Feng, Y, Wang, X, Chen, Q, Yang, Z, Wang, J, Li, H, Xia, G and Liu, Z

Prediction of the severity of marine accidents using improved machine learning

Article

Feng, Y, Wang, X, Chen, Q, Yang, Z, Wang, J, Li, H, Xia, G and Liu, Z (2024) Prediction of the severity of marine accidents using improved machine learning. Transportation Research Part E: Logistics and Transportation Review. 188. ISSN 1366-5545

Transportation Research Part E 188 (2024) 103647

Contents lists available at ScienceDirect

Transportation Research Part E

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

Prediction of the severity of marine accidents using improved machine learning

Yinwei Feng [a,b], Xinjian Wang [a,b,c,d,**], Qilei Chen [e], Zaili Yang [d,f,*], Jin Wang [d], Huanhuan Li [d], Guoqing Xia [a,b], Zhengjiang Liu [a,b]

[a] *Navigation College, Dalian Maritime University, Dalian 116026, PR China*
[b] *Key Laboratory of Navigation Safety Guarantee of Liaoning Province, Dalian 116026, PR China*
[c] *Seafarers Research Institute, Dalian Maritime University, Dalian 116026, PR China*
[d] *Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool L3 3AF, UK*
[e] *Miner School of Computer & Information Sciences, University of Massachusetts Lowell, Lowell, MA 01854, USA*
[f] *Transport Engineering College, Dalian Maritime University, Dalian 116026, PR China*

ARTICLE INFO

Keywords:
Maritime transportation
Marine accident
Machine learning
Mutual information
Feature selection
LightGBM

ABSTRACT

Although many studies have focused on the occurrence likelihood of marine accidents, few have focused on the analysis of the severity of the consequences, and even fewer on the prediction of the severity. To this end, a new research framework is proposed in this study to accurately predict the severity of marine accidents. First, a novel two-stage feature selection (FS) method was developed to select and rank Risk Influential Factors (RIFs) to improve the accuracy of the Machine Learning (ML) model and interpretability of the FS. Second, a comprehensive evaluation method is proposed to measure the performance of the FS methods based on stability, predictive performance improvement, and statistical tests. Third, six well-established ML models were used and compared to measure the performance of different predictors. The Light Gradient Boosting Machine (LightGBM) was found to have the best predictive performance for the severity prediction of marine accidents and was treated as the benchmark model. Finally, LightGBM was used to predict accident severity based on the RIFs selected by the proposed FS method, and the effect of risk control measures was counterfactually analysed from a quantitative perspective. This innovative study on the use of improved ML approaches can effectively analyse and predict the severity of marine accidents, providing a novel methodology for and triggering a new direction for using Artificial Intelligence (AI) technologies in safety assessment and accident prevention studies. **The source code is publicly available at:** https://github.com/FengYinLeo/PGI-SDMI.

1. Introduction

1.1. Background

Despite significant efforts in maritime safety assurance, the frequency of marine accidents has not changed significantly over the

* Corresponding author at: Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool L3 3AF, UK.
** Corresponding author at: Navigation College, Dalian Maritime University, Dalian 116026, PR China.
E-mail addresses: wangxinjian@dlmu.edu.cn (X. Wang), Z.Yang@ljmu.ac.uk (Z. Yang).

https://doi.org/10.1016/j.tre.2024.103647
Received 18 October 2023; Received in revised form 28 May 2024; Accepted 25 June 2024
Available online 2 July 2024
1366-5545/

past decade (Cao et al., 2023; Fang et al., 2024). A marine accident is any or a series of events during a ship's operations that threaten the vessel, occupants, environment, or others, excluding marine casualties (Yan et al., 2023). Between 2014 and 2020, the European Maritime Safety Agency (EMSA) reported 22,532 marine accidents in Europe involving 8,015 ships, causing 6,921 injuries (Cao et al., 2023). Marine accidents have significant consequences on individuals, cargo, vessels, and marine ecosystems, resulting in varying degrees of harm and financial setbacks. Many studies (Rawson et al., 2021; Wang and Yang, 2018; Weng and Yang, 2015) state that maritime safety has become a more critical concern. Therefore, prompt corrective measures must be undertaken to ensure safety at sea, as underscored by the International Safety Management (ISM) code and the International Convention for the Safety of Life at Sea (SOLAS).

In recent years, machine learning (ML) has been used to analyse marine accidents and improve navigation safety and efficiency with the development of Artificial Intelligence (AI) technology (Chen et al., 2020; Xin et al., 2024). Its application is indispensable for accident prevention and establishment of a secure maritime environment. ML is a method of training computers to gain knowledge and expertise through self-learning and constantly enhance their abilities. ML has proven to be more effective in identifying Risk Influential Factors (RIFs) (Weng and Yang, 2015) and predicting accidents (Lan et al., 2023) than traditional research methods. However, its current applications focus predominantly on accident frequency analysis and RIFs identification, neglecting its potential to predict marine accident severity and subsequent risk control measures. In particular, the use of a subset of features to more accurately predict results has become a key scientific issue in improving the prediction of marine accidents. Incorporating the principles and guidelines outlined in the ISM Code and SOLAS Convention into ML-based predictive models can further enhance their effectiveness in ensuring maritime safety and preventing accidents.

1.2. Related work and research gaps

1.2.1. Application of machine learning in marine accident prediction

Extensive research has been conducted to thoroughly analyse the RIFs of marine accidents and predict their likelihood. This comprehensive analysis considers numerous RIFs, including the crew's intentions and physical and mental conditions, the ship's navigational state and performance, and the surrounding waterway and weather conditions (Cao et al., 2023; Wang et al., 2021a; Wang et al., 2022). Sotiralis et al. (2016) employed a Bayesian Network (BN) to scrutinise the acts of the Officer on Watch (OOW) and identify the key RIFs responsible for human errors in marine accidents. Hu et al. (2020) analysed the impact of human error on marine accidents. The analysis considered perception, decision-making, and execution using a BN. Liu et al. (2021a) analysed marine accidents off the coast of China using a data-driven BN. The analysis helped to identify the critical RIFs and their relationships that impact the severity of marine accidents. These findings provide valuable information for local maritime authorities and other stakeholders. Zheng et al. (2020) utilised a Support Vector Machine (SVM) to forecast the likelihood of ship collisions. This research presented risk assessment outcomes for distinct ship conditions and aided ships in autonomously making collision avoidance decisions and actions. Rawson et al. (2021) utilised ML models to forecast the likelihood of shipping accidents under severe weather conditions. These findings indicated that ML models exhibit a promising potential for accident prediction. Zhou et al. (2023) employed a rules-based BN to identify the primary risk factors within the cruise industry, pinpointing "ship accidents", "geopolitics", and "climate change" as the most significant.

In the safety realm of maritime transport system, existing studies (Jin, 2014; Rawson et al., 2021; Sotiralis et al., 2016) have often examined a specific aspect and categorized limited subsets of RIFs linked to marine accidents. Safety System Engineering (SSE) is the application of systems engineering in safety engineering to identify, analyse, evaluate, and eliminate various hazards in a system, with the aim of improving the safety of the entire system or the whole process of system operation (Sultana et al., 2019). The concept of safety engineering shows that accidents are caused by the unsafe act of human, the unsafe state of things, and the defects in management (Xia et al., 2023). Therefore, the SSE divides the system into a human subsystem, machine subsystem, and environment subsystem, of which the human subsystem can also be further extended to the management subsystem to consider the safety issue fundamentally and holistically (Wang et al., 2023b). Based on these concepts, this comprehensive approach involves classifying RIFs into four primary domains: human, ship, environmental, and management factors (Cao et al., 2023; Wang et al., 2023b). These categories, viewed from the SSE perspective, allow for a systemic examination of the intricate interactions among RIFs within the maritime transport system. The predominant RIFs vary across the different types of marine accidents. Jin (2014) examined the influence of ship factors on marine accidents and discovered a strong correlation among accident severity, loss of stability, and ship sinking. Furthermore, they observed that various types of ships have distinct impacts on accidents. Wang and Yang (2018) revealed the significance of the accident type, location, ship type, and service age as RIFs in marine accidents. Rawson et al. (2021) indicated that although extreme weather can influence accident occurrence, accidents primarily stem from a combination of multiple factors. Additionally, addressing risk uncertainty is a key focus of research in marine accident risk analysis. Nguyen et al. (2023) proposed a method based on BN-conditional probability tables to quantify the risk uncertainty associated with container shipping operations, thereby enhancing the quality and reliability of risk analysis. Building upon this, Nguyen et al. (2019) developed a Delphi-based risk communication platform model to comprehensively address contingency and cognitive uncertainty. Subsequently, Nguyen et al. (2021) proposed a flexible uncertainty assessment method, known as the quantitative risk analysis (QRA) model, which minimizes reliance on subjectivity. Furthermore, in the realm of maritime blockchain risk analysis, Nguyen et al. (2022) established a risk network probability indicator tailored for complex scenarios, facilitating effective risk analysis.

However, few studies have attempted to estimate or predict the severity of marine accidents. Zhou et al. (2024) developed a novel global database of maritime casualties, and subsequently constructed a data-driven BN model. This study elucidated the temporal evolution of maritime casualties across various scales, offering enhanced objectivity and predictive insights. Yang et al. (2022)

presented a stacked model that can help identify areas prone to accidents and predict their severity based on the spatial distribution of marine accidents. They analysed the characteristics of accident aggregation and spatial correlation to determine the pattern of accident severity. Using a stacked model, they identified traffic features that were strongly correlated with accident severity and made predictions based on these characteristics. Lan et al. (2023) proposed that a data-driven method could be used to predict the severity of marine casualties. They analysed the relationships between the RIFs of such accidents and identified critical aspects for predicting the severity of ship collision accidents. Rawson et al. (2021) examined accident severity prediction under extreme weather conditions. However, only the effect of weather on accident severity was considered, rather than a wider range of RIFs, and the prediction accuracy was sacrificed to achieve a high recall, resulting in a large number of false positives.

Although the aforementioned studies have provided valuable insights into marine accident analysis, several research gaps remain to be addressed:

(1) Most studies predicting the severity of marine accidents have examined only a limited number of RIFs and have yet to conduct a comprehensive analysis from the perspective of SSE.

(2) There is a scarcity of studies focusing on predicting the severity of marine accidents, and among those that do, shortcomings exist in the data processing methods or the accuracy of the prediction models. These shortcomings include issues such as data leakage and high recall rates.

(3) Few existing studies present both the severity of accidents and associated risk control measures simultaneously, and even fewer have quantitatively analysed the effects of risk control measures.

1.2.2. Research of feature engineering in machine learning

The Feature Selection (FS) is one of the most important and interesting research areas in ML (Chen et al., 2020; Gao et al., 2022). This is a process of selecting the optimal subset of features that can maximise the retention of the most representative and salient features in the observed data (Gao et al., 2022), effectively reducing the computational complexity of ML models and avoiding overfitting (Naik and Kuppili, 2022). Filter, embedded, wrapper, are three broad forms of the FS methods (Chen et al., 2020; Naik and Kuppili, 2022). Filter relies primarily on the information in the data itself to remove low-informative or low-variance features. Depending on the chosen technique, filter-based FS methods can be categorised as similarity-based, informatics-based, sparse-learning-based, or statistics-based (Naik and Kuppili, 2022) ones. Embedded methods require pre-learning using an ML model to select features by measuring their performances of the features (Chen et al., 2020). Wrapper also requires an ML model for FS; however, the process is performed in parallel with model training (Gao et al., 2022). Embedded and Wrapper methods are more suited to specific ML models; hence, they lack generality and reliable interpretability and are often computationally expensive (Naik and Kuppili, 2022).

Recently, various forecasting studies have focused on FS methods in the field of transport. Yang (2013) proposed an FS method based on p-test scores for traffic congestion prediction. Experimental results showed that this method can effectively reduce the number of features and improve traffic congestion prediction accuracy. Mohammadi et al. (2019) developed a Recursive Feature Elimination (RFE) algorithm for the FS of track geometric features and the prediction of track defects using the optimal subset of features. The results showed that RFE was effective in reducing the feature size; however, the study did not conduct further experiments and analyses on the role of RFE in maintaining high-accuracy prediction. Griesbach et al. (2020) used Principal Component Analysis (PCA) to select vehicle lane-changing features and Feedforward Neural Networks (FNN) to predict lane-changing behaviour. They found that most feature combinations improved the classification ability of the model. Gao et al. (2022) conducted a comprehensive FS study to analyse the relationship between aviation and the environment. The study used an unsupervised FS method based on information theory to initially filter out redundant and useless features and a supervised FS method with targets based on the shrinkage method and decision-trees (DTs) to identify important features that are closely related to the environment. Wang et al. (2021b) used the backward feature elimination technique to analyse the importance and impact of features in aircraft glide time prediction. They found that the use of the FS technique enables the ML model to achieve high prediction accuracy with a small number of features, and simultaneously, the optimal subset of features includes important features in the usual sense as well as specific features of the target scene.

In contrast, limited research has been conducted on FS within the maritime domain, leading to diverse approaches with inherent challenges. For instance, Wu et al. (2022) adopted an expert-experience-based Analytic Hierarchy Process (AHP) approach to enhance the accuracy of ship detention prediction. However, this method relied on expert judgment and introducesd subjectivity, potentially leading to the oversight of crucial factors and discrepancies between the selected features and real-world scenarios. Similarly, Khan and Hussain (2022) utilised an FS technique based on the XGBoost model pre-training for container ship tariff prediction. Although this approach emphasised feature importance calculations, it occasionally led to biased selections, neglecting the pivotal interactions necessary for accurate predictions. These occurrences highlight the difficulties associated solely with expert judgment or feature importance rankings, potentially disregarding subtle feature interactions that could be essential for robust modelling in maritime studies.

Although they show some attractiveness, they still have some practical problems, including:

(1) Current FS methods lack research in the maritime domain and fail to address the interactions between features.

(2) FS methods have a limited capability to elucidate selected features and offer inadequate interpretations for a specific issue.

(3) A quantitative comprehensive evaluation of the FS methods is lacking, which poses challenges in determining the systematic validity of their performance.

1.3. Contributions

The accurate prediction of accident severity is pivotal for addressing fundamental scientific challenges in maritime safety. This constitutes the foundation for informed safety decisions, targeted accident prevention strategies, and post-accident analyses (Yang et al., 2022). By analysing the characteristics across accidents of varying severities, ML techniques hold promise for significantly improving maritime safety standards. This study aims to fill this research gap by enhancing ML approaches to predict the severity of marine accidents precisely, thereby enabling informed risk control measures and fostering a safer maritime environment. The specific contributions of this study are as follows:

(1) Development of a new prediction framework for marine accident severity. The framework consists of four main aspects: establishment of a database from an SSE perspective, development of FS methods, performance evaluation of FS methods, and assessment of accident severity prediction.

(2) The development of a pioneering two-stage FS method that integrates the interactions among features and variations in the contributions of different feature states to the target.

(3) The development of a new evaluation method to comprehensively measure the performance of the proposed FS methods from multiple aspects, including stability, performance improvement, and statistical tests. In the stability evaluation process, a new convergent and stable algorithm is proposed to improve the stability evaluation performance.

(4) A comparative analysis of mainstream classical ML models is conducted to evaluate their predictive performances and establish a baseline model for marine accident severity prediction. The effectiveness of the FS methods in enhancing the performance of the ML models is validated using statistical methods.

(5) Counterfactual analysis techniques are employed to evaluate the efficacy of risk control measures by applying the methodology proposed in this study.

The remainder of this paper is structured as follows. Section 2 provides a detailed description of the problem generation and descriptions. Section 3 outlines the methodology of the study. Section 4 presents the experimental analysis, details of the dataset, and validates the performance of the FS method. Section 5 presents the analysis and discussion of the results. Section 6 comprises practice and implications, focusing on quantifying the benefits of risk control and elucidating the potential implications of this study. Section 7 presents the conclusions, limitations, and directions for future research.

2. Problem generation and descriptions

2.1. Problem generation

This study aims to optimise the predictive performance of the model by utilising FS techniques to enhance the prediction accuracy. Among the commonly employed FS algorithms, methods such as Removing Features with Low Variance (RFLV) and Traditional Mutual Information (TMI) are prevalent for feature subset selection in classification problems (Chen et al., 2020). However, factors such as problem dependency, feature complexity, and combinatorial explosion can hinder these methods from yielding the optimal feature subset for a specific problem (Gao et al., 2022), thus limiting improvements in model prediction performance. Consequently, the primary objective of this study, denoted as **Problem 1**, is to develop an advanced FS method aimed at maximising prediction accuracy.

Subsequently, building on Problem 1, this study evaluates the performance of FS methods to compare their advantages and disadvantages. Among the common evaluation metrics, prediction performance and stability are frequently used in FS assessment (Griesbach et al., 2020). However, a trade-off often exists between stability and prediction performance (Naik and Kuppili, 2022). Hence, **Problem 2** aims to propose a comprehensive FS evaluation framework that addresses this inherent contradiction. Furthermore, because of the wide variety of advanced ML models, it is necessary to select a more appropriate model for predicting the severity of marine accidents. **Problem 3** aims to identify a model with optimal performance in terms of prediction accuracy.

2.2. Problem descriptions

Problem 1. (*How to develop an advanced FS method to select features that contain more valuable information to maximize marine accident severity prediction accuracy?*) The RIFs of marine accidents exhibit intricate interrelationships which plays a pivotal roles in accident occurrence (Cao et al., 2023). Additionally, variations in the states of these RIFs lead to different consequences (Wang et al., 2021a). Hence, it is imperative to develop a two-stage advanced FS method that takes into account both the interactions among RIFs and the asymmetric contributions of their different states to accident outcomes. Refer to Section 3.2 for further elaboration.

Problem 2. (*How to comprehensively assess FS methods performance?*) In the assessment of multiple FS methods, the likelihood of any single method being optimal in terms of both stability and predictability is notably low (Naik and Kuppili, 2022). Furthermore, the statistical properties of FS outcomes are frequently overlooked. To address this gap, this study proposes a novel approach to comprehensively evaluate the performance of FS methods, considering stability, predictability, and statistical tests. Refer to Section 3.3 for further elaboration.

Problem 3. (*How to select an appropriate ML model?*) Although advanced ML models exhibit remarkable power, their performances inherently vary when applied to distinct problem domains (Zheng et al., 2020). Therefore, to identify an appropriate model for predicting the severity of marine accidents, this study aims to compare the predictive performances of six well-established ML models.

In this evaluation, five commonly used evaluation metrics were employed to conduct pairwise comparisons among the models and select the optimal model. Refer to Section 3.4 for further elaboration.

3. Methodology

This section provides a comprehensive overview of the dataset construction process, followed by detailed solutions to the three research problems posed in Section 2. The structural layout of this section is illustrated in Fig. 1. Section 3.1 elucidates the specifics of the dataset construction. Section 3.2 delineates a two-stage FS method aimed at identifying the optimal subset of RIFs that is effective in predicting the severity of marine accidents. Section 3.3 outlines the process of evaluating the effectiveness of the FS method. Section 3.4 succinctly introduces six well-established ML models and their evaluation metrics.

3.1. Dataset construction

Marine accident investigation reports (MAIRs) from seven sources (detailed in Section 3.1.1) were used to construct the Marine Accident RIFs Database (MARIFD), as shown in Fig. 2. The MARIFD consists of each accident's RIFs (i.e., accident features) and accident severity labels (i.e., target categories). The construction process of the MARIFD consists of three phases: extraction of marine accident investigation reports, labelling of accident severity, and extraction of accident RIFs.

3.1.1. Accident investigation report extraction

With reference to relevant literature (Chauvin et al., 2013; Coraddu et al., 2020; Puisa et al., 2018; Uğurlu et al., 2018), this study incorporated data from seven marine accident investigation agencies as they are among the most widely used databases in marine accident studies, including China Maritime Safety Administration (China MSA), Federal Bureau of Maritime Casualty Investigation (BSU), National Transportation Safety Board (NTSB), Japan Transportation Safety Board (JTSB), Australian Transportation Safety Board (ATSB), Canadian Transportation Safety Board (CTSB), and the Marine Accident Investigation Branch (MAIB), to ensure valid data sources. Concurrently, given the potential disruptive impact of COVID-19 on the shipping activities (Antão et al., 2023; Zhou et al., 2022), which could result in shifts in the interrelationships among RIFs, uncertainties risk (Nguyen et al., 2023) and potential conceptual drift of accident data (Liu et al., 2021b; Li et al., 2024), this study specifically utilized the 2000–2019 database.

Analysis of marine accident investigation reports from these agencies revealed variations in the level of detail among accident records depending on the country or region. Consequently, records with incomplete information, such as those lacking specific accident causes, were excluded. Following the initial screening, 2,513 accident records were obtained. To ensure accuracy and completeness of the dataset, its authenticity was rigorously evaluated. Information from different databases was cross-checked and discrepancies in accident details from multiple sources were carefully compared. For example, in the process of eliminating duplicate reports, the reports containing detailed accident information were retained. Additionally, for specific details such as ship defects, ship seaworthiness certificates, and ship personnel certificates, data from organizations such as the TOKYO MOU and Paris MOU were utilized to align and validate this information, thereby maintaining the consistency and reliability of the dataset. This rigorous validation process resulted in a refined dataset comprising 1,294 marine accident investigation reports, with comprehensive details

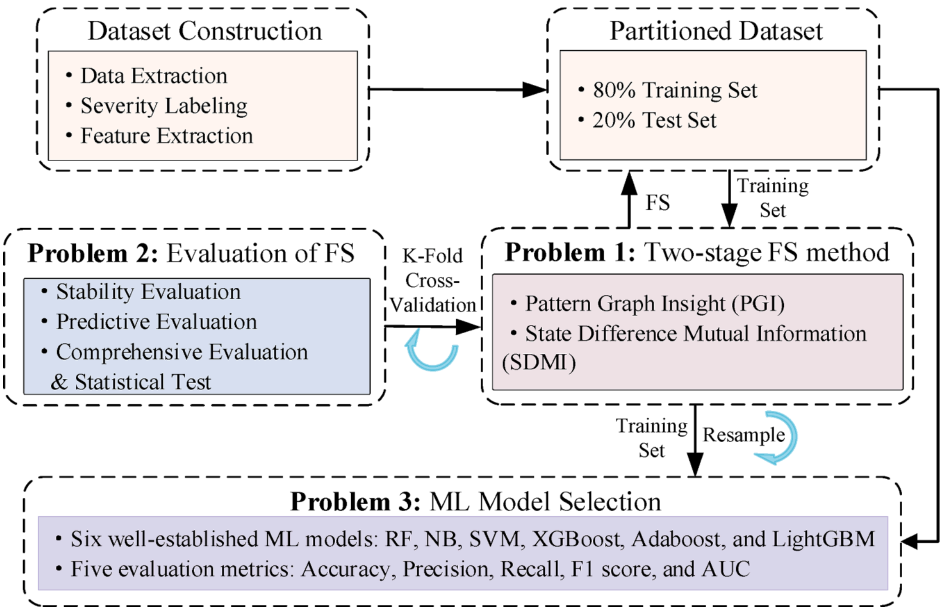

Fig. 1. The proposed prediction framework for marine accident severity.

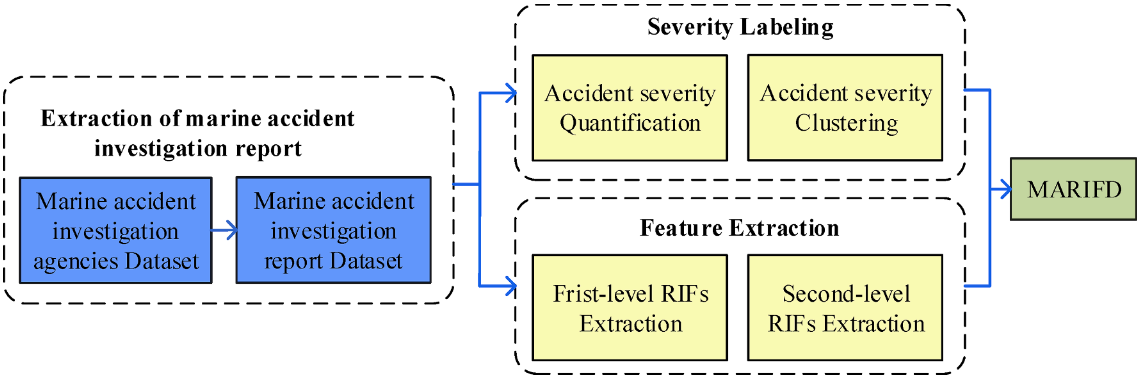

Fig. 2. General procedure of MARIFD construction.

available in previous studies (Cao et al., 2023; Wang et al., 2021a). Fig. 3 illustrates the distribution of the marine accident investigation reports by source. Variations in data quality may lead to differences in data distribution. Among the 1,294 valid reports, ATSB and MAIB accounted for the largest shares at 26 % and 20 %, respectively. In contrast, the JTSB accounts for the smallest number of valid reports, comprising only 7 %. The remaining four marine accident investigation agencies (BSU, NTSB, MSA, and TSB) demonstrated similar frequencies in the number of valid reports.

3.1.2. Accident severity labels

Currently, the International Maritime Organization (IMO) categorises marine accidents according to the extent of ship damage, casualties, and environmental pollution. For instance, based on the degree of casualties or pollution, the IMO classifies accident severity into three categories: less serious, serious, and very serious casualties (Wang et al., 2021a). Accidents involving the loss of any life, serious injury, or specific quantities of environmental pollution (such as a spill of 500 tons of oil) are defined as very serious casualties. In addition,various countries have proposed classification criteria for accident severity. For example, the Ministry of Transport of the People's Republic of China classifies marine accidents into five levels: minor, general, primary, critical, and catastrophic (Cao et al., 2023; Ministry of Transport of China, 2014). In this study, data from diverse marine accident investigation reports across different regions were utilised, which may pose challenges in standardising the data. To ensure analytical consistency and prioritise the mitigation of casualties and pollution in future maritime practices, a simplified approach was adopted to predict the severity of marine accidents. Instead of focusing on the nuanced degrees of casualties or pollution, accidents are categorised into a binary classification: those involving casualties or pollution are considered serious accidents, whereas those without casualties or pollution are deemed non-serious accidents. This simplified classification scheme allows this study underscores the importance of preventing accidents that lead to casualties or pollution, thus contributing to the overarching goal of minimising such occurrences in future maritime practices. Therefore, in this dataset, there were 960 serious accidents (74.19 %) and 334 non-serious accidents (25.81 %). This database exhibits an imbalance of categories, raising concerns regarding its potential impact on predictive modelling performance. Therefore, after extensive pre-experiments evaluating various data balancing techniques, the Support Vector Machine-Synthetic Minority Over-sampling Technique (SVM-SMOTE) (Wang, 2008) was utilised to rectify the imbalance issue. This technique adequately rebalances imbalanced datasets by synthetically augmenting the samples for the minority class, thereby ensuring data equilibrium without compromising or significantly altering the original dataset.

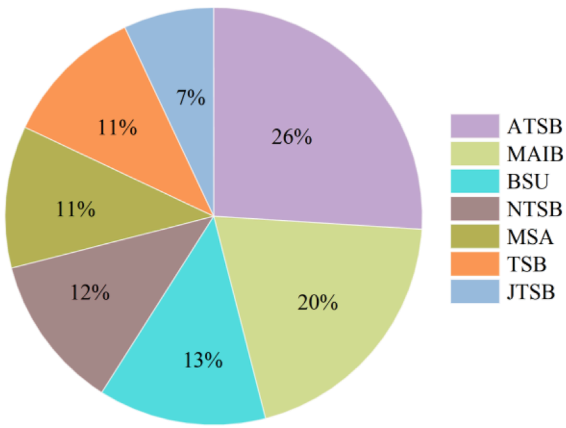

Fig. 3. Source distribution of marine accident reports.

3.1.3. Feature extraction

ML classification models cannot directly learn and classify the text of marine accident investigation reports but need to artificially select the RIFs (i.e. features) of a marine accident from an investigation report and convert the textual information about the RIFs of a marine accident into numerical or categorical information.

Based on the relevant literature (Cao et al., 2023; Wang et al., 2021a) and the description of RIFs from an SSE perspective, this study extracted human, ship, environmental, and management factors and basic accident information from accident investigation reports as criterion-level features of the MARIFD. An analysis of numerous accident investigation reports revealed that the condition of the crew plays a vital role in accident severity, making it the most significant aspect of human factors. For instance, an OOW is typically responsible for carrying out a ship's handling mission, and RIFs such as crew fatigue, psychological state, theoretical knowledge and experience, and adherence to operational standards can influence OOW's performance (Sotiralis et al., 2016). This assertion aligns with the principles outlined in the Maritime Labour Convention (MLC). An analysis of ship factors showed that ship defects can significantly affect navigation safety (Hu et al., 2020). Factors such as the ship type, gross tonnage, and service age can affect the likelihood of these defects. For instance, general dry/ refrigerated cargo ships and passenger ships have a higher probability of equipment failure than other types (Tokyo MOU, 2023). There is also a correlation between the service age of a ship and accident

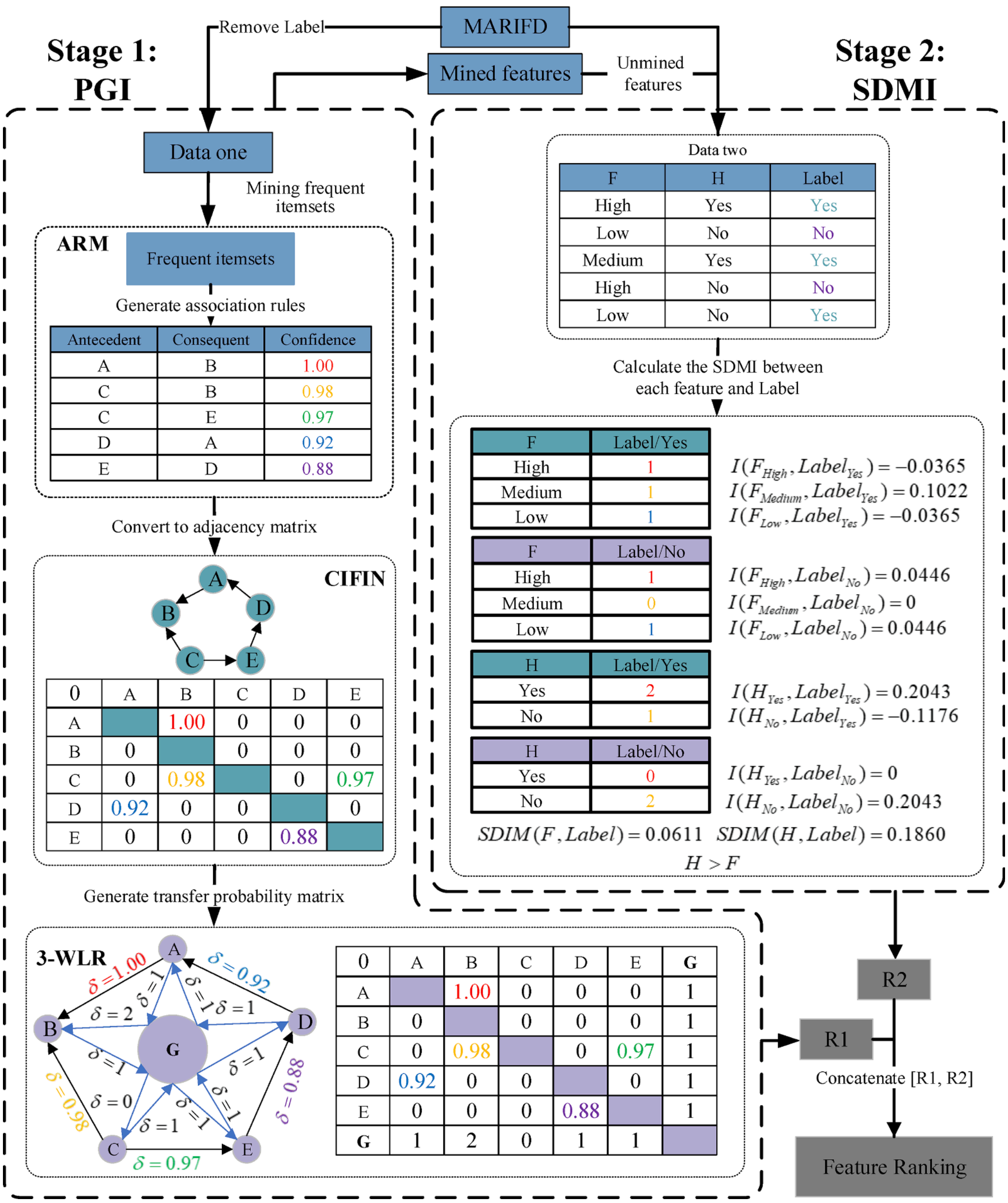

Fig. 4. The flowchart of PGI-SDMI method.

severity (Cao et al., 2023). Furthermore, inspections conducted by Port State Control/Flag State Control (PSC/FSC) play a crucial role in ensuring ship safety (Cao et al., 2023). These inspections can ascertain whether a ship is defective (Tokyo MOU, 2023). Concurrently, the seaworthiness of the vessel, including factors such as the condition of the ship's crew, the validity of its documents, and whether passengers/cargo meet seaworthiness requirements, significantly influences the severity of marine accidents (Wang et al., 2022). In terms of environmental factors, this study mainly considers natural conditions, such as winds, waves, and currents, and waterway conditions, such as the location of navigable waters and traffic density (Wang et al., 2022; Weng and Yang, 2015). Concerning management factors, management deficiencies by companies and authorities may also influence accident severity (Cao et al., 2023). In addition, several studies have shown that both the type and time of an accident are correlated with its severity (Cao et al., 2023; Jin, 2014; Wang et al., 2022; Zhou et al., 2023). For example, Cao et al. (2023) highlighted that capsizing or sinking accidents are positively correlated with a higher accident severity. Similarly, Wang et al. (2022) suggested that the probability of serious accidents occurring at night might be higher than that of daytime accidents. For this reason, basic accident information was used as a criterion-level feature in this study.

In the data processing phase, the five aforementioned criterion-level features were categorised into 68 index-level features. This categorisation, based on expert judgment and textual analysis, involves converting continuous and character-based data into discrete categories. The detailed process is outlined in the previous work conducted by this team (Cao et al., 2023; Wang et al., 2021a). For specific descriptions of these RIFs, please refer to Appendix A.

3.2. Feature selection method

In response to Problem 1, this study proposes an innovative two-stage FS approach called Pattern Graph Insight-State Difference Mutual Information (PGI-SDMI), which aims to address the challenges associated with the interactions between features and the relationships between features and targets. The PGI-SDMI model not only leverages the strengths of both PGI and SDMI but also compensates for their individual limitations. On the one hand, PGI, while effective, is sensitive to parameter settings and may not comprehensively assess the importance of all features. However, SDMI, although parameter-free, exhibited sensitivity to data and lower stability than PGI. By combining these methods in a two-stage process, the PGI-SDMI model provides a comprehensive evaluation of all features under commonly used parameters without the need for fine-tuning. This approach ensures improved robustness, stability, and interpretability of marine accident analysis. The structure of the PGI-SDMI model is illustrated in Fig. 4 and the corresponding pseudocode is detailed in Algorithm 1 of Appendix B.

In the first stage, the PGI represents an innovative fusion of Association Rule Mining (ARM) (Feng et al., 2024) and the Three-Weight LeaderRank (3-WLR) algorithm. This method comprises a three-phase process (as shown in the left part of Fig. 4): initial ARM-based rule extraction, construction of a Complex Influential Factor Interaction Network (CIFIN), and ranking the importance of RIFs based on 3-WLR (the pseudocode is shown in Algorithm 2 of Appendix B).

A distinctive feature of the PGI is its improvement of the WLR algorithm, resulting in the creation of the 3-WLR algorithm tailored for a comprehensive understanding of the importance and interactions among the RIFs within the CIFIN. Notable improvements include the redefinition of the conventional weight definition ($Weight_{ij} = 1$) in the WLR algorithm (Li et al., 2014), which is achieved by introducing a transfer probability matrix, as shown in Equations (1) and (2). This matrix refines the portrayal of the relationships between neighbouring RIFs, providing a more precise representation of their interactions in the marine accident system. Another innovation of the 3-WLR algorithm is the introduction of a novel integrated weight calculation method rooted in graph theory, specifically the out-strength and betweenness centrality coefficients (Wang et al., 2023a), as shown in Equation (3). This computational approach enhances the capacity of the algorithm to assess the importance of RIFs and their interactions within complex networks. The iterative computation of the 3-WLR algorithm and the final computation after weight assignment are shown in Equations (4) and (5), respectively.

$$M = (\delta_{ij})_{(n+1)\times(n+1)} \tag{1}$$

$$\delta_{ij} = \begin{cases} k_j^{in},\ when\ i = g\ and\ j \neq g \\ 1,\ when\ i \neq g\ and\ j = g \\ C(i \Rightarrow j), when\ there\ exists\ i \Rightarrow j\ and\ i \neq j \neq g \\ 0,\ else \end{cases} \tag{2}$$

$$\mu_i = \left(\frac{B_i}{\sum B_m} + \frac{S_i^{out}}{\sum S_m}\right)/2, i \in \{1, 2, ..., n\} \tag{3}$$

$$L_i(t+1) = \sum_{j=1}^{n+1} \frac{M \odot H_{ij}}{Z_j{}^* M^* O} L_j(t), i, j \in \{1, 2, ..., n, g\} \tag{4}$$

$$R_i = L_i(t_{end}) + \mu_i L_g(t_{end}), i \in \{1, 2, ..., n\} \tag{5}$$

where M denotes the (n + 1)-dimensional transfer probability square matrix whose corresponding node in the (n + 1)-th dimension is g

(Ground node); δ_{ij} denotes the transfer probability from node i to node j; n denotes both the number of RIFs mined by ARM and number of nodes of CIFIN; g represents the number assigned to the Ground node added to the CIFIN (Li et al., 2014); k_j^{in} represents the in-degree of node j (Wang et al., 2023a); $C(i \Rightarrow j)$ represents the confidence of $i \Rightarrow j$ (Feng et al., 2024); $i \Rightarrow j$ denotes both the existence of a valid association rule from i to j and the existence of a directed edge from node i to j (Feng et al., 2024); μ_i, B_i and S_i^{out} denote the weight, betweenness centrality coefficient and out-strength of node i, respectively; $L_i(t)$ represents the score of node i after the t $^{\text{th}}$ iteration; t represents the iteration number; H_{ij} represents an $(n+1)-$order square matrix in which only the element of row i and column j is 1 and the rest of the elements are all 0; Z_j denotes an $1 \times (n+1)-$order matrix in which only the element of column j is 1 and the rest of the elements are all 0; O represents an $(n+1) \times 1-$order matrix in which the elements are all 1; $\odot$ and * denote the Hadamard and matrix

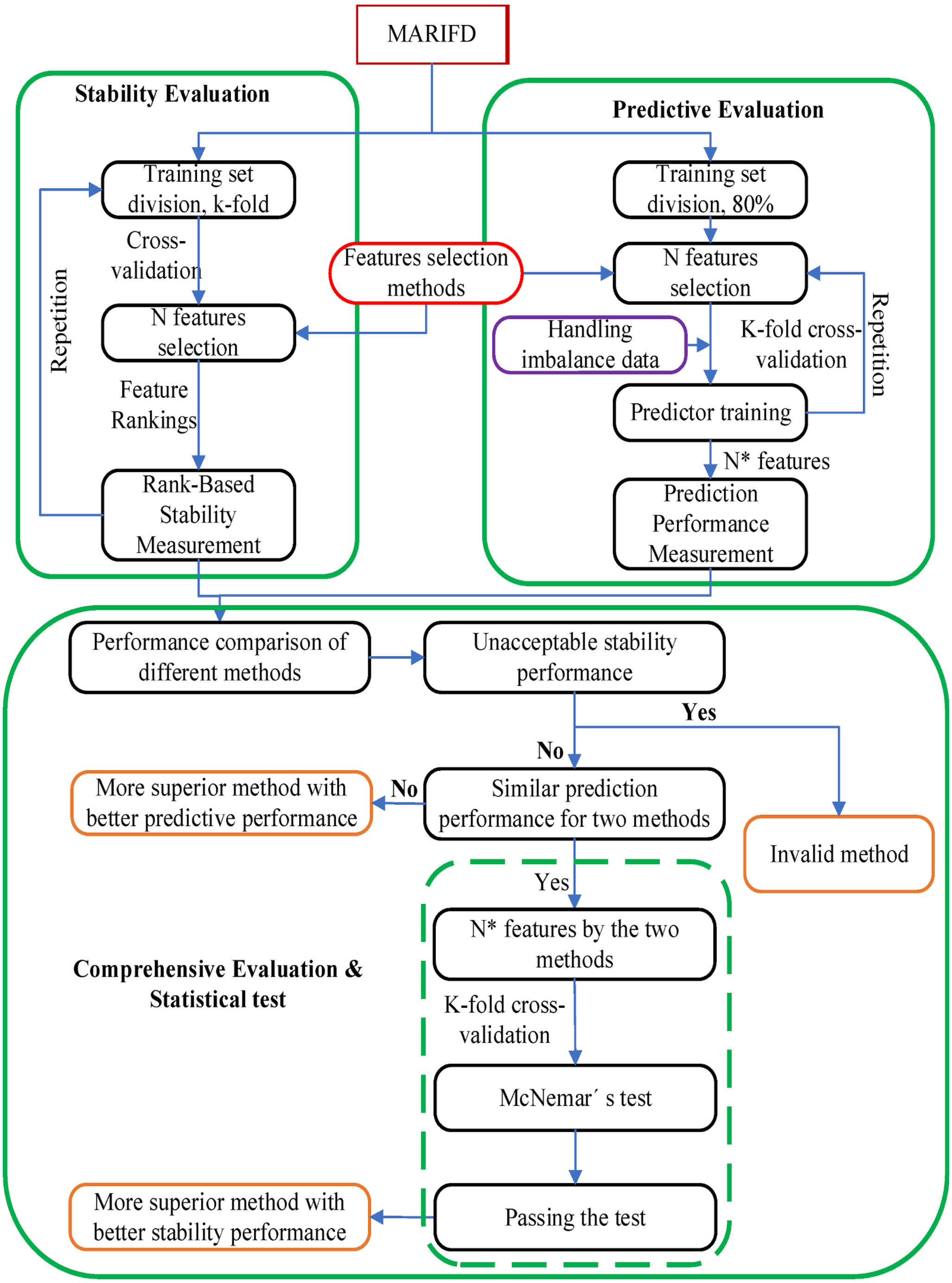

Fig. 5. General procedures for performance evaluation of FS methods.

product, respectively; t_{end} denotes the last iteration; R_i represents the final score of node (RIF) i.

In the second stage, unlike TMI, which primarily focuses on the overall information shared between two variables (Hoque et al., 2014; Ryu et al., 2018), SDMI is proposed to gauge feature importance across diverse states of the target variable, as shown in Equation (6), the pseudo-code is shown in Algorithm 3 of Appendix B. The SDMI places greater emphasis on understanding information changes in features across various states of the target variable. This nuanced approach allows for a more detailed assessment of the feature importance under different states.

$$SDMI(X,Y) = \sum_{a=1}^{A} \frac{\left(\sum_{p=1}^{D}\sum_{q=1}^{D}\left(I(X_p,Y_a) - I(X_q,Y_a)\right)^2\right)^{1/2}}{2D} \tag{6}$$

where $SDMI(X,Y)$ represents the state-differentiated mutual information between RIF X and label Y; A and D denote the number of state of the label Y and the RIF X, respectively; $I(X_p,Y_a)$ indicates the mutual information between state p of RIF X and state a of the label Y.

Drawing inspiration from TMI measures (Ryu et al., 2018) and the Gini coefficient (Gao et al., 2022), SDMI extends its applicability to provide an asymmetric relationship with feature targets, allowing for a more precise evaluation of feature importance within specific states of the target variable, thereby addressing the limitations of conventional methods.

In the holistic PGI-SDMI approach, these two stages play complementary roles. Initially, the PGI examines and ranks features, capturing their intricate interactions within a complex network landscape. Subsequently, the SDMI focuses on assessing the significance of features excluded from the PGI ranking across different target variable states, generating an additional feature ranking. By combining these rankings, the comprehensive approach ensures a holistic consideration of both complex feature interactions (identified by the PGI) and nuanced feature importance across diverse states (analysed by the SDMI), thereby facilitating the identification of crucial features with multifaceted relevance.

3.3. *Evaluation of feature selection*

In response to Problem 2, this study introduces a novel and comprehensive framework for evaluating FS methods by considering three crucial aspects: stability, predictive performance, and conducting a comprehensive evaluation with statistical testing (as illustrated in Fig. 5). By concurrently examining these factors, this study aims to assess the performances of various FS methods in a more comprehensive and thorough manner.

The process begins by assessing the stability of the FS methods. This study introduces an innovative stability assessment algorithm to evaluate the stability of the FS method across varying degrees of dataset fluctuation and to ensure realistic and reliable assessment results. A randomly generated number “k” is incorporated in the algorithm to describe randomness. Following the principles of K-fold cross-validation, the dataset undergoes a degree and type of variation consistent with randomness, thereby simulating real-world scenarios. For repeatability, the algorithm involves multiple repetitions of experiments and averaging the results to mitigate excessive fluctuations. In addition, a reference feature ordering was established to anchor the stability results. To satisfy the conditions of boundedness and monotonicity, the algorithm employed Spearman’s rho for stability evaluation (Chen et al., 2020), as expressed in Equation (7). This indicator adheres to the bounded condition, ranging from -1 to 1. A higher stability tends to yield results closer to 1, ensuring a monotonic relationship with the assessed stability. The pseudocode of the algorithm is provided in Algorithm 4 of Appendix B.

$$Sprr^{R1,R2} = 1 - \frac{6\sum_{i\in V}(R1_i - R2_i)^2}{N(N-1)(N+1)} \tag{7}$$

where $Sprr^{R1,R2}$ is the correlation between sort $R1$ and sort $R2$; $R1$ is the reference ranking; $R2$ is the experimental ranking; N is the number of top RIFs; V represents all the RIFs of $R1$ included in the top RIFs, and if RIF i is not in the top RIFs of $R2$, then $R2_i = N+1$.

After the stability evaluation, the subsequent phase involved assessing the predictive performance of the selected features. The objective of this step was to assess the influence of the identified RIFs on the accuracy of the prediction models. This process entails training the prediction models using the selected RIFs and identifying the subset that achieves the highest accuracy with the fewest features. It involves 5-fold cross-validation using advanced ML models on an 80 % training set. SVM-SMOTE (Wang, 2008) was employed to address the data imbalance, and the accuracy of the model was evaluated using the metrics outlined in Section 3.4.

Ultimately, a comprehensive statistical evaluation is essential to integrate stability and predictive performance assessments. The stability of an FS method is deemed unacceptable if its optimal stability is below 0.6, or if the difference from another FS method possessing optimal stability exceeds 0.3 (Nogueira and Brown, 2015). Methods with acceptable stability were subjected to further analysis to determine their predictive performance. When the performance metrics exhibited variability, a superior method was identified based on the observed performance differences. If the metrics were identical, statistical tests like the 5-fold McNemar test were conducted (Bennasar et al., 2015; Hoque et al., 2014). P-values exceeding 0.05 indicate no significant difference in predictive performance among different methods (Wong and Yeh, 2020). Finally, when the predictive performances were comparable, the method showing superior stability was deemed the best performing method.

3.4. Advanced machine learning prediction models selection

In response to Problem 3, this section succinctly reviews six well-established ML models (as Table 1), delineates five prevalent evaluation metrics, and outlines a methodology for conducting one-to-one comparisons.

To evaluate the predictive performance of the ML models, five evaluation indices, including Accuracy, Precision, Recall, F1 score, and Area Under the Curve (AUC), were used in this study (Chen et al., 2020; Lan et al., 2023). Throughout the evaluation process, a one-to-one comparison strategy was employed to analyse the performance differentials among the models, distinguishing winners from losers.

4. Experiment

First, this section describes the distribution and proportion of RIFs and the different accident severity ratios for the same RIF in the MARIFD. Second, the effectiveness of the proposed FS method for marine accidents is verified.

4.1. MARIFD descriptions

In this study, based on 1294 marine accident investigation reports, 68 RIFs were identified and classified into five categories: human factors, ship factors, environmental factors, management factors, and accident information. Fig. 6 shows the frequency distribution of all the RIFs. Fig. 6 shows that human and management factors occurred more frequently among the five RIF categories. Among the human factors, H5 (operational error) and H6 (violation of rules and regulations) occurred most frequently. Among the management factors, M3 (Defective safety management system) has a higher frequency of occurrence. Fig. 6 shows that Collision (A1) and Stranding/Ground (A2) are the main types of marine accidents (Wang et al., 2021a). Bulk carriers (ST1) and Fishing vessels (ST7) were the main vessels involved in an accident. The traffic conditions in the Channel (ET) are generally busy during most accidents.

Fig. 7 shows the distribution of accident severity for different accident times, accident types, ship types, and ship ages in bar charts. Fig. 7(a) illustrates that serious accidents are more likely to occur between 0000 and 0400 (T6), whereas Fig. 7(b) shows that the highest frequency of serious accidents occurred in capsize/sink accidents (A5). Fig. 7 (c) shows that the severity of accidents varies significantly with the type of vessel. Fishing vessels (ST7) were prone to serious accidents. Fig. 7 (d) shows that the proportion of vessels involved in serious accidents increases with the age of the ship.

4.2. Performance evaluation of feature selection methods

This study presents an innovative FS method, PGI-SDMI, and evaluates its performance by comparing it with four established FS methods, (filter FS based on RFLV and embedded FS based on SVM, RF, and Gradient Boosting Decision Tree (GBDT) respectively), which serve as benchmarks. Moreover, the individual performances of the PGI and SDMI methods were assessed separately to determine the necessity of combining them in a two-stage approach. Additionally, to highlight the enhanced performance of the 3-WLR algorithm within the PGI, this study compared it with the WLR and PageRank algorithms. Furthermore, to demonstrate the superiority of the SDMI algorithm, it was compared with the TMI algorithm. To compare the performances of the FS methods across all RIF ranges, minimum support and confidence thresholds of 0.005 and 0.3 are used as examples in the PGI, whereas the minimum support and confidence thresholds in the PGI-SDMI were set as 0.1 and 0.3, respectively. Notably, the choice of parameters for PGI and PGI-SDMI did not affect the results of the main body of this study. Appendix C provides additional information on the performances of the PGI and PGI-SDMI with other parameters.

(1) Stability evaluation stage

Each FS method was trained on the training set data in this subsection. The stability of each FS method was then measured using the ranking-based stability algorithm proposed in Section 3.3. As shown in Fig. 8 (a), the 3-WLR algorithm proposed in the PGI is more robust in terms of the FS stability. When the number of RIFs selections exceeded 21, the FS stability of the 3-WLR algorithm remained above 0.9. The highest stability values for the PageRank and WLR algorithms were 0.8461 and 0.8541, respectively.

According to Fig. 8 (b), the embedded FS techniques based on the GBDT and SVM do not exhibit sufficient stability when tested

Table 1
Six well-established ML models.

Model	Description	Reference
SVM	SVM is a linear classifier defined on the feature space with maximum intervals.	(Chen et al., 2020)
NB	NB is a prediction model based on the assumption that each feature and sample data are independent.	(Sotiralis et al., 2016)
RF	RF uses sampling techniques to select sample data and train several independent decision trees.	(Li et al., 2023a)
AdaBoost	AdaBoost uses the technique of changing the weights of the samples to focus on poorly trained samples.	(Rawson et al., 2021)
XGBoost	XGBoost mainly uses a gradient boosting method based on a second-order Taylor expansion.	(Rawson et al., 2021)
LightGBM	LightGBM uses techniques such as a gradient boosting method and a histogram optimisation algorithm.	(Chen et al., 2020)

Note: SVM is support vetor machines; NB is Naive Bayesian; RF is random forest; AdaBoost is adaptive boosting; XGBoost is extreme gradient boosting; LightGBM is light gradient boosting machine.

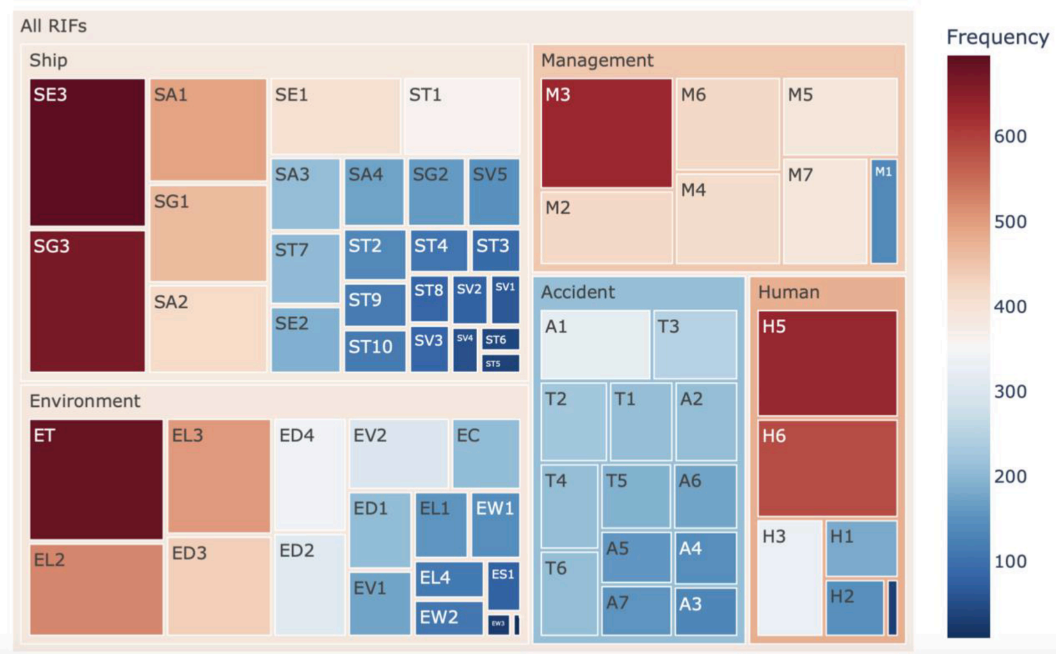

Fig. 6. Tree map of the frequency distribution of all RIFs.

using MARIFD. This may be owing to the lack of robustness and generality of the classifiers, resulting in their inability to demonstrate satisfactory performance in terms of stability (Chen et al., 2020). In comparison, the RFLV-based filtered and RF-based embedded FS methods produce better stability outcomes. This may be because the RFLV variance test is not affected by size variations in identically distributed data (Liu et al., 2020), whereas RF applies bagging techniques that reduce classifier prediction errors and enhance stability (Wager and Athey, 2018). Furthermore, it is evident that the stability of the FS using TMI was lower than that of SDMI from Fig. 8(b). This suggests that considering the contribution of features with different states to the target enhances the stability of the FS compared with a uniform assessment of the overall information relationship.

Compared to the traditional FS methods of ML, such as RF, RFLV, GBDT, SVM, and TMI, the PGI-SDMI method, as well as the individual PGI and SDMI methods, demonstrated higher stability in the FS process. A comparison of the stabilities of PGI and PGI-SDMI revealed interesting nuances. The PGI-SDMI tends to exhibit superior stability to the SDMI alone, primarily because of the FP growth algorithm used in the ARM technique, which is known for its robustness against data perturbations (Zhou, 2021). Notably, when dealing with a smaller number of RIFs, the PGI-SDMI outperformed the PGI in terms of stability. However, as the number of RIFs exceeded 30, PGI exhibited a progressively higher stability than PGI-SDMI. This disparity might arise from the training of the SDMI algorithm within the PGI-SDMI when the number of RIFs surpasses 30, leading to a decrease in the average stability. A comprehensive comparison of PGI-SDMI, PGI, and SDMI indicates that while PGI surpasses SDMI in terms of stability, SDMI does not require parameter settings and can sort all features. Consequently, the integration of the PGI and SDMI into the PGI-SDMI ensures sustained high stability while effectively sorting all features without compromising crucial information. It's noteworthy that due to the parameter settings of PGI-SDMI (minimum support and confidence levels set to 0.1 and 0.3, respectively), PGI can only mine up to 30 RIFs, with SDMI sorting the subsequent unmined RIFs.

(2) Prediction performance evaluation stage

This stage aims to showcase the effectiveness of various FS methods in predicting model performance using the LightGBM model as an example to perform a 5-fold cross-validation on the training data after the FS process. The results of the experiment are presented in Table 2 and Fig. 9. The prediction performances of different FS methods on other ML models are presented in Appendix D.

By analysing Fig. 9 and Table 2, it can be observed that the 3-WLR algorithm selects a smaller number of N* RIFs, resulting in a better prediction performance when compared to PageRank and WLR algorithms. This implies that the improved 3-WLR algorithm was effective in this study. Furthermore, it is evident that in comparison to TMI, the SDMI algorithm proposed in this study can select fewer features while achieving enhanced prediction performance.

The results presented in Fig. 9 and Table 2 also show that the LightGBM predictor trained on the N* RIFs selected by the proposed PGI, SDMI, and PGI-SDMI methods performed better than the five traditional FS methods of ML. PGI-SDMI, in particular, selected the lowest number of N* RIFs and exhibited the best prediction performance.

(3) Comprehensive Evaluation & Statistical test stage

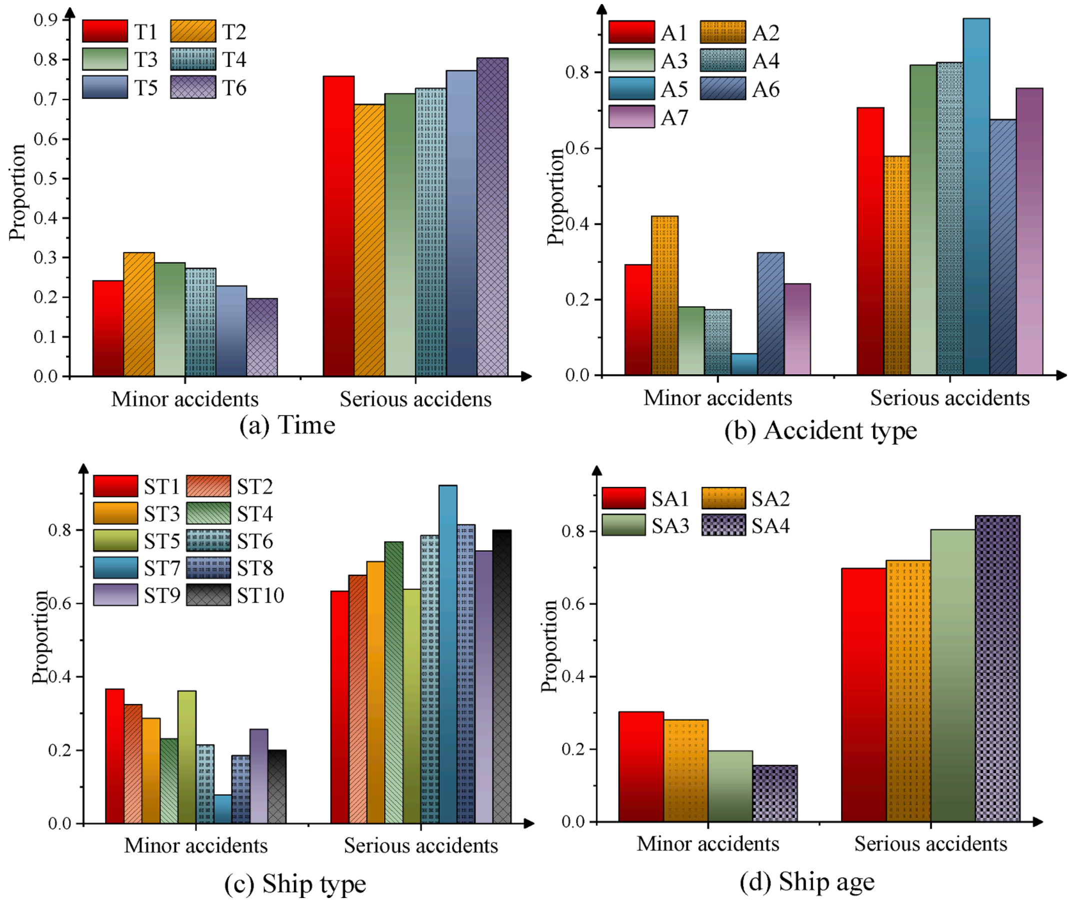

Fig. 7. Distributions of accident severity.

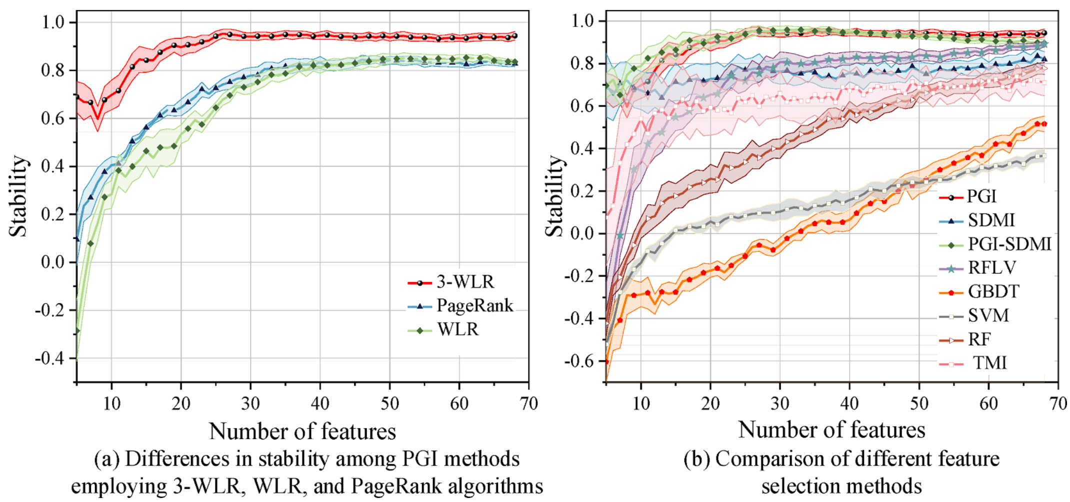

Fig. 8. The stability comparison of FS methods.

The comprehensive assessment of stability and predictive performance shows that although the predictive performance of the embedded FS methods based on GBDT and SVM is close to that of the three FS methods proposed in this study (refer to Fig. 9 and Table 2), their stability is outside the acceptable range (their optimal stability is less than 0.6, as shown in Fig. 8). Therefore, PGI, SDMI, and PGI-SDMI performed better than the embedded FS methods based on the GBDT and SVM.

Table 2
Prediction performance based on N* RIFs selected by different methods.

FS Methods	Accuracy	Precision	Recall	F1	AUC
PGI	0.8490 ± 0.0152	0.8563 ± 0.0236	0.8399 ± 0.0260	0.8475 ± 0.0128	0.9135 ± 0.0159
SDMI	0.8477 ± 0.0128	0.8668 ± 0.0240	0.8230 ± 0.0192	0.8439 ± 0.0082	0.9223 ± 0.0125
PGI-SDMI	0.8555 ± 0.0163	0.8687 ± 0.0182	0.8386 ± 0.0298	0.8528 ± 0.0141	0.9226 ± 0.0135
RF	0.8477 ± 0.0127	0.8496 ± 0.0282	0.8452 ± 0.0192	0.8471 ± 0.0154	0.9178 ± 0.0138
GBDT	0.8464 ± 0.0134	0.8568 ± 0.0242	0.8295 ± 0.0200	0.8426 ± 0.0136	0.9156 ± 0.0115
SVM	0.8464 ± 0.0129	0.8460 ± 0.0228	0.8478 ± 0.0135	0.8466 ± 0.0090	0.9063 ± 0.0130
RFLV	0.8470 ± 0.0128	0.8594 ± 0.0308	0.8317 ± 0.0335	0.8443 ± 0.0134	0.9187 ± 0.0128
PageRank	0.8337 ± 0.0234	0.8570 ± 0.0226	0.8376 ± 0.0293	0.8465 ± 0.0085	0.9210 ± 0.0130
WLR	0.8385 ± 0.0077	0.8478 ± 0.0107	0.8255 ± 0.0240	0.8362 ± 0.0093	0.9218 ± 0.0062
TMI	0.8451 ± 0.0136	0.8587 ± 0.0253	0.8225 ± 0.0211	0.8418 ± 0.0142	0.9186 ± 0.0064

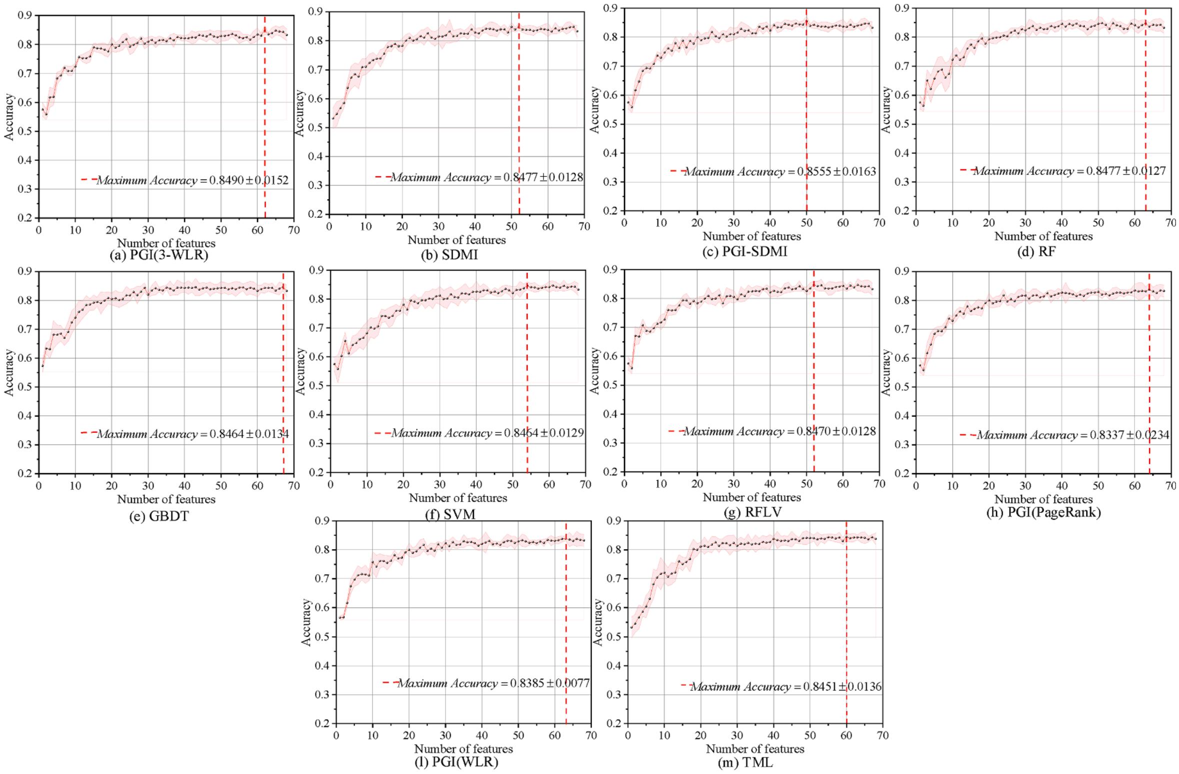

Fig. 9. Prediction accuracy based on different FS methods. Note: The horizontal-axis value of the dash line refers to the least number of RIFs, N*, with which the predictor (lightGBM) is trained to achieve the best predictive performance.

Based on the data presented in Fig. 9 (b), it is evident that as the number of RIFs increases, the stabilities of RFLV-based filtered, FS methods of RF-based embedded and TMI gradually approach those of the FS methods proposed in this study. The stabilities of FS methods of RF-based embedded and TMI is consistently lower than those of PGI, SDMI, and PGI-SDMI. The stability of the RFLV-based filtered FS method was consistently lower than those of PGI and PGI-SDMI. However, when the number of RIFs exceeded 30, the

Table 3
The FS methods proposed in this study were compared with RFLV and RF on different predictors by the p-value of McNemar's test.

FS methods comparison	PGI		SDMI		PGI-SDMI	
	RF	RFLV	RF	RFLV	RF	RFLV
LightGBM	0.6862	0.4534	0.3747	0.5109	0.5902	0.7318
RF	0.5782	0.6096	0.6225	0.2932	0.4721	0.7083
SVM	0.6922	0.6837	0.7251	0.7017	0.7251	0.7017
NB	0.5431	0.5752	0.3747	0.5013	0.3747	0.5013
AdaBoost	0.3140	0.5329	0.3758	0.7496	0.6466	0.4793
XGBoost	0.5266	0.5799	0.5162	0.6166	0.5696	0.6166

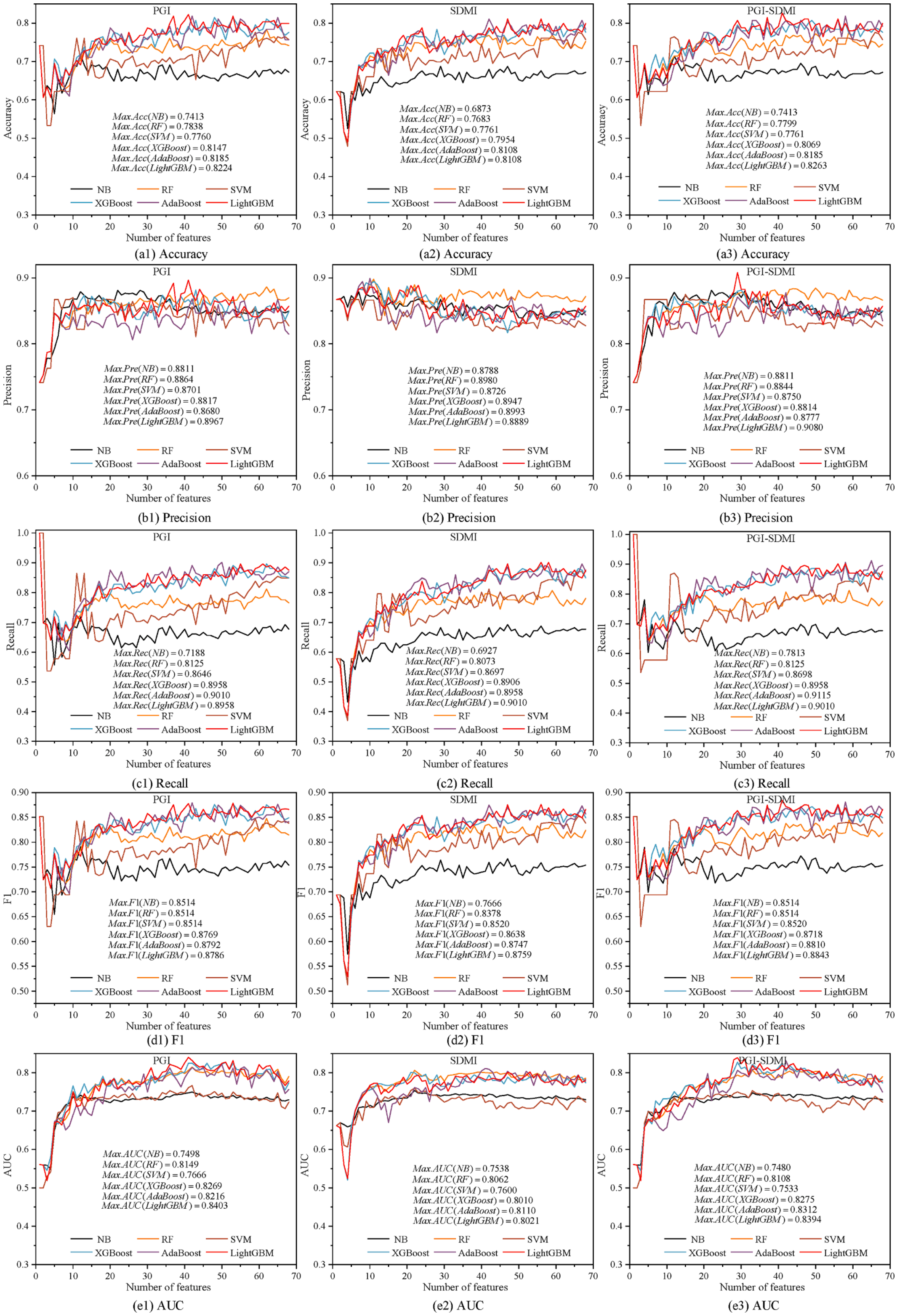

Fig. 10. Generalisation performance of six advanced ML models based on RIF selected by PGI, SDMI and PGI-SDMI. (Note: The maximum recall does not consider the case where the number of RIFs is less than or equal to 2.).

stability of the RFLV-based filtered FS method was slightly higher than that of the SDMI.

The stabilities of the PGI using the WLR and PageRank algorithms, respectively, fell within acceptable limits. However, as detailed above and in Appendix D, it is apparent that the prediction performance of the PGI with the WLR and PageRank algorithms is lower

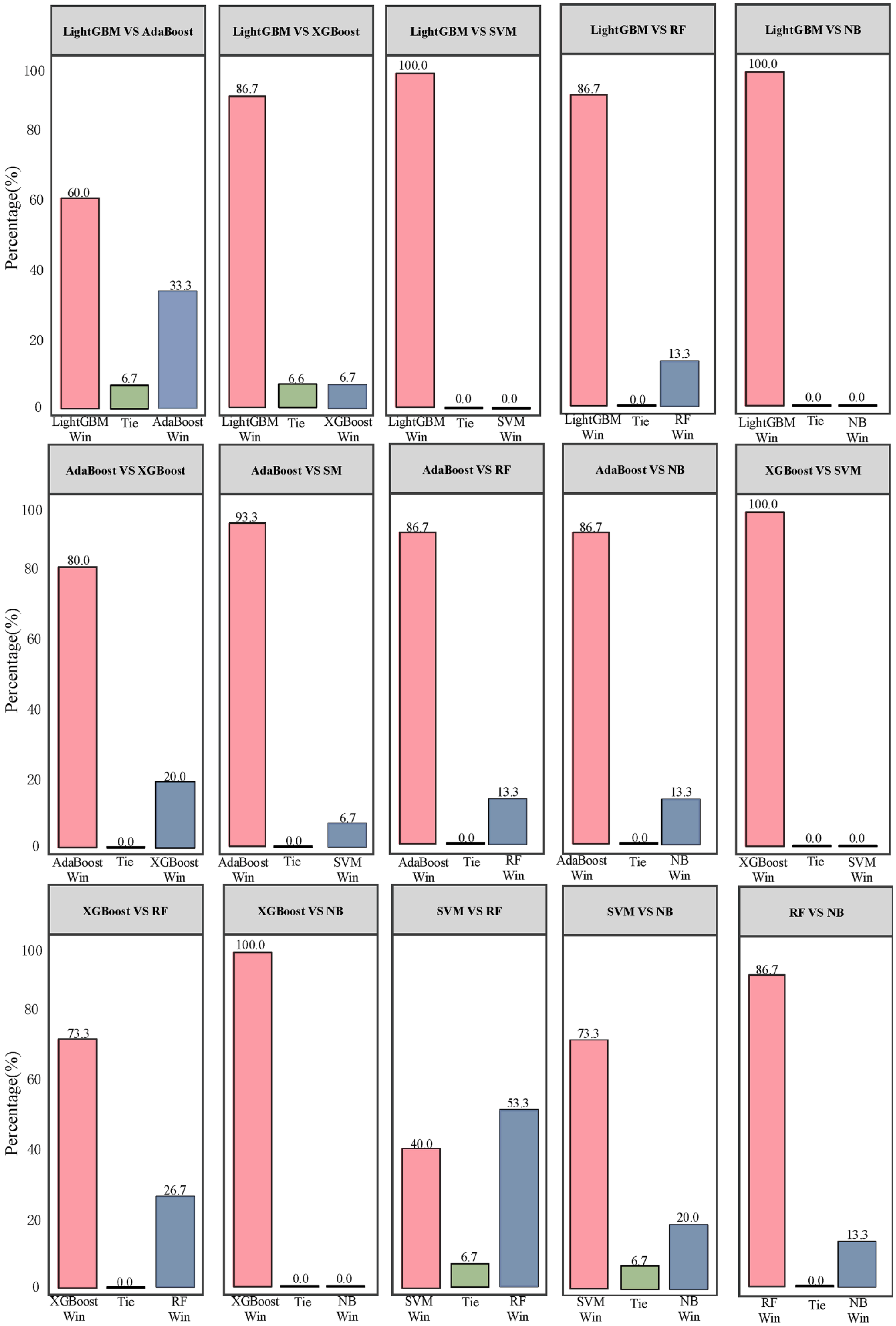

Fig. 11. Percentage of wins, ties and losses between different ML models.

than that of the PGI-SDMI and PGI using the 3-WLR algorithm. Additionally, the prediction performance of the TMI algorithm was lower than that of SDMI and PGI-SDMI. This indicates the effectiveness of the improvements introduced by the two new algorithms.

From the above discussion and Appendix D, it can be seen that the RF-based embedded and RFLV-filtered FS methods do not have the same advantages and disadvantages in predictive performance as PGI, SDMI, and PGI-SDMI for the six different predictors. To demonstrate the effectiveness of the proposed FS methods further, McNemar's test was performed, and the results are presented in Table 3. The test results showed no significant difference in the predictive performance between the FS methods proposed in this study and the filtered RF-based embedded and RFLV methods (Bennasar et al., 2015; Hoque et al., 2014).

Overall, the FS methods introduced in this study outperformed the five benchmark methods. Furthermore, the enhanced 3-WLR algorithm outperformed both WLR and PageRank algorithms. The newly proposed SDMI algorithm also exhibited a better performance than TMI. Combining the distinct PGI and SDMI techniques into a two-stage PGI-SDMI method improved the overall FS performance.

5. Results and discussions

5.1. *Different predictors' generalization performance*

This study aimed to accurately predict the severity of marine accidents and develop a benchmark model for accident severity prediction. To achieve this, the performances of six well-established ML models were compared. First, RIFs were selected from the training data using the PGI, SDMI, and PGI-SDMI. Second, the training data were balanced using the SVM-SMOTE technique. Subsequently, six advanced ML models were trained on the balanced training data. Finally, the generalisation performances of the models were evaluated using the test data. Fig. 10 shows the generalisation performance of the six ML models.

To compare the generalisation performance of each ML model more intuitively, this section compares the winners and losers of each ML model individually based on the maximum value of each performance metric, as shown in Fig. 10. This is achieved by comparing the models one to one, as shown in Fig. 11. Based on these findings, LightGBM exhibited the best generalisation performance, closely followed by AdaBoost and XGBoost. The RF showed a slightly better generalisation performance than the SVM, with the AUC of the RF being stronger than that of the SVM. This indicates that the reliability of the RF model results was higher than that of SVM. In contrast, NB exhibited the worst generalisation performance.

5.2. *Predictor's generalization performance improvement*

As detailed in Section 5.1, LightGBM demonstrates the best generalisation performance after FS. To visually highlight the superiority of the FS methods in this study compared with the algorithms before improvement, the enhancement in the generalisation performance of the LightGBM model by each FS method is presented in Table 4. Notably, the SDMI and PGI-SDMI selected the fewest key features, indicating their ability to eliminate redundant information. Furthermore, the PGI, utilising the 3-WLR algorithm proposed in this study, outperformed the WLR and PageRank algorithms in enhancing the generalisation performance of the LightGBM model. The SDMI algorithm, also proposed in this study, exhibited a greater improvement in the generalisation performance of the LightGBM model than TMI. Additionally, the integration of the PGI and SDMI into the two-stage PGI-SDMI in this study substantially enhanced the generalisation performance of LightGBM compared with using the PGI and SDMI individually.

Beyond evaluating the standalone use of LightGBM as a predictor to illustrate the broad applicability of FS algorithms, this study compared the generalisation performance improvements of five other ML models. The analysis demonstrates that the improved 3-WLR algorithm enhances the overall generalisation performance of the predictors by 0.03 % to 3.09 % and 0.03 % to 2.25 % compared with PageRank and WLR, respectively. The SDMI algorithm proposed in this study improves the overall generalisation performance of the predictors by 1.36 %–6.67 % more compared to TMI. Furthermore, compared with using PGI and SDMI individually, the two-stage PGI-SDMI in this study enhanced the overall generalisation performance of the predictors by 0.08 % to 1.53 % and 0.21 % to 6.14 %, respectively. These findings demonstrate the superior performance of the proposed FS methods in this study.

The ANOVA statistical test was used to analyse the prediction results of the six ML models (Bennasar et al., 2015). The goal was to verify that the improvement in model generalisation performance from the FS methods proposed in this study was not random but systematic. After examining the results of the initial training data and the data that underwent FS, the data were analysed (Bennasar et al., 2015). As shown in Table 5, P value was used, which is the probability that the improvement in generalisation performance is due to chance and MS, which is the mean square error. According to the results, the P value was less than 0.05 for all three FS methods.

Table 4
Improvement of predictor (LightGBM) generalization performance by FS methods.

Improvement	PGI	SDMI	PGI-SDMI	WLR	PageRank	TMI
N*	**41**	47	**41**	49	50	54
Accuracy	2.32 %	1.16 %	**2.70 %**	1.55 %	1.93 %	1.16 %
Precision	**3.96 %**	−0.14 %	1.60 %	2.18 %	3.90 %	−0.86 %
Recall	−1.56 %	**2.08 %**	**2.08 %**	−0.52 %	−2.08 %	0.00 %
F1	1.17 %	0.93 %	**1.83 %**	0.83 %	0.87 %	−0.45 %
AUC	**6.15 %**	2.23 %	4.10 %	4.87 %	4.20 %	1.39 %
Mean	2.41 %	1.25 %	**2.46 %**	1.78 %	1.76 %	0.25 %

Table 5
The ANOVA test of prediction results of three FS methods and six ML models.

FS Methods	MS	F	P
PGI	0.003128	11.9338	0.008637
SDMI	0.004073	20.1157	0.002042
PGI-SDMI	0.002967	24.9227	0.001063

Note: MS is the mean square error; F-cirtical value = 5.31766; $P < 0.05$.

This indicates that the improvement in the model generalisation performance from the FS methods proposed in this study is systematic and not random (Haque et al., 2016).

5.3. Discussions

5.3.1. Interpretation of feature selection

As outlined above, the PGI-SDMI algorithm introduced in this study demonstrated superior performance in enhancing model prediction accuracy. This section elucidates the factors contributing to the exceptional performance of the PGI-SDMI. This was achieved by examining the specific RIFs underlying the selection of PGI-SDMI and contrasting the feature-ordering differences between PGI-SDMI and the other methods.

(1) Specific RIFs underlying the selection of PGI-SDMI

In this section, the interpretability of the PGI-SDMI is demonstrated by explaining the selected key RIFs, highlighting its ability to identify crucial RIFs. The results of the RIF ordering for PGI-SDMI with a minimum support of 0.1 and a minimum confidence of 0.3 is shown in Table 6. The order of the first 30 RIFs in Table 6 was obtained using PGI, and the last 38 RIFs were ordered using the SDMI algorithm.

In the interaction process of marine accident RIFs, the PGI-SDMI considers those RIFs that interact with more RIFs and have high transferability of such interactions to be more critical RIFs. Therefore, for M3 (inadequate safety management system) and H5 (operational error), which ranks the highest among the 30 RIFs, the number of effects interacting with them and the transferability of the interaction information are much higher than those for the lower-ranked RIFs. For example, an inadequate safety management system can lead to a lack of safety culture in a shipping company and increase the likelihood that training will not be conducted as planned, leading to a change in accident severity (Cao et al., 2023). Furthermore, the PGI-SDMI suggests that the greater the SDMI between the RIFs and accident severity, the greater the utility of the RIF in discriminating accident consequences within the model. In particular, for decision-tree models, RIFs with large SDMI are beneficial for determining the consequences of accidents. Based on this, five critical conclusions were drawn.

1) Human factors are leading causes of marine accidents. Operational errors, rule violations, lack of theoretical knowledge, and crew members' inexperience significantly affect the severity of marine accidents (Wang et al., 2021a). Therefore, human factors such as H5, H6, H3, and H2 are generally more critical.

2) Management factors and accident severity also have a solid causal relationship (Cao et al., 2023). When the state of the management factors changes, the probability of a change in accident severity also increases. Therefore, the management factors M3, M2, M6, M7, M5, and M4 are generally essential.

3) The location of waterways and traffic conditions are also essential environmental factors. In line with numerous studies (Wang

Table 6
RIF ranking based on PGI-SDMI.

No.	RIF	No.	RIF	No.	RIF	No.	RIF
1	M3	18	ED3	35	M1	52	T5
2	H5	19	ST1	36	EC	53	ST8
3	H6	20	AE1	37	EL1	54	EW3
4	SG3	21	ED4	38	SA3	55	ST5
5	AE3	22	ED2	39	T6	56	T1
6	M2	23	ST7	40	A4	57	T4
7	ET	24	A1	41	A6	58	ST4
8	EL2	25	H2	42	AE2	59	ST3
9	M6	26	T2	43	ES1	60	A7
10	M7	27	T3	44	SV2	61	ST6
11	M5	28	A2	45	A3	62	SV4
12	M4	29	EV2	46	SG2	63	ES2
13	SA1	30	EV1	47	ST2	64	EL4
14	SG1	31	ED1	48	SV3	65	H4
15	EL3	32	A5	49	EW2	66	H1
16	H3	33	SV5	50	ST10	67	ST9
17	SA2	34	SA4	51	SV1	68	EW1

et al., 2023b; Yang et al., 2022), it is believed that regions characterised by high traffic density exhibit heightened susceptibility to accidents. Traffic density is typically more pronounced in harbours and coastal waters than in other water areas. However, coastal waters are more prone to serious accidents, and port waters are more prone to general accidents (Cao et al., 2023). Therefore, the environmental factors ET, EL2, and EL3 are also important.

4) Bulk carriers and fishing vessels were the most important RIFs in the ship factors. First, bulk carriers and fishing vessels were the primary types of vessels (as shown in Fig. 6). Second, when analysing the degree of lift of bulk carriers on the probability of occurrence of different accident severities, it was found that the degree of lift of bulk carriers on the likelihood of an event of general accidents is 1.45, and it has an inhibiting effect on the probability of occurrence of serious accidents. Furthermore, this study highlights bulk carriers as exhibiting the most pronounced impact on collision occurrence probability, thus categorising them among vessel types with elevated collision risks. This finding contrasts with some studies that designated specialised cargo ships (Antão et al., 2023) or tankers (Li et al., 2023b) as having the highest collision risks, potentially owing to disparities in the dataset sources or assessment methodologies. Finally, accidents involving fishing vessels are notably more likely to result in serious consequences than those involving other vessel types, as indicated in previous research (Bye and Aalberg, 2018). Thus, the ship factors ST7 and ST1 are also relatively significant.

5) Collision, Stranding/Ground, Capsize/Sinking, and 4 a.m. to 12p.m. are important RIFs in accident information. On the one hand, collisions, Stranding/Ground accidents are the main types of accidents (as shown in Fig. 6). At the same time, in alignment with the consensus of most studies (Jin, 2014; Weng et al., 2016), this study underscores capsize/sinking accidents as the most severe incidents leading to casualties. On the other hand, this study reveals that the period from 4 a.m. to 12p.m. is the most frequent (as shown in Fig. 6), with this period having an increasing effect on the probability of general accidents and inhibiting the occurrence of serious accidents and the impact of this period on the severity of accidents is greater than that of the other periods. Hence, A1, A2, A5, T2, and T3 are relatively more important than the other RIFs for accident information.

(2) Differences in feature ordering

To explain the variations in the effectiveness of PGI-SDMI compared to other methods, we trained these diverse FS methods on the training dataset and calculated the correlation among the resulting RIF rankings using Equation (7). Fig. 12 shows the correlation plots of the different RIF rankings.

In Fig. 12, it is evident that 3-WLR is similar to both PageRank and WLR in terms of ranking RIFs by PGI. This suggests that the 3-WLR algorithm proposed in this study alters the sequence of RIFs to a lesser degree but significantly improves the stability and predictive capability of the FS method. Furthermore, the results of SDMI and TMI for sorting the RIFs were comparable. This indicates that, compared to the perspective of overall similar-information FS, selecting features based on the difference in the contribution of feature states to the target can accurately adjust the ordering of specific features, thereby substantially improving the FS performance.

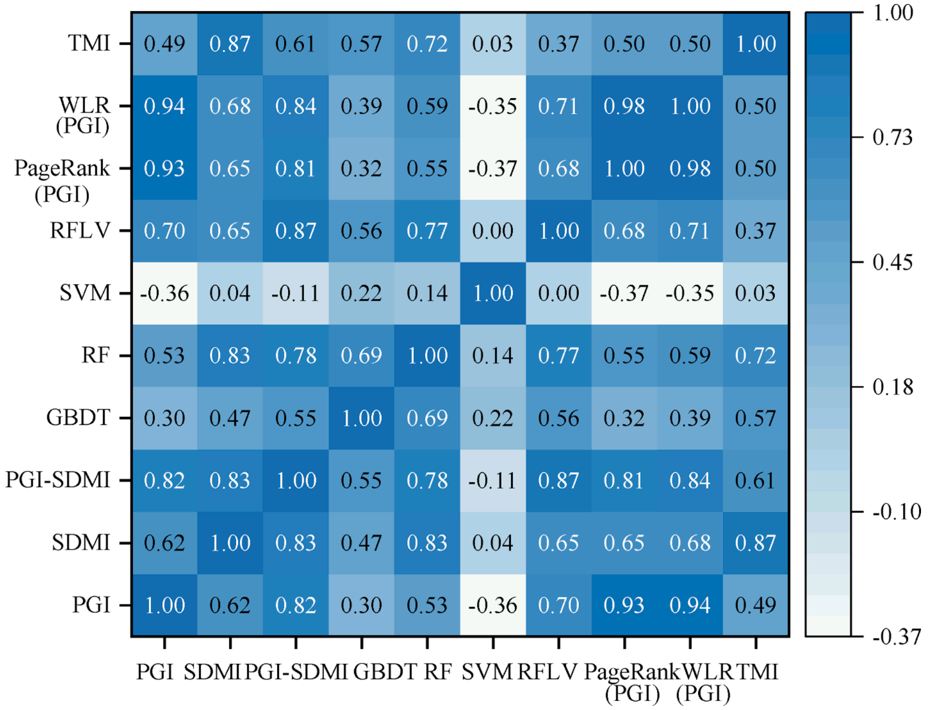

Fig. 12. The thermodynamic chart of RIF rankings' correlation coefficients based on different FS methods.

However, the variance in RIF rankings between PGI and PGI-SDMI is primarily linked to the parameter configurations and the fact that SDMI ranks specific RIFs in PGI-SDMI.

In Fig. 12, there is a weak correlation between the RIFs selected by the proposed method and those selected by SVM. This suggests a significant difference between the technique for scoring RIFs using hyperplane feature coefficients (Chen et al., 2020) and the ranking results of the FS method proposed in this study. This also indicates that the FS by the proposed methods does not contribute significantly to building SVM prediction models with hyperplanes.

In Fig. 12, it is evident that there is a higher correlation between SDMI and embedded FS based on RF than between PGI, PGI-SDMI and embedded FS based on RF. This can be attributed to the similarity in operational logic between the SDMI algorithm proposed in this study and the Gini coefficient in the RF. The prediction performance of the SDMI was also more significant than that of the PGI when RF was used as a predictor (see Appendix D). PGI-SDMI, which incorporates the SDMI, shares a higher similarity in RIF ranking with embedded FS based on RF than PGI. As a result, the PGI-SDMI has a better prediction performance than the PGI when RF is used as

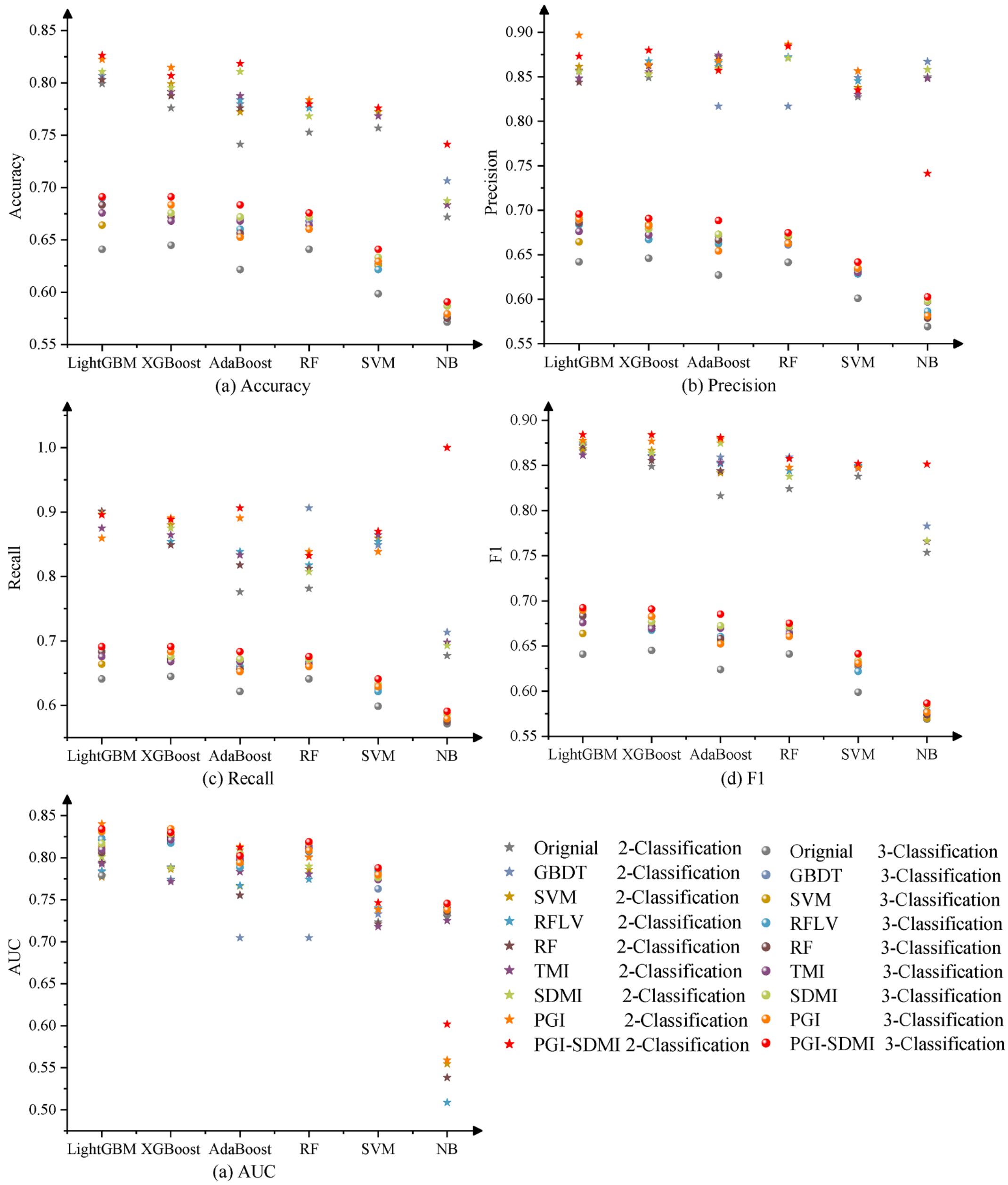

Fig. 13. Comparison of experimental results for predicting the severity of three-classification of accidents versus the severity of two-classification of accidents. (Note: The accuracy, precision, recall, and F1 metrics for the three-classification experimental results are computed using weighted calculations, while the AUC metrics are based on a "one vs one" strategy.).

a predictor (refer to Appendix D).

5.3.2. Impact of multi-classified accident severity on research results

The main goal of this study was to provide technical support for stakeholders in developing risk control measures through accurate predictions of accident severity, with the aim of reducing casualties and pollution occurrence probabilities. Consequently, accident severity was simplified into a binary classification problem, distinguishing between non-serious accidents (lacking casualties or pollution) and serious accidents (involving casualties or pollution). This simplification enabled us to focus on essential indicators without delving into specific levels of casualties or pollution. Although this simplification may not fully capture the nuanced nature of the accident severity, it aligns with the objectives of effective risk control and prevention strategies. However, this simplification may not fully align with maritime practices. Hence, to delve into the distinctions between the binary and multi-classification prediction results, this section refers to the IMO standards (Cao et al., 2023; Wang et al., 2021a) and categorises accident severity into three classes: less serious accidents (25.81 %), serious accidents (34.70 %), and very serious accidents (39.49 %). The results of the two- and three-classification comparison experiments are presented in Fig. 13.

Based on the experimental results in Fig. 13, regardless of the predictor and FS method employed, the evaluation metrics such as Accuracy, Precision, Recall, and F1 for the three-classification, compared to the two-classification prediction, exhibited a significant decrease in the model's generalisation ability. This suggests that the capacity of the ML model to discriminate among various specific categories diminishes when performing multi-classification prediction. This phenomenon is attributed to the greater feature space of the multi-classification issue, which leads to challenges such as sample overlap and conflicts between multiple categories. Consequently, it becomes more challenging for the model to establish a precise decision boundary for the accurate discrimination between different categories. These results underscore the primary necessity for this study to focus on binary prediction to enhance model accuracy and target accident severity reduction.

6. Implications

6.1. Theoretical implications

To demonstrate the applicability of the methodology proposed in this study, this section evaluates the benefits that can be achieved by controlling RIFs using Single-Learner (Dai et al., 2021) as an example of controlling human or management factors. The analysis utilised LightGBM, a ML model with superior predictive performance, along with an optimal subset of features comprising 41 RIFs selected by the PGI-SDMI, as the foundation for evaluation.

6.1.1. Theoretical insights for single-risk control measures

The human and managerial factors in the 41 RIFs of the PGI-SDMI were controlled to discern specific post-control benefits, as depicted in Fig. 14. The graphical representation in Fig. 14 illustrates that shipping companies can achieve optimal benefits, denoted by control H3, by engaging experienced crew members. Furthermore, adherence to International Convention the Standards of Training, Certification and Watchkeeping for Seafarers (STCW) is imperative for shipping companies, as it explicitly mandates the assignment of experienced and competent crew members to safety–critical operational duties and facilitates the proactive enhancement of crew safety management competencies. This measure contributed to averting 10.12 % of serious accidents. The rationale behind this outcome lies in the superior adaptability of experienced crews to varying conditions, their ability to minimise risks in unfamiliar waters or harbours, mitigate psychophysical discomfort resulting from ship movements, and respond effectively to accidents (Jasionowski, 2011). For example, the sinking of MV Estonia saw 137 crews and passengers spared due to the extensive

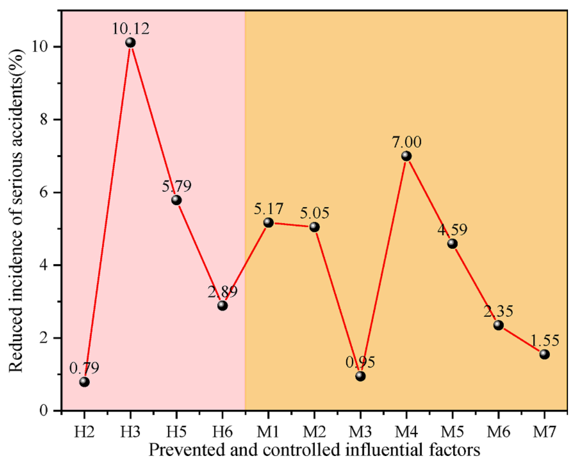

Fig. 14. The benefit from a single-risk control measure.

experience and swift response of the crew, leading to rapid intervention within the initial 10–20 min (Jasionowski, 2011).

Another effective strategy involved the prompt correction of safety issues (Control M4), resulting in a 7.00 % reduction in serious accidents. Timely addressing of comprehensive problems can help reduce hazards, such as ship or equipment failures, thereby enhancing the overall safety performance and diminishing the occurrence of serious accidents. This approach aligns with the SOLAS convention and ISM code, which prioritize the establishment of a fundamental safe environment for ship operations. Moreover, promptly addressing issues contributes to enhancing crew training, improving safety awareness, fostering a positive safety culture within a company, and generating positive feedback.

Conversely, controlling either H2 or M3 was identified as the least effective strategy. This observation may stem from the fact that this study advocates an approach that focuses on individual estimation of the whole. Consequently, controlling RIFs like "Operational error" and "Violation of rules and regulations" is more likely to yield a positive impact at an individual level, significantly reducing the severity of marine accidents.

An interesting phenomenon can be found that M3 is shown to be the most important RIF in Table 6, but controlling M3 is one of the least effective strategies in Fig. 14. The low effectiveness of controlling M3 in reducing serious accidents is mainly due to the fact that M3 promotes the occurrence of non-serious accidents and inhibits the occurrence of serious accidents. Regarding the result that M3 is the most important in prediction, first, it is most important when viewed from the perspective of the entire accident system, which contains both serious and non-serious accidents. Second, improving the accuracy of the prediction is mainly dependent on the ability of RIFs to distinguish between serious and non-serious accidents rather than relying on whether controlling RIFs can effectively reduce the severity of the accident.

6.1.2. Theoretical insights for multi-risk control measures

In this section, critical human and management factors are controlled together in a multifactorial manner and their benefits are analysed. First, if the shipping company employs experienced and low operational error rate crew (control H3 and H5), the serious accident rate will be reduced by 16.18 %; therefore, if the authority improves the relevant regulations to control M1, 21.05 % of serious accidents will not occur. Second, if the authority improves the applicable regulations and implements adequate supervision, the shipping company corrects the safety problems in time, establishes a good safety culture, reinforces the training of the crew on board and conducts drills according to the plan with the requirements of the SOLAS; this indicates that if M1, M2, M4, M5, M6, and M7 are effectively controlled, 25 % of serious accidents will not occur.

In areas with heavy traffic (ET), the most common environmental factor, the incidence of serious accidents can be reduced by 61.54 % if the authority improves relevant regulations (control M1), the shipping company employs experienced crew (control H3), promptly addresses safety issues (control M4), establishes a strong safety culture (control M5), and conducts training according to the plan (control M7). In coastal waters (EL3), which have a high traffic density, the incidence of serious accidents can be reduced by 17.39 % if the authority improves the relevant regulations (control M1), provides adequate supervision (control M2), employs crews with a low error rate (control H5), promptly addresses safety issues (control M4), establishes a strong safety culture (control M5), conducts onboard crew training (control M6), and schedules drills (control M7).

Considering the vessel type, the probability of serious accidents is much higher for fishing vessels than for other vessels. For fishing vessels, if the shipping company employs an experienced crew, low error rate, and rule-abiding crew and establishes a sound safety management system, it corrects safety problems promptly, establishes a good safety culture, strengthens onboard crew training, and conducts scheduled drills; that is, H3, H5, H6, M4, M5, M6, and M7 are effectively controlled, and the serious accident rate will be reduced by 40.00 %.

In summary, in the vast majority of cases, following the requirements of the SOLAS convention, ISM code and STCW convention, effectively reducing the severity of accidents can be achieved by employing experienced crew members, clearly assigning key operations to competent personnel, actively conducting on-board training and drills to enhance responsiveness to emergencies, and improving the safety of the ship's operating environment.

6.2. Practical implications

This study has significant practical implications for various stakeholders and researchers in the maritime industry, particularly in terms of the following five aspects: risk control, accident investigation, emergency rescue, maritime education, and accident data.

6.2.1. Practical insights for integrated risk control strategies

This encompasses both Macro-Level risk control optimisation and Micro-Level navigation plan adjustment. At the macro-level, the entire risk system can be evaluated using the accident severity prediction model developed in this study. This evaluation leads to the formulation of comprehensive risk management policies and decisions aimed at enhancing the overall risk resilience of the organization. Conversely, the navigation plan adjustment operates at the micro level, concentrating on specific operational adjustments to the navigation plan. Accident severity prediction models play a crucial role in minimising the probability of potentially serious accidents.

At the macro-level, stakeholders such as shipping companies or authorities can leverage the accident severity prediction model developed in this study along with counterfactual analyses to evaluate the effectiveness of risk control measures (as detailed in Section 6.1). Building on this analysis, stakeholders can optimise strategy combinations to identify macro-optimal strategies that minimise accident severity. As an illustration, fishing vessel companies can apply the model from this study to determine the efficacy of strategies, such as enhancing crew experience and improving company safety culture, in reducing accident severity. Subsequently, the vessel company can elevate its company-wide safety culture at the macro level, enhance the overall experience level of the crew, and recruit experienced crew members.

At the micro level, by segmenting the navigational area and integrating RIFs, such as the month, shipping companies or crews can leverage the model developed in this study to evaluate the potential severity of accidents in various regions, seasons, and time periods. This strategic approach offers precise decision support to shipping companies or their crews. With real-time monitoring and dynamic adjustment, ships can respond more effectively to changing environments and risks, thereby enhancing maritime safety. For instance, at sea, the crew can anticipate the severity of the next likely accident based on weather conditions, time of day, month, and sailing area. This enables them to modify the sailing route to avoid high-risk areas prone to serious accidents, or prepare in advance to manage the expected severity of the accident.

6.2.2. Practical insights for accident liability analysis

Marine accident investigation agencies and relevant authorities can employ the models and methods developed in this study to aid the identification and quantification of liabilities in marine accidents. By analysing the influence of RIFs on accident severity, it is possible to identify the main causes of marine accidents and ascertain the potential liability. Simultaneously, counterfactual analyses of individual accidents can be conducted to evaluate the impact of altering the values of specific RIFs on accident severity, thereby offering a foundation for quantifying liabilities. The model and methodology proposed in this study can provide marine accident investigation agencies or experts with a scientific basis and comprehensive perspective to determine accident liability more accurately. For example, after an accident occurs, investigators can use the model of this study to analyse the various RIFs existing in the accident, quantitatively determine the impact of the RIFs on the severity of the accident, and the greater the impact on the severity of the factors involved in the responsible party, the greater the responsibility for the accident.

6.2.3. Practical insights for emergency rescue

This encompasses emergency rescue capability assessments and emergency rescue plan formulations. The model and methodology presented in this study will enable maritime authorities to gain a deeper understanding of the effectiveness of emergency rescues in various scenarios. This understanding facilitates the continuous optimisation and improvement of rescue capabilities.

Given that this study did not explicitly consider the impact of emergency response on accident severity in severity prediction, emergency response departments and stakeholders could compare the severity of an accident with an active emergency response to the severity predicted by this study's model. This comparison facilitates a preliminary assessment of the effectiveness of the emergency response capabilities. For instance, if an emergency response department actively addresses a marine accident, the model can be utilised to predict the severity of the accident, allowing for a comparison with the post-response severity. This evaluation aids in assessing the effectiveness of the response efforts.

In addition, by building on the evaluation of emergency response capacity mentioned above, this study can serve as a foundation for crafting emergency response plans. By simulating various marine accident scenarios and demonstrating the influence of diverse emergency rescue strategies on accident severity, maritime authorities can develop more rational and efficient emergency response plans. This ensures swift implementation of the most suitable rescue measures during accidents, thereby enhancing the overall rescue efficiency. For instance, maritime authorities can utilise the abundance of maritime rescue accident data to evaluate the effectiveness of each response and subsequently summarise the optimal response strategy for each scenario.

6.2.4. Practical insights for maritime education and training

Maritime education and training institutions, along with other stakeholders, can leverage this model and methodology to develop instructional materials for simulating accident scenarios. These materials can be used to demonstrate the actual consequences and severity of various RIFs to trainees and crew members, thereby enhancing their safety awareness. Moreover, the predictive models presented in this study can facilitate counterfactual analyses, enabling rehearsal of the actual effects of the measures taken by trainees under different risk scenarios. This process contributes to an improved understanding of the risk control strategies.

6.2.5. Practical insights for enhanced data enrichment for accident analysis

The predictive models proposed in this study offer stakeholders and researchers an approach for enhanced data refinement and supplementation in accident analysis. Although the primary focus is accident severity prediction, the methodology and model presented herein provide a generalisable framework with the potential to address missing or ambiguous information in accident data. Conversely, using predictive modelling and FS, stakeholders can repair and supplement the accident data, thereby enhancing its accuracy and completeness. For instance, in current marine accident research, errors or missing valid information may occur because of variations among different investigating organisations or omissions by recorders. In cases of missing information, researchers often resort to deleting data, leading to a significant deviation from the actual situation. Therefore, during dataset construction, researchers can employ the model developed in this study and existing information to predict missing or incorrect information, thereby ensuring data completeness. Accident investigation organisations can utilise this study's model to identify the key RIFs in accidents. This enabled them to conduct targeted investigations, thereby enhancing data completeness and reliability. This practical application opens new opportunities for improving the quality of marine accident data and offers more reliable support for data-driven decision-making processes.

In conclusion, this study not only contributes significantly to enhancing risk control, accident investigation, emergency rescue capabilities, maritime education and training, and accident data quality in the maritime industry but also provides a comprehensive framework for stakeholders and researchers to make informed decisions and advancements.

7. Conclusion

This study presents a comprehensive framework that employs an enhanced ML approach to predict the severity of marine accidents. This framework comprises four crucial aspects: dataset construction, RIFs selection, FS performance evaluation, and accident severity prediction.

A novel two-stage FS method, PGI-SDMI, was proposed considering both the interactions among RIFs and their contributions to severity. Furthermore, a framework to assess the performance of these methods and a convergence algorithm for evaluating the FS stability were introduced. In addition, this study evaluated the predictive performance of six advanced ML models: LightGBM, XGBoost, AdaBoost, RF, SVM, and NB.

The results indicate a substantial improvement in the performance of the enhanced 3-WLR algorithm compared with PageRank and WLR. Similarly, the SDMI algorithm proposed in this study, which considers the differences in the contributions of different feature states to the target, exhibited considerable improvement compared to TMI. Moreover, the two-stage model PGI-SDMI surpasses other FS methods, outperforming the individual uses of the PGI and SDMI. Among the ML models, LightGBM emerged as the top performer in terms of prediction performance.

From a practical standpoint, upon analysing the counterfactual scenario of reducing accident severity, this study concludes that priority should be given to enhancing the experience of crew members and proactively addressing issues by shipping companies.

This study proposes an advanced data-driven FS algorithm and an algorithm performance evaluation method. It makes a notable contribution by employing counterfactual techniques to assess the benefits achievable through controlling RIFs. However, there are still some limitations in this study that warrant attention and resolution in future research. First, the creation of a dataset requires subjective input (e.g., classification). In the future, text mining techniques based on deep learning can be utilised to enhance the construction of datasets. Second, the dataset used in this study did not entirely encompass the latest accident report data, which introduces certain limitations. Future research may explore changes in the relationship between RIFs over an extended time scale, variations in their impact on accident severity over time, and alterations in the influence of RIFs on accident severity during emergencies, such as the COVID-19 pandemic. Concurrently, future investigations could also contemplate the aspect of risk uncertainty, particularly in unforeseen circumstances. Third, the work has yet to take full advantage of the interpretability of the new FS method owing to word limitations. In the future, when more space becomes available, a more detailed analysis of interpretability can be provided; hence, the results of FS can be analysed extensively from the perspective of marine accidents. Finally, this study only compared the performances of six advanced ML models and did not introduce any new models. Therefore, further efforts can be made to develop more advanced ML or deep learning models to improve the prediction performance in the future.

CRediT authorship contribution statement

Yinwei Feng: Writing – original draft, Validation, Methodology, Conceptualization. **Xinjian Wang:** Writing – original draft, Methodology, Funding acquisition, Conceptualization. **Qilei Chen:** Writing – review & editing, Validation, Data curation. **Zaili Yang:** Writing – review & editing, Validation, Funding acquisition, Formal analysis. **Jin Wang:** Writing – review & editing, Validation, Formal analysis. **Huanhuan Li:** Writing – review & editing, Visualization, Validation. **Guoqing Xia:** Writing – review & editing, Validation, Data curation. **Zhengjiang Liu:** Validation, Funding acquisition, Conceptualization.

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

Data will be made available on request.

Acknowledgements

The authors gratefully acknowledge support from the National Natural Science Foundation of China [Grant No. 52101399] and the Bolian Research Funds of Dalian Maritime University (Grant No. 3132023617). This work was also supported by EU H2020 ERC Consolidator Grant program (TRUST Grant No. 864724).

Appendix A. Feature description of MARIFD

Table A
Description of the 68 RIFs.

RIFs		Variables	Description
Human		H1	Poor state of physical & psychological
		H2	Insufficient education and training
		H3	Lack of experience
		H4	Poor communication
		H5	Operational error
		H6	Violation rules and regulations
Ship	Ship type	ST1	Bulk carrier
		ST2	Container ship
		ST3	Oil tanker
		ST4	Passenger ship (including cruise and ro-ro passenger ship)
		ST5	Chemical tanker
		ST6	General cargo ship
		ST7	Fishing vessel
		ST8	Yacht and sailing vessel
		ST9	Tug and port traffic boat
		ST10	Others
	Ship age	SA1	0–10 years
		SA2	10–20 years
		SA3	20–30 years
		SA4	>=30 years
	Gross tonnage	SG1	<500 t
		SG2	500–3000 t
		SG3	>=3000 t
	Engine power	SE1	<750KW
		SE2	750-3000KW
		SE3	>=3000KW
	Voyage data	SV1	Incomplete or invalid ship certificates
		SV2	Insufficient manning
		SV3	Incomplete or invalid seafarer certificates
		SV4	Poor seaworthiness (passenger/ cargo not meet the requirements)
		SV5	Uncertain PSC/FSC inspection (Defective ship)
Management		M1	Inadequate regulation
		M2	Inadequate supervision
		M3	Defective safety management system
		M4	Unresponsive rectification of problems
		M5	Poor company safety culture
		M6	Inadequate ship training
		M7	Not drill by sticking to the schedule
Environment	Location	EL1	Inland waters
		EL2	Port
		EL3	Coastal waters
		EL4	Open sea
	Visibility	EV1	Poor visibility (0.5–2 nm)
		EV2	Very Poor(<0.5 nm)
	Wind force	EW1	Strong wind (6–7)
		EW2	Very strong wind (8–9)
		EW3	Strom (10–12)
	Sea State	ES1	Poor sea state (6–7)
		ES2	Very poor sea state (8–9)
	Current	EC	Current speed >=4kn
	Traffic density	ET	Heavy traffic
	Depth-draft ratio (h/d)	ED1	h/d < 1.2
		ED2	1.2 <= h/d < 1.5
		ED3	1.5 <= h/d < 3
		ED4	h/d >= 3
Accident	Accident type	A1	Collision
		A2	Stranding/Ground
		A3	Fire/Explosion
		A4	Contact
		A5	Capsize/Sinking
		A6	Hull/Machinery damage
		A7	Other
	Time	T1	0000–0400
		T2	0400–0800

(continued on next page)

Table A (*continued*)

RIFs	Variables	Description
	T3	0800–1200
	T4	1200–1600
	T5	1600–2000
	T6	2000–2400

Note: The unit for wind force is "Beaufort scale"; the unit for sea state is "Douglas Sea Scale.".

Appendix B. Pseudo-codes of the algorithms of this study

Table B1
Pseudo-code of PGI-SDMI algorithm.

Algorithm 1. Two-stage PGI-SDMI Approach

Require: Dataset *DS*
Ensure: Final Feature Ranking List *FinalRankings*

Stage 1: PGI
Initialize feature ranking list: *Rankings* = [].
for each feature f in the dataset **do**
  Calculate the importance of feature f using PGI: *Importance*(f).
  Append the feature and its importance to *Rankings*.
end for
Sort *Rankings* in descending order based on feature importance.
Stage 2: SDMI
Initialize complementary feature ranking list: *ComplementaryRankings* = [].
for each feature f not included in *Rankings* **do**
  Initialize *SDMI_score*(f) = 0.
  for each target variable state s **do**
    Calculate state difference mutual information between feature f and target variable state s: *SDMI_score*(f_s).
    Add *SDMI_score*(f_s) to *SDMI_score*(f).
  end for
  Append the feature and its *SDMI_score*(f) to *ComplementaryRankings*.
end for
Sort *ComplementaryRankings* in descending order based on *SDMI_score*(f).
Concatenate Rankings
Initialize final feature ranking list: *FinalRankings* = *Rankings* + *ComplementaryRankings*.
return *FinalRankings*

Table B2
Pseudo-code of PGI algorithm.

Algorithm 2. Pattern Graph Insight (PGI)

Require: RIFs extracted by ARM
Ensure: Ranked importance of RIFs based on 3-WLR algorithm

Construct Complex Influential Factor Interaction Network (CIFIN) based on association rule mining results.
Calculate weights of RIFs:

$$\mu_i = \left(\frac{B_i}{\sum B_m} + \frac{S_i^{out}}{\sum S_m}\right)/2, i \in \{1, 2, ..., n\}$$

Add a Ground node g to the CIFIN
Initialize the transfer probability matrix:

$$M = (\delta_{ij})_{(n+1)\times(n+1)}$$

Define transfer probabilities:

$$\delta_{ij} = \begin{cases} k_j^{in}, \text{ when } i = g \text{ and } j \neq g \\ 1, \text{ when } i \neq g \text{ and } j = g \\ C(i \Rightarrow j), \text{ when there exists } i \Rightarrow j \text{ and } i \neq j \neq g \\ 0, \text{ else} \end{cases}$$

Initialize scores of all nodes to 1.
while not converged **do**
  Compute H_{ij} matrix.
  Compute normalization matrix Z_j^*.
  Update scores of all nodes:

$$L_i(t+1) = \sum_{j=1}^{n+1} \frac{M \odot H_{ij}}{Z_j{}^* M^* O} L_j(t), i, j \in \{1, 2, ..., n, g\}.$$

end while
Compute final scores of RIFs:

$$R_i = L_i(t_{end}) + \mu_i L_g(t_{end}), i \in \{1, 2, ..., n\}.$$

Rank RIFs based on their importance R_i.
return Ranking of RIFs

Table B3
Pseudo-code of SDMI algorithm.

Algorithm 3. State Difference Mutual Information (SDMI)

Require: RIF, X; Label, Y
Ensure: State Difference Mutual Information, $SDMI(X,Y)$

Initialize $SDMI(X,Y) = 0$
Determine the number of states of the label, A.
Determine the number of states of the RIF, D.
for $a = 1$ to A **do**
  for $p = 1$ to D **do**
    Calculate mutual information between state p of RIF X and state a of label Y: $I(X_p, Y_a)$.
    for $q = 1$ to D **do**
      Calculate mutual information between state q of RIF X and state a of label Y: $I(X_p, Y_a)$.
      Calculate the difference of mutual information: $(I(X_p, Y_a) - I(X_q, Y_a))^2$.
      Add the difference to a running total: $\sum_{p=1}^{D}\sum_{q=1}^{D}(I(X_p, Y_a) - I(X_q, Y_a))^2$.
    end for
  end for
  Calculate the square root of the running total divided by $2D$: $\frac{\left(\sum_{p=1}^{D}\sum_{q=1}^{D}(I(X_p, Y_a) - I(X_q, Y_a))^2\right)^{1/2}}{2D}$.
  Add the result to $SDMI(X,Y)$.
end for
return $SDMI(X,Y)$

Table B4
Pseudo-code of Stability measurement algorithm.

Algorithm 4 Stability measurement algorithm for feature selection

Require: Dataset DS, number of copies K, number of features to select N, number of experiments T, feature selection method FS
Ensure: Stability of feature selection when N features are selected

Select the top N features from DS using FS and obtain the reference ranking RR.
Generate a $T-$dimensional vector AS with all elements set to 0.
for $t = 1$ to T **do**
  Generate a random number $k \in K$.
  Randomly divide DS into k parts: $DS_1^{(t)}, DS_2^{(t)}, \ldots, DS_k^{(t)}$, where each part has $\frac{|DS|}{k}$ samples and $DS_i^{(t)}$ is the i^{th} part.
  for $j = 1$ to k **do**
    Let $DSA = DS_j^{(t)}$ and $DSB = \bigcup_{l \neq j} DS_l^{(t)}$.
    Select the top N features from DSA and DSB using FS and obtain the ranking $R_{DSA}^{(j)}$ and $R_{DSB}^{(j)}$.
    Compute the Spearman's rho of between $R_{DSA}^{(j)}$ and RR: $Sprr^{R_{DSA}^{(j)},RR} = 1 - \frac{6\sum_{i \in V}(R_{DSA_i}^{(j)} - RR_i)^2}{N(N-1)(N+1)}$
    Compute the Spearman's rho of between $R_{DSB}^{(j)}$ and RR: $Sprr^{R_{DSB}^{(j)},RR} = 1 - \frac{6\sum_{i \in V}(R_{DSB_i}^{(j)} - RR_i)^2}{N(N-1)(N+1)}$
    Let $Sprr_j = Sprr^{R_{DSA}^{(j)},RR} + Sprr^{R_{DSB}^{(j)},RR}$.
  end for
  Calculation the average stability score: $S_t = \frac{1}{k}\sum_{j=1}^{k} Sprr_j$.
  Update $AS_t = S_t$.
end for
return Stability of feature selection AS.

Appendix C. Performance of PGI and PGI-SDMI based on different parameters

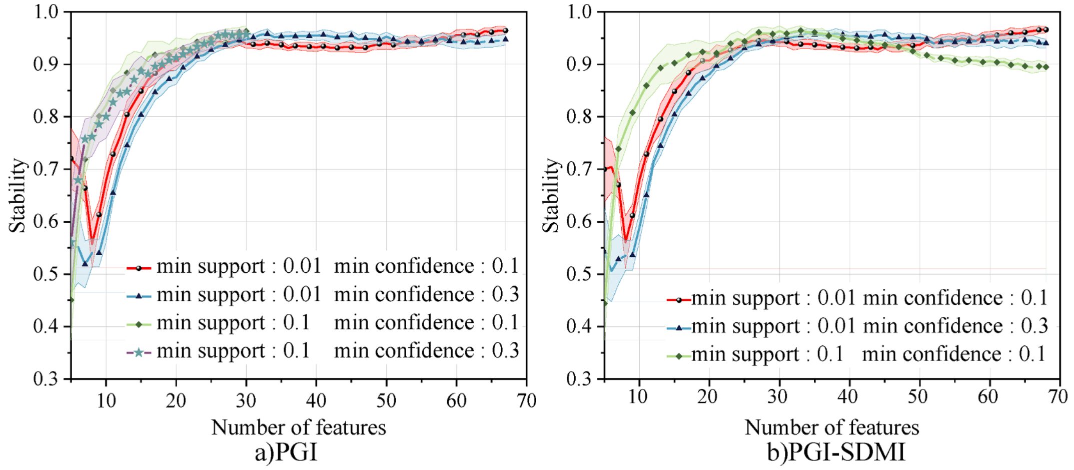

Fig. C1. Stability of PGI and PGI-SDMI.

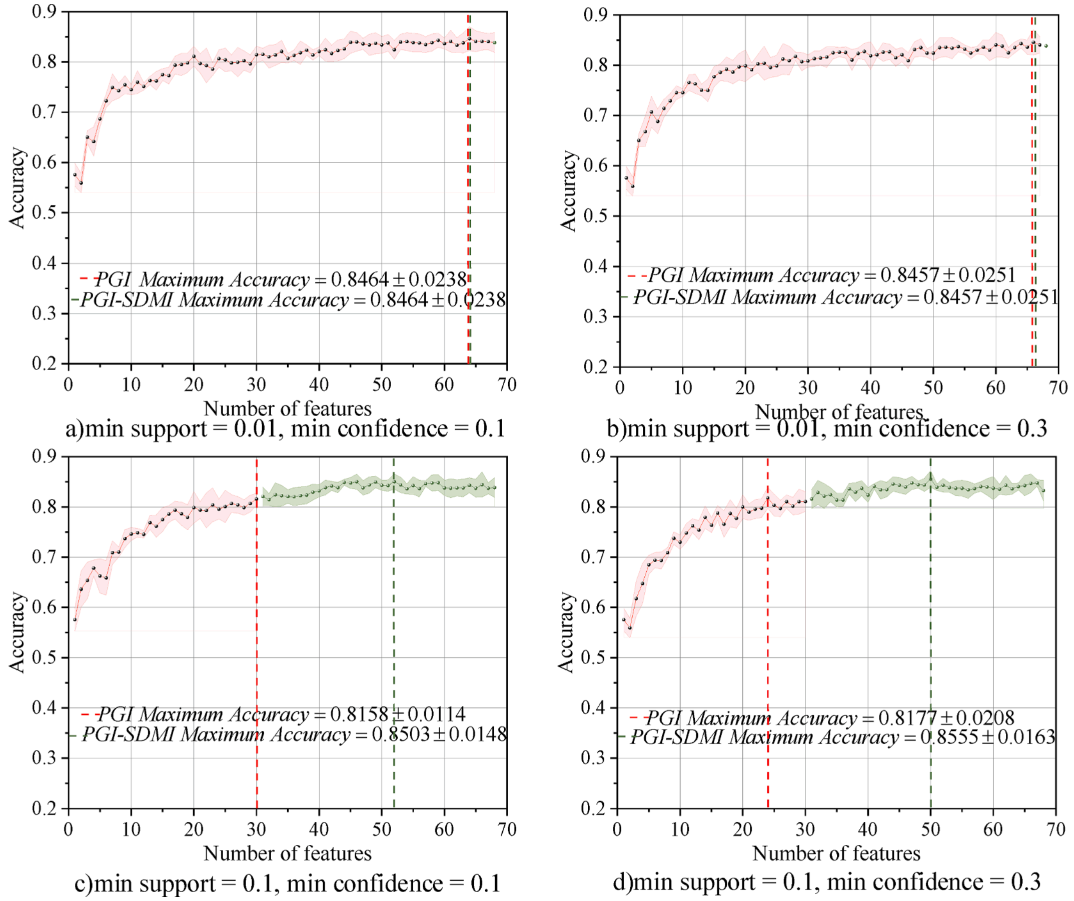

Fig. C2. Prediction accuracy based on PGI and PGI-SDMI (LightGBM).

Appendix D. Tables of predictors' accuracy based on different FS methods

Table D1
Prediction accuracy based on RF.

FS Methods	Accuracy	Precision	Recall	F1	AUC
PGI	0.8073 ± 0.0328	0.8523 ± 0.0434	0.7449 ± 0.0327	0.7944 ± 0.0311	0.8697 ± 0.0256 ≠
SDMI	0.8145 ± 0.0176	0.8557 ± 0.0356	0.7594 ± 0.0301	0.8036 ± 0.0147	0.8849 ± 0.0185
PGI-SDMI	0.8099 ± 0.0194	0.8622 ± 0.0327	0.7400 ± 0.0266	0.7957 ± 0.0152	0.8858 ± 0.0189
RF	0.8158 ± 0.0288	0.8719 ± 0.0351	0.7416 ± 0.0304	0.8010 ± 0.0266	0.8819 ± 0.0213
GBDT	0.8138 ± 0.0282	0.8586 ± 0.0371	0.7584 ± 0.0447	0.8042 ± 0.0292	0.8789 ± 0.0171
SVM	0.8027 ± 0.0232	0.8408 ± 0.0228	0.7495 ± 0.0460	0.7912 ± 0.0232	0.8849 ± 0.0180
RFLV	0.8132 ± 0.0313	0.8490 ± 0.0286	0.7632 ± 0.0303	0.8034 ± 0.0281	0.8849 ± 0.0120
PageRank	0.8014 ± 0.0209	0.8302 ± 0.0339	0.7593 ± 0.0320	0.7925 ± 0.0225	0.8692 ± 0.0181
WLR	0.8067 ± 0.0357	0.8562 ± 0.0379	0.7372 ± 0.0465	0.7916 ± 0.0374	0.8755 ± 0.0206
TMI	0.8066 ± 0.0185	0.8459 ± 0.0347	0.7529 ± 0.0423	0.7953 ± 0.0215	0.8823 ± 0.0177

Table D2
Prediction accuracy based on SVM.

FS Methods	Accuracy	Precision	Recall	F1	AUC
PGI	0.8255 ± 0.0208	0.8090 ± 0.0233	0.8534 ± 0.0272	0.8302 ± 0.0178	0.8901 ± 0.0199
SDMI	0.8223 ± 0.0145	0.8069 ± 0.0225	0.8482 ± 0.0243	0.8266 ± 0.0135	0.8939 ± 0.0213
PGI-SDMI	0.8223 ± 0.0145	0.8069 ± 0.0225	0.8482 ± 0.0243	0.8266 ± 0.0135	0.8939 ± 0.0213
RF	0.8203 ± 0.0230	0.8052 ± 0.0245	0.8621 ± 0.0203	0.8326 ± 0.0209	0.8902 ± 0.0256
GBDT	0.8126 ± 0.0311	0.8170 ± 0.0279	0.8282 ± 0.0363	0.8224 ± 0.0301	0.8861 ± 0.0236
SVM	0.8346 ± 0.0124	0.8238 ± 0.0243	0.8516 ± 0.0096	0.8373 ± 0.0133	0.8938 ± 0.0210
RFLV	0.8201 ± 0.0224	0.8044 ± 0.0295	0.8462 ± 0.0180	0.8247 ± 0.0231	0.8910 ± 0.0224
PageRank	0.8203 ± 0.0233	0.8114 ± 0.0358	0.8547 ± 0.0284	0.8323 ± 0.0301	0.8891 ± 0.0233
WLR	0.8203 ± 0.0233	0.8114 ± 0.0358	0.8547 ± 0.0284	0.8323 ± 0.0301	0.8891 ± 0.0233
TMI	0.8197 ± 0.0159	0.8030 ± 0.0234	0.8471 ± 0.0317	0.8243 ± 0.0158	0.8894 ± 0.0212

Table D3
Prediction accuracy based on NB.

FS Methods	Accuracy	Precision	Recall	F1	AUC
PGI	0.7174 ± 0.0379	0.7597 ± 0.0352	0.6357 ± 0.0679	0.6904 ± 0.0476	0.8104 ± 0.0288
SDMI	0.7265 ± 0.0474	0.7690 ± 0.0259	0.6451 ± 0.0808	0.6999 ± 0.0581	0.8071 ± 0.0313
PGI-SDMI	0.7265 ± 0.0474	0.7690 ± 0.0259	0.6451 ± 0.0808	0.6999 ± 0.0581	0.8071 ± 0.0313
RF	0.7265 ± 0.0595	0.7614 ± 0.0621	0.6609 ± 0.0817	0.7054 ± 0.0646	0.8160 ± 0.0301
GBDT	0.7337 ± 0.0285	0.7812 ± 0.0244	0.6505 ± 0.0535	0.7084 ± 0.0342	0.8174 ± 0.0218
SVM	0.7304 ± 0.0376	0.7757 ± 0.0250	0.6477 ± 0.0666	0.7045 ± 0.0455	0.8209 ± 0.0277
RFLV	0.7214 ± 0.0500	0.7584 ± 0.0587	0.6523 ± 0.0683	0.6996 ± 0.0528	0.8107 ± 0.0326
PageRank	0.7122 ± 0.0444	0.7427 ± 0.0398	0.6488 ± 0.0742	0.6908 ± 0.0520	0.8092 ± 0.0292
WLR	0.7122 ± 0.0444	0.7427 ± 0.0398	0.6488 ± 0.0742	0.6908 ± 0.0520	0.8092 ± 0.0292
TMI	0.7194 ± 0.0346	0.7573 ± 0.0385	0.6434 ± 0.0669	0.6956 ± 0.0465	0.8039 ± 0.0290

Table D4
Prediction accuracy based on AdaBoost.

FS Methods	Accuracy	Precision	Recall	F1	AUC
PGI	0.8223 ± 0.0263	0.8555 ± 0.0264	0.7762 ± 0.0446	0.8131 ± 0.0282	0.9039 ± 0.0185
SDMI	0.8327 ± 0.0248	0.8652 ± 0.0362	0.7908 ± 0.0362	0.8253 ± 0.0234	0.9053 ± 0.0153
PGI-SDMI	0.8288 ± 0.0205	0.8690 ± 0.0264	0.7764 ± 0.0337	0.8193 ± 0.0172	0.9031 ± 0.0091
RF	0.8281 ± 0.0189	0.8734 ± 0.0384	0.7713 ± 0.0343	0.8178 ± 0.0148	0.9030 ± 0.0125
GBDT	0.8275 ± 0.0261	0.8704 ± 0.0281	0.7717 ± 0.0252	0.8177 ± 0.0194	0.9023 ± 0.0220
SVM	0.8229 ± 0.0178	0.8590 ± 0.0291	0.7736 ± 0.0316	0.8133 ± 0.0193	0.8986 ± 0.0193
RFLV	0.8294 ± 0.0265	0.8709 ± 0.0173	0.7712 ± 0.0298	0.8175 ± 0.0162	0.9055 ± 0.0107
PageRank	0.8183 ± 0.0224	0.8583 ± 0.0178	0.7635 ± 0.0333	0.8077 ± 0.0206	0.8956 ± 0.0164
WLR	0.8229 ± 0.0245	0.8698 ± 0.0226	0.7612 ± 0.0446	0.8107 ± 0.0256	0.8999 ± 0.0128
TMI	0.8216 ± 0.0147	0.8549 ± 0.0298	0.7764 ± 0.0209	0.8132 ± 0.0126	0.8999 ± 0.0108

Table D5
Prediction accuracy based on XGBoost.

FS Methods	Accuracy	Precision	Recall	F1	AUC
PGI	0.8483 ± 0.0171	0.8630 ± 0.0296	0.8295 ± 0.0212	0.8454 ± 0.0145	0.9256 ± 0.0178
SDMI	0.8483 ± 0.0159	0.8655 ± 0.0332	0.8270 ± 0.0248	0.8450 ± 0.0126	0.9284 ± 0.0132
PGI-SDMI	0.8483 ± 0.0159	0.8655 ± 0.0322	0.8270 ± 0.0248	0.8450 ± 0.0126	0.9284 ± 0.0132
RF	0.8477 ± 0.0123	0.8479 ± 0.0261	0.8386 ± 0.0283	0.8428 ± 0.0189	0.9259 ± 0.0149
GBDT	0.8483 ± 0.0189	0.8623 ± 0.0346	0.8306 ± 0.0243	0.8454 ± 0.0172	0.9270 ± 0.0141

(continued on next page)

Table D5 (*continued*)

FS Methods	Accuracy	Precision	Recall	F1	AUC
SVM	0.8457 ± 0.0265	0.8533 ± 0.0345	0.8358 ± 0.0293	0.8439 ± 0.0243	0.9150 ± 0.0207
RFLV	0.8477 ± 0.0187	0.8612 ± 0.0363	0.8307 ± 0.0182	0.8450 ± 0.0163	0.9270 ± 0.0120
PageRank	0.8477 ± 0.0120	0.8618 ± 0.0148	0.8280 ± 0.0171	0.8444 ± 0.0120	0.9248 ± 0.0145
WLR	0.8477 ± 0.0171	0.8568 ± 0.0305	0.8361 ± 0.0216	0.8458 ± 0.0156	0.9178 ± 0.0142
TMI	0.8477 ± 0.0167	0.8594 ± 0.0363	0.8216 ± 0.0205	0.8371 ± 0.0174	0.9219 ± 0.0171

Appendix E. Supplementary material

Supplementary source code to this article can be found online at https://github.com/FengYinLeo/PGI-SDMI.

References

Antão, P., Sun, S., Teixeira, A.P., Guedes Soares, C., 2023. Quantitative assessment of ship collision risk influencing factors from worldwide accident and fleet data. Reliab. Eng. Syst. Saf. 234, 109166 https://doi.org/10.1016/j.ress.2023.109166.

Bennasar, M., Hicks, Y., Setchi, R., 2015. Feature selection using Joint Mutual Information Maximisation. Expert Syst. Appl. 42 (22), 8520–8532. https://doi.org/10.1016/j.eswa.2015.07.007.

Bye, R.J., Aalberg, A.L., 2018. Maritime navigation accidents and risk indicators: An exploratory statistical analysis using AIS data and accident reports. Reliab. Eng. Syst. Saf. 176, 174–186. https://doi.org/10.1016/j.ress.2018.03.033.

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

Chauvin, C., Lardjane, S., Morel, G., Clostermann, J.-P., Langard, B., 2013. Human and organisational factors in maritime accidents: Analysis of collisions at sea using the HFACS. Accid. Anal. Prev. 59, 26–37. https://doi.org/10.1016/j.aap.2013.05.006.

Chen, T., Shi, X., Wong, Y.D., Yu, X., 2020. Predicting lane-changing risk level based on vehicles' space-series features: A pre-emptive learning approach. Transp. Res. Part C: Emerg. Technol. 116, 102646 https://doi.org/10.1016/j.trc.2020.102646.

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

Dai, L., Liu, Y., Hansen, M., 2021. Modeling go-around occurrence using principal component logistic regression. Transp. Res. Part C: Emerg. Technol. 129, 103262 https://doi.org/10.1016/j.trc.2021.103262.

Fang, S., Liu, Z., Wang, X., Cao, Y., Yang, Z., 2024. Dynamic analysis of emergency evacuation in a rolling passenger ship using a two-layer social force model. Expert Syst. Appl. 247, 123310 https://doi.org/10.1016/j.eswa.2024.123310.

Feng, Y., Wang, H., Xia, G., Cao, W., Li, T., Wang, X., Liu, Z., 2024. A machine learning-based data-driven method for risk analysis of marine accidents. Journal of Marine Engineering & Technology 1–12. https://doi.org/10.1080/20464177.2024.2368914.

Gao, Z., Kampezidou, S.I., Behere, A., Puranik, T.G., Rajaram, D., Mavris, D.N., 2022. Multi-level aircraft feature representation and selection for aviation environmental impact analysis. Transp. Res. Part C: Emerg. Technol. 143, 103824 https://doi.org/10.1016/j.trc.2022.103824.

Griesbach, K., Hoffmann, K.H., Beggiato, M., 2020. Prediction of lane change by echo state networks. Transp. Res. Part C: Emerg. Technol. 121, 102841 https://doi.org/10.1016/j.trc.2020.102841.

Haque, M.M., Oviedo-Trespalacios, O., Debnath, A.K., Washington, S., 2016. Gap acceptance behavior of mobile phone-distracted drivers at roundabouts. Transp. Res. Rec. 2602 (1), 43–51. https://doi.org/10.3141/2602-06.

Hoque, N., Bhattacharyya, D.K., Kalita, J.K., 2014. MIFS-ND: A mutual information-based feature selection method. Expert Syst. Appl. 41 (14), 6371–6385. https://doi.org/10.1016/j.eswa.2014.04.019.

Jasionowski, A., 2011. Decision support for ship flooding crisis management. Ocean Eng. 38 (14), 1568–1581. https://doi.org/10.1016/j.oceaneng.2011.06.002.

Jin, D., 2014. The determinants of fishing vessel accident severity. Accid. Anal. Prev. 66, 1–7. https://doi.org/10.1016/j.aap.2014.01.001.

Hu, Y., Zhang, Q., G. -K, P., Yang, X., 2020. Automatic Identification of Ship Navigation Risk for Collision Accidents Using Uncertain Regression Model, *2020 Chinese Automation Congress (CAC), Shanghai, China*, 06-08 November 2020, pp. 3111-3116.

Khan, I.A., Hussain, F.K., 2022. Regression Analysis Using Machine Learning Approaches for Predicting Container Shipping Rates, in: Barolli, L., Hussain, F., Enokido, T. (Eds.), *Advanced Information Networking and Applications*. Springer International Publishing, Cham, pp. 269-280.

Lan, H., Ma, X., Qiao, W., Deng, W., 2023. Determining the critical risk factors for predicting the severity of ship collision accidents using a data-driven approach. Reliab. Eng. Syst. Saf. 230, 108934 https://doi.org/10.1016/j.ress.2022.108934.

Li, Q., Zhou, T., Lu, L., Chen, D., 2014. Identifying influential spreaders by weighted LeaderRank. Physica A-Statistical Mechanics and Its Applications 404, 47–55. https://doi.org/10.1016/j.physa.2014.02.041.

Li, H., Çelik, C., Bashir, M., Zou, L., Yang, Z., 2024. Incorporation of a global perspective into data-driven analysis of maritime collision accident risk. Reliability Engineering & System Safety 249, 110187. https://doi.org/10.1016/j.ress.2024.110187.

Li, H., Jia, P., Wang, X., Yang, Z., Wang, J., Kuang, H., 2023a. Ship carbon dioxide emission estimation in coastal domestic emission control areas using high spatial-temporal resolution data: A China case. Ocean Coast. Manag. 232, 106419 https://doi.org/10.1016/j.ocecoaman.2022.106419.

Li, H., Ren, X., Yang, Z., 2023b. Data-driven Bayesian network for risk analysis of global maritime accidents. Reliab. Eng. Syst. Saf. 230, 108938 https://doi.org/10.1016/j.ress.2022.108938.

Liu, Y., Tong, L.C., Zhu, X., Du, W., 2021b. Dynamic activity chain pattern estimation under mobility demand changes during COVID-19. Transp. Res. Part C: Emerg. Technol. 131, 103361 https://doi.org/10.1016/j.trc.2021.103361.

Liu, J., Wu, D., Wang, Z., Jin, X., Dong, F., Jiang, L., Cai, C., 2020. Automatic sleep staging algorithm based on random forest and hidden Markov model. Comput. Model. Eng. Sci. 123 (1) https://doi.org/10.32604/cmes.2020.08731.

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

Ministry of Transport of China, 2014. Regulation of water transportation accident statistics, Beijing, China.

Mohammadi, R., He, Q., Ghofrani, F., Pathak, A., Aref, A., 2019. Exploring the impact of foot-by-foot track geometry on the occurrence of rail defects. Transp. Res. Part C: Emerg. Technol. 102, 153–172. https://doi.org/10.1016/j.trc.2019.03.004.

Naik, A.K., Kuppili, V., 2022. Dynamic relevance and interdependent feature selection for continuous data. Expert Syst. Appl. 191, 116302 https://doi.org/10.1016/j.eswa.2021.116302.

Nguyen, S., Chen, P.-S.-L., Du, Y., Shi, W., 2019. A quantitative risk analysis model with integrated deliberative Delphi platform for container shipping operational risks. Transp. Res. Part E: Log. Transp. Rev. 129, 203–227. https://doi.org/10.1016/j.tre.2019.08.002.

Nguyen, S., Chen, P.-S.-L., Du, Y., Thai, V.V., 2021. An operational risk analysis model for container shipping systems considering uncertainty quantification. Reliab. Eng. Syst. Saf. 209, 107362 https://doi.org/10.1016/j.ress.2020.107362.

Nguyen, S., Shu-Ling Chen, P., Du, Y., 2022. Risk assessment of maritime container shipping blockchain-integrated systems: An analysis of multi-event scenarios. Transp. Res. Part E: Log. Transp. Rev. 163, 102764 https://doi.org/10.1016/j.tre.2022.102764.

Nguyen, S., Chen, P.S.-L., Du, Y., 2023. A methodological framework for quantitative risk analysis in container shipping operations. Maritime Business Review 8 (2), 139–155. https://doi.org/10.1108/MABR-11-2021-0088.

Nogueira, S., Brown, G., 2015. Measuring the Stability of Feature Selection with Applications to Ensemble Methods, In: Schwenker, F., Roli, F., Kittler, J. (Eds.), *MULTIPLE CLASSIFIER SYSTEMS (MCS 2015). Springer International Publishing, Cham*, pp. 135-146.

Puisa, R., Lin, L., Bolbot, V., Vassalos, D., 2018. Unravelling causal factors of maritime incidents and accidents. Saf. Sci. 110, 124–141. https://doi.org/10.1016/j.ssci.2018.08.001.

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

Ryu, U., Wang, J., Kim, T., Kwak, S., U, J., 2018. Construction of traffic state vector using mutual information for short-term traffic flow prediction. Transportation Research Part C: Emerging Technologies 96, 55–71. https://doi.org/10.1016/j.trc.2018.09.015.

Sotiralis, P., Ventikos, N.P., Hamann, R., Golyshev, P., Teixeira, A.P., 2016. Incorporation of human factors into ship collision risk models focusing on human centred design aspects. Reliab. Eng. Syst. Saf. 156, 210–227. https://doi.org/10.1016/j.ress.2016.08.007.

Sultana, S., Andersen, B.S., Haugen, S., 2019. Identifying safety indicators for safety performance measurement using a system engineering approach. Process Saf. Environ. Prot. 128, 107–120. https://doi.org/10.1016/j.psep.2019.05.047.

Tokyo MOU, 2023. ANNUAL REPORT ON PORT STATE CONTROL IN THE ASIA-PACIFIC REGION.

Uğurlu, Ö., Yıldız, S., Loughney, S., Wang, J., 2018. Modified human factor analysis and classification system for passenger vessel accidents (HFACS-PV). Ocean Eng. 161, 47–61. https://doi.org/10.1016/j.oceaneng.2018.04.086.

Wager, S., Athey, S., 2018. Estimation and inference of heterogeneous treatment effects using random forests. J. Am. Stat. Assoc. 113 (523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839.

Wang, X., Brownlee, A.E.I., Woodward, J.R., Weiszer, M., Mahfouf, M., Chen, J., 2021b. Aircraft taxi time prediction: Feature importance and their implications. Transp. Res. Part C: Emerg. Technol. 124, 102892 https://doi.org/10.1016/j.trc.2020.102892.

Wang, J., Wang, X., Feng, Y., Cao, Y., Guo, Z., Liu, Z., 2023a. Assessing the Connectivity Reliability of a Maritime Transport Network: A Case of Imported Crude Oil in China. Journal of Marine Science and Engineering 11 (8), 1597. https://doi.org/10.3390/jmse11081597.

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

Wang, H., Liu, Z., Wang, X., Huang, D., Cao, L., Wang, J., 2022. Analysis of the injury-severity outcomes of maritime accidents using a zero-inflated ordered probit model. Ocean Eng. 258, 111796 https://doi.org/10.1016/j.oceaneng.2022.111796.

Wang, X., Xia, G., Zhao, J., Wang, J., Yang, Z., Loughney, S., Fang, S., Zhang, S., Xing, Y., Liu, Z., 2023b. A novel method for the risk assessment of human evacuation from cruise ships in maritime transportation. Reliab. Eng. Syst. Saf. 230, 108887 https://doi.org/10.1016/j.ress.2022.108887.

Wang, L., Yang, Z., 2018. Bayesian network modelling and analysis of accident severity in waterborne transportation: A case study in China. Reliab. Eng. Syst. Saf. 180, 277–289. https://doi.org/10.1016/j.ress.2018.07.021.

Wang, H.Y., 2008. Combination approach of SMOTE and biased-SVM for imbalanced datasets, *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 01-08 June 2008*, pp. 228-231.

Weng, J., Ge, Y.E., Han, H., 2016. Evaluation of shipping accident casualties using zero-inflated negative binomial regression technique. J. Navig. 69 (2), 433–448. https://doi.org/10.1017/S0373463315000788.

Weng, J., Yang, D., 2015. Investigation of shipping accident injury severity and mortality. Accid. Anal. Prev. 76, 92–101. https://doi.org/10.1016/j.aap.2015.01.002.

Wong, T.T., Yeh, P.Y., 2020. Reliable accuracy estimates from k-fold cross validation. IEEE Trans. Knowl. Data Eng. 32 (8), 1586–1594. https://doi.org/10.1109/TKDE.2019.2912815.

Wu, S., Chen, X., Shi, C., Fu, J., Yan, Y., Wang, S., 2022. Ship detention prediction via feature selection scheme and support vector machine (SVM). Maritime Policy & Management 49 (1), 140–153. https://doi.org/10.1080/03088839.2021.1875141.

Xia, G., Wang, X., Feng, Y., Cao, Y., Dai, Z., Wang, H., Liu, Z., 2023. Navigational Risk of Inland Water Transportation: A Case Study in the Songhua River, China. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 9 (4), 04023042. https://doi.org/10.1061/AJRUA6.RUENG-1158.

Xin, X., Liu, K., Li, H., Yang, Z., 2024. Maritime traffic partitioning: An adaptive semi-supervised spectral regularization approach for leveraging multi-graph evolutionary traffic interactions. Transportation Research Part C: Emerging Technologies 164, 104670. https://doi.org/10.1016/j.trc.2024.104670.

Yan, K., Wang, Y., Jia, L., Wang, W., Liu, S., Geng, Y., 2023. A content-aware corpus-based model for analysis of marine accidents. Accid. Anal. Prev. 184 https://doi.org/10.1016/j.aap.2023.106991.

Yang, S., 2013. On feature selection for traffic congestion prediction. Transp. Res. Part C: Emerg. Technol. 26, 160–169. https://doi.org/10.1016/j.trc.2012.08.005.

Yang, Y., Shao, Z.P., Hu, Y., Mei, Q., Pan, J.C., Song, R.X., Wang, P., 2022. Geographical spatial analysis and risk prediction based on machine learning for maritime traffic accidents: A case study of Fujian sea area. Ocean Eng. 266, 113106 https://doi.org/10.1016/j.oceaneng.2022.113106.

Zheng, K., Chen, Y.B., Jiang, Y., Qiao, S.H., 2020. A SVM based ship collision risk assessment algorithm. Ocean Eng. 202, 107062 https://doi.org/10.1016/j.oceaneng.2020.107062.

Zhou, S., 2021. Data mining and analysis of the compatibility law of Traditional Chinese Medicines Based on FP-growth algorithm. J. Math. (Wuhan) 2021, 1045152. https://doi.org/10.1155/2021/1045152.

Zhou, J., Chen, S.-L., Shi, W., Nguyen, S., Kanrak, M., Ge, J., 2023. A belief rule-based bayesian network approach for assessing risks in the cruise supply chain: An empirical study in Shanghai, China. Ocean & Coastal Management 232, 106443. https://doi.org/10.1016/j.ocecoaman.2022.106443.

Zhou, Y., Li, X., Yuen, K.F., 2022. Holistic risk assessment of container shipping service based on Bayesian Network Modelling. Reliab. Eng. Syst. Saf. 220, 108305 https://doi.org/10.1016/j.ress.2021.108305.

Zhou, K., Xing, W., Wang, J., Li, H., Yang, Z., 2024. A data-driven risk model for maritime casualty analysis: A global perspective. Reliab. Eng. Syst. Saf. 244, 109925 https://doi.org/10.1016/j.ress.2023.109925.