Risk Analysis, Vol. 42, No. 10, 2022 DOI: 10.1111/risa.13866

Spatial Modeling of Maritime Risk Using Machine Learning

Andrew Rawson,[1,*] **Mario Brito**,[2] **and Zoheir Sabeur**[3]

Managing navigational safety is a key responsibility of coastal states. Predicting and measuring these risks has a high complexity due to their infrequent occurrence, multitude of causes, and large study areas. As a result, maritime risk models are generally limited in scale to small regions, generalized across diverse environments, or rely on the use of expert judgement. Therefore, such an approach has limited scalability and may incorrectly characterize the risk. Within this article a novel method for undertaking spatial modeling of maritime risk is proposed through machine learning. This enables navigational safety to be characterized while leveraging the significant volumes of relevant data available. The method comprises two key components: aggregation of historical accident data, vessel traffic, and other exploratory features into a spatial grid; and the implementation of several classification algorithms that predicts annual accident occurrence for various vessel types. This approach is applied to characterize the risk of collisions and groundings in the United Kingdom. The results vary between hazard types and vessel types but show remarkable capability at characterizing maritime risk, with accuracies and area under curve scores in excess of 90% in most implementations. Furthermore, the ensemble tree-based algorithms of XGBoost and Random Forest consistently outperformed other machine learning algorithms that were tested. The resultant potential risk maps provide decisionmakers with actionable intelligence in order to target risk mitigation measures in regions with the greatest requirement.

KEY WORDS: machine learning; Maritime risk assessment; risk mapping

1. INTRODUCTION

Maritime accidents such as collisions and groundings can result in significant loss of life, pollution, and economic losses. Accurately characterizing maritime risk within an area is a critical task for decisionmakers. Coastal states need to determine the requirement for different risk mitigation measures such as traffic routing measures or pilotage (IMO, 2004). The offshore renewables or oil and gas industries need to ensure that the risks to their developments, and impact on navigation safety are acceptable. Ports and harbors need to ensure that their waterways are safe for trading vessels and appropriate risk controls are in place. Maritime safety assessments are often framed in the context of the International Maritime Organization's (IMO) Formal Safety Assessment (FSA) (Montewka, Goerlandt, & Kujala, 2014), which provides a structured and systematic methodology for risk analysis (IMO, 2018). The FSA is goal-based and proactive rather than reactive, identifying hazards, assessing risks, identifying mitigation measures, and performing a cost-benefit assessment before providing recommendations.

Numerous quantitative methods are proposed for use in maritime risk studies (OpenRisk, 2018).

[1]Electronics and Computer Science, University of Southampton, Highfield, Southampton, UK.

[2]Centre for Risk Research, Southampton Business School, University of Southampton, Highfield, Southampton, UK.

[3]Department of Computing and Informatics, Talbot Campus, University of Bournemouth, Bournemouth, UK.

*Address correspondence to Andrew Rawson, School of Electronics and Computer Science, University of Southampton, SO17 1BJ, UK; tel: +44(0)2380 595000; fax: +44(0)2380 593131; A.Rawson@soton.ac.uk.

0272-4332/22/0100-2291$22.00/1

These include statistical analysis of accident data and incident rates (Bye & Almklov, 2019), the use of expert judgement in the form of Bayesian Networks (Hanninen, 2014; Montewka, Goerlandt, & Kujala, 2015) and analytical risk modeling, such as geometric route models (Li, Meng, & Qu, 2012; Mazaheri & Ylitalo, 2010; Pedersen, 1995) or time–domain simulations (Pietrzykowski & Uriasz, 2009; Wang, 2013). While these methods have made a significant contribution to understanding maritime safety, limitations have been identified in probabilistic risk assessments more generally (Aven & Zio, 2011), and maritime risk assessments specifically (EMSA, 2018; Hoorn & Knapp, 2015; Psaraftis, 2012). Maritime accidents occur infrequently, limiting the available sample size, and while some have proposed using near misses (Du, Goerlandt, & Kujala, 2020) or expert judgement (Mazaheri, Montewka, Kotilainen, Sormunen, & Kujala, 2014), this may not accurately reflect the circumstances of previous incidents. Most accidents are the result of human error (Weng, Yang, Chai, & Fu, 2019), which requires a detailed understanding of human factors and organizational influences that can be challenging to model. Furthermore, often there is a complex interplay of different factors that leads to an accident, which leads to difficulty in diagnosing the root cause (Brito, Smeed, & Griffiths, 2014).

While predicting each individual accident occurrence is challenging, over time accidents tend to occur more frequently in some places than others (Hoorn & Knapp, 2015). Spatially dependent variables include traffic volume, bathymetry, weather, and a myriad of other factors (Mazaheri et al., 2014). By mapping the presence of these risk factors, it may be possible to develop national scale, high resolution, strategic risk maps to support decisionmakers. While some work has attempted such an exercise, several key challenges need to be addressed. First, how can the multitude of maritime risk factors be quantified and integrated into a model given their heterogenous formats? Second, how can the complex relationships between vessel traffic, incidents, and relevant contributory factors be determined? Third, what is the effect of aggregating the input data, both spatially and temporally, on the effectiveness of the model? Finally, how can the performance of these models be evaluated and what is their utility to navigation authorities, given their potentially high cost.

To address these challenges, this article proposes a novel approach to maritime risk assessment by using machine learning methods to predict the likelihood of collisions and groundings in U.K. waters. Several key contributions are made. First, a pipeline is proposed for integrating massive and heterogenous maritime data sets into a common spatial data structure, recognized as an emerging but important trend (Kulkarni, Goerlandt, Li, Banda, & Kujala, 2020), partly due to the inherent challenges of achieving this (Lensu & Goerlandt, 2019). Second, the strength of different machine learning methods to predict the spatial distribution of maritime accidents is tested. In situations with complex, nonlinear relationships between heterogenous data sets, machine learning methods have been shown to be effective but there are few examples for maritime risk assessment (Adland, Jia, Lode, & Skontorp, 2021; Jin, Shi, Yuen, Xiao, & Li, 2019) and it is a growing method in risk analysis more generally (Nateghi & Aven, 2021). Finally, it has been noted that little scientific attention has been given to implementing maritime risk models to support decision making (Kulkarni et al., 2020), and this article sets out a practical, structured framework for achieving this.

The remainder of this article is set out as follows: Section 2 provides an overview of existing methods to develop spatial risk models. Section 3 describes the general methodological approach, which is proposed in developing the spatial model, including the variables, algorithms, and data preprocessing. Section 4 implements this methodology for a case study on assessing navigational safety in the United Kingdom. Finally, Section 5 discusses the implications of this work and proposes several areas of future research.

2. LITERATURE REVIEW

2.1. Spatial Models of Maritime Risk

Mapping the relative likelihood of maritime risk are often reliant on either expert judgement or historical accidents. First, in practice risk assessment is performed using expert judgement using hazard workshops and risk matrices (see for example Port of Dover, 2016). Limitations of risk matrices have been widely discussed (Cox, 2008; Hubbard, 2009; Kontovas & Psaraftis, 2009) and include a fixed spatial scale of assessment. Second, the sparsity of historical accident data might result in regions without previous accidents being incorrectly interpreted as having zero risk (Rawson, Sabeur, & Correndo, 2019). One method to overcome this is to generalize the study areas into much larger regions (Bye & Almklov, 2019), but doing so compromises the important insights from localized risk factors.

To incorporate the spatial element, several studies have utilized Geographical Information Systems (GIS). In general, regions are subdivided into a grid of smaller spatial units, data sets are mapped across this grid, and some form of calculation conducted to derive a risk score. This might include assigning weightings derived from expert judgement (Halpern, Walbridge, & Selkoe, 2008), fuzzy logic, or statistical techniques (Hoorn & Knapp, 2015, 2019). Table I provides an evaluation of several major regional scale studies into maritime risk, demonstrating a great variation in purpose, data sets, methods, and spatial units. Several limitations are common between these studies on each of these aspects, which should be addressed.

First, some studies seek to guide national policy (Marico, 2015; Safetec, 1999), while others are tailored to evaluating the impacts of new terminals (DNV, 2012; Van Dorp, Harrald, & Merrick, 2008, 2014). The scope of the studies is limited to one specific issue, omitting many other potentially significant hazards. Furthermore, the models have a high cost and are conducted in isolation (EMSA, 2018), becoming quickly out of date, superseded by changing shipping patterns or new developments.

Second, the principal inputs used in these assessments include expert judgement, vessel traffic data collected from the Automatic Identification System (AIS), historical incident data, and various environmental data sets. Some have criticized maritime risk assessments as overly qualitative (Psaraftis, 2012) and limitations of expert judgment have been widely discussed (Kahneman et al., 1982). In addition, issues with the quality of the accident (Hassel, Asbjornslett, & Hole, 2011; Qu, Meng, & Li, 2012) and vessel traffic data (Harati-Mokhtari, Brooks, Wall, & Wang, 2007) have been highlighted.

Third, a variety of different techniques are utilized to calculate risk. These range from applying international accident rates (Genivar, 2013), mapping traffic flows with risk factors using a Bayesian Network (DNV, 2012, 2013) to comprehensive traffic simulations (Van Dorp et al., 2008, 2014). It is notable that many studies utilize proprietary models developed by the consultancies involved and therefore lack transparency (EMSA, 2018; Psaraftis, 2012). While these models have their genesis within the academic literature (Fowler & Sorgard, 2000; Merrick & van Dorp, 2006), several recent studies have challenged the underlying assumptions of these models, such as the relationship between traffic flows, grounding accidents (Mazaheri et al., 2014), and near misses (Rawson & Brito, 2021).

Fourth, different approaches have been devised to tackle the spatial element of the assessments with routes, grid cells, or large regions used. Each of these seek to reduce the complexity of the assessment from effectively infinite spatial variation of maritime risk to a manageable scale of assessment. Where grid cells are utilized, none have considered the potentially significant implications on the derived results of spatial distortion due to mapping a square grid on a spherical globe (Barnes, 2016, 2019; Battersby, Stebe, & Finn, 2016). Furthermore, as a result of different methodological implementations and assumptions, each study is conducted in isolation for a specific waterway which prevents comparison.

Lensu and Goerlandt (2019) noted that a compromise is generally required on either the study area size and data set volume or the methodological complexity. In this article, a machine learning approach to maritime risk assessment is proposed that addresses this compromise as well as the aforementioned methodological challenges.

2.2. Machine Learning in Spatial Risk Assessment

Machine learning techniques for risk assessment are an emerging field of study (Hedge & Rokseth, 2020; Nateghi & Aven, 2021). Supervised learning of accident data, whereby a model is constructed on data containing both input and outputs, has two key applications within transportation safety. First, predicting the severity of accidents based on the accident characteristics (Lee, Yoon, Kwon, & Lee, 2019; Li, Liu, Wang, & Xu, 2012; Zhang & Mahadevan, 2019). Second, predicting the likelihood of accidents based on identified risk factors such as driving style, personal descriptive characteristics (Fang, Qiu, Zhao, & Jin, 2018; Wang, Liu, Xu, & Lv, 2019), and environmental conditions (Yuan, Zhou, Yang, Tamerius, & Mantilla, 2017).

Within the maritime domain, this topic has been rarely addressed though their potential was recognized some time ago (Wang et al., 2004). Several studies have sought to identify which vessels are likely to have accidents given their characteristics

Table I. Review of Selected Major Spatial Risk Models

N	Study	Description/Scope	Spatial Units	Principal Data Sources	Summary of Model
1	MEHRAs Safetec (1999)	Identify marine environmental high-risk areas around the U.K. Coast based on the risk of pollution and environmental sensitivity	Varied: between 4.5 × 75 nm and 60 × 32 nm	Vessel routes database Incident data Environmental data	For each cell and route, the geometric probability of vessel/shoreline interactions was multiplied by a causation factor representing different local factors/risk controls to produce an incident probability. Spill probabilities were derived from conditional tables given accident type and vessel types. Environmental sensitivity was mapped based on weighted scores of different receptors. Risk is presented as incident probability × spill size × environmental sensitivity.
2	Prince Rupert Marine Risk Assessment DNV (2012)	A risk assessment for potential introduction of LNG and oil tanker traffic and any possible associated risks or hazards.	7 "Routes"	Traffic numbers Environmental data Operational data Generic incident rates	MARCS Model: creates vessel routes from AIS data, multiplies by base accident frequencies from incident data which is modified with environmental data sets to account for localized risk. Risk controls are then applied as a percentage effectiveness.
3	Risk Assessment for Marine Spills in Canadian Waters Genivar (2013)	Estimate the relative risk for ship-source spills of oil in Canadian waters.	77 broad regions	Transit data Incident data Environmental data sets	For each region, calculate volume of oil transiting through, multiply by a generic accident rate and oil spill size distributions. An environmental sensitivity index (ESI) per area is then defined, risk being the product of oil spill volume and ESI.
4	North East Shipping Risk Assessment DNV (2013)	Assessment of navigational risks due to shipping in open waters	1 nm grid	AIS Environmental data Operational data Generic incident rates	MARCS Model: creates vessel routes from AIS data, multiplies by base accident frequencies from incident data which is modified with environmental data sets to account for localized risk. Risk controls are then applied as a percentage effectiveness.
5	Assessment of Marine Oil Spill Risk and Environ-mental Vulnerability for the State of Alaska RPSasa et al. (2014)	Determine the probabilities of spills occurring with respect to geographic region, oil type, and season, as well as the potential impacts from an oil spill.	14 large broad regions	Incident data	Spill likelihood is calculated based on analysis of historical incidents per region and per oil type. These were increased based on future traffic projections. In a similar fashion, historical incident analysis was used for spill volumes. Where no incidents occurred in a region, the rates were manually altered to reflect the project team's opinion. Environmental sensitivities were mapped to derive vulnerability.
6	BE-AWARE (2014)	Gain a better understanding of the regional and subregional risk of accidents and the potential for marine pollution events in the North Sea.	Route Network	AIS data Incident data	AIS data are compressed into thousands of network routes. For each route, the geometric probability and causation probability are multiplied by the number of passages. An oil spill output model is then applied.
7	LINZ Hydro-graphic Risk Assessment Marico (2015)	Assess navigational risk to prioritize the need for nautical charting.	Varied: 500 m to 3 km cells	AIS S57 Charts Environmental data sets	Within each grid cell, for each vessel type, the total number of transits/year are multiplied by a generic causation factor and then an event tree to derive annualized potential consequence for people, environment, and monetary impacts. A localized weighted causation modifier is applied (e.g., complexity, chart age, hazards, etc.). This is then multiplied by a weighted consequence factor for each consequence type (e.g., response complexity, World Heritage Sites, tourist sites, wetlands, etc.).

(*Continued*)

Table I. (Continued)

N	Study	Description/Scope	Spatial Units	Principal Data Sources	Summary of Model
8	Marine Environmental Risk Assessment – Greenland DNV (2015)	Quantify and describe the likelihood of marine accidents with and without pollution around Greenland.	10km grid	AIS Incident data Ice coverage	Risk modeling is undertaken per each grid cell, by multiplying likelihood and consequence. Accident frequency is derived by calculating annual distance traveled, multiplying by base accident frequencies/nm per vessel type globally and an adjustment factor (e.g., cells within 2 nm of coast have a 10× adjustment factor for grounding). Finally, a table of fuel spill likelihood is used per accident scenario. Spill volumes are similarly derived based on vessel size and type tables. Environmental sensitivities are derived on a per species basis, using their vulnerability and mortality to score their sensitivity.
9	VTRA Van Dorp et al. (2008, 2014)	The Vessel Traffic Risk Assessment (VTRA) assesses the likelihood of collisions, allisions, and groundings in the Puget Sound (USA).	0.5nm grid	AIS data Incident data Environmental data Expert elicitation	A simulation is constructed from AIS data that counts interactions between vessels, projects future courses and models possible drifting patterns. A base accident rate is used, which is modified using a Bayesian pairwise expert elicitation model to account for other causal factors. Finally, oil spill outflow for each accident situation is modeled.

using ship details (Jin et al., 2019) or inspection outcomes (Heij & Knapp, 2018). However, such models use only descriptive variables of the vessels such as age, flag, or size and therefore lacks any spatial element. Others have incorporated spatial and temporal data sets such as weather conditions to improve the discrimination of accidents between vessel transits (Adland et al., 2021; Rawson, Brito, Sabeur, & Tran-Thanh, 2021; Wu, Pelot, & Hilliard, 2009). By aggregating these data sets, a strategic risk tool can be developed to identify which regions have a higher propensity for accidents.

There are two approaches that can be taken in order to frame the research question; namely, regression to predict some target variable in an area based on input features; or classification of whether a variable occurs or not in that location. Accident frequency as a continuous variable is naturally a regression problem, and some studies have sought to map crime frequency (McClendon & Meghanathan, 2015), road accident frequency (Pan, Fu, & Thakali, 2017), and susceptibility of landslides (Lee, Hong, & Jung, 2017). However, the presence or absence of a certain event can also be predicted spatially using classification algorithms if we consider that all locations and times where an incident has occurred is a positive case whereas all other locations and times are negative cases. For example, the presence or absence of forest fires can be modeled as a classification exercise using historical fire events and various spatial variables such as elevation, land use, and rainfall (Agarwal, Tang, Narayanan, & Zhuang, 2020; Nguyen et al., 2018; Rodrigues & Riva, 2014). Other examples of this approach include air pollution (Choubin et al., 2020), avalanches (Choubin et al., 2019), flooding disasters (Li et al., 2019; Mojaddadi, Pradhan, Nampak, Ahmad, & Ghazali, 2017; Tetri et al., 2019), or road traffic accidents (Moosavi, Samavaian, Parthasarathy, Teodorscu, & Ramnath, 2019; Yuan et al., 2017). The trained model can then be applied to the entire data set in order to produce regional maps of the relative likelihood of certain events (Li et al., 2019; Mojaddadi et al., 2017; Nguyen et al., 2018). High resolution and data driven impact maps can be generated, far exceeding the detail and scale of what other methods might achieve, including areas where no historical events have previously occurred.

Section 3 describes a general methodological framework for achieving these aims in the context of maritime safety before testing its effectiveness using a U.K. case study in Section 4.

3. GENERAL METHODOLOGY

The proposed methodology consists of four steps (Fig. 1), which are described in the following subsections.

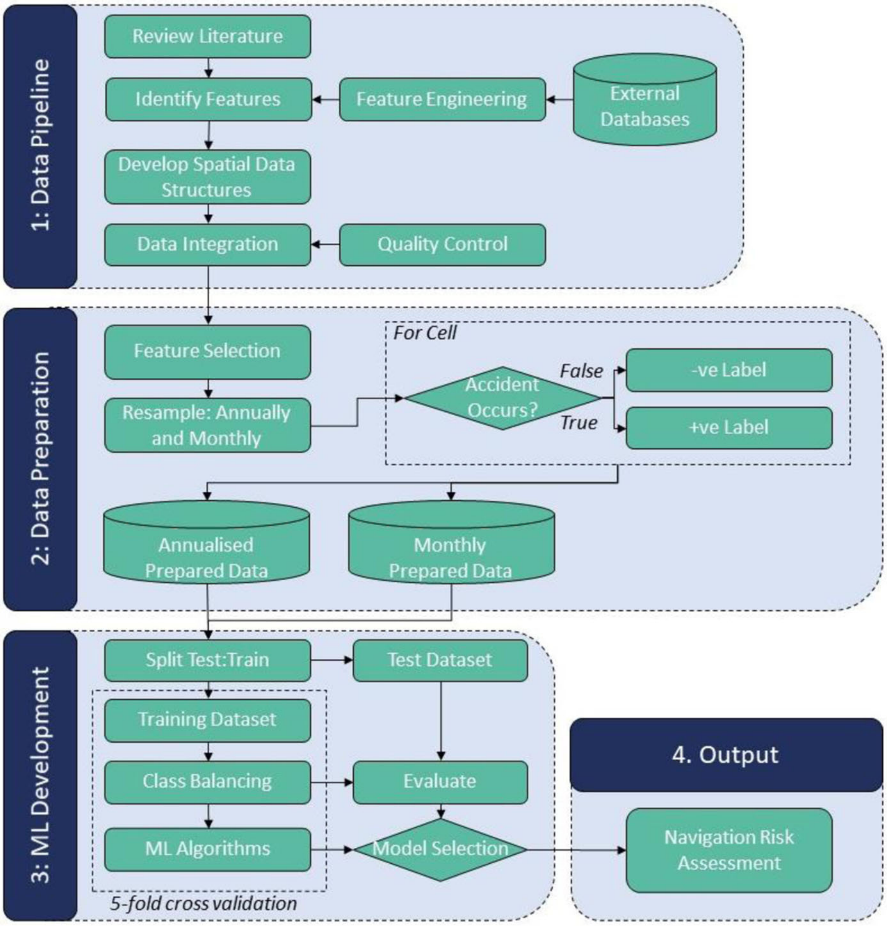

Fig 1. Methodological approach.

3.1. Step 1: Data Pipeline

Step 1 requires the development of a data pipeline to identify, quantify, and integrate multiple risk factors as features into a common data model. Three principal sources of data are required. First, accident data are the class label and can be obtained from national administrations or commercial proprietary data sets. Second, all other factors being equal, we might naturally expect more accidents in locations where more vessels transit and therefore a key input is a measure of vessel activity. Multiple standardized units of measurement can be developed using AIS data, including the number of transits, hours of transit, or distance traveled (Bye & Almklov, 2019).

Third, other independent variables related to the relative likelihood of maritime accidents have been proposed (Bye & Aalberg, 2018; Hoorn & Knapp, 2019; Kite-Powell, Jin, Jebsen, Papkonstantinou, & Patrikalakis, 1999; Kristiansen, 2005; Mazaheri et al., 2014; Mazaheri, Montewka, & Kujala, 2016; Olba, Daamen, Vellinga, & Hoogendoorn, 2019; USCG, 2010). These can be categorized into human, mechanical, and external factors and an overview is provided in Table II. Not all of these causal factors are well suited to integration into a spatial model, but different methods have been proposed to model weather (Adland et al., 2021; Knapp, Kumar, Sakurada, & Shen, 2011; Rezaee, Pelot, & Ghasemi, 2016) or ship characteristics (Bye & Aalberg, 2018; Heij & Knapp, 2018; Jin et al., 2019) among others. Some features might need to be engineered or obtained from experts. For example, mapping the perceived complexity of navigation as characterized by ship's masters (Mazaheri et al., 2014).

Having identified and obtained relevant data sets, a base spatial model is required to fuse the heterogeneous data sets with different geometries,

Table II. Significant Causes of Maritime Accidents

Category	Cause
Human and Organizational Factors	Inattention and Fatigue
	Bridge Resource Management
	Communication
	Position Monitoring
	Training and experience
	Regulation
Vessel and Mechanical Factors	Ship Dimensions and Maneuverability Characteristics
	Vessel Age
	Vessel Flag State and Safety Regime
	Maintenance
	Vessel Speed
External Factors	Traffic Density and distribution
	Waterway geometry
	Depth
	Weather (Visibility, wave, ice, darkness, etc.)
	Hydrodynamic Effects (e.g., Tidal, Bank Effect)
	Support Availability (VTS, Tugs, TSS, Pilotage, Aids to Navigation, etc.)

scales, and accuracy. Binning of spatial data into a discrete number of cells reduces complexity and enables standardized statistical methods to be applied. Conventionally, cartesian grid systems, with fixed *x*-*y* dimensions, are widely employed in maritime studies (Filipiak, Strozyna, Wecel, & Abramowicz, 2018; Wu, Xu, Wang, Wang, & Xu, 2017). However, such a structure attempts to map a regular lattice onto a spherical globe, inevitably introducing a number of distortions in cell size and shape that could limit the validity of analysis (Battersby et al., 2016). Within this study, a form of equal-area, hexagonal, and global tessellation is implemented known as the Discrete Global Grid System (DGGS) (OGC, 2019). The final data set should then be quality checked to identify and correct any missing or spurious values.

3.2. Step 2: Data Preparation

At this juncture there are two methods through which to frame the problem, namely regression and classification. If taken as the number of incidents per year, the relative sparsity of incidents means that in any individual waterway almost all records will take the form of 1 or 0. For example, there are few locations where multiple collisions of commercial vessels happen every year. As a result, classification models may be better suited, and have been implemented in this case study. Each DGGS cell is sampled on accident occurrence at different temporal scales. The data have also been aggregated annually, but also compared by month to test the importance of seasonal factors. For example, the occurrence of an accident in a cell in one month but not the other 11, is represented as one positive and 11 negative samples.

3.3. Step 3: Machine Learning Model Development

The prepared data set consists of a number of exploratory features, and a binary label of accident occurrence or not occurrence. The data set is randomly split into a training and testing data set with a ratio of 70% to 30%, respectively. Model development and tuning is conducted utilizing the training data set and evaluated on the test set.

Classification algorithms have a natural tendency toward the majority class, incorrectly treating the minority class as noise (Leevy, Khoshgoftaar, Bauder, & Seliya, 2018). Strategies to redress this include data rebalancing, the use of class distribution sensitive models, and the use of cost-sensitive learning approaches. In this study, a powerful technique to generate realistic new samples to balance the training data sets called Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002) is used. As shown in Equation 1, SMOTE searches k-nearest minority neighbors of each minority instance (X_i), selecting one of the neighbors as a reference point (X_{KNN}) and generating a new value (X_{new}) by multiplying the difference with a random value, *r*, between 0 and 1 (*r*).

$$X_{\mathrm{new}} = X_i + (X_{\mathrm{KNN}} - X_i) \times r. \tag{1}$$

There are a significant number of possible machine learning algorithms suitable for classification. This work implements four that have been shown to achieve good performance in previous studies, namely Logistic Regression, Support Vector Machines (SVM), Random Forest, and Gradient Boosted Trees. Each algorithm has hyperparameters which impact prediction performance and require tuning by iteratively retraining the model with different input hyperparameters and comparing performance. The best performing model is then chosen for further evaluation and implementation using the test data set.

3.3.1. *Logistic Regression*

Logistic Regression has been utilized by many for maritime risk modeling due to its suitability of using multiple independent variables, greater transparency, low computational requirements, and capability to provide a probabilistic output between 0 and 1 (Jin et al., 2019; Knapp et al., 2011; Rezaee et al., 2016). Independent variables $X\,(x_1, x_2, x_3\ldots)$ explain a binary outcome y_i, where y_i is 1 if an accident occurs, otherwise y_i is 0:

$$\text{Logit }(p_i) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k. \tag{2}$$

Where Logit(p_i) is the logit transformation of the odds and β_0 and β_{n} are the bias and feature coefficients respectively, and can be rewritten as:

$$P_i = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}. \tag{3}$$

3.3.2. *Support Vector Machines*

An SVM can perform linear and nonlinear classification by constructing a hyperplane or set of hyperplanes in high dimensional space to maximize the margin between training examples (Kecman, 2005). A linear SVM's decision function is created from the feature weights vector (w) plus a bias term (b):

$$w^T X + b = 0. \tag{4}$$

To maximize the margin, it is necessary to minimize the weight vector by solving the optimization problem of: $\underset{w,b}{\text{minimize}}\ \frac{1}{2} w^T w$

$$\text{subject to } y_i \left(w^T x_i + b\right) \geq 1 \text{ for } i = 1, 2, \ldots, m. \tag{5}$$

In cases where the data are not linearly separable, the hyperplane margins can be made soft. In addition, it is possible to apply a kernel function to transform the data into a higher dimensional space, where it is linearly separable.

3.3.3. *Random Forest*

Random Forests develop an ensemble of decision trees and has attractive properties such as training speed and robustness when using high-dimensional and unbalanced data sets (Brieman, 2001). Decision trees are constructed in a top-down recursive manner that partitions the data into different groups. At each step, a feature k is split by a threshold value t_k so as to maximize the purity of each subset such that each node is as homogenous as possible. The cost function (J) that is optimized can be represented as below, where G and m represent the impurity and number of instances of each subset, respectively.

$$J\ (k, t_k) = \frac{m_{\text{left}}}{m}\, G_{\text{left}} + \frac{m_{\text{right}}}{m} G_{\text{right}}. \tag{6}$$

Gini impurity (G) is used to measure the proportion of training instances that belong to the same class, where $p_{i,k}$ is the ratio of class k among the training instances in the i-th node.

$$G_i = 1 - \sum_{k=1}^{n} {p_{i,k}}^2. \tag{7}$$

Decision trees are prone to overfitting and random forest introduces several features (Breimen, 2001). First, bagging (bootstrap aggregating) involves the training data set being sampled with replacement. Second, randomly selecting attribute variables when splitting the data set. This leads to decorrelation of each model. The model prediction is the aggregated majority decision of the ensemble of individual trees.

3.3.4. *Gradient Boosted Trees (XGBoost)*

Overfitting of decision trees can also be addressed through boosting, which generates an ensemble of weaker models that seek to correct the residual errors in previous models to create a stronger classifier (Friedman, 2001). New models are iteratively trained on the gradient of the loss function through gradient descent. For training data x_i and labels y_i, a tree ensemble model takes the form where K is the number of trees, f is a function in the functional space F of the set of all possible trees. The prediction scores of each tree are summed so as to reach a final score and the predicted value ($\hat{y}_i$).

$$\hat{y}_i = \sum_{k=1}^{K} f_k\,(x_i)\,,\ \ f_k\ \in F. \tag{8}$$

In order to learn the functions that describe the structure of the tree and leaf scores, an additive strategy is undertaken to sequentially add trees and fix the errors in what has been learnt. The prediction value at step t is given as $\hat{y}_i^{(t)}$, with n number of predictions, then at each stage of training we want to add the tree that optimizes a regularized objective function where l is the loss function between prediction $\hat{y}$ and target y. A regularization term (Ω) is used to prevent overfitting and measures the complexity of the model, consisting of T as the number of leaves in a tree, the vector of leaf scores w, complexity parameter γ and λ as the parameter to scale the penalty.

$$\text{obj}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + \text{constant}, \quad (9)$$

where

$$\Omega\ (f) = \ \gamma T + \ \frac{1}{2}\lambda w^2. \quad (10)$$

Through second order Taylor expansion, where g_i and h_i are the first and second order gradient statistics of the loss functions, respectively, and removing the constants, the specific objective at step t becomes as presented in Equation 11.

$$\text{obj}^{(t)} = \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \Omega(\mathrm{f}_t). \quad (11)$$

The optimal weights and tree structure functions are computed using Equations 12 and 13, respectively:

$$w_j^* = \frac{G_j}{H_j + \lambda}, \quad (12)$$

$$\text{obj}^* = \ -\frac{1}{2}\sum_{j\,=\,1}^{T}\frac{G_j^2}{H_j + \lambda} + \ \gamma T. \quad (13)$$

Finally, each tree is optimized one level at a time, splitting a leaf with the score it gains defined in Equation 14, namely, the score on the left leaf, the score on the new right left, the score on the original leaf, and regularization on the additional leaf. If the gain is smaller than γ, then the branch should not be split. The gain for each feature indicates its relative contribution toward the model making accurate predictions, therefore implying its importance.

$$\text{Gain} \ = \ \frac{1}{2}\left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda}\right] - \gamma \quad (14)$$

Extreme Gradient Boosting (XGBoost) introduces several innovations including parallel learning to improve performance, while being computationally less costly and therefore fast (Chen & Guestrin, 2016). In addition, XGBoost utilizes the techniques of shrinkage and subsampling to prevent overfitting, the former reducing the influence of each tree through training, and the latter training on a random subset of data columns. XGBoost has been shown to have good predictive capabilities in accident prediction, often exceeding the accuracy of other models (Leevy et al., 2018) in both road transportation (Parsa, Movahedi, Taghipour, Derrible, & Mohammadian, 2020; Shi, Wong, Li, Palanisamy, & Chai, 2019; Wang et al., 2019) and maritime risk (Adland et al., 2021; Jin et al., 2019). It has also been shown to be highly efficient and therefore scalable to massive data sets (Leevy et al., 2018), which is a significant advantage when analyzing large maritime traffic data sets.

3.4. Step 4: Results Evaluation and Implementation

Multiple performance measures for machine learning classification algorithms are available. These include model accuracy, recall (ratio of true positives to false negatives and true positives), specificity (ratio of true negatives to true negatives and false positives), precision (ratio of true positives to true positives and false positives), and F_1 score (harmonic mean of precision and recall). In this study, the Area Under Curve (AUC) of the Receiver Operating Characteristics (ROC) Curve is used as the primary performance measure. The ROC curve plots the true positive rate (recall) against the false positive rate (1 – specificity). The resulting score measures the model's ability to separate positive and negative samples. A score of 0.5 indicating random performance and a score of 1.0 indicating perfect performance.

Predicted class probabilities need to be calibrated as upsampling the minority class through SMOTE causes high probabilities for the majority class. This can be corrected by adjusting for the ratio of positive to negative classes in the data (Pozzolo, Caelen, Johnson, & Bontempi, 2015), where p_s is the probability of selecting a positive or negative sample and β is the ratio of positive to negative samples during sampling.

$$\text{Calibrated } P = \frac{\beta p_s}{\beta p_s - p_s + 1}. \quad (15)$$

Following this the probability of accidents in each cell can be displayed, and the relative significance of each feature to accident occurrence considered.

4. UNITED KINGDOM EEZ CASE STUDY

4.1. U.K. Strategic Management of Vessel Safety

The Maritime and Coastguard Agency (MCA) is responsible for managing the safety of shipping in U.K. waters. The MCA must ensure that existing waterways are safe and that any proposed future developments do not compromise the safety of navigation. At a regional or local level, risk assessments are also required by developers for offshore wind farms (MCA, 2021) and oil and gas infrastructure (DECC, 2012). For U.K. ports and harbors, the Port Marine Safety Code (PMSC), states that a risk assessment should "ensure that marine risks are formally assessed and are eliminated or reduced to the lowest possible level" (DfT, 2016, p. 5). Such activities are hampered by the scale of the task, with the United Kingdom responsible for more than 10,000 miles of coastline and two million square miles of sea. As a result, there is no overarching and holistic maritime risk assessment for U.K. waters utilized by the MCA. To address this, the methodological framework developed in Section 3 is evaluated within the context of the U.K.'s EEZ in the following section.

4.2. Data sets

In this case study a spatial model is constructed using DGGS at resolution 11 (Barnes, 2016), which consists of 2,941 hexagonal cells (excluding those entirely on land) each with an area of 290 km^2 and a diameter of 22 km. This resolution was found to exhibit a balance between spatial precision and real-world implementation, providing similar cell sizes as used in other studies. However, the approach described above can be conducted at any resolution, utilizing a finer or coarser spatial scale as required.

4.2.1. Vessel Traffic

Anonymized monthly AIS data in the United Kingdom for 2017 is available from the Marine Management Organization (MMO) in ESRI shapefile polylines format (MMO, 2014). These tracks have been grouped into unique daily vessel movements and annualized for each grid cell for each vessel type utilizing a spatial join (see Fig. 2). Within the MMO data set, vessels are classified into 11 type categories based on their AIS message information. Vessels have been recategorized into five principal types based on their size and purpose (Table III). Some vessel types have been overwritten based on their operational activities, for example freight ferries and oil and gas supply vessels are described as cargo vessels but have been assigned passenger and tug and service vessels, respectively.

4.2.2. Incident Data

Incident data were provided by the U.K.'s Marine Accident Investigation Branch (MAIB) under a Freedom of Information Request for the years 2010–2020. The data have been filtered to the U.K. EEZ and filtered to collisions ($n = 1{,}226$) and groundings ($n = 901$), shown in Fig. 3. The same vessel type categorizations shown in Table III have been applied. While every effort has been made to ensure the accuracy of the underlying accident database, several examples were found of erroneous locations assigned to accidents. Where these have been identified the incident attributes were corrected manually, however, we must accept some degree of uncertainty in the quality of the underlying accident data sets which have been commented on previously (Hassel et al., 2011; Qu et al., 2012).

4.2.3. Other Data Sets and Feature Engineering

Based on the risk factors identified in Table II, the following model features were developed (Figs. 4 and 5):

(1) *Annual Vessel Movements for Vessel Class I* (n): as described in Section 4.2.1.
(2) *Annual Vessel Movements for all other Vessel Classes* (n): collision incidents would include collisions between the target vessel type and other vessel types.
(3) *Average Depth of grid cell* (meters): developed as part of the United Kingdom's renewable energy atlas (https://www.renewables-atlas.info/) and indicates increased probability of encountering shallow waters.

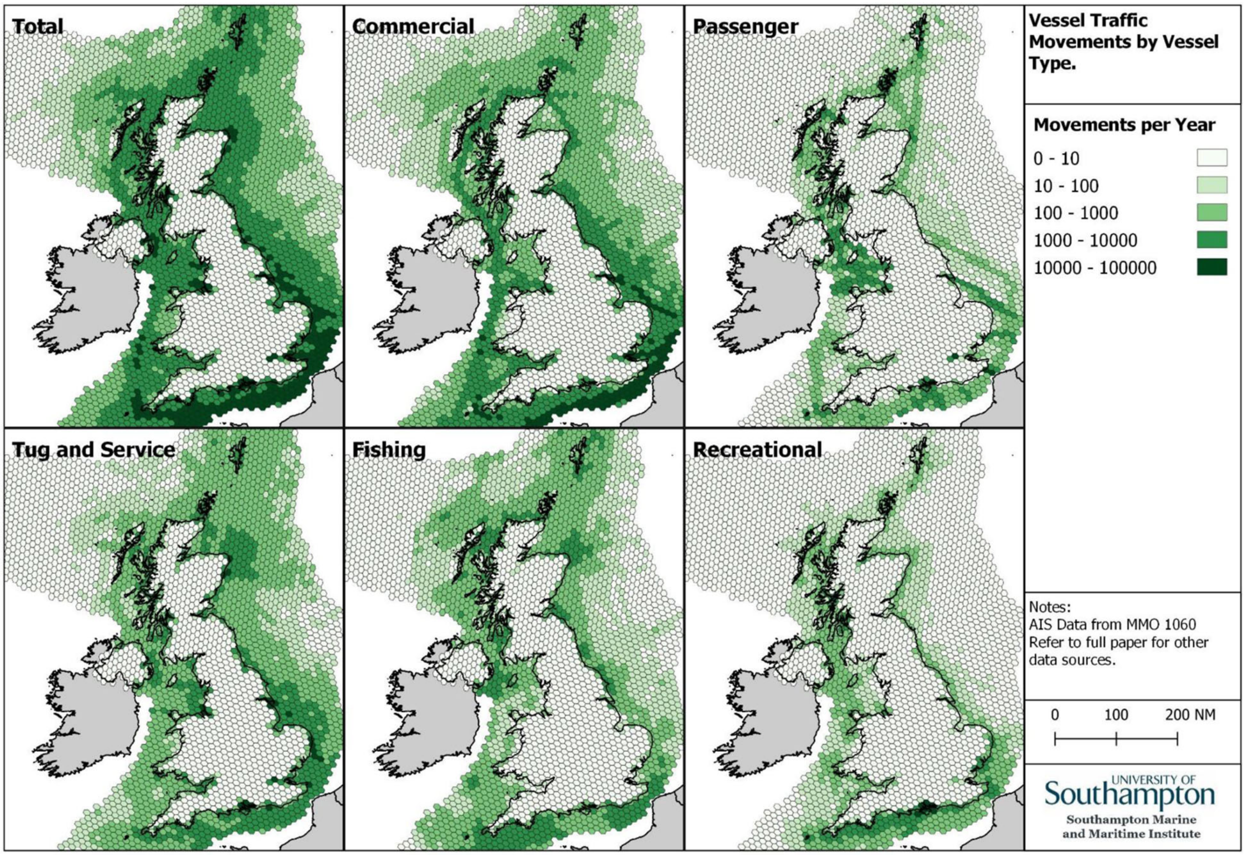

Fig 2. Vessel traffic movements by vessel type

Table III. Vessel Categories

Vessel Category	MMO Vessel Types
Commercial (Dry or liquid bulk)	Cargo (*), Tankers
Passenger (Cruise Ships and Ferries)	Passenger Vessels, high speed craft (*).
Fishing	Fishing
Recreational (Leisure)	Sailing and Pleasure Craft
Tug and Service Craft	Cargo (*), Unknown, Port Service Craft, Vessels engaged in dredging or underwater operations, high speed craft (*), military or law enforcement.

Note: * indicates some vessels split into different categories.

(4) *Mean Wind Speed* (m/s) and *Mean Significant Wave Height* (meters): obtained from the EU Copernicus earth observation system (CERSAT-GLO-REP_WIND_L4-OBS and GLOBAL-REANALYSIS-WAV-001-032). Data consist of a NetCDF format grid that contained hourly and monthly values for 2017, which are aggregated. Feature indicates presence of metocean factors which might compromise ship handling performance.

(5) *Mean Spring Tidal Current* (knots): developed as part of the United Kingdom's renewable energy atlas (https://www.renewables-atlas.info/) and indicates presence of hydrodynamic factors which might increase navigational complexity.

(6) *Distance from Shore* (km) and *Inland Waterways* (binary): calculated for each DGGS cell using a spatial query from the GADM world landmass shapefile under free academic

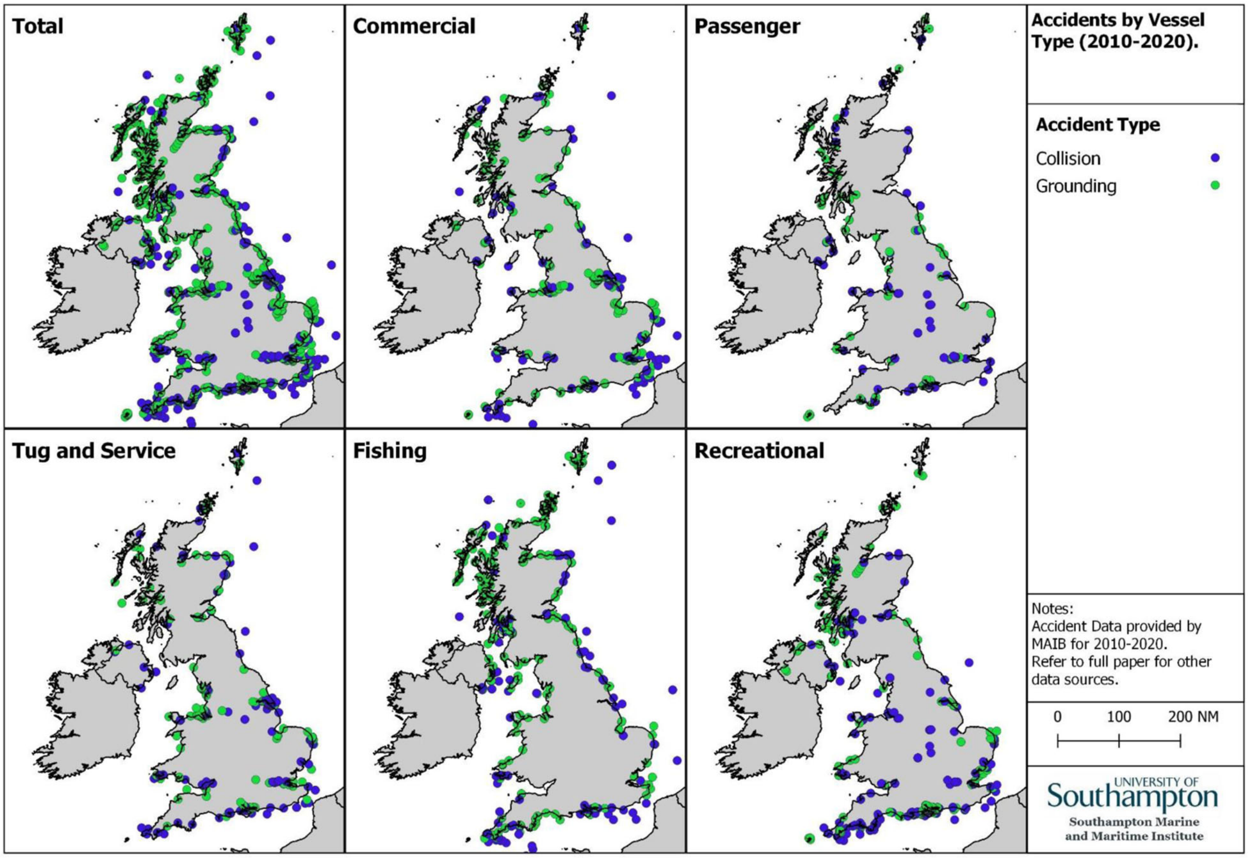

Fig 3. Historical Accidents (2010-2020)

license (https://gadm.org/download_world.html). Feature indicates relative navigational complexity of a waterway and distance to shore.

(7) *Average vessel density* (0 –1): the proportion of navigable waters in a cell which has high density traffic. A 1,000 m high resolution grid was generated of the entire study area, and the total number of vessel transits calculated per grid cell. Where the number of transits was greater than 100/year, this cell was classed as high density (N_{100}). The average density of each DGGS cell was then calculated and compared against the percentage of that cell which is not-land (P_{Area}). Feature indicates presence of major shipping routes or concentrated traffic flows, which might increase the likelihood of vessel interactions.

$$\text{Density} = \frac{N_{100}}{P_{\text{Area}}} \tag{16}$$

(8) *Presence or absence of major ports* (binary): The Department for Transport publishes annual statistics for what are classed "major" or "minor" ports in the United Kingdom (DfT, 2020). All 53 major ports were located as a five nautical mile circular buffer, with the presence or absence of a port within a DGGS cell taken as a feature. Feature indicates locations of increased navigational complexity and vessel activity.

(9) *Presence or absence of major shipping routes* (binary): Feature indicates the presence of IMO mandated traffic schemes as identified from the relevant nautical publications.

4.3. Model Development

Given the identified model features, the processing steps identified in Section 3 are undertaken. First, the data set is split into a training and testing data set with the ratio of 70:30. Second, each spatial unit

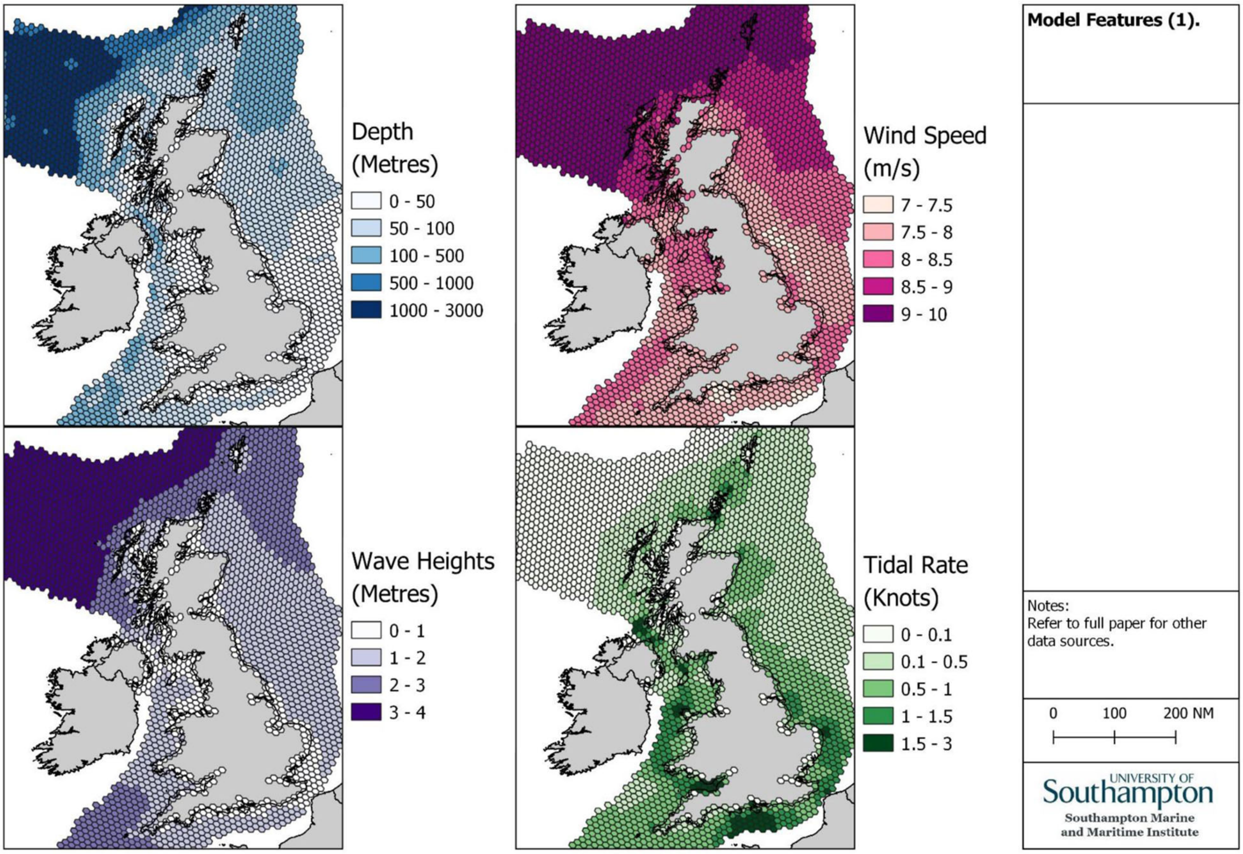

Fig 4. Model features (1).

was resampled using expected annual and monthly accident frequencies to form a classification problem. For example, if during the 10 years of accident data, one incident occurred in a grid cell, this could be expressed as one positive sample and nine negative samples. It was necessary to undertake this resampling after splitting into train and test data sets in order to prevent identical samples appearing in both data sets. Third, the data set was split into 12 different implementations, six collision, and six grounding implementations for each of the five vessel categories and a combined total. Fourth, positive and negative samples in the training data set were balanced using SMOTE.

Each of the four algorithms described in Section 3.3 are implemented using the Scikit-Learn and XGBoost python libraries. Each model was trained independently to reflect the different factors that might influence a commercial vessel as opposed to a fishing vessel, and a collision as opposed to a grounding. Hyperparameter tuning was conducted using a parameter grid with randomized search using 25 iterations with five-fold cross validation used to determine the optimal parameters. Each model was assessed to maximize the AUC of the ROC curve. The best model in each case was then used on the set aside test set to evaluate their effectiveness. The models were developed using aggregated annual data sets and monthly data sets, evaluate the significance of temporal factors.

4.4. Results

4.4.1. Algorithm Performance

Fig. 6 compares the accuracies achieved by each of the machine learning algorithms using the same features and general methodology. Of the 12 models, XGBoost exceeded Logistic Regression and a Linear SVM in all 12, and Random Forest in 75% of cases. This supports the findings of other studies that XGBoost is a powerful multi-application machine learning model that achieves high performance

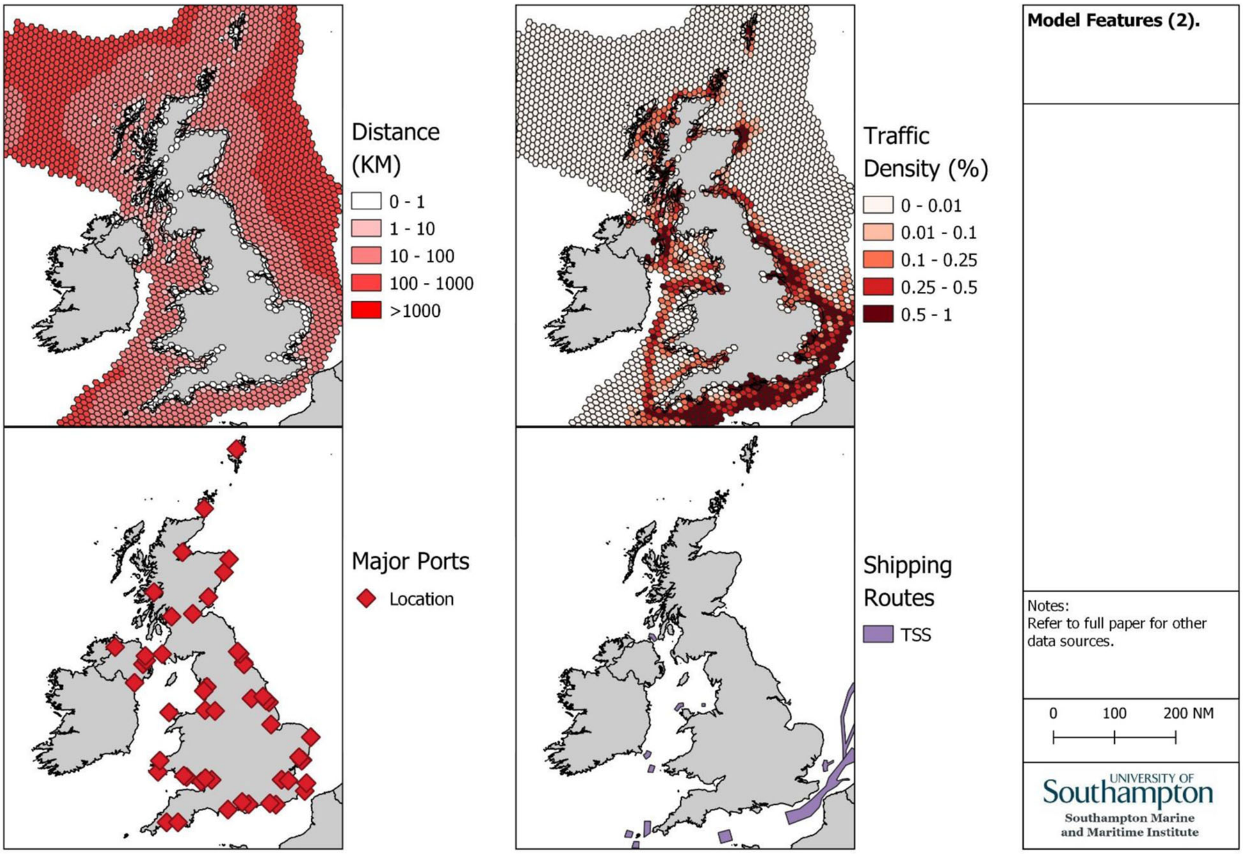

Fig 5. Model features (2).

(Adland et al., 2021; Jin et al., 2019; Wang, Deng, & Wang, 2020). Some have argued that logistic regression models will rarely perform as well given their inherent statistical assumptions and static relationships between variables, but their greater transparency can be an advantage (Adland et al., 2021). Linear SVM's performed marginally better than Logistic Regression but did not achieve the performance of the tree-based algorithms. Random Forest by contrast was consistently marginally weaker than XGBoost, however, achieved the highest accuracies in the case of collision risk for commercial vessels, recreational, and all vessel types combined.

Fig. 7 further differentiates the XGBoost performance scores utilizing the AUC ROC value for accident type, vessel type and temporal units. All 12 models achieved AUC ROC scores in excess of 0.87, indicating both an overall strong performance and which particular model configurations perform more or less well. First, the model performance is on average higher for ship groundings as opposed to ship collisions. This suggests that spatial factors are more important for ship groundings, which given the necessity for shallow water is to be expected. While ship collisions often occur in areas of shallow water, where ship navigation is more constrained and they navigate closer together, they can also occur offshore in deep water.

Second, model performance varied significantly by vessel type. For example, fishing vessels obtained the lowest scores and may be due to underrepresentation of their movements as they are not required to carry AIS. While this is also true for recreational craft, their activities are much more concentrated in specific regions such as the Solent and are therefore more predictable.

Third, in most cases, aggregating the data into annualized Figs. reduced performance as opposed to using monthly figures (Fig. 7). Monthly figures allow for temporal variations in vessel activity and weather conditions to be captured, which would otherwise be lost. Furthermore, providing monthly fig-

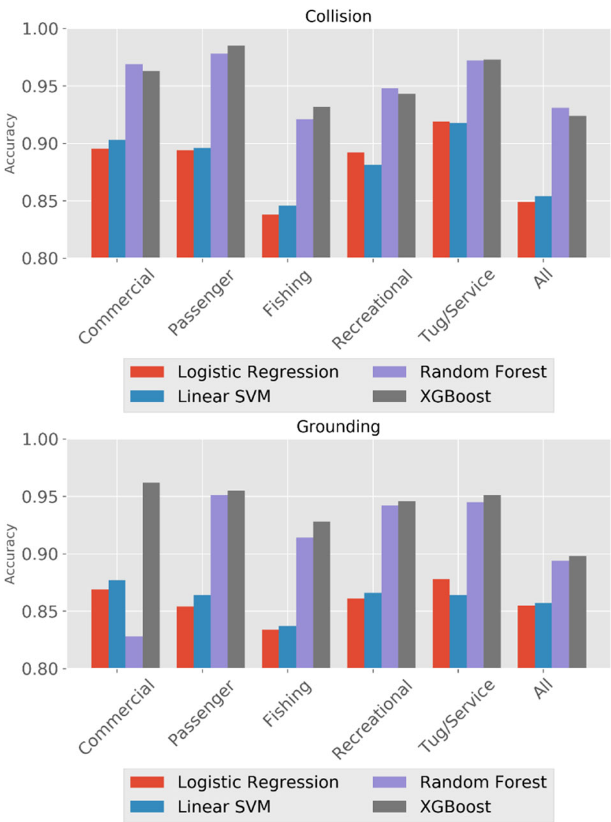

Fig 6. Comparison of algorithm performance.

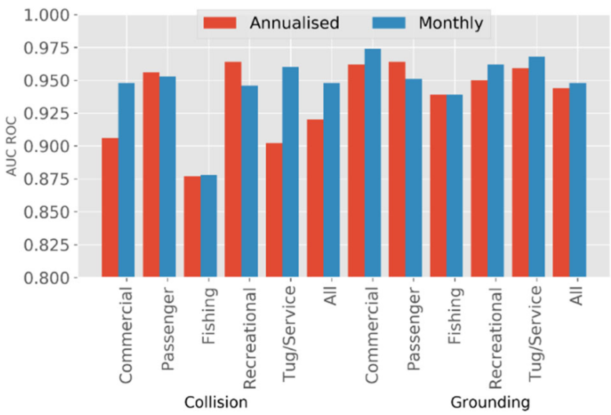

Fig 7. XGBoost performance between annualized and monthly data sets.

ures increases the sample size and training data by 12, likely improving performance. A notable exception is recreational craft collision, which while highly seasonal, did not achieve as high performance as annualized figures.

4.4.2. *Implementation*

Given the consistently superior performance of XGBoost, the trained models utilizing this algorithm have been implemented in order to generate risk maps for collision and grounding of different vessel types. The annualized collision and grounding frequencies for each vessel type are shown in Figs. 8 and 9, respectively. Several observations can be made. First, the accident data presented in Fig. 3 show that accident distribution is not uniform with a small sample size, and therefore there are large regions where no accidents have occurred. By contrast, the trained models have both correctly learnt which locations have frequent accidents and identified regions where it predicts could have accidents in the future, given the presence of certain risk factors.

Second, for some vessel types the risk is highly concentrated in specific waterways. For example, passenger vessel collisions and groundings are shown to be most likely in ferry ports such as Southampton/Portsmouth, Liverpool, Belfast, and inland waterways such as the Thames. Tug and service vessels are highly concentrated in ports and harbors as they include tugs and pilot vessels, but the risk of collision involving North Sea oil and gas fields is evident in Fig. 8. The risks associated with recreational craft are concentrated inshore near popular cruising destinations such as the Solent.

Third, the distribution of risk for groundings is much more concentrated inshore than for collisions, given the significance of water depth in hazard occurrence. However, the model has identified some waterways further offshore, such as within the Thames Estuary and Dover Straits which contain numerous sand banks where historical ship groundings have occurred.

4.4.3. *Variable Importance*

Table IV compares the XGBoost feature importance rankings for collision and grounding between the vessel types. First, the overwhelmingly most important feature is distance from shore, and it can be seen in Fig. 3 that the majority of accidents are coastal. While other types of maritime accident do occur offshore, such as sinkings and fires, collisions and groundings are principally near shore hazards. As depths of water are averaged per cell, some shallow water shoals may be underrepresented contributing a minor influence on the model.

Second, the frequency of vessel movements is a key contribution to accident frequency. The

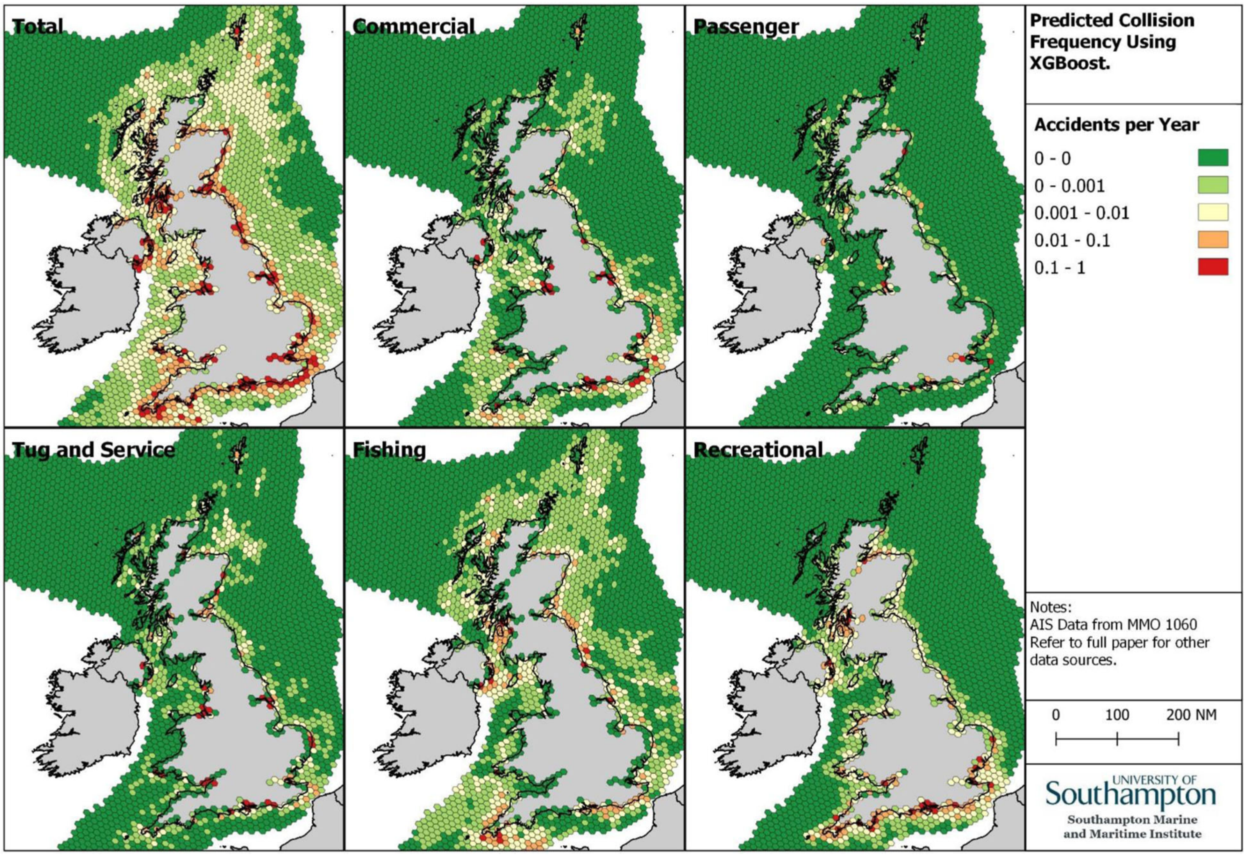

Fig 8. Predicted collision frequency using XGBoost.

statistical relationship between accidents and vessel movements have been challenged by some (Mazaheri et al., 2014) but supported by others to varying degrees of significance (Bye & Aalberg, 2018; Rawson & Brito, 2021). However, the analysis shows that most accidents occur in the busiest locations such as ports. In addition, the density of traffic increases the proximity between vessels and has an influence on collision risk for some vessel types. It is interesting that the presence of noncommercial traffic is a more significant contributor to collision risk for commercial and passenger vessels, suggesting that these vessel types are more likely to collide with other vessel types than similar vessels.

Finally, the influence of metocean conditions such as wind, wave, and tide are among the least important features. This is contrary to the findings of other works that metocean features are critical for predicting certain accident types such as insurance claims (Adland et al., 2021), hurricane impacts (Rawson et al., 2021), and fishing vessel casualties (Rezaee et al., 2016). There are two likely reasons for this difference. First, the aforementioned works include a variety of other hazards such as capsize or cargo damage which are much more related to weather conditions than collisions or groundings, which are far more commonly associated with human error. Second, by aggregating the metocean conditions into monthly and annual averages, the most exposed locations are typically furthest from shore where accidents are much less likely.

5. DISCUSSION AND FUTURE RESEARCH DIRECTIONS

The case study demonstrates that a machine learning approach to map navigation safety is a powerful and effective tool for strategic maritime risk assessment. In particular, this approach has several attractive qualities over conventional methods, which make them well suited for practical implementation as a decision support tool for navigational

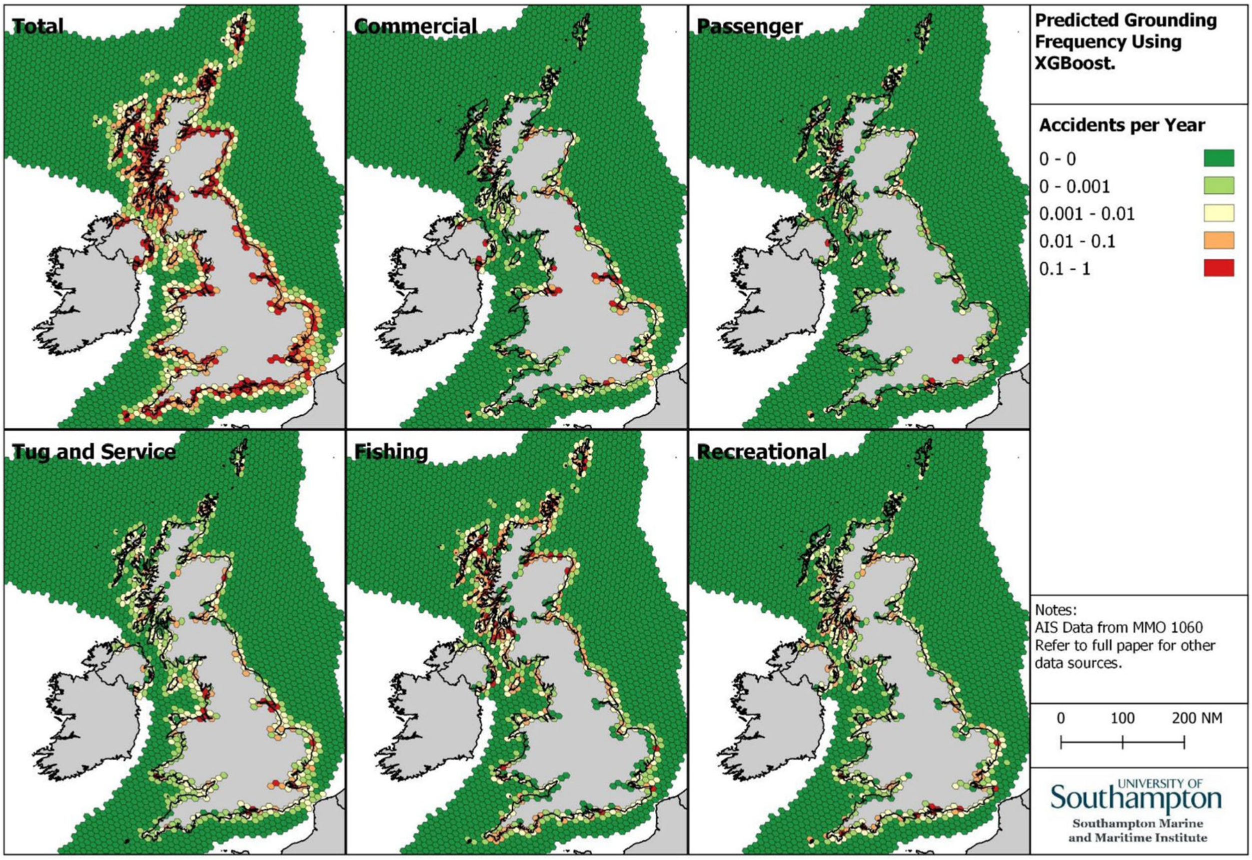

Fig 9. Predicted grounding frequency using XGBoost.

authorities. First, the proposed model provides a high resolution, standardized, and visual tool for decision-makers to effectively plan the safety of navigation between different waterways. Coastal states have an obligation to assess the degree of risk within their waterways and determine the requirement for risk control measures (IMO, 2004). These measures can be expensive, such as Emergency Towage Vessels (ETVs) which cost tens of millions of pounds (Transport Committee, 2011) and require both justification and allocation to the regions where they would be most effective. Furthermore, marine spatial planning of new developments such as offshore wind farms to deconflict with maritime risk can be more effectively undertaken. A national, data-driven risk map can support these activities, providing an important visual appreciation of the spatial distribution of risk (Hoorn & Knapp, 2015). Second, the model can be quickly and cost effectively updated with new traffic conditions at regular intervals without the need to commission new studies, overcoming a major limitation of conventional models (EMSA, 2018). Furthermore, the model can be tested with future case scenarios by adding additional shipping routes or mitigation measures and evaluating the impact on risk.

Whist this case study has shown significant promise in achieving these goals, several limitations and areas of further work remain. First, there are challenges associated with the availability and representativeness of the training data (Guikema, 2020). For example, accidents are underreported (Hassel et al., 2011; Qu et al., 2012) and not all risk factors identified as important (Table II) can be easily quantified into aggregated spatial models. Furthermore, a machine learning approach cannot predict accidents for which there are no previous examples. For example, allisions between commercial ships and offshore wind farms is a credible hazard (MCA, 2021) but is absent from the training data.

Second, only a single spatial scale of assessment has been utilized for model development. Aggregating data into spatial units improves the scalabil-

Table IV. Feature Ranking Per XGBoost Model (Top 3 Highlighted)

Variable	Collision						Grounding					
	Commercial	Passenger	Fishing	Recreational	Tug and Service	All	Commercial	Passenger	Fishing	Recreational	Tug and Service	All
Target Traffic	10	7	2	2	1	2	5	2	2	2	6	2
Non-Target Traffic	3	2	11	7	5	N/A	9	9	7	7	8	N/A
Distance	5	1	5	1	2	1	1	1	1	1	1	1
Depth	4	3	3	4	4	5	3	4	11	3	2	5
Density%	1	8	1	6	6	4	6	5	5	9	4	4
Port	2	9	10	5	11	6	2	8	8	6	3	6
TSS	6	11	8	9	7	10	10	7	3	5	9	10
Inland	7	4	4	11	3	3	11	6	6	11	5	3
Wind	8	5	7	3	9	8	7	10	10	4	11	8
Wave	11	6	9	10	10	9	4	3	4	8	7	9
Tidal	9	10	6	8	8	7	8	11	9	10	10	7

ity of subsequent analysis but can lead to statistical challenges associated with the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1984; Rawson et al., 2019). While a spatial data structure such as a DGGS inherently supports multiresolution analysis, the influence of changing resolutions on machine learning model performance requires further exploration. Furthermore, it is likely that many features exhibit positive spatial autocorrelation whereby risk is clustered into specific waterways. To support model development and integration of massive data sets, this work has aggregated data into discrete spatial units without regard for the relationships between neighboring cells. Further work could be undertaken to better understand these spatial and statistical relationships.

Finally, while a strategic maritime risk map is valuable to decisionmakers, a tactical decision support tool could be envisaged which identifies risk in real-time to enable coastguards to intervene to prevent an accident. In such a case, the model would be trained on individual ship positions or transits and the specific conditions that the vessel is exposed to (Adland et al., 2021; Rawson et al., 2021). Furthermore, the characteristics of each vessel, such as age, flag state, or size, could be included as new features in the model development (Jin et al., 2019). However, given the requirement for massive vessel traffic, metocean, and other data sets to achieve this, a big data processing solution would be required (Abualhaol, Falcon, Abielmona, & Petriu, 2018; Lensu & Goerlandt, 2019). While some work has considered the application of architectures such as Apache Spark or Hadoop for maritime risk analysis (Chatzikokolakis, Zissis, Vodas, Spiliopoulos, & Kotopoulos, 2019; Filipiak et al., 2018; Zhang, Meng, & Fwa, 2019), further research is required to integrate machine learning processes with big data infrastructure for maritime risk assessment.

6. CONCLUSIONS

Spatial maritime risk assessments can support decisionmakers by comparing and monitoring the risk profile between different waterways and enabling the targeted deployment of risk mitigation. However, previous risk models have been criticized due to their significant cost, methodological assumptions, and limited scale. Within this article, a methodological framework utilizing machine learning models to map the risk of collisions and groundings across the United Kingdom has been presented and tested. The results demonstrate that a data processing pipeline using a DGGS, combined with a machine learning algorithm can predict which areas of the United Kingdom are likely to experience maritime accidents to a high degree of accuracy. Furthermore, the model is able to assess the relative risk of regions where no historical accidents have previously occurred, overcoming challenges with conventional quantitative models.

This work demonstrates how heterogenous spatial data sets such as bathymetry, vessel traffic, metocean, and accidents can be effectively integrated into a common spatial model, overcoming the static nature of previous works (Jin et al., 2019). While many of these data sets are aggregated, by differentiating between monthly and annualized samples, the model performance is shown to be more effective with a finer temporal resolution. Furthermore, the use of XGBoost is shown to have consistently greater predictive performance than alternative machine learning algorithms. From these outputs, the contribution of different risk factors is discerned, providing additional transparency as to how the model is developed. Finally, the practical and operational benefits to coastal states and navigational authorities of implementing such a strategic risk tool are described. High resolution, visual and automated risk assessment tools can support decisionmakers with better managing the safety of navigation and reducing the risk of loss of life and pollution at sea.

ACKNOWLEDGMENTS

This work is partly funded by the University of Southampton's Marine and Maritime Institute (SMMI) and the European Commission, under Horizon 2020, Research and Innovation project SEDNA, Grant agreement number: 723526.

References

Abualhaol, I. Y., Falcon, R., Abielmona, R. S., & Petriu, E. M.(2018). Mining port congestion indicators from big AIS data. In *Proceedings of the International Joint Conference on Neural Networks* (pp. 1–8). Rio de Janeiro, Brazil: IEEE.

Adland, R., Jia, H., Lode, T., & Skontorp, J. (2021). The value of meteorological data in marine risk assessment. *Reliability Engineering and System Safety*, *209*, 107480.

Agarwal, P., Tang, J., Narayanan, A., & Zhuang, J. (2020). Big data and predictive analytics in fire risk using weather data. *Risk Analysis*, *40*(7), 1438–1449.

Aven, T., & Zio, E. (2011). Some considerations on the treatment of uncertainties in risk assessment for practical decision making. *Reliability Engineering and System Safety*, *96*(1), 64–74.

Barnes, R. (2016). dggridR: Discrete global grid systems for R. Retrieved from https://github.com/r-barnes/dggridR

Barnes, R. (2019). Optimal orientations of discrete global grids and the poles of inaccessibility. *International Journal of Digital Earth*, *13*(7), 803–816.

Battersby, S., Stebe, D., & Finn, M. (2016). Shapes on a Plane: Evaluating the impact of projection distortion on spatial binning. *Cartography and Geographic Information Science*, *44*(5), 410–421.

BE-AWARE. (2014). *BE-AWARE Summary Report*. Retrieved from: https://ec.europa.eu/echo/sites/default/files/be_aware_final_report.pdf.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Brito, M., Smeed, D., & Griffiths, G. (2014). Analysis of causation of loss of communication with marine autonomous systems: A probability tree approach. *Methods in Oceanography*, *10*, 122–137.

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.

Bye, R., & Almklov, P. (2019). Normalization of maritime accident data using AIS. *Marine Policy*, *109*, 103675.

Chatzikokolakis, K., Zissis, D., Vodas, M., Spiliopoulos, G., & Kotopoulos, I. (2019). *A distributed lighting fast maritime anomaly detection service*. Proceedings of Oceans 2019, June 17–20, Marseille, France.

Chawla, N., Bowyer, K., Hall, L., & Keglemeyer, W. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research 16*, 321–357.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of KDD* (pp. 785–794). San Francisco, CA: ACM.

Choubin, B., Borji, M., Mosavi, A., Sahedi-Hosseini, F., Singh, V., & Shamshirband, S. (2019). Snow avalanche hazard prediction using machine learning methods. *Journal of Hydrology*, *577*, 123929.

Choubin, B., Abdolshahnejad, M., Moradi, E., Querol, X., Mosavi, A., Shamshirband, S., & Ghamisi, P. (2020). Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. *Science of the Total Environment*, *701*, 134474.

Cox, L.A. (2008). What's wrong with risk matrices. *Risk Analysis*, *28*(2), 497–512.

DECC. (2012). *Consultation on guidance relating to the consent to locate process under Part 4a of the Energy Act 2008*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/66572/6655-cons-guidance-relating-consent-locate-pro.pdf

DfT. (2016). *Port marine safety code*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/564723/port-marine-safety-code.pdf

DfT. (2020). *Port and domestic waterborne freight statistics: Data tables (PORT)*. Retrieved from https://www.gov.uk/government/statistical-data-sets/port-and-domestic-waterborne-freight-statistics-port

DNV. (2012). *Prince Rupert marine risk assessment*. (Report No./DNV Reg No. /13JIMVK-8 Rev 3). Bærum, Norway: DNV.

DNV. (2013). *North East shipping risk assessment*. (Report No./DNV Reg No. 0/14OFICX-4 Rev 1).Bærum, Norway: DNV.

DNV. (2015). *Marine environmental risk assessment – Concerning waters in and adjacent to Greenland the Arctic*. (Report No. 2014-0951 Rev E). Bærum, Norway: DNV.

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*, 106933.

EMSA. (2018). *Joint workshop on risk assessment and response planning in Europe*. 14 April 2018, London.

Fang, A., Qiu, C., Zhao, L., & Jin, Y. (2018). Driver risk assessment using traffic violation and accident data by machine learning approaches. In *Proceedings of 3rd international conference on intelligent transportation engineering* (pp. 291–295). Singapore: IEEE.

Filipiak, D., Strozyna, M., Wecel, K., & Abramowicz, W. (2018). *Anomaly detection in the maritime domain: comparison of traditional and big data approach*. Proceedings of NATO IST-160-RSM specialists' meeting on big data and artificial intelligence for military decision making. Bordeaux, France.

Fowler, T., & Sorgard, E. (2000). Modeling ship transportation risk. *Risk Analysis*, *20*(2), 225–244.

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232.

Genivar. (2013). *Risk assessment for marine spills in canadian waters: Phase 1 oil spills south of 60th parallel*. Retrieved from https://www.wcel.org/sites/default/files/file-downloads/131-17593-00_ERA_Oil-Spill-South_150116_pp1-124.pdf

Guikema, S. (2020). Artificial intelligence for natural hazards risk analysis: Potential, challenges, and research needs. *Risk Analysis*, *40*, 1117–1123.

Halpern, B., Walbridge, S., & Selkoe, K. (2008). A global map of human impact on marine ecosystems. *Science*, *319*, 948–952.

Hanninen, M. (2014). Bayesian networks for maritime traffic accident prevention: Benefits and challenges. *Accident Analysis and Prevention*, *73*, 305–312.

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.

Hassel, M., Asbjornslett, B., & Hole, L. (2011). Underreporting of maritime accidents to vessel accident databases. *Accident Analysis and Prevention*, *43*, 2053–2063.

Hedge, J., & Rokseth, B. (2020). Applications of machine learning methods for engineering risk assessment – A review. *Safety Science*, *122*, 104492.

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.

Hoorn, S., & Knapp, S. (2015). A multi-layered risk exposure assessment approach for the shipping industry. *Transportation Research Part A*, *78*, 21–33.

Hoorn, S., & Knapp, S. (2019). Predicting traffic and risk exposure in the maritime industry. *Safety*, *5*(3), 42.

Hubbard, D. W. (2009). *The failure of risk management: Why it's broken and how to fix it*. Hoboken, NJ: Wiley.

IMO. (2004). *International convention for the safety of life at sea*. London, UK: IMO.

IMO. (2018). *Revised guidelines for formal safety assessment (FSA) for use in the IMO rule-making process*. Retrieved from MSC-MEPC.2/Circ.12/Rev.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. *Marine Policy*, *108*, 103655.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge UK.

Kecman, V. (2005). Support vector machines– An introduction. In L. Wang (Ed.), *Support vector machines: Theory and applications* (pp. 1–48). New York: Springer.

Kite-Powell, H., Jin, D., Jebsen, J., Papkonstantinou, V., & Patrikalakis, N. (1999). Investigation of potential risk factors for groundings of commercial vessels in U.S. ports. *International Journal of Offshore and Polar Engineering*, *9*(1), 16–21.

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.

Kontovas, C., & Psaraftis, H. (2009). Formal safety assessment A critical review. *Marine Technology*, *46*(1), 45–59.

Kristiansen, S. (2005). *Maritime transportation safety management and risk analysis*. London, UK: Routledge.

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*, 104798.

Lee, S., Hong, S., & Jung, H. (2017). A support vector machine for landslide susceptibility mapping in Gangwon province, Korea. *Sustainability*, *9*(1), 48.

Lee, J., Yoon, T., Kwon, S., & Lee, J. (2019). Model evaluation for forecasting traffic accident severity in rainy seasons using machine learning algorithms: Seoul city study. *Applied Sciences*, *10*, 129.

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.

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. Retrieved from: https://doi.org/10.1016/j.marpol.2019.02.038.

Li, S., Meng, Q., & Qu, X. (2012). An overview of maritime waterway quantitative risk assessment models. *Risk Analysis*, *32*(3), 496–512.

Li, Z., Liu, P., Wang, W., & Xu, C. (2012). Using Support vector machine models for crash injury severity analysis. *Accident Analysis and Prevention*, *45*, 478–486.

Li, X., Yan, D., Wang, K., Weng, B., Qin, T., & Liu, S. (2019). Flood risk assessment of global watersheds based on multiple machine learning models. *Water*, *11*(8), 1654.

Marico. (2015). *LINZ Hydrography risk assessment methodology update*. Retrieved from https://iho.int/uploads/user/Inter-Regional%20Coordination/WEND-WG/WENDWG%20Repository/SWPRHP-Hydrography-Risk-Assessment-Methodology.pdf

Mazaheri, A., & Ylitalo, J. (2010). Comments on geometrical modelling of ship grounding. Proceedings of 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. *Journal of Navigation*, *68*(1), 89–106.

Mazaheri, A., Montewka, J., & Kujala, P. (2016). Towards an evidence-based probabilistic risk model for ship-grounding accidents. *Safety Science*, *86*, 195–210.

MCA. (2021). *Safety of navigation: Offshore renewable energy installations (OREIs) - Guidance on UK navigational practice, safety and emergency response*. MGN 654. Retrieved from https://www.gov.uk/guidance/offshore-renewable-energy-installations-impact-on-shipping

McClendon, L., & Meghanathan, N. (2015). Using machine learning algorithms to analyze crime data. *Machine Learning and Applications: An International Journal*, *2*(1), 1–12.

Merrick, J., & van Dorp, R. (2006). Speaking the truth in maritime risk assessment. *Risk Analysis*, *26*(1), 223–237.

MMO. (2014). *Mapping UK shipping density and routes technical annex: MMO 1066*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/317771/1066-annex.pdf

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.

Moosavi, S., Samavaian, M., Parthasarathy, S., Teodorscu, R., & Ramnath, R. (2019). Accident risk prediction based on heterogeneous sparse data: New dataset and insights. In F. Banaei-

Kashani, G. Trajcevski, R. H. Güting, L. Kulik, S. Newsam (Eds.),*Proceedings of SIGSPATIAL 19* (pp. 33–42). Chicago, USA: ACM.
Montewka, J., Goerlandt, J., & Kujala, P. (2014). On a systematic perspective on risk for formal safety assessment (FSA). *Reliability Engineering and System Safety*, *127*, 77–85.
Montewka, J., Goerlandt, J., & Kujala, P. (2015). Towards probabilistic models for the prediction of a ship performance in dynamic ice. *Cold Regions Science and Technology*, *112*, 14–28.
Nateghi, R., & Aven, T. (2021). Risk analysis in the age of big data: Promises and pitfalls. *Risk Analysis*, https://doi.org/10.1111/risa.13682
Nguyen, N., Dang, B., Pham, X., Nguyen, H., Bui, H., Hoangm, N., & Bui, D. (2018). Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: A comparative study. *Ecological Informatics*, *46*, 74–85.
OGC. (2019). *Discrete global grid systems SWG*. Retrieved from https://www.opengeospatial.org/projects/groups/dggsswg
Olba, X., Daamen, W., Vellinga, T., & Hoogendoorn, S. (2019). Risk assessment methodology for vessel traffic in ports by defining the nautical port risk index. *Journal of Marine Science and Engineering*, *8*(1), 10.
OpenRisk. (2018). *OpenRisk guideline for regional risk management to improve european pollution preparedness and response at sea*. Retrieved from https://helcom.fi/media/publications/OpenRisk-Guideline-for-pollution-response-at-sea.pdf
Openshaw, S. (1984). *The modifiable areal unit problem. CATMOG 38*. Norwich, UK: Geobooks.
Pan, G., Fu, L., & Thakali, L. (2017). Development of global road safety performance function using deep neural networks. *International Journal of Transportation Science and Technology*, *6*(3), 159–173.
Parsa, A., Movahedi, A., Taghipour, H., Derrible, S., & Mohammadian, A. (2020). Toward safer highways. Application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accident Analysis and Prevention*, *136*, 105405.
Pedersen, P. T. (1995). Collision and grounding mechanics. In *Proceedings of WEMT 95* (pp. 125–157). Copenhagen, Denmark: The Danish Society of Naval Architecture and Marine Engineering.
Pietrzykowski, Z., & Uriasz, J. (2009). The ship domain – A criterion of navigational safety assessment in an open sea area. *Journal of Navigation*, *62*, 93–108.
Port of Dover. (2016). *Navigation risk assessment*. Retrieved from https://www.doverport.co.uk/administrator/tinymce/source/PDF/Port%20of%20Dover%20-%20Navigational%20Risk%20Assessment.pdf
Pozzolo, A., Caelen, O., Johnson, R., & Bontempi, G. (2015). Calibrating probability with undersampling for unbalanced classification. In *Proceedings of IEEE Symposium Series on Computational Intelligence* (pp. 159–166.) Cape Town, South Africa: IEEE.
Psaraftis, H.N. (2012). Formal safety assessment: An updated review. *Journal of Marine Science and Technology*, *11*, (3), 390–402.
Qu, X., Meng, Q., & Li, S. (2012). Analyses and implications of accidents in Singapore strait. *Journal of Transportation Research Board*, *2273*, 106–111.
Rawson, A., Sabeur, Z., & Correndo, G. (2019). Spatial challenges of maritime risk analysis using big data. In C. G. Soares (Ed.), *Proceedings of the 8th International Conference on Collision and Grounding of Ships and Offshore Structures (ICCGS 2019)* (pp. 275–283). Boca Raton, FL: CRC Press/Balkema.
Rawson, A., & Brito, M. (2021). A critique of the use of domain analysis for spatial collision risk assessment. *Ocean Engineering*, *219*, 108259.
Rawson, A., Brito, M., Sabeur, Z., & Tran-Thanh, L. (2021). A machine learning approach for monitoring ship safety in extreme weather events. *Safety Science*, *141*, 105336.
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.
Rodrigues, M., & Riva, J. (2014). An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling and Software*, *57*, 192–201.
RPSasa, Environmental Research Consulting, RPI and the Louis Berger Group. (2014). *Assessment of marine oil spill risk and environmental vulnerability for the state of Alaska*. NOAA Contract Number WC133F-11-CQ-0002.
Safetec. (1999). *Identification of marine environmental high risk areas (MEHRAs) in the UK*. Doc No. ST-8639-MI-1-Rev 01.
Shi, X., Wong, Y., Li, M., Palanisamy, C., & Chai, C. (2019). A feature learning approach based on XGBoost for driving assessment for risk prediction. *Accident Analysis and Prevention*, *129*, 170–179.
Tetri, G., Ruin, I., Gourley, J., Kirstetter, P., Flamig, Z., Blanchet, J., … Anquetin, S. (2019). Toward probabilistic prediction of flash flood human impacts. *Risk Analysis*, *39*(1), 140–161.
Transport Committee. (2011). *The Coastguard, emergency towing vessels and the maritime incident response group. Sixth Report of Session 2010–2012*. House of Commons, HC 948.
USCG. (2010). *Ports and waterways safety assessment workshop guide*. Retrieved from https://www.navcen.uscg.gov/?pageName=pawsaGuide
Van Dorp, J., Harrald, J., & Merrick, J. Grabowski. (2008). *Assessment of oil spill risk due to potential vessel traffic at Cherry Point, Washington*. Washington, DC: George Washington University.
Van Dorp, J., & Merrick, J. (2014). *VTRA 2010 final report*. Washington, DC: George Washington University.
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.
Wang, N. (2013). A novel analytical framework for dynamic quaternion ship domains. *Journal of Navigation*, *66*, 265–281.
Wang, C., Liu, L., Xu, C., & Lv, W. (2019). Predicting future driving risk of crash-involved drivers based on a systematic machine learning framework. *International Journal of Environmental Research and Public Health*, *16*, 334.
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.
Weng, J., Yang, D., Chai, T., & Fu, S. (2019). Investigation of occurrence likelihood of human errors in shipping operations. *Ocean Engineering*, *182*, 28–37.
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.
Wu, L., Xu, Y., Wang, Q., Wang, F., & Xu, Z. (2017). Mapping global shipping density from AIS data. *Journal of Navigation*, *70*(1), 67–81.
Yuan, Z., Zhou, X., Yang, T., Tamerius, J., & Mantilla, R. (2017). Predicting traffic accidents through heterogeneous urban data: A case study. In *Proceedings of 23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Halifax, Canada*. New York: Association for Computing Machinery.
Zhang, L., Meng, Q., & Fwa, T. (2019). Big AIS data based spatial-temporal analyses of ship traffic in Singapore port waters. *Transportation Research Part E*, *129*, 287–304.
Zhang, X., & Mahadevan, S. (2019). Ensemble machine learning models for aviation incident risk prediction. *Decision Support Systems*, *116*, 48–63.