Turk. J. Fish.& Aquat. Sci. 24(SI), TRJFAS27015
https://doi.org/10.4194/TRJFAS27015

Turkish Journal of FISHERIES and AQUATIC SCIENCES

RESEARCH PAPER

A Prediction Model for Maritime Accidents and Associated Risks in Arctic Routes Using Deep Learning

Mengü Demir[1], Sercan Erol[2], Serdar Yildiz[3,*]

[1]Department of Data Processing Center, Karadeniz Technical University, Trabzon, Turkiye.
[2]Maritime Transportation and Management Engineering Department, Karadeniz Technical University, Trabzon, Turkiye.
[3]Maritime Transport Department, Sharjah Maritime Academy, Sharjah, United Arab Emirates.

How to Cite

Demir, M., Erol S., Yildiz, S. (2024). A Prediction Model for Maritime Accidents and Associated Risks in Arctic Routes Using Deep Learning. *Turkish Journal of Fisheries and Aquatic Sciences*, *24(SI)*, *TRJFAS27015*. https://doi.org/10.4194/TRJFAS27015

Article History
Received 28 October 2024
Accepted 11 December 2024
First Online 20 December 2024

Corresponding Author
E-mail: serdar.yildiz@sma.ac.ae

Keywords
Arctic navigation
Deep learning
Machine learning
Marine casualties

Abstract

Climate change led to a global effort to seek for new trade routes in the Arctic, significantly altering the maritime sector and increasing maritime activities. Attempt to search and navigate on new routes brought likelihood to a rise in ship accidents, therefore presenting substantial associated risks. Study employed deep learning techniques to predict accidents and outcomes, focusing on the likelihood of occurrences. Accident dataset covering 2005-2017, including variables such as vessel length, age, tonnage, and weather conditions. Dataset was divided into 70% training and 30% testing, using a k-fold cross-validation approach on 511 input-output combinations with 1000 trials each. Results demonstrate that ship tonnage, length, and age are crucial predictors. Highest F1 score (0.89) and lowest standard deviation were achieved using all features. Removing features like minimum daily temperature significantly reduced model performance, reliability improved when combined with weather forecasts. Model can aid planning and management of Arctic maritime operations by predicting associated risks and optimizing insurance premiums. Future research should incorporate additional data sources, test the model under diverse maritime conditions, and focus on specific ship types to develop specialized mitigation strategies. Implementing Polar Code regulations into model predictions can expand the model's applicability and offer insights for policymakers.

Introduction

Today, climate change has led to the emergence of new trade routes, particularly in the Arctic region, prompting significant changes in the maritime sector (Bai, 2015; Chircop, 2020; Hebbar et al., 2020; Sakhuja, 2014). These changes, along with the utilization of new trade routes, have resulted in an increase in maritime activities such as fishing and aquaculture in the region, which in turn has led to an increase in the number of ship accidents. This increase carries with it a range of risks, primarily environmental impacts. The growing commercial interest in Arctic waters and the accompanying increase in ships due to climate change have continually risen. According to a 2024 report by Allianz Global Corporate & Specialty (AGCS), the number of ships entering the Arctic Polar Code area increased by 37% (approximately 500 ships) between 2013 and 2023. The conditions in polar waters are challenging due to sea ice and hazardous weather conditions; additionally, the remoteness of the region limits access to rescue services when a ship encounters problems, as evidenced by a passenger ship that ran aground in Greenland in September 2023 and was unable to be rescued for three days (Allianz, 2024). The use of non-ice-class tankers in Arctic waters is concerning due to the insufficient number of icebreakers to meet the increasing demand. In the past, incidents of ships becoming trapped in ice have occurred. If a non-ice-class ship becomes trapped, this not only poses a risk to the potential loss of the ship and crew safety but can also lead to a potential environmental disaster. Measures critical to enhancing navigational safety and maritime skills in Arctic sea routes are imperative (Council, 2024). Accidents result in environmental damage such as loss of life and oil spills, and at best, cause delays and disruptions in maritime trade (Allianz, 2024; Boylan, 2021). Therefore, a detailed analysis of accidents, their prevention, and the mitigation of potential damages are crucial for accurately predicting accident risks in advance.

Few studies have been conducted on the Arctic regions using both machine learning techniques and other statistical methods. In their study, Fu et al. (2016) focused on predicting ship besetting in Arctic waters using Bayesian belief networks. Faury and Cariou (2016) examined the competitiveness of the Northern Sea Route for oil tankers, developing a decision system for the optimal route from Russia to Asia. Afenyo et al, (2017) also used Bayesian Networks to model collisions between icebergs and ships in Arctic waters. Gunnarsson (2021) explored the latest ship traffic and emerging maritime trends on the Northern Sea Route. Andersson et al. (2021) predicted icing in the Arctic region using deep neural networks. Study conducted by Kandel and Baroud (2024), focusing on eight classes of accident types, which were then further reduced from a total of 81 to 10 sub-classes. These categories included Equipment Failure, Loss of Control, Grounding/ Stranding, Collision, Contact, Fire/Explosion, Damage to Ship/Equipment, Capsizing/Listing, and Non-incidental Event. The study explored various machine learning techniques, although deep learning was not among them.

Other than the studies focused on maritime safety in Arctic waters, the literature include various methodologies and hybrid models for analysing marine accidents, with an emphasis on human factors and probabilistic approaches. Uğurlu et al. (2020a) analysed collision, grounding, and sinking accidents in the Black Sea using the Human Factors Analysis and Classification System (HFACS) combined with Bayesian Networks, emphasized the importance of geographic and human factors in accident formation and suggested a Bayesian Network model to use preventing accidents in the Black Sea. Sarıalioğlu et al. (2020) adopted a hybrid approach involving HFACS and fuzzy fault tree analysis (FFTA) to analyse engine-room fires on ships, highlighted contributing factors such as ship age and mechanical fatigue. Uğurlu et al. (2020b) used Bayesian Networks and Chi-square methods to investigate fishing vessel accidents, identified a significant relationship between accident severity, vessel length, and loss of life, proposed a network model to predict the occurrence of fishing vessel accidents. Özaydın et al. (2022) utilized a hybrid model integrating Bayesian Networks and Association Rule Mining (ARM) to analyse unreported occupational accidents on Turkish fishing vessels, focusing on latent factors, active failures, and environmental conditions influencing accident occurrence. Göksu et al. (2024) conducted a probabilistic assessment of ship blackout incidents using Fault Tree Analysis (FTA) and Bayesian Networks, identifying critical causes such as voltage regulator failures and mechanical faults. These studies, focusing various areas other than Arctic waters, collectively underlined the importance of hybrid models and probabilistic approaches in understanding and mitigating risks in marine accidents, with an emphasis on human error, environmental conditions, and machinery failures.

This study focused on effectively modelling and predicting ship accidents and their outcomes. In high-risk scenarios, taking appropriate measures can ensure safer navigation, thereby preventing potential accidents and enhancing safety sustainability (Uğurlu et al., 2016; Yıldırım et al., 2019; Yildiz et al., 2022). Due to the challenges and limitations of mathematical modelling in simulating ship accidents, AI-based approaches, especially deep learning, real-time analysis of big data, and deep neural networks, offer significant potential (Erol et al., 2018; Fu et al., 2022; Kim and Park, 2023). Hence, this study employs deep learning techniques using neural networks developed and tested on large datasets. Furthermore, for the first time, a deep learning model has been proposed to predict whether accidents in the Arctic region will result in oil spills or physical damages in this study. The findings provide valuable outputs for insurance companies in calculating premiums, and also in creating emergency response plans, determining the locations of search and rescue stations, and developing action plans. The results of this study will significantly contribute to the effective management and reduction of risks associated with ship accidents, thereby helping create a safer and more sustainable future in the maritime sector. The deep learning method used in this study, although time-consuming due to training on numerous datasets, results in models that once trained and shared, can successfully operate in new situations and produce accurate results. With the presented model, current meteorological data and the physical values of the ship can be used to accurately predict whether an accident leading to oil pollution will occur.

The paper is structured as follows: after the introduction section, the dataset and methodology in the second section describe the data collection process and the development of the predictive model used for analysing ship accidents in the Arctic region. Then, in the model setup and training section, the architecture and training process of the deep neural network are detailed. The performance measurement section outlines the evaluation metrics and validation approach used to assess the model's effectiveness in the third section. Finally, the results section presents the experimental findings, and a conclusion section summarizes the key insights, implications, and recommendations for future research.

Dataset and methodology

Dataset

In this study, ship accident data created by the Protection of Arctic Marine Environment (PAME) covering the years 2005-2017 were utilized (PAME, 2023). The mentioned data encompass records from Canada, Russia, Iceland, Denmark, Norway, and the United States (TSBC, 2024; EMCIP, 2024; GISIS, 2024; DMAIB, 2024; NSIA, 2024; Homeport USCG, 2024). This study specifically used accidents that occurred above the 58-degree latitude. The dataset included variables such as vessel length, vessel age, vessel tonnage, latitude, longitude, consequence of incident, and accident date. While vessel flag and vessel type were also part of the dataset, they were examined not for model training but for gaining insights about the data. The aim of the study was to develop a predictor, or classifier, that determines whether an oil pollution incident will occur or not, thus the 'Consequence of Incident' value was used as the output feature to differentiate between two classes (oil-related accidents and other maritime incidents occurred or not occurred). Since this study focused on the occurrence (whether it occurs or not) of marine casualties and pollutions the dataset did not specifically categorize types of marine oil pollutants as separate variables, future extensions of the study could incorporate this detail if relevant data becomes available. Such categorization would allow for the demonstration of the impact of different pollutant types on arctic marine ecosystem which may be useful to develop pollutant type specific pollution response strategies. Figure 1 presents the distribution of accidents according to vessel flags, types of accidents, vessel types, and vessel age brackets on a map.

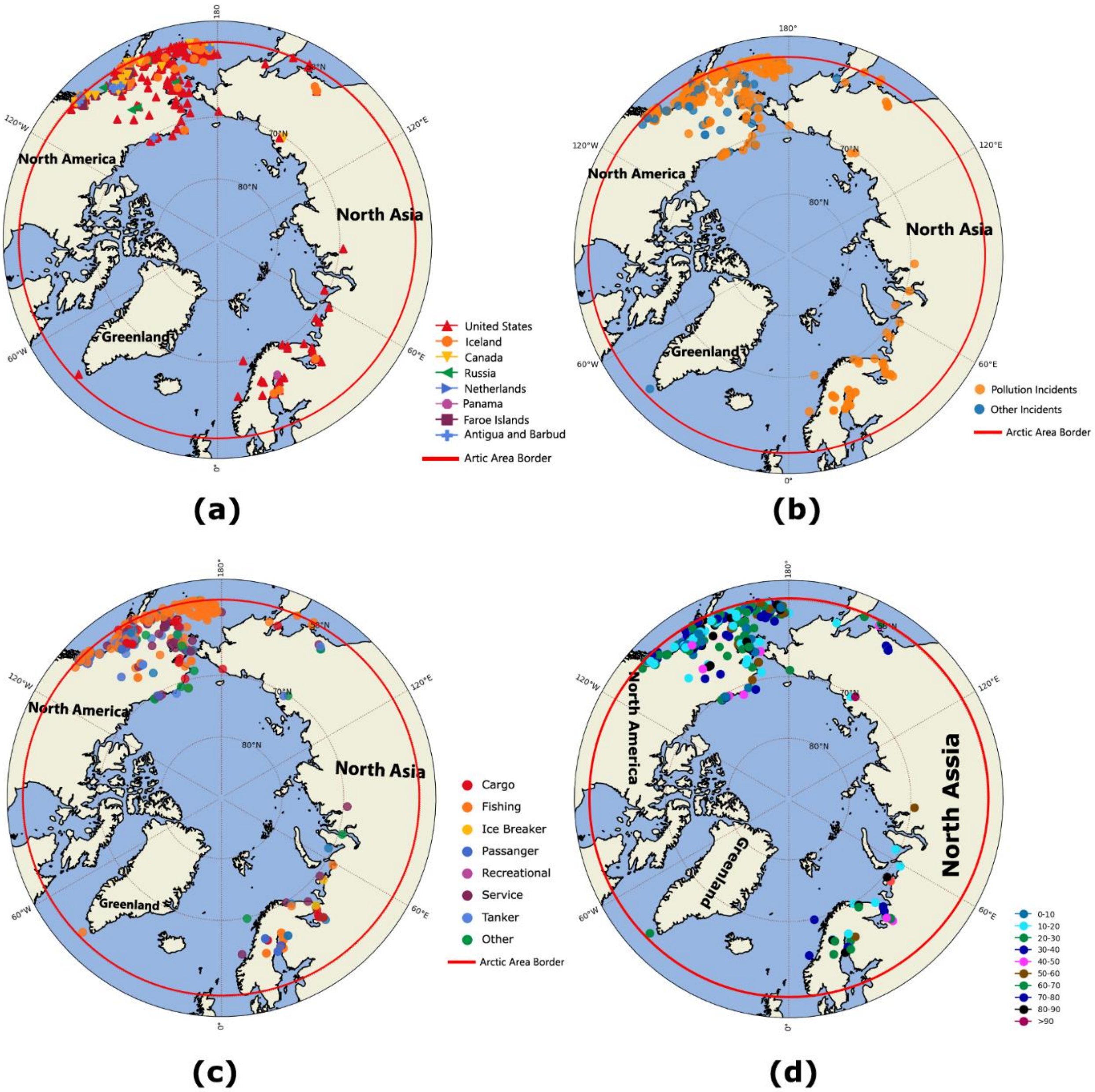


Figure 1. Map representation of accidents; (a) Distribution by vessel flag, (b) Distribution by types of accidents, (c) Distribution by vessel types, (d) Distribution by vessel age brackets.

As can be seen from the maps in Figure 1, maritime accidents have occurred over a wide geographic area. In the dataset, the values for latitude, longitude, and accident date were used to retrieve weather data for the week prior to each accident from https://open-meteo.com/. Table 1 displays the variables used as input features in the construction of the model.

As output of the system, the values of Consequence of Incident were used. These values are divided into two categories: Marine Casualty and Discharge of Oil. In collision accidents, if one of the ships has a Discharge of Oil, it is assumed that both records have Discharge of Oil.

Scatter plots are graphs that clearly show how two-dimensional data is distributed. Because of this feature, they are very useful in classification problems to see whether a linear classifier can be used (Schulz et al., 2019). Figure 2 contains six scatter diagrams selected from ship tonnage, age, and temperature values. When the data are examined in pairs on the scatter diagrams, it can be seen that the data cannot be separated linearly. This indicates that the data can only be separated by a non-linear classifier (Ghosh et al., 2019).

This graph also indicates that the feature values are close to each other, making separation difficult. All these observations have led us to use deep neural networks, a non-linear classifier, for classification.

Model Setup and Training

In this study, a network structure trained with deep neural network architecture was used. A deep neural network is a general term for networks that have many hidden layers with special transfer functions and optimization methods (Abdolrasol et al., 2021).

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction (Goodfellow et al., 2016; LeCun et al., 2015). To test the success of the generated models, the F1 score was used. The F1 score is a method often used in situations where class values are imbalanced, meaning one class has many elements while the other class has few elements (Wardhani et al., 2019).

Deep Neural Network Structure

In the study, a neural network structure with two hidden layers containing 500 and 100 cells respectively was developed. The activation functions for these layers are Relu and Tanh, respectively. The Sigmoid function was chosen for the output layer (Szandała, 2021; Nwankpa et al., 2018). The formulas for Relu (1), Tanh (2), and Sigmoid (3) are expressed as follows.

Table 1. Features used for training the neural network

Feature Name	Explanation
Vessel Length	The length of ship in feet (ft).
Vessel Age	Ship's age in years.
Vessel Tonnage	Ships capacity in gross tonnage.
Last Day Lowest Temperature	Lowest temperature at the accident site on the day of accident (centigrade)
Last Day Highest Temperature	Highest temperature at the accident site on the day of accident (centigrade)
Last Day Average Temperature	Average temperature at the accident site on the day of accident (centigrade)
Weekly Lowest Temperature	Lowest temperature at the accident site in the week of accident (centigrade)
Weekly Highest Temperature	Highest temperature at the accident site in the week of accident (centigrade)
Weekly Average Temperature	Average temperature at the accident site in the week of accident (centigrade)

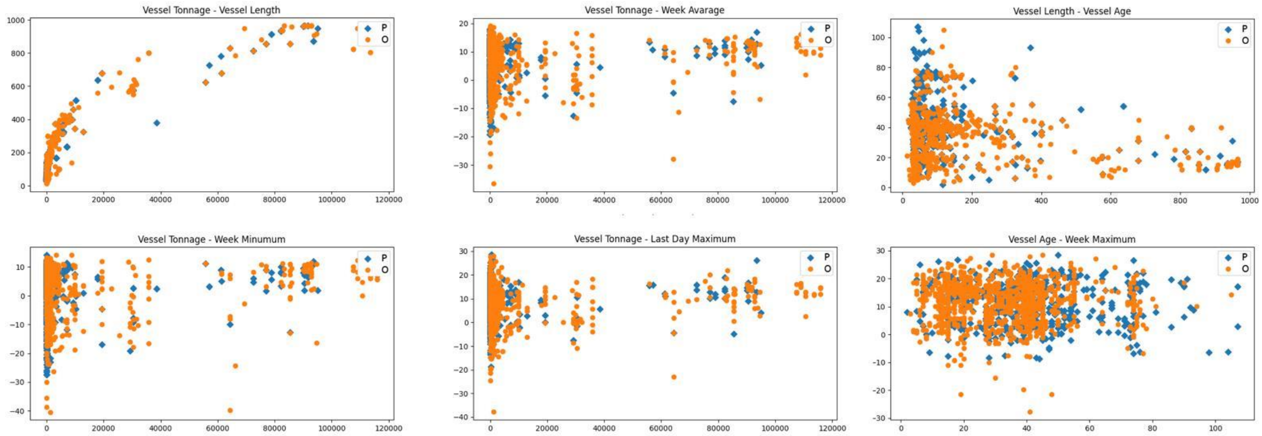


Figure 2. Scatter values for marine casualty - discharge of oil for sample input values.

$$f(x) = \boldsymbol{max}(0, x) = \frac{x + |x|}{2} = \begin{cases} x \; if \; x > 0 \\ 0 \; otherwise \end{cases} \quad (1)$$

$$f(x) = \boldsymbol{tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2)$$

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

The graphical representation of these functions are shown in Figure 3.

During the training of the network, Adaptive Moment Estimation (Adam) optimization was used. This optimization technique is one of the most commonly used methods in modern deep network architectures. The binary cross-entropy method was used as the loss function (Wang and Lo, 2020). The formula for this function is shown in Eq. 4, where y is the label (1 for green points and 0 for red points) and p(y) is the predicted probability of the point being green for all N points.

$$\mathrm{H_p(q)} = -\frac{1}{N}\sum_{i=1}^{N} \mathrm{yi.\log(p(yi)) + (1 - yi).\log(1 - p(yi))} \quad (4)$$

The configuration of the network developed for four inputs is shown in the Figure 4. The intermediate layer structure and the number of neurons is the same for all input combinations

K-Fold Cross Validation

In the machine learning approach, the dataset is generally divided into two parts: training and testing. Then, the training set is further divided into sub-training and validation sets (Figure 5). The system is trained on the sub-training data, and the success of this model is measured on the validation data set. This process can be repeated until the desired success rate is achieved. Finally, the model that has achieved the desired success is applied to the test data to obtain the actual performance. The true performance is the success on this test data.

There are different approaches to training the model with sub-training and validation sets. In this study, one of the most commonly used methods, the K-fold cross-validation approach, was adopted (Alrumaidhi et al., 2023). In this approach, the dataset is divided into k equal parts. Each time, $k-1$ parts are

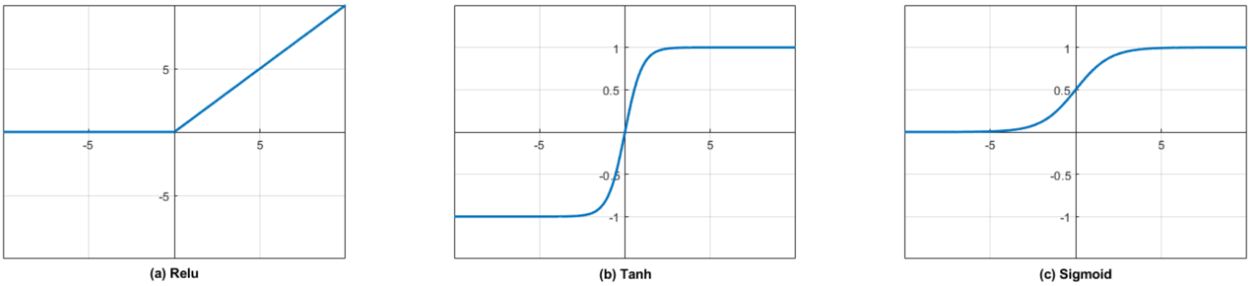


Figure 3. Activation functions.

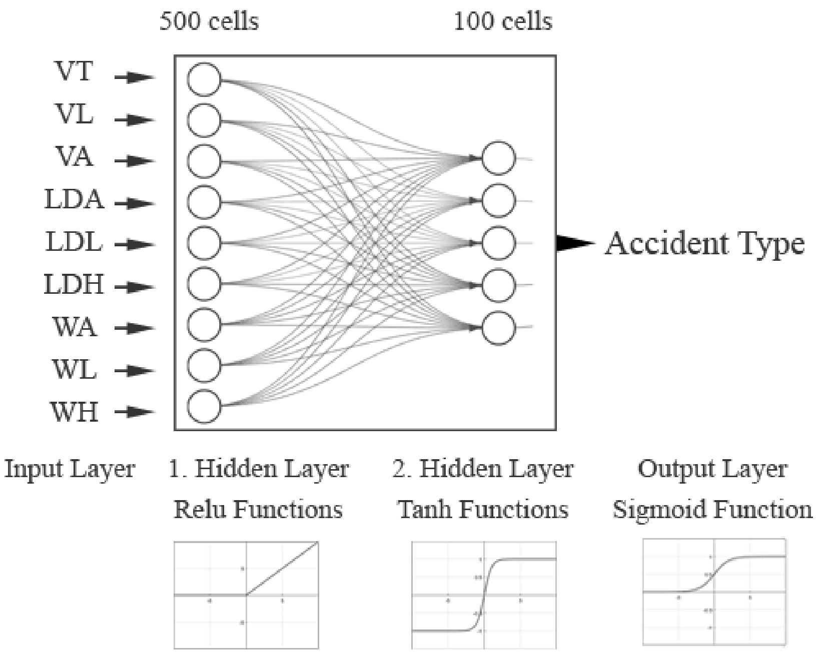


Figure 4. The architecture of the network generated for the input.

used for sub-training and 1 part is used for validation (Pal and Patel, 2020). The average of the validation successes obtained at each step is considered the overall performance of the model.

Performance Measurement

The F1 score, developed by Van Rijsbergen (1979), is a reliable measure frequently used in statistical analysis. The F1 score was also employed in this study, relying on four measurements: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). True positive denotes instances where the model correctly diagnoses positivity, and true negative represents accurate negative diagnoses. False positives occur when the model wrongly identifies positivity, whereas false negatives arise from incorrect negative diagnoses. The F1 score calculation involves two essential metrics: precision (5) and sensitivity (6). And also flow chart of the methodology is presented in Figure 6.

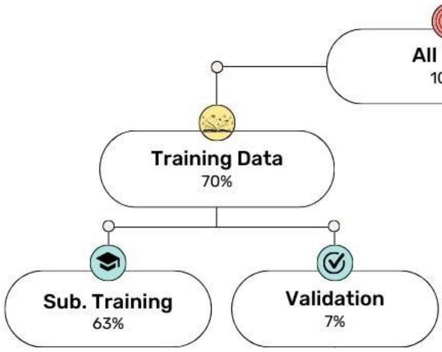


Figure 5. Splitting the dataset.

Step 1. Dataset Collection
Collecting accident data from databases covering the years 2005-2017, in total 511 data from various sources

Step 2. Data Preprocessing
Cleaning and preprocessing the data to ensure data quality.

Step 3. Feature Selection
Selecting features for model training; vessel length, age, tonnage, and weather conditions, etc.

Step 4. Data Visualization
Visualizing data distribution using scatter plots and maps to reveal relationships between variables.

Step 5. Model Setup and Training
Developing a deep neural network with two hidden layers containing 500 and 100 cells.

Step 6. K-Fold Cross-Validation
Training the model on sub-training data and validating on validation data to optimize model performance; 70% of the total dataset (358) used for training and 30% of the dataset (153) for with 1000 trials each.

Step 7. Model Evaluation
Evaluating model performance using the F1 score, which balances precision and sensitivity.

Step 8. Performance Analysis
Analysing the results to identify the most critical features affecting model performance.

Step 9. Model Application
Applying the trained model to predict whether an accident will result in oil pollution or other incidents.

Figure 6. Flow chart of the methodology.

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Sensivity = \frac{TP}{TP + FN} \quad (6)$$

$$Accuracy = \frac{TN + TP}{TP + FN + FP + TN} \quad (7)$$

Utilizing these considerations, the F1 criterion is computed using formula (8).

$$F1 = 2 * \frac{Precision * Sensivity}{Precision + Sensivity} = \frac{2 * TP}{2TP + FP + FN} \quad (8)$$

Results

The source code was written in Python using the Keras library (Lynch, 2023). The data was split into two parts: 70% for training and 30% for testing. The training data was used to train the model with a 10-fold cross-validation approach for $k = 10$. Experiments were conducted for 511 different input-output combinations of the nine input values (from 1 to 9) shown in Table 1. Figure 7 shows the changes in F1 and accuracy metrics during training for a sample with nine inputs.

For each combination, 1000 experiments were conducted, and the average and standard deviation of the F1 scores on the test data were reported. The average test F1 values, the standard deviations of the F1 values, and the average accuracy values of the networks with a single input are presented in Table 2.

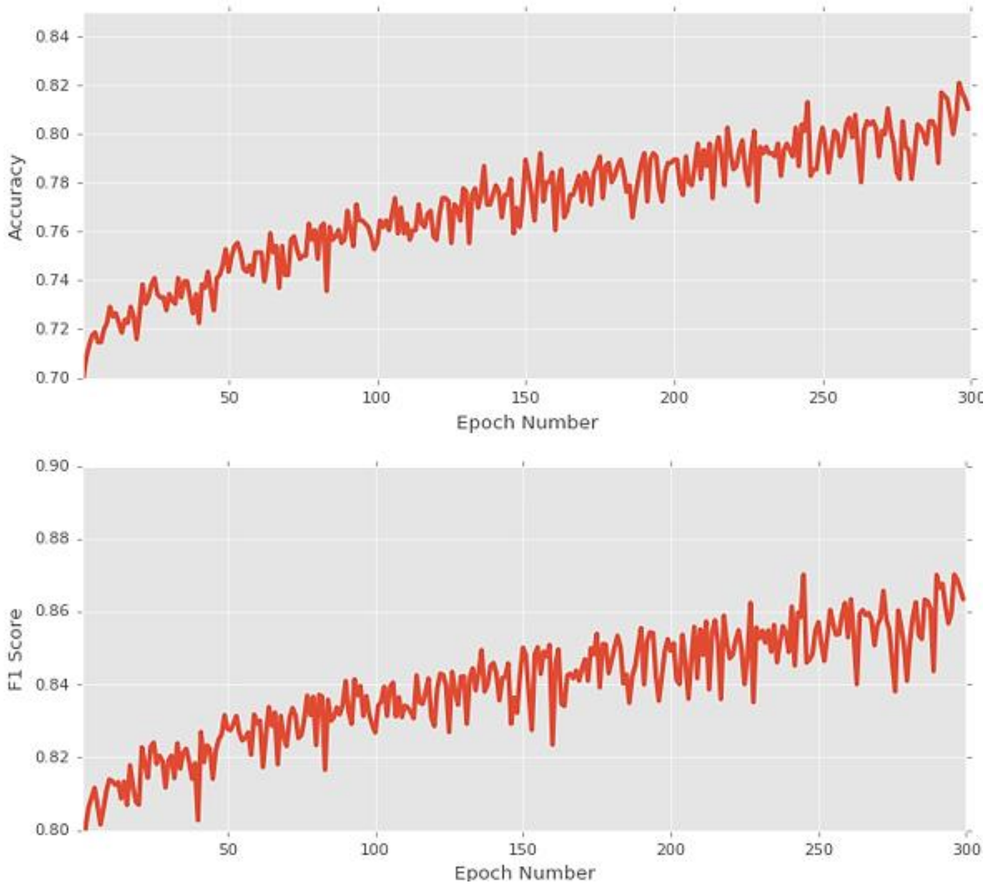


Figure 7. F1 and accuracy values obtained during a sample training session.

Table 2. Performance metrics for single-input deep networks

Feature	Average Test F1	Standard Deviation	Average Accuracy
VT	0.28	0.39	0.12
VL	0.29	0.52	0.14
VA	0.31	0.45	0.15
WL	0.35	0.40	0.16
WH	0.32	0.39	0.16
WA	0.33	0.40	0.13
LDL	0.32	0.39	0.14

VT: Vessel Tonnage, VL: Vessel Length, VA: Vessel Age, WL: Week Low, WH: Week High, WA: Week Average, LDL: Last Day Low, LDH: Last Day High, LDA: Last Day Average

From these values, it can be seen that the highest performance is given by the Vessel Age (VA) value. However, this performance is very low. Each feature alone has a high standard deviation value and is not sufficient to successfully represent the system on its own. Figure 8 shows the average F1, accuracy, and standard deviation values obtained by removing each attribute individually from the entire set of attributes.

As expected, the combination using all features (VT, VL, VA, LDA, LDL, LDH, WA, WL, WH) achieved the highest success rate. Additionally, the standard deviation value is quite low. When the WA and WH parameters were removed, the success rate decreased by 0.89 to 0.87, and the standard deviation value approximately doubled. This indicates that the obtained success became more unstable.

Subsequently, with the removal of the Weekly Lowest Temperature (WL) parameter, the success rate increased by 0.03 to 0.84, and the standard deviation increased to 0.35. A dramatic fall in success was observed with the removal of the LDH parameter. Removing the SGED value caused the success rate to decline to 0.78.

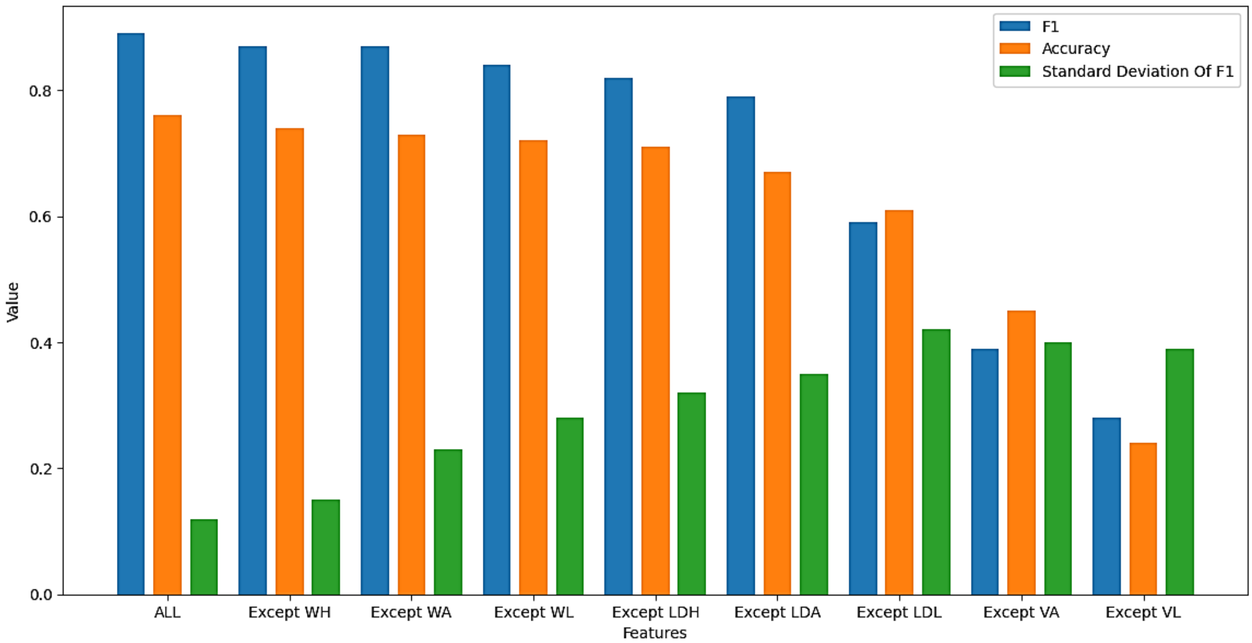
In general, as examined the situation, and also as expected, removing parameters decreased the success rate and increased the standard deviation, in other words, the instability.

As shown in Figure 8, when all input parameters are used, it can be predicted the likelihood of an environmental pollution or physical malfunction occurring in a ship accident in the Arctic region with an 89% F1 success rate.

Kandel and Baroud (2024) have used various machine learning models in their studies. They developed models to predict the type of damage a ship accident would cause based on the damage to the ship. In their study, a maximum success rate of 0.514 was achieved for 8 class values. In contrast, this study focused on predicting the environmental damage caused by the accident and achieved a maximum F1 success rate of 0.89 for 2 classes using deep learning.

Figure 8 shows the F1, Accuracy, and Standard Deviation of F1 values together. Generally, it can be observed that as features are removed, the success rate decreases, and the standard deviation increases.

It is notable that when the LDL value is also removed from the features, although the success rate decreases, the standard deviation does not increase. This is a thought-provoking situation. It can be explained by the fact that the system remains consistently at a



Feature Names (X: used, -: not used)									Average Test Ac	Average Test F1	Standard Deviation
VT	**VL**	**VA**	**LDL**	**LDA**	**LDH**	**WL**	**WA**	**WH**			
X	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	0.76	**0.89**	0.12
X	**X**	**X**	**X**	**X**	**X**	**X**	**X**	-	0.74	0.87	0.15
X	**X**	**X**	**X**	**X**	**X**	**X**	-	-	0.73	0.87	0.23
X	**X**	**X**	**X**	**X**	**X**	-	-	-	0.72	0.84	0.28
X	**X**	**X**	**X**	**X**	-	-	-	-	0.71	0.82	0.32
X	**X**	**X**	**X**	-	-	-	-	-	0.67	0.79	0.35
X	**X**	**X**	-	-	-	-	-	-	0.61	**0.59**	0.42
X	**X**	-	-	-	-	-	-	-	0.45	0.39	0.40
X	-	-	-	-	-	-	-	-	0.24	0.28	0.39

Figure 8. Top six input combinations by performance metrics, accuracy and standard deviation values obtained by removing features.

lower performance level and high performance is less frequently achieved, leading to a more stable (but lower-performing) model.

The study presented the feasible application of deep learning methods to predict oil pollution incidents from ship accidents in the Arctic region with a high degree of accuracy. The developed neural network model achieved a maximum F1 score of 0.89, which indicates strong predictive performance when all input features were utilized. Importantly, the results highlight the critical role of specific features such as vessel age, vessel length, and environmental conditions in determining the likelihood of pollution incidents. Removing key features not only reduced the F1 score but also increased the standard deviation, emphasizing the instability introduced by excluding relevant parameters. The results suggest that the methodology and insights may serve multiple practical purposes as follows.

Predictive modelling: The developed model can be utilized by maritime regulatory bodies and insurers to assess risk levels and prioritize monitoring efforts in high-risk scenarios.

Multi factor evaluation: The analysis of feature importance provides actionable insights for improving data collection practices, particularly emphasizing the inclusion of high-impact features like weather conditions and vessel characteristics in future datasets to use the model for decision-making support tool. The findings support the development of a robust decision-support system that integrates predictive capabilities with real-time data to reduce the number of accidents and mitigate the environmental impact of shipping operations in vulnerable regions like the Arctic. The model can be adapted to various other regions such as narrow straits, close coastal navigation where heavy weather and sea conditions is a prevalent factor.

Higher accuracy: Proposed approach outperformed existing methodologies cited in the literature, such as those by Kandel and Baroud (2024), which achieved a maximum F1 score of 0.514 for a multi-class problem. By focusing on binary classification and leveraging deep learning techniques, this study offers a significantly improved predictive framework for addressing environmental challenges in maritime operations.

Discussion

The findings indicate the promising potential of deep learning methods in predicting environmental risks associated with maritime accidents, particularly in vulnerable regions like the Arctic waters. The developed model demonstrated strong predictive performance, achieving an F1 score of 0.89, which are higher than the earlier research results, i.e. Kandel and Baroud (2024) reported a maximum F1 score of 0.514 when predicting damage categories in maritime accidents using traditional machine learning methods. This contrast highlights the advantage of leveraging deep learning for binary classification tasks, where the focus on specific outcomes such as pollution likelihood enables more precise predictions. The comparison to Kandel and Baroud's work particularly highlights the advantage of the proposed two-stage feature selection process, which effectively captures critical Risk Influential Factors (RIFs), allowing for higher predictive performance and improved model robustness.

The analysis of feature importance revealed the critical role of variables such as vessel age, vessel length, and environmental conditions in the model's success. Removing high-impact features, such as Weekly Average (WA) and Weekly High (WH) temperatures, resulted in notable reductions in the F1 score and increased standard deviation, reflecting heightened instability in model predictions. These findings align with research by Goodfellow et al. (2016), which emphasized the importance of comprehensive feature sets in achieving consistent performance in machine learning models. The findings also corroborate with Feng et al. (2024), who introduced a novel two-stage feature selection approach, combining feature interaction analysis and state-differentiated mutual information, to ensure the stability and robustness of predictive models by retaining the most influential features. However, the removal of the Lowest Daily Level (LDL) did not increase variability despite a reduction in accuracy, suggesting a unique interaction among features. Feng et al. (2024) similarly highlighted the complex interplay between feature redundancy and model stability, underscoring the need for robust feature selection processes.

The presented model can serve as a decision-support tool for maritime regulators, insurers, and operators, aiding in risk assessment and resource allocation for high-risk scenarios. The emphasis on feature importance also highlights the necessity of improving data collection practices in maritime safety, particularly regarding environmental conditions and vessel-specific characteristics. Furthermore, the model's adaptability to diverse operational contexts, such as narrow straits or heavy-weather navigation zones, offers additional utility, as demonstrated in context-aware predictive models for maritime safety (Zhang et al., 2021). The approach parallels the methodology outlined by Feng et al. (2024), wherein the LightGBM model was used as a benchmark for accurately predicting the severity of marine accidents, focusing on a holistic feature selection and stability evaluation framework to optimize decision-support capabilities.

Despite its success, the study's limitations must be acknowledged. The binary classification framework, while effective, simplifies the complexity of real-world scenarios. Future research could explore multi-class problems to capture more granular environmental outcomes. Additionally, the dataset's focus on Arctic conditions limits its generalizability to other regions. Expanding the model to diverse maritime contexts will require further validation and adaptation. Feng et al.

(2024) also pointed out similar limitations in their study. Furthermore, the application of explainable techniques, as suggested by Gunning et al. (2019), could enhance the interpretability of model predictions, making them more accessible to practitioners and regulators.

Conclusion

This study presents the results of experimental analyses conducted on models written in Python using the Keras library, with data split into 70% training and 30% testing, employing a k-fold cross-validation approach. The analyses were performed on 511 different input-output combinations, with 1000 trials for each. The findings indicate that certain features, such as the ship's tonnage, length, and age, play a critical role in predicting ship accidents. The results show that the highest F1 score and lowest standard deviation were achieved when all features were used. Notably, when some features, like the minimum daily temperature in the navigation area, were removed, a significant decrease in model performance was observed, indicating their importance for the model. However, the impact on the model's overall stability was observed as an increasing standard deviation with the removal of more features.

Whilst examining the results in Figure 8 in detail, it is evident that considering the ship's physical size and age independently of the climate results in lower F1 scores and average performance. This indicates that the ship's physical size alone is not sufficient. A model that evaluates data together with daily minimum temperature values produces more reliable results. This suggests that a model produced in conjunction with weather forecasts is more successful in Arctic regions.

Ships traveling to the Arctic region and organizations inspecting these ships can use this model for predictions. It can measure whether accidents will have an environmental impact. This enables organizations to assign ships with lower environmental risks. Moreover, it can significantly contribute to calculating accident risks and determining insurance premiums by insurance companies.

The limitations of the study are related to the scope and quality of the dataset used. Missing or incorrect data in the dataset can affect the model's overall accuracy and reliability. Therefore, the marine casualty investigation reports should be as comprehensive as possible as specified in the revised guidelines of the IMO (IMO, 2014). Additionally, the generalizability of the model may be limited by the diversity of the dataset used. Since this study was conducted on a specific dataset, the model's performance may vary under different datasets or conditions. To address these data quality issues in the future, several specific measures could be implemented. Data augmentation techniques could be employed to artificially increase the size and diversity of the training dataset, thereby enhancing the model's robustness. Additionally, imputation techniques, such as K-Nearest Neighbors (KNN) or Multiple Imputation by Chained Equations (MICE), could be used to handle missing data more effectively, reducing the potential biases introduced by incomplete information. Collaborations with maritime organizations, regulatory bodies, or data providers could also be explored to obtain more granular and diverse datasets, further improving the model's generalizability and resilience to different operational conditions. Integrating real-time data feeds from monitoring systems or sensor networks could provide higher-quality, up-to-date information, contributing to more accurate predictions and a more reliable decision-support tool. Furthermore, sea ice conditions are a critical factor influencing maritime navigation, particularly in Arctic regions. The presence and extent of sea ice can significantly affect vessel manoeuvrability, operational risks, and accident causality. While this study analysed 511 accident reports to explore contributing factors to maritime incidents, the variability in the level of detail across these reports limited the ability to incorporate sea ice conditions as a separate variable. Many reports did not include sufficient detail on ice conditions and its specific role in the accidents, resulting in a difficulty to add it in machine learning models.

For future studies, it is recommended to integrate additional data sources to increase the model's robustness and test its applicability on broader datasets. More extensive tests should be conducted to evaluate the model's effectiveness under different maritime conditions and for various types of ships. By focusing on specific ship types, such as non-ice-class ships, special risk mitigation strategies can be developed for these vessels. Finally, implementing regulations such as the Polar Code and integrating these regulations into model predictions could significantly enhance the model's applicability and provide valuable insights for policymakers. Incorporating specific aspects of the Polar Code, such as requirements related to ship structure, equipment, crew training, and voyage planning, could improve the model's capacity to predict safety risks in Arctic navigation. For instance, integrating structural regulations aimed at strengthening vessels to withstand ice conditions could refine the model's assessment of vessel resilience in polar environments. Likewise, incorporating the Polar Code’s mandatory training standards for crew would enable the model to account for human factors, such as the competency of the crew in handling emergencies unique to polar waters. Moreover, adding voyage planning requirements that consider ice conditions, weather, and available search and rescue resources could improve the predictive accuracy of potential navigational hazards. By including these detailed regulatory elements, the model could assist policymakers in evaluating the effectiveness of current safety measures, identifying areas for regulatory enhancement, and developing informed policies that support safer Arctic maritime operations.

Ethical Statement

This study involves no human or animal subject therefore holds no ethical concerns.

Funding Information

This study received no funding.

Author Contribution

Mengü Demir: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Visualization and Writing -original draft, Sercan Erol: Supervision, Writing -original draft, Writing -review and editing, Serdar Yildiz: Methodology, Visualization, Writing -original draft, Writing -review and editing.

Conflict of Interest

The authors declare that they have no known competing financial or non-financial, professional, or personal conflicts that could have appeared to influence the work reported in this paper.

Acknowledgements

Not applicable.

References

Abdolrasol, M. G. M., Hussain, S. M. S., Ustun, T. S., Sarker, M. R., Hannan, M. A., Mohamed, R., Ali, J. A., Mekhilef, S., & Milad, A. (2021). Artificial neural networks-based optimization techniques: A review. *Electronics, 10*(21), 2689. https://doi.org/10.3390/electronics10212689

Afenyo, M., Khan, F., Veitch, B., & Yang, M. (2017). Arctic shipping accident scenario analysis using Bayesian Network approach. *Ocean Engineering, 133*, 224–230. https://doi.org/10.1016/j.oceaneng.2017.02.002

Allianz. (2024). *Safety and Shipping Review 2024*. In A. G. C. Specialty (Ed.), *An annual review of trends and developments in shipping losses and safety* (pp. 1–40).

Alrumaidhi, M., Farag, M. M., & Rakha, H. A. (2023). Comparative analysis of parametric and non-parametric data-driven models to predict road crash severity among elderly drivers using synthetic resampling techniques. *Sustainability, 15*(13), 9878. https://doi.org/10.3390/su15139878

Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., Law, S., Jones, D. C., Wilkinson, J., & Phillips, T. (2021). Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nature Communications, 12*(1), 5124. https://doi.org/10.1038/s41467-021-25257-4

Bai, J. (2015). The IMO Polar Code: The emerging rules of Arctic shipping governance. *The International Journal of Marine and Coastal Law, 30*(4), 674–699. https://doi.org/10.1163/15718085-12341376

Boylan, B. M. (2021). Increased maritime traffic in the Arctic: Implications for governance of Arctic sea routes. *Marine Policy, 131*, 104566. https://doi.org/10.1016/j.marpol.2021.104566

Chircop, A. (2020). The polar code and the Arctic marine environment: Assessing the regulation of the environmental risks of shipping. *The International Journal of Marine and Coastal Law, 35*(3), 533–569. https://doi.org/10.1163/15718085-BJA10033

Council, A. (2024). *Arctic Shipping Status Report (ASSR) #1: The increase in Arctic shipping 2013–2023*. In Arctic Council (Ed.), *Protection of Arctic Marine Environment (PAME)* (pp. 1–26).

DMAIB. (2024). *Danish Maritime Accident Investigation Board-Reports*. Accessed January 20, 2024. https://dmaib.com/reports

EMCIP. (2024). *European Marine Casualty Information Platform*. European Maritime Safety Agency (EMSA). Accessed January 25, 2024. https://emsa.europa.eu/emcip.html

Erol, S., Demir, M., Çetişli, B., & Eyüboğlu, E. (2018). Analysis of ship accidents in the Istanbul Strait using neuro-fuzzy and genetically optimized fuzzy classifiers. *The Journal of Navigation, 71*(2), 419–436. https://doi.org/10.1017/S0373463317000601

Faury, O., & Cariou, P. (2016). The Northern Sea Route competitiveness for oil tankers. *Transportation Research Part A: Policy and Practice, 94*, 461–469. https://doi.org/10.1016/j.tra.2016.09.026

Feng, Y., Wang, X., Chen, Q., Yang, Z., Wang, J., Li, H., Guaqing, X., & Liu, Z. (2024). Prediction of the severity of marine accidents using improved machine learning. *Transportation Research Part E: Logistics and Transportation Review, 188*, 103647. https://doi.org/10.1016/j.tre.2024.103647

Fu, S., Yu, Y., Chen, J., Xi, Y., & Zhang, M. (2022). A framework for quantitative analysis of the causation of grounding accidents in Arctic shipping. *Reliability Engineering & System Safety, 226*, 108706. https://doi.org/10.1016/j.ress.2022.108706

Fu, S., Zhang, D., Montewka, J., Yan, X., & Zio, E. (2016). Towards a probabilistic model for predicting ship besetting in ice in Arctic waters. *Reliability Engineering & System Safety, 155*, 124–136. https://doi.org/10.1016/j.ress.2016.06.010

Ghosh, S., Dasgupta, A., & Swetapadma, A. (2019). A study on support vector machine-based linear and non-linear pattern classification. In *2019 International Conference on Intelligent Sustainable Systems (ICISS)* (pp. 24–28). IEEE. https://doi.org/10.1109/ISS1.2019.8908018

GISIS. (2024). *Global Integrated Shipping Information System, Marine Casualties and Incidents Module*. International Maritime Organization (IMO). Accessed January 20, 2024. https://gisis.imo.org/Public/MCIR/Search.aspx

Göksu, B., Şakar, C., & Yüksel, O. (2024). A probabilistic assessment of ship blackout incidents with Fault Tree Analysis (FTA) and Bayesian Network (BN). *Journal of Marine Engineering & Technology, 1–16*. https://doi.org/10.1080/20464177.2024.2423425

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Gunnarsson, B. (2021). Recent ship traffic and developing shipping trends on the Northern Sea Route—Policy implications for future Arctic shipping. *Marine Policy, 124*, 104369. https://doi.org/10.1016/j.marpol.2020.104369

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics, 4*(37), eaay7120.

https://doi.org/10.1126/scirobotics.aay7120

Hebbar, A. A., Schröder-Hinrichs, J.-U., Mejia, M. Q., Deggim, H., & Pristrom, S. (2020). The IMO regulatory framework for Arctic shipping: Risk perspectives and goal-based pathways. In *Governance of Arctic Shipping: Rethinking Risk, Human Impacts, and Regulation* (pp. 229–247). https://doi.org/10.1007/978-3-030-44975-9

Homeport USCG. (2024). *United States Coast Guard: Marine casualty and pollution data for researchers*. Accessed January 9, 2024. https://homeport.uscg.mil/Lists/Content/DispForm.aspx?&ID=211&Source=https://homeport.uscg.mil/missions/investigations/marine-casualty-pollution-investigations

IMO. (2014). *International Maritime Organization: Casualty-related matters: Reports on marine casualties and incidents: Revised harmonized reporting procedures – Reports required under SOLAS regulations I/21 and XI-1/6, and MARPOL, articles 8 and 12*. International Maritime Organization. Accessed May 3, 2024. https://wwwcdn.imo.org/localresources/en/OurWork/MSAS/Documents/MSC-MEPC3/MSC-MEPC.3-Circ.4%20Rev%201%20%20Revised%20harmonized%20reporting%20procedures%20-%20Reports%20required%20under%20SOLAS%20regulations%20I21.pdf

Kandel, R., & Baroud, H. (2024). A data-driven risk assessment of Arctic maritime incidents: Using machine learning to predict incident types and identify risk factors. *Reliability Engineering & System Safety, 243*, 109779. https://doi.org/10.1016/j.ress.2023.109779

Kim, J.-K., & Park, D.-J. (2023). Determining the proper times and sufficient actions for the collision avoidance of navigator-centered ships in the open sea using artificial neural networks. *Journal of Marine Science and Engineering, 11*(7), 1384. https://doi.org/10.3390/jmse11071384

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lynch, S. (2023). *Python for scientific computing and artificial intelligence*. Chapman and Hall/CRC.

NSIA. (2024). *Norwegian Safety Investigation Authority: Marine Published Reports*. Accessed February 10, 2024. https://nsia.no/Marine/Published-reports

Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*. https://doi.org/10.48550/arXiv.1811.03378

Özaydın, E., Fışkın, R., Uğurlu, Ö., & Wang, J. (2022). A hybrid model for marine accident analysis based on Bayesian Network (BN) and Association Rule Mining (ARM). *Ocean Engineering, 247*, 110705. https://doi.org/10.1016/j.oceaneng.2022.110705

Pal, K., & Patel, B. V. (2020). Data classification with k-fold cross-validation and holdout accuracy estimation methods with 5 different machine learning techniques. In *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 83–87). IEEE. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00016

PAME. (2023). *Protection of the Arctic Marine Environment*. https://pame.is/

Sakhuja, V. (2014). The polar code and Arctic navigation. *Strategic Analysis, 38*(6), 803–811. https://doi.org/10.1080/09700161.2014.952943

Sarıalioğlu, S., Uğurlu, Ö., Aydın, M., Vardar, B., & Wang, J. (2020). A hybrid model for human-factor analysis of engine-room fires on ships: HFACS-PV&FFTA. *Ocean Engineering, 217*, 107992. https://doi.org/10.1016/j.oceaneng.2020.107992

Schulz, A., Hinder, F., & Hammer, B. (2019). Deepview: Visualizing classification boundaries of deep neural networks as scatter plots using discriminative dimensionality reduction. *arXiv preprint arXiv:1909.09154*. https://doi.org/10.48550/arXiv.1909.09154

Szandała, T. (2021). Review and comparison of commonly used activation functions for deep neural networks. In *Bio-inspired Neurocomputing* (pp. 203–224). https://doi.org/10.1007/978-981-15-5495-7_11

TSBC. (2024). *Transportation Safety Board of Canada: Marine transportation safety investigations and reports*. Government of Canada. Accessed January 25, 2024. https://www.tsb.gc.ca/eng/rapports-reports/marine/index.html

Uğurlu, F., Yıldız, S., Boran, M., Uğurlu, Ö., & Wang, J. (2020b). Analysis of fishing vessel accidents with Bayesian Network and Chi-square methods. *Ocean Engineering, 198*, 106956. https://doi.org/10.1016/j.oceaneng.2020.106956

Uğurlu, Ö., Erol, S., & Başar, E. (2016). The analysis of life safety and economic loss in marine accidents occurring in the Turkish Straits. *Maritime Policy & Management, 43*(3), 356–370. https://doi.org/10.1080/03088839.2014.1000992

Uğurlu, Ö., Yıldız, S., Loughney, S., Wang, J., Kuntchulia, S., & Sharabidze, I. (2020a). Analyzing collision, grounding, and sinking accidents occurring in the Black Sea utilizing HFACS and Bayesian networks. *Risk Analysis, 40*(12), 2610–2638. https://doi.org/10.1111/risa.13568

Van Rijsbergen, C. (1979). Information retrieval: Theory and practice. In *Proceedings of the Joint IBM/University of Newcastle Upon Tyne Seminar on Data Base Systems* (pp. 1–14).

Wang, H., & Lo, J. T. H. (2020). Adaptively solving the local-minimum problem for deep neural networks. *arXiv preprint arXiv:2012.13632*. https://doi.org/10.48550/arXiv.2012.13632

Wardhani, N. W. S., Rochayani, M. Y., Iriany, A., Sulistyono, A. D., & Lestantyo, P. (2019). Cross-validation metrics for evaluating classification performance on imbalanced data. In *2019 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)* (pp. 14–18). IEEE. https://doi.org/10.1109/IC3INA48034.2019.8949568

Yıldırım, U., Başar, E., & Uğurlu, Ö. (2019). Assessment of collisions and grounding accidents with human factors analysis and classification system (HFACS) and statistical methods. *Safety Science, 119*, 412–425. https://doi.org/10.1016/j.ssci.2017.09.022

Yildiz, S., Uğurlu, Ö., Loughney, S., Wang, J., & Tonoğlu, F. (2022). Spatial and statistical analysis of operational conditions influencing accident formation in narrow waterways: A case study of Istanbul Strait and Dover Strait. *Ocean Engineering, 265*, 112647. https://doi.org/10.1016/j.oceaneng.2022.112647

Zhang, L., Wang, X., & Chen, Q. (2021). Context-aware predictive models for maritime safety. *Marine Policy, 130*, 104567.