Li, H, Çelik, C, Bashir, M, Zou, L and Yang, Z

Incorporation of a global perspective into data-driven analysis of maritime collision accident risk

Article

Reliability Engineering and System Safety 249 (2024) 110187

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Incorporation of a global perspective into data-driven analysis of maritime collision accident risk

Huanhuan Li [a], Cihad Çelik [a], Musa Bashir [a], Lu Zou [b,c], Zaili Yang [a,*]

[a] Liverpool Logistics, Offshore and Marine Research Institute, Liverpool John Moores University, Liverpool, UK
[b] School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, China
[c] State Key Laboratory of Ocean Engineering, Shanghai Jiao Tong University, Shanghai, China

ARTICLE INFO

Keywords:
Maritime safety
Global maritime transportation
Maritime collision accidents
Bayesian network

ABSTRACT

Ship collision accidents are one of the most frequent accident types in global maritime transportation. Nevertheless, conducting an in-depth analysis of collision prevention poses a formidable challenge due to the constraints of limited Risk Influential Factors (RIFs) and available datasets. This paper aims to incorporate a global perspective into a new data-driven risk model, scrutinize the root causes of collision accidents, and advance measures for their mitigation. Additionally, it seeks to analyze the spatial distribution and conduct a comprehensive comparative study on collision characteristics for both pre- and post-COVID-19, utilizing the real accident dataset collected from two reputable organizations: Global Integrated Shipping Information System (GISIS) and Lloyd's Register Fairplay (LRF). The research findings and implications encompass several crucial aspects: 1) the constructed model demonstrates its reliability and accuracy in predicting collision accidents, as evident from its prediction performance and various scenario analysis; 2) the most hazardous voyage segment for collision accidents is identified to provide valuable guidance to different stakeholders; and 3) the hierarchical significance of various ship types in the context of collision accident is highlighted regarding the most probable scenario for collision occurrences; 4) During the pandemic, the rise in collision probabilities, particularly involving older vessels and bulk carriers, implies heightened operational challenges or maintenance issues for these ship types; (5) The prominence of favorable and adverse sea conditions in collision reports underscores the significant influence of weather on accidents during the pandemic. These findings and implications help enhance safety protocols, ultimately reducing the frequency of collision accidents in the global maritime domain.

1. Introduction

Collision accidents rank among the most frequent accident types in global maritime transport [1,2]. As reported by the European Maritime Safety Agency (EMSA) [3], ship collision accidents have resulted in dire consequences, including substantial economic damages, loss of lives, and significant environmental degradation. These severe outcomes underscore the paramount importance of maritime safety and the pressing need for proactive prevention measures. In response to this critical issue, the International Maritime Organization (IMO) has introduced the Formal Safety Assessment (FSA) procedure, aiming to assess and mitigate risks within the maritime domain effectively. The primary objective is to implement preventive measures against potential accidents and enhance overall safety. Furthermore, as part of its commitment to reducing the risk of maritime collision accidents, the IMO has put forth the Convention on the International Regulations for Preventing Collisions at Sea (COLREGs). These essential regulations, mandated for adherence by all maritime vessels engaged in waterborne transportation, play a pivotal role in safeguarding against collisions and promoting safe navigation practices.

Ship collision accidents can occur in various environments, encompassing open seas, coastal waters, ports, straits, inland waterways, Arctic regions, trade corridors, and other pertinent locations. The surging demand for maritime transportation has led to complex and high-density traffic, especially in restricted and narrow waterways, highlighting navigational safety concerns. For instance, Liu et al. [4] found that collision accidents form a substantial portion of maritime accidents in China's coastal waters, a trend evident in both historical and global annual analyses. Along the Yangtze River, recognized as the world's busiest inland waterway, collision accidents accounted for 59.18 % of

* Corresponding author.
E-mail address: Z.Yang@ljmu.ac.uk (Z. Yang).

https://doi.org/10.1016/j.ress.2024.110187
Received 25 October 2023; Received in revised form 8 March 2024; Accepted 4 May 2024
Available online 15 May 2024

the reported accidents from 2012 to 2021 [5]. In the İstanbul Strait, one of the narrowest and most congested maritime channels globally, the Kandilli sector exhibits the highest collision likelihood, as highlighted by Kamal and Çakır [6]. In inland rivers, collision accidents are a common occurrence accident type. Collision accidents are also common in inland rivers, prompting Kandel and Baroud [7] to propose intelligent technologies like ship speed optimization and situational awareness to decrease their probability. Furthermore, the warming of the Arctic region has opened up alternative routes in maritime transportation, leading to an uptick in collision-related accidents, particularly for longer vessels, as predicted by Zhang et al. [8]. The main challenge in regional risk analysis studies lies in the identification of Risk Influential Factors (RIFs) tailored to each unique geographical area. Although such studies shed light on accident risk and severity, they might fall short of offering a comprehensive view of global maritime collision accidents.

The literature comprises both qualitative and quantitative studies for conducting maritime accident risk analysis. The primary goals of these analyses are risk assessment and the implementation of necessary precautions to prevent adverse consequences in the maritime domain. Qualitative studies typically rely on expert opinions, but it is essential to manage the uncertainties stemming from such subjective data effectively. In contrast, quantitative research doesn't entail these uncertainties. However, the primary constraint for quantitative studies is adequate databases, as a substantial sample size is vital for training models effectively. Advanced techniques like fuzzy logic, the Dempster-Shafer (D-S) theory, and the Bayesian probabilistic method are employed for these analyses. Among these methods, the Bayesian Network (BN) model is prominent for its comprehensive risk analysis capabilities, allowing for dual-direction interpretation both from cause-to-root and root-to-cause [9]. Furthermore, BN can reveal relationships among RIFs by integrating sophisticated algorithms like the Naïve Bayesian network (NBN) and the Tree Augmented Bayesian Network (TAN).

Collision accident-related studies can be classified into three groups depending on the granularity of collision events and the data type utilized [10]. The first category focuses on reducing collision accidents by evaluating both singular and multi-ship encounter situations [11,12]. The second category concentrates on diminishing collision accidents by analyzing ship traffic patterns using data from the Automatic Identification System (AIS) [13]. The third category's objective is not only to reduce the frequency of collisions but also to mitigate their consequences by assessing risks using data gathered from historical accident records [14,15]. A comparative analysis of the three categories reveals the distinct focuses within the field: the first two primarily target reducing the likelihood of collision accidents, while the third adopts a more comprehensive approach to collision prevention. By leveraging historical accident data, it offers a broader perspective derived from past occurrences, aiming to decrease both the number of collisions and the impact of such events. Consequently, it strengthens proactive prevention and effective management of the aftermath, ensuring a robust and all-encompassing safety strategy. While some studies have explored the impact of the COVID-19 pandemic on the mental health of seafarers [16], maritime pilots [17], and other stakeholders [18], a comprehensive examination of the repercussions of COVID-19 on maritime accidents is still lacking. Furthermore, detailed research on the spatial and temporal characteristics of collisions, both before and during COVID-19, remains limited.

According to the abovementioned research status, this paper delves into a global analysis of maritime collision accidents from a data-driven and macro perspective, drawing from historical data sourced from the Global Integrated Shipping Information System (GISIS) and Lloyd's Register Fairplay (LRF). Utilizing the latitude and longitude information from collision accidents in the IMO dataset, this paper advances the analysis by examining the spatial distribution characteristics. It identifies and extracts significant features related to the geographical spread of these accidents. The generated dataset, which is among the most recent, includes 402 accident records between 2017 and 2021. The methodology employs a comprehensive quantitative approach rooted in historical accident data, thereby eliminating uncertainties tied to expert judgments. Leveraging the strengths of the BN model using the TAN algorithm, a thorough examination has been conducted to investigate the complex interplay among RIFs and facilitate root-cause analysis. Additionally, this paper provides a comprehensive analysis aimed at reducing both the frequency of collisions and their ensuing consequences, providing readers with an overall risk perspective. Finally, this paper further compares the characteristics of collision accidents before and during COVID-19, revealing significant findings and implications, thereby providing effective guidance for the entire maritime industry.

The paper is structured as follows. Section 2 introduces the literature review of BN-based maritime accident studies, focusing on collision accidents in particular, and highlights the research gaps, addressing them with new contributions. Section 3 provides information on the mathematical background of the data-driven BN model and identifies the details of the RIFs. The validation of the constructed model has been conducted using fundamental analysis in Section 4. Section 5 discusses the implications through scenario analysis. A thorough comparative analysis is performed in Section 6 to offer a comprehensive insight into the collision characteristics both pre- and during the pandemic. Additionally, the findings of this research are compared with the literature studies, specifically concerning collision accidents in terms of RIFs and ship types. Finally, the conclusions are highlighted in Section 7.

2. Literature review

2.1. BN-based maritime accident risk analysis

The IMO established the FSA to systematically evaluate maritime risks and enhance safety measures. Leveraging the principles of FSA, numerous global maritime risk studies, spanning both qualitative and quantitative evaluations, have been undertaken. Prominent methods for qualitative risk assessments encompass the Functional Resonance Analysis Method (FRAM) [19], Root Cause Analysis (RCA), the Human Factor Analysis and Classification System (HFACS) [20,21], as well as Strength Weakness Opportunity and Threat (SWOT) analysis, among others. A primary concern with using qualitative methods is the potential uncertainty in their results, mainly due to their reliance on expert views that can be subjective. In contrast, quantitative risk assessment models address this issue by utilizing objective data [22]. Fuzzy Logic (FL), Evidence Reasoning (ER), Fault Tree Analysis (FTA), Event Tree Analysis (ETA), and BN methods are frequently employed within the maritime domain.

Each model has its own set of strengths and limitations based on its design and implementation. For instance, the FTA and ETA models yield binary outcomes from their analyses, rendering the interpretation of root-cause relationships less adaptable and detailed. To enhance interpretive coherence, some models are synergistically combined. Within this framework, the Hybrid Causal Logic (HCL) model incorporates the FTA method to investigate the events in the Even Sequence Diagram (ESD), as well as the BN to investigate the human factors, as demonstrated by Zhang et al. [23]. The primary objective was to discern event sequences leading to accidents for Maritime Autonomous Surface Ships (MASS). Drawing from expert opinions, Yu et al. [24] formulated an integrated model based on BN and ER to evaluate the ship-specific risks associated with geographical-dependent factors. In this context, ER aids in mitigating uncertainties stemming from expert judgment.

The application of BN analysis in maritime accident research can incorporate objective and subjective data or a combination of both. For instance, Zhang et al. [25] created a BN model for predicting accident outcomes at Tianjin Port, integrating historical accident records with expert knowledge. Nevertheless, the subjective nature of expert insights introduces uncertainties. To mitigate such uncertainties, leveraging robust databases is recommended to reduce reliance solely on expert experience. Recent studies have investigated maritime accidents by directly mining data from historical records of maritime organizations.

Despite the availability of abundant training data for BN, selecting the suitable algorithm for specific challenges remains an obstacle. The BN model offers various algorithms, encompassing the NBN, Augmented Naïve Bayesian Network (ABN), and TAN. Fan et al. [26] adopted the NBN to analyze maritime accident risks using a relatively new historical accident dataset, providing quantitative insights. However, due to limitations, the NBN model couldn't capture intricate relationships among RIFs. The limitation was addressed by adopting the ABN algorithm, which enriches the naïve structure by establishing ties between leaf nodes based on the value of the target node [27]. Li et al. [1] employed the TAN approach to analyze global maritime accidents spanning 2017–2021, facilitating the identification of RIF relationships without solely relying on expert perspectives. Using the same approach, Cao et al. [28] evaluated accident severity using historical accident data from various national databases, identifying key RIFs such as ship type, voyage segment/location, deadweight/gross tonnage, and engine power. Additionally, Zhou et al. [29] developed a data-driven BN model to observe changes in RIF roles through yearly analysis of maritime accidents.

Furthermore, the application of the BN model, which utilizes historical accident data, is prevalent in maritime risk analysis. Zhao et al. [30] introduced an accident prevention model using a data-driven BN framework specially tailored for autonomous ships, aiming to mitigate human factors' influence on accidents. Combining BN with TOPSIS, Fan et al. [31] tackled this challenge of human factors in accident prevention, leveraging data from the Marine Accident Investigation Branch (MAIB) and the Transportation Safety Board (TSB) database.

This comprehensive BN analysis pinpointed information, clear order, and safety culture as crucial factors for advancing maritime safety, particularly regarding human factors. Additionally, there has been an emphasis on regional studies, with BN-based risk analyses conducted for various regions such as the Maritime Silk Road [32], Suez Canal [33], China Coasts [27], İstanbul Straight [6], and Artric Waters [15,34]. However, challenges in regional risk analysis studies include region-based influential factors, potentially limiting insights applicable to global maritime accidents. Some researchers have also utilized Port

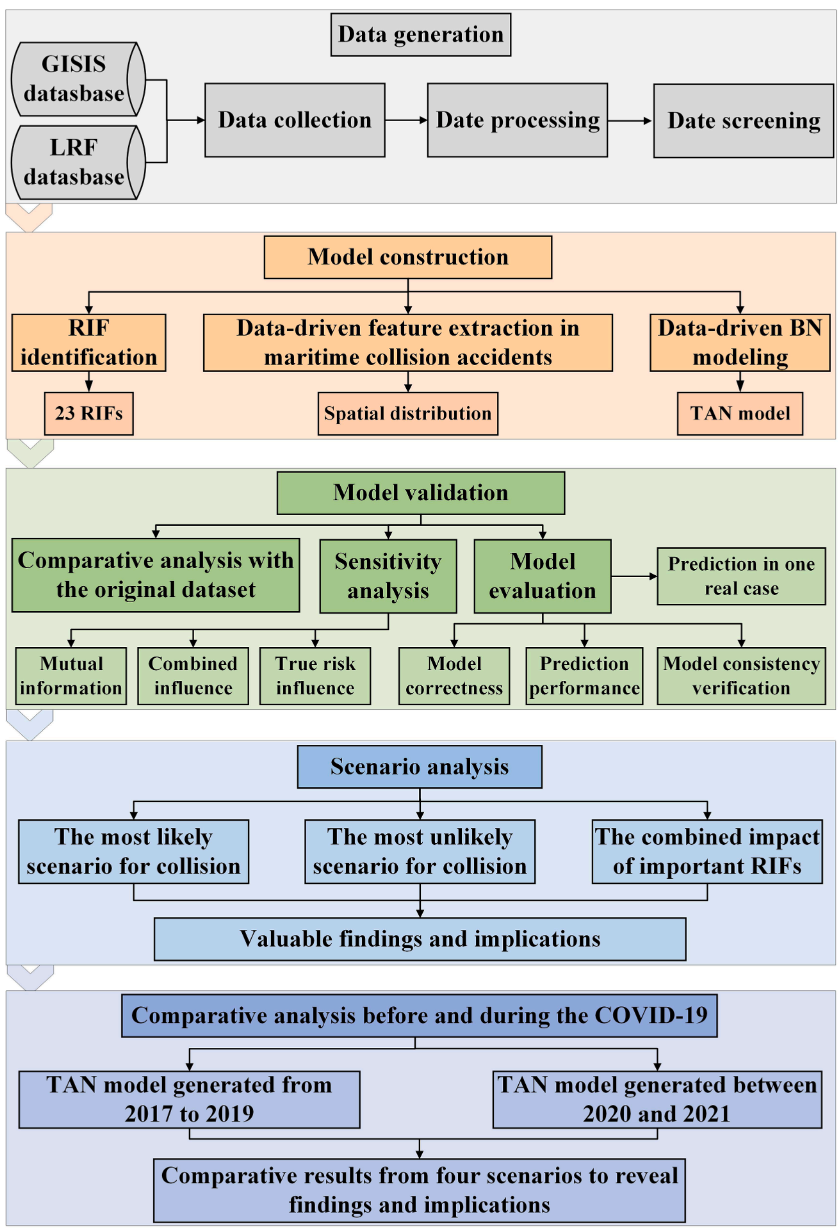

Fig. 1. The flowchart of the framework.

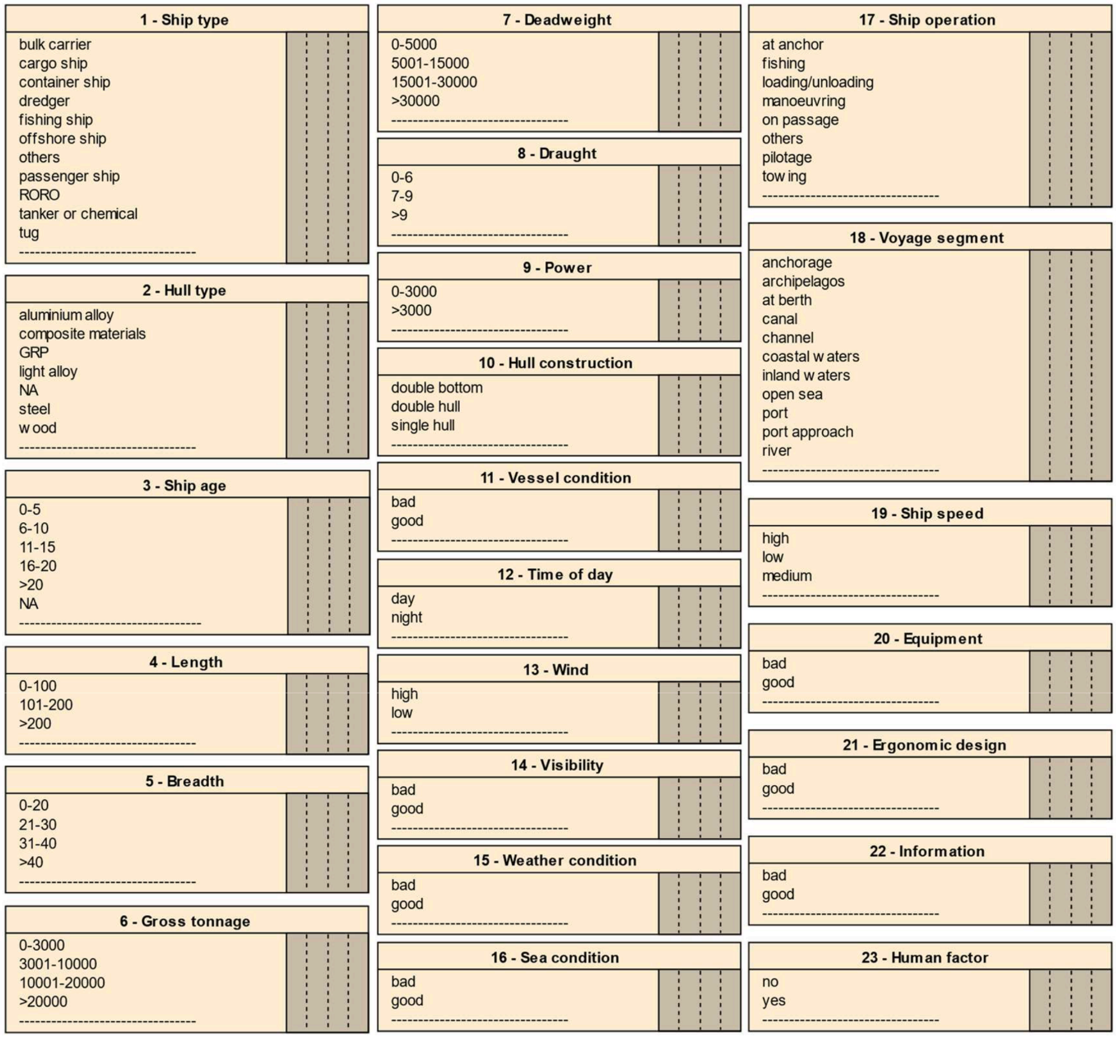

Fig. 2. RIFs and their states.

State Control (PSC) inspection data within the BN model framework to explore ship accidents and their consequences [35–38].

Despite the adaptability and versatility of BN in modeling maritime accidents, challenges such as collecting real accident datasets persist. To address these challenges, this study compiles a comprehensive database from authoritative institutions, GISIS and LRF, comprising 402 accident records for a detailed exploration using the TAN algorithm within the BN framework.

2.2. Collision accident risk analysis

Collision accidents represent a significant portion of global maritime transportation accidents, leading to extensive research efforts to understand and mitigate their occurrence and impact [1,4,34,39]. These studies can be broadly categorized into geometric collision models and collision causation models, each with its own focus and methodology [39]. Geometric collision models employ Euclidean methods to quantify Time to Closest Point Approach (TCPA) and Distance Closest Point of Approach (DCPA) between vessels and potential collision targets. Nonetheless, these methods fall short in analyzing ship collision scenarios from an accident evolution perspective. In contrast, causation collision models aim to mitigate both the likelihood and impact of collision by evaluating risks based on historical accident records.

Numerous studies have been carried out to evaluate the risk of navigational accidents regionally, aligning with the FSA framework proposed by the IMO [40–42]. For instance, the Yangtze River experiences frequent collision accidents, leading to innovative methodologies like BN analysis to address these challenges. Wu et al. [40] improved BN by using mutual information calculations to understand relationships among key factors like ship type, visibility, distress personnel count, and emergency resources. Wu et al. [10] constructed a decision-making model for real-time response strategies, considering challenges in densely populated areas near the Yangtze River. They also simulated collision consequences downstream, while Zhao et al. [42] integrated navigational accident reports at Qinzhou Port to create a comprehensive fault tree model using fuzzy set theory and BN. Aydin et al. [43] identified significant risk factors in narrow waters using BN, including maneuvering failures, collision detection risks, and communication breakdowns. Similarly, Hänninen and Kujala [44] explored ship collision likelihood using BN, monitoring causation probabilities. Trucco et al. [45] merged FTA with BN to understand collision risk factors in open sea conditions, recognizing connections between accident occurrences and organizational contexts.

In addition to studies analyzing the risk of collision accidents, some scholars have focused on the severe consequences of such accidents. For instance, Goerlandt and Montewka [46] investigated an oil spill scenario resulting from a ship collision, aiming to improve response strategies and evaluate environmental risks. Chen et al. [47] investigated the causes of collision accidents leading to oil spills, identifying factors like operational errors, lookout negligence, and vessel size. Similarly,

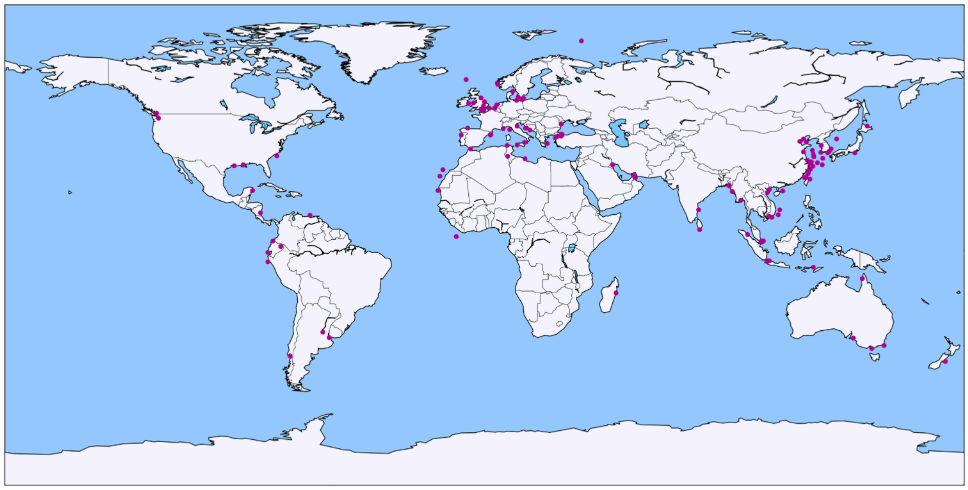

Fig. 3. The spatial distribution of global maritime collision accidents.

Montewka et al. [48] conducted a case analysis of collisions involving RoPax vessels in open waters, presenting a methodical framework for evaluating risk in maritime transportation. This approach aligns with the conventional definition of risk assessment. Moreover, Pilatis et al. [49] examined the impact of collision accidents on vessel hulls through statistical analysis of ship accidents.

The literature emphasizes the importance of collision causation models in understanding the causes associated with predefined RIFs. However, there is a notable absence of analysis regarding the temporal and spatial characteristics of collision accidents, as well as a scarcity of studies comparing collision accidents before and during the COVID-19 pandemic. While regional collision causation analysis is widespread, our research takes a global perspective. Through the application of bi-directional BN analysis, this paper aims to explore both cause-to-root and root-to-cause relationships. As a result, this paper addresses these limitations and offers valuable insights to maritime stakeholders. Findings from this study will enable preventive measures to be enacted to decrease collision accidents and alleviate their severe consequences.

2.3. Research contributions

Given the backdrop of the abovementioned review, the utilization of current accident data is crucial for investigating emerging RIFs and their contributions to global collision accidents. An up-to-date accident dataset paves the way for more robust data-driven models in assessing maritime collision risks, reducing dependence on expert judgments. The paper presents the following key contributions:

(1) Employing a comprehensive quantitative approach rooted in historical accident data, reducing uncertainties associated with expert judgments.
(2) Leveraging the strengths of the BN model using the TAN algorithm to conduct a thorough examination of the complex interplay among RIFs and facilitate root-cause analysis.
(3) Providing a comprehensive analysis aimed at reducing both the frequency of collisions and their ensuing consequences from an overall risk perspective.
(4) Comparing the characteristics of collision accidents before and during COVID-19, revealing significant findings and implications for the maritime industry and providing effective guidance.

3. Methodology

The framework proposed in this paper is illustrated in Fig. 1, encompassing data generation, model construction, model validation, discussion by scenario analysis, and comparative analysis before and during the pandemic. The framework examines the spatial distribution, as well as the risk model analysis of maritime collision accidents, uncovering significant insights and implications for a range of stakeholders.

3.1. RIF identification

In risk analysis, the factors that directly or indirectly impact safety are referred to as RIFs. They have been derived through a deep examination of the literature review and the data obtained from GISIS and LRF. This paper focuses on collision accidents using a set of 23 identical RIFs. The states of these RIFs are illustrated in Fig. 2, and their detailed definitions are available in Li et al. [1]. One notable advantage of this research lies in the detailed states of factors recognized to exert a direct impact on collision accidents. For instance, RIFs such as ship operation, voyage segment, and ship type are each divided into more than eight discrete states. This statement offers a refined comprehension of the underlying causes of collision accidents.

3.2. Dataset generation and feature extraction

3.2.1. Data aggregation

The data were initially gathered from the latest IMO GISIS accident reports spanning from 2017 to 2021. These reports provide comprehensive details such as the ship's specifications, the time and location of the accident, the sequence of events, and the surrounding environmental conditions. The detailed process is listed below.

Initially, a collection of 1105 historical accident records was compiled. Yet, a significant issue with these reports is the lack of specific static information about the ships involved.

By correlating the ships' Maritime Mobility Service Identity (MMSI) and IMO numbers, the missing details including type of ship, type of hull, age of ship, dimensions (length and breadth), gross tonnage, deadweight, and the material of the hull construction were filled in using the LRF database.

Nevertheless, some data on fishing ship accidents were found to be lacking due to their oversimplification, absence of detailed accident reports, and the removal of IMO numbers.

After a thorough process of data cleaning, missing information filling, and filtering, 402 accident records encompassing types of accidents and 23 RIFs were retained to effectively construct the BN model.

3.2.2. Spatial distribution of collision accidents

To deepen the analysis of the spatial characteristics of maritime

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0	0.0193	0.0198	0.0267	0.0125	0.0205	0.0164	0.0158	0.0084	0.0072	0.0068	0.0032	0.0109	0.0253	0.0121	0.014	0.012	0.017	0.03	0.012	0.011	0.014	0.014
2	0.0193	0	0.1991	0.2923	0.4461	0.3982	0.4613	0.5062	0.3116	0.3433	0.3348	0.0423	0.3336	0.3153	0.0527	0.033	0.045	0.084	0.241	0.103	0.082	0.057	0.087
3	0.0198	0.1991	0	0.1244	0.107	0.0785	0.1055	0.1083	0.0981	0.0837	0.0796	0.0178	0.0688	0.1318	0.0209	0.018	0.02	0.038	0.079	0.035	0.038	0.027	0.046
4	0.0267	0.2923	0.1244	0	0.1616	0.1512	0.1738	0.1871	0.1095	0.1305	0.1126	0.0507	0.1268	0.1429	0.0117	0.017	0.011	0.062	0.137	0.074	0.023	0.026	0.05
5	0.0125	0.4461	0.107	0.1616	0	0.5753	0.6991	0.6275	0.4424	0.4452	0.223	0.026	0.1109	0.102	0.0112	0.009	0.008	0.014	0.062	0.024	0.022	0.033	0.032
6	0.0205	0.3982	0.0785	0.1512	0.5753	0	0.613	0.5704	0.3521	0.3857	0.167	0.0286	0.1061	0.0998	0.0143	0.018	0.016	0.018	0.086	0.04	0.028	0.037	0.049
7	0.0164	0.4613	0.1055	0.1738	0.6991	0.613	0	0.7739	0.4581	0.5059	0.2501	0.0303	0.1159	0.1157	0.0123	0.012	0.014	0.038	0.093	0.028	0.044	0.055	0.086
8	0.0158	0.5062	0.1083	0.1871	0.6275	0.5704	0.7739	0	0.4312	0.5411	0.2184	0.0342	0.1156	0.1177	0.019	0.019	0.026	0.029	0.093	0.046	0.027	0.023	0.017
9	0.0084	0.3116	0.0981	0.1095	0.4424	0.3521	0.4581	0.4312	0	0.3458	0.1875	0.0148	0.0841	0.0875	0.0084	0.004	0.005	0.013	0.044	0.008	0.023	0.051	0.059
10	0.0072	0.3433	0.0837	0.1305	0.4452	0.3857	0.5059	0.5411	0.3458	0	0.1626	0.0232	0.0875	0.1226	0.0085	0.004	0.008	0.027	0.062	0.026	0.012	0.004	0.007
11	0.0068	0.3348	0.0796	0.1126	0.223	0.167	0.2501	0.2184	0.1875	0.1626	0	0.0187	0.0977	0.0986	0.0036	0.009	0.004	0.035	0.062	0.027	0.022	0.028	0.029
12	0.0032	0.0423	0.0178	0.0507	0.026	0.0286	0.0303	0.0342	0.0148	0.0232	0.0187	0	0.0234	0.031	0.0006	0	0.002	0.028	0.033	0.214	0.048	0.015	0.027
13	0.0109	0.3336	0.0688	0.1268	0.1109	0.1061	0.1159	0.1156	0.0841	0.0875	0.0977	0.0234	0	0.5878	0.027	0.021	0.026	0.111	0.353	0.076	0.074	0.092	0.085
14	0.0253	0.3153	0.1318	0.1429	0.102	0.0998	0.1157	0.1177	0.0875	0.1226	0.0986	0.031	0.5878	0	0.0295	0.028	0.033	0.136	0.361	0.115	0.034	0.058	0.061
15	0.0121	0.0527	0.0209	0.0117	0.0112	0.0143	0.0123	0.019	0.0084	0.0085	0.0036	0.0006	0.027	0.0295	0	0.118	0.418	0.303	0.055	0.025	0.008	0.015	0.017
16	0.0138	0.0329	0.0177	0.0166	0.0088	0.018	0.0121	0.0192	0.0043	0.0039	0.0091	0.0002	0.0209	0.0284	0.1183	0	0.175	0.072	0.029	0.027	0.008	0.014	0.019
17	0.012	0.0452	0.0203	0.0105	0.0079	0.0158	0.0143	0.0263	0.0053	0.008	0.004	0.0025	0.0258	0.0335	0.4184	0.175	0	0.252	0.044	0.019	0.009	0.009	0.026
18	0.0136	0.0416	0.0313	0.0099	0.0124	0.0158	0.0154	0.0222	0.013	0.0092	0.0088	0.0027	0.0508	0.035	0.3563	0.1	0.307	0	0.076	0.02	0.009	0.008	0.017
19	0.0024	0.1356	0.0194	0.0393	0.0478	0.0594	0.0543	0.0595	0.0472	0.0456	0.0342	0.0174	0.3624	0.2963	0.0067	0.009	0.005	0.076	0	0.029	0.023	0.036	0.039
20	0.0037	0.0136	0.0113	0.0233	0.0017	0.0167	0.0043	0.0115	0.001	0.0015	0.0045	0.22	0.0219	0.0194	0.0201	0.009	0.015	0.02	0.029	0	0.055	0.047	0.019
21	0.0017	0.0313	0.0054	0.0124	0.0088	0.0049	0.015	0.0102	0.003	0.0049	0.0066	0.0135	0.012	0.015	0.0022	0.002	0.002	0.009	0.023	0.039	0	0.027	0.029
22	0.0018	0.0274	0.0124	0.0157	0.004	0.0154	0.001	0.0046	0.0031	0.0066	0.0054	0.0031	0.0308	0.0373	0.0039	0.001	0	0.015	0.039	0.048	0.027	0	0.039
23	0.0016	0.0363	0.0185	0.0226	0.0088	0.0226	0.0168	0.0294	0.0028	0.0189	0.0091	0.0435	0.009	0.0206	0.0111	0.007	0.017	0.018	0.034	0.015	0.029	0.039	0

Fig. 4. The conditional mutual information value between each RIF.

collision accidents, latitude and longitude record information was extracted from the IMO dataset. This refined dataset facilitated a focused analysis, with the results illustrated in Fig. 3. Fig. 3 maps the spatial distribution of global maritime collision accidents over a five-year period. The key spatial features discerned from this data are summarized below.

(1) Concentration of collision accidents

There are visible clusters of accidents in certain regions. For example, there is a consistent concentration of accidents along the coastlines of East Asia, particularly around the South China Sea, Yellow Sea, and East China Sea. Other notable clusters are seen around the North Sea, the Mediterranean Sea, and along the east coast of North America.

(2) Geographical spread

Collisions are not uniformly spread out across the oceans but are concentrated along busy shipping routes, in proximity to ports, and in areas with heavy maritime traffic.

(3) Coastal waters versus open waters

The majority of collisions occur near coastlines rather than in open oceans, which suggests that the complexity of navigation near the coast and the higher density of vessels have a higher influence on ship collisions.

These features are reflected prominently in the 'voyage segment' RIF, indicating a close correlation between geographic regions and navigational waters. These findings also serve to validate the significance of the 'voyage segment' RIF.

3.3. Data-driven BN modeling

Following the comprehensive discussion in Section 2.2, BN is an effective and powerful graphical model that reveals probabilistic relationships between variables and uses rules for reasoning and learning [6]. Maritime risk analysis studies employing BN generally adhere to a sequence that encompasses data collection, the identification of variables, the acquisition of structural knowledge, model validation, and sensitivity analysis. In addition to examining the methodological framework used in the existing studies, an in-depth scenario analysis has been conducted pertaining to collision accidents involving various ship types on a global scale.

Data-driven refers to the utilization of machine learning algorithms for acquiring knowledge and constructing a BN model from a given dataset. Prominent data-driven techniques in the context of maritime risk analysis involve the K2 algorithm, Markov Chain Monte Carlo (MCMC), NBN, ABN, and TAN. TAN represents an enhancement over the NBN approach by eliminating the assumption of attribute independence, thereby accounting for interdependencies between attributes. Consequently, TAN preserves the resilience of NBN while rendering the network structure more reflective to real-world scenarios. Within the TAN network, each attribute relies on both the class variable and another attribute.

Let $X_1, \cdots, X_n$ be the attribute variables (i.e., RIFs) and C indicates the class variable (i.e., collision accident in this paper). ΠC denotes the parent variable of C. The model is defined as a TAN model if $\Pi C = \varnothing$ and there exists a function φ which defines a tree over $X_1, \cdots, X_n$ such that if $\varphi(i) > 0$, then $\Pi X_i = \{C, X\varphi(i)\}$, and if $\varphi(i) = 0$, then $\Pi X_i = \{C\}$. This optimization problem finds a tree-defining function φ on $X_1, \cdots, X_n$ which maximizes the log-likelihood, and the TAN model under this function is taken as the final BN structure.

The "Construct-TAN' program proposed by Friedman et al. [50] is applied to solve the above optimization problem. The method utilizes the conditional mutual information between attributes to help construct the TAN structure, and the calculation formula is defined by

$$I_P(X_i, X_j | C) = \sum_{x_{ii}, x_{ji}, c_i} P(x_{ii}, x_{ji}, c_i) \log \frac{P(x_{ii}, x_{ji} | c_i)}{P(x_{ii} | c_i) P(x_{ji} | c_i)} \tag{1}$$

where I_P represents the conditional mutual information, x_{ii} is the ith state of the attribute variable X_i, x_{ji} is the ith state of the attribute variable X_j, and c_i is the ith state of the class variable C.

I_P between each RIF is illustrated in Fig. 4, displaying a gradual shift in color from blue to red to simplify the intensity of their mutual relationships.

Leveraging the learning network function, the initial step in this paper involves the construction of the TAN network structure. After establishing the qualitative framework of the network, the Bayesian learning method is employed using the Netica software to learn the

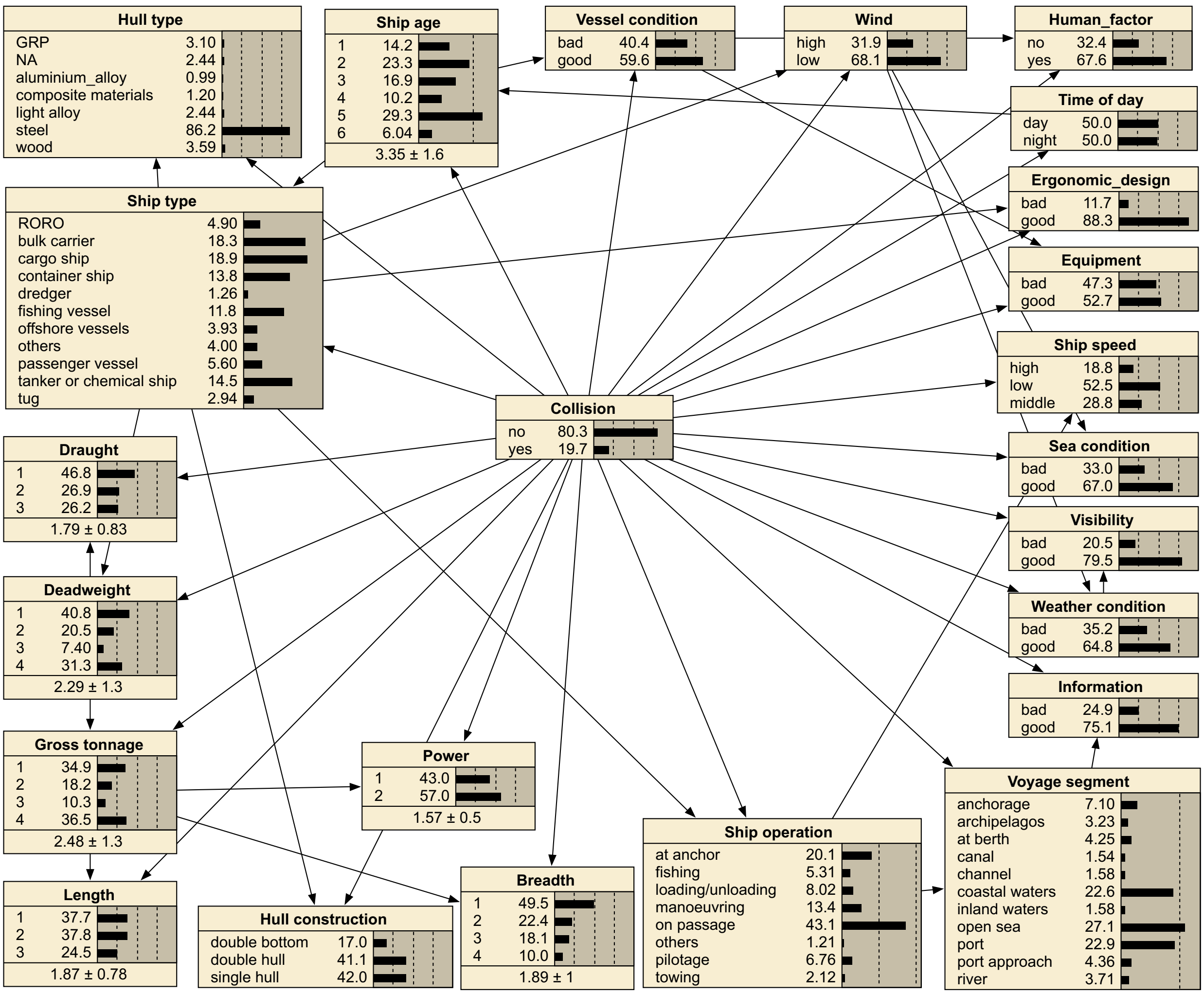

Fig. 5. The final TAN model.

parameters and establish the conditional probability distribution of the nodes. The resulting TAN model, following this learning process, is illustrated in Fig. 5.

4. Model validation

4.1. Comparative analysis using the original dataset

As an initial step, it is necessary to validate the predicted results against the historical statistical results. The analysis of historical data indicates an occurrence probability of collision accidents at 19.70 %, while the trained BN estimates it at 19.75 %. This comparison result demonstrates that the prediction accuracy of the TAN-based model when using the historical data as the test data.

4.2. Sensitivity analysis

4.2.1. Mutual information

The mutual information values between the 'Collision' target node and 23 RIFs are presented in Table 1. A higher mutual information value indicates a stronger impact of the respective RIF on the 'collision accident'. The average value is calculated to be 0.02138. RIFs with mutual information values greater than 0.02138 are deemed more important, including Ship operation, Information, Voyage segment, Ship speed, Sea condition, Wind, Ship type, and Human factor.

These factors are thoroughly examined in the following subsections to provide a comprehensive collision risk analysis.

4.2.2. Combined impact of multiple RIFs

To further investigate the detailed impact of the most important RIFs identified through mutual information calculation, additional sensitivity analysis methods are utilized to examine how these RIFs influence collision accidents. To accomplish this, the probability of each RIF's different states is incrementally increased to 100 %, generating the joint probability of each variable and "collision accident' [27]. The results of these joint probabilities are presented in Table 2. For instance, when a ship is in an anchored state, denoting its stationary mooring at a specific location, the probability of a collision occurrence exhibits a decrease from 19.67 % to 2.27 %. Conversely, when the ship is actively maneuvering, the likelihood of a collision event escalates from 19.67 % to 36.77 %. These outcomes align remarkably well with anticipated trends related to critical operational scenarios.

From the perspective of the information factor, if seafarers receive inadequate and ineffective information, the probability of a collision occurrence approximately doubles from 19.67 % to 42.98 %. Within the context of distinct voyage segments, collision accidents have a higher likelihood of occurring in canals but are less likely to happen when a vessel is berthed. For the remaining important RIFs, a comparison with the original probabilities listed in the first row of Table 2 demonstrates the alterations in collision accident probabilities when a particular RIF is associated with a specific state. This analysis highlights the states where each variable has the most significant influence and the least impact.

Table 1
Mutual information shared with 'Collision'.

Node	Mutual Information	Entropy Reduction Percent	Variance of Beliefs
Collision	0.71521	100	0.1579899
Ship operation	0.0866	12.1	0.015872
Information	0.07367	10.3	0.018041
Voyage segment	0.04173	5.83	0.0088661
Ship speed	0.03332	4.66	0.0071858
Sea condition	0.03034	4.24	0.0059462
Wind	0.02922	4.09	0.0057025
Ship type	0.02603	3.64	0.0053992
Human factor	0.02587	3.62	0.0051097
Time of day	0.02087	2.92	0.0044969
Deadweight	0.0181	2.53	0.0038502
Vessel condition	0.0166	2.32	0.0034748
Draught	0.0147	2.05	0.0031661
Gross tonnage	0.012	1.68	0.0025856
Ship age	0.01029	1.44	0.0021763
Equipment	0.00976	1.36	0.0021075
Weather condition	0.00929	1.3	0.0019448
Length	0.00842	1.18	0.0018093
Visibility	0.006	0.839	0.0013934
Breadth	0.00591	0.826	0.0013947
Hull type	0.00462	0.645	0.0009937
Hull construction	0.00298	0.416	0.000658
Power	0.00284	0.396	0.0006139
Ergonomic design	0.00247	0.345	0.0004993

While most of the findings are in good harmony with the best practice and/or results from previous studies in the field, two interesting findings involving sea conditions and wind attract further investigation. The marginal probability of the vessel operating in bad sea conditions or high wind indicates a lower probability of ship collision. Further investigation finds that when the sea and wind conditions are bad, the other influenced factors (e.g., Information, Ship Speed and Ship Operation) change towards a safer status, which jointly leads to a lower probability of collisions. It mirrors some similar findings in transportation research, suggesting that operators could develop better risk awareness and possibly take countermeasures (e.g., reducing speed) when the external environment becomes harsh (e.g., poor visibility). It also highlights the value of utilizing joint probability in BN for assessing the combined impact of multiple RIFs in risk analysis, as opposed to analyzing individual factors in isolation.

4.2.3. True risk influence

True Risk Influence (TRI) represents an innovative technique for verifying sensitivity, as introduced by Alyami et al. [51]. The concept behind TRI is to assess the hierarchy of importance among the most influential factors identified through mutual information analysis. The detailed steps are listed below:

(1) Following the TAN-based model analysis, original influence values for both 'collision occurrence - yes' and 'collision occurrence - no.' Subsequently, the joint probability of each variable and 'collision accident' was computed as each factor's different states were incrementally increased to 100 %.
(2) The High Risk Influence (HRI) value is determined by calculating the difference between the highest occurrence probability of a certain state and the original influence.
(3) Conversely, the Low Risk Influence (LRI) value is calculated by quantifying the difference between the original influence and the lowest occurrence probability of a certain state.
(4) In the final step, the TRI is calculated as follows;

$$TRI = \frac{HRI + LRI}{2} \tag{2}$$

Table 2
The joint probability (100 %).

	no	yes
original	80.33	19.67
Ship operation		
at anchor	**97.73**	**2.27**
fishing	95.65	4.35
loading/unloading	94.25	5.75
manoeuvring	**63.23**	**36.77**
on passage	72.65	27.35
others	80.85	19.15
pilotage	86.38	13.62
towing	68.94	31.06
Information		
bad	**57.02**	**42.98**
good	**88.07**	**11.93**
Voyage segment		
anchorage	88.12	11.88
archipelagos	80.79	19.21
at berth	**95.90**	**4.10**
canal	**59.60**	**40.40**
channel	88.96	11.04
coastal waters	70.38	29.62
inland waters	88.93	11.07
open sea	79.38	20.62
port	90.54	9.46
port approach	59.95	40.06
river	77.01	22.99
Ship speed		
high	**71.31**	**28.69**
low	**88.40**	**11.60**
middle	71.51	28.49
Sea condition		
bad	**91.31**	**8.69**
good	**74.92**	**25.08**
Wind		
high	**91.35**	**8.65**
low	**75.16**	**24.84**
Ship type		
RORO	87.84	12.16
bulk carrier	71.35	28.65
cargo ship	83.42	16.58
container ship	84.00	16.00
dredger	89.52	10.48
fishing vessel	87.03	12.98
offshore vessels	**96.65**	**3.35**
others	**70.32**	**29.69**
passenger vessel	85.44	14.56
tanker	71.83	28.17
tug	87.84	12.16
Human factor		
no	**90.66**	**9.34**
yes	**75.39**	**24.61**

Table 3
TRI of RIFs for 'Collision'.

	no	yes
Ship operation	17.251	17.250
Information	15.525	15.525
Voyage segment	18.151	18.151
Ship speed	8.549	8.549
Sea condition	8.197	8.197
Wind	8.098	8.098
Ship type	13.168	13.168
Human factor	7.638	7.637

Table 4
The most important RIFs for ‘Collision’.

	no	yes
Ship operation	2	2
Information	3	3
Voyage segment	1	1
Ship speed	5	5
Sea condition	6	6
Wind	7	7
Ship type	4	4
Human factor	8	8

Table 3 provides the corresponding TRI results for the influential factors, while Table 4 illustrates the hierarchy of their significance. The ordering of the most influential factors is as follows: Voyage segment, Ship operation, Information, Ship type, Ship speed, Sea condition, Wind, and Human factor.

4.3. Model evaluation

4.3.1. Model correctness verification

The results of the aforementioned sensitivity analysis can also help validate the correctness of the model. Table 5 demonstrates that increasing the prior probabilities of variable nodes leads to corresponding increases in the posterior probabilities of the target node, thus confirming Theorem 1 [52]. Furthermore, Table 5 indicates that the probability values of the target node progressively increase as the RIFs are continuously updated, showing a rising trend in the probability values from left to right. This observation validates Theorem 2. The cumulative increments of the probability values have also been provided in the lower rows of Table 5. Specifically, the increment for ‘collision occurrence – no’ does not exceed 0.714 %, while for ‘collision occurrence – yes,’ it remains below 3.401 %. Additionally, the absence of abrupt transitions between the updated values underscores the inherent robustness in the trained model. In conclusion, the model proposed in this paper satisfies both Theorem 1 and Theorem 2, thus validating the correctness and reliability of the model.

4.3.2. Evaluation of the predictive performance of the model

This paper utilizes a range of predictive performance evaluation metrics to gauge the accuracy and reliability of the constructed model's predictions. Initially, the data for training and testing the BN model is randomly selected from the new dataset of historical accidents. The model is built using the training set, and the testing set is employed for model evaluation. The overall accuracy serves as a straightforward and effective metric for gauging the prediction accuracy of the constructed model, defined as the percentage of correctly predicted instances in the total sample. However, it may not be suitable when dealing with unbalanced samples. To address these issues, precision, recall, and F-measure are chosen as measurement techniques to validate the model's reliability and robustness.

In this paper, 80 % of the data from the dataset was randomly selected for training the model, leaving the remaining 20 % (comprising 80 accident records) for use as the test dataset. The test results are presented in the form of a confusion matrix, as shown in Table 6.

The calculation formula of the overall accuracy indicator is shown in Eq. (3).

$$overallaccuracy = \frac{T_P + T_N}{T_P + F_P + T_N + F_N} \tag{3}$$

The overall accuracy of the mode is 97.5 % (78/80) from the results in the confusion matrix. Other performance indicators for collision accidents are calculated and presented in Table 7. The closer the value approaches 1, the more favorable the outcome.

Drawing from the aforementioned results, it is evident that this model displays robust predictive performance and yields dependable outcomes.

4.3.3. Model consistency verification

In this paper, Cohen's Kappa statistic is used to evaluate the model's consistency in predicting collision accidents. The kappa statistic k value closer to 1 indicates stronger consistency of the model [53], which is given as follows:

$$k = \frac{p_o - p_e}{1 - p_e} \tag{4}$$

where p_o is the overall accuracy. According to the confusion matrix, the overall accuracy p_o of the model is 0.975.

Assuming that the total number of samples in the test dataset is n. In these samples, ‘$a1$’ instances where the actual collision occurrence is ‘yes,’ and ‘$a2$’ instances where the actual collision occurrence is ‘no’ are accounted for. In the prediction results, ‘$b1$’ instances where a collision occurrence is predicted as ‘yes,’ and ‘$b2$’ instances where it is predicted as ‘n’ are recorded. Then, the definition of p_e is

$$p_e = \frac{a1 \times b1 + a2 \times b2}{n \times n} = \frac{64 \times 62 + 16 \times 8}{80 \times 80} = 0.665 \tag{5}$$

Table 6
The predicted results.

Actual Predicted	no	yes	Actual total	Accuracy rate (%)
no	62 (T_P)	2 (F_P)	64	96.875
yes	0 (F_N)	16 (T_N)	16	100
Predicted total	62	18	80	97.500

Table 7
Performance results for “Collision’.

	no	yes
Precision	1.000	0.889
Recall	0.969	1.000
F-measure	0.984	0.941

Table 5
The analysis result of minor changes in RIFs.

Human factor		+2 %	+2 %	+2 %	+2 %	+2 %	+2 %	+2 %	+2 %
Wind			+2 %	+2 %	+2 %	+2 %	+2 %	+2 %	+2 %
Sea condition				+2 %	+2 %	+2 %	+2 %	+2 %	+2 %
Ship speed					+2 %	+2 %	+2 %	+2 %	+2 %
Ship type						+2 %	+2 %	+2 %	+2 %
Information							+2 %	+2 %	+2 %
Ship operation								+2 %	+2 %
Voyage segment									+2 %
No	80.33	80.64	80.96	81.29	81.62	82.11	82.69	83.32	83.66
Increment (%)	–	0.31	0.62	0.96	1.29	1.77	2.36	2.99	3.33
Yes	19.67	19.97	20.30	20.62	20.98	21.54	22.20	22.95	23.72
Increment (%)	–	0.31	0.63	0.96	1.31	1.88	2.53	3.29	4.05

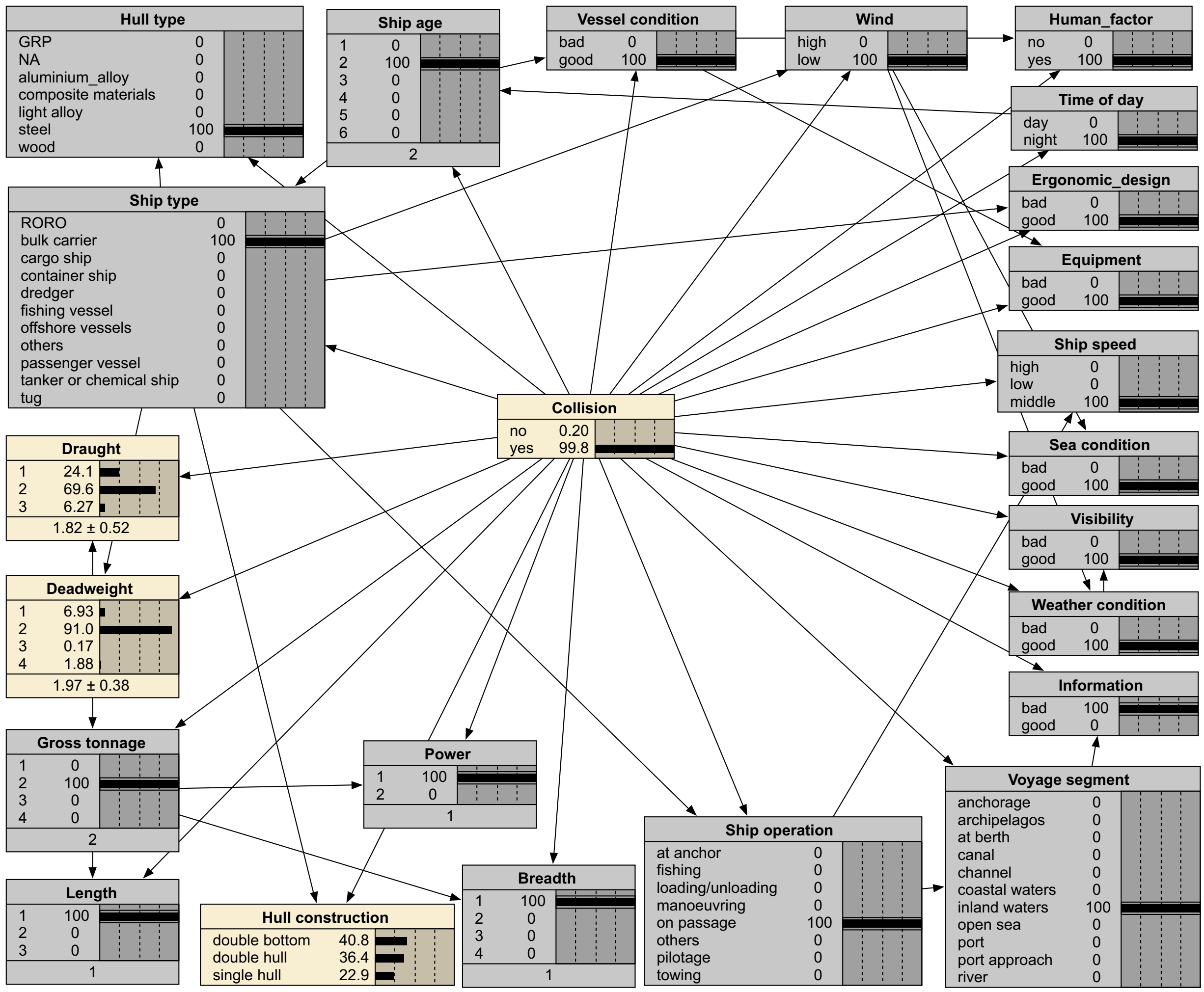

Fig. 6. The prediction result.

Therefore, the calculation of the Kappa coefficient k of the model using Eqs. (4) and (5) yields a value of 0.925, indicating strong consistency in the proposed model.

4.3.4. *Prediction performance measurement by real cases*

To provide additional evidence of the model's effectiveness, a real-world verification is conducted by using a maritime accident that occurred in 2022. This accident was not part of the training and testing datasets used in this paper, and its details are shown in Fig. 6. Despite the presence of three unknown nodes, the model predicted a remarkably high probability of a ship collision, reaching 99.8 %. This real-case analysis further confirms the robustness and reliability of the constructed risk analysis model. Moreover, the risk model presented in this paper can be used as a predictive tool, offering valuable insights and effective measures for preventing future collision accidents in the maritime domain.

5. Discussions and implications by scenario analysis and comparison analysis

Scenario analysis involves modifying the states of nodes to simulate various scenarios, enabling the exploration of the impact of specific conditions on collision accidents [54]. By conducting scenario analysis, it is possible to uncover the risks associated with specific scenarios leading to a collision occurrence. This information can effectively assist maritime authorities in formulating appropriate collision prevention strategies.

5.1. *Scenario one: The most likely scenario for collision*

The most important influential factors (depicted as grey boxes in Fig. 7) identified through mutual analysis, are selected to simulate the most probable collision scenario. Then, the most pivotal state for each factor is set to 100 %. For instance, in the event of inadequate information for the seafarer and the highest likelihood of human error leading to a potential collision accident, when the ship is navigating in a canal at a high cruising speed with low wind conditions, the probability of a collision occurrence reaches its peak, registering at 96.7 %. The specific state values within the BN model prediction, alongside the outcome for the most probable collision occurrence, are presented in Fig. 7.

5.2. *Scenario two: The most unlikely scenario for collision*

The determination of the least likely scenario is reaffirmed by taking into account the most influential factors, given their substantial influence on the absence of collision accidents. The state with the least influence for each factor, as determined through joint probability analysis, is assigned a value of 100 % in Fig. 8. Following the analysis, it becomes evident that even under high wind conditions, collision accidents are highly unlikely for offshore vessels engaged in low-speed sailing or anchoring operations at the berth. There are no instances of human error, and the available information resources are deemed sufficient to adequately support the seafarers.

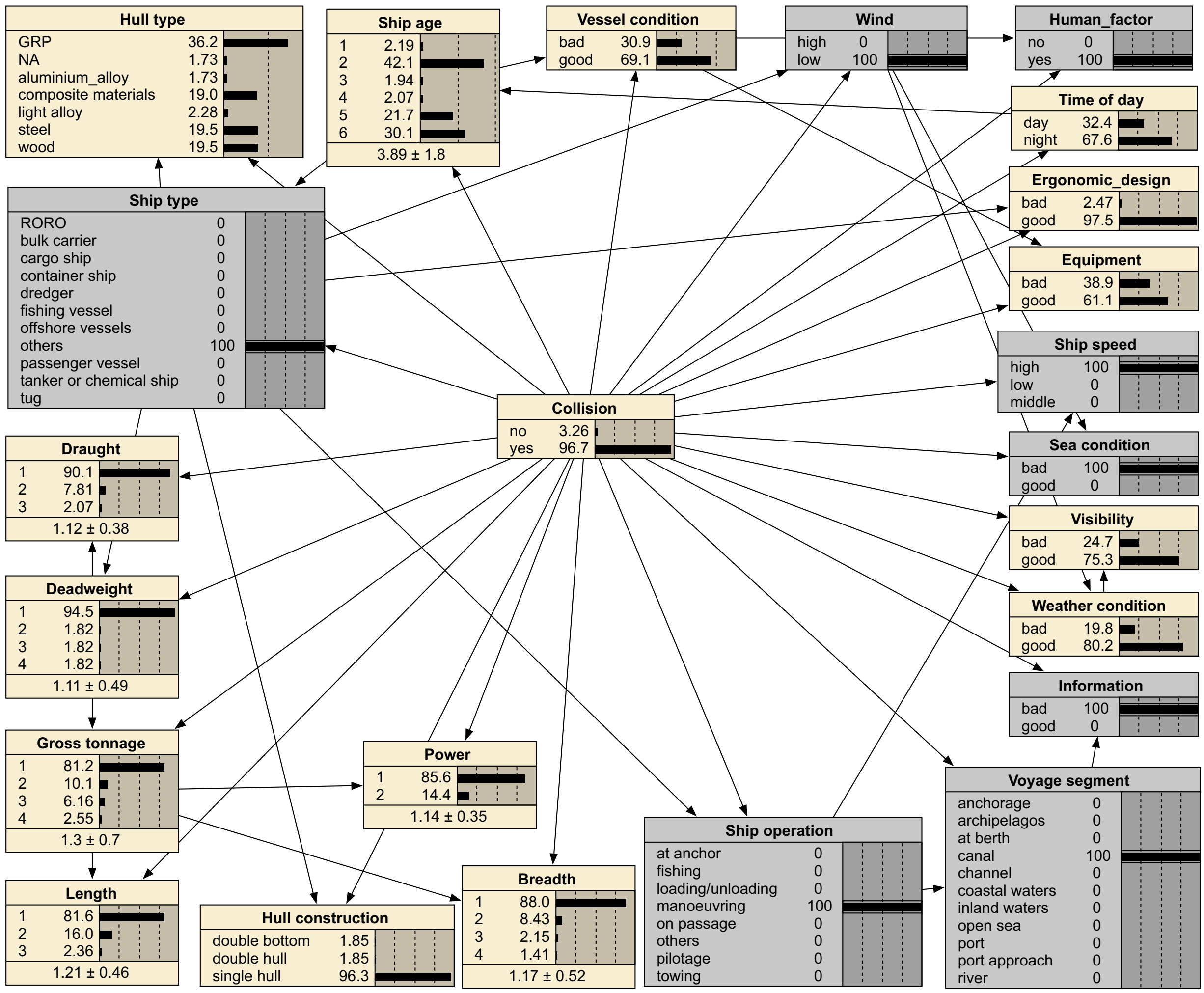

Fig. 7. The most likely scenario for collision.

5.3. Scenario three: The combined impact of RIFs

The TAN-based model offers the Most Likely Explanation (MPE) for determining collision occurrence probability and finding the most likely state of occurrence in a node. This MPE mode enables the observation of the most likely RIFs in the current scenario and updates known scenarios through manual evidence input. This approach enhances the analysis of maritime collision accidents, predicts their probability, and aids in collision prevention. The combined impact of RIFs has been further investigated to present prominent aspects of collision accidents, utilizing a relatively updated database that holds significant interest for maritime authorities, shipping companies, and seafarers. The findings of this paper have been compared with those from pertinent research conducted within the domains of maritime accidents, maritime accident severity, and collision accidents, as presented in Table 8. In contrast to our research, the limitations of these referenced studies are emphasized. For instance, Wu et al. [40] developed a model to estimate navigational accidents specifically within the Yangtze River region. However, it is noteworthy that their analysis was conducted at a regional scale, thereby lacking a comprehensive global perspective. Furthermore, their investigation focused on a total of 11 RIFs, with only 4 of them being identified as highly significant, a scope narrower than that of the present paper. Fan et al. [55] and Fan et al. [33] conducted a comprehensive maritime accident analysis employing a data-driven BN model, considering a substantial number of RIFs. Nevertheless, the analysis primarily focused on determining the most influential factors related to maritime accidents, including but not limited to collision, grounding, capsizing, and so on, rather than exclusively concentrating on collision accidents. The hierarchy of importance among different ship types within their research was extracted manually from their joint probability table. Liu et al. [4] conducted a systematic investigation into the causation of maritime collision accidents in the coastal waters of China. However, it should be noted that their study involved a smaller set of RIFs and a narrower range of ship types compared to the scope of the current research. In recent studies conducted by Antão et al. [14] and Liao et al. [2], a detailed analysis of maritime collision accidents and broader maritime accidents was undertaken to identify RIFs and their associated outcomes, respectively. It is worth noting, however, that the historical data employed in their research predates the timeframe considered in the present study. Besides, the number of RIFs considered and the identification of their most influential factors were found to be more extensive in our investigation in comparison to their work. As a kernel, this paper has undertaken an examination of the most influential factors by treating collision accidents as a target node within a BN model. Notably, the historical data utilized in the BN training process exhibits a significant degree of contemporaneity compared to previous available studies. Furthermore, this research has provided a valuable analysis of the hierarchical importance of various ship types. This analysis carries practical implications for ship owners and shipping companies seeking to assess the global collision risk associated with their vessels. Through a rigorous joint probability analysis involving the occurrence probability of collision and ship types, the study has identified a hierarchy of risk among vessel categories. Following the ship type state ‘others’, the riskiest categories of vessels in descending order of collision accident

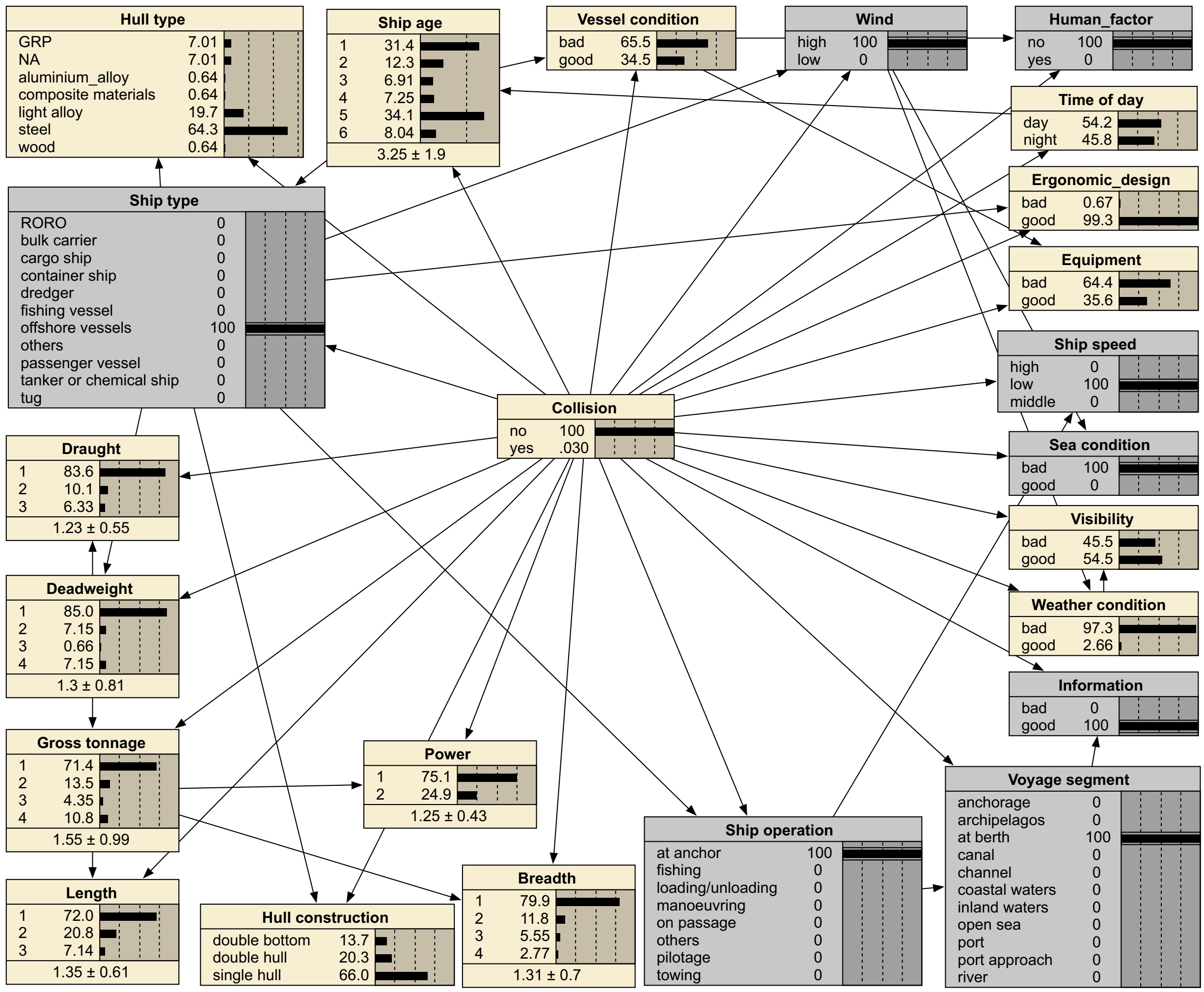

Fig. 8. The most unlikely scenario for collision.

occurrence were identified as bulk carriers, tankers, cargo vessels, container ships, passenger vessels, fishing vessels, Ro-Ro ships, tugboats, dredgers, and offshore vessels.

Besides, a comprehensive investigation has been conducted concerning joint probabilities. For instance, one of the most likely collision scenarios for tanker and bulk carrier ships involves the vessel being in the process of maneuvering within a canal, given that the probability of collision is above 95 %. The level of collision occurrence is similar for cargo ships under the same operational conditions, particularly in the high-speed region exceeding 12 knots. If the speed of cargo ships is reduced from the high-speed region to the low-speed region, the probability of collision occurrence decreases significantly, by approximately 65.5 %. In addition to ship type, a thorough investigation of the voyage segment is also conducted. As previously emphasized, a critical scenario for bulk carrier ships involves maneuvering while navigating canals and coastal waters. Interestingly, even within the open-sea voyage segment, these vessels demonstrate consistent levels of collision occurrence. Container ships typically operate within a high-speed regime, defined as exceeding 12 knots in this paper. The high-speed container ships pose a considerable challenge in maneuvering operations. Consequently, the probability of collision occurrence is further investigated across various voyage segments, including berth, canal, channel, coastal water, port approach, and notably, open sea. These segments emerge as the most critical ones among the others, with collision accident probabilities consistently exceeding 80 %. An additional analysis of collision accidents is conducted, specifically focusing on fishing vessels engaged in loading/unloading operations in coastal waters. It is noteworthy that regardless of the vessel's speed category, the collision probability consistently exceeds 97 %. This finding holds significant implications for seafarers and warrants attention.

6. Comparative analysis before and during the COVID-19 pandemic for collision

Research on collision accidents in the shipping industry before and during the COVID-19 pandemic holds significant importance on various fronts. It is however difficult to achieve due to the constraints in terms of data availability and modelling capacity. Due to the advances that this work brings both the updated accident data in 2017–2021 and the new model containing comprehensive RIFs, the comparative impact analysis of the COVID-19 pandemic (before and during) on ship collisions is conducted in this section. It provides crucial insights into the resilience of the industry amidst crisis, identifies the impact of external factors such as changes in shipping patterns or regulations, and informs the development of targeted safety measures and policies. Additionally, the analysis guides risk management strategies, contributes to global resilience efforts, and enhances preparedness for future disruptions. By understanding how collision accidents have evolved during the pandemic, stakeholders can make informed decisions about resource allocation, route planning, and safety protocols, thus ensuring the industry's continued safety and operational efficiency.

The comparative analysis results in terms of collision accidents in the shipping industry before and during the COVID-19 pandemic are presented in Table 9, with regard to between the periods 2017–2019 and

Table 8
Comparison of findings specific to collision accidents.

Refs	Region	Period of historical data	Total of RIFs	Most influential RIFs	Most risky ship types
Current Research	Global	2017–2021	23	Ship operation Information Voyage segment Ship speed Sea condition Wind Ship type Human factor	Others Bulk carrier Tanker Cargo Container Passenger Fishing Ro-Ro Tug Dredger Offshore vessel
[14]	Global	2005–2017	6	Ship type Geographic area	Cargo Container Ro-Ro
[2]	Global	2005–2020	7	Quarter of accident Ship type Gross tonnage	Bulk carrier Chemical/oil tanker Container Others General cargo Passenger Fishing
[33]	Global	2005–2021	25	Ship type Ship age Passage plan Gross tonnage Weather condition Risk assessment Information	Tanker/Chemical tanker Container Ro-Ro Others Cargo Bulk carrier Passenger Tug/barge
[4]	China coastal waters	2007–2020	19	Ship type Season Ship speed Length Ship age	General cargo Others Fishing Oil and gas Service
[52]	Global	2012–2017	25	Ship operation Voyage segment Ship age Vessel condition Information	Fishing Tug Others Container Bulk carrier Cargo Passenger Tanker/Chemical tanker Barge Ro-Ro
[39]	Yangtze River	2006–2013	11	Ship type VisibilityThe number of people in distress Emergency source used	–

2020–2021, revealing key findings.

(1) The ranking of risk factors changed during the pandemic, with visibility and ship type becoming more significant.
(2) The reduced number of data records could indicate fewer shipping movements during the pandemic, due to restricted operations or prioritization of resources.
(3) There was a dramatic increase in the probability of collisions due to ship-related factors during the pandemic, indicating that ship operations became dangerous, possibly due to crewing or maintenance challenges.
(4) Relatively speaking, the probability of collisions from environmental factors remains more stable compared to ship factors.
(5) Among all the environmental factors, bad sea conditions became more influential in collisions during the pandemic than other ones.

Some valuable implications for the shipping industry include:

(1) The rise in crucial RIFs during the pandemic indicates the emergence of new hazards, potentially stemming from modifications in maritime logistics or environmental shifts. It is imperative to update risk assessment frameworks to account for the identified risk elements of increasing importance, such as visibility and the type of vessel involved.

(2) Implementing stricter maintenance inspections and/or more efforts on old ships could serve as a preventive measure against collision accidents.

(3) The consistently high risk of collisions across different periods highlights an ongoing challenge within maritime operations, underscoring the essential need for enhanced safety protocols and risk management efforts to mitigate the risk of vessel collisions.

(4) The categorization of "fishing type" as a vessel category and "maneuvering" as an operation during the pandemic period points to an increase in risk associated with fishing vessels and their maneuvers, likely due to shifts in fishing practices or more frequent interactions with other ships.

(5) Policies regarding coastal navigation should be examined with the aim of implementing tighter regulations or improving navigational aids to ensure safer operations.

7. Conclusions

This paper contributes significantly to the understanding of collision

Table 9
Comparison of findings specific to collision accidents before and during the COVID-19.

	2017–2019	2020–2021
RIFs	24 RIFs in total, of which 4 are important RIFs	24 RIFs in total, of which 5 are important RIFs
Data records	337	65
Scenario one: The combined impact of important RIFs	Important RIFs setting: (1) ship operation: at anchor (2) information: bad (3) voyage segment: open sea (4) ship speed: middle The state probabilities of collision: (1) no: 1.0 % (2) yes: 99.0 %	Important RIFs setting: (1) visibility: low (2) information: bad (3) sea condition: good (4) ship type: fishing type (5) ship operation: manoeuvring The state probabilities of collision: (1) no: 0.19 % (2) yes: 99.8 %
Scenario two: The combined impact of ship-related factors	The state probabilities of collision: (1) no: 99.5 % (2) yes: 0.49 %	The state probabilities of collision: (1) no: 55.0 % (2) yes: 45.0 %
Scenario three: The combined impact of environment-related factors	The state probabilities of collision: (1) no: 89.9 % (2) yes: 10.1 %	The state probabilities of collision: (1) no: 97.9 % (2) yes: 2.09 %
Scenario four: The most likely scenario for collision (yes=100 %)	(1) Type of casualty (very serious): 56.0 % (2) Ship type (bulk carrier): 25.1 % (3) Ship operation (on passage): 57.2 % (4) Ship age [6, 10]: 33.4 % (5) Voyage segment (coastal waters): 35.3 % (6) Sea condition (good): 81.0 % (7) Weather condition (good): 81.1 %	(1) Type of casualty (serious): 53.4 % (2) Ship type (bulk carrier): 28.3 % (3) Ship operation (on passage): 66.6 % (4) Ship age (>20): 39.9 % (5) Voyage segment (coastal waters): 31.6 % (6) Sea condition (good): 98.5 % (7) Weather condition (bad): 51.5 %

accidents in global maritime transport. By identifying the most influential factors and assessing their impact, it offers valuable information to maritime stakeholders, ultimately contributing to safer and more efficient maritime operations regarding collision accidents. The findings of this research have the potential to improve safety measures, reducing the occurrence frequency of collision accidents in the maritime domain.

(1) Model validation and sensitivity analysis demonstrated the accuracy and reliability of the constructed model, identifying key RIFs, such as Ship operation, Information, Voyage segment, Ship speed, Sea condition, Wind, Ship type, and Human factor. TRI analysis further ranked these factors, highlighting their relative importance.
(2) Bulk carriers, tankers, cargo vessels, container ships, passenger vessels, fishing vessels, Ro-Ro ships, tugboats, dredgers, and offshore vessels presented the highest levels of risk in terms of collision accidents, sorted in descending order of occurrence.
(3) Typically, canals are considered the most hazardous voyage segment for collision accidents in comparison to other voyage states. Furthermore, among the states of the ship operation, ship maneuvering exerts a significant influence on collision accidents in almost all scenarios.
(4) When engaged in loading and unloading operations in coastal waters, fishing vessels are notably susceptible to collision accidents.
(5) Container ships predominantly operate within high-speed regions, and the collision occurrence probability at maneuvering operation is predicted to be approximately 89.6 %, even in the open sea.
(6) During the pandemic, the rise in collision probabilities, particularly involving older vessels and bulk carriers, implies heightened operational challenges or maintenance issues for these ship types.
(7) The prominence of favorable and adverse sea conditions in collision reports underscores the significant influence of weather on accidents during the pandemic.

The data-driven BN model developed in this paper provides a macro-scale assessment of collision accident risks by leveraging global historical accident data. However, the model's global focus on overarching themes means that it reveals a limitation in addressing local density details, which is compounded by the absence of density data in the IMO dataset and the current unavailability of global AIS data. These constraints highlight areas for future investigation and improvement.

The integration of accident data with AIS data presents a promising opportunity for advanced maritime accident analysis in real time. Moreover, there is potential for further exploration of machine learning methods combined with BN to investigate maritime accident risk analysis in future research endeavors.

CRediT authorship contribution statement

Huanhuan Li: Writing – review & editing, Writing – original draft, Visualization, Validation, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Supervision. **Cihad Çelik:** Writing – review & editing, Writing – original draft, Visualization, Validation, Formal analysis, Investigation, Resources, Software. **Musa Bashir:** Writing – review & editing, Supervision. **Lu Zou:** Writing – review & editing, Supervision. **Zaili Yang:** Methodology, Funding acquisition, Conceptualization, Project administration, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests

Zaili Yang reports financial support was provided by European Research Council. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by a European Research Council project (TRUST CoG 2019 864724).

References

[1] Li H, Ren X, Yang Z. Data-driven Bayesian network for risk analysis of global maritime accidents. Reliab Eng Syst Saf 2023;230:108938.
[2] 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. https://doi.org/10.3390/jmse11081513.
[3] Accident Investigation - Annual Overview - EMSA - European Maritime Safety Agency. Annu Overv Mar Casualties Incid 2019;2019.
[4] Liu K, Yu Q, Yuan Z, Yang Z, Shu Y. A systematic analysis for maritime accidents causation in Chinese coastal waters using machine learning approaches. Ocean Coast Manag 2021. https://doi.org/10.1016/j.ocecoaman.2021.105859.
[5] Weng J, Du J, Shi K, Liao S. Effects of ship domain shapes on ship collision risk estimates considering collision frequency and severity. Ocean Eng 2023;283:115070.
[6] Kamal B, Çakır E. Data-driven Bayes approach on marine accidents occurring in Istanbul strait. Appl Ocean Res 2022. https://doi.org/10.1016/j.apor.2022.103180.
[7] Kandel R, Baroud H. A data-driven risk assessment of Arctic maritime incidents: Using machine learning to predict incident types and identify risk factors. Reliab Eng Syst Saf 2024;243:109779.
[8] 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.
[9] Vojković L, Kuzmanić Skelin A, Mohovic D, Zec D. The development of a Bayesian network framework with model validation for maritime accident risk factor assessment. Appl Sci 2021;11:10866.
[10] Wu B, Zhao C, Yip TL, Jiang D. A novel emergency decision-making model for collision accidents in the Yangtze River. Ocean Eng 2021. https://doi.org/10.1016/j.oceaneng.2021.108622.
[11] Goerlandt F, Montewka J, Kuzmin V, Kujala P. A risk-informed ship collision alert system: Framework and application. Saf Sci 2015:77. https://doi.org/10.1016/j.ssci.2015.03.015.
[12] Ozturk U, Cicek K. Individual collision risk assessment in ship navigation: A systematic literature review. Ocean Eng 2019:180. https://doi.org/10.1016/j.oceaneng.2019.03.042.
[13] Chen P, Huang Y, Mou J, van Gelder PHAJM. Probabilistic risk analysis for ship-ship collision. State-of-the-art. Saf Sci 2019;117:108–22. https://doi.org/10.1016/j.ssci.2019.04.014.
[14] Antão P, Sun S, Teixeira AP, Soares CG. Quantitative assessment of ship collision risk influencing factors from worldwide accident and fleet data. Reliab Eng Syst Saf 2023;234:109166.
[15] Fu S, Zhang Y, Zhang M, Han B, Wu Z. An object-oriented Bayesian network model for the quantitative risk assessment of navigational accidents in ice-covered Arctic waters. Reliab Eng Syst Saf 2023:238. https://doi.org/10.1016/j.ress.2023.109459.
[16] Pauksztat B, Andrei DM, Grech MR. Effects of the COVID-19 pandemic on the mental health of seafarers: A comparison using matched samples. Saf Sci 2022:146. https://doi.org/10.1016/j.ssci.2021.105542.
[17] Zhao Z, Tang L. The impact of COVID-19 on maritime pilots: Evidence and lessons. Mar Policy 2023:153. https://doi.org/10.1016/j.marpol.2023.105664.
[18] Brooks SK, Greenberg N. Mental health and wellbeing of seafaring personnel during COVID-19: Scoping review. J Occup Health 2022;64. https://doi.org/10.1002/1348-9585.12361.
[19] Salihoglu E, Bal Beşikçi E. The use of Functional Resonance Analysis Method (FRAM) in a maritime accident: A case study of Prestige. Ocean Eng 2021. https://doi.org/10.1016/j.oceaneng.2020.108223.
[20] Qiao W, Liu Y, Ma X, Liu Y. Human factors analysis for maritime accidents based on a dynamic fuzzy Bayesian network. Risk Anal 2020;40:957–80.
[21] Uğurlu Ö, Yıldız S, Loughney S, Wang J, Kuntchulia S, Sharabidze I. Analyzing Collision, Grounding, and Sinking Accidents Occurring in the Black Sea Utilizing HFACS and Bayesian Networks. Risk Anal 2020. https://doi.org/10.1111/risa.13568.
[22] Li KX, Yin J, Bang HS, Yang Z, Wang J. Bayesian network with quantitative input for maritime risk analysis. Transp Transp Sci 2014;10:89–118.
[23] Zhang D, Han Z, Zhang K, Zhang J, Zhang M, Zhang F. Use of Hybrid Causal Logic Method for Preliminary Hazard Analysis of Maritime Autonomous Surface Ships. J Mar Sci Eng 2022. https://doi.org/10.3390/jmse10060725.
[24] Yu Q, Teixeira ÂP, Liu K, Rong H, Soares CG. An integrated dynamic ship risk model based on Bayesian Networks and Evidential Reasoning. Reliab Eng Syst Saf 2021;216:107993.
[25] Zhang J, Teixeira ÂP, Guedes Soares C, Yan X, Liu K. Maritime Transportation Risk Assessment of Tianjin Port with Bayesian Belief Networks. Risk Anal 2016. https://doi.org/10.1111/risa.12519.
[26] 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. https://doi.org/10.1177/1748006X19900850.
[27] Wang L, Yang Z. Bayesian network modelling and analysis of accident severity in waterborne transportation: A case study in China. Reliab Eng Syst Saf 2018. https://doi.org/10.1016/j.ress.2018.07.021.
[28] Cao Y, Wang X, Wang Y, Fan S, Wang H, Yang Z, et al. Analysis of factors affecting the severity of marine accidents using a data-driven Bayesian network. Ocean Eng 2023;269:113563.
[29] Zhou K, Xing W, Wang J, Li H, Yang Z. A data-driven risk model for maritime casualty analysis: A global perspective. Reliab Eng Syst Saf 2024;244:109925.
[30] Zhao X, Yuan H, Yu Q. Autonomous vessels in the Yangtze river: A study on the maritime accidents using data-driven bayesian networks. Sustainability 2021;13:9985.
[31] 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. https://doi.org/10.1016/j.oceaneng.2020.107544.
[32] Jiang M, Lu J, Yang Z, Li J. Risk analysis of maritime accidents along the main route of the Maritime Silk Road: a Bayesian network approach. Marit Policy Manag 2020. https://doi.org/10.1080/03088839.2020.1730010.
[33] 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.
[34] Baksh A-A, Abbassi R, Garaniya V, Khan F. Marine transportation risk assessment using Bayesian Network: Application to Arctic waters. Ocean Eng 2018;159:422–36.
[35] Fan L, Wang M, Yin J. The impacts of risk level based on PSC inspection deficiencies on ship accident consequences. Res Transp Bus Manag 2019;33:100464.
[36] Fan L, Zhang Z, Yin J, Wang X. The efficiency improvement of port state control based on ship accident Bayesian networks. Proc Inst Mech Eng Part O J Risk Reliab 2019;233:71–83.
[37] Yang Z, Yang Z, Yin J. Realising advanced risk-based port state control inspection using data-driven Bayesian networks. Transp Res Part Policy Pract 2018:110. https://doi.org/10.1016/j.tra.2018.01.033.
[38] Dinis D, Teixeira AP, Soares CG. Probabilistic approach for characterising the static risk of ships using Bayesian networks. Reliab Eng Syst Saf 2020;203:107073.
[39] Wu B, Yip TL, Yan X, Mao Z. A mutual information-based Bayesian network model for consequence estimation of navigational accidents in the Yangtze River. J Navig 2020;73:559–80.
[40] Wu B, Tian H, Yan X, Guedes Soares C. A probabilistic consequence estimation model for collision accidents in the downstream of Yangtze River using Bayesian Networks. Proc Inst Mech Eng Part O J Risk Reliab 2020;234:422–36.
[41] 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.
[42] Zhao C, Yip TL, Wu B, Lyu J. Use of fuzzy fault tree analysis and Bayesian network for occurrence likelihood estimation of navigational accidents in the Qinzhou Port. Ocean Eng 2022;263:112381.
[43] Aydin M, Akyuz E, Turan O, Arslan O. Validation of risk analysis for ship collision in narrow waters by using fuzzy Bayesian networks approach. Ocean Eng 2021;231:108973.
[44] Hänninen M, Kujala P. Influences of variables on ship collision probability in a Bayesian belief network model. Reliab Eng Syst Saf 2012;102:27–40.
[45] Trucco P, Cagno E, Ruggeri F, Grande O. A Bayesian Belief Network modelling of organisational factors in risk analysis: A case study in maritime transportation. Reliab Eng Syst Saf 2008;93:845–56.
[46] Goerlandt F, Montewka J. A framework for risk analysis of maritime transportation systems: A case study for oil spill from tankers in a ship–ship collision. Saf Sci 2015;76:42–66.
[47] Chen J, Chen H, Shi J, Yan T, Gu M, Huang X. Factor diagnosis and governance strategies of ship oil spill accidents based on formal concept analysis. Mar Pollut Bull 2023;196:115606.
[48] 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.
[49] Pilatis AN, Pagonis D-N, Serris M, Peppa S, Kaltsas G. A Statistical Analysis of Ship Accidents (1990–2020) Focusing on Collision, Grounding, Hull Failure, and Resulting Hull Damage. J Mar Sci Eng 2024;12:122. https://doi.org/10.3390/jmse12010122.
[50] Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. Mach Learn 1997;29:131–63.
[51] Alyami H, Yang Z, Riahi R, Bonsall S, Wang J. Advanced uncertainty modelling for container port risk analysis. Accid Anal Prev 2019;123:411–21.
[52] Li H, Zhou K, Zhang C, Bashir M, Yang Z. Dynamic evolution of maritime accidents: Comparative analysis through data-driven Bayesian Networks. Ocean Eng 2024;303:117736. https://doi.org/10.1016/j.oceaneng.2024.117736.
[53] Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76:378. https://doi.org/10.1037/h0031619.
[54] 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.
[55] Cao Y, Wang X, Yang Z, Wang J, Wang H, Liu Z. Research in marine accidents: A bibliometric analysis, systematic review and future directions. Ocean Eng 2023;284:115048.