes	8	8%
7	Decision Tree	7	7%
8	XGBoost	3	3%
9	LSTM RNN	2	2%
10	Convolutional Neural Network (CNN)	1	1%

experts (Zhang et al., 2020). In addition, Recurrent Neural Networks (RNNs) expand ANNs with loops such that previous outputs can be used as inputs, making them especially useful in timeseries data and the fields of natural language processing. Long Short-Term Memory networks (LSTMs) are a popular form of RNNs that address the long-term dependency problems of RNNs, enabling better retention of information over long periods through the use gates. Within the reviewed studies, LSTMs are implemented by Ma, Li, Jia, Zhang, and Zhang (2020) to predict future collision risk based on the paths of vessels or to predict trends in accident numbers (Chai, Xue, Sun, & Weng, 2020). A challenge of neural networks is their complexity which can lead to long development times, overfitting, poor generalisations and low interpretability. Furthermore, they often require significant datasets to train which might not be possible with rare events such as maritime accidents.

A support vector machine (SVM) can perform linear and non-linear classification by constructing a hyperplane or set of hyperplanes in high dimensional space to maximise the margin between training examples and have proved popular due to their classification capability (Kecman, 2005). Ozturk et al. (2019) utilise SVM for semi-supervised labelling of a simulator dataset using expert labels. Several authors have seen potential of SVMs for use in collision risk assessment (Gang et al., 2016; Zheng et al., 2020). However, SVMs have no native capability to identify the relative contribution of different features to accident prediction which might limit their application in risk studies.

Decision trees are a non-parametric supervised learning method that seeks to predict the value of a target variable by learning simple decision rules inferred from data features. For example, Wu et al. (2009) use a regression tree to predict the relative incident rate of fishing vessels based on various weather-related factors. Such methods are relatively simple, easy to understand and are highly transparent, which allow for the interrogation of why and how predictions are made. Furthermore, tree-based algorithms are able to calculate the relative importance of each feature in producing the predictions. However, decision trees rarely achieve high accuracies, partly due to methodological weaknesses when using a mixture of continuous and categorical variables.

Decision trees are prone to overfitting, but an ensemble of decision trees such as Random Forests or XGBoost can result in more accurate predictions. Random Forests (Breiman, 2001) introduce bagging and random subspace for resampling and feature selection, to produce a diversity of trees, with the final prediction the majority average. Random forests have widespread popularity and availability in numerous languages as well as inherent properties such as training speed and robustness when using high-dimensional and unbalanced datasets (Breiman, 2001). In Jin, Shi, Yuen, et al. (2019) random forests achieved the second highest accuracy score (96%) to predict ship accidents. XGBoost (Chen & Guestrin, 2016) further expands tree-based learning through boosting and gradient descent, such that new models are iteratively trained on the residual errors of previous models. As a result, this method achieves some of the highest predictive accuracies in comparative studies (Jin, Shi, Yuen, et al., 2019; Adland et al., 2021). XGBoost is argued to be more accurate, more scalable to large datasets and highly efficient, and is predicted to become one of the most widely applied methods in data science (Wang, Deng, & Wang, 2020).

Naïve Bayes is an implementation of Bayesian logic in that it contains a naïve assumption of conditional independence between every pair of features given the value of the class variable. A key advantage of Naïve Bayes is its speed, without the need for expensive hyper-parameter tuning and often works well with little training data. However, in the context of maritime risk assessment with large datasets of vessel movements, their application is limited. Dorsey et al. (2020) utilise Naïve Bayes to classify incidents based on their descriptions to aid dataset verification. Bayesian Networks are a technique for graphically representing a joint probability distribution of a selected set of variables (Yang et al., 2018). Each node in the network has a finite number of states and a probability of occurrence that is dependent on the states of its parent nodes. Whilst the structures and probabilities can be defined qualitatively, conditional probability tables can be learnt quantitatively through various implementations, such as Bayesian Search or Greedy Thick Thinning (Fan, Zhang, et al., 2019; Fan, Wang, et al., 2019). The use Bayesian Networks to model expert elicited risk values is arguably more transparent than the use of complex machine learning methods such as neural networks (Zhang et al., 2020).

5.2. Class imbalance and results evaluation

Maritime accidents are rare events which pose significant challenges to both model development and results evaluation. For example, in Adland et al. (2021), a dataset of 41,717 transits across the Pacific Ocean resulted in only 86 insurance claims, a rate of 0.21%. Even where expert labelling is utilised, it would be expected that the number of high-risk compared to low-risk samples would remain imbalanced (Pietrzykowski et al., 2020).

Firstly, many machine learning algorithms are developed to assume a relatively well-balanced distribution of data, and therefore, class imbalance may challenge their effectiveness (Leevy, Khoshgoftaar, Bauder, & Seliya, 2018). Two principal methods to address this are at a data-level by balancing the training data or through the use of specific algorithms that weight the classes appropriately during training (Leevy et al., 2018). At a data-level, the negative class might be downsampled to match the frequency in the positive class, or the positive class upsampled, artificially generating new accidents would reduce the disparity between accident and non-accident data. One method to achieve this is SMOTE, Synthetic Minority Oversampling Technique (Chawla, Bowyer, Hall, & Keglemeyer, 2002), which generates artificial minority classes based on the distribution of actual minority classes. In He, Hao, and Wang (2021), the use of SMOTE increased the overall accuracy from 97.8% to 99.3%, as a result of reducing the number of false negatives.

Secondly, the nature of class imbalance also introduces challenges with evaluating the outputs of machine learning models. In the case of Adland et al. (2021), a model which assumed accidents could not occur would achieve an accuracy of 99.79%. Given this imbalance, it is common to assess machine learning classification results with alternative metrics such as a confusion matrix, F-scores or the Receiver Operating Characteristic (ROC) curve. Firstly, a confusion matrix compares the actual labels to the predicted labels, and can be used to assess recall, the ratio of True Positives to False Negatives and True Positives, and precision, the ratio of True Positives to True Positives and False Positives. In such cases, where rare events need to be predicted, recall assumes a primary role in model utility (Paltrinieri, Comfort, & Reniers, 2019). Secondly, F-scores, are a useful means of

combining precision and recall, to provide an overall measure of model accuracy. Thirdly, the ROC curve plots the true positive rate (recall) against the false positive rate. A trade-off exists between a higher recall and more false positives. Different classifiers can be compared by measuring the area under the curve (AUC), with a perfect classifier having an AUC score of 1 and a random classifier having a score of 0.5.

Thirdly, it can be difficult to gauge the degree to which the calculated risk numbers from these models are accurate compared to the underlying risk (Aven & Heide, 2009). Whilst holdout validation, with a testing and training dataset can be used to assess the quality of an algorithm, predictions made on new data are difficult to validate, particularly given the rarity at which maritime accidents occur. As a result, the degree to which machine learning methods might be more accurate, or at least more useful, than conventional methods is difficult to assess.

Finally, many machine learning algorithms can output class probabilities, which could be interpreted as the relative likelihood of accident occurrence, and therefore, as a risk score. However, due to the inherent class imbalance and use of resampling strategies, this probability is not calibrated and will generate higher probabilities for the majority class. Whilst there are methods to correct this (Pozzolo, Caelen, Johnson, & Bontempi, 2015), in many circumstances there may be insufficient data to calibrate the results. Therefore, a non-dimensional output may be desirable, such as a ship safety index (Li et al., 2014) or qualitative risk levels (Zhang et al., 2020). This allows some delineation between the risk of different scenarios but lacks quantitative metrics of probability or consequence.

6. Concluding remarks and future research opportunities

The use of machine learning in safety and risk engineering is increasing (Hedge & Rokseth, 2020), yet their application within the maritime domain has been limited. Therefore, this review demonstrates the state-of-the-art within this field, highlighting opportunities and challenges for further adoption. It is clear that through leveraging machine learning and big data, new insights into accident likelihood and consequence can be achieved, with predictive accuracies in excess of 90% (Jin, Shi, Yuen, et al., 2019; He et al., 2021). In particular, such methods enable complex models to be constructed that are scalable to massive datasets, overcoming a key compromise of conventional approaches (Lensu & Goerlandt, 2019). The use of machine learning methods with powerful learning capabilities can contribute to vessel safety by providing decision makers with reliable and accurate tools for real-time and strategic maritime risk assessment (Dorsey et al., 2020). However, several key issues have been identified and summarised below that warrant further research, many of which are not unique to maritime transportation and equally applicable across other transportation disciplines.

- The relationship between accidents and their casual factors is highly complex which makes prediction inherently challenging. Many have argued that machine learning methods are, therefore, more suited to these research questions. Firstly, it does not require any a priori assumptions on the underlying relationship between dependent and exploratory variables (Jin, Shi, Yuen, et al., 2019). Secondly, machine learning methods are well suited to dealing with complex, non-linear relationships and multi-

dimensional datasets, which are often inherent characteristics of maritime safety (Gang et al., 2016; Jin, Shi, Yuen, et al., 2019; Qiao et al., 2020). Whilst high accuracies have been achieved previously, the significant importance of human factors in accident events might limit the inherent capabilities of this approach. By contrast, predicting the fuel consumption of a vessel involves a well understood mechanical process with easily available datasets and well-balanced features, that can achieve accuracies in excess of 99.99% (Uyanik et al., 2020). It is perhaps unlikely that the same level of accuracy could ever be achieved in accident prediction.

- The review has demonstrated the breadth of different datasets and features which have been utilised to develop maritime risk models. Two key areas of further development are necessary to improve the predictive capabilities of such models. Firstly, the greater inclusion of spatial temporal datasets such as metocean conditions, bathymetry and the behaviours of the vessel. These pose significant processing challenges that have as yet curtailed the application of machine learning towards certain hazards such as groundings. Secondly, human factors are inherently difficult to quantify but account for a considerable number of accidents. Methods to incorporate such features require greater exploration.
- Many have noted that the use of machine learning methods can overcome potential bias from experts in qualitative risk assessments that might reduce accuracy (Fan, Zhang, et al., 2019; Coraddu, Oneto, Maya, & Kurt, 2020) or are expensive and time-consuming (Zhang et al., 2020). However, where there is limited quantitative accident data, machine learning methods can learn to replicate expert labelling achieving similar effectiveness (Ozturk et al., 2019; Zhang et al., 2020). Therefore, there is significant potential for model development towards automated decision support tools to monitor vessel safety in real-time through replicating the expertise of master mariners.
- Many have criticised machine learning models as “black-box” which reduces transparency and interpretability, undermining risk communication (Zhang et al., 2020; Guikema, 2020). Whilst these criticisms have been given to maritime risk assessments more generally (Psaraftis, 2012) the complexity in certain models makes understanding the risk contribution of different factors less clear (Wen et al., 2021). It is notable that whilst Jin, Shi, Yuen, et al. (2019) found that logistic regression was not the most effective tool, the authors dedicated significant discussion to its results given its greater interpretability. Recent research into model interpretability has led to new methods such as SHAP (He et al., 2021; Adland et al., 2021), and it is encouraged that future works explore these developments.
- Whilst proponents of such studies have argued that the use of machine learning exhibits several advantages, as yet there is little comparative evidence that they outperform conventional maritime risk models. This review has been limited to scoping the state-of-the-art in this topic, and further research is encouraged to compare and contrast conventional and machine learning methods for maritime risk analysis.

Given the increasing digitalisation of shipping and development towards autonomous shipping, it is likely that there will be greater interest of machine learning technologies within maritime transportation. As this value is increasingly recognised, more data collection and research will no doubt address some of the challenges highlighted within this review. Furthermore, new applications and opportunities will be developed which will

solidify this approach as a reliable and effective method for performing maritime risk analysis.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Horizon 2020 Framework Programme [grant number 723526]; Southampton Marine and Maritime Institute.

ORCID

Andrew Rawson  <http://orcid.org/0000-0002-8774-2415>

Mario Brito  <http://orcid.org/0000-0002-1779-4535>

References

- Adland, R., Jia, H., Lode, T., & Skontorp, J. (2021). The value of meteorological data in marine risk assessment. *Reliability Engineering and System Safety*, 209. doi:10.1016/j.res.2021.107480
- Allianz. (2012). Safety and shipping 1912–2012: From Titanic to Costa Concordia [online]. Accessed 18th May 2020. <https://www.agcs.allianz.com/content/dam/onemarketing/agcs/agcs/reports/AGCS-Safety-Shipping-Review-2012.pdf>
- Aven, T., & Heide, B. (2009). Reliability and validity of risk analysis. *Reliability Engineering and System Safety*, 94, 1862–1868. doi:10.1016/j.res.2009.06.003
- Baksh, A., Abbassi, R., Garaniya, V., & Khan, F. (2018). Marine transportation risk assessment using Bayesian Network: Application to Arctic waters. *Ocean Engineering*, 159, 422–436. doi:10.1016/j.oceaneng.2018.04.024
- Blanc, L., Hashemi, R., & Rucks, C. (2001). Pattern development for vessel accidents: A comparison of statistical and neural computing techniques. *Expert Systems with Applications*, 20, 163–171. doi:10.1016/S0957-4174(00)00056-7
- Bowen, H. (2020). The shipping losses of the British East India Company, 1750–1813. *International Journal of Maritime History*, 32(2), 323–336. doi:10.1177/0843871420920963
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. doi:10.1023/A:1010933404324
- Bye, R., & Aalberg, A. (2018). Maritime navigation accidents and risk indicators: An exploratory statistical analysis using AIS data and accident reports. *Reliability Engineering and System Safety*, 176, 174–186. doi:10.1016/j.res.2018.03.033
- Bye, R., & Almklov, P. (2019). Normalization of maritime accident data using AIS. *Marine Policy*, 109. doi:10.1016/j.marpol.2019.103675
- Chai, T., Xue, H., Sun, K., & Weng, J. (2020). Ship accident prediction based on improved quantum-behaved PSO-LSSVM. *Mathematical Problems in Engineering*, 2020. doi:10.1155/2020/8823322
- Chang, Y., & Park, H. (2019). The impact of vessel speed reduction on port accidents. *Accident Analysis and Prevention*, 123, 422–432. doi:10.1016/j.aap.2016.03.003
- Chawla, N., Bowyer, K., Hall, L., & Keglemeyer, W. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. doi:10.1613/jair.953
- Chen, C., Chen, X., Ma, F., Zeng, X., & Wang, J. (2019). A knowledge-free path planning approach for smart ships based on reinforcement learning. *Ocean Engineering*, 189. doi:10.1016/j.oceaneng.2019.106299
- Chen, P., Huang, Y., Mou, J., & van Gelder, P. (2019). Probabilistic risk analysis for ship-ship collision: State of the art. *Safety Science*, 117, 108–122. doi:10.1016/j.ssci.2019.04.014

- Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. San Francisco: KDD. doi:10.1145/2939672.2939785
- Chen, Z., Xue, J., Wu, C., Qin, L., Liu, L., & Cheng, X. (2018). Classification of vessel motion pattern in inland waterways based on automatic identification system. *Ocean Engineering*, 161, 69–76. doi:10.1016/j.oceaneng.2018.04.072
- Coraddu, A., Oneto, L., Maya, B., & Kurt, R. (2020). Determining the most influential human factors in maritime accidents: A data-driven approach. *Ocean Engineering*, 211. doi:10.1016/j.oceaneng.2020.107588
- Dorsey, C., Wang, B., Grabowski, M., Merrick, J., & Harrald, J. (2020). Self-healing databases for predictive risk analytics in safety-critical systems. *Journal of Loss Prevention in the Process Industries*, 63. doi:10.1016/j.jlp.2019.104014
- Du, L., Goerlandt, F., & Kujala, P. (2020). Review and analysis of methods for assessing maritime waterway risk based on non-accident critical events detected from AIS data. *Reliability Engineering and System Safety*, 200. doi:10.1016/j.res.2020.106933
- EMSA. (2018). *Joint workshop on risk assessment and response planning in Europe*. London.
- EMSA. (2019). *Annual overview of marine casualties and incidents 2019 [online]*. Accessed 18th May 2020. <http://www.emsa.europa.eu/emsa-homepage/2-news-a-press-centre/news/3734-annual-overview-of-marine-casualties-and-incidents-2019.html>
- Fan, L., Wang, M., & Yin, J. (2019). The impacts of risk level based on PSC inspection deficiencies on ship accident consequences. *Research in Transportation Business and Management*, 33. doi:10.1016/j.rtbm.2020.100464
- Fan, L., Zhang, Z., Yin, J., & Wang, X. (2019). The efficiency improvements of port state control based on ship accident Bayesian networks. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 233(1), 71–83. doi:10.1177/1748006X18811199
- Fan, S., Yang, Z., Blanco-Davis, E., Zhang, J., & Yan, X. (2020). Analysis of maritime transport accidents using Bayesian Networks. *Proceedings of the Institution of Mechanical Engineers Part O: Journal of Risk and Reliability*, 234(3), 439–454. doi:10.1177/1748006X19900850
- Gang, L., Wang, Y., Sun, Y., Zhou, L., & Zhang, M. (2016). Estimation of vessel collision risk index based on support vector machine. *Advances in Mechanical Engineering*, 8(11), 1–10. doi:10.1177/1687814016671250
- Goerlandt, F., & Kujala, P. (2014). On the reliability and validity of ship-ship collision risk analysis in light of different perspectives on risk. *Safety Science*, 62, 348–365. doi:10.1016/j.ssci.2013.09.010
- Goerlandt, F., & Montewka, J. (2015). Maritime transportation risk analysis: Review and analysis in light of some foundational issues. *Reliability Engineering & System Safety*, 138, 115–134. doi:10.1016/j.res.2015.01.025
- Guikema, S. (2020). Artificial intelligence for natural hazards risk analysis: Potential, challenges, and research needs. *Risk Analysis*, 40, 1117–1123. doi:10.1111/risa.13476
- Harati-Mokhtari, A., Brooks, P., Wall, A., & Wang, J. (2007). Automatic Identification System (AIS): Data reliability and human error implications. *Journal of Navigation*, 60, 373–389. doi:10.1017/S0373463307004298
- Hashemi, R., Blanc, L., Rucks, C., & Shearry, A. (1995). A neural network for transportation safety modelling. *Expert Systems with Applications*, 9(3), 247–256.
- Hassel, M., Asbjornslett, B., & Hole, L. (2011). Underreporting of maritime accidents to vessel accident databases. *Accident Analysis and Prevention*, 43, 2053–2063. doi:10.1016/j.aap.2011.05.027
- He, J., Hao, Y., & Wang, X. (2021). An interpretable aid decision-making model for flag state control ship detention based on SMOTE and XGBoost. *Journal of Marine Science and Technology*, 9, 2. doi:10.3390/jmse9020156
- Hedge, J., & Rokseth, B. (2020). Applications of machine learning methods for engineering risk assessment – A review. *Safety Science*, 122. doi:10.1016/j.ssci.2019.09.015
- Heij, C., & Knapp, S. (2018). Predictive power of inspection outcomes for future shipping accidents – An empirical appraisal with special attention for human factor aspects. *Maritime Policy and Management*, 45(5), 604–621. doi:10.1080/03088839.2018.1440441
- Hoorn, S., & Knapp, S. (2015). A multi-layered risk exposure assessment approach for the shipping industry. *Transportation Research Part A*, 78, 21–33. doi:10.1016/j.tra.2015.04.032

- IALA. (2002). *IALA guidelines on the universal automatic identification system (AIS)*. Volume 1, Part II – Technical Issues. Edition 1.1.
- Iphar, C., Ray, C., & Napoli, A. (2020). Data integrity assessment for maritime anomaly detection. *Expert Systems with Applications*, 147. doi:10.1016/j.eswa.2020.113219
- ITOPF. (2020). *Oil tanker spill statistics 2019* [online]. Accessed 18th May 2020. <https://www.itopf.org/knowledge-resources/data-statistics/statistics/>.
- Jin, M., Shi, W., Lin, K., & Li, K. (2019). Marine piracy prediction and prevention: Policy implications. *Marine Policy*, 108. doi:10.1016/j.marpol.2019.103528
- Jin, M., Shi, W., Yuen, K., Xiao, Y., & Li, K. (2019). Oil tanker risks on the marine environment: An empirical study and policy implications. *Marine Policy*, 108. doi:10.1016/j.marpol.2019.103655
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Karakasnaki, M., Vlachopoulos, P., Pantouvakis, A., & Bouranta, N. (2018). ISM code implementation: An investigation of safety issues in the shipping industry. *WMU Journal of Maritime Affairs*, 17, 461–474. doi:10.1007/s13437-018-0153-4
- Kecman, V. (2005). Support vector machines – An introduction. In L. Wang (Ed.), *Support vector machines: Theory and applications* (pp. 1–48). New York: Springer.
- Knapp, S., & Franses, P. (2008). Econometric analysis to differentiate effects of various ship safety inspections. *Marine Policy*, 32(4), 653–662. doi:10.1016/j.marpol.2007.11.006
- Knapp, S., & Heij, C. (2020). Improve strategies for the maritime industry to target vessels for inspection and to select inspection priority areas. *Safety*, 6, 2. doi:10.3390/safety6020018
- Knapp, S., Kumar, S., Sakurada, Y., & Shen, J. (2011). Econometric analysis of the changing effects in wind strength and significant wave height on the probability of casualty in shipping. *Accident Analysis and Prevention*, 43, 1252–1266. doi:10.1016/j.aap.2011.01.008
- Kretschmann, L. (2020). Leading indicators and maritime safety: Predicting future risk with a machine learning approach. *Journal of Shipping and Trade*, 5. doi:10.1186/s41072-020-00071-1
- Kulkarni, K., Goerlandt, F., Li, J., Banda, O., & Kujala, P. (2020). Preventing shipping accidents: Past, present and future of waterway risk management with Baltic Sea focus. *Safety Science*, 129. doi:10.1016/j.ssci.2020.104798
- Leevy, J., Khoshgoftaar, T., Bauder, R., & Seliya, N. (2018). A survey on addressing high-class imbalance in big data. *Journal of Big Data*, 5, 42. doi:10.1186/s40537-018-0151-6
- Lensu, M., & Goerlandt, F. (2019). Big maritime data for the Baltic Sea with a focus on the winter navigation system. *Marine Policy*, 104, 53–65. doi:10.1016/j.marpol.2019.02.038
- Li, J., Goerlandt, F., & Reniers, G. (2021). An overview of scientometric mapping for the safety science community: Methods, tools, and framework. *Safety Science*, 134. doi:10.1016/j.ssci.2020.105093
- Li, K., Yin, J., Bang, H., Yang, Z., & Wang, J. (2012). Bayesian network with quantitative input for maritime risk analysis. *Transportmetrica A: Transport Science*, 10(2), 89–118. doi:10.1080/18128602.2012.675527
- Li, K., Yin, J., & Fan, L. (2014). Ship safety index. *Transportation Research Part A*, 66, 75–87. doi:10.1016/j.tra.2014.04.016
- Li, S., Meng, Q., & Qu, X. (2012). An overview of maritime waterway quantitative risk assessment models. *Risk Analysis*, 32(3), 496–512. doi:10.1111/j.1539-6924.2011.01697.x
- Likun, W., & Zaili, Y. (2018). Bayesian networking modelling and analysis of accident severity in waterborne transportation: A case study in China. *Reliability Engineering and System Safety*, 180, 277–289. doi:10.1016/j.ress.2018.07.021
- Lim, G., Cho, J., Bora, S., Biobaku, T., & Parsaei, H. (2018). Models and computational algorithms for maritime risk analysis: A review. *Annals of Operational Research*, 271, 765–786. doi:10.1007/s10479-018-2768-4
- Lisowski, J. (2016). Dynamic optimisation of safe ship trajectory with neural representation of encountered ships. *Scientific Journals of the Maritime University of Szczecin*, 47, 91–97.
- Liu, D., Wang, X., Cai, Y., Liu, Z., & Liu, Z. (2020). A novel framework of real time regional collision risk prediction based on the RNN approach. *Journal of Marine Science and Engineering*, 8(3), 224. doi:10.3390/jmse8030224

- Liu, Y., Duan, W., Huang, L., Duan, S., & Ma, X. (2020). The input vector space optimisation for LSTM deep learning model in real-time prediction of ship motions. *Ocean Engineering*, 213. doi:10.1016/j.oceaneng.2020.107681
- Ma, J., Li, W., Jia, C., Zhang, C., & Zhang, Y. (2020). Risk prediction for ship encounter situation awareness using long short-term memory based deep learning on internship behaviours. *Journal of Advanced Transportation*. doi:10.1155/2020/8897700
- Mazaheri, A., Montewka, J., Kotilainen, P., Sormunen, O. E., & Kujala, P. (2014). Assessing grounding frequency using ship traffic and waterway complexity. *Journal of Navigation*, 68(1), 89–106. doi:10.1017/S0373463314000502
- Mazaheri, A., Montewka, J., & Kujala, P. (2016). Towards an evidence-based probabilistic risk model for ship-grounding accidents. *Safety Science*, 86, 195–210. doi:10.1016/j.ssci.2016.03.002
- MCA. (2010). *The human element: A guide to human behaviour in the shipping industry*. The Stationary Office, London.
- McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133. doi:10.1007/BF02478259
- Mehdizadeh, A., Cai, M., Hu, Q., Yazdi, M., Mohabbati-Kalejahi, N., Vinel, A., ... Megahed, F. (2020). A review of data analytic applications in road traffic safety. Part 1: Descriptive and predictive modelling. *Sensors*, 20(4), 1107. doi:10.3390/s20041107
- Merghadi, A., Yunus, A., Dou, J., Whiteley, J., ThaiPham, B., Tien Bui, D., ... Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, 207, 103225. doi:10.1016/j.earscirev.2020.103225
- Mojaddadi, H., Pradhan, B., Nampak, H., Ahmad, N., & Ghazali, A. (2017). Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. *Geomatics, Natural Hazards and Risk*, 8(2), 1080–1102. doi:10.1080/19475705.2017.1294113
- Norrington, L., Quigley, J., Russell, A., & Meer, R. (2008). Modelling the reliability of search and rescue operations with Bayesian Belief Networks. *Reliability Engineering and System Safety*, 93(7), 940–949. doi:10.1016/j.ress.2007.03.006
- OpenRisk. (2018). OpenRisk guideline for regional risk management to improve European pollution preparedness and response at sea [online]. Accessed 17th October 2020. <https://helcom.fi/media/publications/OpenRisk-Guideline-for-pollution-response-at-sea.pdf>.
- Ozturk, U., Birbil, S., & Kadir, C. (2019). Evaluating navigational risk of port approach manoeuvres with expert assessments and machine learning. *Ocean Engineering*, 192. doi:10.1016/j.oceaneng.2019.106558
- Paltrinieri, N., Comfort, L., & Reniers, G. (2019). Learning about risk: Machine learning for risk assessment. *Safety Science*, 118, 475–486. doi:10.1016/j.ssci.2019.06.001
- Pietrzykowski, Z. (2001). The analysis of a ship fuzzy domain in a restricted area. *IFAC Proceedings Volumes*, 34(7), 45–50. doi:10.1016/S1474-6670(17)35057-7
- Pietrzykowski, Z., Wielgosz, M., & Breitsprecher, M. (2020). Navigators behaviour analysis using data mining. *Journal of Marine Science and Engineering*, 8, 1. doi:10.3390/jmse8010050
- Pozzolo, A., Caelen, O., Johnson, R., & Bontempi, G. (2015). Calibrating probability with undersampling for unbalanced classification. *IEEE Symposium Series on Computational Intelligence, Cape Town* (pp. 159–166). doi:10.1109/SSCI.2015.33
- Psaraftis, H. N. (2012). Formal safety assessment: An updated review. *Journal of Marine Science and Technology*, 11(3), 390–402. doi:10.1007/s00773-012-0175-0
- Qiao, W., Liu, Y., Ma, X., & Liu, Y. (2020). A methodology to evaluate human factors contributed to maritime accident by mapping fuzzy FT into ANN based on HFACS. *Ocean Engineering*, 197. doi:10.1016/j.oceaneng.2019.106892
- Rawson, A., & Brito, M. (2021a). Developing contextually aware ship domains using machine learning. *Journal of Navigation*. doi:10.1017/S0373463321000047
- Rawson, A., & Brito, M. (2021b). A critique of the use of domain analysis for spatial collision risk assessment. *Ocean Engineering*, 219. doi:10.1016/j.oceaneng.2020.108259
- Razavi, A., Inkpen, D., Falcon, R., & Abielmona, R. (2014). Textual risk mining for maritime situational awareness. *IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support*. doi:10.1109/CogSIMA.2014.6816558

- Rezaee, S., Pelot, R., & Finnis, J. (2016). The effect of extratropical cyclone weather conditions on fishing vessel incidents' severity level in Atlantic Canada. *Safety Science*, 85, 33–40. doi:[10.1016/j.ssci.2015.12.006](https://doi.org/10.1016/j.ssci.2015.12.006)
- Rezaee, S., Pelot, R., & Ghasemi, A. (2016). The effect of extreme weather conditions on commercial fishing activities and vessel incidents in Atlantic Canada. *Ocean and Coastal Management*, 130, 115–127. doi:[10.1016/j.ocecoaman.2016.05.011](https://doi.org/10.1016/j.ocecoaman.2016.05.011)
- Riveiro, M., Pallotta, G., & Vespe, M. (2018). Maritime anomaly detection: A review. *Data Mining and Knowledge Discovery*, 8, 5. doi:[10.1002/widm.1266](https://doi.org/10.1002/widm.1266)
- Tang, L., Tang, Y., Zhang, K., Du, L., & Wang, M. (2019). Prediction of grades of ship collision accidents based on random forests and bayesian networks. *5th International Conference on Transportation Information and Safety*, July 14–17, Liverpool.
- Tsou, M. (2018). Big data analytics of safety assessment for a port of entry: A case study in Keelung Harbour. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, 233(4), 1260–1275. doi:[10.1177/1475090218805245](https://doi.org/10.1177/1475090218805245)
- Uyanik, T., Karatug, C., & Arslanoglu, Y. (2020). Machine learning approach to ship fuel consumption: A case study of container vessel. *Transportation Research Part D*, 84. doi:[10.1016/j.trd.2020.102389](https://doi.org/10.1016/j.trd.2020.102389)
- Wang, C., Deng, C., & Wang, S. (2020). Imbalance-XGBoost: Leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. *Pattern Recognition Letters*, 136, 190–197. doi:[10.1016/j.patrec.2020.05.035](https://doi.org/10.1016/j.patrec.2020.05.035)
- Wang, H., Liu, Z., Wang, X., Graham, T., & Wang, J. (2021). An analysis of factors affecting the severity of marine accidents. *Reliability Engineering and System Safety*, 210. doi:[10.1016/j.res.2021.107513](https://doi.org/10.1016/j.res.2021.107513)
- Wang, J., Sii, H., Yang, J., Pillay, A., Yu, D., Liu, J., & Saajedi, A. (2004). Use of advances in technology for maritime risk assessment. *Risk Analysis*, 24(4), 1041–1063. doi:[10.1111/j.0272-4332.2004.00506.x](https://doi.org/10.1111/j.0272-4332.2004.00506.x)
- Wang, L., Huang, R., Shi, W., & Zhang, C. (2021). Domino effect in marine accidents: Evidence from temporal association rules. *Transport Policy*, 103, 236–244. doi:[10.1016/j.tranpol.2021.02.006](https://doi.org/10.1016/j.tranpol.2021.02.006)
- Wang, Y., Wang, L., Jiang, J., Wang, J., & Yang, Z. (2020). Modelling ship collision risk based on the statistical analysis of historical data: A case study in Hong Kong waters. *Ocean Engineering*, 197. doi:[10.1016/j.oceaneng.2019.106869](https://doi.org/10.1016/j.oceaneng.2019.106869)
- Wen, X., Xie, Y., Jiang, L., Pu, Z., & Ge, T. (2021). Applications of machine learning methods in traffic crash severity modelling: Current status and future directions. *Transport Reviews*. doi:[10.1080/01441647.2021.1954108](https://doi.org/10.1080/01441647.2021.1954108)
- Wu, S., Chen, X., Shi, C., Fu, J., Yan, Y., & Wang, S. (2021). Ship detention prediction via feature selection scheme and support vector machine (SVM). *Maritime Policy and Management*. doi:[10.1080/03088839.2021.1875141](https://doi.org/10.1080/03088839.2021.1875141)
- Wu, Y., Pelot, R., & Hilliard, C. (2009). The influence of weather conditions on the relative incident rate of fishing vessels. *Risk Analysis*, 29(7), 985–999. doi:[10.1111/j.1539-6924.2009.01217.x](https://doi.org/10.1111/j.1539-6924.2009.01217.x)
- Xiao, Y., Wang, G., Lin, K., Qi, G., & Li, K. (2020). The effectiveness of the new inspection regime for port state control: Application of the Tokyo MoU. *Marine Policy*, 115. doi:[10.1016/j.marpol.2020.103857](https://doi.org/10.1016/j.marpol.2020.103857)
- Yang, D., Wu, L., Wang, S., Jia, H., & Li, K. (2019). How big data enriches maritime research – A critical review of Automatic Identification System (AIS) data applications. *Transport Reviews*, 39(6), 755–773. doi:[10.1080/01441647.2019.1649315](https://doi.org/10.1080/01441647.2019.1649315)
- Yang, Z., Yang, Z., & Yin, J. (2018). Realising advanced risk-based port state control inspection using data-driven Bayesian networks. *Transportation Research Part A*, 110, 38–56. doi:[10.1016/j.tra.2018.01.033](https://doi.org/10.1016/j.tra.2018.01.033)
- Yuan, Z., Zhou, X., Yang, T., Tamerius, J., & Mantilla, R. (2017). Predicting traffic accidents through heterogeneous urban data: A case study. *23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Halifax*. doi:[10.475/123_4](https://doi.org/10.475/123_4)
- Zhang, W., Feng, X., Goerlandt, F., & Liu, Q. (2020). Towards a convolutional neural network model for classifying regional ship collision risk levels for waterway risk analysis. *Reliability Engineering and System Safety*, 204. doi:[10.1016/j.res.2020.107127](https://doi.org/10.1016/j.res.2020.107127)
- Zhang, Y., Sun, X., Chen, J., & Cheng, C. (2021). Spatial patterns and characteristics of global maritime accidents. *Reliability Engineering and System Safety*, 206, 107310. doi:[10.1016/j.res.2020.107310](https://doi.org/10.1016/j.res.2020.107310)
- Zheng, K., Chen, Y., Jiang, Y., & Qiao, S. (2020). A SVM based ship collision risk assessment algorithm. *Ocean Engineering*, 202. doi:[10.1016/j.oceaneng.2020.107062](https://doi.org/10.1016/j.oceaneng.2020.107062)