



A machine learning approach for monitoring ship safety in extreme weather events

Andrew Rawson^{a,*}, Mario Brito^b, Zoheir Sabeur^c, Long Tran-Thanh^d

^a Electronics and Computer Science, University of Southampton, Highfield, Southampton SO17 1BJ, UK

^b Centre for Risk Research, Southampton Business School, University of Southampton, Highfield, Southampton SO17 1BJ, UK

^c Department of Computing and Informatics, Talbot Campus, University of Bournemouth, Bournemouth BH12 5BB, United Kingdom

^d Department of Computer Science, University of Warwick, Warwick CV4 7AL, United Kingdom

ARTICLE INFO

Keywords:

Maritime risk assessment
Navigation safety
Machine learning
Severe weather events

ABSTRACT

Extreme weather events can result in loss of life, environmental pollution and major damage to vessels caught in their path. Many methods to characterise this risk have been proposed, however, they typically utilise deterministic thresholds of wind and wave limits which might not accurately reflect risk. To address this limitation, we investigate the potential of machine learning algorithms to quantify the relative likelihood of an incident during the US Atlantic hurricane season. By training an algorithm on vessel traffic, weather and historical casualty data, accident candidates can be identified from historic vessel tracks. Amongst the various methods tested, Support Vector Machines showed good performance with Recall at 95% and Accuracy reaching 92%. Finally, we implement the developed model using a case study of Hurricane Matthew (October 2016). Our method contributes to enhancements in maritime safety by enabling machine intelligent risk-aware ship routing and monitoring of vessel transits by Coastguard agencies.

1. Introduction

Extreme weather events such as hurricanes are a significant danger to commercial shipping. Whilst adverse wind and waves can impede a vessel's progress requiring a reduction of speed or deviation of route, accidents may occur which can result in significant loss of life, vessel damage and environmental pollution. Severe weather contributed to the some of the most significant shipping losses of the 20th Century including the Derbyshire (1980), the Prestige (2002) and the Estonia (1994). Most recently, on the 1st October 2015, the 241 m container vessel SS El Faro and its 33 crew were lost in the North Atlantic during Hurricane Joaquin (NTSB, 2017). In 2017, at least 21 vessels were lost as a result of adverse weather conditions amongst the global fleet (Allianz, 2018). In addition, more than 1500 containers are lost at sea on average each year (World Shipping Council, 2017), many the result of forces exerted by wind and waves.

Such conditions can result in a number of dangerous phenomena for vessels which might result in capsize or significant structural damage to the ship or its equipment (IMO, 2007a; Swedish Club, 2014):

- Stability – surf-riding, broaching-to, synchronous and parametric rolling.
- Physical Impact – slamming, shipping seas.
- Mechanical – Propeller Racing, Torque Rich effect on engine.

For commercial shipping, the recommended advice is simply to avoid these storms and their forecast trajectories (NTSB, 2017). Mariners can utilise extensive weather forecasting services and their own expertise to identify and assess potentially hazardous conditions. In addition, a number of commercial routing packages assist the navigator in this regard using pre-set limits of wind or wave conditions to plot the optimal route to avoid the worst of the weather (StormGeo, 2020). Yet, as accidents still occur and these conditions continue to pose a significant risk to the safety of commercial shipping, a probabilistic and risk-based method of monitoring ship safety would be of value to enable intervention in potentially hazardous situations and timely response to incidents. Yet, given the significant number of vessels at sea and the pressures on coastguards, it is not possible to manually monitor all vessels at all times. Unless some method of intelligent prioritisation is employed, hazardous situations may be missed, and warnings to vessels not given. By way of example, in 2015 the 83 m cement carrier Cemfjord

* Corresponding author.

E-mail address: A.Rawson@soton.ac.uk (A. Rawson).

<https://doi.org/10.1016/j.ssci.2021.105336>

Received 22 July 2020; Received in revised form 10 March 2021; Accepted 4 May 2021

Available online 28 May 2021

0925-7535/© 2021 Elsevier Ltd. All rights reserved.

was lost with all eight lives during extreme conditions in the Pentland Firth, north of Scotland (MAIB, 2016). The vessel capsized within a ship reporting system monitored by Shetland Coastguard, but the first indication of the incident was 25 h later that the capsized hull was discovered. At the time, three persons were responsible for operations across a significant portion of the North Sea with potentially hundreds of vessels, and therefore it was not practical to monitor each individual passage. Some means to automatically determine the high-risk nature of the Cemford's passage and therefore prioritise coastguard monitoring of its passage could have enabled prompt deployment of search and rescue assets and may have resulted in a different outcome. Furthermore, algorithms which can predict vessel risk can be utilised in weather routing algorithms to determine the least-risk path and strategic planning of search and rescue assets or risk controls (Razi and Karatas, 2020).

Therefore, given the need to improve monitoring of the safety of navigating vessels, we propose a novel approach to maritime risk assessment through the use of machine learning. Few studies have sought to apply machine learning to vessel traffic data (Fujino et al. 2018; Tang et al. 2019) or specifically to maritime risk assessment (Jin et al. 2019; Dorsey et al. 2020). However, many have recognised that by combining vessel traffic, accident and other datasets, greater insights into maritime safety can be achieved (Lensu and Goerlandt, 2019; Kul-karni et al. 2020). In particular, we outline how vessel traffic data and historical incident data can be combined and models constructed to produce a probabilistic classifier of vessel accident candidates for use in maritime risk analysis. By way of example, we implement this approach to model the probability of an accident during the Atlantic hurricane season on the East Coast of the United States. This region was chosen due to the relatively low historical frequency of weather-related accidents that prevent alternative modelling approaches and the high availability of vessel traffic data.

The remainder of this paper is set out as follows: Section 2 provides an overview of existing methods to model the risk in severe weather and discusses recent approaches to risk assessment using machine learning. Section 3 describes the methodology, variables, models and pre-processing used within this paper. Section 4 describes the results of the assessment, evaluating the strengths and limitations of this approach, including future work proposals.

2. Literature review

2.1. Modelling voyage risk in adverse conditions

Given commercial pressures within the shipping industry, there has been significant interest in offshore weather routing to improve voyage efficiency and reduce fuel cost. Adverse wind and wave conditions can result in increased fuel consumption and reduced transit speed and therefore should be avoided. Many have sought to leverage the increasing availability of maritime datasets for offshore weather routing and modelling fuel consumption (Lee et al. 2018a; Grifoll et al. 2018; Li et al., 2018; Lee et al. 2018b; Liang et al. 2019). Such studies do not routinely consider ship safety, although some have proposed hard constraints based on deterministic wind or wave limits (Cui et al. 2016; Krata and Szlapczynska, 2018; Szlapczynski and Krata, 2018), much as if they were routing through an archipelago or mined waters (Babel and Zimmermann, 2015). Others have sought to demonstrate rough weather avoidance by ships through quantitative analysis of historical data (Vettor and Soares 2016).

To better model ship safety, some authors have attempted to use historical accident and metocean data to model the probability (Soares et al. 2001; Zhang and Li, 2017) or consequences (Rezaee et al., 2016a) of accidents in different conditions. Others have proposed Bayesian Networks as a more probabilistic method for human error or ship failure in extreme conditions (Hinz et al. 2016; Antao and Soares, 2019). However, without a measure of vessel exposure, the work has limited utility in determining the relative propensity for accidents under

different conditions (Bye and Almklov, 2019). For example, it may be that incidents occur most frequently closer to shore, but this might reflect where vessels spend more of their time rather than being inherently more hazardous.

Knapp et al (2011) combine the International Comprehensive Ocean-Atmosphere Dataset with Lloyds incident data to analyse the impact of changing weather conditions on the probability of ship accidents using a binary regression model. Whilst the majority of results are not-significant, significant results were identified for wind and wave conditions over time. This conclusion is supported in further work by Heij and Knapp (2015). For fishing vessels in the Canadian Atlantic, Rezaee et al. (2016b) utilises a logistic regression model with vessel data, accident data and metocean conditions to model the influence of these factors, with a one mile per second increase in wind speed causing a 3.25% increase in incident rates.

Whilst these studies have made significant contributions to our understanding of maritime risk in adverse weather, several key limitations exist. Firstly, many studies lack a spatial/temporal approach and use aggregated vessel traffic or incident data to derive statistical relationships. This approach likely omits a number of important features relevant to the accident scenario. Secondly, the use of Bayesian Networks or other qualitative models have significant uncertainties in deriving the priors and may be subject to bias by the contributing experts (Zhang and Thai, 2016). Therefore, greater analysis of historical accident data can mitigate this. Thirdly, the use of fixed limits and thresholds of weather conditions may be overly prescriptive and cannot provide probabilistic outcomes which better reflects risk. Finally, the approaches described above have poor scalability and cannot be applied simultaneously between different voyages, geographic regions and storm systems.

To overcome these challenges, we propose the use of a supervised form of machine learning, leveraging significant volumes and varieties of data, in order to probabilistically classify the risk of an incident as a result of adverse weather conditions.

2.2. From conventional to machine learning methods for maritime risk assessment

The field of maritime risk analysis consists of a wide body of work that aims to apply quantitative modelling techniques to better understand the likelihood and consequences of maritime accidents, a summary of which has been conducted by others (Li et al. 2012a; Chen et al. 2019). Such work is often framed in the context of the International Maritime Organisation's (IMO) Formal Safety Assessment (FSA) which provides a structured and systematic methodology for risk analysis and cost benefit assessment (IMO, 2007b). The FSA aims to be goal-based and proactive rather than reactive, identifying hazards, assessing risks, identifying risk mitigation measures, performing a cost-benefit assessment, before providing recommendations. Whilst criticisms of the FSA approach have been articulated by many authors (Psarafitis 2012), it remains the most prevalent structure for maritime risk analysis within the industry (Montewka et al. 2014).

Assessing the risks of hazards within an FSA context is dominated by several approaches. Firstly, statistical analysis of accident data and aggregated vessel traffic data to derive incident rates (Bye and Almklov, 2019). Secondly, the use of expert judgement in the form of Bayesian Networks or Event Trees (Hanninen, 2014). Thirdly, the development of models to represent navigation safety, such as geometric route models (Pedersen, 1995; Mazaheri and Ylitalo, 2010; Li et al. 2012a) or time-domain simulations (Pietrzykowski and Uriasz, 2009). Much of the work of this latter approach takes the form (Pedersen, 1995):

$$P = N_a * P_c \quad (1)$$

where P is the probability of an accident per a defined unit measurement, N_a are the number of accident candidates and P_c is the causation probability or failure rate. This reflects a concept that as vessels

navigate, they encounter potentially hazardous situations, an accident occurring is conditional on a failure to take adequate steps to avoid that situation, principally through either mechanical failure or human error. Vessels will encounter far more hazardous situations than have accidents, as vessels are crewed by trained professionals and equipped with modern equipment, the conditional probability is therefore low. This reflects a similar approach to risk as a “Swiss Cheese Model” (Reason, 1990). By determining methods to predict near-miss or non-accident critical events (N_a), valuable information on the likelihood of accident occurrence can be achieved (Du et al. 2020).

Many have described inherent limitations in traditional maritime risk models. Firstly, models developed using historical incident data is a poor predictor of future incident occurrence (Bye and Almklov, 2019), partly due to a low sample size and significant underreporting (Hassel et al. 2011). Secondly, many of the existing models’ aggregate flows of vessel traffic, losing valuable information on vessel behaviour. Thirdly, the system under study is complex, with numerous interrelating human and environmental factors that contribute to an accident (Kristiansen, 2005). As a result, some have argued that that existing maritime risk models have limited predictive capability (Goerlandt and Kujala, 2014). Finally, little scientific attention has been given to the implementation of risk models for practical applications to support decision making (Kulkarni et al. 2020). Given these challenges, some have proposed that machine learning methods might achieve superior results (Jin et al. 2019).

Machine learning techniques for risk assessment are an emerging field of study (Hedge and Rokseth, 2020). Within the wider transportation discipline, many have demonstrated how supervised machine learning can be used to predict both the severity (Li et al. 2012b; Zhang and Mahadevan, 2019) and likelihood of accidents (Yuan et al. 2017; Wang et al. 2019), based on the presence of different risk features across multiple datasets.

Within the maritime domain, whilst machine learning is routinely used in applications such as fuel consumption (Uyanik et al. 2020) or computer vision (Kim et al. 2019), risk assessment has received little attention (Jin et al. 2019; Dorsey et al. 2020), even though their potential was recognised some time ago (Wang et al. 2004). Often this work utilises aggregated or static representations of risk, such as a list of vessels (Jin et al. 2019) or ship inspections and detention outcomes (Heij and Knapp, 2012), omitting the dynamic conditions and behaviours that might have contributed to that accident. By utilising vessel traffic data such as the Automatic Identification System (AIS), models can be constructed with spatial-temporal features that might enable real-time monitoring of vessel risk.

The development of tactical models of situational awareness using vessel traffic data and machine learning has been a significant area of study in recent years. Anomaly detection models allow for the characterisation of positional (location), contextual (unexpected behaviour) or kinematic (speed or course) based analysis of historical vessel traffic datasets (Riveiro et al. 2018). Developing all a priori events and situations which may be of interest to operators is not possible, and therefore the use of some form of unsupervised machine learning is widespread for this purpose (Laxhammar, 2008). Riveiro et al. (2018) provide a thorough review of the approaches provided in the literature. These include Gaussian Mixture Models (Laxhammar, 2008), density-based clustering techniques (Liu et al. 2015, Arguedas et al. 2018), Bayesian Networks (Lane et al. 2010) and K Nearest Neighbours (Tan et al. 2018). Given the significant volume of AIS data and complexity of vessel navigation, some have argued that deep neural networks will inherently have more success (Kim and Lee, 2018), however there are few examples of application. In many cases the models have limited success at matching the accuracy of an expert labelled dataset (Liu et al. 2015).

Whilst sometimes presented as such, an anomalous vessel is not necessarily at risk. Firstly, all vessels have the potential to be involved in an accident, such as following engine failure, even if they do not behave abnormally. Secondly, in some conditions the safest route might be an

abnormal one, such as avoiding a storm. Thirdly, without the inclusion of historical accident data, it is difficult to calibrate such approaches to predict risk. Therefore, the use of supervised classification methods may be better suited to probabilistically predicting the likelihood of an accident occurring.

We propose that the advancements in machine learning can be applied to assess navigational risk for vessels. Within this paper we demonstrate the capability of machine learning to produce a high resolution and scalable risk assessment for hurricanes using large heterogeneous datasets.

3. Methodology

To conduct this analysis, the datasets, processing methods, algorithms and evaluation metrics are described in the following section. The incident, vessel traffic and metocean datasets are described in Section 3.1. Data pre-processing, including feature selection, normalisation and balancing is described in Section 3.3. Finally, the classification algorithms and evaluation methods are described in Sections 3.2 and 3.4 respectively.

3.1. Datasets

Two principal datasets were developed consisting of historical accident data which is interpreted as a positive class and vessel traffic data interpreted as a negative class. For each instance, seven independent variables were identified from the literature that might impact weather related risks and are described in the following section:

1. Wind Speed (metres/second)
2. Significant wave height (metres)
3. Vessel Category (Cargo or Tanker)
4. Vessel Length (metres)
5. Vessel Flag of Convenience binary classifier (1/0)
6. Distance from Shore (Nautical Miles)
7. Vessel Age (Year of Build)

3.2. Incident data

Incident data was sourced from the IMO’s Global Integrated Shipping Information System (GISIS), a description of which is available in Zhang et al. (2021). Under the Safety of Life at Sea Convention (SOLAS) Part 1 Regulation 21, national administrations should investigate marine casualties and supply the IMO with any pertinent information concerning the investigation’s findings (IMO, 2004), which are recorded in GISIS. Whilst this therefore includes significant details of global accidents, there will be variation in reporting standards by county and it is likely that the dataset is bias towards high consequence events, with more minor accidents unreported (Hassel et al. 2011).

All Incidents between January 2005 and December 2018 (inclusive) were extracted from the system, and filtered to records containing key information such as date, latitude and longitude, a total of 3099 incidents. To identify incidents relating to severe weather conditions, keyword filtering was conducted on the incident description. Where one of the 30 keywords listed in Table 1 was found, the incident was

Table 1
Keywords used to identify incidents related to severe weather conditions.

| Type | Keyword |
|-------------|---|
| General | Deteriorating, Deteriorated, Severe Conditions. |
| Ocean | Heavy Sea(s), Rough Sea(s), freak/High/Big/Large/Rogue Wave(s), Very Rough, Heavy Swell, Pitching Heavily. |
| Atmospheric | Cyclone, Gale, Storm, Hurricane, Bad/Adverse/Strong/Severe/Rough/Heavy/Extreme/Poor Weather, Weather was bad, Strong wind(s), Due to Weather. |

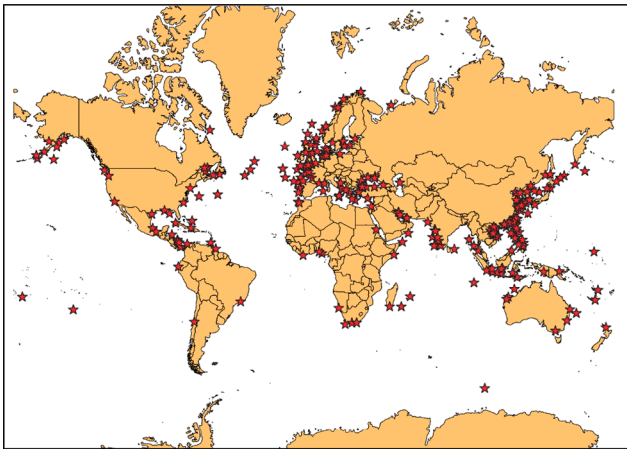


Fig. 1. Incident locations.

extracted to a separate dataset. Additional filtering was then undertaken to remove all personal injuries, man overboard or other crew related incidents and a manual review of the incident description undertaken to remove any remaining false matches. Finally, the analysis was conducted with only commercial cargo and tanker vessels, resulting in 207 unique events (see Fig. 1) which were then manually checked to ensure consistency between reported location and time with their description. Errors were corrected where identified, however, it was not possible to verify the accuracy of each incident's details against secondary sources.

3.3. Vessel traffic data

Vessel traffic data from the Automatic Identification System was used to model the navigation of vessels within the study area. AIS is a transponder system required to be fitted to all commercial vessels, under SOLAS Chapter V, and voluntarily carried by smaller craft, that sends and receives information about the movement (location, speed, course etc.) and identification (type, size, name etc.) of navigating vessels (IALA, 2002). Whilst the system was principally developed for improving maritime safety, AIS data enables high resolution analysis across a wide range of applications (Fournier et al. 2018; Yang et al. 2019).

The *Marine Cadastre* (2020), a joint project by the Bureau of Ocean Management and the National Oceanic Atmospheric Administration, publish AIS data collected by the US Coast Guard's national network of AIS receivers. Data was extracted for August, September and October for the years 2016 to 2017 in the Universal Transverse Mercator (UTM) Zones 17 to 18, approximately between Florida and Boston.

The dataset was temporally downsampled by the *Marine Cadastre* to approximately one-minute resolution. As this was significantly higher than our requirements, further downsampling was undertaken using linear interpolation to one position per vessel per hour. The interpolation allowed new positions to be created provided the time between sequential positions of the same vessel did not exceed six hours. As AIS often contains missing or erroneous descriptions of vessels (Harati-Mokhtari et al. 2007), additional datasets were used to verify vessel details such as vessel type and length. Missing values were either filled in or if no vessel type information was found, the vessel was omitted from the final dataset. The resulting vessel traffic dataset contained 735,000 positions, each representing one hour of transit.

3.4. MetOcean and topographic data

The European Union's (EU) Copernicus Marine Environment Monitoring Service provided wind and wave datasets for the analysis. Wind data was extracted from the Global Ocean Wind L4 Reprocessed 6 Hourly Observations dataset which contains global six hourly mean

wind speeds and directions at 0.25 degree resolution. Wave datasets were extracted from the Global Ocean Waves Reanalysis which contains three hourly mean significant wave heights and wave directions at a 0.2 degree resolution.

The dataset is provided in NetCDF format, a machine-independent format for representing scientific data which can be easily converted into arrays and databases. To combine the vessel traffic and incident data with the metocean data, a spatial join was conducted using a Discrete Global Grid System (DGGS). DGGS have been developed that tessellate the world into equal-area cells of platonic spatial objects that enable fast and effective combination of heterogeneous spatial data (Sahr and White, 1998). In this case, an aperture 4 hexagonal DGGS at resolution 7 (Barnes, 2018), closely resembling the units of the Copernicus data, was used as an index to conduct a spatial join between the vessel traffic, incident and metocean data.

A measure of distance from shore was sought to represent risk of grounding related hazards. The data was inputted into a Geographical Information System (GIS) and a high resolution world landmass shapefile was extracted (https://gadm.org/download_world.html). A spatial query was used to determine the geodetic distance of all data from land at 0.1, 1, 10, 100 and 1000 nautical mile distances.

3.5. Ship attribute information

Vessel length and type are provided by both the AIS and accident databases, but two additional features had to be derived. Firstly, each vessel was grouped by Flag of Convenience using the 2019 list by the International Transport Workers' Federation (ITF). Flags of Convenience are vessels which are registered in countries which offer minimal regulation or low cost and tax. This may indicate that a vessel is poorly maintained, or workers are poorly paid with long periods of work. Secondly, the year of build of a vessel was obtained for all incidents from the GISIS. This information is not available from AIS, and therefore external datasets of ship characteristics were used. Where age was not available, the IMO number was used as a proxy measure of vessel age based on a regression model with a Pearson coefficient of 0.8. Where no IMO number was available, the mean year of build of the dataset was applied.

4. Machine learning based classifiers

Multiple machine learning classification algorithms were tested on the input datasets; namely Logistic Regression (LR), Support Vector Machines (SVM), Random Forest (RF), XGBoost, SVMs optimised using Stochastic Gradient Descent (SGD) and a Multi-layer Perception (MLP). These methods have been chosen as they have been widely utilised in other transportation contexts and shown to have good predictive capabilities using widely available software libraries. In this case, these methods been implemented in python using Scikit-Learn and xgboost libraries.

1. SVMs 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; Li et al. 2012b). Whilst SVMs can support non-linear kernel functions, the training time proved excessive due to limitations on computing power and therefore a Linear-SVM model only was developed. SVMs were also tested using stochastic gradient descent as a means to improve the training speed.
2. RFs (Ho, 1995) are a popular ensemble tree learning algorithm (Jin et al. 2019; Wang et al. 2019). Within RF, a large number of uncorrelated decision trees are constructed through bagging (bootstrap aggregating), whereby the training dataset is sampled with replacement to avoid overfitting and to produce a diversity of trees.

RF then uses these trees as majority voting classifiers to produce the class prediction.

3. XGBoost (Extreme Gradient Boosting) is based on tree learning algorithms that combines boosting and gradient descent such that an ensemble of weaker models are developed that seek to correct the residual errors in previous models (Chen and Guestrin, 2016). Each decision tree makes a prediction and is then weighted based on the accuracy of their predictions, with the final prediction the weighted average of their estimates. XGBoost has been shown to have high predictive capabilities (Leevy et al. 2018; Wang et al. 2019) and scalability to massive datasets (Leevy et al. 2018), such as vessel traffic data (Jin et al. 2019).
4. LR have been widely applied for risk modelling due to their suitability of using multiple independent variables and capability to provide a probabilistic output between 0 and 1 (Rezaee et al., 2016a; Knapp et al. 2011; Jin et al. 2019).
5. A MLP is a simple feed-forward neural network architecture composed of an input layer, one or more hidden layers and an output layer. The use of this simple neural network architecture has proven successful in many applications (Wang et al. 2004; Liang et al. 2019). In this case as only a binary output is required, a single output neuron using the logistic activation function is used.

Each method contains a number of hyper parameters which can be used to optimise model accuracy. Given the dataset size and breadth of methods utilised, a grid search of all possible hyperparameter computations would be time-consuming and impractical. Previous research has shown that similar or better optimisation can be achieved far more quickly using random trials across the same grid (Bergstra and Bengio, 2012). Therefore, a parameter grid was created and a randomized search using 30 iterations with 5-fold cross validation used to determine the optimal parameters. Each method was assessed to maximise the model recall, prioritising the number of true positives such that the most likely conditions in which accidents occurred are determined. The best method in each case was then used on the set aside test set to evaluate their effectiveness.

5. Data processing and feature selection

The resultant dataset contains 735,000 positions vessel positions and 207 incidents with associated attributes, representing a highly unbalanced classification problem (Leevy et al. 2018). Feature normalisation is necessary to mitigate the impact of varying magnitudes for different features. A standard scaler (Z-normalisation) was used to standardise the input data vector in the form:

$$z_{norm} = \frac{x - \mu}{\sigma} \quad (2)$$

where μ is the mean and σ is the standard deviation of the original feature vector x . Subsequently, the dataset was split into 80% training and 20% testing datasets.

Standard classification algorithms have a natural bias towards the majority class, whereby the minority class may be incorrectly identified as noise (Fernandez et al 2017; Leevy et al 2018). Three general approaches are available to address this, namely data rebalancing, the use of class distribution sensitive models and the use of cost-sensitive learning approaches. One popular method to rebalance training data is Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al. 2002). SMOTE has shown good results in improving the capability of machine learning models in imbalanced applications such as drug trials (Saad et al 2019) and driving risk classification (Wang et al 2019), and therefore it has been implemented in this case. SMOTE searches k-nearest minority neighbours of each minority instance, selecting one of the neighbours as a reference point and generating a new value by multiplying the difference with a random value between 0 and 1 (r).

$$X_{new} = X_i + (X_{KNN} - X_i) \times r \quad (3)$$

This alters the training dataset from 587,267 non-accidents and 173 accidents to 587,440 instances of both accident and non-accident data.

5.1. Performance evaluation metrics

Simple measures of classification method performance such as the accuracy (ratio of correct to incorrect predictions) are limited when evaluating on highly imbalanced datasets. Approximately 100,000 commercial ships are active each year (Equasis, 2019), yet 21 vessels were lost as a result of bad weather conditions amongst the global fleet in 2017 (Allianz, 2018). A model which assumed no accidents occurred would only fail to predict 21 of the 100,000 vessel outcomes, an accuracy of 99.98%.

Better evaluation metrics are achieved through a confusion matrix (Table 2) which compares the predicted class values with the actual class values. From these scores, recall can be calculated as the ratio of True Positives to False Negatives and True Positives. Furthermore, specificity is the ratio of True Negatives to True Negatives and False Positives. Precision is the ratio of True Positives to True Positives and False Positives. To combine these scores into an overall measure, the F-Score is given as:

$$FScore = \frac{2(Precision * Recall)}{Precision + Recall} \quad (4)$$

In addition, the Area Under Curve (AUC) is calculated for Receiver Operating Characteristic (ROC) curves.

6. Results and discussion

6.1. Model results

A variety of results are obtained representing the strengths and weaknesses of each classifier to this particular task. Table 3 provides a summary of the model results for each of the classifiers tested. In general, high accuracy results were obtained due to the class imbalance within the testing dataset. The implementation of RF achieved the highest absolute accuracy; however, the resulting recall is low, limiting their application at identifying accident candidates. Conversely, SVM, XGBoost, LR, SGD-SVM and MLP all achieved high recall but at the expense of a number of false positives and therefore F-scores.

The high number of false positives reflects the similarity between the vessel traffic data and accident data in many circumstances. Many vessels in the dataset were recorded transiting in poor weather conditions, with significant wind speeds, yet no accident took place. Conversely, several accidents occurred during conditions less severe than would be expected. This overlap in conditions makes it challenging for a classifier to correctly identify the two classes and some reasons as to why this is the case are discussed in Section 4.4. In our case, we are more concerned with achieving high recall, so as not to falsely classify accidents as non-accidents and improve the accuracy of our accident candidate classifier.

To demonstrate the necessity of class balancing and the effectiveness of SMOTE, the training was repeated using the imbalanced dataset of 1:3400 accidents to non-accidents. The results show that whilst overall accuracy increase, a significant reduction in recall occurred. For example, XGBoost and Random Forest predicted only 13 and 8 positive records respectively, and therefore the majority of accident events have

Table 2
Confusion matrix for binary classification.

| Actual Class | Predicted Class | |
|--------------|---------------------|---------------------|
| | No Accident | Accident |
| No Accident | True Negative (TN) | False Positive (FP) |
| Accident | False Negative (FN) | True Positive (TP) |

Table 3
Classification Results on Test Set.

| Algorithm | True Positives | False Negatives | False Positives | True Negatives | Precision | Recall | F Score | Accuracy | AUC-ROC |
|---------------------|----------------|-----------------|-----------------|----------------|-----------|--------|---------|----------|---------|
| SVM | 39 | 2 | 12,056 | 134,764 | 0.003 | 0.95 | 0.006 | 0.92 | 0.98 |
| SGD-SVM | 39 | 2 | 12,112 | 134,708 | 0.003 | 0.95 | 0.006 | 0.92 | 0.98 |
| Random Forest | 12 | 29 | 7 | 146,813 | 0.632 | 0.29 | 0.400 | 0.99 | 0.94 |
| XGBoost | 31 | 10 | 1,441 | 145,379 | 0.021 | 0.76 | 0.041 | 0.99 | 0.79 |
| Logistic Regression | 38 | 3 | 12,077 | 134,743 | 0.003 | 0.93 | 0.006 | 0.92 | 0.98 |
| MLP | 33 | 8 | 5,622 | 141,198 | 0.006 | 0.80 | 0.012 | 0.96 | 0.94 |

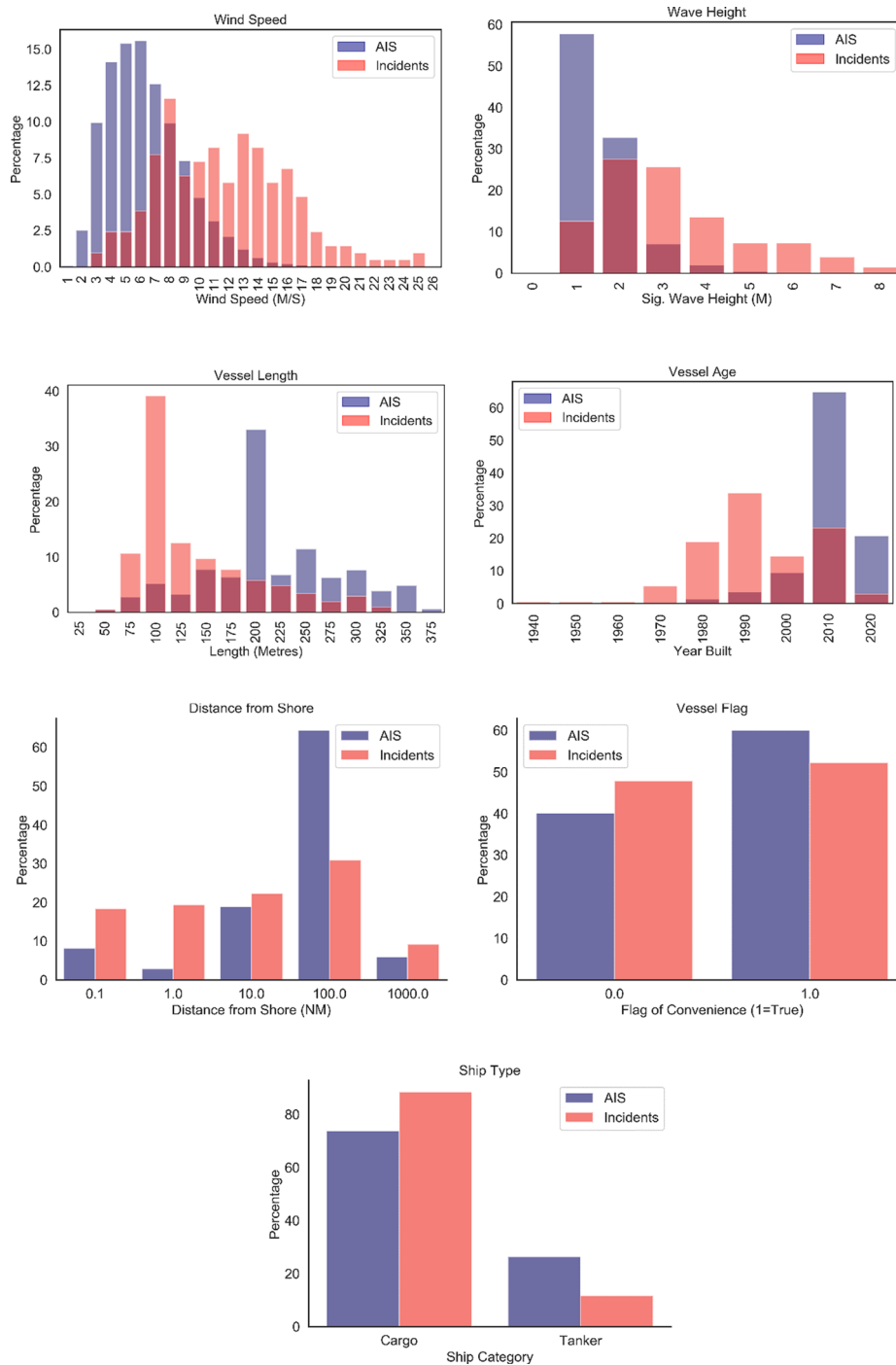


Fig. 2. Descriptive Metrics of Vessel Traffic (AIS) and Accident Datasets.

been missed, limiting its suitability as a risk model.

Based on the performance of these methods, SGD optimised SVM was used in the subsequent case study (Section 4.3). SGD-SVM showed very high recall compared to other methods whilst maintaining an acceptable number of false positives. In addition, SGD-SVM far exceeded the efficiency of other methods for both training time, completing hyperparameter tuning more than 30 times faster than the linear SVM. As such, its capability to be scaled to larger datasets is more appealing.

6.2. Analysis of variables

Fig. 2 compares a number of the features for the vessel traffic and incident datasets and Table 4 conducts a two-sided T Test to compare the two populations. In particular, it is evident that, on average, accidents involve situations with a higher wind speed, a vessel which is both smaller and older, are generally closer to shore and involve disproportionately more cargo than tanker vessels. In addition, it can be seen that a minority of accidents occur at low wind speeds, therefore signifying that either the resolution of the metocean data has not captured localised squalls or storms or that some of the accident data may not be accurate. Fig. 3 shows a correlation matrix between the exploratory variables. Whilst wind and wave conditions are highly correlated, distance from shore has some effect at wave characteristics that justify their separation as two features.

6.3. Hurricane matthew case study

Following the selection of the chosen algorithm, the model was implemented on the passage of Hurricane Matthew during October 2016. Hurricane Matthew was a Category 5 hurricane which formed on the 28th September 2016 in the North Atlantic and dissipated on the 10th October 2016 east of North Carolina. Matthew was responsible for 585 direct fatalities and significant economic damage with peak wind speeds of 145 knots (NHC, 2017). The hurricane led to both the rerouting of a most vessel traffic and the temporary closure of many ports.

Within this case study, we implement the trained SGD-SVM algorithm on AIS data for the 1st to the 10th October 2016 to monitor the risk to navigating vessels as the hurricane passes. This therefore serves as a real time vessel monitoring system that predicts which vessels have a higher likelihood of having an accident. In Fig. 4 the class probabilities for navigating vessels at a moment in time are shown. Whilst the highest potential risk vessels are shown near to the centre of the hurricane, the model is outputting varying risk scores for vessels at similar distances, reflecting the relative accident propensity of smaller vessels and vessels which are closer to shore. It is also immediately apparent that the hurricane has changed vessel behaviour, with the majority of vessels having navigated clear of the hurricane's path, particularly waiting south of Florida for Matthew to pass before continuing their transit.

Fig. 5 tracks the cumulative risk per hour during the course of the hurricane. Prior to the 5th October, the hurricane is outside the study area and therefore the risk is relatively low, representing a background level of risk with typical metocean conditions. The hurricane passes between Bahamas and Florida on the 5th which brings it close to a

significant number of vessels before passing the Atlantic seaboard which has been largely evacuated by ships, hence a decrease. Interestingly, the risk increases significantly as the storm reduces in intensity but passes close to New Jersey, New York and Delaware, due to the greater number of vessels exposed to the storm. This is reflective of a trade-off between greater inherent risk to an individual vessel as the storm is strongest, which the majority of vessels avoid, and the greater cumulative risk of a weaker storm with many more vessels exposed to it.

The preceding analysis considers the residual risk experienced by vessels, thereby partly mitigated by their actions to avoid such adverse conditions. This approach can also be applied to assess the relative risk in all locations for a given vessel at a given time. To achieve this, a mesh of metocean conditions was extracted for each hour of analysis and the characteristics of a sample vessel (Cargo vessel, length 200 m, built 2000) inputted. The trained model therefore predicts the relative risk at each grid cell at each timestamp if the sample vessel were in that location, which can be interpolated using kriging to produce a risk layer. Fig. 6 shows the output of this exercise for a certain time of Hurricane Matthew with the greatest risk at the storm centre, which most vessels are avoiding (Fig. 4). This risk map would be different for different vessels, depending on their characteristics and could be used as an input layer for a weather routing algorithm to determine the safest route.

6.4. Limitations and future work

This approach has demonstrated numerous advantages in identifying accident candidates based on historical data through machine learning. However, there are several challenges and limitations where improvements are required. Most notably, the model results achieved high recall scores at the expense of low precision scores, with many false positives. To some extent this is inevitable given that in most cases, vessels are exposed to severe conditions for some time before an accident occurs. For example, the losses of El Faro and Derbyshire spent many hours in severe conditions before succumbing. Therefore, for each data point of an accident, we would expect many more non-accident data points in similar conditions. This could be addressed through coupling this approach with a Bayesian Network to address the uncertainty of other factors, and thereby produce an output risk score. Challenges with the datasets might also undermine the results. The EU Copernicus metocean model has a 6-hour resolution, that might omit fast moving storms or isolated local squalls. In addition, the effects of wave period or risks associated with rogue waves are not included. This might partly explain why notable proportion of the accidents in Fig. 2 occur in relatively benign conditions. Similar results have been obtained by others, with half of all swell-related incidents analysed by Zhang and Li (2017) occurring in relatively low sea states (significant wave height < 3 m). Similarly, the global accident might be unrepresentative of the specific conditions experienced on the Atlantic East Coast, with benchmarking against the fewer local accidents a possible means to understand this.

Furthermore, low precision could indicate the omission of some relevant features. Accidents are rarely caused by conditions alone and there are many other contributory factors, particularly related to human error and vessel condition (Mazaheri et al. 2014; Allianz, 2018; Olba et al. 2019). Factors such as company culture, crew experience and hours of rest have been shown to have a significant influence on the propensity for accidents but cannot be observed from data. Some have proposed that other factors are associated with human error and vessel condition, such as inspection deficiencies (Heij and Knapp 2012) and classification society (Jin et al. 2019) and their inclusion in the models may improve performance.

Most significantly, the risks in adverse weather are dependent upon the actions taken by the crew through reduction of speed and maintaining a safe angle to incoming swell (Swedish Club, 2014). Collecting sufficient vessel traffic data of accident events imposes significant cost and processing requirements, and therefore an asymmetry of information between accident and non-accident datasets often occurs

Table 4

T test for exploratory variables between AIS data and incident data.

| Variable | T test statistic | P value |
|--------------------------------------|------------------|-------------------------|
| Wind Speed (Metres/Second) | 28.92 | 7.49×10^{-184} |
| Significant Wave Height (Metres) | 32.97 | 3.79×10^{-238} |
| Distance from Shore (Nautical Miles) | 0.004 | 9.97×10^{-01} |
| Vessel Length (Metres) | -16.69 | 1.49×10^{-62} |
| Flag of Convenience (True/False) | -2.28 | 2.23×10^{-02} |
| Vessel Age (Year) | -35.28 | 1.77×10^{-272} |
| Cargo Vessel | 4.80 | 1.52×10^{-06} |
| Tanker Vessel | -4.80 | 1.52×10^{-06} |

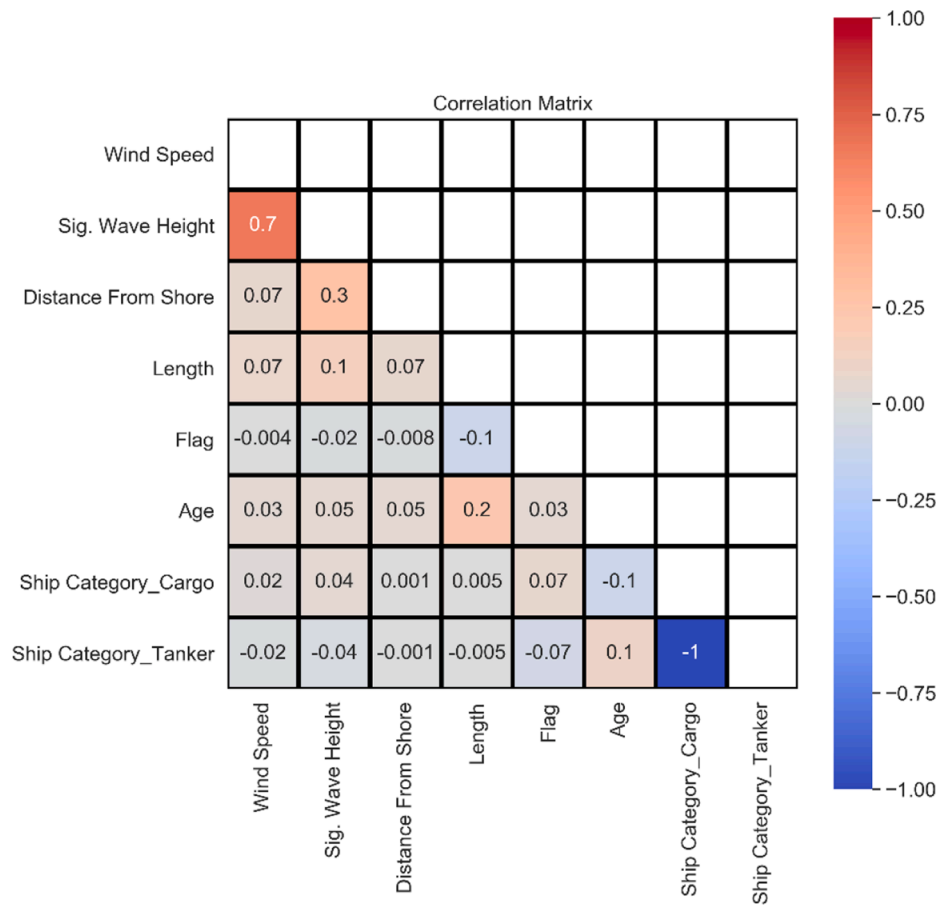


Fig. 3. Correlation matrix between exploratory variables.

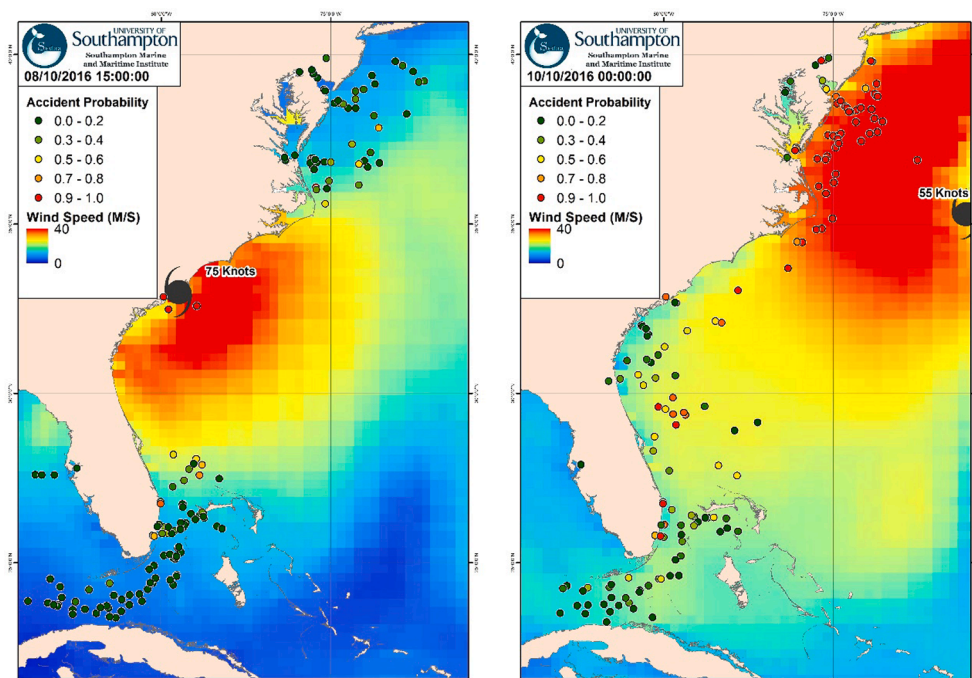


Fig. 4. Model snapshot of 8th and 10th October 2016 during Hurricane Matthew.

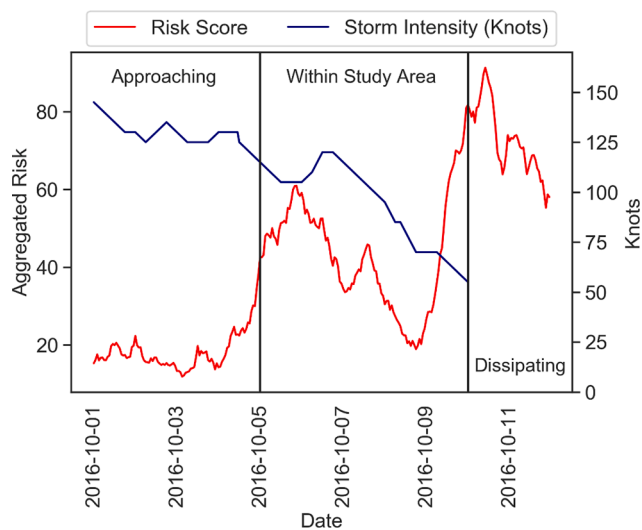


Fig. 5. Change in Relative Risk during course of Hurricane Matthew and October 2016.

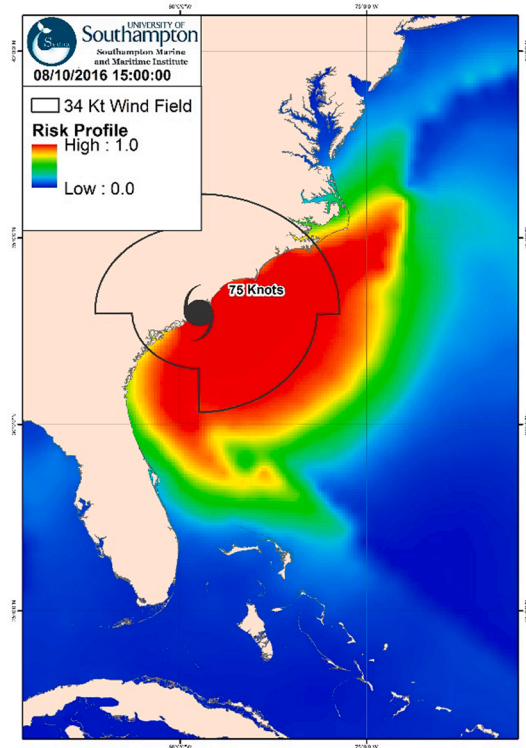


Fig. 6. Generated Risk Map for 200 m Cargo vessel built 2000 (15:00 08/10/2016).

(Mehdizadeh et al. 2020). Some research has leveraged big data analytics to process this data (AbuAlhaol et al. 2018; Filipiak et al. 2018) but to this date little of this research has been focused on risk analysis. The inclusion of vessel actions, speed and course might improve predictive capability and warrants further work. Whilst unsupervised clustering of ship behaviours might serve this purpose (Rawson and Brito, 2020), the omission of historical accident data undermines its utility as a risk analysis tool.

Given the rarity of accident data, this research has utilised a single accident category that combines many specific events which have different contributory factors and relationships with exploratory

variables, such as loss of containers, groundings or capsizes. Whilst it is desirable to have a multi class classification method that can differentiate between them, the sample size was shown to be significantly diminished as the model became more specific and resulted in over-fitting. One method to overcome this is through the use of non-accident events (Du et al. 2020) or the use of expert labelling (Zhang et al. 2020a). A further possible method to overcome this is to implement deep learning or transfer learning methodologies, whereby a model trained on one task is applied to another task. Whilst this approach is popular in computer vision and natural language processing contexts, this is an open area of research in accident prediction.

Finally, the models have outputted class probabilities which are non-dimensional risk scores rather than hazard likelihoods. Whilst such outputs have been shown to be useful in maritime risk studies (Zhang et al. 2020b), a probabilistic hazard likelihood would be preferred. The models tested assume that the input data is balanced when presenting class probabilities, which is not the case due to the significant imbalance of the dataset or the use of sampling methods which distort the class probabilities. Whilst it is possible to calibrate to correct for this (see for example Pozzolo et al. 2015), in this case the actual ratio between vessel traffic and historical accidents is unknown. Similarly, the outputted probabilities do not reflect the aleatory and epistemic uncertainties inherent in risk modelling (Aven, 2016). Nor can these results be easily validated on our case study of Hurricane Matthew, albeit challenges with reliability and validity are consistent with conventional maritime risk models (Goerlandt and Kujala, 2014).

7. Conclusions

Predicting navigational accidents through maritime risk models is an important area of research to reduce loss of life and pollution at sea. Traditional models focus on the use of expert judgement, deterministic models of ship manoeuvring characteristics, or analytical models. These models have some inherent limitations which limit their predictive capability. This paper has presented a novel approach to maritime risk analysis through the use of supervised classification machine learning algorithms. Through the processing of historical vessel accident data, AIS data and metocean data, a probabilistic classifier has been developed to assess the likelihood of a weather-related incident to transiting vessels. Whilst only the risk posed by hurricanes has been investigated in this case, this approach has good potential to enable intelligent monitoring of maritime risk by navigation authorities, automatically determining which vessels warrant enhanced monitoring or intervention based on the characteristics of previous incidents.

The model results show modest success at accident prediction. In general, the techniques are able to distinguish accidents from non-accidents, however, there are a significant number of false positives which reduce the precision and F-scores. As the methods output class probabilities, we can interpret these as potential near miss situations in traditional maritime risk models and apply causation probabilities or Bayesian Networks, to output a likelihood score of an accident. This enables a quantitative, intelligent method of identifying high-risk transits by commercial shipping that can improve our understanding of maritime safety. In addition, this may overcome some scaling limitations with traditional techniques by leveraging much greater volumes and varieties of data than is typically used in existing methods. Further work is however necessary to extend this approach to other hazard types and produce probabilistic classifiers.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is partly funded by the University of Southampton's Marine and Maritime Institute (SMMI) and the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement number: 723526: SEDNA).

Appendix A. Supplementary material

Supplementary data to this article and example model outputs can be found online at <https://doi.org/10.1016/j.ssci.2021.105336>.

References

- Abualhaol, I.Y., Falcon, R., Abielmona, R.S., Petriu, E.M., 2018. Mining Port Congestion Indicators from Big AIS Data. International Joint Conference on Neural Networks. Rio de Janeiro.
- Allianz, 2018. Safety and Shipping Review 2018. <https://www.agcs.allianz.com/content/dam/onemarketing/agcs/agcs/reports/AGCS-Safety-Shipping-Review-2018.pdf>. Accessed 19 April 2019.
- Antao, P., Soares, C., 2019. Analysis of the influence of human errors on the occurrence of coastal ship accidents in different wave conditions using Bayesian Belief Networks. *Accid. Anal. Prev.* 133 <https://doi.org/10.1016/j.aap.2019.105262>.
- Arguedas, V., Pallotta, G., Vespe, M., 2018. Maritime traffic networks: from historical positioning data to unsupervised maritime traffic monitoring. *IEEE Trans. Intell. Transp. Syst.* 19 (3), 722–732. <https://doi.org/10.1109/TITS.2017.2699635>.
- Aven, T., 2016. Risk assessment and risk management: review of recent advances on their foundation. *Eur. J. Oper. Res.* 253, 1–13. <https://doi.org/10.1016/j.ejor.2015.12.023>.
- Babel, L., Zimmermann, T., 2015. Planning safe navigation routes through mined waters. *Eur. J. Oper. Res.* 241, 99–108. <https://doi.org/10.1016/j.ejor.2014.08.002>.
- Barnes, R., 2018. dggridR: Discrete Global Grids for R. <https://CRAN.R-project.org/package=dggridR>. Accessed 24 November 2019.
- Bergstra, J., Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13, pp. 281–305. <https://doi.org/10.5555/2188385.2188395>.
- Bye, R., Almklov, P., 2019. Normalization of maritime accident data using AIS. *Marine Policy* 109. <https://doi.org/10.1016/j.marpol.2019.103675>.
- Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W., 2002. SMOTE: synthetic minority over-sampling technique. *J. Artificial Intell. Res.* 16, 321–357. <https://doi.org/10.1613/jair.953>.
- Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. *KDD*, San Francisco.
- Chen, P., Huang, Y., Mou, J., van Gelder, P., 2019. Probabilistic risk analysis for ship-ship collision: state of the art. *Saf. Sci.* 117, 108–122. <https://doi.org/10.1016/j.ssci.2019.04.014>.
- Cui, T., Turan, O., Boulougouris, E., 2016. Development of a ship weather routing system for energy efficient shipping. *Annual conference of the International Association of Maritime Economists*, Hamburg.
- Dorsey, C., Wang, B., Grabowski, M., Merrick, J., Harrald, J., 2020. Self-healing databases for predictive risk analytics in safety-critical systems. *J. Loss Prev. Process Ind.* 63, 104014 <https://doi.org/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. *Reliab. Eng. Syst. Saf.* 200 <https://doi.org/10.1016/j.res.2020.106933>.
- Equisis, 2019. The World Merchant Fleet in 2018: Statistics from Equisis. <http://emsa.europa.eu/publications/technical-reports-studies-and-plans/item/472-annual-statistical-report-on-the-world-merchant-statistics-from-equisis-from-equisis.html>. Accessed 21 November 2019.
- Fernandez, A., Rio, S., Chawla, N., Herrera, F., 2017. An insight into imbalanced Big Data classification: outcomes and challenges. *Complex Intell. Syst.* 3, 105–120. <https://doi.org/10.1007/s40747-017-0037-9>.
- Filipiak, D., Strozyna, M., Wecl, K. and Abramowicz, W., 2018. Anomaly Detection in the maritime Domain: Comparison of Traditional and Big Data Approach. *NATO IST-160-RSM Specialists' Meeting on Big Data and Artificial Intelligence for Military Decision Making*. Bordeaux, France.
- Fournier, M., Hilliard, R., Rezaee, S., Pelot, R., 2018. Past, present and future of the satellite-based automatic identification system: areas of application (2004–2016). *WMU Journal of Maritime Affairs* 17, 311–345. <https://doi.org/10.1007/s13437-018-0151-6>.
- Fujino, I., Claramunt, C. and Boudraa, A. (2018). Extracting Courses of Vessels from AIS Data and Real-Time Warning Against Off-Course. *2nd International Conference on Big Data Research*, Weihai, China.
- Grifoll, M., Martorell, L., Castells, M., Oses, F., 2018. Ship Weather Routing Using Pathfinding Algorithms: The Case of Barcelona – Palma de Mallorca. *Transp. Res. Procedia* 33, 299–306. <https://doi.org/10.1016/j.trpro.2018.10.106>.
- Goerlandt, F., Kujala, P., 2014. On the reliability and validity of ship-ship collision risk analysis in light of different perspectives on risk. *Saf. Sci.* 62, 348–365. <https://doi.org/10.1016/j.ssci.2013.09.010>.
- Hanninen, M., 2014. Bayesian networks for maritime traffic accident prevention: benefits and challenges. *Accid. Anal. Prev.* 73, 305–312. <https://doi.org/10.1016/j.aap.2014.09.017>.
- Harati-Mokhtari, A., Brooks, P., Wall, A., Wang, J., 2007. Automatic Identification System (AIS): Data Reliability and Human Error Implications. *J. Navig.* 60, 373–389. <https://doi.org/10.1017/S0373463307004298>.
- Hassel, M., Asbjørnslett, B., Hole, L., 2011. Underreporting of maritime accidents to vessel accident databases. *Accid. Anal. Prev.* 43, 2053–2063. <https://doi.org/10.1016/j.aap.2011.05.027>.
- Hedge, J., Rokseth, B., 2020. Applications of machine learning methods for engineering risk assessment – a review. *Saf. Sci.* 122 <https://doi.org/10.1016/j.ssci.2019.09.015>.
- Heij, C., Knapp, S., 2012. Evaluation of safety and environmental risk at individual ship and company level. *Transp. Res. Part D* 17, 228–236. <https://doi.org/10.1016/j.trd.2011.12.003>.
- Heij, C., Knapp, S., 2015. Effects of wind strength and wave height on ship incident risk: regional trends and seasonality. *Transport. Res. D: Transp. Environ.* 37, 29–39. <https://doi.org/10.1016/j.trd.2015.04.016>.
- Hinz, T., Krata, P., Montewka, J., 2016. The use of Bayesian networks for risk assessment of ship capsizing as an alternative way of evaluating the operational safety of ship at sea. *13th International Conference on Probabilistic Safety Assessment and management*.
- Ho, T., 1995. Random decision forests. *Proceedings of the Third International Conference on Document Analysis and Recognition*.
- IALA, 2002. IALA Guidelines on the Universal Automatic Identification System (AIS). Volume 1, Part II – Technical Issues. Edition 1 (1).
- IMO, 2004. International Convention for the Safety of Life at Sea (SOLAS). IMO, London.
- IMO, 2007a. Revised Guidelines to the Master for Avoiding Dangerous Situations in Adverse Weather and Sea Conditions. *MSC.1/Circ.1228*.
- IMO, 2007b. Formal Safety Assessment. *MSC.83/INF.2*.
- Jin, M., Shi, W., Yuen, K., Xiao, Y., Li, K., 2019. Oil tanker risks on the marine environment: an empirical study and policy implications. *Mar. Pol.* 108 <https://doi.org/10.1016/j.marpol.2019.103655>.
- Kecman, V., 2005. Support Vector Machines – an Introduction. In: Wang, L. (Ed.), *Support Vector Machines: Theory and Applications*. Springer, New York, pp. 1–48.
- Kim, K., Lee, K., 2018. Deep learning-based caution area traffic prediction with automatic identification system sensor data. *Sensors* 18. <https://doi.org/10.3390/s18093172>.
- Kim, H., Koo, J., Kim, D., Park, B., Jo, Y., Myung, H., Lee, D., 2019. Vision-based real-time obstacle segmentation algorithm for autonomous surface vehicles. *IEEE Access* 7, 179420–179428. <https://doi.org/10.1109/ACCESS.2019.2959312>.
- 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. *Accid. Anal. Prev.* 43, 1252–1266. <https://doi.org/10.1016/j.aap.2011.01.008>.
- Krata, P., Szlapczynska, J., 2018. Ship weather routing optimisation with dynamic constraints based on reliable synchronous roll prediction. *Ocean Eng.* 150, 124–137. <https://doi.org/10.1016/j.oceaneng.2017.12.049>.
- Kristiansen, S., 2005. *Maritime Transportation: Safety Management and Risk Analysis*. Elsevier, Oxford.
- 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. *Saf. Sci.* 129 <https://doi.org/10.1016/j.ssci.2020.104798>.
- Lane, R.O., Nevell, D.A., Hayward, S.D., Beane, T.W., 2010. *Maritime Anomaly Detection and Threat Assessment*. 13th International Conference on Information Fusion. Edinburgh, UK.
- Laxhammar, R., 2008. *Anomaly Detection for Sea Surveillance*. 11th International Conference on Information Fusion, Cologne, Germany.
- Lee, S., Roh, M., Kim, K., Jung, H., Park, J., 2018a. Method for a simultaneous determination of the path and the speed for ship route planning problems. *Ocean Eng.* 157, 301–312. <https://doi.org/10.1016/j.oceaneng.2018.03.068>.
- Lee, H., Aydin, N., Choi, Y., Lekhavat, S., Irani, Z., 2018b. A decision support system for vessel speed decision in maritime logistics using weather archive big data. *Comput. Oper. Res.* 98, 330–324. <https://doi.org/10.1016/j.cor.2017.06.005>.
- Leevy, J., Khoshgoftaar, T., Bauder, R., Seliya, N., 2018. A survey on addressing high-class imbalance in big data. *J. Big Data* 5 (42). <https://doi.org/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. <https://doi.org/10.1016/j.marpol.2019.02.038>.
- Li, S., Meng, Q., Qu, X., 2012a. An overview of maritime waterway quantitative risk assessment models. *Risk Anal.* 32 (3), 496–512. <https://doi.org/10.1111/j.1539-6924.2011.01697.x>.
- Li, Z., Liu, P., Wang, W., Xu, C., 2012b. Using support vector machine models for crash injury severity analysis. *Accid. Anal. Prev.* 45, 478–486. <https://doi.org/10.1016/j.aap.2011.08.016>.
- Li, X., Sun, B., Zhao, Q., Li, Y., Shen, Z., Du, W., Xu, N., 2018. Model of speed optimization of oil tanker with irregular winds and waves for given route. *Ocean Eng.* 164, 628–639. <https://doi.org/10.1016/j.oceaneng.2018.07.009>.
- Liang, Q., Tvette, H., Brinks, H., 2019. Prediction of vessel propulsion power using machine learning on AIS data, ship performance measurements and weather data. *MTEC/ICMASS 2019*, Norway.
- Liu, B., Souza, E., Hilliard, C., Matwin, S., 2015. *Ship Movement Anomaly Detection Using Specialized Distance Measures*. 18th International Conference on Information Fusion, Washington DC.
- MAIB, 2016. Report on the investigation of the capsizing and sinking of the cement carrier Cemfirth in the Pentland Firth. https://assets.publishing.service.gov.uk/media/571760fee5274a22d300001e/MAIBInVrpt_8_2016.pdf. Accessed 19 February 2020.

- MarineCadastre, 2020. Vessel Traffic Data. <https://marinecadastre.gov/ais/>. Accessed 12 March 2020.
- Mazaheri, A., Ylitalo, J., 2010. Comments on Geometrical Modelling of Ship Grounding. 5th Conference on Collision and Grounding of Ships, Espoo, Finland.
- Mazaheri, A., Montewka, J., Kotilainen, P., Sormunen, O.E., Kujala, P., 2014. Assessing grounding frequency using ship traffic and waterway complexity. *J. Navig.* 68 (1), 89–106. <https://doi.org/10.1017/S0373463314000502>.
- Mehdizadeh, A., Cai, M., Hu, Q., Yazdi, M., Mohabbati-Kalejahi, N., Vinel, A., Rigdon, S., Davis, K. and Megahed, F. (2020). A review of data analytic applications in road traffic safety. Part 1: Descriptive and predictive modelling. *Sensors*, 20. <https://dx.doi.org/10.3390/s20041107>.
- Montewka, J., Goerlandt, J., Kujala, P., 2014. On a systematic perspective on risk for formal safety assessment (FSA). *Reliab. Eng. Syst. Saf.* 127, 77–85. <https://doi.org/10.1016/j.res.2014.03.009>.
- NHC, 2017. National hurricane center Tropical Cyclone Report: Hurricane Matthew. https://www.nhc.noaa.gov/data/tcr/AL142016_Matthew.pdf. Accessed 18 November 2019.
- NTSB, 2017. Marine Accident Report: Sinking of US Cargo Vessel SS El Faro. MAR-17/01 PB2018-100342.
- Olba, X., Daamen, W., Vellinga, T., Hoogendoorn, S., 2019. Risk assessment methodology for vessel traffic in ports by defining the nautical port risk index. *J. Mar. Sci. Eng.* 8 <https://doi.org/10.3390/jmse8010010>.
- Pedersen, P.T., 1995. Collision and Grounding Mechanics. Proceedings of WEMT 95, Denmark.
- Pietrzykowski, Z., Uriasz, J., 2009. The ship domain – a criterion of navigational safety assessment in an open sea area. *J. Navig.* 62, 93–108. <https://doi.org/10.1017/S0373463308005018>.
- Pozzolo, A., Caelen, O., Johnson, R., Bontempi, G., 2015. Calibrating probability with undersampling for unbalanced classification. *IEEE Symposium Series on Computational Intelligence, Cape Town*.
- Psarafitis, H.N., 2012. Formal safety assessment: an updated review. *J. Mar. Sci. Technol.* 11 (3), 390–402. <https://doi.org/10.1007/s00773-012-0175-0>.
- Rawson, A., Brito, M., 2020. Modelling of ship navigation in extreme weather events using machine learning. *Research Publishing*, 8.
- Razi, N., Karatas, M., 2020. A multi-objective model for locating search and rescue boats. *Eur. J. Oper. Res.* 254, 279–293. <https://doi.org/10.1016/j.ejor.2016.03.026>.
- Reason, J., 1990. The contribution of latent human failures to the breakdown of complex systems. *Philos. Trans. Roy. Soc. Lond. B: Biol. Sci.* 327, 475–484. <https://doi.org/10.1098/rstb.1990.0090>.
- Rezaee, S., Pelot, R., Finnis, J., 2016a. The effect of extratropical cyclone weather conditions on fishing vessel incidents' severity level in Atlantic Canada. *Saf. Sci.* 85, 33–40. <https://doi.org/10.1016/j.ssci.2015.12.006>.
- Rezaee, S., Pelot, R., Ghasemi, A., 2016b. The effect of extreme weather conditions on commercial fishing activities and vessel incidents in Atlantic Canada. *Ocean Coast. Manage.* 130, 115–127. <https://doi.org/10.1016/j.ocecoaman.2016.05.011>.
- Riveiro, M., Pallotta, G., Vespe, M., 2018. Maritime anomaly detection: a review. *Data Min. Knowl. Disc.* 8 (5) <https://doi.org/10.1002/widm.1266>.
- Saad, A., Omar, Y., Maghraby, F., 2019. Predicting drug interaction with adenosine receptors using machine learning and SMOTE techniques. *IEEE Access* 7, 146953–146983. <https://doi.org/10.1109/ACCESS.2019.2946314>.
- Sahr, K., White, D., 1998. Discrete Global Grid Systems. *Computing Science and Statistics*, 30. <https://doi.org/10.1.1.136.7897>.
- Soares, C.G., Bitner-Gregersen, E.M., Antao, P., 2001. Analysis of the Frequency of Ship Accidents Under Severe North Atlantic Weather Conditions. Design and Operation for Abnormal Conditions II, London.
- StormGeo, 2020. Voyage Planning and Route Optimisation. <https://www.stormgeo.com/solutions/shipping/on-board-services/>. Accessed 12 March 2020.
- Swedish Club, 2014. Heavy Weather. https://www.swedishclub.com/media_upload/files/Publications/Loss%20Prevention/Heavy%20Weather_2014-06-30.pdf. Accessed 12 March 2020.
- Szlapczynski, R., Krata, P., 2018. Determining and visualising safe motion parameters of a ship navigating in severe weather conditions. *Ocean Eng.* 158, 263–274. <https://doi.org/10.1016/j.oceaneng.2018.03.092>.
- Tan, W., Weng, C., Zhou, Y., Chua, K., Chen, I., 2018. Historical Data is Useful for Navigation Planning: Data Driven Route Generation for Autonomous Ship. International Conference on Robotics and Automation, Brisbane, Australia.
- 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. In: 5th International Conference on Transportation Information and Safety, July 14-17 2019, Liverpool.
- Uyanik, T., Karatug, C., Arslanoglu, Y., 2020. Machine learning approach to ship fuel consumption: a case of container vessel. *Transport. Res. D: Transp. Environ.* 84, 102389 <https://doi.org/10.1016/j.trd.2020.102389>.
- Vettor, R., Soares, C.G., 2016. Rough weather avoidance effect on the wave climate experienced by oceangoing vessels. *Appl. Ocean Res.* 59, 606–615. <https://doi.org/10.1016/j.apor.2016.06.004>.
- Wang, J., Sii, H., Yang, J., Pillay, A., Yu, D., Liu, J., Maistralis, E., Saajedi, A., 2004. Use of advances in technology for maritime risk assessment. *Risk Anal.* 24 (4), 1041–1063. <https://doi.org/10.1111/j.0272-4332.2004.00506.x>.
- Wang, C., Liu, L., Xu, C., Lv, W., 2019. Predicting future driving risk of crash-involved drivers based on a systematic machine learning framework. *Int. J. Environ. Res. Public Health* 16, 334. <https://doi.org/10.3390/ijerph16030334>.
- World Shipping Council, 2017. Containers Lost at Sea - 2017 Update. http://www.worldshipping.org/industry-issues/safety/Containers_Lost_at_Sea_-_2017_Update_FINAL_July_10.pdf. Accessed 30 August 2019.
- 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. *Transp. Rev.* 39, 755–773. <https://doi.org/10.1080/01441647.2019.1649315>.
- 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.
- Zhang, Z., Li, X., 2017. Global ship accidents and ocean swell-related sea states. *Nat. Hazards Earth Syst. Sc.* 17, 2041–2051. <https://doi.org/10.5194/nhess-17-2041-2017>.
- Zhang, G., Thai, V., 2016. Expert Elicitation and Bayesian Network modelling for shipping accidents: a literature review. *Saf. Sci.* 87, 53–62. <https://doi.org/10.1016/j.ssci.2016.03.019>.
- Zhang, X., Mahadevan, S., 2019. Ensemble machine learning models for aviation incident risk prediction. *Decis. Support Syst.* 116, 48–63. <https://doi.org/10.1016/j.dss.2018.10.009>.
- Zhang, W., Feng, X., Goerlandt, F., Liu, Q., 2020a. Towards a Convolutional Neural Network model for classifying regional ship collision risk levels for waterway risk analysis. *Reliab. Eng. Syst. Saf.* 204 <https://doi.org/10.1016/j.res.2020.107127>.
- Zhang, J., He, A., Fan, C., Yan, X., Soares, C.G., 2020b. Quantitative analysis of risk influencing factors in the Jiangsu segment of the Yangtze River. *Risk Anal.* <https://doi.org/10.1111/risa.13662>.
- Zhang, Y., Sun, X., Chen, J., Cheng, C., 2021. Spatial patterns and characteristics of global maritime accidents. *Reliab. Eng. Syst. Saf.* 206, 107310 <https://doi.org/10.1016/j.res.2020.107310>.