Wang, Y, Li, P, Hong, C and Yang, Z

Causation analysis of ship collisions using a TM-FRAM model

Reliability Engineering and System Safety 260 (2025) 111035

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Causation analysis of ship collisions using a TM-FRAM model

Yuhong Wang [a], Pengchang Li [a], Cheng Hong [a], Zaili Yang [b,c,*]

[a] Faculty of Maritime and Transportation, Ningbo University, Ningbo, China
[b] Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool L3 3AF, UK
[c] School of Engineering, Liverpool John Moores University, L3 3AF, UK

ARTICLE INFO

Keywords:
Causation analysis
Ship collision
Text mining
Functional resonance analysis method
Maritime risk

ABSTRACT

Ship collisions pose a significant threat to life and property, presenting a major challenge in maritime safety. Current risk analysis methods have been criticized in terms of a lack of capacity of quantifying the risks of different features and a standardized database reflecting the multidimensional risks of human, mechanical, environmental, and management factors. Additionally, traditional analysis sometimes involves strong assumptions that 1) the established and widely used databases can capture all the essential features of ship collisions and 2) the modelling of ship collision process can be simplified by focusing the analysis on a single causal relationship at once. This paper aims to develop a new approach to enabling multi-dimensional analysis of the causation of ship collisions through the establishment of a new database for ship collisions by innovatively combining Text Mining (TM), Association Rule (AR), and Functional Resonance Analysis Method (FRAM). The new approach enables to construct a risk analysis network based on FRAM, and the model's practicality and effectiveness are validated through expert reviews and case studies. As a result, thirty-eight key risk factors have been successfully identified as per their influence to ship collision incidents, encompassing human error, mechanical failures, adverse environmental conditions, and operational issues. The findings not only offer a new perspective and methodology for ship collision risk analysis, but also enrich the theoretical framework of ship safety management, providing valuable guidance for ensuring ship navigation safety.

1. Introduction

Shipping plays a pivotal role in the security of global supply chains and economic growth. However, its safety is challenged by the intricate nature of maritime transport, still drawing widespread concern despite the technological advancement in recent years. The International Maritime Organization (IMO) data indicates that annually, over 500 ship collisions occur worldwide, resulting in more than 300 fatalities [1]. A study of Canadian waters from 2012 to 2022 revealed 2,456 ship accidents, with collisions accounting for 34.57% (849 accidents) [2]. In response to the frequent ship collisions, the IMO has formulated the International Regulations for Preventing Collisions at Sea (COLREGs) and International Convention for the Safety of Life at Sea (SOLAS). Although laws and regulations are in place to ensure navigational safety, they only provide the minimum requirements for safety. These requirements are passive in nature. To take a proactive approach, collision risk analysis should be conducted to prevent and anticipate ship collisions.

Collision risks have been analysed by scholars from various perspectives, including ship collision avoidance [2–6], simulation [4,6–8], AIS data-driven [4,7,9–11], and qualitative or quantitative risk assessment based on probabilistic models [1,12–14]. In terms of probabilistic models, Yu et al. [15] utilized Failure Mode and Effect Analysis (FMEA) and Bayesian Network (BN) to analyse collision risks between ships and offshore facilities, elucidating risk factor causality. Sokukcu et al. [16] focused on collision risks for ships at berth, developing a collision risk Fault Tree (FT) and BN model to identify primary causes. Zhang et al. [17] employed Human Factors Analysis and Classification System (HFACS) and FT models to quantify collision risks for icebreakers in ice-laden waters. Commonly, the studies in the field leverage methods such as FT [16,18,19], Event Tree (ET) [20,21], FMEA [15,22,23], BN [24–26], and fuzzy theory [16,27,28] for qualitative and quantitative risk assessments of ship collisions. However, these studies often rely on a single source of data which may result in less comprehensive in identifying risk factors. Meanwhile, when dealing with a large amount of accident data, there is a strong subjective element, making it difficult to accurately and objectively identify those key factors of collision risk. Moreover, previous studies are found to be associated with

* Corresponding author at: Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool L3 3AF, UK.
E-mail address: z.yang@ljmu.ac.uk (Z. Yang).

https://doi.org/10.1016/j.ress.2025.111035
Received 24 July 2024; Received in revised form 2 January 2025; Accepted 14 March 2025
Available online 17 March 2025

strong assumptions against which the complex ship collision process is simplified into a single or linear causal relationship, ignoring its complexity and variability [29,30]. It is worth noting that with the continuous advancement of information technology and data mining, TM provides new solutions to risk identification. It has been used in such areas as financial analysis [31,32], medical field [33], and mining safety [34,35] to extract valuable information from massive text data and to support accurate and comprehensive risk factor identification.

Given this, it is therefore crucial to develop a TM-based method enabling new exploration of the causes of ship collision risk and better understanding of the relationship between risk factors in the collision process for ship navigation safety. To achieve it, this reveals three challenges and gaps wanting new solutions to be found.

(G1): Due to the limited coverage of collision risk factors in a single database and the non-uniformity of accident data formats across different databases, it is challenging to encompass all factors involved in ship collision risk, including those from multiple dimensions, such as human errors, ship machinery and equipment factors, management factors, environmental factors.

(G2): Current risk causation analysis methods depend on subjective judgment or manual statistics, which makes it challenging to accurately and efficiently analyse extensive unstructured accident data and investigate the correlation between risk factors. Consequently, there is a need for a mature and efficient risk identification algorithm capable of handling extensive accident data, delving into the hidden risk factors, and further exploring the complex correlation relationships behind these risk factors.

(G3): Current research fails to comprehensively model the causation of ship collisions when employing a single or linear causal relationship analysis method. It is due to the fact that not all risks follow a linear pattern, and accidents of random causation can occur. Therefore, it is necessary to develop a critical risk causation analysis method that can take into account non-linear or non-single causality when dealing with complex ship collision processes.

To address the above three questions, this paper aims to develop a new TM-FRAM method to conduct deep mining and visualisation analysis of ship collision risk factors. It can provide new theoretical perspectives for ship navigation risk assessment and implications for making safe navigation policy. The paper's contributions are as follows:

C1: Techniques that can leverage big data, information and data mining technologies are continuously evolving, requiring the establishment of a specialized database before taking their advantages for ship collisions analysis. This paper employs maritime accident data sources from different regions/countries to develop a new standardized database involving all the key factors influencing ship collisions in the current literature. As data entries become richer and the volume grows, this database will progressively become more comprehensive and exhaustive.

C2: With the advancement in data mining technology, TM is now being used to effectively identify the risk factors of ship collision. Application of this method for causation analysis of ship collision can effectively address the deficiencies of the previous studies which rely on subjective judgment and manual statistics and hence introduce subjective bias and inaccuracy into the identification of risk factors. Additionally, it can uncover a direct link between the identified key risk factors, providing insights for in-depth analysis of the causes of ship collision risk.

C3: Ship collisions often present complex and constantly evolving processes. FRAM can be used to formulate a functional resonance network capable of dealing with such complication. Unlike traditional linear models, it captures the emergent behaviors and variability inherent in complex systems, making it an ideal tool for understanding the multifaceted nature of ship collision risks. Additionally, it has been successfully applied in multiple risk analysis domains. However, In the current literature, there are few studies combining TM and FRAM in risk causation analysis, and fewer in the application of TM-FRAM into the maritime sector. This innovative approach effectively addresses the limitations of traditional single or linear causality, overcoming the challenges of constructing risk causality in complex and variable collision accident processes. As a result, it reveals both methodological and applied research innovation.

The rest of this paper is organized as follows. Section 2 provides an overview of the current research progress and defines the-state-of-the-art in the field of risk assessment for ship collision accidents. Section 3 presents the data sources utilized in this paper and outlines the methodology employed, encompassing TM, AR, and FRAM. The findings and discussions are described in Section 4. Section 5 draws the conclusion.

2. Literature review

2.1. Ship collision data

Accident data underpins ship collision risk studies, and a well-structured and uniform database is pivotal for thorough risk factor analysis. Scholars have harnessed historical data to advance ship collision risk assessment significantly. Baksh et al. [36] assessed collision and grounding risks by using Arctic Portal data, while Afenyo et al. [24] drew from the ArcticData database to pinpoint key collision accident causes through BN. Uflaz et al. [19] investigated human error's role in collision risk by integrating evidential inference with FTA, using Marine Accident Investigation Branch (MAIB) database insights. Further, Liao et al. [37] crafted a scenario probability model for ship collisions from Global Integrated Shipping Information System (GISIS) data, and Ung et al. [38] scrutinized GISIS data from 2005 to 2015, identifying personnel fatigue and irregular operations as main risk contributors. Moreover, Zhang et al. [39] exploited data from the Changjiang Maritime Safety Administration (CMSA) to evaluate navigation safety, model collision risk, and deduce scenarios. However, many current studies only rely on a single database like MAIB [19] or GISIS [37,38], making it difficult to comprehensively cover the diversity of risk factors when assessing the risk of ship collisions by using a single database. Table 1 displays the literature about the GISIS, MAIB, NTSB, and TSB databases utilized in the past decade.

Variations in risk factors included in accident reports across these databases. For instance, MAIB furnishes detailed accident scene information, ship and navigation data, accident analysis, and recommendations [56]. In contrast, National Transportation Safety Board (NTSB) covers accidents in US waters, but its reports mainly focus on basic ship information, time and place, and causes of accidents, omitting details such as crew qualifications, training, and safety management [85]. Transportation Safety Board (TSB) provides comprehensive information including crew qualifications, communication records, training, and ship speed [85]. GISIS, on the other hand, compensates for the lack of geographical coverage of the former three by integrating ship accident data from various countries and providing extensive data support for global ship safety research [45,86].

Table 1
Data sources and relevant literature utilising the database in the last decade.

Database	Relevant literature			
GISIS	[40–52]	[53]	[54–58]	[59–61]
TSB	[62–65]			
MAIB	[66–72]	[64,73]		
NTSB	[74–84]		/	

2.2. Identification of risk factors

In exploring the factors contributing to ship collision risk, traditional methodologies frequently utilise tools such as FT [16,18,87], ET [20,21], FMEA [15,22,], and the HFACS [17]. Moreover, the literature analysis [16,88] and manual statistical evaluations [26,85] are prevalent approaches. These conventional methods predominantly focus on the identification of risk factors through manual statistical analysis, and largely involve personal experience and judgment [34]. As a result, they may have limited ability to handle large amounts of data and may introduce error and bias [89]. In contrast, TM provides a more objective analytical framework, facilitating the extraction of pertinent information from voluminous text data [90]. This approach enhances the precision and breadth of risk assessments, offering a more comprehensive and accurate evaluation of ship collision risks.

With the rise of big data and increased computing power, TM has seen significant improvements in accuracy and efficiency [33,31]. They are now widely applied in fields such as accounting [31], sentiment analysis [33], healthcare [33,91], and mining engineering [34]. These studies effectively showcase TM's advances in handling vast amounts of data, extracting accident information, and objectively identifying risk factors.

In the field of water transport safety, TM has yielded research findings. Xia et al. [90] employed TM to effectively pinpoint primary risk factors in inland vessel accidents and subsequently developed a comprehensive risk assessment index system for such accidents. Similarly, Dorsey et al. [92] applied TM to dissect the causal elements of collision accidents involving inland vessels. Furthermore, Wang et al. [89] leveraged TM to identify and extract risk factors pertinent to inland ship accidents. Their research underscores the method's capacity to significantly reduce uncertainty and subjectivity, thereby affirming the practicality and efficacy of TM in the realm of maritime risk evaluation [32,33]. Nonetheless, these studies are still limited by the use of singular nature of accident data sources, and the identification of risk factors is hence affected. Further, they have yet incorporated the TM with other approaches in a holistic way to deal with the non-linear relationship of risk factors and the associated complication.

2.3. Inter-relationships between risk factors

In the current literature of ship collision risk analysis, BN is identified as a primary approach to capturing the inter-relationships between the risk factors. In terms of BN inference, the causal relationship between parent and child nodes is reflected through conditional probabilities, and changes in the state of the parent node will directly affect the state of the child node [4,36,85]. Similarly, the FT starts from a specific accident and analyses the causes of its occurrence layer by layer until revealing the basic causes at the lowest level, and uses logic gates to connect these causes, to reveal the causal relationship of the event [19,21,87]. It is worth noting that despite the difference in expression between FT and inference BN, they both embody linear causality [16,18].

This approach, however, does not account for the non-linear nature of risk transfer in various hazard scenarios, where accidents can occur unexpectedly [93]. Recognizing the complexity and temporal evolution inherent in ship collisions, FRAM has been validated as a potent analytical tool for dissecting the multifaceted causes of such accidents [93,94]. FRAM's adaptability to complex and unpredictable scenarios has facilitated its extensive application across diverse sectors, including aerospace [95,96]. This is evidenced by the work of Salihoglu et al [97], who employed FRAM in qualitative evaluations of marine oil spills, thereby elucidating the nuanced risk factors involved. Similarly, Guo et al. [94] merged FRAM with a dynamic BN, formulating a model to assess the developmental trajectory of ship collision risks. Salihoglu et al. [97] employed FRAM for the Prestige Oil Spill analysis, facilitating the modelling of intricate interactions inherent in such events and enhancing comprehension of the system's operational dynamics via the functional modelling of maritime accidents. Furthermore, it aids in identifying prospective alterations in system functionalities and provides elucidations for potential causative factors of associated accidents. This underscores FRAM's capability to reveal more intricate causative interactions than those typically reported in accident reports. Ma et al. [98] advanced a hybrid methodology that synthesizes FRAM, fuzzy sets, and risk matrices to quantitatively evaluate the risks instigated by the Critical Links (CLs) throughout the continuum of the Human-Centered Maritime Transport System (HCMTS). Through the application of FRAM, a functional network delineating the standard operation of the HCMTS was established, enabling the identification of CLs linking upstream and downstream functionalities. The advantages of FRAM, as documented in various studies [94,95,99], encompass its efficacy in complex system assessments, even with limited data availability. It is particularly proficient in describing and analysing systems that involve human factors and inherent uncertainties. The method's emphasis on the analysis of entire system functionalities facilitates a deeper comprehension of interrelated actions and potential risks, making it a valuable tool in safety and risk management fields.

FRAM applications currently depend largely on expert knowledge or subjective judgment. For instance, Fu et al. [99] integrated FMEA with FRAM to assess the risk of nuclear-powered icebreakers in polar regions. This subjectively constructed FRAM method arguably lacks objective data support. Thus, FRAM should be combined with other methods to ensure its construction is adequately supported by data. This paper applies an AR algorithm to thoroughly analyse the connection between collision risk factors, laying a solid data foundation for constructing the FRAM network.

AR, as a machine learning-based algorithm, have been widely used in maritime accident data mining [25,100]. Lan et al. [101] analysed 1554 maritime accidents resulting in a total loss by utilising AR. They identified that the age of the ship was the primary factor causing casualties. This demonstrated the effectiveness of AR in discovering the relationships among risk factors. Jia et al. [88] used the same method to investigate information risk in maritime logistics services. They emphasized that AR were useful in exposing the interrelationships among risk factors. Yu et al. [102] examined the temporal and spatial correlation of risk factors leading to ship collisions by AR. They revealed the collision risk relationships in different regions. Özaydın et al. [103] combined AR and BN to construct a hybrid model of fishing vessel accidents. They identified the key risk factors. Wang et al. [100] used a time-based AR approach to study the domino effect in different types of maritime accidents. They revealed the hidden relationships between risk factors. These studies provide confirmation that AR is crucial in analysing the risks associated with maritime activities.

Although AR is advantageous in identifying frequent relationships in large datasets, it has also revealed some limitations. Firstly, AR mining can only reveal correlations among item sets, but it cannot capture causal relationships among risk factors [88]. This is where FRAM excels as it can more accurately describe the functional and causal pulse within a system [94,99]. Secondly, AR mining is often used for analysing static data and may not be suitable for studying time series data or time-varying analysis of system behavior [103,104]. However, FRAM concentrates on how different system functions interact with each other, and adapts to the dynamic changes in the internal and external environment of the system (Salihoglu, 2020). Lastly, AR mining is primarily based on analysing frequent item sets, which can make it challenging to capture the complex relationships among multidimensional risk factors [101,102]. In contrast, FRAM can comprehensively consider all of the multidimensional factors and reveal the intricate linkages among risk factors in a deeper way [105,106].

3. Data and methodology

3.1. Data sources

According to the ship collision database mentioned above, a total of 975 ship collision reports were reviewed from 2014 to 2023. After removing duplicate and invalid reports, 210 reports were selected for the new database by the three steps stated below. The criteria for valid reports include [46,107]: (1) Authority, issued by official maritime agencies and safety investigation bureaus to ensure data reliability; (2) Completeness, includes essential details such as time, location, and weather, along with a thorough account of the accident's cause, progression, and outcome. (3) Uniform Format, written in English, conforming to international maritime standards, and provided in a non-image text format to facilitate effective data extraction and analysis. These reports analysed various risk factors that contribute to ship collision accidents, such as ship information, navigational data, accident history, accident causes analysis, personnel qualifications and training, ship machinery and equipment, safety inspections, weather conditions, sea state, hydrography, and navigational environment. Table 2 provides a detailed comparison between different databases and the newly constructed database in this paper [45,46]. The data filtering steps are as follows:

Step 1: Data searching

The search for ship collision accidents was limited to accidents that occurred between January 1, 2014 and December 31, 2023. Preliminary results of 975 records were obtained from four databases GISIS, MAIB, NTSB, and TSB as public users.

Step 2: Data cleaning

After removing inconsistent, duplicate, incomplete, irrelevant, and inaccessible reports totaling 353, all reports that were not consistent with the objectives of the study were eliminated to ensure the accuracy and completeness of the data.

Step 3: Data screening

During the data screening stage, standardisation was carried out to address the differences in formatting and terminology between different databases, and the formatting and naming of all data fields were unified. At the same time, files in image format were filtered, totalling 210 after retaining only files in pdf, doc and docx formats. This step aimed to improve the consistency and analysability of the data, laying a solid foundation for subsequent data mining and analysis.

Table 2
Information covered by MAIB, NTSB, TSB, and new databases.

	MAIB	NTSB	TSB	GISIS	New database
Ship particulars	✓	✓	✓	✓	✓
Voyage particulars	✓		✓	✓	✓
Marine casualty information	✓	✓	✓	✓	✓
Communication	✓		✓		✓
Personnel certification and experience			✓		✓
Safety Staffing	✓	✓	✓		✓
Emergency response	✓				✓
Organizational and management factors			✓	✓	✓
Vessel certification and inspection	✓		✓		✓
Safety system				✓	✓
Accident analysis	✓	✓	✓	✓	✓
Ship machinery and equipment		✓	✓		✓
Weather and sea conditions	✓	✓	✓	✓	✓
Hydrology			✓		✓
Navigation environment		✓		✓	✓

3.2. Methodology

In this paper, TM techniques are employed to analyse 210 ship collision accident reports. This helps to identify the risk factors associated with ship collisions. Subsequently, utilizing the AR, a rule-based association mining technique [103,104], to conduct a more in-depth analysis of the identified risk factors. The aim is to investigate the relationships between various factors and comprehend their influence on accidents involving ship collisions. Using the identified risk factors and their interrelations, a FRAM network is established to qualitatively analyse the causal factors behind such accidents. The flow of the TM-FRAM model is depicted in Fig. 1.

Integrating TM with AR can effectively extract the underlying causes responsible for ship collisions from the unstructured and extensive accident reports. Further, the analysis of AR can unveil correlations among various risk factors, which serve as the foundation for constructing a FRAM network and evaluating the risk associated with ship collision accidents. This integrated approach helps identify the key risk factors contributing to ship collisions, providing a more comprehensive perspective and aiding decision-making processes [100,103].

3.2.1. Text mining

For accurate analysis of ship collision causation, it is necessary to clean the collected accident accidents before building a table to record cleaned data [33,34]. The dataset for ship collision analysis is created by isolating the segment that details the causes of accidents. R language and micro-wordcloud are used for TM and visualization [101]. Subsequent steps involve establishing tables for stopwords, professional terms, and merging words. Stopwords, including but not limited to intonational words, adverbs, prepositions, conjunctions, and punctuation marks, are identified as having minimal impact on the sentence's overall meaning [33,34]. Stopwords is based on lists provided by the "Natural Language Toolkit", "Harbin Institute of Technology", "spaCy", and "Baidu", further refined by incorporating insights from accident reports, the above thesauruses feature comprehensive and reliable stopwords, proven as efficient resources for text preprocessing through extensive testing and application [30,73]. Additionally, they are particularly suited for scientific and technology-driven text analysis projects like text mining and sentiment analysis [34,91]. Professional words aim to prevent the misclassification of specialized terminology during textual analysis. It is achieved by amalgamating terms from the domains of water transport engineering, safety engineering, shipping, meteorology, and more, augmented by details from accident reports. The inconsistency in accident report formats often results in varied expressions for identical causes of accidents, generating an abundance of synonymous terms. To improve the precision of the analysis, these terms are standardized, culminating in the creation of a thesaurus, as outlined in Table 3.

The calculation of Term Frequency-Inverse Document Frequency (TF-IDF) involves the integration of TF and IDF [30,108]. This index serves to gauge a term's relevance and scarcity within a dataset. TF measures the frequency of a term within a document, normalized by the total term count to counterbalance biases caused by document length. Accordingly, a term's significance increases with its frequency of occurrence. Conversely, IDF assesses the term's rarity across a collection of documents, with higher values indicating less common terms [29,109]. Furthermore, risk factors identified in accident reports are depicted through a word cloud visualization. The mathematical formulas for calculating TF, IDF, and TF-IDF are explicitly outlined in Eqs. (1)–(3):

$$\mathrm{TF} = \frac{n_{i,j}}{\sum_k n_{i,j}} \tag{1}$$

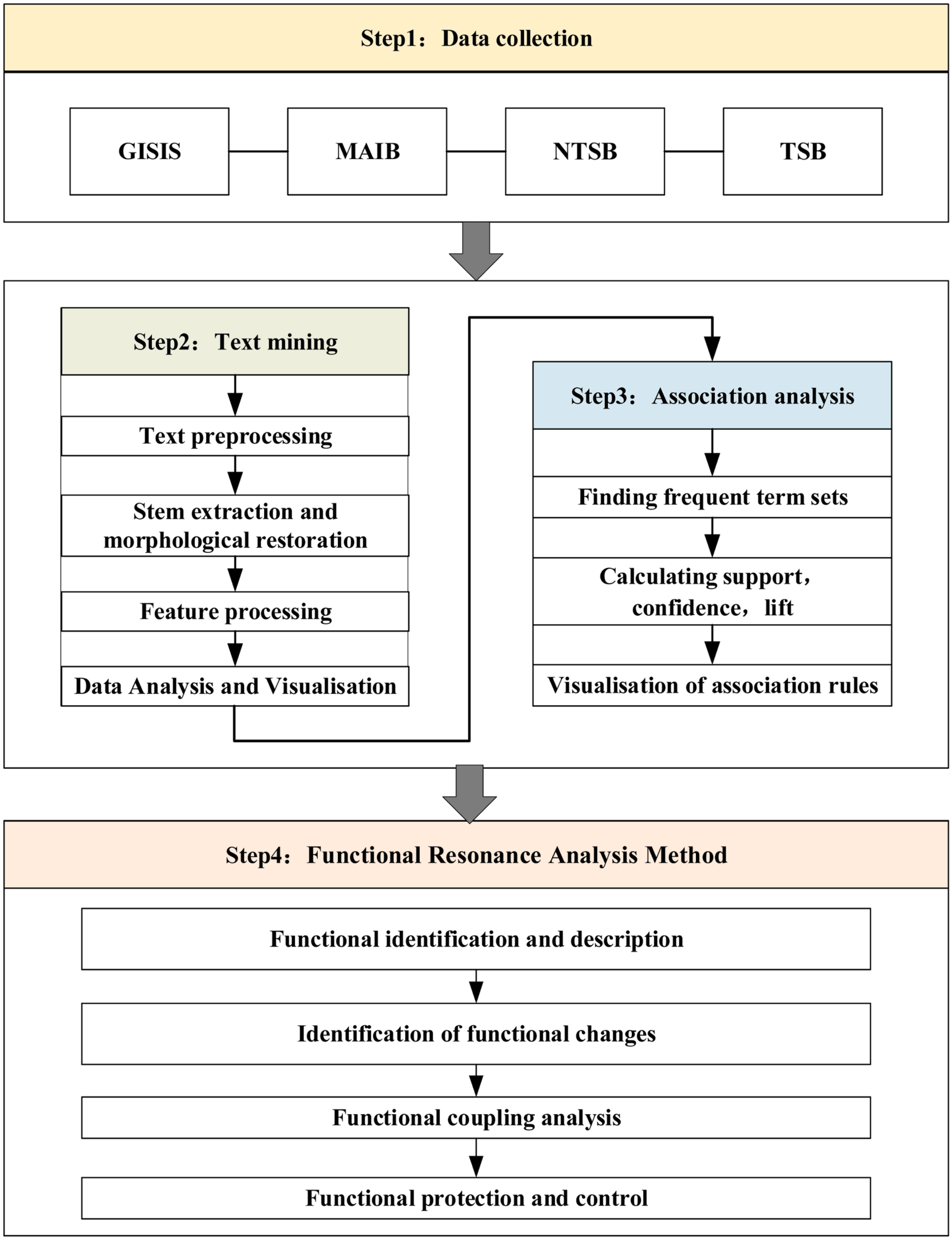


Fig. 1. The flowchart of the TM-FRAM model.

$$IDF = \log \frac{|D|}{|\{j : t_i \in d_j\}|} \tag{2}$$

$$TF - IDF = TF \times IDF \tag{3}$$

In ship collision risk identification, where $n_{i,j}$ denotes the frequency of collision risk factor t_i in the j ship collision incident report d_j, and $\sum_k n_{i,j}$ indicates the aggregate frequency of all risk factors in accident report d_j. Furthermore, $|D|$ represents the total number of all collision accident reports in the corpus, and $|\{j : t_i \in d_j\}|$ signifies the count of accident reports in the corpus that contain the risk factor t_i.

For example, if the term " Insufficient training " appears 2 times in a report containing 30,000 words, TF would be calculated as 2/30,000 = 0.00015. If " Insufficient training" appears in 36 out of 210 reports, IDF is calculated using the formula log(210/36) = 0.766. Therefore, the TF-IDF value for "inadequately trained" is the TF-IDF = 0.00015 * 0.766 = 0.015. 0.766 = 0.0113%.

3.2.2. *Deep mining of ship collisions utilising an AR approach*

Ship collision accidents pose a significant challenge to maritime

Table 3
Merging words for ship collision accidents mining.

Risk factors	Semantic description in accident reports
Negligence in lookout	Failure to maintain proper lookout, serious neglect in lookout, lookout negligence, failure to utilize all available means to maintain continuous and effective lookout.
Insufficient qualifications	Ship personnel lacks competency, insufficient ship handling skills and emergency response abilities, absence of appropriate ship handling qualifications, lack of professional knowledge and skills for fishing vessel operations.
Failure to take evasive action	Failed to take evasive action, did not take the most effective action to avoid collision, improper evasive measures, did not take correct evasive measures, failed to take any evasive action, not prompt in evasive measures.
Poor management	Inadequate safety management system, significant non-compliance in safety management system, failure to establish safety management procedures, deficiencies in wave prevention measures, ineffective implementation of safety management system, failure to comply with safety management system requirements.
Judgment errors	Failed to make correct judgments, did not accurately assess collision danger, failed to properly evaluate collision danger, insufficient judgment regarding collision danger, did not adequately assess the situation and collision danger.
……	……

transportation safety. This paper utilizes AR mining algorithms to conduct in-depth analysis of ship collision accident databases, aiming to discover potential AR among risk factors and provide a basis for constructing the FRAM for ship collision accidents. AR mining algorithms explore the dataset for frequent itemsets, revealing strong associations that highlight underlying correlations among risk factors. Support, confidence, and lift are crucial metrics in the process of AR mining [100,102]. The lift is a key indicator of the relevance and dependence between two itemsets within an AR. A lift value greater than 1 signifies a strong association, equal to 1 suggests no association, and less than 1 is considered to lack practical significance. Calculation formulas are provided [88,101]:

Consider the set $I = \{I_1, I_2, I_3, \dots, I_n\}$ representing all items in the database D, with t as a non-empty subset of the itemset I. The support for the AR $A \Rightarrow B$ is defined as the probability of A and B occurring together in a transaction, and is calculated as follows [100,103]:

$$\mathrm{support}(A \Rightarrow B) = P(A \cup B) = \frac{\mathrm{count}(A \cup B)}{\mathrm{count}(D)} \tag{4}$$

Where both A and B belong to I and $A \cap B = \varnothing$.

$$\mathrm{confidence}(A \Rightarrow B) = P(B|A) = \frac{\mathrm{count}(A \cup B)}{\mathrm{count}(A)} \tag{5}$$

$$\mathrm{lift}(A \Rightarrow B) = \frac{P(B|A)}{p(B)} = \frac{\mathrm{support}(A \cup B)}{\mathrm{support}(A) \cdot \mathrm{support}(B)} \tag{6}$$

A and B represent the set of collision risk factors respectively, and $\mathrm{count}(A \cup B)$ denotes the number of simultaneous occurrences of set A and B. The support reflects the probability of simultaneous occurrences of all risk factors in the AR. For instance, in a collision report, if "A" represents "insufficient training" and "B" denotes "improper emergency response," the confidence measures the probability of an improper emergency response occurring when insufficient training is already known to have happened. The term "lift" refers to the extent to which the presence of risk factor A increases the probability of risk factor B occurring, it helps to discover strong correlations between rules.

3.2.3. *FRAM*

FRAM is a method widely used for system safety analysis. It aims to help identify and understand the coupling relationships among diverse functional modules within a system, as well as the factors that may lead to unintended outcomes. This method analyses the characteristics and functional relationships of each functional module in the system to identify functional changes and couplings, and provides protective and control strategies [106,110,111], as shown in Fig. 2. An analysis of functional resonance networks, specifically focusing on ship collision accidents, can be executed following the outlined steps:

Step 1: Functional Identification and Description

Following a comprehensive review of the literature and analysis of accident reports, this paper delineates three primary phases in the ship navigation system: pre-navigation preparation, navigation, and collision avoidance response [94,112]. The pre-navigation phase encompasses crew training (F1), organization and management (F2), decision-making and planning (F3), prepare to set sail (F4). The navigation phase covers actual navigation (F5), observations and lookouts (F6), and discovering collision risk (F7). Finally, the collision avoidance phase includes steps such as emergency handling (F8), communication and exchange (F9), ship manoeuvring (F10) and emergency collision avoidance (F11). This paper extends the functional module of pre-navigation preparation based on the FRAM network delineation on ship collision accidents by Guo et al. [94]. An in-depth examination of accident reports and TM reveal that key risk factors for collisions include insufficient personnel training, defective safety management systems, crossing channels, and deviating from routes, which pose a serious threat to the navigation safety. Consequently, these findings were integrated into the FRAM network, augmenting its framework. Additionally, there are similarities between ship and aviation collisions, Carvalho et al. [95] identified pre-flight preparation as a critical module in the FRAM network for aviation collisions, addressing key factors such as route planning and approvals. Recognizing the distinct differences in navigational conditions, speeds, and emergency responses between maritime and aviation domains, this paper selectively incorporates applicable FRAM elements from aviation, like pre-flight preparation, while excluding those irrelevant to maritime settings in developing ship FRAM networks [95]. Aviation and ship collisions are massively different, as most of air crafts collisions occur in airport, and it is rare to have collisions in sky, while ships do collide each other at sea.

The 11 functional modules delineate the sequence of events leading to ship collisions and encompass risk factors including human error, organizational deficiencies, and environmental influences, which are derived from the mining results in Section 3.2.1. The FRAM for ship collisions is constructed by linking the inputs, outputs, prerequisites and others utilizing the AR findings from Section 3.2.2 along with comprehensive literature review. Amendments were also made with reference

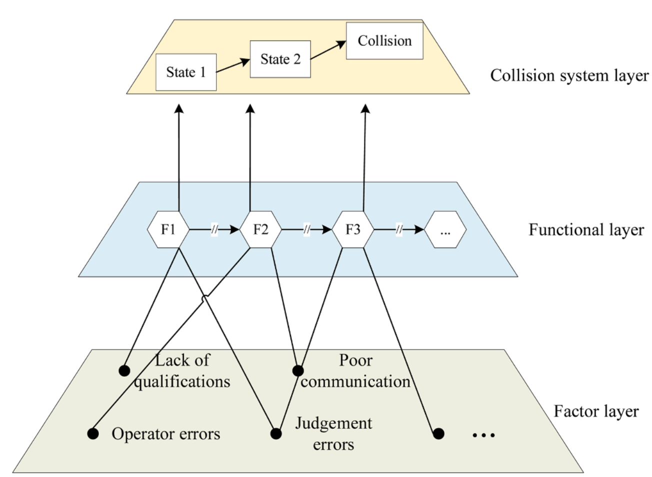


Fig. 2. Hierarchical model for collision coupling.

to literature, collision avoidance guide and accident reports to ensure the accuracy of the network. Appendix 1 offers an exhaustive account of each functional module along with its relevant indicators. Taking F2 as an example, which focuses on the organizational management level and covers risk factors such as insufficient safety inspections, improper maintenance of ship equipment, inadequate safety management system, insufficient manning, among others. The inputs to F2 are derived from the outputs of F1, and its outputs contribute directly to F3. When the output of F2 is the improvement of the safety management system, effective implementation of safety inspection, and sufficient manning, it will be the premise of F4 [94]. Additionally, the output of F2 serves as a resource for F6, F7, and F9. Crew lookout, collision risk discovery, and communication exchanges covered by these functional modules depend on the support of the ship's equipment, such as radar, AIS system, collision warning system, and various types of communication equipment [99]. Consequently, the output of F2 is highly linked to the resource of F6, F7, and F9 within the FRAM framework, and the constructed FRAM is shown in Fig. 3.

Step 2: Identifying changes in functions

Analysis of the 11 identified functional modules related to ship collision accidents offers insights into potential changes during actual operation. Examining these changes enhances understanding of the causes and risk factors for ship collisions, facilitating functional coupling analysis and the development of preventive and control strategies. FRAM consists of three forms of causing functional changes: internal changes, external changes, and functional coupling (outlined in step 3), and the detailed changes are shown in Table 4.

Table 4
Identification of functional changes in ship collisions.

Function	Function Type	Source of Variation	Relevant Factors
F1	Personnel	Internal Variation	Crew proficiency, safety awareness
F2	Organizational	External Variation	Management system, work arrangement, navigation planning
F4	Environment	External Variation	Severe weather, poor sea conditions
F5	Personnel, Environment	External Variation	Negligent lookout, low visibility
F6	Personnel, Vessel Equipment	External Variation	Judgment errors, vessel equipment failures
F8	Personnel	External Variation	Operation errors, improper emergency response
F9	Personnel, Vessel Equipment	Internal Variation	Communication breakdown, radar and sensing equipment failures
F10	Personnel	Internal Variation	Failure to perform emergency avoidance, communication breakdown

1) Internal changes: the effectiveness of F1 is influenced by the requirements for training, personnel turnover, and training quality, which in turn can alter crew proficiency, safety consciousness. F2's efficacy hinges on the appropriateness and adaptability of management decisions and operational plans, which should align with real-world sailing demands and the current state of the crew [94,106]. Additionally, F9's functionality is impacted by crew communication effectiveness and the reliability of communication devices, including radar systems. Finally, F10 operates under constraints related to ship navigation, communication efficacy, and the selection of risk

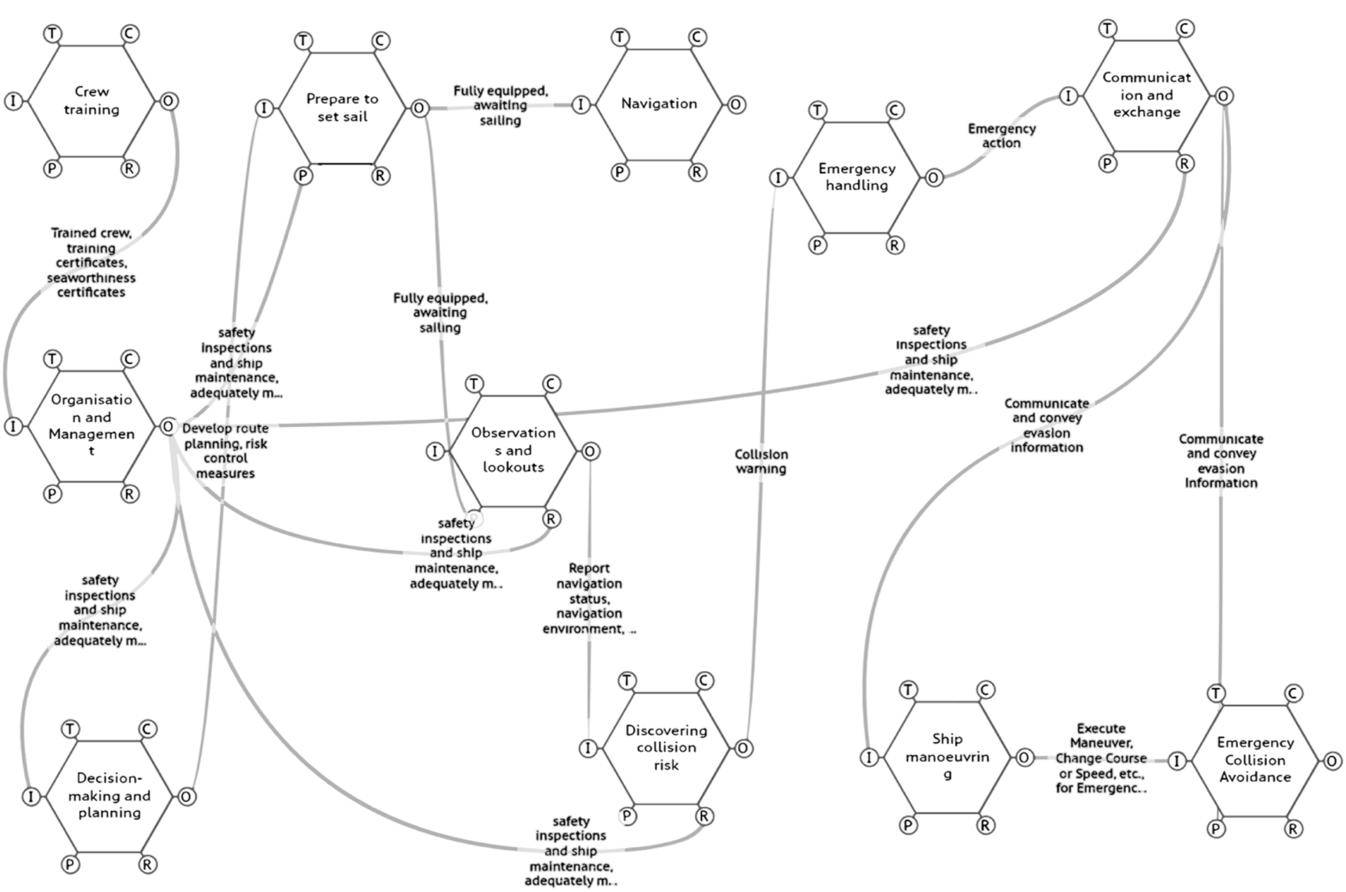


Fig. 3. Ship collision function resonance network.

mitigation strategies, necessitating vigilant monitoring of response adaptations in practice.

2) External changes: the F5 module, responsible for observation, may undertake variations in its capability, lookout periods, and equipment status during operations, affecting risk detection and reporting. F6 depends on various information sources, including radar and lookout reports, which may be compromised by equipment malfunctions or inaccurate information [106]. F4′s decision-making and planning are shaped by operational schedules and route planning, relying on external inputs such as nautical charts and weather data [94,106]. F8, focused on emergency handling, operates within the constraints of navigational directives and the vessel's immediate condition [93,111].

Step 3: Functional coupling

Functional coupling refers to the relationship between different functional modules that depend on and influence each other. In analyzing ship collisions using the FRAM, this paper reviews relevant literature and accident reports to identify the subsequent changes and effects of functional coupling.

Firstly, there is a strong coupling between the F1 and the F2 modules. As the need for new or additional crew members arises, the personnel and training module will make inputs on requirements of the crew training [111], while the organisation and management module will make outputs on management decisions and operational arrangements based on the navigation tasks and crew status [110]. Such functional interplay necessitates modifications in crew training and management practices, ultimately influencing crew quality and proficiency [112].

Secondly, F4 is coupled with the F6 module. Operational scheduling requirements serve as inputs for F4, whereas F6 generates collision warnings utilizing radar data and lookout reports [93,105]. Modifications in operational scheduling prompt adjustments in route planning and risk control strategies [106]. These adjustments consequently influence inputs and outputs of F6, impacting the efficacy of collision risk detection [99].

Finally, the interaction between F8 and F10 demonstrates a significant functional coupling. F8 processes inputs from navigation commands and ship status to execute movements, course adjustments, or speed changes [94]. Subsequently, F10 initiates collision avoidance maneuvers, relying on collision warnings and the actions taken by F8 [106]. Alterations in this coupling can necessitate modifications in ship manoeuvring strategies, thereby influencing the operational responses of F10 for collision avoidance [96,106].

Step 4: Functional Protection and Control

Conducting a functional protection and control analysis requires a detailed examination of the functional modules and their interconnections within the FRAM. With insights into the functional dependencies and changes outlined in Step 3, strategies and measures can be developed to mitigate the risks associated with each functional module [94,106].

Firstly, according to the coupling analysis of Step 3 between the F1 module and the F2 module, specific measures should be taken to enhance the crew's skills, safety awareness, and risk management capabilities. This includes implementing a comprehensive training program, strengthening crew competency assessments, conducting an in-depth study of the International Regulations for Preventing Collisions at Sea, continuously assessing crew performance, and establishing effective communication channels to convey safety information [106,111]. Secondly, regarding the coupling between the F2 module and the modules F1, F3, F4, F6, F7, and F9, the following measures should be implemented: 1) Formulate a comprehensive training plan to ensure crew members undergo extensive training in safety operations, emergency responses, and regulatory compliance, regularly updating their skills and qualifications according to the latest international standards [99]. 2) Optimize the voyage planning process by establishing an efficient information flow mechanism, ensuring all decision-making levels operate based on the most current and accurate data [94]. 3) Thoroughly examine and optimize pre-sailing preparations, taking into account all potential risks, such as mechanical failures and meteorological conditions, with timely notifications to the relevant crew [93,106]. 4) Enhance emergency response capabilities by regularly rehearsing collision risk responses, ensuring crews can swiftly and effectively execute emergency collision avoidance measures, and introducing decision-support systems like collision warning systems. 5) Strengthen communication efficiency by optimizing the information transfer and feedback mechanisms to ensure the rapid and accurate exchange of critical information [113].

In the F4 module, coupling analysis reveals that it is coupled with the F2, F3, F5, and F6 modules. Consequently, when the F4 module experiences functional resonance, it may induce resonance in these modules. Therefore, the following measures are recommended: 1) ensure organizational support for a comprehensive pre-sailing inspection, covering hull integrity, mechanical performance, navigational equipment, and safety equipment [113,96]. 2) Maintain an unimpeded flow of information before and during the voyage, facilitating communication among crew members and with the ship's management team [110,111]. For the Navigation Module (F5), its coupling with F2, F4, and F7 modules means that functional resonance in F5 could also resonate with these related modules. The suggested actions include: 1) complete all necessary preparations before departure and establish a multi-stage final confirmation process to mitigate safety risks [93]. 2) Ensure that management directly supervises key voyage decisions and conducts regular safety reviews. Regularly update and optimize navigational procedures and policies in line with international standards [105]. 3) Equip ships with modern collision warning systems and provide effective training for crews [99]. For the F10 module, due to its coupling with the F9 and F11 modules, the following measures are essential: 1) regularly train crew members in communication and conflict resolution skills, enhancing efficiency under stress or in emergencies. 2) Conduct regular emergency avoidance drills to ensure crew proficiency in critical maneuvers. 3) Utilize data recording and analysis tools to monitor ship maneuvering performance and crew response efficiency. Regularly review safety policies and practices to reduce collision risks effectively [94].

3.2.4. *Methodology validation*

To verify that the identified risk factors for ship collisions are not the result of random occurrence, a multi-stage random sampling of accident reports at intervals of 5%, 10%, 15%, and 20% is performed. The purpose of this multi-stage sampling is to evaluate the stability and consistency of the TM method across varying extents of data coverage. The method's reliability in pinpointing critical risk factors across datasets of different sizes was gauged by comparing the outcomes at various sampling rates. Each sampling acts as an independent assessment on a distinct dataset, thereby strengthening the validation process and enhancing understanding of how the analytical method responds to data variations. Additionally, risk factors are determined manually on an individual case basis to maintain objectivity and precision. By drawing comparisons with prior studies, pivotal risk factors in ship collisions are pinpointed, laying a robust groundwork for future risk management, prevention, and control initiatives. Should the key factors prove insufficiently representative, refining the data preprocessing steps and augmenting risk identification accuracy through consultations with domain experts may be required.

Variations in outcomes at distinct sampling rates offer a crucial foundation for evaluating the model's robustness. A significant discrepancy in results may suggest heightened sensitivity of the model to dataset size or quality, necessitating appropriate adjustments. Through in-depth examination of these variations, the most stable and reliable risk factors, as well as those influenced by data volume or quality, can be pinpointed. These insights enable the refinement of models and analytical approaches, incorporating robust statistical methods to improve result reliability and validity.

In the development of the FRAM for ship collisions, this paper infused AR results, accident reports and literature to refine and amplify its efficacy, resulting in marked enhancements in both practicality and adaptability. To assess the actual impact of FRAM, representative accident cases have been carefully selected for in-depth analysis. This process confirmed the efficiency and reliability of the FRAM in real-world applications.

4. Results and discussions

4.1. Preliminary results by applying TM

In the analysis outlined in Section 3.1, a total of 96 keywords were initially extracted. However, this set included terms not directly relevant to the risk factor analysis of ship collision accidents. A manual review was conducted to filter out unrelated terminology, such as "classification society," "container," and "certificate of registry." Following this meticulous review, 38 keywords were identified as significant risk factors for ship collision accidents. The findings, including these 38 risk factor keywords ranked by their TF-IDF frequency, are displayed in Table 5, while Fig. 4 illustrates the occurrence frequency of these keywords in the accident reports. An examination of Table 5 and Fig. 4 reveals that, although the ranking of risk factors by TF-IDF generally aligns with their frequency of occurrence, discrepancies are observed with terms like "Helm failure," "Mismanagement," as well as "Bad weather," "Obstacles," and "Poor boating skills."

Firstly, risk factors with higher weights warrant special attention. Notably, issues like "Negligent lookout", "Inadequate safety checks", "Low visibility", and "Inadequate management system" have received high frequency. These elements are prevalent in navigation safety concerns and significantly influence navigation safety. Secondly, it is important not to overlook risk factors with lower weights. While factors such as "Failure of ship equipment" and "Failure to maintain safe distance" might receive lower scores, the possible impact of these elements on navigation safety should still be considered.

TM quantitatively evaluates maritime risk factors utilizing both frequency analysis and TF-IDF, revealing discrepancies between the two methodologies. According to Fig. 4 and Table 5, risk factors such as "Negligent lookout", "Inadequate safety inspections", "Low visibility", and "Lack of management system" have been identified as highly significant. This highlights their prevalence in real-world scenarios and their potential to significantly compromise navigation safety. To mitigate these risks, it is recommended that safety oversight and training be enhanced by the appropriate authorities. On the other hand, the ranking of certain risk factors, like "Rudder failure" and "Improper management," varies when analyzed through frequency as opposed to TF-IDF. This variation suggests that the significance of these factors may differ across different contexts and accident types. Frequency analysis captures the total occurrences of risk factors across all incidents, whereas TF-IDF assesses their relative significance within specific accident categories, suggesting that TF-IDF may offer a more context-specific understanding of risk factor importance.

Fig. 5 presents a word cloud generated from the TF-IDF analysis of risk factors, where the size of the font corresponds to the frequency of occurrence within the dataset. Words situated closer to the center of the cloud denote a more substantial causal association with collision risks. The examination of both Table 5 and Fig. 5 demonstrates that certain risk factors, including "Negligent lookout," "Insufficient safety inspections," "Low visibility," and "Imperfect management system," occur frequently and exert a notable influence on navigation safety. Chief among these, "Negligent lookout" is mainly ascribed to the crew's failure to maintain continuous vigilance over the vessel's surroundings, thereby impeding their capacity to promptly identify potential hazards. "Insufficient safety inspections" signify lapses in the enforcement of comprehensive safety protocols, thereby permitting potential threats to remain undetected. The risk factor of "Low visibility," often resulting from adverse weather conditions such as fog or heavy rain, is a critical element in maritime accidents by impeding navigational visibility. Moreover, "Imperfect management systems" are associated with insufficiencies in operational oversight and safety practices, typically stemming from the inadequate implementation of safety measures by shipping companies or regulatory bodies. Additional significant factors include poor communication, which is frequently emphasized due to its role in engendering misunderstandings and navigational errors among crew members. Fatigue, which detrimentally impacts crew alertness and decision-making, is also underscored, indicating the necessity for enhanced management of work schedules and rest periods. Furthermore, insufficient manning is identified as a substantial cause of operational overload and error, highlighting the need for adequate crew staffing on vessels. Collectively, these factors, frequently accentuated in the word cloud, delineate critical areas where interventions can be made to mitigate the risks associated with maritime accidents.

4.2. Mining results utilising Apriori algorithm

Utilizing the AR mining based on the Apriori algorithm, as detailed in

Table 5
Risk factors in English accident reports.

Risk factor	Frequency	TF-IDF	Risk factor	Frequencyy	TF-IDF
Low visibility	597	0.1404%	Inadequate security precautions	53	0.0172%
Poor communication	652	0.1331%	Failure to use light or sound signals	51	0.0170%
Safety management system	513	0.1251%	Aids to navigation system failure	44	0.0145%
Fatigue	372	0.0847%	Poor sailing plans	39	0.0141%
Improper emergency response	354	0.0806%	Heavy traffic on route	41	0.0137%
Negligence in lookout	236	0.0717%	Adverse weather	37	0.0126%
Safety awareness	206	0.0679%	Information exchange	38	0.0119%
Ship equipment failure	204	0.0507%	Insufficient training	36	0.0113%
Failure to adequately consider risks	159	0.0410%	Lack of experience	29	0.0094%
Insufficient manning	157	0.0406%	Inspection and maintenance	21	0.0073%
Break the rule	146	0.0381%	Insufficient qualifications	20	0.0069%
Not using safe speed	106	0.0334%	Strong tidal streams	15	0.0056%
Safety culture	109	0.0298%	Failure to wear a pfd	13	0.0046%
Drug and alcohol	100	0.0279%	Crew competence	12	0.0043%
Situational awareness	92	0.0266%	Poor concentration	12	0.0043%
Improper operation	90	0.0252%	Sleep debt	10	0.0038%
Conversation	84	0.0237%	Stability and structure	5	0.0020%
Omission	69	0.0221%	Manoeuvrability of the vessel	4	0.0016%
Failure to take emergency evasion	55	0.0187%	Wind	5	0.0013%

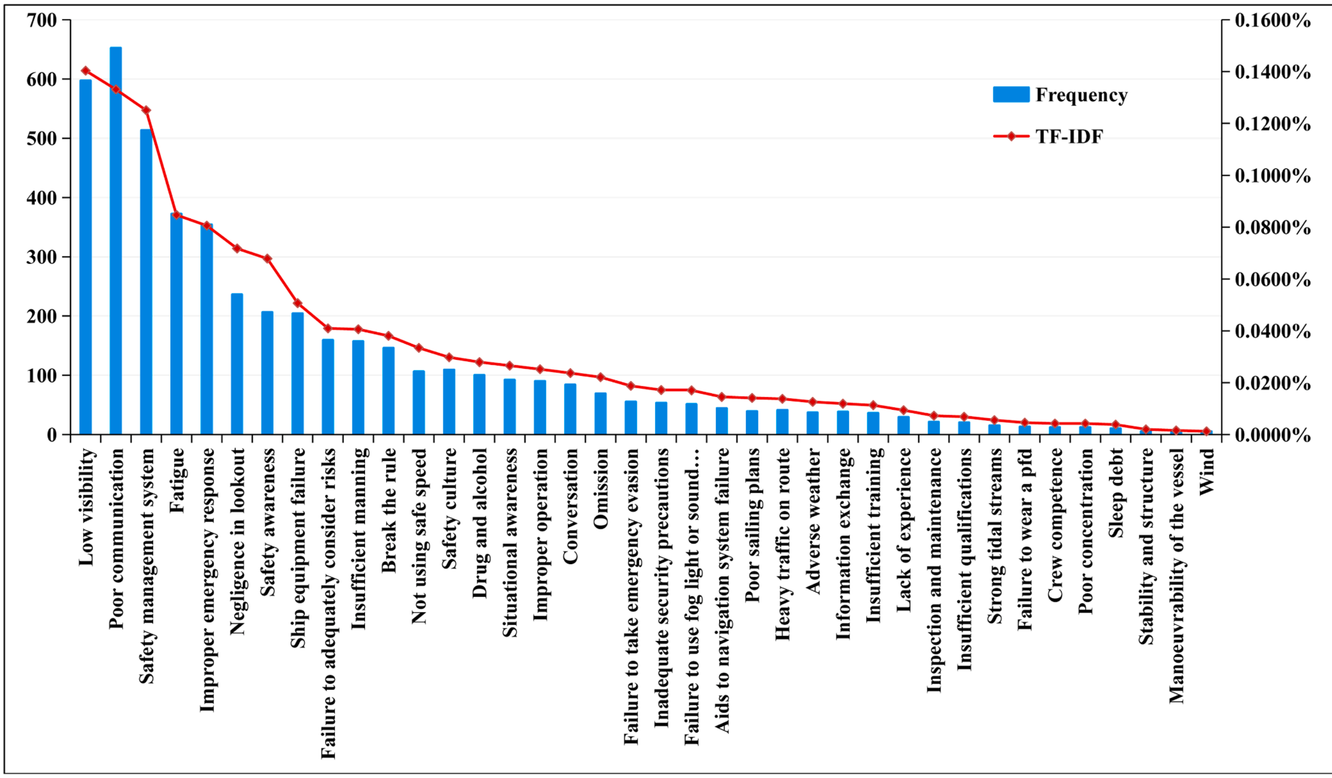


Fig. 4. Risk factors in the mining of accident reports.

Fig. 5. Wordcloud of risk factors using TM.

Section 3.2.2, this paper extensively analyses the relationships among 38 risk factors. This analysis serves as the foundational groundwork for the development of the FRAM. Rstudio 4.0, equipped with the arules and arulesViz packages from the R programming language, was employed for conducting the AR mining. Within the Apriori algorithm, support and confidence are critical conditional metrics, with their threshold settings significantly influencing the outcomes of the AR mining process. The existing literature on AR mining does not prescribe a standard for determining the minimum thresholds for support and confidence [100, 103]. To balance between avoiding the generation of superfluous rules and overlooking crucial ones, this paper adopts the average TF-IDF value of the risk factors as a reference point for setting the minimum support threshold [102,104]. Following this approach, adjustments to different levels of confidence were made, leading to the determination of the final thresholds for minimum support and confidence after numerous iterations.

For the determination of thresholds, it is established that the mean value of the frequency of occurrence of each attribute feature serves as the reference for the minimum support threshold [88,103,104]. Based on this reference, various combinations of minimum support and confidence are determined through many trials. After repeated trials and integration of prior studies [100–102], the minimum support and confidence thresholds are ultimately established. After extensive trials, thresholds of 0.1 for minimum support and 0.7 for minimum confidence are established. This process yielded a total of 705 initial rules. Following Eq. (6), a higher lift indicates a stronger AR, whereas a lower value suggests a weaker association. Following the analysis, it was determined that the average lift value was 1.22. Subsequent experiments established the lift threshold at 1.22. The range of lift values observed after screening fell between 1.22 and 2.53. Consequently, rules that were duplicate, invalid, or had a lift less than 1.22 are excluded, resulting in 92 robust ARs being identified. Tables 5 and 6 display selected ARs along with their support, confidence, and lift. Additionally,

Table 6
AR between risk factors by using the apriori algorithm (only a part).

Rules	Support	Confidence	Lift
{Insufficient training} => {Negligence in lookout}	0.1842	0.7778	1.2315
{Improper operation} => {Improper emergency response}	0.2105	0.8000	2.1714
{Improper operation} => {Failure to adequately consider risks}	0.1842	0.7000	1.6625
{Failure to use fog light or sound signals} => {Not using safe speed}	0.2895	0.9167	2.0490
{Failure to use fog light or sound signals} => {Break the rule}	0.2368	0.7500	1.7813
{Poor communication} => {Failure to take emergency evasion}	0.3158	0.8000	1.5200
{Break the rule,Failure to use fog light or sound signals} => {Failure to take emergency evasion}	0.2105	0.8889	1.6889
{Break the rule,Not using safe speed} => {Failure to use fog light or sound signals}	0.2105	0.8000	2.5333
{Not using safe speed,Judgment error} => {Failure to take emergency evasion}	0.1842	0.8750	1.6625
……	……	……	……

Fig. 7 illustrates a scatter plot visualizing the relationship among support, confidence, and lift values.

The lift is a crucial metric for evaluating the effectiveness of association rules. Analysis of Table 6 and Fig. 6 indicates that in scenarios involving "Failure to use safe speed" and "Violation of collision avoidance rules", The situation of "Not using fog lights or audible signals" exhibits a significantly higher lift (support 0.2105, confidence 0.8, lift 2.533). This indicates that under foggy or low visibility conditions, the absence of sound or light signals, compounded by unsafe speeds and disregard for collision avoidance regulations, elevates the risk of ship collisions. Conversely, the link between "insufficient training" and "negligent lookout" shows a less pronounced lift (support 0.1842, confidence 0.7778, lift 1.2315), suggesting that inadequate training results in a failure to maintain vigilant observation, thereby increasing collision risks. Consequently, for the enhancement of maritime safety, it is imperative for shipping companies to improve crew training and underscore the adherence to international collision avoidance protocols. It is essential to instruct crews on maintaining safe speeds and effectively utilizing navigational signals, including auditory and visual cues.

Analysis of mining data reveals a high confidence level for the co-occurrence of a lack of safety awareness with violations of collision avoidance rules, and the failure to execute emergency avoidance maneuvers (support 0.2632, confidence 0.9091, lift 1.7273). This indicates that scenarios involving crew members' deficiency in safety awareness and non-compliance with navigation rules are likely to result in a delayed or absent emergency response. Conversely, the confidence for the combined occurrence of inadequate emergency response, negligent lookout, and insufficient safety awareness is comparatively lower (support 0.1842, confidence 0.7, lift 1.33), suggesting a lesser likelihood of these events happening together. These findings offer significant insights into improving crew safety awareness and adherence to navigation rules.

In addition, some of the AR involve more risk factors, such as "Poor emergency response" caused by the combination of five risk factors, "Helm failure, bad weather, unsuitable personnel, faulty on-board equipment, and high and low tides". This indicates that the occurrence of navigational risks is often the result of a combination of risk factors rather than a single factor.

4.3. FRAM case analysis results

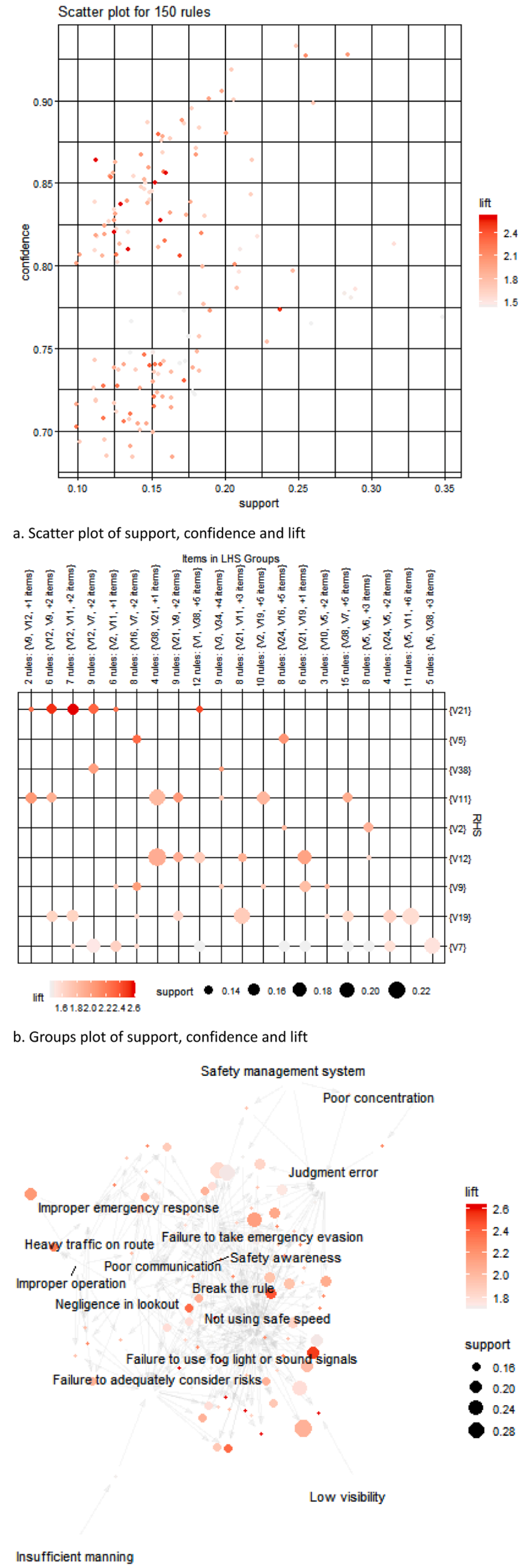
In this paper, FRAM model was validated using a ship collision accident data from Irish waters in 2015 as a case study. The incident, which involved two fishing vessels, resulted in the severe damage and sinking of one vessel, while the other sustained moderate damage. It was examined in depth utilizing the FRAM network as described in Section 4.3.

Step 1:Functional Identification and Functional Description

Employing the FRAM network outlined in Section 3.2.3, the analysis condensed the ship collision process into eleven critical functional modules. Through comprehensive examination of accident reports, essential modules associated with collisions were pinpointed, including: F1 personnel training, F2 organizational management, F5 navigation, F6 observation and lookout, F7 detection of collision risks, F8 emergency response, F9 communication and exchange, F10 ship manoeuvring, and F11 emergency avoidance.

Step 2: Identify changes in functions

The analysis of the accident report indicates that the primary cause of the collision was crew fatigue, which impaired the ability for prompt risk detection and situational awareness. Additionally, the crew's lack of adequate training, deficient seamanship skills, and unfamiliarity with the vessel's AIS and navigation tools hindered their capacity to



a. Scatter plot of support, confidence and lift

b. Groups plot of support, confidence and lift

c. Graph for 150 rules

Fig. 6. Visualisation results for AR.

maneuver effectively and prevent the collision. Moreover, the absence of collision risk alarms, mandatory under international maritime regulations, was a critical oversight. Detailed variations in each function during the accident are elaborated in Table 7.

Step 3: Functional coupling

Utilizing the FRAM network from Section 3.2.3 and the identified functional changes, Fig. 7 illustrates the functional coupling network for the collision. Key functions involved are shown as blue modules, with purple lines (1-9) indicating failure points in the links between them. Line 1 failures denote inadequate crew training affecting operational procedures and safety knowledge, impacting organizational management, like insufficient route planning. Lines 2-6 failures highlight shortcomings in safety inspections, ship maintenance, and safety management, jeopardizing navigation. These failures could amplify damage and impede correct crew response in a collision, aligning with the accident's primary risk factors. The failure associated with line 7 points to insufficient lookout measures, which delay timely maneuvers to avoid imminent collision risks. Line 8′s failure highlights the absence of essential alarm equipment, a critical component for enabling prompt crew action in the face of collision risks. Furthermore, the breakdown in line 9, attributable to poor seamanship skills and inadequate crew training, obstructs effective emergency maneuvers to circumvent collision.

Step 4: Functional Protection and Control

1) For Line 1 failure (insufficient crew training)

Shipping companies should enhance safety and operational training for crew by developing regular and systematic training programs that ensure all crew receive comprehensive training, including emergency response and accident prevention. Maritime authorities should intensify the oversight of training quality provided by shipping companies to ensure that all crew training meets international maritime standards.

2) In response to the failure of Lines 2-6 (safety inspection, ship maintenance, and inadequacy of the safety management system)

Maritime Administration should regularly conduct comprehensive safety reviews and ship inspections to ensure that all equipment and operations comply with the latest safety standards. Shipping companies should implement an internal safety audit program, carry out regular self-inspections, and respond swiftly to identified issues to prevent potential safety hazards from becoming actual risks.

3) In response to Line 7 failure (negligence in lookout)

Shipping companies should enhance lookout training for crew members, especially in complex or busy waters. Advanced surveillance technologies should be adopted to support manual lookouts, such as the installation of enhanced radar and Automatic Identification Systems (AIS). Maritime authorities should establish stricter lookout requirements and conduct regular checks to ensure compliance.

4) In response to Line 8 failure (failure to install necessary alarm equipment)

Shipping companies should ensure all ships are equipped with alarm systems that meet the latest standards, such as collision warning systems and emergency response systems, and maintain and upgrade these systems regularly. Maritime authorities should regularly inspect ships for the proper installation of alarm systems and take punitive actions against those with inadequately installed or maintained systems.

5) In response to Line 9 failures (inadequate seafaring skills and lack of training for crew)

Shipping companies should increase training in advanced seamanship skills, such as advanced navigation and emergency management, and provide ongoing support for career development. Insurance companies could incentivize shipping companies to enhance their training quality by offering insurance discounts to those with exemplary training records and low accident rates.

Table 7
Functional changes identified for collisions in Northern Ireland waters in 2015.

Functional module	Function type	Variation	Relevant factors
F1	Personnel	Internal	Insufficient crew training and lack of familiarity with the ship's equipment.
F2	Organizational	External	Failure to maintain the ship on a regular basis
F5	Personnel	Internal	Failure to follow COLREGs rules
F6	Personnel, environment	External	Failure to maintain proper lookout due to fatigue and other factors
F7	Personnel, equipment	External	Failure of a ship to install required navigation alarm equipment.
F10	Personnel	External	Poor sailing skills and emergency response

4.4. Comparative analysis for model and result validation

4.4.1. Case study validation of FRAM

This paper developed the FRAM through AR and used the mined rules to improve the accuracy and consistency of the analysis. However, there is still some subjectivity and uncertainty present in the system analysis and modelling process. To evaluate the constructed FRAM (as shown in Fig. 3), it was compared with litertures, accident reports and ship avoidance guidelines issued by IMO. The evaluation results indicate that the FRAM is effective in the pilotage collision accident scenario of the case study and its functional modules and coupling structure are consistent with experience. Furthermore, the case study detailed in Section 4.3 demonstrates the utility of FRAM in identifying primary risk factors for ship collisions through the FRAM network analysis. This evaluation is based on the historical data and detailed account of a ship collision from an accident report, ensuring consistency with the report's findings and recommendations.

4.4.2. Validation of risk factors

To verify the accuracy of the mining results, a manual sampling method was used to sample and validate 5%, 10%, 15%, and 20% of the accident reports. The results confirmed that the identified risk factors are consistent with the mining results, as shown in Fig. 8. In addition, to enhance the identification accuracy, this paper conducted an exhaustive comparison and analysis with other relevant literature. The studies reveal that there is consistency in identifying risk factors [62,87]. However, it also found significant differences upon closer analysis. Notably, the use of alcohol and drugs was scarcely mentioned in previous studies and was uncommon but infrequent in accident reports. This variation is mainly due to differences in culture and legal regulations across countries. For instance, Chinese laws explicitly prohibit the use of alcohol and drugs during navigation, which is a contributing factor to the difference.

The results of the sample study showed consistent trends for 5%, 10%, 15%, and 20% of the samples. None of these samples showed "Stability and structure", indicating that most ships are stable before they set sail. The most common issue found was the "Negligence in lookout", which was attributed to the traditional reliance on human lookout, judgment, and maneuvering in ship navigation. Crew members' psycho-physiological and subjective interferences were identified as the main causes of such negligence. Secondly, "Safety awareness" and "Failure to take emergency evasion" are also noteworthy. Long voyages and extended crew operations can lead to a lapse in vigilance, coupled with insufficient safety education and training, resulting in ineffective emergency response when confronted with collision risks. This aligns with Sotiralis's [21] research highlighting the increased likelihood of collisions caused by human errors such as negligent lookout and

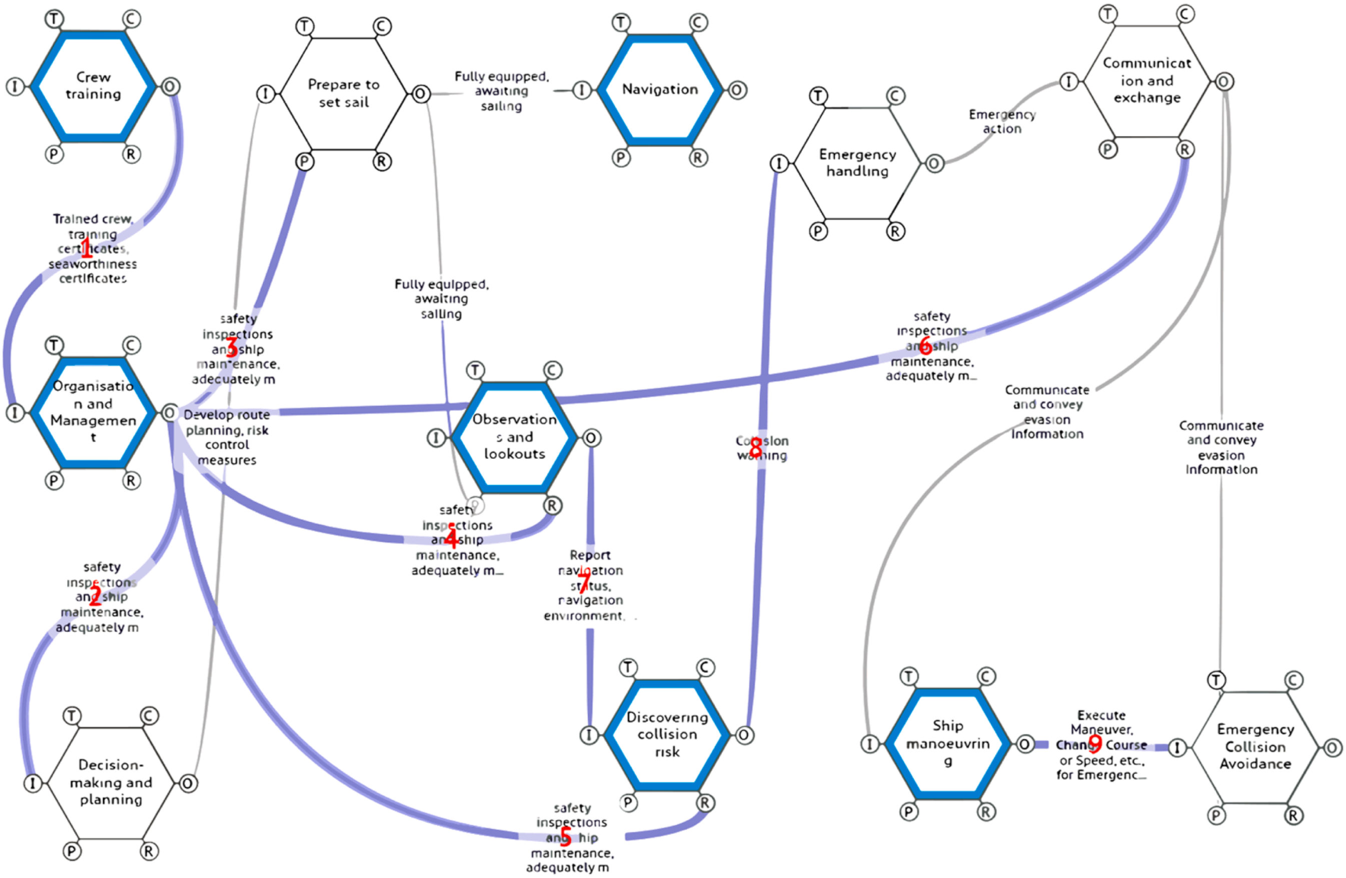


Fig. 7. Functional module failures for collisions in Northern Ireland waters in 2015.

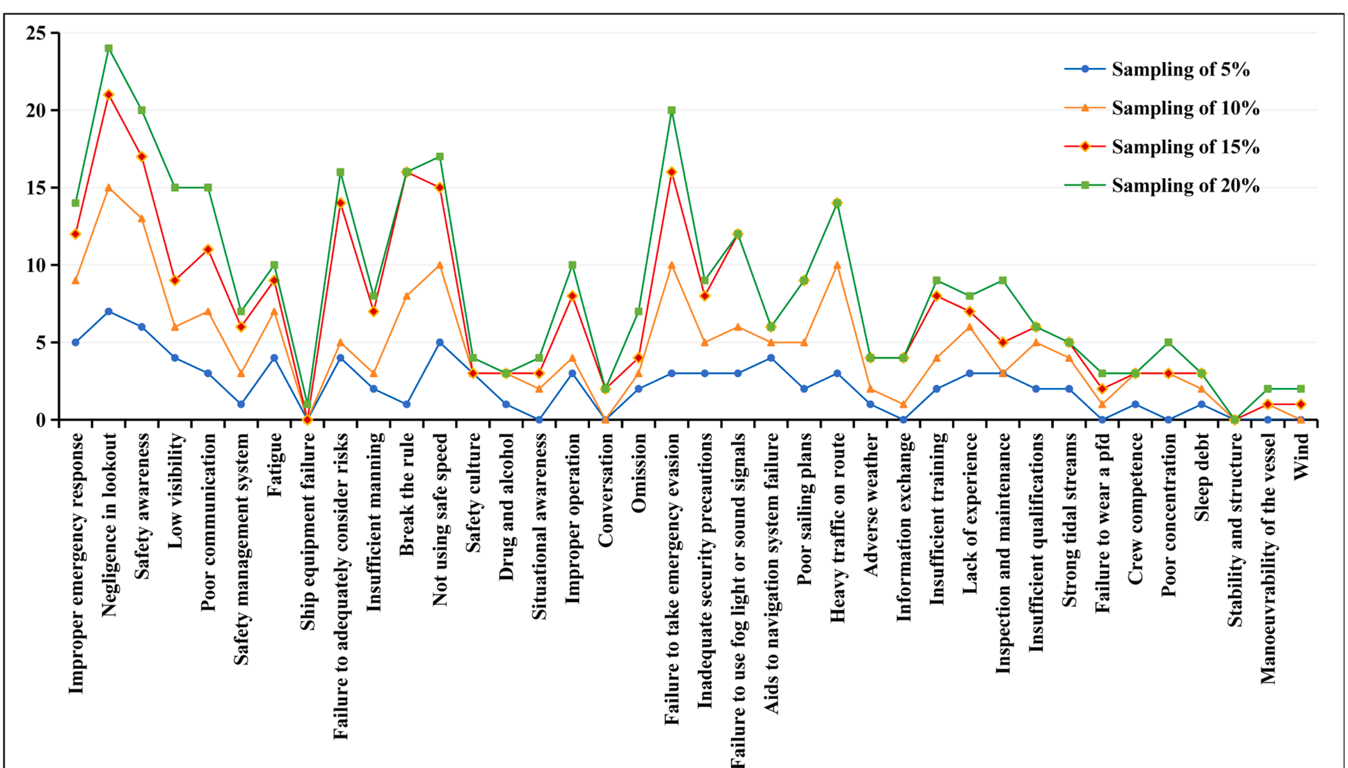


Fig. 8. Validation results for manual sampling of 5%, 10%, 15% and 20%.

operational mistakes on ships.

It is worth noting that factors such as “Ship equipment failure”, “Conversation” and “Heavy wind” were not involved in either the 5% or 10% samples, suggesting that the risks posed by these factors are relatively small during ship navigation. However, risk factors such as miscommunication and high winds were present in 15% and 20% of the samples. In particular, 10% of the sample did not include factors like “Ship equipment failure”, “Conversation”, “Adverse weather”, and “Stability and structure”. This implies that these factors may have a lower probability of contributing to ship collisions. In contrast, “Conversation” was involved in 15% and 20% of the samples, indicating a higher potential impact on safety in actual navigation compared to factors like “Ship equipment failure”, “Conversation”, and “Stability and structure”.

4.5. Implications and discussions

Analysis of accident reports reveal inconsistencies in the accident reporting standards across different national maritime administrations, including report format, content, and accident descriptions. These discrepancies complicate manual statistical analyses and data mining efforts due to the variance in documentation. It is advisable for the IMO and other maritime entities to establish uniform accident reporting standards applicable to all nations. The structure of these reports should encompass detailed descriptions and specifics of the involved vessels, an examination of the accident’s causative factors including human elements, ship machinery and equipment, environmental and navigational conditions, as well as organizational and managerial aspects. Additionally, the reports should detail the findings of the accident, actions taken for safety improvement, and further recommendations. Ensuring the comprehensiveness of accident reports is vital, including all pertinent information for an effective safety analysis.

The methodology of integrating TM and AR overcomes the subjectivity inherent in traditional inductive statistical approaches, offering a novel theoretical framework for identifying maritime and other sector-specific security risks. Building upon the risk factors such as human and management factors mentioned in previous studies [87,101], it enables the discovery of various risk factors not commonly highlighted in prior studies, including drug and alcohol use, omission of emergency evasion actions, neglect in using light or sound signals, and sleep deprivation among others. These factors, although less frequently cited in prior research and underrepresented in conventional databases, emerge as crucial elements in our extensive analysis, significantly contributing to ship collisions. Specifically, the impairment of tactile abilities, judgment, and operational skills due to drug and alcohol consumption among crew members can lead to such collisions. Additionally, in conditions of low visibility, such as fog, the failure to activate lights or sirens in line with collision avoidance regulations can hinder the detectability of the vessel by others, increasing the likelihood of accidents.

In the AR-based deep mining study, appropriate support and confidence thresholds were determined based on the specific context of the research. The determination of these thresholds considers the study’s objectives, the characteristics of the dataset, and insights from previous studies. Adjusting these thresholds facilitates the extraction of ARs at varying levels, enabling a more detailed analysis of the factors contributing to ship collision accidents. It is important to acknowledge that varying support and confidence thresholds yield different AR findings. Lower thresholds might uncover a broader array of ARs, including some that are less significant or contain noise. Conversely, higher thresholds could overlook certain potential associative relationships. Hence, in applying AR mining techniques, the selection of thresholds should be carefully considered to ensure the derivation of accurate and meaningful ARs, tailored to the study’s specific needs.

The application of the FRAM in examining ship collision accidents enhances comprehension of the underlying mechanisms and contributing factors. Unlike traditional linear analysis methods [85] and the AR algorithms, which can only capture correlations between risk factors [88], FRAM addresses the challenges of dissecting complex and dynamic systems by moving beyond conventional linear analyses, offering insights into accident complexity. It underscores the criticality of training crew members in effective lookout procedures and the timely updating of collision warning systems, particularly in intricate and heavily trafficked waterways, a consideration that has been comparatively underemphasized in traditional studies, facilitating, and enables the identification of potential negative resonances between functions. This method elucidates the intricate factors and their interactions leading to accidents, crucial for developing effective prevention and mitigation strategies. Nevertheless, applying FRAM to ship collision accidents presents certain limitations. Constructing and analyzing FRAM networks necessitates extensive data and expertise, with practical applications often hampered by data acquisition challenges, particularly concerning novel risk factors and functional modules. To address these challenges, this study incorporates the AR mining with the FRAM. This integration utilises both the data mining capabilities of AR to provide data support for FRAM, and the interpretive capabilities of FRAM to help understand the possible causal mechanisms of the mined rules in the context of ship collisions. Therefore, this innovative framework further extends and improves ship collision accident analysis, offering a data-driven and objective benchmark in this domain.

To mitigate potential risk factors including low visibility, communication barriers, and inadequate safety management systems, shipping companies can enhance risk management to diminish accidents, reduce economic losses, and augment the efficiency and safety of ship operations. It can also lower the costs associated with repairs, insurance, and legal fees due to accidents. Maritime administrations and authorities bear the responsibility to improve the oversight of ship safety management systems, develop more effective policies and regulations, and establish legal standards regarding issues like alcohol and drug use. Implementing safety awareness education, training, and compliance checks can decrease accidents resulting from negligence, fatigue, or improper handling. Moreover, ensuring optimal crew training and certification processes is essential for crew members to acquire necessary skills and qualifications. Insurance companies can motivate shipping companies to adopt preventive measures through accurate risk assessment and setting reasonable premium rates, potentially influencing insurance product pricing strategies and reducing claim risks. For researchers in ship collision analysis, the integration of TM and FRAM offers innovative research methodologies and encourages interdisciplinary collaboration, deepening the understanding of the systemic nature of ship collisions. Manufacturers of maritime equipment should prioritize the development of advanced technologies and equipment, including enhancements to ship stability, manoeuvrability, and auxiliary navigation systems, while ensuring reliability and minimal maintenance requirements. These manufacturers are also tasked with providing comprehensive technical support to shipping companies, ensuring the effective use of such advanced equipment.

5. Conclusions

This paper constructs a new maritime accident database to comprehensively explore the risk factors of ship collision, which not only fills the gap of the lack of risk data but also overcomes the limitation of previous studies that were largely based on standardised forms of secondary data source. Furthermore, the database also provides robust data support for combining TM and FRAM to analyse the causes of ship collision risks. During the review of various accident reports (Section 2.1), it was discovered that there are significant differences in the format and content of the reports, depending on the country. The findings can serve as a valuable reference for IMO and other maritime organisations, to standardize the format and content of accident reports.

This paper utilizes TM to identify 38 risk factors that are associated with ship collision accidents. These risk factors are multi-dimensional and beyond the state-of-the-art in the literature, including human error, lookout negligence, lack of safety awareness, failure of ship machinery and equipment, steering gear failure, Inadequate safety management system, insufficient training, adverse environment, limited visibility, storms, and dense navigational environment. Additionally, the study introduced a machine learning algorithm based on AR to discover strong associations between these risk factors. This analytical approach lays the groundwork for investigating the causes of ship collision accidents and constructing the FRAM network of ship collision.

Employing the Apriori deep mining technique, it constructs a time series-based FRAM network to analyse ship collision accidents. The collision process is segmented into 11 functional modules, facilitating a comprehensive qualitative analysis of the causative risk factors inherent in each module. The findings indicate that the primary causal factors in the human module include negligence of lookout, operational errors, and insufficient safety awareness. Concurrently, rudder failure emerges as a significant factor within the ship machinery and equipment module. The proposed TM-FRAM model facilitates a comprehensive analysis of the factors contributing to ship collisions. By segmenting the collision process into distinct functional modules, the model yields critical and new insights for ensuring the safe navigation of ships. This methodology not only strengthens the theoretical underpinnings of maritime safety risk assessment but also provides insight for pioneering a methodological framework applicable across various domains.

Future study should be devoted to collecting a wide range of data on ship collision accidents, continuously improving the data collection mechanism, constructing a more complete specialised thesaurus, and merging the thesaurus, to further improve the accuracy of risk causation mining. With ongoing advancements in data mining technology, incorporating time series analysis could enable researchers to examine the dynamic evolution of risk elements over time, thereby enhancing the promptness of risk assessments. The effectiveness of the FRAM methodology is contingent upon the thoroughness and reliability of the data. To mitigate these constraints, it is advisable to augment data standardization and incorporate sophisticated validation techniques to refine future analyses. Moreover, subsequent efforts should contemplate its integration with other risk assessment methodologies, such as FTA and BN, to construct a more exhaustive risk assessment framework. Furthermore, a detailed exploration of the functional resonance mechanism within the FRAM is crucial for understanding the interplay and impact of various functions on ship collision incidents, ultimately facilitating quantitative analysis within the FRAM framework.

CRediT authorship contribution statement

Yuhong Wang: Resources, Project administration, Funding acquisition, Formal analysis. **Pengchang Li:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Cheng Hong:** Validation, Formal analysis. **Zaili Yang:** Writing – review & editing, Writing – original draft, Resources, Project administration, Methodology, Funding acquisition.

Declaration of competing interest

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

Acknowledgements

This project has received funding from the Natural Science Foundation of Ningbo, China (Grant No. 2021J111), the "Innovation Yongjiang 2035" Key R&D Programme (Grant No. 2024H032) and the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement No. 864724).

Appendix

Table 8
FRAM function module detailed description.

Functional module	Input	Output	Precondition	Resource	Time	Control
Crew training	Untrained crew	Trained crew, training certificates, seaworthiness certificates	When new or additional crew is required	Trainers, training facilities, teaching materials	Before working on the ship	Shipping companies
Organisation and Management	Navigation tasks, crew status	Completed safety inspections and ship maintenance, adequately manned and with a sound management system	Ship in operation with crew on duty	Managers, communication systems, etc.	Before the ship sets sail	Managers, management system regulations for shipping companies
Decision-making and planning	Ship operating arrangement requirements	Develop route planning, risk control measures	Sailing plans need to be determined	Captain, navigation equipment, risk assessment tools, etc.	Pre-departure and continuous updates	Captain, navigation system
Prepare to set sail	Sailing command, ship status	Fully equipped, awaiting sailing	Safety inspection and ship equipment maintenance have been carried out, and the ship is seaworthy	Ship equipment, crew, navigation information	Before sailing, when the ship arrives at the port	Notice of sailing
Navigation	Ship ready, awaiting sailing	Navigate following maritime rules (International Maritime Rules (COLREGS)), maintain safe speed	Ship ready and received sailing commands	Navigational equipment, crew, communication system	During navigation	Captain, navigation system
Observations and lookouts	Current navigation status, observation	Report navigation status, navigation environment, natural conditions, etc.	During vessel operation	Telescope, radar, AIS system, etc.	During lookout, in real-time	Lookout personnel

(continued on next page)

Table 8 (*continued*)

Functional module	Input	Output	Precondition	Resource	Time	Control
	requirements, lookout equipment					
Discovering collision risk	Lookout report, radar, AIS, and other data	Collision warning	When there is a potential collision risk	Radar, collision warning system, crew	As soon as there is a potential collision risk	Collision warning system
Emergency handling	Collision warning, emergency plan	Emergency action	Collision warning issued	Emergency equipment, crew	Immediately in an emergency	Emergency plan, captain
Communication and exchange	Communicating and conveying evasive actions when emergency action is inconvenient	Communicate and convey evasion information	When Information Exchange is Needed	Communication Equipment, Crew	During Communication, Real-Time	Communication Equipment, Communication Protocol
Ship manoeuvring	Navigation Commands, Ship Avoidance Information	Execute Maneuver, Change Course or Speed, etc., for Emergency Collision Avoidance	When Need to Change Course or Speed	Helmsman, Ship Steering Equipment	Execute Upon Command Delivery	Captain, Helmsman
Emergency Collision Avoidance	Emergency Collision Avoidance Instructions, Emergency Reports, etc.	Actions to Adjust Ship Position to Avoid	Avoidance Instructions Issued	Emergency steering system, emergency avoidance manual, etc.	Immediate Action in Emergency	Avoidance Guide, Captain

Data availability

Data will be made available on request.

References

[1] Dong Y, Frangopol DM. Probabilistic ship collision risk and sustainability assessment considering risk attitudes. Struct Saf 2015;53:75–84. https://doi.org/10.1016/j.strusafe.2014.10.004.

[2] Montewka J, Ehlers S, Goerlandt F, Hinz T, Tabri K, Kujala P. A framework for risk assessment for maritime transportation systems–a case study for open sea collisions involving RoPax vessels. Reliab Eng Syst Saf 2014;124:142–57. https://doi.org/10.1016/j.ress.2013.11.014.

[3] Altan YC. Collision diameter for maritime accidents considering the drifting of vessels. Ocean Eng 2019;187:106158. https://doi.org/10.1016/j.oceaneng.2019.106158.

[4] Xin X, Liu K, Loughney S, Wang J, Li H, Ekere N, Yang Z. Multi-scale collision risk estimation for maritime traffic in complex port waters. Reliab Eng Syst Saf 2023;240:109554. https://doi.org/10.1016/j.ress.2023.109554.

[5] Xin X, Liu K, Yang Z, Zhang J, Wu X. A probabilistic risk approach for the collision detection of multi-ships under spatiotemporal movement uncertainty. Reliab Eng Syst Saf 2021;215:107772. https://doi.org/10.1016/j.ress.2021.107772.

[6] Yoo Y, Lee JS. Evaluation of ship collision risk assessments using environmental stress and collision risk models. Ocean Eng 2019;191:106527. https://doi.org/10.1016/j.oceaneng.2019.106527.

[7] Rong H, Teixeira AP, Guedes Soares C. Spatial correlation analysis of near ship collision hotspots with local maritime traffic characteristics. Reliab Eng Syst Saf 2021;209:107463. https://doi.org/10.1016/j.ress.2021.107463.

[8] Yu H, Fang Z, Murray AT, Peng G. A direction-constrained space-time prism-based approach for quantifying possible multi-ship collision risks. IEEE Trans Intell Transp Syst 2021;22:131–41. https://doi.org/10.1109/TITS.2019.2955048.

[9] Liu Z, Zhang B, Zhang M, Wang H, Fu X. A quantitative method for the analysis of ship collision risk using AIS data. Ocean Eng 2023;272:113906. https://doi.org/10.1016/j.oceaneng.2023.113906.

[10] Nguyen M, Zhang S, Wang X. A novel method for risk assessment and simulation of collision avoidance for vessels based on AIS. Algorithms 2018;11:204. https://doi.org/10.3390/a11120204.

[11] Zhao L, Fu X. A novel index for real-time ship collision risk assessment based on velocity obstacle considering dimension data from AIS. Ocean Eng 2021;240:109913. https://doi.org/10.1016/j.oceaneng.2021.109913.

[12] Wang M, Wang Y, Cui E, Fu X. A novel multi-ship collision probability estimation method considering data-driven quantification of trajectory uncertainty. Ocean Eng 2023;272:113825. https://doi.org/10.1016/j.oceaneng.2023.113825.

[13] Wang T, Wu Q, A Diaconeasa M, Yan X, Mosleh A. On the use of the hybrid causal logic methodology in ship collision risk assessment. J Mar Sci Eng 2020;8:485. https://doi.org/10.3390/jmse8070485.

[14] Wang YF, Wang LT, Jiang JC, Wang J, Yang ZL. Modelling ship collision risk based on the statistical analysis of historical data: a case study in Hong Kong waters. Ocean Eng 2020;197:106869. https://doi.org/10.1016/j.oceaneng.2019.106869.

[15] Yu Q, Liu K, Yang Z, Wang H, Yang Z. Geometrical risk evaluation of the collisions between ships and offshore installations using rule-based Bayesian reasoning. Reliab Eng Syst Saf 2021;210:107474. https://doi.org/10.1016/j.ress.2021.107474.

[16] Sokukcu M, Sakar C. Risk analysis of collision accidents during underway STS berthing maneuver through integrating fault tree analysis (FTA) into Bayesian network (BN). Appl Ocean Res 2022;126:103290. https://doi.org/10.1016/j.apor.2022.103290.

[17] Zhang M, Zhang D, Goerlandt F, Yan X, Kujala P. Use of HFACS and fault tree model for collision risk factors analysis of icebreaker assistance in ice-covered waters. Saf Sci 2019;111:128–43. https://doi.org/10.1016/j.ssci.2018.07.002.

[18] Sakar C, Toz AC, Buber M, Koseoglu B. Risk analysis of grounding accidents by a fault tree into a bayesian network. Appl Ocean Res 2021;113:102764. https://doi.org/10.1016/j.apor.2021.102764.

[19] Uflaz E, Akyuz E, Arslan O, Gardoni P, Turan O, Aydin M. Analysing human error contribution to ship collision risk in congested waters under the evidential reasoning SPAR-H extended fault tree analysis. Ocean Eng 2023;287:115758. https://doi.org/10.1016/j.oceaneng.2023.115758.

[20] Li Z, Zhu X, Li R, Liao S, Gao K. A comprehensive method for causation analysis of ship–ice collision risk in Arctic waters. Environ Sci Pollut Res 2023. https://doi.org/10.1007/s11356-023-28249-7.

[21] Sotiralis P, Ventikos NP, Hamann R, Golyshev P, Teixeira AP. Incorporation of human factors into ship collision risk models focusing on human centred design aspects. Reliab Eng Syst Saf 2016;156:210–27. https://doi.org/10.1016/j.ress.2016.08.007.

[22] Chang CH, Kontovas C, Yu Q, Yang Z. Risk assessment of the operations of maritime autonomous surface ships. Reliab Eng Syst Saf 2021;207:107324. https://doi.org/10.1016/j.ress.2020.107324.

[23] Wan C, Yan X, Zhang D, Qu Z, Yang Z. An advanced fuzzy Bayesian-based FMEA approach for assessing maritime supply chain risks. Transp Res Part E Logist Transp Rev 2019;125:222–40. https://doi.org/10.1016/j.tre.2019.03.011.

[24] Afenyo M, Khan F, Veitch B, Yang M. Arctic shipping accident scenario analysis using Bayesian Network approach. Ocean Eng 2017;133:224–30. https://doi.org/10.1016/j.oceaneng.2017.02.002.

[25] Yu Q, Liu K, Chang CH, Yang Z. Realising advanced risk assessment of vessel traffic flows near offshore wind farms. Reliab Eng Syst Saf 2020;203:107086. https://doi.org/10.1016/j.ress.2020.107086.

[26] Zhang D, Yan XP, Yang ZL, Wall A, Wang J. Incorporation of formal safety assessment and Bayesian network in navigational risk estimation of the Yangtze River. Reliab Eng Syst Saf 2013;118:93–105. https://doi.org/10.1016/j.ress.2013.04.006.

[27] Yang Z, Bonsall S, Wang J. A fuzzy bayesian reasoning method to realise interactive failure analysis. In: Proceedings of the 8th international conference on reliability, maintainability and safety (ICRMS 2009). IEEE; 2009. p. 403–6. https://doi.org/10.1109/ICRMS.2009.5270163.

[28] Yang ZL, Bonsall S, Wang J. Use of hybrid multiple uncertain attribute decision making techniques in safety management. Expert Syst Appl 2009;36:1569–86. https://doi.org/10.1016/j.eswa.2007.11.054.

[29] Ma X, Lan H, Qiao W, Han B, He H. On the causation correlation of maritime accidents based on data mining techniques. Proc Inst Mech Eng Part O J Risk Reliab 2022. https://doi.org/10.1177/1748006×221131717. 1748006×2211317.

[30] Shah SM, Lütjen M, Freitag M. Text mining for supply chain risk management in the apparel industry. Appl Sci 2021;11:2323. https://doi.org/10.3390/app11052323.

[31] Senave E, Jans MJ, Srivastava RP. The application of text mining in accounting. Int J Account Inf Syst 2023;50:100624. https://doi.org/10.1016/j.accinf.2023.100624.

[32] Wei L, Li G, Zhu X, Li J. Discovering bank risk factors from financial statements based on a new semi-supervised text mining algorithm. Account Financ 2019;59:1519–52. https://doi.org/10.1111/acfi.12453.
[33] Jonnagaddala J, Liaw ST, Ray P, Kumar M, Chang NW, Dai HJ. Coronary artery disease risk assessment from unstructured electronic health records using text mining. J Biomed Inform 2015;58:S203–10. https://doi.org/10.1016/j.jbi.2015.08.003.
[34] Li S, You M, Li D, Liu J. Identifying coal mine safety production risk factors by employing text mining and Bayesian network techniques. Process Saf Environ Prot 2022;162:1067–81. https://doi.org/10.1016/j.psep.2022.04.054.
[35] Shi D, Guan J, Zurada J, Manikas A. A data-mining approach to identification of risk factors in safety management systems. J Manag Inf Syst 2017;34:1054–81. https://doi.org/10.1080/07421222.2017.1394056.
[36] Baksh AA, Abbassi R, Garaniya V, Khan F. Marine transportation risk assessment using Bayesian Network: application to Arctic waters. Ocean Eng 2018;159:422–36. https://doi.org/10.1016/j.oceaneng.2018.04.024.
[37] Liao S, Weng J, Zhang Z, Li Z, Li F. Probabilistic modeling of maritime accident scenarios leveraging Bayesian network techniques. J Mar Sci Eng 2023;11:1513. https://doi.org/10.3390/jmse11081513.
[38] Ung ST. Evaluation of human error contribution to oil tanker collision using fault tree analysis and modified fuzzy Bayesian Network based CREAM. Ocean Eng 2019;179:159–72. https://doi.org/10.1016/j.oceaneng.2019.03.031.
[39] Zhang J, Jin M, Wan C, Dong Z, Wu X. A Bayesian network-based model for risk modeling and scenario deduction of collision accidents of inland intelligent ships. Reliab Eng Syst Saf 2024;243:109816. https://doi.org/10.1016/j.ress.2023.109816.
[40] Acharya TD, Yoo KW, Lee DH. GIS-based spatio-temporal analysis of marine accidents database in the coastal zone of Korea. J Coast Res 2017;79:114–8. https://doi.org/10.2112/SI79-024.1.
[41] Antão P, Sun S, Teixeira AP, Guedes Soares C. Quantitative assessment of ship collision risk influencing factors from worldwide accident and fleet data. Reliab Eng Syst Saf 2023;234:109166. https://doi.org/10.1016/j.ress.2023.109166.
[42] Aydin M, Uğurlu Ö, Boran M. Assessment of human error contribution to maritime pilot transfer operation under HFACS-PV and SLIM approach. Ocean Eng 2022;266:112830. https://doi.org/10.1016/j.oceaneng.2022.112830.
[43] Fan H, Lyu J, Chang Z, He X, Guo S. Spatial patterns and characteristics of global piracy analyzed using a geographic information system. Mar Policy 2023;157:105816. https://doi.org/10.1016/j.marpol.2023.105816.
[44] He Z, Wang C, Gao J, Xie Y. Assessment of global shipping risk caused by maritime piracy. Heliyon 2023;9:e20988. https://doi.org/10.1016/j.heliyon.2023.e20988.
[45] Li H, Ren X, Yang Z. Data-driven Bayesian network for risk analysis of global maritime accidents. Reliab Eng Syst Saf 2023;230:108938. https://doi.org/10.1016/j.ress.2022.108938.
[46] Li H, Yang Z. Towards safe navigation environment: the imminent role of spatio-temporal pattern mining in maritime piracy incidents analysis. Reliab Eng Syst Saf 2023;238:109422. https://doi.org/10.1016/j.ress.2023.109422.
[47] Li YZ, Hu H, Huang DZ. Data fusion of maritime incident databases with Dempster–Shafer theory. Transp Res Rec 2014;2426:27–32. https://doi.org/10.3141/2426-04.
[48] Lu J, Su W, Jiang M, Ji Y. Severity prediction and risk assessment for non-traditional safety events in sea lanes based on a random forest approach. Ocean Coast Manag 2022;225:106202. https://doi.org/10.1016/j.ocecoaman.2022.106202.
[49] Pristrom S, Yang Z, Wang J, Yan X. A novel flexible model for piracy and robbery assessment of merchant ship operations. Reliab Eng Syst Saf 2016;155:196–211. https://doi.org/10.1016/j.ress.2016.07.001.
[50] Uğurlu Ö, Köse E, Yıldırım U, Yüksekyıldız E. Marine accident analysis for collision and grounding in oil tanker using FTA method. Marit Policy Manag 2015;42:163–85. https://doi.org/10.1080/03088839.2013.856524.
[51] Wang J, Zhou Y, Zhang S, Zhuang L, Shi L, Chen J, Hu D. Societal risk acceptance criteria of the global general cargo ships. Ocean Eng 2022;261:112162. https://doi.org/10.1016/j.oceaneng.2022.112162.
[52] Zhang Y, Zhai Y, Fu S, Shi M, Jiang X. Quantitative analysis of maritime piracy at global and regional scales to improve maritime security. Ocean Coast Manag 2024;248:106968. https://doi.org/10.1016/j.ocecoaman.2023.106968.
[53] Wróbel K, Montewka J, Kujala P. Towards the assessment of potential impact of unmanned vessels on maritime transportation safety. Reliab Eng Syst Saf 2017;165:155–69. https://doi.org/10.1016/j.ress.2017.03.029.
[54] Demirci SE, Canımoğlu R, Elçiçek H. Analysis of causal relations of marine accidents during ship navigation under pilotage: a DEMATEL approach. Proc Inst Mech Eng Part M J Eng Marit Environ 2023;237:308–21. https://doi.org/10.1177/14750902221127093.
[55] Fan S, Blanco-Davis E, Fairclough S, Zhang J, Yan X, Wang J, Yang Z. Incorporation of seafarer psychological factors into maritime safety assessment. Ocean Coast Manag 2023;237:106515. https://doi.org/10.1016/j.ocecoaman.2023.106515.
[56] Fan S, Blanco-Davis E, Yang Z, Zhang J, Yan X. Incorporation of human factors into maritime accident analysis using a data-driven Bayesian network. Reliab Eng Syst Saf 2020;203:107070. https://doi.org/10.1016/j.ress.2020.107070.
[57] Fan S, Zhang J, Blanco-Davis E, Yang Z, Yan X. Maritime accident prevention strategy formulation from a human factor perspective using Bayesian Networks and TOPSIS. Ocean Eng 2020;210:107544. https://doi.org/10.1016/j.oceaneng.2020.107544.
[58] Özkan U, Umut Y, Ersan B. Analysis of grounding accidents caused by human error. J Mar Sci Technol 2015;23. https://doi.org/10.6119/JMST-015-0615-1.
[59] Fan S, Yang Z, Blanco-Davis E, Zhang J, Yan X. Analysis of maritime transport accidents using Bayesian networks. Proc Inst Mech Eng Part O J Risk Reliab 2020;234:439–54. https://doi.org/10.1177/1748006×19900850.
[60] Fan S, Yang Z, Wang J, Marsland J. Shipping accident analysis in restricted waters: lesson from the Suez Canal blockage in 2021. Ocean Eng 2022;266:113119. https://doi.org/10.1016/j.oceaneng.2022.113119.
[61] Kum S, Sahin B. A root cause analysis for Arctic Marine accidents from 1993 to 2011. Saf Sci 2015;74:206–20. https://doi.org/10.1016/j.ssci.2014.12.010.
[62] Arici SS, Akyuz E, Arslan O. Application of fuzzy bow-tie risk analysis to maritime transportation: the case of ship collision during the STS operation. Ocean Eng 2020;217:107960. https://doi.org/10.1016/j.oceaneng.2020.107960.
[63] Ebrahimi H, Sattari F, Lefsrud L, Macciotta R. Analysis of train derailments and collisions to identify leading causes of loss incidents in rail transport of dangerous goods in Canada. J Loss Prev Process Ind 2021;72:104517. https://doi.org/10.1016/j.jlp.2021.104517.
[64] Graziano A, Teixeira AP, Guedes Soares C. Classification of human errors in grounding and collision accidents using the TRACEr taxonomy. Saf Sci 2016;86:245–57. https://doi.org/10.1016/j.ssci.2016.02.026.
[65] Rudin-Brown CM, Harris S, Rosberg A. How shift scheduling practices contribute to fatigue amongst freight rail operating employees: findings from Canadian accident investigations. Accid Anal Prev 2019;126:64–9. https://doi.org/10.1016/j.aap.2018.01.027.
[66] Çakır E. Fatal and serious injuries on board merchant cargo ships. Int Marit Health 2019;70:113–8. https://doi.org/10.5603/IMH.2019.0018.
[67] Coraddu A, Oneto L, Navas De Maya B, Kurt R. Determining the most influential human factors in maritime accidents: a data-driven approach. Ocean Eng 2020;211:107588. https://doi.org/10.1016/j.oceaneng.2020.107588.
[68] Djurhuus R. Fumigation on bulk cargo ships: a chemical threat to seafarers. Int Marit Health 2021;72:206–16. https://doi.org/10.5603/IMH.2021.0039.
[69] Fan S, Yang Z. Analysing seafarer competencies in a dynamic human-machine system. Ocean Coast Manag 2023;240:106662. https://doi.org/10.1016/j.ocecoaman.2023.106662.
[70] Hasanspahić N, Vujicic S, Francic V, Campara L. The role of the human factor in marine accidents. J Mar Sci Eng 2021;9(3):261. https://doi.org/10.3390/jmse9030261.
[71] Majumdar A, Manole I, Nalty R. Analysis of port accidents and calibration of Heinrich's pyramid. Transp Res Rec 2022;2676:476–89. https://doi.org/10.1177/03611981211044447.
[72] Mazaheri A, Montewka J, Nisula J, Kujala P. Usability of accident and incident reports for evidence-based risk modeling–a case study on ship grounding reports. Saf Sci 2015;76:202–14. https://doi.org/10.1016/j.ssci.2015.02.019.
[73] Ma XF, Shi GY, Liu ZJ. TAR-based domino effect model for maritime accidents. J Mar Sci Eng 2022;10:788. https://doi.org/10.3390/jmse10060788.
[74] De Voogt A, St Amour E. Safety of twin-engine helicopters: risks and operational specificity. Saf Sci 2021;136:105169. https://doi.org/10.1016/j.ssci.2021.105169.
[75] DeLoughery EP. Characteristics of fatal marine accidents. Int Marit Health 2022;73:115–6. https://doi.org/10.5603/IMH.2022.0022.
[76] Dinh-Zarr TB, Shuey R, Mooren L. The Five "I" Framework of crash investigation: linking investigation practices to safety reform. Accid Anal Prev 2024;194:107296. https://doi.org/10.1016/j.aap.2023.107296.
[77] Laukkala T, Bor R, Budowle B, Sajantila A, Navathe P, Sainio M, Vuorio A. Attention-deficit/hyperactivity disorder and fatal accidents in aviation medicine. Aerosp Med Hum Perform 2017;88:871–5. https://doi.org/10.3357/AMHP.4919.2017.
[78] Marcus JH, Rosekind MR. Fatigue in transportation: NTSB investigations and safety recommendations. Inj Prev 2017;23(4):232–8. https://doi.org/10.1136/injuryprev-2015-041791.
[79] McKay MP, Poland K, Barth TH. Medical investigation after a motorcoach strikes the rear of a tractor semitrailer parked on the highway; 13 fatal. Traffic Inj Prev 2018;19:S162–4. https://doi.org/10.1080/15389588.2018.1532215.
[80] Nix S, Buckner S, Cercone R. A review of risk analysis and helicopter air ambulance accidents. Air Med J 2014;33:218–21. https://doi.org/10.1016/j.amj.2014.06.011.
[81] Rao AH, Marais K. A state-based approach to modeling general aviation accidents. Reliab Eng Syst Saf 2020;193:106670. https://doi.org/10.1016/j.ress.2019.106670.
[82] Sorrentino CJ, Schulz J, Chaudry A, Cohen E, Friedman Y. The legal limit: differences in injury severity score for vehicular crashes. J Trauma Nurs 2016;23:184–8. https://doi.org/10.1097/JTN.0000000000000214.
[83] Yang H, Ozbay K, Ozturk O, Xie K. Work Zone safety analysis and modeling: a state-of-the-art review. Traffic Inj Prev 2015;16:387–96. https://doi.org/10.1080/15389588.2014.948615.
[84] Zhang X, Mahadevan S. Bayesian network modeling of accident investigation reports for aviation safety assessment. Reliab Eng Syst Saf 2021;209:107371. https://doi.org/10.1016/j.ress.2020.107371.
[85] Cao Y, Wang X, Wang Y, Fan S, Wang H, Yang Z, Liu Z, Wang J, Shi R. Analysis of factors affecting the severity of marine accidents using a data-driven Bayesian network. Ocean Eng 2023;269:113563. https://doi.org/10.1016/j.oceaneng.2022.113563.
[86] Ung ST. Navigation Risk estimation using a modified Bayesian Network modeling-a case study in Taiwan. Reliab Eng Syst Saf 2021;213:107777. https://doi.org/10.1016/j.ress.2021.107777.
[87] Ugurlu H, Cicek I. Analysis and assessment of ship collision accidents using fault tree and multiple correspondence analysis. Ocean Eng 2022;245:110514. https://doi.org/10.1016/j.oceaneng.2021.110514.

[88] Jia X, Zhang D. Prediction of maritime logistics service risks applying soft set based association rule: an early warning model. Reliab Eng Syst Saf 2021;207:107339. https://doi.org/10.1016/j.ress.2020.107339.

[89] Wang Z, Yin J. Risk assessment of inland waterborne transportation using data mining. Marit Policy Manag 2020;47:633–48. https://doi.org/10.1080/03088839.2020.1738582.

[90] Xia G, Wang X, Feng Y, Cao Y, Dai Z, Wang H, Liu Z. Navigational risk of inland water transportation: a case study in the Songhua River, China. ASCE-ASME J Risk Uncertain Eng Syst Part A Civ Eng 2023;9(4). https://doi.org/10.1061/ajrua6.rueng-1158.

[91] Boggs J, Kafka J. A critical review of text mining applications for suicide research. Curr Epidemiol Rep 2022;9(3):126–34. https://doi.org/10.1007/s40471-022-00293-w.

[92] Dorsey LC, Wang B, Grabowski M, Merrick J, Harrald JR. Self healing databases for predictive risk analytics in safety-critical systems. J Loss Prev Process Ind 2020;63:104014. https://doi.org/10.1016/j.jlp.2019.104014.

[93] Anvarifar F, Voorendt MZ, Zevenbergen C, Thissen W. An application of the Functional Resonance Analysis Method (FRAM) to risk analysis of multifunctional flood defences in the Netherlands. Reliab Eng Syst Saf 2017;158:130–41. https://doi.org/10.1016/j.ress.2016.10.004.

[94] Guo Y, Jin Y, Hu S, Yang Z, Xi Y, Han B. Risk evolution analysis of ship pilotage operation by an integrated model of FRAM and DBN. Reliab Eng Syst Saf 2023;229:108850. https://doi.org/10.1016/j.ress.2022.108850.

[95] De Carvalho PVR. The use of Functional Resonance Analysis Method (FRAM) in a mid-air collision to understand some characteristics of the air traffic management system resilience. Reliab Eng Syst Saf 2011;96:1482–98. https://doi.org/10.1016/j.ress.2011.05.009.

[96] Zinetullina A, Yang M, Khakzad N, Golman B, Li X. Quantitative resilience assessment of chemical process systems using functional Resonance Analysis Method and dynamic Bayesian Network. Reliab Eng Syst Saf 2021;205:107232. https://doi.org/10.1016/j.ress.2020.107232.

[97] Salihoglu E, Besikci EB. The use of Functional Resonance Analysis Method (FRAM) in a maritime accident: a case study of Prestige. Ocean Eng 2021:219.

[98] Ma L, Ma X, Liu Y, Deng W, Lan H. Risk assessment of coupling links in hazardous chemicals maritime transportation system. J Loss Prev Process Ind 2023:82.

[99] Fu S, Yu Y, Chen J, Han B, Wu Z. Towards a probabilistic approach for risk analysis of nuclear-powered icebreakers using FMEA and FRAM. Ocean Eng 2022;260:112041. https://doi.org/10.1016/j.oceaneng.2022.112041.

[100] Wang L, Huang R, Shi W, Zhang C. Domino effect in marine accidents: evidence from temporal association rules. Transp Policy 2021;103:236–44. https://doi.org/10.1016/j.tranpol.2021.02.006.

[101] Lan H, Ma X, Ma L, Qiao W. Pattern investigation of total loss maritime accidents based on association rule mining. Reliab Eng Syst Saf 2023;229:108893. https://doi.org/10.1016/j.ress.2022.108893.

[102] Yu Y, Chen L, Shu Y, Zhu W. Evaluation model and management strategy for reducing pollution caused by ship collision in coastal waters. Ocean Coast Manag 2021;203:105446. https://doi.org/10.1016/j.ocecoaman.2020.105446.

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

[104] Chen Y, Lou N, Liu G, Luan Y, Jiang H. Risk analysis of ship detention defects based on association rules. Mar Policy 2022;142:105123. https://doi.org/10.1016/j.marpol.2022.105123.

[105] Peng C, Zhen X, Huang Y. A multi-layer Fram based approach to short-term human-automation resilience assessment: a case study on dynamic positioning system in offshore tandem offloading process. Ocean Eng 2022;266:112728. https://doi.org/10.1016/j.oceaneng.2022.112728.

[106] Qiao W, Ma X, Liu Y, Deng W. Resilience evaluation of maritime liquid cargo emergency response by integrating FRAM and a BN: a case study of a propylene leakage emergency scenario. Ocean Eng 2022;247:110584. https://doi.org/10.1016/j.oceaneng.2022.110584.

[107] Fan S, Yang Z. Accident data-driven human fatigue analysis in maritime transport using machine learning. Reliab Eng Syst Saf 2024;241:109675. https://doi.org/10.1016/j.ress.2023.109675.

[108] Uronen L, Moen H, Teperi S, Martimo KP, Hartiala J, Salanterä S. Towards automated detection of psychosocial risk factors with text mining. Occup Med 2020;70:203–6. https://doi.org/10.1093/occmed/kqaa022.

[109] Deiva Ganesh A, Kalpana P. Supply Chain Risk Identification: a real-time data-mining approach. Ind Manag Data Syst 2022;122(5):1333–54. https://doi.org/10.1108/imds-11-2021-0719.

[110] Li W, He M, Sun Y, Cao Q. A proactive operational risk identification and analysis framework based on the integration of ACAT and FRAM. Reliab Eng Syst Saf 2019;186:101–9. https://doi.org/10.1016/j.ress.2019.02.012.

[111] Peng C, Zhen X, Huang Y. Human-automation interaction centered approach based on Fram for systemic safety analysis of dynamic positioning operations for offshore tandem offloading. Ocean Eng 2023;267:113249. https://doi.org/10.1016/j.oceaneng.2022.113249.

[112] Togawa T, Morita H, Tsuji T. Applying the functional resonance analysis method (FRAM) to flood risk management at a community level: Koriyama City's emergency-response process during Typhoon Hagibis. Prog Disaster Sci 2023;19:100291. https://doi.org/10.2139/ssrn.4370049.

[113] Yu M, Erraguntla M, Quddus N, Kravaris C. A data-driven approach of quantifying function couplings and identifying paths towards emerging hazards in complex systems. Process Saf Environ Prot 2021;150:464–77. https://doi.org/10.1016/j.psep.2021.04.037.