Reliability Engineering and System Safety 248 (2024) 110148

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Predicting maritime accident risk using Automated Machine Learning

Ziaul Haque Munim [a,*], Michael André Sørli [a], Hyungju Kim [b], Ilan Alon [c,d]

[a] Faculty of Technology, Natural and Maritime Sciences, University of South-Eastern Norway, Campus Vestfold, Horten, Norway
[b] Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[c] Department of Economics and Business Administration, Ariel University, Ariel, Israel
[d] School of Business and Law, University of Agder, Kristiansand, Norway

ARTICLE INFO

Keywords:
Maritime safety
Maritime accident
Machine learning
Classification tree
Artificial intelligence

ABSTRACT

Machine learning (ML), particularly, Automated machine learning (AutoML) offers a range of possibilities for analysing large volumes of historical maritime accidents data with advanced algorithms for integrating predictive analytics in operational and policy decision-making for improving maritime safety. This study explores historical data of maritime accidents in Norwegian waters over 40 years. The data has been utilised for analysing five major maritime accident categories: grounding, contact damage, fire or explosion, collision, and heavy weather damage. A total of 29 classification ML algorithms were trained, and the *Light Gradient Boosted Trees Classifier* was found to be the best-performing with the highest predictive accuracy. The three most impactful factors for accident risk are the *category of navigation waters, phase of operation, and gross tonnage of the vessel.* Based on the feature effect results, vessels sailing in narrow coastal waters, in the along-the-way operational phase, and fishing vessels are highly vulnerable to grounding relative to other types of accidents. The results can be used as input for the entire procedure of risk analysis, from hazard identification to quantification of accident consequences, and the best-performing ML algorithm can be utilized in developing a decision support system for real-time maritime accident risk assessment.

1. Introduction

Maritime activities at sea and in coastal areas, ranging from shipping to offshore installations, pose a degree of risk to the surroundings and hold the possibility to result in accidents. Maritime accidents entail the risk of damaging vessels, equipment, goods, and human lives. In 2021 alone, as many as 2637 occurrences led to 36 fatalities and 621 injuries with vessels sailing in the EU waters ([1], p. 7). In shipping, it can also be the economic risk shipowners take by transporting something of value. The grounding of the ship Ever Given in the Suez Canal in March 2021 accounted for an estimated economic loss of USD 6 to 8 billion per week [2].

A large number of factors are at play causing maritime accidents. These factors can be categorized into six groups: vessel and equipment related, navigation and operations related, human factors, environmental factors, traffic related, and shipping market condition related [3]. Vessel type, length, and tonnage are found important predictors of maritime accident risk [4–6]. Poor visibility, strong wind, heavy sea, and strong current are key navigation and environment related factors [4,6]. Human factors include inadequate manning, invalid competency certificates, and less sea experience of crew among others [6]. Congested shipping routes or locations and fluctuations in freight rates are relevant factors too. For instance, when shipping freight rates are high, it is common to increase sailing speed to make more trips before the rates go down. To avoid maritime accidents, understanding the most significant factors that contribute to maritime accidents is essential for operational and policy decision making.

Maritime conventions regulated internationally by the International Maritime Organization (IMO), such as SOLAS, MARPOL, STCW, and COLREG were developed and amended over time to ensure maritime safety [7]. To enhance maritime safety further, particularly from an operational point of view, predicting future events and implementing warning systems can be a useful operational tool [8]. Today relevant data associated with maritime accidents is available in a large number of sources, such as Automatic Identification System (AIS) data, weather information, vessel related data from sensors including Revolutions per minute (RPM), temperature, pressure, operating time, vibration, and so on from equipment and machineries onboard. Vessel related data are usually available to shipowners and ship management companies, while various service companies can provide AIS and weather-related data.

* Corresponding author.
E-mail address: ziaul.h.munim@usn.no (Z.H. Munim).

https://doi.org/10.1016/j.ress.2024.110148
Received 17 August 2023; Received in revised form 26 March 2024; Accepted 20 April 2024
Available online 21 April 2024

However, large volumes of available data are not utilized by ship operators and other relevant stakeholders such as the coast guard in operational decision making. The continuous development in machine learning (ML) modelling approaches can offer valuable insights on the accident predictabily of vessels through utilizing available data [9].

Rawson and Brito [10] provide a detailed review of studies using ML in maritime risk analysis in terms of context, datasets and methodologies used. Previously several frameworks for maritime safety have been published within different domains and in the maritime transportation field [11]. Several studies cover accidents in certain sea areas, such as the Yangtze River [12], navigation risks in Taiwan [13], in Chinese coastal waters [14], visibility estimation for safer navigation in the strait of Istanbul [15], a risk prediction system for early warning in the Lower Mississippi River [8], and a case study for vessel traffic management of the Western port of Shenzhen City in China [16].

This study explores maritime accidents in the Norwegian coastal waters, which has not been explored much in the literature [10]. Further, the Norwegian Maritime Authority (NMA) provides one of the most structured historical accident records data spanning over more than 40 years. Although, Bye & Aalberg [4] show several factors that lead to an increased probability of navigation accidents in the Norwegian coastal waters, their approach and the one of this study vary significantly. Bye & Aalberg [4] predicted whether an accident is navigation related or other type of accident. Similarly, Wang et al., [6] used a worldwide accident dataset for the period of 2010 to 2019 to predict maritime accident severity. The focus of this study is on predicting the probability of a particular accident type such as collision, contact damage, fire/explosion, grounding, and heavy weather damage using a dataset covering more than 40 years of accident records.

Moreover, application of recent developments in ML research such as the use of Automated Machine Learning (AutoML) to predict future maritime accidents is not evident in the existing literature. AutoML allows testing a wide-range of ML algorithms in a standalone as well as hybrid ensemble setting with optimised parameters to find the best-performing prediction model. Using AutoML, testing 20 or more relevant ML models for a particular dataset can be achieved within a manageable timeline. Further, AutoML, depending on the available tool used, facilitates robust cross validation of the model parameter as well as updating model parameters and hyperparameter optimization in a dynamic mode. The feature impact and feature effects of the AutoML approach further increase interpretability of the results. Considering these gaps in the literature, this study addresses two research questions in the nexus between maritime accidents and ML applications. First, which factors are most significantly associated with maritime accidents? Second, which ML models are most accurate in predicting maritime accidents?

This study is structured into six sections. In Section 2, the relevant literature on this topic is reviewed—also, theoretical background for maritime accidents to show the topic's history. The methodology is presented in Section 3 and also the data collection. The study results are presented in Section 4, with a discussion following in Section 5. A conclusion that summarizes the answers to the research questions comes in Section 6.

2. Literature review

The purpose of this section is not to provide a comprehensive review of maritime accident literature but to identify the most relevant studies that used ML approaches to accident risk prediction. For a detailed review of maritime accident literature, see Luo and Shin [3] and Cao et al., [17]. For a comprehensive review of ML applications in maritime accident risk analysis, see Rawson and Brito [10]. This study has reviewed a number of relevant journal articles on maritime accidents analysis that employed ML using a systematic approach. The literature search was conducted in the Web of Science (WOS) database using the Boolean expression: ("maritime accident*" AND "machine learning") OR "maritime accident* analysis". By manually screening the retrieved studies in WOS, 11 most relevant studies were identified. A summary of the most relevant studies is presented in Table 1.

2.1. Machine learning approaches used in maritime accident risk

ML is used to explore beyond the limitations of conventional methods to characterise risk. Studies use several different models to predict accidents, to mention some: regression analysis, grey system model (GM), exponential smoothing, Markov model [19]. In recent years, the use of ML has opened new doors to predicting accident risk by using models such as XGBoost, LightGBN, KNN, LinearRegression, DummyRegression, Logistic Regression (LR), Support Vector Machines (SVM), Random Forest (RF) [20,21,24]. Rawson, Brito, and Sabeur [22] use ML on historical casualty, weather, and vessel traffic data with an accuracy of 92 % and a recall of 95 %. However, most studies experimented between three and five ML algorithms, which limits the potential for exploring a wider range of algorithms.

The major accidents at sea can be prevented by various measures during the entire lifecycle of vessels, and one of the operational measures is implementing systems for early warning. This way, predicted risk probabilities can become inputs to an automated system. Merrick et al. [8] investigated the prediction accuracy in a warning system for maritime accidents using machine learning. Seven algorithms were used to test the data: logistic regression (LR), decision trees (DT), random forest (RF), k-Nearest neighbour (kNN), neural network (NN), gradient boosted trees (GBT), and Stochastic gradient boosted trees (SGBT) [8]. SGBT performed with the highest accuracy of 98.3 % action precision. Zhao & Lv [19] used regression analysis, grey system model (GM), exponential smoothing, and Markov model. It is evident that model or method comparions were performed by most studies to reveal the best-performing approaches. However, the same ML model might perform with lower or higher accuracy if applied to a different context.

Application of Bayesian networks were observed in two studies [14, 18], while the classic Fault Tree Analysis (FTA) in one [23]. While FTA is widely used in risk assessment, as the systems under consideration become larger, application of FTA becomes complex and time consuming. Also, FTA is a static approach and not appropriate for the dynamic analysis of a system. Bayesian network allows for models with factors that are multi-state and have non-linear relationships between them [26]. In recent years, dynamic bayesin network models have been used in real-time risk analysis, too [27].

ML models have been used in other contexts of the maritime sector. Atak & Arslanoğlu [21] used it to predict accidents in ports. In their study, they used classification models, SVM, Kernel SVM, KNN, Light GBM, XGBoost, Logistic and Naïve Bayes. A historical dataset from a terminal was used as the training and test data, and a dataset from another terminal was used to validate the models. The results show that three classification methods had an accuracy of 97 % for prediction. Fan and Yang [28] have used ANN model to predict stress levels of ship navigators using psychophysiological data. Such an approach is novel and contributes to understanding human factors in maritime safety. However, this approach is mostly feasible in simulated environments, and available historical accident records do not have any psychophysiological data at the time of accidents.

ML models were widely used in accidents prediction, management, and assessment in other industry contexts such as in aviation, nuclear, oil and gas, and transport sectors. For instance, the Deep Neural Network (DNN) model was found the best-in-class while predicting helicopter accidents [29]. Using deep learning models in combination with decision trees, an operator tool was proposed for management of accidents in nuclear power plants [30]. Graph-based ML models has been used in post-accident risk assessment for a urban gas pipeline network [31]. The Convolution neural network (CNN) model was found suitable for high-rail grade crossings accident, particularly while dealing with imbalanced data [32].

Table 1
Summary of relevant literature on maritime accidents.

No	Study	Data type	Data source	Methods	Best-performing
1	Jiang et al. [18]	413 marine accidents	LRF, IMO, 2010–2017	Bayesian network (BN)	Bayesian network (BN)
2	Zhao & Lv [19]	328 Maritime accidents	In Tianjin, 2003–2013	Regression analysis, grey system models (GM), exponential smoothing, Markov model	Markov GM (1,1)
3	Rawson, Brito, Sabeur, & Tran-Thanh [20]	207 unique accident data	GISIS, AIS, MetOcean, 2005–2018	Logistic Regression (LR), Support Vector Machines (SVM), Random Forest (RF), XGBoost	Support Vector Machines (SVM),
4	Atak & Arslanoğlu [21]	16 accidents and 16 pre/post-accident	Accident reports from terminals in Turkey	SVM, Kernel SVM, KNN, LightGBM, XGBoost, Logistic, Naïve Bayes classificatoion	XGBoost, LioghtGBM, KNN
5	Rawson, Brito, & Sabeur [22]	2 127 Incident data	MAIB, AIS (MMO), 2021–2020	Logistic Regression, Support Vector Machines, Random Forest, Gradient Boosted Trees (XGBoost)	XGBoost, Random Forest
6	Ugurlu & Cicek [23]	513 ship collision accidents	MAIB, GISIS 1977–2020	Fault tree analysis (FTA)	FTA
7	Liu et al. [14]	414 maritime accidents	China Maritime Safety Adm., 2013–2020	Bayesian Network (BN)	Bayesian Network (BN)
8	Bye & Aalberg [4]	931 maritime accidents	NMA, 2010–2016	Correspondence analysis, F-tests, Multivariate logistic regression	N/A
9	Kretschmann [24]	185 accident data on 544 container vessels	N/A	DummyRegressor, LinearRegression, RandoForest(RF)	RandimForest(RF)
10	Uyanık et al. [15]	Metrological data	Local weather stations in the Strait of Instanbul	AdaBoost Reg., Bayesian Ridge Reg.,Gradient Boosting Reg.,	Gradient Boosting Regression
11	Park & Jeong [25]	737 marine accidents	KMST, AIS-data, 2014–2018	Support Vector Machine (SVM), Relevance Vector Machine (RVM)	Relevance Vector Machine (RVM)

The existing body of literature encompasses a diverse array of methodologies for analyzing maritime accidents, including both ML models and analytical techniques such as FTA, BN, and so on. It is observed in Table 1 that most studies used three to five ML algorithms in their analysis. Some of the literature have adopted a monolithic approach, applying a singular method to dissect and provide insights into marine accidents. The other studies have embarked on a comparative analysis, leveraging multiple ML algorithms to identify the most suitable algorithm tailored for marine accident analysis. Nonetheless, a noticeable gap in the literature is evident, as these investigations have only explored a limited repertoire of ML models without conducting a comprehensive comparison across a wide spectrum of ML algorithms. This study aims to bridge this gap by experimenting with AutoML to systematically compare the suitability of a wide range of ML algorithms. From this perspective, the novelty of this study lies in its methodological approach, utilizing AutoML to identify the best-performing ML models in the context of maritime accident risk analysis.

2.2. Factors affecting maritime accident risk

While use of ML models is evident in accident analysis within maritime and other industry contexts, the use of AutoML is rare. Many studies focus on maritime accidents, factors affecting them, prevention and prediction. The input variables in the models are of high importance for ensuring higher accuracy of predictions. The most relevant variables for maritime accidents are summarized in Table 2, which faciliates the choice of variables in this study. The variables found in the literature can be categorised as vessel properties, weather-related, route properties, and mechanical.

Vessel properties related variables affecting maritime accidents are age of the vessel, length, deadweight tonnage, vessel type, vessel engine and materials, sailing speed, flag of convenience, flag state, maintenance records, classification society, ownership nationality, and number of crew onboard. Weather related factors are wind speed, wave height, visibility, and water depth. Route properties mainly relate to distance from shore and geographical location. Mechanical properties mainly include hydrodynamic effects. Several human and organizational factors are relevant such as safety culture, training and experience of the crew on board, communication in the bridge team etc. The majority of the published studies focused on vessel properties as listed in Table 2, and not all the variables in the table were available in the dataset used in this study, e.g., human and organizational related factors, vessel speed, distance from shore. While prior studies have made significant efforts in identifying critical factors influencing maritime accidents, none of them explored multiple accident class analysis. Most of them focused on one accident type or a binary target feature. This study addresses this gap by leveraging multi-class classification ML algorithms. This innovative approach allows for a nuanced examination of how accident risks differ across five types of marine accidents. This novelty is a key in creating better and more focused ways to prevent marine accidents, giving us a clearer and comprehensive picture of the complex factors that lead to

Table 2
Factors affecting maritime accidents.

Variable category	Variables	Reference
Vessel properties	• Vessel speed • Vessel age • Vessel category • Vessel length • Vessel flag of convenience • Vessel flag state and safety regime • Maintenance • IMO number • Class society • Nationality • Tonnage • Building material • Number of people onboard	Jiang et al. [18], Rawson, Brito, Sabeur, et al. [20], Rawson, Brito, & Sabeur [22], Bye & Aalberg [4], Mullai & Paulsson [33], B. Li et al. [34], Wang & Yang [35]
Weather-related	• Wind speed • Significant wave height • Visibility • Water depth	Rawson, Brito, Sabeur, et al. [20], Bye & Aalberg [4], Rawson, Brito, & Sabeur [22]
Route properties	• Location of accident • Distance from shore	Jiang et al. [18], Rawson, Brito, Sabeur, et al. [20]
Human and organizational	• Bridge Resource Management (BRM) • Communication • Position monitoring • Training and experience • Regulation • Inadequate procedures • Deviation from Standard Operating Procedures (SOP) • Lack of knowledge • Information • Clear order • Safety culture	Rawson, Brito, & Sabeur [22], Coraddu et al. [36], Fan, Zhang, et al. [37]
Mechanical	• Hydrodynamic effects	Rawson, Brito, & Sabeur [22]

these accidents.

3. Data and methodology

This study uses AutoML to find the optimal model for predicting maritime accidents. Typically extensive knowledge and understanding of mathematics, statistics, and computer science is expected in utilizing ML models. AutoML platforms make ML models accessible to researchers and practitioners without extensive training requirements. Of course, some understanding of mathematics and statistics are still required. There are several AutoML platforms that exist today such as BigML, DataRobot, H2O etc. In this study, the ML models are trained and tested using the cloud artificial intelligence (AI) platform DataRobot [38]. This study adopted a five-step process as depicted in Fig. 1.

3.1. Data

To predict maritime accidents, it is important to use data that are relevant and representative. The data used in this study are collected from the Norwegian Maritime Authority's (NMA) register for maritime accidents [39], available publicly. NMA's responsibility is all vessel-related issues, for all "vessels flying under the Norwegian flag and foreign ships in Norwegian waters" [40]. The dataset is available in Excel format, where variables are coded in Norwegian. Only the facts regarding the accident events such as tonnage, location, type of injury, severity, data, etc. are available in the dataset.

Before using the data in DataRobot, it was pre-processed by sorting out the unimportant factors and missing values. Pre-processing includes:

- Translating data to English
- Selecting the dependent variable *Accident type* with the parameters *Collision, Contact damage, Fire/explosion, Grounding and Heavy weather damage.* Based on the FSA reports by IMO, the accident types were selected. This reduced the dataset from 37 515 to 9 281 datapoints.
- Removing:
  - *Missing* and *unknown* vessels
  - Length and tonnage with the values of zero
  - Datapoints with two or more missing/unknown variables
  - Reduced the dataset to 9 025 datapoints
- The selection of independent variables. Important are factors that do not consist of too many categories (less than 20) and factors that occurred or were known before the accidents happened and not after. For example, unimportant information in this study includes the IMO number, vessel name, evacuation of the vessel, number of fatalities, etc.

The data are maritime accidents in Norwegian coastal waters from 1981 to 2020. It consists of 13 different accident types. However, we focus on five accident types based on IMO's FSA reports ([41], p. 13; [42], p. 12). In this study, considered maritime accidents are *Collision, Contact damage, Fire/explosion, Grounding*, and *Heavy weather damage*. Cao et al., [17] found that collision, machinery or hull damage, and capsizing lead to more severe maritime accidents. The dataset of this study consists of 9025 accidents from 1981 to 2020. The number of accidents distributed for each year is shown in Fig. 2. The total number of accidents during this period is relatively consistent with about 250 occurrences per year. Accident frequencies were decreasing steadily from 1995 to 2004, and then kept increasing again. The year 2020 reports the lowest number of accidents, most likely due to low shipping activity during the COVID pandemic. The different accident types distribution is also relatively stable over the years. Contact damage has a larger amount of instances in the last ten years compared to earlier.

Further, the annual percentage distribution of accidents for vessel types is presented in Fig. 3. Cargo vessels account for the majority of the accidents, while mobile facility and pleasure vessels have considerable lower counting than the other vessel types. The distribution shows a slight change after 2004 with increasing percentage of accidents for pleasure and passenger vessels and a decreasing trend in fishing vessels. The accident data distribution of Norwegian coastal waters is identical to the data of European waters, where cargo ships too account for the majority of the accidents followed by passenger and fishing vessels [1].

The variables used in this study are presented in Table 3 (after pre-processing), which are chosen based on literature review and availability of data. The goal is to use the independent variables in AutoML to classify and predict the outcomes of the dependent variable, that is, accident type.

3.2. Methodological approach

ML can be defined as "a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty" ([43], p. 1). Based on the learning approach, ML is typically classified into three categories: supervised, unsupervised, and reinforcement learning [43]. While supervised learning is suitable for labelled data, unsupervised is used when dealing with unlabeled data, and reinforcement learning learns from the interactions within an environment [44]. Some of the most common supervised algorithms are SVM, neural networks, random forests, decision trees, and genetic algorithms, while unsupervised algorithms include auto-encoders, deep learning, principal component analysis, and deep belief networks [44]. Moreover, based on output, ML has three methodological approaches: classification,

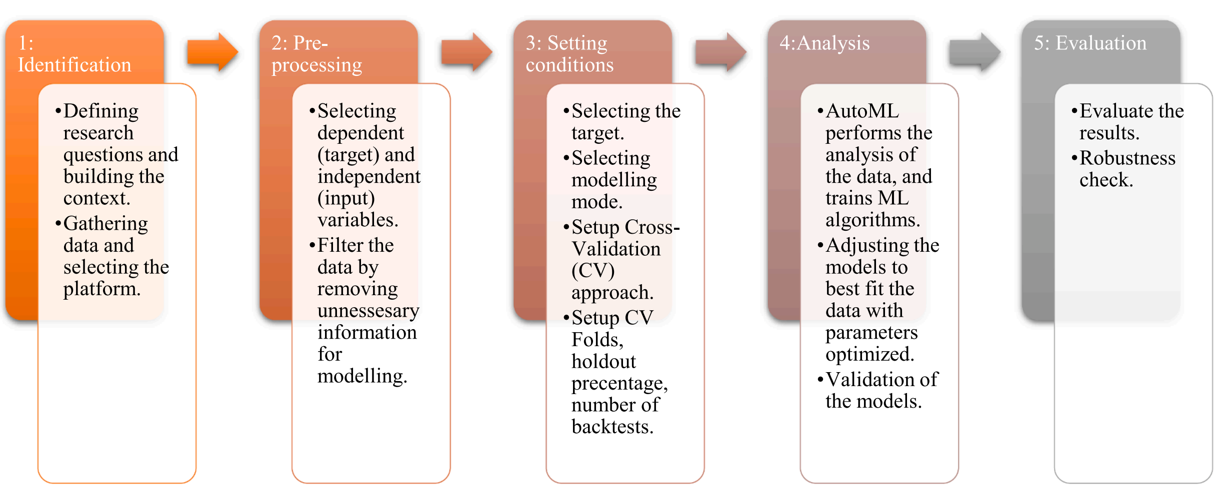


Fig. 1. Research workflow in implementing AutoML.

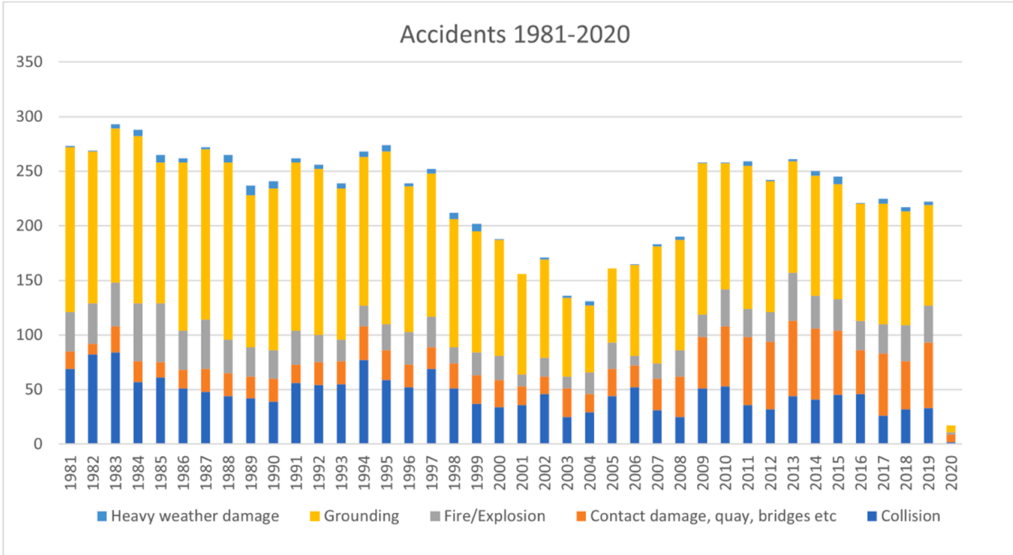


Fig. 2. Maritime accidents in Norwegian Coastal Waters, 1981–2020.

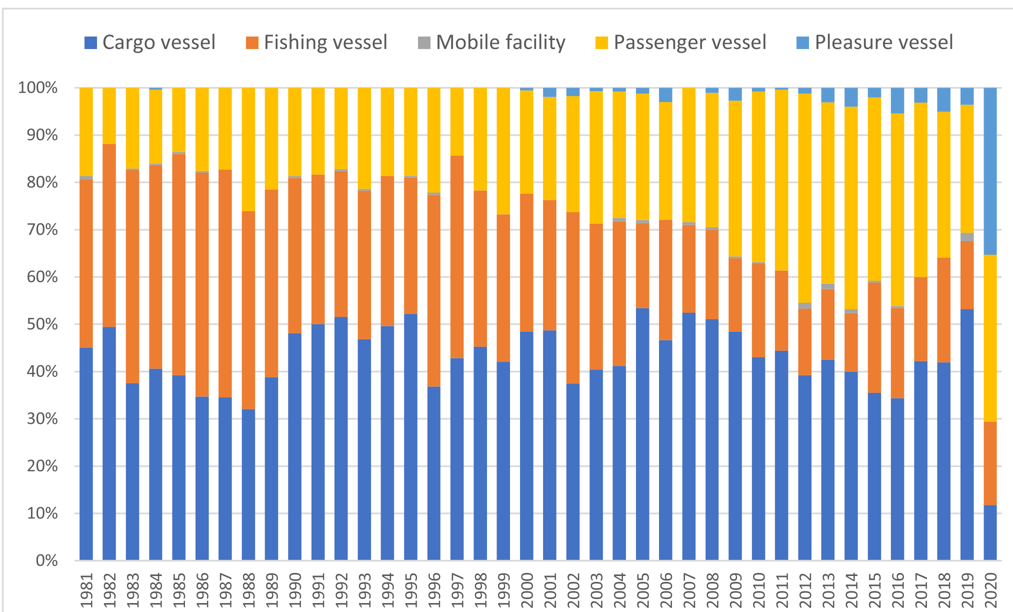


Fig. 3. Annual percentage of maritime accidents by vessel type 1981–2020.

clustering, and regression for class or condition detection, grouping or pattern recognition, and number prediction output, respectively [43].

The ML models utilized in this study can be referred to as a multiclass classification problem. Multiclass, because the outcome variable has more than two possibilities, and classification, because the output is a choice between classes, not numbers [45]. Although dates of the accidents are available in the data used in this study, it is not a time-series data structure since accidents occur irregularly. Hence, time series modelling approaches were not considered. In DataRobot, classification models were used to predict the accident type. However, two time features — month of the year and day of the week — were extracted through feature engineering from the date variable. These types of data, which are time-relevant but not a time series, should be evaluated in a Cross-Validation (CV) setting. A five-fold CV is considered for the robustness of the model estimates. For evaluation of the predictive performance of the ML models the *Accuracy* measure is used [46]. When selecting the suitable ML model for the dataset, four factors were considered: *Speed of training, Memory usage, Predictive accuracy* and *Interpretability*.

3.3. Automated machine learning (AutoML)

AutoML is about deciding the best ML model that fits data without any coding or major human manipulation, thereby automated [47]. Being an emerging research domain, there are relatively few published studies using AutoML yet. For complex analyses of large data sets, AutoML offers possibilities to test endless numbers of algorithms with optimized parameters in a manageable timeframe. Several studies discussed the unexplored potentials of the AutoML approach [48,49].

3.4. DataRobot

DataRobot is a cloud AI platform that offers advanced AI capabilities. With easy access and a user-friendly interface, both statisticians and

Table 3
Variables with descriptives.

Variable	Label	Descriptive values and frequencies
Date	• Year – Month – Week – Day	• 1981–2020
Accident type (Target variable)	• Collision • Contact damage (quay, bridge, etc.) • Fire/explosion • Grounding • Heavy weather damage	• 1 850 (20.5 %) • 1 229 (13.5 %) • 1 084 (12 %) • 4 722 (52.5 %) • 140 (1.5 %)
Type of vessel	• Cargo vessel • Passenger vessel • Fishing vessel • Pleasure vessel • Mobile facility	• 3 915 • 2 277 • 2 703 • 99 • 31
Waters	• Narrow coastal waters • In port area • Lakes • Canal, river, etc. • Separation and caution area • Oil field • Along quay • Archipelagic • Other • Unknown • Outer coastal waters • Open sea area	• 3 284 • 2 153 • 17 • 491 • 22 • 80 • 437 • 1 • 17 • 6 • 1 764 • 651
Length	• 3–366 m	• Mean: 61.20 • SD: 55.89 • Median 46.6 • Min: 3 • Max: 366
Gross tonnage	• 0.5–169 658 tonnes	• Mean: 4 510 • SD: 13 923 • Median: 429 • Min: 0.5 • Max: 169 658
NOR	• Norwegian nationality	• YES: 8 347 • NO: 678
Phase of operation	• Arrival at port • At quay • Departure at port • Along the way • At dock • Unknown • Other • Safety condition • By installation • In operation • Dynamic positioning • During fishing • During towing • By loading buoy • In drilling position • Anchored • In circulation	• 1 334 • 432 • 446 • 5 300 • 35 • 54 • 104 • 1 • 81 • 143 • 18 • 353 • 21 • 7 • 9 • 125 • 17

scientists can perform data analysis and predictions on all desired data types. The platform uses AutoML solutions that identifies best-performing model, in performance and up to date features. DataRobot explores the data provided and builds a number of models that best fit the data based on data features and accuracy. DataRobot uses open-source algorithms to build models that fit the data [38].

3.4.1. Data processing

The data are provided to the platform, and the target (dependent factor) to predict is selected by the user. The process can be divided into three phases, pre-processing, machine learning, and optimisation (see Fig. 4). The data are processed before being applied to algorithms and the training of models. This includes working with the "messy" data and cleaning it by transforming and reducing bad datapoints, finding the dataset's features and selecting the training and validation data. The ML phase uses algorithms to train and validate various ML models. The automation optimises the parameters, and identifies the best-performing model in terms of accuracy [49].

3.4.2. Training, validation and holdout

For robustness of ML models in predicting future events with unknown datasets with the same variables, the dataset at hand was divided into training, validation, and holdout samples. A part of the dataset needs to be selected as training data for estimating ML model prameters. The validation sample uses another part of the training data to test the trained model's performance in a new sample which was not used in the training of model parameters. This procedure is known as Cross Validation (CV), which controls for the overfitting issue in ML model estimation [50]. The holdout sample is separated from the training and validation process. The intention is to use the holdout information after selecting the best-performing model for a final evaluation of the performance. By not using the validation data in building models, validation accuracy can be considered unbiased [51].

It is common to use five or ten fold cross validation in ML studies to control for the overfitting issue as these have been found to achieve a good balance between computing time and reliable estimation of model performance. Studies have found that ten fold CV does not have any advantage over five fold (Feng et al., 2005; [52]). Hence, considering the sample size and findings from past studies we implemented five-fold CV. In CV, the default settings by DataRobot divide the data into 80 % training and validation, and 20 % holdout. Out of the 80 %, validation sample uses 20 % (1/5). This means that the training sample size is 64 % [100 %-(80 %/5)−20 %]. Five-fold cross validation is applied providing five sub-datasets (folds) that test the analysed model's performance individually (Fig. 5). The mean accuracy of the five CVs is the value presented as the model's accuracy leaderboard [51]. This CV approach reduces the overfitting and sample bias problem in ML.

3.4.3. Evaluation metrics

Evaluating the predictions of an algorithm is essential to validate their performance. Several metrics can be used to measure the performance of classification models, e.g. accuracy, precision, recall, F1-score, Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) [47]. The accuracy is relatable and describes the percentage of correctly classified accidents, with the ratio of correct predictions over correct and wrong predictions. The calculation is described as follows [46]:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{1}$$

Here, correct predictions include TP – True positives, TN – True negatives. Incorrect predictions include FP – False positives, FN – False negatives. In Multiclass Confusion Matrix (MCM), True Positive Rate (TPR), Positive Predictive Value (PPV), and the F1-score are presented as accuracy measures.

TPR or Recall, also known as sensitivity, shows the values correctly identified by the model:

$$\text{True Positive Rate (TPR)} = \frac{TP}{TP + FN} \tag{2}$$

PPV or Precision shows the correctly identified predicted values by the model:

$$\text{Positive Predictive Value (PPV)} = \frac{TP}{TP + FP} \tag{3}$$

A model's accuracy described by F1-score is the relative importance of precision vs recall:

$$\text{F1} - \text{score} = \frac{2(PPV{*}TPR)}{PPV + TPR} \tag{4}$$

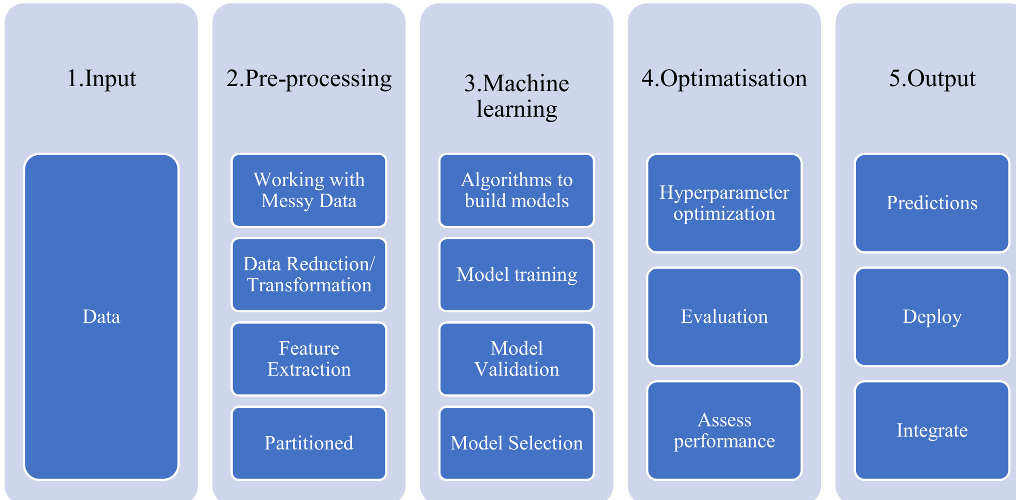


Fig. 4. AutoML process, based on [49].

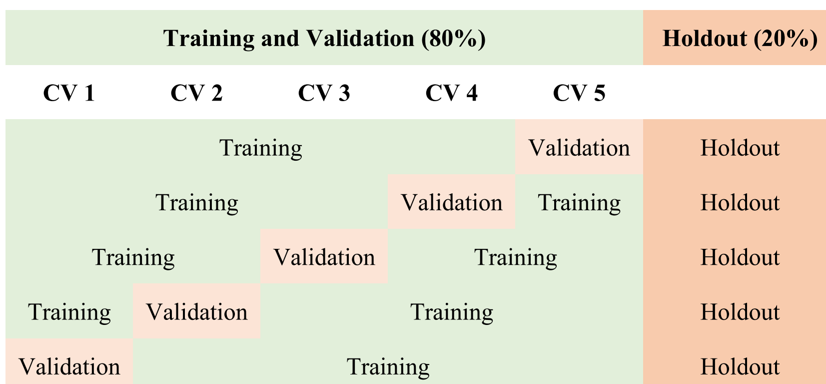


Fig. 5. Five-fold cross-validation illustration.

4. Results

The results based on the analysis in DataRobot are presented in this section with the ML model's accuracy of prediction. The data used for training and validation (80 %) are distributed with 52 % *Grounding*, 20.5 % *Collision,* 14 % *Contact damage,* 12 % *Fire/explosion* and 1,5 % *Heavy weather damage.* Fig. 6 presents the DataRobot model training initiation window.

4.1. Prediction models

In the comprehensive modelling mode, DataRobot trained a total of 29 ML algorithms. The performance information of the 29 estimated algorithms are presented in the Appendix (see Table 6). The accuracy of the top 5 models is presented in Table 4 in high to low accuracy order.

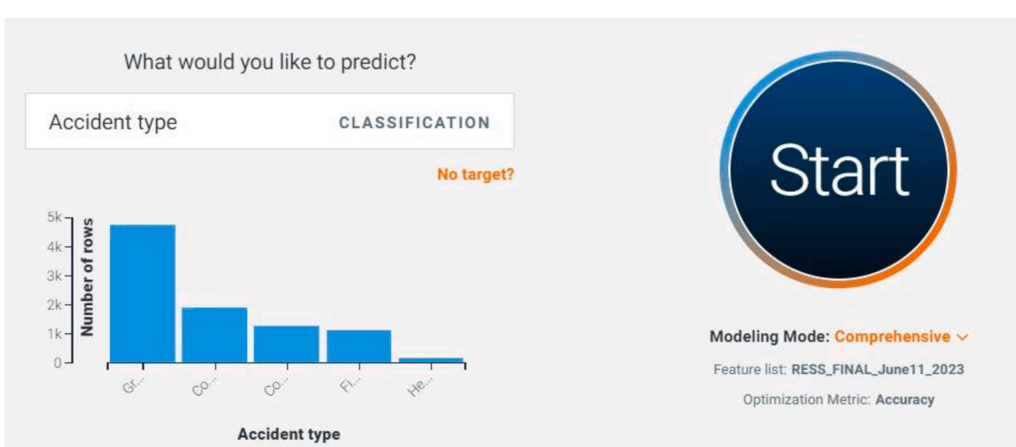


Fig. 6. DataRobot training initiation window.

Table 4
Top 5 ML models for maritime accident prediction.

No.	Model	Sample size	Validation	Cross-Validation	Holdout
1	Light Gradient Boosted Trees Classifier with Early Stopping (SoftMax Loss) (64 leaves)	100 %	0.6496	0.6427	0.6249
2	Light Gradient Boosted Trees Classifier with Early Stopping (SoftMax Loss) (64 leaves)	80 %	0.6406	0.6375	0.6294
3	Light Gradient Boosted Trees Classifier with Early Stopping (SoftMax Loss) (64 leaves)	64 %	0.6434	0.6349	0.6194
4	Light Gradient Boosted Trees Classifier with Early Stopping (SoftMax Loss) (16 leaves)	64 %	0.6475	0.6328	0.6294
5	Gradient Boosted Trees Classifier with Early Stopping	64 %	0.6413	0.6323	0.6183

The top three models are the same model with varying sample size. The blueprint of the best-performing model, that is, Light Gradient Boosted Trees Classifier with Early Stopping (SoftMax Loss) (64 leaves), is reported in Fig. 15 in the Appendix. The results also show that tree-based algorithms are top-performing for maritime accident risk prediction with a relatively stable accuracy of around 64 %. For the robustness check, the best-performing model is validated through Python (notebook copy provided as supplementary material). First, the algorithm is trained on 80 % data and tested on 20 % with accuracies of 89.60 % and 61.22 %, respectively. Further, five-fold cross validation was applied to the training sample, similar to the approach adopted in DataRobot. The accuracy in CV folds were between 59.49 % and 62.26 %, with an average accuracy of 61.18 %, which are identical to DataRobot accuracy percentages.

The model recommended by DataRobot is the Light Gradient Boosted Trees Classifier with Early Stopping (SoftMax Loss) (64 leaves) (hereafter referred to as LightGBM). This tree-based algorithm has some efficiency advantages in its design compared to standard Gradient Boosting Machines (GBM). It can handle large datasets faster and more accurately without too much memory [53,54]. GBM is known as one of the most versatile prediction algorithms and is based on the AdaBoost algorithm initially proposed by Freund & Schapire [55] with close connections to random forests models.

The LightGBM algorithm was developed in 2017 to improve the efficiency of the XGBoost algorithm on large datasets and multiple features. XGBoost is a time-consuming algorithm as it has to make the decision trees with all the possibilities for all factors. LightGBM uses three methods to improve efficiency compared to other gradient boosted methods: histogram-based decision tree algorithms, Gradient-Based One-Side Sampling (GOSS), and Exclusive Feature Bundling (EFB) [54]. The histogram-based decision trees to reduce the use of memory and increase speed, GOSS for increasing accuracy by including more gradients (in extensive range), and EFB for reducing the number of features.

4.2. Speed and accuracy

One of the important factors in selecting ML models is the speed of a model estimation and its relative accuracy. Fig. 7 presents the speed versus accuracy chart of the most relevant estimated models. It is observed that the LightGBM model (Model: M74, Blueprint: BP10) is providing the highest accuracy in lowest time. Hence, it was selected as the best-performing model.

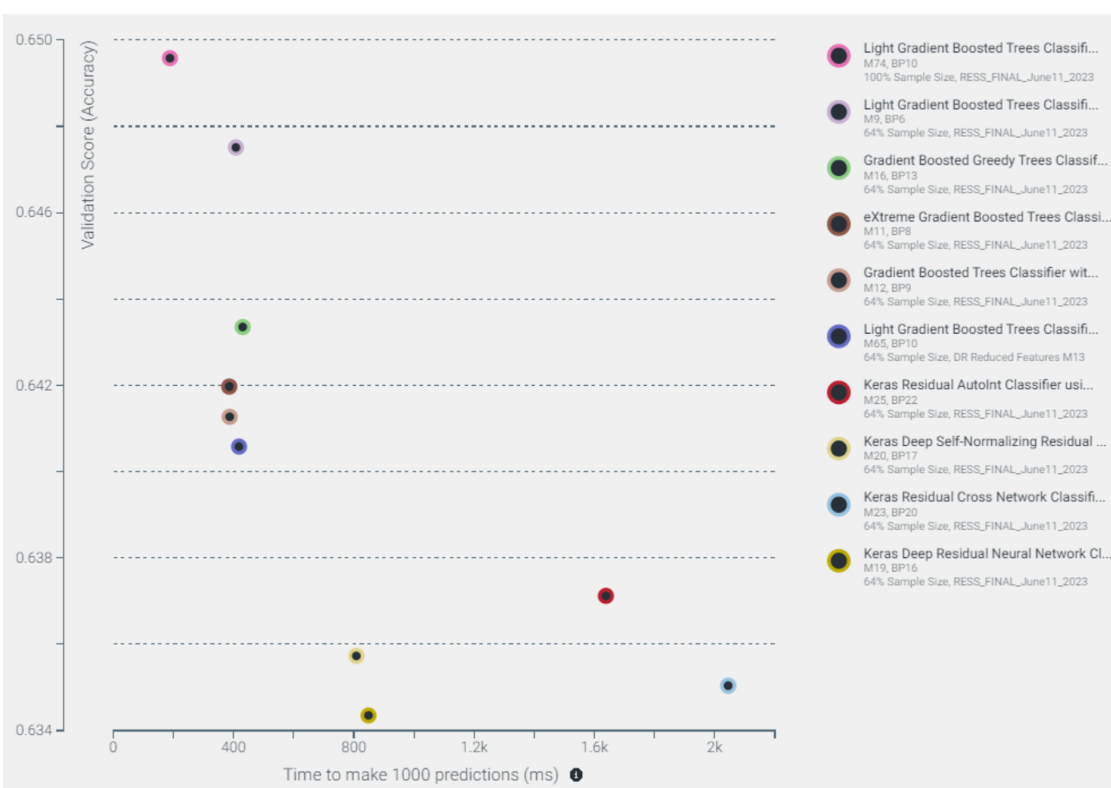


Fig. 7. Speed and accuracy of top performing ML models.

4.3. Feature impact

Feature impact shows the impact of each feature (or independent variable) under the selected ML model. Each factor is described with a percentage that represents the impact. The first factor has a default of 100 %, and the others are relative to that top factor. The value is a measurement of how much the predictive accuracy of the model is affected by changes in that factor [56]. Fig. 8 shows the aggregate feature impact diagram (a) — the overall impact of the features on the model, and the feature impact for each accident type (b-f). The results show that a change in the features *Waters, Phase of operation* and *Gross tonnage* significantly impacts the likelihood of accident type on an aggregate level (Fig. 8a).

For a particular accident type, the most significant features vary. The three most impactful features for collisions are waters, gross tonnage, and length of vessel (Fig. 8b); for contact damage, phase of operation, waters, and type of vessel (Fig. 8c); for fire and explosion phase of operation, waters, and gross tonnage (Fig 8d); for grounding, waters, phase of operation, and gross tonnage (Fig 8e); and for heavy weather damage, waters, gross tonnage, and length (Fig 8f). The ranking of the three most impactful features are the same for collision and heavy weather damage accidents. Waters is the most impactful for collision, grounding, and heavy weather damange, while phase of operation is the most impactful for contact damage and fire/explosion.

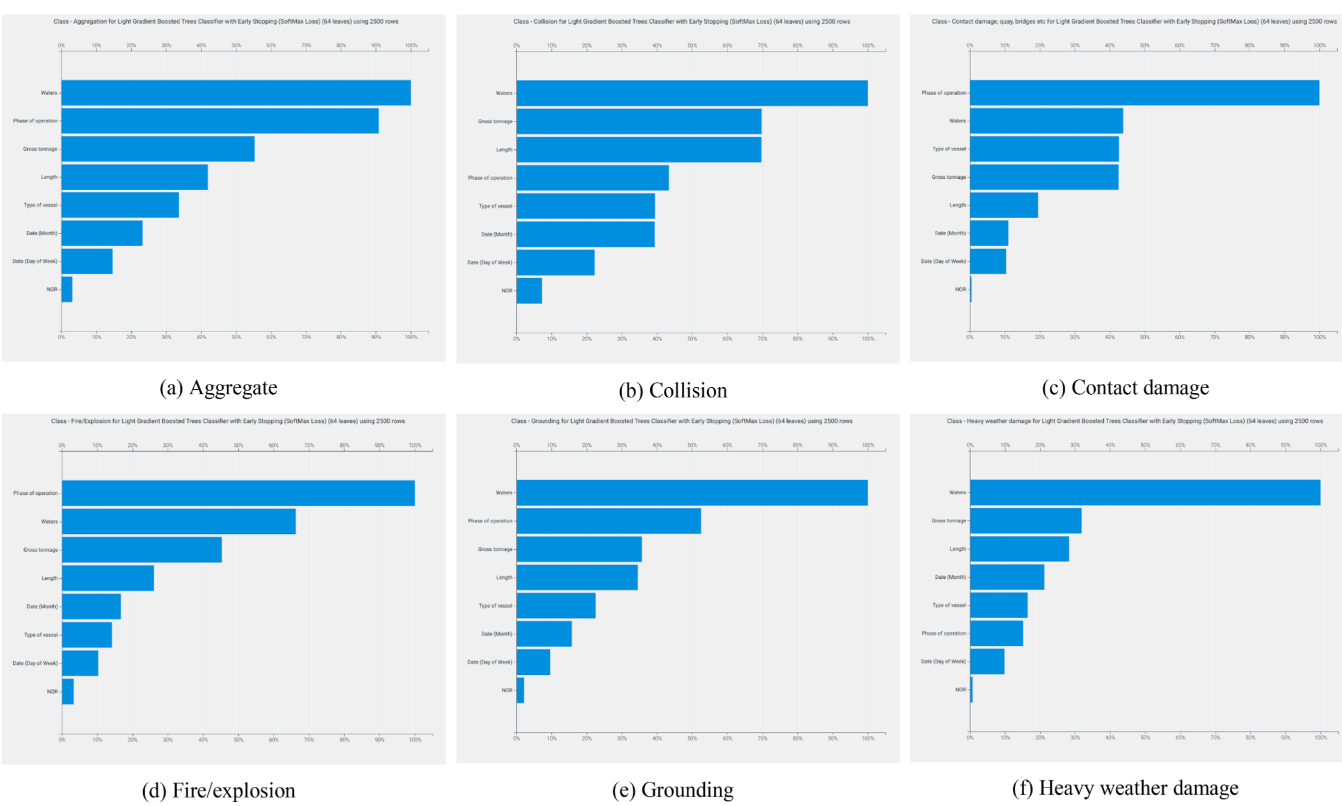
4.4. Feature effect

Feature effect shows how each input feature (or independent variable) effects the target (or dependent variable) under the best-performing ML model. The results in Fig. 9 to 13 show the effect of each input feature on the accident types *collision, grounding, contact damage, fire and explosion*, and *heavy weather damage*, respectively.

Feature effect figures depict three values: actual, predicted, and partial dependence. Acutal represents the original data in the sample dataset. The predicted values are based on the prediction of the best-performing model. The partial dependence values represent the impact of a feature on the target while accounting the impact of other features in the model, and its value ranges between 0 and 1. The categories *Other Unseen* and *Missing* are a collection of data that are not in the training set or missing in total [56].

Based on the partial dependence values in Fig. 9, among the categories of waters feature, open sea area and oil field are highly associated with collisions (Fig 9a). For gross tonnage and length of vessel features, the risk of collision increases steadily with an increase in vessel tonnage and length (Fig. 9b-c). Some cut-off points are observed as well. For example, length of vessel exceeding 100 meters increases the risk of collision to a greater extent. For phase of operation, arrival and departure at port, and at quay, dock, or circulation are associated with low risk of collision (Fig. 9d). Mobile facilities and pleasure vessels exhibit a higher risk of collision (Fig 9e). In some months of the year the risk of collision is higher than others, for instance, April, June, and July (Fig 9f). In the day of the week feature, weeks start from Sunday (0) and finish on Saturday (6). Later days of the week have a lower association with collision (Fig 9g). Finally, non-norwegian vessels and vessels of which ownership information is missing are more associated with collision risk (Fig. 9h). Similarly, the feature effect figures of grounding (Fig 10a-h), contact damage (Fig 11a-h), fire and explosion (Fig 12a-h), and heavy weather damage (Fig 13a-h) can be interpreted.

The values in feature effect figures are relative to the y-axis, the exact partial dependence values were extracted as CSV files. The impact values are between 0 and 1, which can be classified as low, medium, and high risk based on the effect value. The exact feature effect values for the five accident categories for three impactful features are reported in Table 5. Feature effect values of gross tonnage and length of vessels are not reported here since each numeric value in the data is associated with a feature effect score. Feature effect scores for month of the year, day of the week, and Norwegian vessel features are not reported as they are not among the most impactful features. The results show that the different



(a) Aggregate (b) Collision (c) Contact damage

(d) Fire/explosion (e) Grounding (f) Heavy weather damage

Fig. 8. Feature impacts.

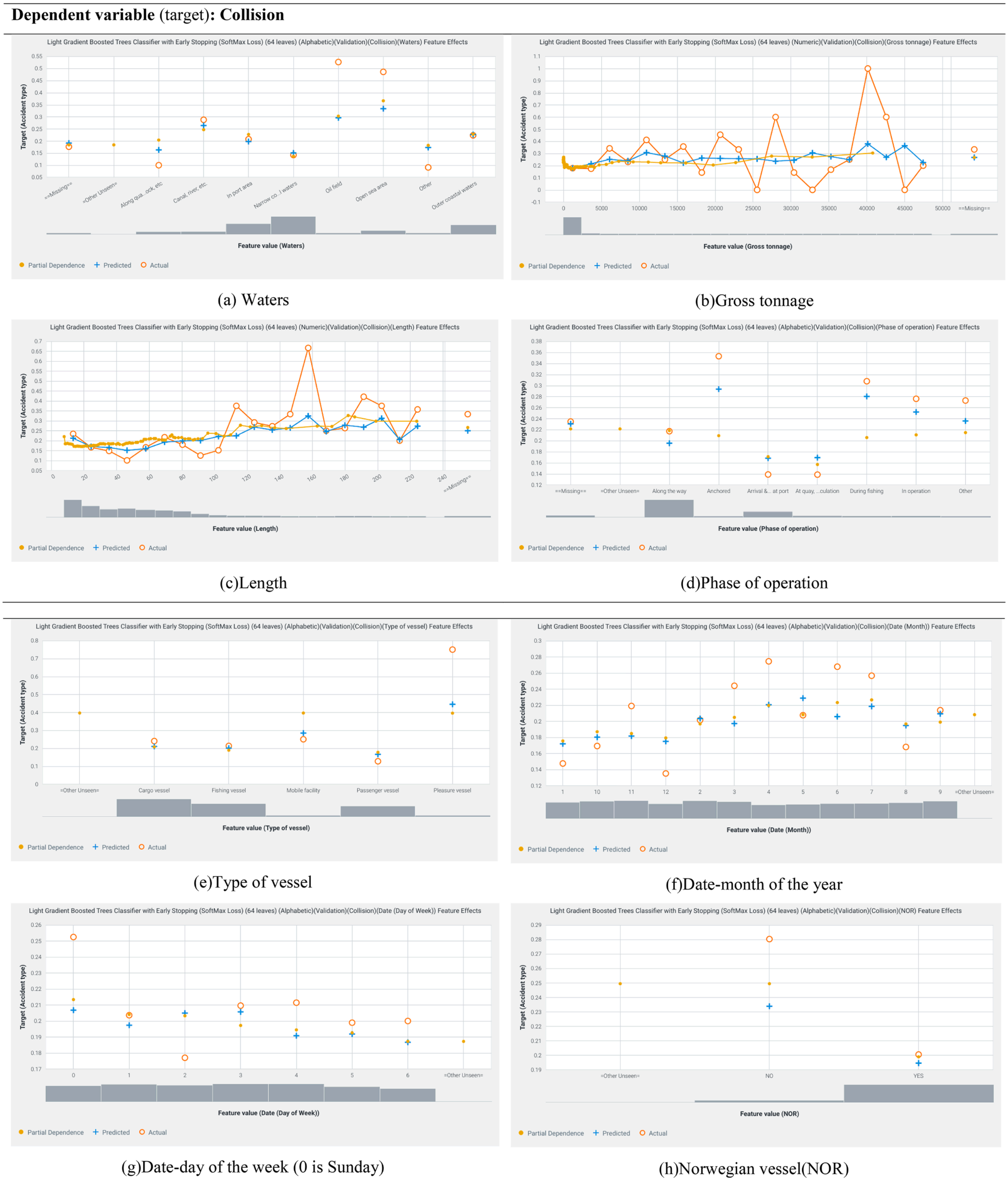


Fig. 9. Feature effects on collision.

accident type is associated differently with the same feature. For example, among the categories of *Waters, Narrow coastal waters (0.681)* has a high association with *grounding*, while very low with other types of accidents. Among the categories of the *Phase of Operation* variable, *Along the way (0.581)* has a high association with *grounding*, low (0.218) with *collision*, and very low with others. Similarly, association with other feature categories and accident types can be interpreted.

4.5. Multiclass confusion matrix (MCM)

The MCM visualises the best-performing model's prediction performance in a matrix structure, where correct and incorrect prediction

Fig. 10. Feature effects on grounding.

frequencies are presented. Green represents correct prediction and red represents incorrect. The exact values for each feature's performance (F1 score, recall and precision) are shown in Fig. 14 (b-f).

5. Discussion

The *Light Gradient Boosted Trees Classifier with Early Stopping (SoftMax Loss) (64 leaves)* has been found best-performing ML model for predicting maritime accidents in the Norwegian coastal waters. LightGBM has shown good performance in Atak & Arslanoğlu [21] for accident

(a) Waters (b)Gross tonnage

(c)Length (d)Phase of operation

(e)Type of vessel (f)Date-month of the year

(g)Date-day of the week (0 is Sunday) (h)Norwegian vessel(NOR)

Fig. 11. Feature effects on contact damage.

prediction in ports with an accuracy of 97 %. The LightGBM model has also shown good performance in other transportation areas, such as the emergency evacuation of metro stations [57] and predicting road traffic injury severity with an accuracy of 73 % [58]. While the latter accuracy is similar to this study, the higher accuracy in Atak & Arslanoğlu [21] could be due to their dataset features and contexts.

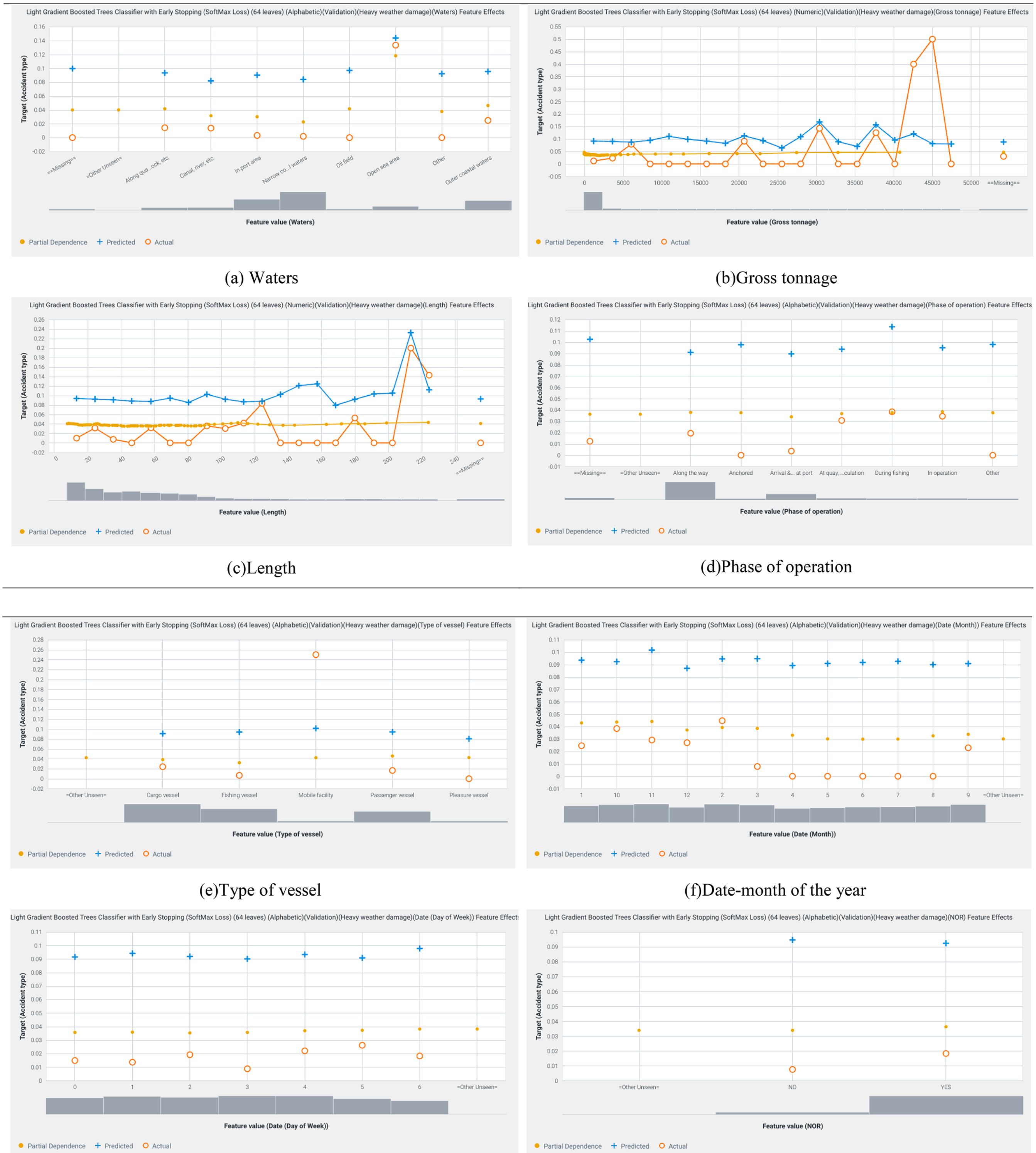
In general, a trend has been observed in the result of feature impact. *Waters* and *Phase of operation* have been the most impactful for most accident type predictions. This indicates that these factors are the most important to consider in the accident risk management of vessels operating in Norwegian waters. Other studies have found similar results, where location was one of the main influencing factors [18,35]. *Gross*

Dependent variable (target)**: Fire and explosion**

Fig. 12. Feature effects on fire and explosion.

tonnage, Length and *Type of vessel* are considered contributing factors for maritime accidents in this study. This is consistent with results found by other researchers [4,20,27,33]. Meanwhile, the feature *Phase of operation,* to the best of authors' knowledge, was not mentioned in the literature as a contributing factor to maritime accidents. This feature should be explored further as it has been found to be one of the most impactful in this study. Further, it is interesting to observe the impacts of month of the year, and day of week (see Fig. 9 and 10). These variables vary within a scope that is the same for the future as for the past. These trends and factors can be utilized for the entire process of risk analysis. For

Dependent variable (target)**: Heavy weather damage**



(a) Waters (b)Gross tonnage

(c)Length (d)Phase of operation

(e)Type of vessel (f)Date-month of the year

(g)Date-day of the week (0 is Sunday) (h)Norwegian vessel(NOR)

Fig. 13. Feature effects on heavy weather damage.

instance, the most impactful factors of each accident type can be used as guidewords for hazard identification (HAZID), and the trend can be an input to quantify accident scenarios in both frequency and consequence analysis.

In terms of ML models performance, the results show that the accuracy is stable at around 66 % for validation, cross-validation and holdout for the best-performing model. The results of feature impact show the importance of the different features in predicting the accident types. This can be used in both planning of and during the operation to see where to put the effort of reducing risk. If specific values of the

Table 5
Feature effect categorization by risk probability.

Validation sample							
Feature: Waters	**Collision**	**Contact damage**	**Fire/explosion**	**Grounding**	**Heavy weather damage**		
Feature Impact Rank (out of 8)	**1**	**2**	**2**	**1**	**1**		
##Missing##	0.184	0.187	0.115	0.474	0.040	High	0.681
#Other Unseen#	0.184	0.187	0.115	0.474	0.040	Medium	0.352
Along quay in dock etc.	0.204	0.297	0.164	0.293	0.042	Low	0.023
Canal river etc.	0.247	0.166	0.106	0.449	0.031		
In port area	0.227	0.180	0.104	0.459	0.030		
Narrow coastal waters	0.137	0.088	0.072	0.681	0.023		
Oil field	0.303	0.208	0.135	0.313	0.042		
Open sea area	0.366	0.123	0.242	0.151	0.118		
Other	0.182	0.180	0.110	0.489	0.038		
Outer coastal waters	0.228	0.111	0.144	0.471	0.046		

Feature: Phase of operation	**Collision**	**Contact damage**	**Fire/explosion**	**Grounding**	**Heavy weather damage**		
Feature Impact Rank (out of 8)	**4**	**1**	**1**	**2**	**6**		
##Missing##	0.221	0.114	0.162	0.467	0.036	High	0.581
#Other Unseen#	0.221	0.114	0.162	0.467	0.036	Medium	0.308
Along the way	0.218	0.073	0.089	0.581	0.038	Low	0.034
Anchored	0.209	0.118	0.184	0.451	0.038		
Arrival & departure at port	0.171	0.260	0.070	0.465	0.034		
At quay dock circulation	0.157	0.144	0.405	0.257	0.037		
During fishing	0.206	0.136	0.261	0.359	0.038		
In operation	0.210	0.182	0.165	0.404	0.039		
Other	0.215	0.125	0.170	0.452	0.038		

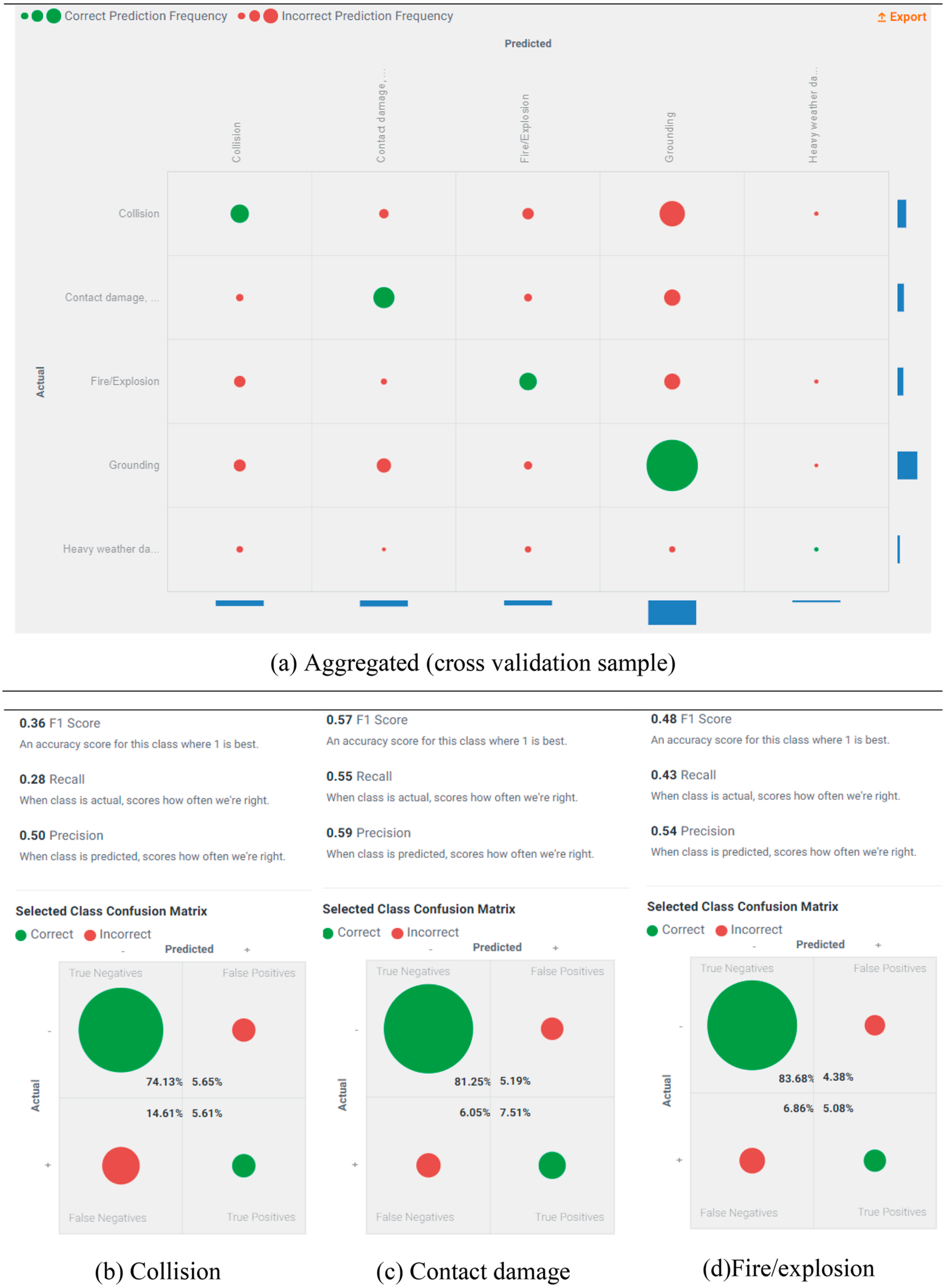
Feature: Type of vessel	**Collision**	**Contact damage**	**Fire/explosion**	**Grounding**	**Heavy weather damage**		
Feature Impact Rank (out of 8)	**5**	**3**	**6**	**5**	**5**		
#Other Unseen#	0.396	0.154	0.129	0.278	0.042	High	0.549
Cargo vessel	0.208	0.108	0.119	0.528	0.039	Medium	0.291
Fishing vessel	0.189	0.096	0.133	0.549	0.032	Low	0.032
Mobile facility	0.396	0.154	0.129	0.278	0.042		
Passenger vessel	0.177	0.176	0.138	0.463	0.046		

features are explored, the association between feature values and accident types are noticed, which can guide identifying "safer" operational states. An example from the *Grounding* data is: *Along the way* in *Narrow coastal waters* with a *Fishing* or *Cargo vessel* has a higher likelihood of *Grounding* than the other accidents. On the other hand, there is a lower likelihood of *Contact damage* in *Open sea areas, During fishing* with *Fishing vessels*.

Another observation is that *Narrow coastal waters* have a relatively low association with *Collision,* although the expectation would be the opposite. The expectations would be that narrow waters mean less distance between vessels and thus harder to manoeuvre, leading to more collisions. On the other hand, narrow waters require more caution and attention to correct vessel handling, with relatively lower ship speed. It may be that these situations raise the alert level and trigger procedures, e.g. more personnel on the bridge, extensive use of the radar and AIS, several lookouts etc., with more available time to react to the situations. *Type of vessel* generally has a lower association with accident types, about 33 %. This is an interesting result because others have found that vessel type is one factor that affects accidents the most [4].

The multiclass confusion matrix shows the correct and incorrect classification done by the model. TP/TN and FP/FN describe the model's correct and incorrect predictions for each selected variable. Fig. 11 shows that the model's accuracy varies much for the different variables. This needs to be considered together with the data distribution. *Grounding* has the best F1 score of 0.77 with 52 % of the data in the sample compared to *Heavy weather damage* F1 score of 0.09 with 1.5 % of the data. This indicates a connection between the model accuracy and the number of datapoints available.

The marine traffic system is a safety system composed of human, ship, environment and management factors. Therefore, maritime accidents are the results of the comprehensive interaction of these factors. The variables used in this study are limited to ship related factors. A ship is likely to have multiple decision support systems (DSS). The proposed model in this study could be one of the DSS. Although human factors are not incorporated into the proposed model, it can predict maritime accident risk in real-time when implemented, and the outcome can be used in a further, human factor based DSS. For example, when high risk is indicated based on our proposed model, the crew on board can get an alert.



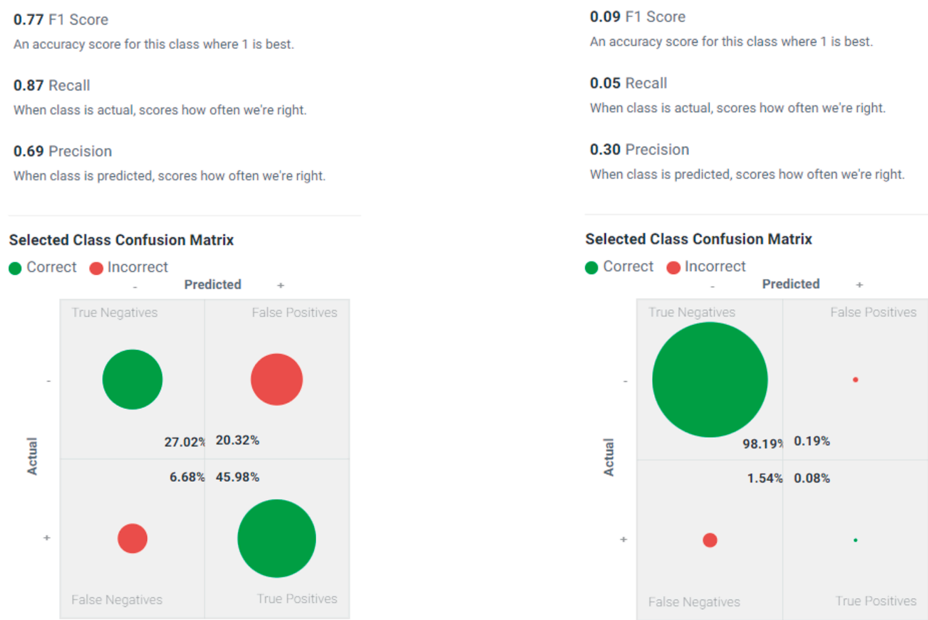
(a) Aggregated (cross validation sample)

(b) Collision (c) Contact damage (d)Fire/explosion

Fig. 14. Multiclass confusion matrix.

6. Conclusion

AutoML offers a wide-range of possibilities for analysing data using advanced ML models. This study used AutoML for predicting maritime accidents while addressing two research questions. Regarding the first research question, some factors have higher association with different accident types—especially *Waters* (i.e. *area of navigation at the time of accident)* and *Phase of operation*. The *Phase of operation* is not mentioned in the existing literature, hence, a contributing factor to maritime accident literature and should be considered in operational decision-



(e) Grounding   (f) Heavy weather damage

Fig. 14. (*continued*).

making and future research. Regarding second research question, tree-based models perform best for predicting maritime accidents in Norwegian coastal waters, particularly, the LightGBM family models. The accident data analysis results can be utilized for various steps in risk analysis of vessels, and the identified models can be used in developing a DSS for warning vessels about their predicted accident risk. Inclusion of further variables such as weather data and vessel sensor data can improve the model accuracy and might be useful in developing real-time accident risk assessment DSS. Applying AutoML to "live" data on different factors can function as a risk indicator that can alert the crew to avoid accidents.

There are a few limitations to this study. The dataset uses only Norwegian data from Norwegian waters. One of the reasons for choosing the dataset is that the Norwegian Maritime Authority provides one of the most structured historical accident records data spanning over more than 40 years. However, data does not include any information about the cause of the accident, and the exact position of the vessels is not considered, only the area. Traffic, topographic conditions and water depths in the area are also not considered. The recommendations for further research would be to follow up on this study and explore the limitations. For example, to apply AutoML to other datasets with additional variables such as weather-related factors. Further, adding spatial integration, GPS data, AIS history, topographic conditions, etc., to get better accuracy of the models and define areas of increased risk would be interesting. Another limitation is that the ML models are trained using data of only Norwegian coastal waters, which might not fit other context. Accident records data collected from other countries or regions need to be re-trained before deploying for accident prediction. Of course, AutoML simplifies the re-training process.

CRediT authorship contribution statement

Ziaul Haque Munim: Conceptualization, Formal analysis, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Michael André Sørli:** Data curation, Formal analysis, Methodology, Writing – original draft. **Hyungju Kim:** Conceptualization, Supervision, Validation, Writing – review & editing. **Ilan Alon:** Methodology, Writing – review & editing.

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.

Data availability

The data is publicly available through Norwegian Maritime Authority (NMA).

Acknowledgement

This research was supported by the 'Development of Autonomous Ship Technology (20200615)' funded by the Ministry of Oceans and Fisheries (MOF, Korea).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ress.2024.110148.

Appendix

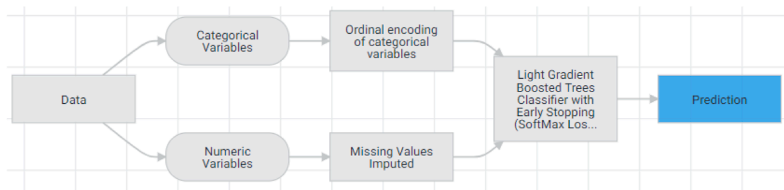
Classifier	Model (Blueprint)
Light Gradient Boosted Trees Classifier with Early Stopping (SoftMax Loss) (64 leaves)	

Fig. 15. Blueprint of light gradient boosted trees classifier.

Table 6
Predictive performance of the 29 trained ML algoritms.

No	ML Algorithms	Sample %	Holdout	Area Under the Curve	Accuracy
1	Gradient Boosted Trees Classifier with Early Stopping	64	1805	AUC: [0.81036, 0.8039]	Accuracy: [0.64127 0.63227]
2	Majority Class Classifier	64	1805	AUC: [0.5]	Accuracy: [0.51939]
3	Keras Deep Residual Neural Network Classifier using Training Schedule (2 Layers: 512, 512 Units)	64	1805	AUC: [0.73315]	Accuracy: [0.5831]
4	Keras Slim Residual Neural Network Classifier using Adaptive Training Schedule (1 Layer: 64 Units)	64	1805	AUC: [0.81027]	Accuracy: [0.63296]
5	Stochastic Gradient Descent Classifier	64	1805	AUC: [0.80302]	Accuracy: [0.62465]
6	Keras Deep Residual Neural Network Classifier using Training Schedule (3 Layers: 512, 64, 64 Units)	64	1805	AUC: [0.8145]	Accuracy: [0.63435]
7	Regularized Logistic Regression (L2)	64	1805	AUC: [0.79738]	Accuracy: [0.61704]
8	RandomForest Classifier (Gini)	64	1805	AUC: [0.8086]	Accuracy: [0.62327]
9	Keras Deep Self-Normalizing Residual Neural Network Classifier using Training Schedule (3 Layers: 256, 128, 64 Units)	64	1805	AUC: [0.81214, 0.8042]	Accuracy: [0.63573 0.626592]
10	Light Gradient Boosted Trees Classifier with Early Stopping (SoftMax Loss) (16 leaves)	64	1805	AUC: [0.81385, 0.8061]	Accuracy: [0.64751 0.632826]
11	Stochastic Gradient Descent Classifier	64	1805	AUC: [0.74579]	Accuracy: [0.58172]
12	Light Gradient Boosted Trees Classifier with Early Stopping (SoftMax Loss) (64 leaves)	16	1805	AUC: [0.79037]	Accuracy: [0.62535]
13	Light Gradient Boosted Trees Classifier with Early Stopping (SoftMax Loss) (64 leaves)	32	1805	AUC: [0.8002]	Accuracy: [0.6392]
14	**Light Gradient Boosted Trees Classifier with Early Stopping (SoftMax Loss) (64 leaves)**	**100**	**1805**	**AUC: [0.77782, 0.7823]**	**Accuracy: [0.64958 0.642656]**
15	Keras Residual Neural Factorization Machine Classifier using Training Schedule (2 Layers: 96, 96 Units)	64	1805	AUC: [0.7928	Accuracy: [0.61011]
16	Gradient Boosted Greedy Trees Classifier with Early Stopping	64	1805	AUC: [0.80989, 0.8039]	Accuracy: [0.64335 0.629364]
17	Decision Tree Classifier (Gini)	64	1805	AUC: [0.79513]	Accuracy: [0.62673]
18	Keras Wide Residual Neural Network Classifier using Training Schedule (1 Layer: 1536 Units)	64	1805	AUC: [0.73024]	Accuracy: [0.56371]
19	Light Gradient Boosted Trees Classifier with Early Stopping (SoftMax Loss) (64 leaves)	64	1805	AUC: [0.80706, 0.80289]	Accuracy: [0.64335 0.634902]
20	Light Gradient Boosted Trees Classifier with Early Stopping (SoftMax Loss) (64 leaves)	64	1805	AUC: [0.80824, 0.80132]	Accuracy: [0.64058 0.632269]
21	Light Gradient Boosted Trees Classifier with Early Stopping (SoftMax Loss) (64 leaves)	80	1805	AUC: [0.77619, 0.77507]	Accuracy: [0.64058 0.6375339]
22	ExtraTrees Classifier (Gini)	64	1805	AUC: [0.7921	Accuracy: [0.62673]
23	eXtreme Gradient Boosted Trees Classifier with Early Stopping	64	1805	AUC: [0.80946, 0.80376]	Accuracy: [0.64197 0.630746]
24	Stochastic Gradient Descent Classifier	64	1805	AUC: [0.79663]	Accuracy: [0.62396]
25	Stochastic Gradient Descent Classifier	64	1805	AUC: [0.79802]	Accuracy: [0.61911]
26	Keras Residual AutoInt Classifier using Training Schedule (2 Attention Layers with 2 Heads, 2 Layers: 96, 96 Units)	64	1805	AUC: [0.81104, 0.80374]	Accuracy: [0.63712 0.626594]
27	Keras Residual Cross Network Classifier using Training Schedule (3 Cross Layers, 4 Layers: 96, 96, 72, 72 Units)	64	1805	AUC: [0.81113, 0.80327]	Accuracy: [0.63504 0.62964]
28	Keras Slim Residual Neural Network Classifier using Training Schedule (1 Layer: 64 Units)	64	1805	AUC: [0.81485]	Accuracy: [0.63296]
29	Keras Slim Residual Neural Network Classifier using Adaptive Training Schedule (1 Layer: 64 Units)	64	1805	AUC: [0.80557]	Accuracy: [0.62465]

References

[1] EMSA. Annual overview of marine casualties and incidents 2022. Eur Marit Saf Agency 2022. https://www.emsa.europa.eu/newsroom/latest-news/download/7362/4867/23.html.

[2] Russon, M. (2021). The cost of the Suez canal blockage. BBCNews. https://www.bbc.com/news/business-56559073, accessed on June 11, 2023.

[3] Luo M, Shin SH. Half-century research developments in maritime accidents: future directions. Accid Anal Prevent 2019;123:448–60.

[4] Bye RJ, Aalberg AL. Maritime navigation accidents and risk indicators: an exploratory statistical analysis using AIS data and accident reports. Reliab Eng Syst Saf 2018;176:174–86. https://doi.org/10.1016/j.ress.2018.03.033.

[5] Cao Y, Wang X, Wang Y, Fan S, Wang H, Yang Z, Shi R. Analysis of factors affecting the severity of marine accidents using a data-driven Bayesian network. Ocean Eng 2023;269:113563. https://doi.org/10.1016/j.oceaneng.2022.113563.

[6] Wang H, Liu Z, Wang X, Graham T, Wang J. An analysis of factors affecting the severity of marine accidents. Reliab Eng Syst Saf 2021;210:107513. https://doi.org/10.1016/j.ress.2021.107513.

[7] Listing of current IMO publications. https://www.imo.org/en/publications/Pages/CurrentPublications.aspx.

[8] Merrick JRW, Dorsey CA, Wang B, Grabowski M, Harrald JR. Measuring prediction accuracy in a maritime accident warning system. Prod Oper Manag 2022;31(2):819–27. https://doi.org/10.1111/poms.13581.

[9] Rawson A, Brito M, Sabeur Z, Tran-Thanh L. From conventional to machine learning methods for maritime risk assessment. Trans-Int J Mar Navig Saf Sea Transp 2021;15(4):757–64. https://doi.org/10.12716/1001.15.04.06.

[10] Rawson A, Brito M. A survey of the opportunities and challenges of supervised machine learning in maritime risk analysis. Transp Rev 2022. https://doi.org/10.1080/01441647.2022.2036864.

[11] Hänninen M, Valdez Banda OA, Kujala P. Bayesian network model of maritime safety management. Expert Syst Appl 2014;41(17):7837–46. https://doi.org/10.1016/j.eswa.2014.06.029.

[12] Wu B, Wang Y, Zhang J, Savan EE, Yan X. Effectiveness of maritime safety control in different navigation zones using a spatial sequential DEA model: yangtze River case. Accid Anal Prevent 2015;81:232–42. https://doi.org/10.1016/j.aap.2014.12.024.

[13] Ung ST. Navigation risk estimation using a modified Bayesian Network modeling-a case study in Taiwan. Reliab Eng Syst Saf 2021;213:107777. https://doi.org/10.1016/j.ress.2021.107777.

[14] Liu K, Yu Q, Yuan Z, Yang Z, Shu Y. A systematic analysis for maritime accidents causation in Chinese coastal waters using machine learning approaches. Ocean Coast Manag 2021;213:105859. https://doi.org/10.1016/j.ocecoaman.2021.105859.

[15] Uyanık T, Karatuğ Ç, Arslanoğlu Y. Machine learning based visibility estimation to ensure safer navigation in strait of Istanbul. Appl Ocean Res 2021;112:102693. https://doi.org/10.1016/j.apor.2021.102693.

[16] Li M, Mou J, Liu R, Chen P, Dong Z, He Y. Relational model of accidents and vessel traffic using AIS data and GIS: a case study of the Western Port of Shenzhen City. J Mar Sci Eng 2019;7(6):163. https://doi.org/10.3390/jmse7060163.

[17] Cao Y, Wang X, Yang Z, Wang J, Wang H, Liu Z. Research in marine accidents: a bibliometric analysis, systematic review and future directions. Ocean Eng 2023;284:115048.

[18] Jiang M, Lu J, Yang Z, Li J. Risk analysis of maritime accidents along the main route of the Maritime Silk Road: a Bayesian network approach. Mariti Policy Manag 2020;47(6):815–32. https://doi.org/10.1080/03088839.2020.1730010.

[19] Zhao J, Lv J. Comparing prediction methods for maritime accidents. Transp Plan Technol 2016;39(8):813–25. https://doi.org/10.1080/03081060.2016.1231901.

[20] Rawson A, Brito M, Sabeur Z, Tran-Thanh L. A machine learning approach for monitoring ship safety in extreme weather events. Saf Sci 2021;141:105336. https://doi.org/10.1016/j.ssci.2021.105336.

[21] Atak Ü, Arslanoğlu Y. Machine learning methods for predicting marine port accidents: a case study in container terminal. Ships Offshore Struct 2021:1–8. https://doi.org/10.1080/17445302.2021.2003067. 0(0).

[22] Rawson A, Brito M, Sabeur Z. Spatial modeling of maritime risk using machine learning. Risk Anal 2021. https://doi.org/10.1111/risa.13866. 0(0).

[23] Ugurlu H, Cicek I. Analysis and assessment of ship collision accidents using fault tree and multiple correspondence analysis. Ocean Eng 2022;245:110514. https://doi.org/10.1016/j.oceaneng.2021.110514.

[24] Kretschmann L. Leading indicators and maritime safety: predicting future risk with a machine learning approach. J Shipp Trade 2020;5(1):19. https://doi.org/10.1186/s41072-020-00071-1.

[25] Park J, Jeong J-S. An estimation of ship collision risk based on relevance vector machine. J Mar Sci Eng 2021;9(5):538. https://doi.org/10.3390/jmse9050538.

[26] Chen P, Huang Y, Mou J, van Gelder PHaJM. Probabilistic risk analysis for ship-ship collision: state-of-the-art. Saf Sci 2019;117:108–22. https://doi.org/10.1016/j.ssci.2019.04.014.

[27] Chen J, Zhong PA, An R, Zhu F, Xu B. Risk analysis for real-time flood control operation of a multi-reservoir system using a dynamic Bayesian network. Environ Modell Softw 2019;111:409–20.

[28] Fan S, Yang Z. Towards objective human performance measurement for maritime safety: a new psychophysiological data-driven machine learning method. Reliab Eng Syst Saf 2023:109103.

[29] Xu Z, Saleh JH, Subagia R. Machine learning for helicopter accident analysis using supervised classification: inference, prediction, and implications. Reliab Eng Syst Saf 2020;204:107210.

[30] Lee JH, Yilmaz A, Denning R, Aldemir T. An online operator support tool for severe accident management in nuclear power plants using dynamic event trees and deep learning. Ann Nucl Energy 2020;146:107626.

[31] Yang Y, Li S, Zhang P. Data-driven accident consequence assessment on urban gas pipeline network based on machine learning. Reliab Eng Syst Saf 2022;219:108216.

[32] Gao L, Lu P, Ren Y. A deep learning approach for imbalanced crash data in predicting highway-rail grade crossings accidents. Reliab Eng Syst Saf 2021;216:108019.

[33] Mullai A, Paulsson U. A grounded theory model for analysis of marine accidents. Accid Anal Prevent 2011;43(4):1590–603. https://doi.org/10.1016/j.aap.2011.03.022.

[34] Li B, Lu J, Lu H, Li J. Predicting maritime accident consequence scenarios for emergency response decisions using optimization-based decision tree approach. Marit Policy Manag 2021:1–23. https://doi.org/10.1080/03088839.2021.1959074.

[35] Wang L, Yang Z. Bayesian network modelling and analysis of accident severity in waterborne transportation: a case study in China. Reliab Eng Syst Saf 2018;180:277–89. https://doi.org/10.1016/j.ress.2018.07.021.

[36] Coraddu A, Oneto L, Navas de Maya B, Kurt R. Determining the most influential human factors in maritime accidents: a data-driven approach. Ocean Eng 2020;211:107588. https://doi.org/10.1016/j.oceaneng.2020.107588.

[37] Fan S, Zhang J, Blanco-Davis E, Yang Z, Yan X. Maritime accident prevention strategy formulation from a human factor perspective using Bayesian Networks and TOPSIS. Ocean Eng 2020;210:107544. https://doi.org/10.1016/j.oceaneng.2020.107544.

[38] DataRobot. Automated Machine Learning. DataRobot AI Cloud 2022. https://www.datarobot.com/platform/automated-machine-learning/.

[39] Sjøfartsdirektoratet. (2022b, April 15). Accident statistics. https://www.sdir.no/en/shipping/accidents-and-safety/ulykkesstatistikk/Accident-statistics/.

[40] Sjøfartsdirektoratet. (2022a, April 15). About us—Norwegian Maritime Authority. https://www.sdir.no/en/about-us/.

[41] Marine Environment Protection Committee. (2008). Formal safety assessment—FSA crude oil tankers MEPC 58/INF.2. IMO.

[42] Maritime Safety Committee. (2007). Formal seafety assessment—FSA container vessels MSC 83/INF.8. IMO.

[43] Murphy KP. Machine learning: a probabilistic perspective. Cambridge, Massachusetts: MIT press; 2012.

[44] Bertolini M, Mezzogori D, Neroni M, Zammori F. Machine Learning for industrial applications: a comprehensive literature review. Expert Syst Appl 2021;175:114820.

[45] MathWorks. (2020). Machine learning with MATLAB. 65.

[46] DataRobot. Optimization metrics: dataRobot docs https://app2.datarobot.com/docs/modeling/reference/model-detail/opt-metric.html#accuracybalanced-accuracy.

[47] Wachowicz M, Gao S. Machine learning approaches. Geogr Inf Sci Technol Body Knowl 2020;2020(Q2). https://doi.org/10.22224/gistbok/2020.2.5.

[48] Bahri M, Salutari F, Putina A, Sozio M. Automl: state of the art with a focus on anomaly detection, challenges, and research directions. Int J Data Sci Anal 2022;14(2):113–26.

[49] He X, Zhao K, Chu X. AutoML: a survey of the state-of-the-art. Knowl Based Syst 2021;212:106622. https://doi.org/10.1016/j.knosys.2020.106622.

[50] Wong TT, Yeh PY. Reliable accuracy estimates from k-fold cross validation. IEEE Trans Knowl Data Eng 2019;32(8):1586–94.

[51] DataRobot. Data partitioning and validation: dataRobot docs https://app2.datarobot.com/docs/modeling/reference/model-detail/data-partitioning.html.

[52] Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. J Cheminform 2014;6(1):1–15.

[53] DataRobot. Light gradient boosted trees classifier with early stopping documentation—datarobot model documentation. https://app2.datarobot.com/model-docs/tasks/ESLGBMTC-Light-Gradient-Boosted-Trees-Classifier-with-Early-Stopping.html.

[54] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y. LightGBM: a highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst 2017;30. https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html.

[55] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 1997;55(1):119–39. https://doi.org/10.1006/jcss.1997.1504.

[56] DataRobot. Feature effects: dataRobot docs. https://app2.datarobot.com/docs/modeling/analyze-models/understand/feature-effects.html.

[57] Guo K, Zhang L. Adaptive multi-objective optimization for emergency evacuation at metro stations. Reliab Eng Syst Saf 2022;219:108210. https://doi.org/10.1016/j.ress.2021.108210.

[58] Dong S, Khattak A, Ullah I, Zhou J, Hussain A. Predicting and analyzing road traffic injury severity using boosting-based ensemble learning models with SHAPley additive explanations. Int J Environ Res Public Health 2022;19(5):2925. https://doi.org/10.3390/ijerph19052925.