

Article

Not peer-reviewed version

Data-driven Analysis on the Causal Chain of Waterborne Traffic Accidents: a hybrid framework based on the improved HFACS and Bayesian Network

Xiangyu Yin , Yan Yan , Jiahao Wang , Hongzhuan Zhao , Qingyan Wu , [Qi Xu](#) *

Posted Date: 28 January 2025

doi: 10.20944/preprints202501.2098.v1

Keywords: Waterborne Traffic Accidents; Causal Chain Analysis; HFACS; Bayesian Network; Data-driven Analysis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Data-Driven Analysis on the Causal Chain of Waterborne Traffic Accidents: A Hybrid Framework Based on the Improved HFACS and Bayesian Network

Xiangyu Yin ¹, Yan Yan ², Jiahao Wang ³, Hongzhuan Zhao ⁴, Qingyan Wu ⁵ and Qi Xu ^{6,*}

- ¹ China Waterborne Transport Research Institute, Beijing, 100088, PR China; yxy730622@163.com
- ² Guangxi Key Laboratory of ITS (Guilin University of Electronic Technology), Guilin, 541004, PR China; 1085157124@qq.com
- ³ Guangxi Key Laboratory of ITS (Guilin University of Electronic Technology), Guilin, 541004, PR China; 19106444843@163.com
- ⁴ School of Architecture and Transportation Engineering, Guilin University of Electronic Technology, Guilin, 541004, PR China; zhaohongzhuan@guet.edu.cn
- ⁵ Department of Infrastructure, Guilin University of Electronic Technology, Guilin, 541004, PR China; peterwuqy@163.com
- ⁶ School of Architecture and Transportation Engineering, Guilin University of Electronic Technology, Guilin, 541004, PR China; xuqi@guet.edu.cn
- * Correspondence: xuqi@guet.edu.cn

Abstract: In the context of economic globalization, waterborne transportation plays an important role in international trade and logistics. However, the occurrence of waterborne traffic accidents poses a severe threat to life, property safety, and the environment. To gain a deeper understanding of the causal mechanisms behind waterborne traffic accidents, this study conducts a data-driven analysis on the causal chain of waterborne traffic accidents. By constructing a hybrid framework integrating the improved HFACS (Human Factors Analysis and Classification System) with Bayesian Network model, the study conducts a multi-dimensional analysis of accident causes. The constructed model is quantitatively analyzed using genie software by the accident samples collected from China MSA. Results indicate that there are 12, 3, 6, 2, 4, and 7 causal chains leading to collisions, contacts, fire/explosion, windstorms, sinking, and other types of accidents respectively. The research results can provide a reference for the enhancement on the safe operation of waterborne transportation.

Keywords: waterborne traffic accidents; causal chain analysis; HFACS; bayesian network; data-driven analysis

1. Introduction

With the acceleration of economic globalization and the expansion of global trade, waterborne transportation has assumed an irreplaceably important role. It has significant advantages such as large carrying capacity, low costs, and capability of long-haul high transport efficiency. Nevertheless, the waterborne transportation system is often exposed to accidents, particularly since the 1990s, when the changes in marine shipping and the international environment have sharply highlighted the importance of shipping safety [1]. These accidents pose significant threats to people's lives, property, and the environment [2].

Waterborne traffic accidents are not caused by a single factor but the result of the interaction of multiple factors. Therefore, it is especially important to analyze the causal chain of waterborne traffic accidents, so as to reduce the possibility of these accidents. The HFACS (Human Factors Analysis

and Classification System) model, as an effective system for analyzing and classifying human factors contributing to accidents, can help to identify and categorize human factors that lead to accidents. Besides, HFACS can integrate various methods to explore the causal relationships among factors leading to accidents [3]. As a probabilistic graphical model, a Bayesian network can learn the probabilistic dependencies between variables from statistical data. By constructing a Bayesian network model, one can identify and analyze the influence of various factors in accidents, providing a new method for accident risk analysis and prediction [4]. Therefore, constructing a hybrid framework combining HFACS and Bayesian Networks, especially under a data-based setting, is a feasible and useful tool to comprehensively understand the mechanisms of accident occurrence, and to provide scientific basis for accident prevention and quick-response.

2. Literature Review

2.1. Identification on Causality of Waterborne Accidents

Within the academic field, research on the causality of waterborne traffic accidents has made remarkable progress. These studies, through the comprehensive application of various methods and theories, have deeply analyzed the root causes of accidents. Wang et al. [5] conducted an in-depth analysis of multiple influencing factors affecting the severity of marine accidents, including human, vessel, environmental, and management factors. They used global accident data and statistical models to reveal the independent and interactive effects of these factors on accident severity. Liu et al. [6] employed machine learning techniques to comprehensively analyze the causes of marine accidents in China's coastal waters. They utilized algorithms including decision trees, random forests, and neural networks to uncover the complex relationships between human, vessel, environmental, and management factors and accident severity. Sun et al.[7] applied the complex network theory to the analyze the causes of waterborne traffic accidents. They constructed a causal network model, analyzed its characteristics, and identified critical causal factors such as human negligence and management loopholes. Wu et al.[8] analyzed ship collision accidents in the Yangtze River from 2013 to 2017 through text mining techniques, identifying 33 causal features covering human, vessel, environmental, and management factors. They used a Bayesian Network model and the text mining methods to predict marine risks, and discovered that human factors were the primary caused. Using GIS technology, Wang et al.[9] explored the spatial distribution patterns of global maritime accidents, including accident frequency and severity through density analysis and clustering analysis methods.

The HFACS model, an effective tool in the field of aviation safety, has been introduced into the waterborne transportation field by many scholars. Kaptan et al.[10] applied the HFACS to the maritime sector to uncover the root causes and causal chains of waterborne traffic accidents. Yildiz et al.[11] applied the HFACS-PV method to explore HOFs in maritime accidents. Their analysis revealed its applicability to collisions, groundings, sinkings, and other accident types. Wang et al.[12] adopted the HFACS-FCMs model to investigate human factors in ship grounding accidents. They found that inadequate ship safety management organizations is the most significant factors on grounding accidents, followed by organizational influences and unsafe behavior antecedents. Yıldırım et al.[13] utilized HFACS and the statistical methods to analyze human factors in collisions and groundings. GIS and FTA of Black Sea accidents revealed COLREG violations and communication lapses as primary collision causes, while grounding accidents stemmed from watchkeeper errors and inadequate bridge resource management communication. Huang [14] utilized the HFACS framework combined with expert scoring methods and grey theory to analyze human factors in marine traffic accidents. The HFACS-MTAI system achieved precise identification, systematic classification, and quantitative ranking of human factors. Chen et al.[15] analyzed the causes of marine traffic accidents using the HFACS-MA framework, to comprehensively assess the human and organizational factors. Through case studies and public accident reports, they validated the framework's effectiveness and identified key causal factors. Chauvin et al. [16] improved the

HFACS model with the classification tree method, to analyze the causal factors at each level of the collision accidents.

The Bayesian Network model has also been widely applied in analyzing the causes of waterborne traffic accidents. Fan et al. [17] employed a data-driven Bayesian Network model to focus on human factors, exploring their impact on the probability and outcomes of various waterborne traffic accidents when interacting with non-human factors. Antão et al.[18] utilized Bayesian Belief Network (BBN) models to investigate the impact of human error on coastal shipping accidents under diverse sea conditions. Their findings reveal a significant influence of human error and variations in risk acceptance and perception among crew of different vessel types. Wang et al.[19] used a Bayesian Network model to investigate the severity of global waterborne transportation accidents, to reveal the complex relationships between weather, ship characteristics, human factors, and accident severity. Meng et al. [20] adopted a data-driven Bayesian network model integrated with physical knowledge to explore risk factors influencing ship collision accidents. The research results demonstrate the model's effectiveness in identifying and analyzing key risk factors of ship collisions. Fan et al.[21] employed a Bayesian network approach to explore risk factors influencing maritime transportation accidents. They constructed a network model reflecting the interdependencies among the influencing factors. Tian et al. [22] utilized a Bayesian Network model to investigate ship collision accidents recorded by the Zhejiang Maritime Safety Administration of China. Their research revealed that human factors such as improper lookout, inadequate collision assessment, and improper collision avoidance measures are primary causes of accidents, and predicting accident probabilities. Hänninen et al.[23] conducted a statistical analysis of ship collisions in the Gulf of Finland and utilized the Bayesian Network to investigate the causal relationships between human factors and the final collision outcomes. Meng et al.[24] employed a combined N-K model and Bayesian Networks to assess coupled risks in Chinese ship collisions. They found that multifactor coupling was more effective than bifactor coupling, with human and management factors being vital, and ship/environment factors' impact becoming more significant with varying probabilities.

Wang et al.[25] integrated HFACS with BN to investigate marine accident root causes, emphasizing the importance of human and organizational factors. Results showed that combining HFACS with BN can effectively uncover complex human factor chains and organizational deficiencies in marine accidents. Rostamabadi et al.[26] developed an FBN-HFACS model to analyze human and organizational factors (HOFs) in process accidents. The model offers a robust tool for accident prevention and safety management. Jiang et al. [27] applied the HFACS model, Bayesian Network model, and path analysis model to analyze the causal pathways of waterborne traffic accidents. Their research revealed the key causal pathways for accidents such as collisions, sinkings, and contacts. Li et al. [28] studied the impact of HOFs on ship collision accidents using HFACS and the Bayesian Network model. The HFACS model identified multi-level causes of accidents, while the Bayesian Network revealed the probabilistic dependencies between causes and identified critical causal pathways. The study revealed that crew operational errors and equipment failures were the primary causes of ship collision accidents. Wang et al.[29] employed the HFACS-BN model to explore HOFs in collisions between merchant ships and fishing vessels. The study revealed the interactions among various human and organizational factors in collision accidents and their degrees of influence on accident occurrence. Özkan et al. [30] analyzed nearly two decades of ship accidents in the Black Sea region using HFACS and the Bayesian Network model. Their study found that frequent accidents in the coastal areas of the Black Sea were primarily due to crew operational errors, equipment failures, and inadequate management.

2.2. Research Gap Analysis

Scholars have conducted extensive research in the field of causation analysis for waterborne traffic accidents. Nevertheless, despite significant progress in this field, there still exists opportunity for improvement. Firstly, current research is mainly based on data from a specific region or a single type of waterborne traffic accident, and the applicability of the method under different situation

needs further validation. Furthermore, waterborne traffic accidents often involve the intricate interplay of multiple complex factors, and current research efforts in deciphering these interactions are still inadequate. Lastly, the tendency towards single-framework methodologies in existing research has somewhat limited the breadth and depth of inquiry.

Therefore, this study conducts a data-driven analysis on the causal chains of waterborne traffic accidents which can be more general to different situations. Firstly, a hybrid framework integrating improved HFACS and Bayesian Network model is constructed to conduct a multi-dimensional analysis of accident causes. Secondly, based on the constructed framework, the study analyses the underlying causes from multiple dimensions and the chain analysis of the causal factors, quantitatively assessing the correlations among factors and their degrees of influence on accident occurrence. Thirdly, the study identifies the global causal chain of waterborne traffic accidents for different types of accidents, which can provide general reference for practical waterborne traffic accident prevention.

3. Data-Driven Hybrid Framework Based on the HFACS and Bayesian Network

3.1. Basic Data-Driven Analysis Framework

In this paper, we propose a data-driven hybrid framework to analyze the causal chain of waterborne traffic accidents, to get a better understanding of the interaction mechanisms between possible causal factors in the waterborne transportation system. The basic framework is shown in Figure 1. As Figure 1 shows, the main procedure includes the following steps.

Step 1: Initialize the accident data set. For all the accident samples collected at the beginning of the procedure, add them into the accident data set one by one.

Step 2: Calculate the number of occurrences for causal factors. Therein, the number of occurrences for causal factors is calculated according to the novel MTAACS (Maritime Traffic Accident Analysis and Classification System, an improved Classification method based on HFACS) proposed in this paper.

Step 3: For every causal factor pair, conduct the correlation analysis. The analysis is conducted by the Chi-Square Test. If the causal factor pair passes the test, and the pair is the last factor pair, then go to **Step 4**; otherwise, test the next causal factor pair.

Step 4: Construction and application of Bayesian Networks. This step includes five steps, including calculation of the conditional probability table, identification of key factors, sensitivity analysis of key causes, cause path identification, and global cause chain analysis.

Step 5: Updating the accident data set and return to Step 2 if there is any new accident data; if not, output the analysis result.

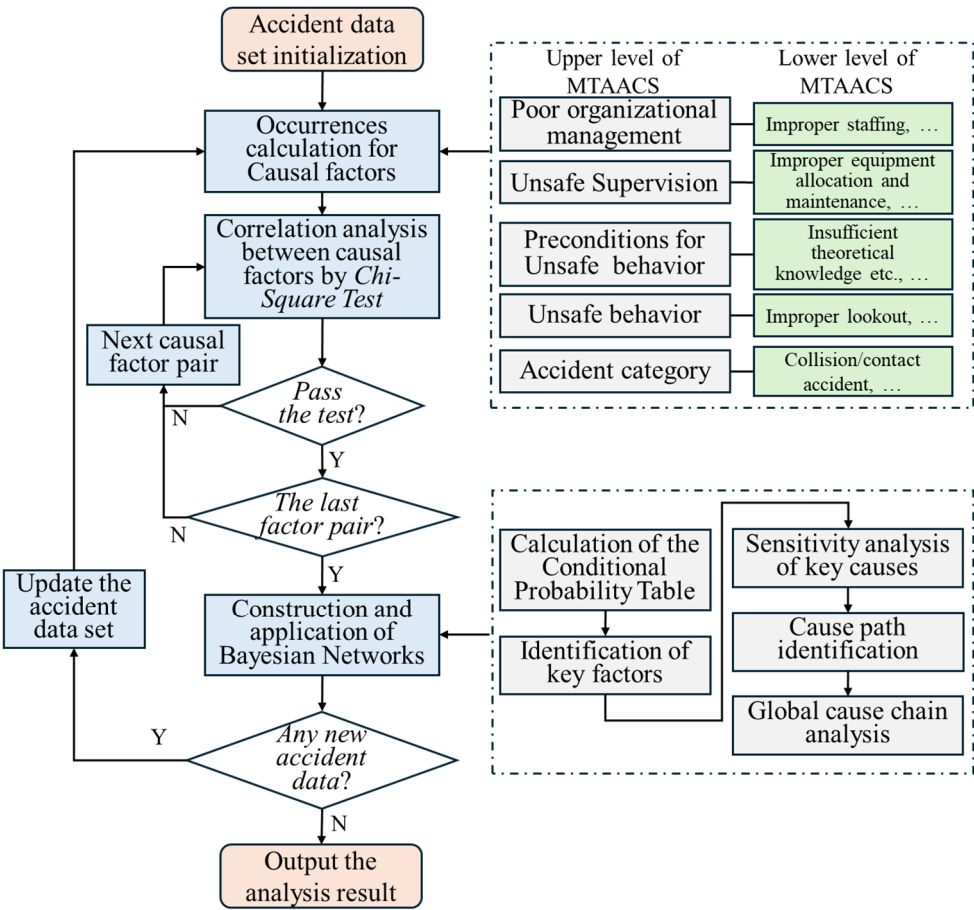


Figure 1. Illustration of the basic data-driven analysis framework.

3.2. Data Collection and Basic Analysis

To ensure the accuracy and representativeness of the data, this study primarily sourced its information from the water traffic accident investigation reports publicly released on the official website of the China Maritime Safety Administration from June 2015 to February 2023. which comprehensively document various types of accidents including collisions, contacts, fires/explosions, windstorms, sinkings, and others. However, upon in-depth analysis of the causal chains of various accidents, it was observed that there were relatively few sample data for strandings and contacts. Given the impact of sample size on research accuracy, strandings and contacts are excluded in this study. The other types of accidents have relatively abundant sample data in the publicly available reports, sufficient to provide the necessary information for an in-depth analysis and understanding of their causal chains.

After systematic data collection and screening, this study has gathered a total of 756 investigation reports on water traffic accidents along China's coast, which will serve as the core data foundation for this research. For collision accidents, each vessel is considered an independent research sample. The study has classified several common types of accidents and recorded their respective occurrence frequencies, as shown in Table 1.

Table 1. Description and Frequency of Accident Types.

Accident Type	Accident Description	Accident Frequency
---------------	----------------------	--------------------

Collision accident	An accident in which two or more ships are directly impacted and cause damage at the same time in the same space.	315
Contact accident	An accident in which a ship collides with structures above or below the water, such as quay walls, docks, navigation aids, bridge piers, floating facilities, or obstacles to navigation such as sunken ships, sunken objects, and wooden piles, causing damage.	49
Fire/explosion accident	A fire or explosion on a ship caused by the uncontrolled ignition source due to some reason during navigation, berthing, or operation.	34
Windstorm accident	An accident in which a ship suffers losses due to a strong storm.	32
Sinking accident	An accident in which a ship sinks or capsizes due to its own reasons.	135
Other accidents	An accident that does not belong to the specific categories of collision, contact, fire/explosion, windstorm, or sinking, but still results in the sinking or damage of the ship, casualties of crew members and passengers, and environmental pollution.	191

The statistical data in Figure 2 clearly shows that there are significant differences in the proportions of different types of accidents in the overall dataset. Collision accidents, other accidents, and sinking accidents account for a relatively high percentage, indicating that these types of accidents are more prone to occur in reality. In contrast, the proportions of contact accidents, fire/explosion accidents, and windstorm accidents are relatively low, suggesting that these types of accidents have a lower probability of occurrence in the statistical context. This data pattern provides us with an intuitive overview of accident occurrences.

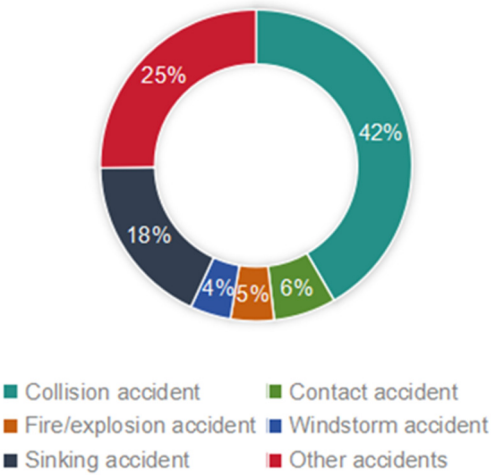


Figure 2. Distribution Chart of Accident Types.

3.3. Improved HFACS (MTAACS) for Causes of Waterborne Traffic Accidents

Although the HFACS model provides a systematic logical framework for causal linkages, its direct application to classifying factors exhibits several limitations. Firstly, the factor hierarchy set in the HFACS model, may overlook the unique nature of causal structures in waterborne transportation accidents. Secondly, the HFACS model lacks clarity in accident classification, which is particularly crucial for waterborne traffic accident analysis. Therefore, incorporating a specific "accident layer" to

explicitly classify various accident types is an indispensable aspect of the analytical framework for waterborne traffic accidents. Furthermore, in the "preconditions for unsafe behavior" level, the influence of the external environment must be analyzed in addition to internal factors. Lastly, the HFACS model assumes a linear causal relationship between adjacent levels, which may be different from the nonlinear causal relationships in waterborne traffic accidents. This complexity necessitates a more nuanced examination of the interactions and associations between factors in the analysis of waterborne traffic accidents.

Thus this paper proposes a novel Maritime Traffic Accident Analysis and Classification System (MTAACS) based on the improved HFACS. The proposed system fully accounts for the unique characteristics and complexities of waterborne traffic accidents. It categorizes the causative factors of waterborne traffic accidents into five hierarchical levels: organizational influences, inadequate supervision, preconditions for unsafe behaviors, unsafe behaviors, and the accident layer. Factors within the same level are considered as parallel relationships, implying no direct causal linkages. This arrangement aims to avoid oversimplifying the complexity of waterborne traffic accidents and allows analysts to comprehensively consider the interactions among various factors at the same level. Concurrently, the relationships between factors in adjacent levels are described as nonlinear. This indicates that a single factor may simultaneously impact multiple lower-level factors, or multiple factors may collectively act on the same lower-level factor. This nonlinear relationship can more accurately reflect the complexity and diversity of the various factors involved in waterborne traffic accidents, providing a deeper and more comprehensive analytical perspective for accident prevention.

Based on waterborne traffic accident investigation reports, causative factors relevant to each dimension are extracted. The accident risk causative factors extracted in this study stem from descriptions in accident investigation reports and can be classified into 38 factors, as shown in Table 2 and Figure 3.

Table 2. Integrated Factors Causing Water Transportation Accidents.

Factor Level	Num	Causal Factor	OC
Poor organizational management	A1	Improper staffing	94
	A2	Inadequate education and training	87
	A3	Lack of communication within the team	41
	A4	Incomplete safety management system	46
	A5	Inadequate implementation of safety management	224
	A6	Improper provision of chart information	10
	A7	Incomplete or invalid ship certificates	52
Unsafe supervision	B1	Improper equipment allocation and maintenance	81
	B2	Lack of supervision and guidance	76
	B3	Improper voyage planning	74
	B4	Improper selection of navigation areas/Improper operation and management	87
	B5	Failure to correct mistakes	11
Preconditions for unsafe behavior	C1	Insufficient theoretical knowledge, work experience, skill level, and other abilities	275
	C2	Insufficient safety awareness	285
	C3	Poor health status	43
	C4	Incomplete ship operation, navigation performance, information, and equipment	117
	C5	Improper cargo securing and stowage	70
	C6	Insufficient visibility	67
	C7	Excessive wind, waves, and currents	135
	C8	Complex navigation environment	79

Unsafe behavior	D1	Improper lookout	393
	D2	Improper watchkeeping	141
	D3	Misjudgment of danger	286
	D4	Improper collision avoidance actions	183
	D5	Failure to detect collision objects early	2
	D6	Improper use of signals	117
	D7	Improper use of equipment	90
	D8	Failure to use a safe speed	150
	D9	Failure to take early action	220
	D10	Improper location of anchorage	5
	D11	Improper emergency measures	73
	D12	Improper typhoon prevention measures	35
Accident category	E1	Collision accident	521
	E2	Contact accident	50
	E3	Fire/explosion accident	34
	E4	Windstorm accident	32
	E5	Sinking accident	135
	E6	Other accidents	192

¹ "Num" is an abbreviation for "Numbering", and "OC" is an abbreviation for "Occurrences".

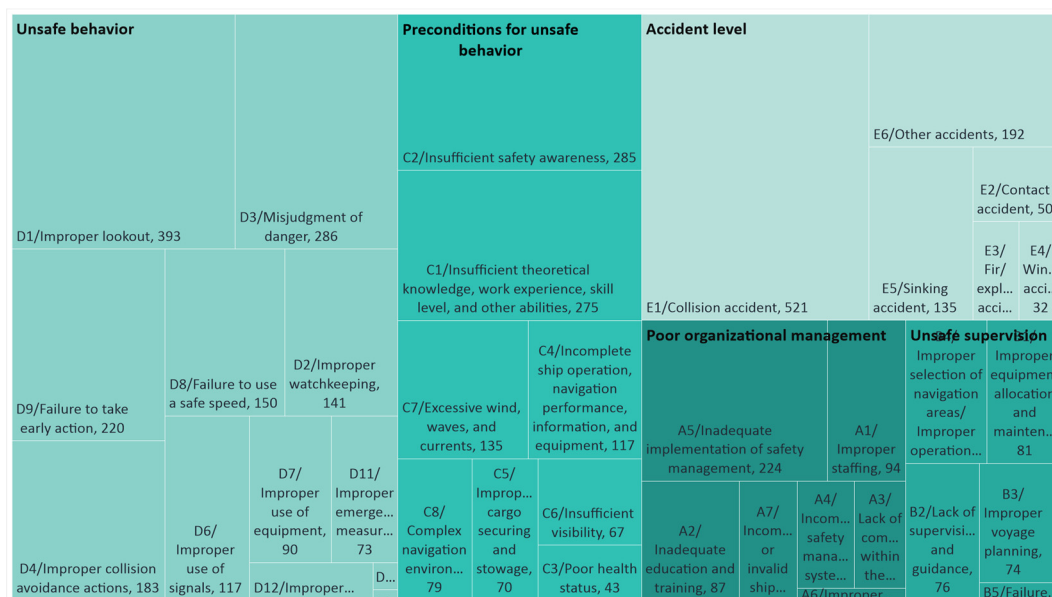


Figure 3. Statistical Chart of Layered Causes of Water Traffic Accidents.

3.4. Correlation Analysis for Causal Factor Pairs

In analyzing the causal factor chain of waterborne traffic accidents, structural learning becomes a crucial initial stage, with the core objective of clarifying the topological structure of the Bayesian network. As the foundation for constructing the Bayesian network model, structural learning determines how various nodes in the model are interconnected and their dependency relationships.

In this process, the constraint-based search method plays a decisive role. This method employs various techniques from statistics and information theory, such as the chi-square test, mutual information, and G2 test. In this study, the chi-square test is selected as the core analytical tool for structural learning. By analyzing the statistical correlation in accident data through the chi-square

test, potential links between different causal factors are identified, thereby providing a basis for constructing the topological structure of the Bayesian network model.

Taking C2 (insufficient safety awareness) and D1 (improper lookout) as an example, the chi-square test is applied to examine whether there is a significant statistical correlation between these two factors. The results of applying the chi-square test to analyze the potential link between C2 (insufficient safety awareness) and D1 (improper lookout) are shown in Figure 4. The figure indicates that the expected count (116.19) is significantly greater than 5, and the total sample size reaches 964, satisfying the conditions for conducting Pearson's Chi-Squared Test. The test results show that the P-value is less than 0.05. According to conventional statistical norms, this result rejects the null hypothesis (i.e., C2 and D1 are independent of each other) and accepts the alternative hypothesis, indicating that there is a statistically significant relationship between C2 (insufficient safety awareness) and D1 (improper lookout).

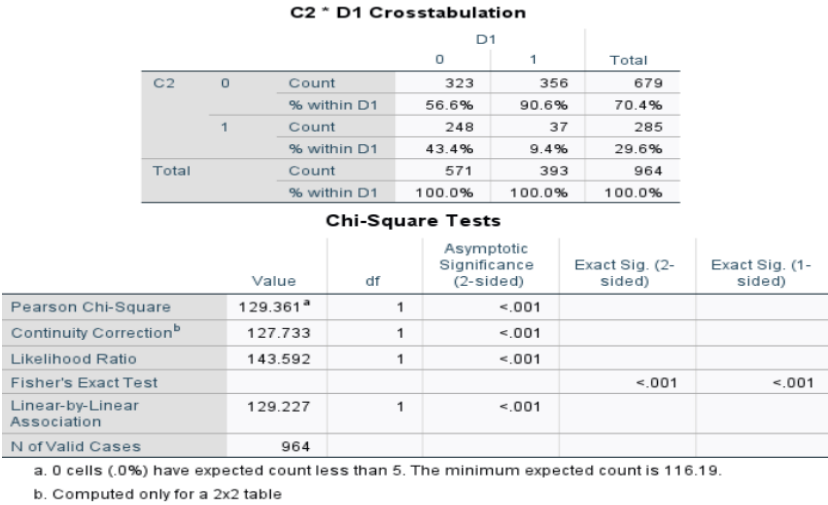


Figure 4. Chi-Square Test Result Chart for C2*D1.

3.5. Construction of Bayesian Network for Causes of Waterborne Traffic Accidents

3.5.1. Basic Concepts and Formulas

A Bayesian network is a probabilistic graphical model used to represent dependencies between variables. It is presented in the form of a Directed Acyclic Graph (DAG), a graphical structure that enables intuitive understanding and analysis of causal relationships and conditional dependencies among variables. In the model, nodes and represent random variables, directed edges represent the interrelationships between variables, and their relationships are quantified through conditional probabilities.

The theoretical basics of Bayesian networks lie primarily in probability theory and graph theory. The basic concepts and formulas for the construction of the Bayesian network are as follows.

(1) Prior Probability

Prior probability is the initial assessment of the probability of a random event or a variable taking on a specific value before observing any relevant data or evidence. In Bayesian statistics and Bayesian networks, prior probability reflects the initial belief about the state of an event or a variable based on historical data, expert knowledge, or background information.

(2) Posterior Probability

Posterior probability is the reassessment of the probability of a random event or a variable taking on a specific value after observing new data or evidence. The calculation of posterior probability is based on Bayes' theorem, which combines the prior probability (the probability before observing new data) and the likelihood function (the probability of observing the data given that the event occurred).

Posterior probability represents the updated belief about the state of an event or a variable given the new evidence or observations.

(3) Conditional Probability

The conditional probability formula is a fundamental concept in probability theory that describes the probability of one event occurring given that another event has already occurred. The conditional probability formula is defined as follows: Let A and B be two events, and $P(B) > 0$ (i.e., the probability of event B occurring is not zero). Then, the probability of event A occurring given that event B has already occurred is denoted as $P(A|B)$.

(4) Joint Probability

The joint probability formula describes the probability of two or more events occurring simultaneously. For two events A and B , the joint probability formula is given by:

$$P(A \cap B) = P(A, B) \quad (1)$$

However, if events A and B are independent, their joint probability can be obtained by multiplying their individual probabilities:

$$P(A \cap B) = P(A) \cdot P(B) \quad (2)$$

In Bayesian networks, joint probabilities are not usually given directly but are calculated indirectly through conditional probability tables. These conditional probability tables describe the probability of a child node taking each possible value given the values of its parent nodes.

(5) Total Probability Formula

The total probability formula is an important theorem in probability theory that describes the probability of a complex event occurring in terms of the probabilities of a series of simpler events and their conditional probabilities.

Let the full sample space be S . If the events $E_1, E_2, E_3, \dots, E_n$ form a complete event group (i.e., they are mutually exclusive and exhaustive, and $E_1 \cup E_2 \cup E_3 \cup \dots \cup E_n = S$), and $P(E_i) > 0$ ($i = 1, 2, 3, \dots, n$), then for any event A , we have:

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \dots + P(A|E_n)P(E_n) = \sum_{i=1}^n P(A|E_i)P(E_i) \quad (3)$$

(6) Bayes' Formula

Bayes' Theorem provides a formula that describes the relationship between two conditional probabilities.

Let the full sample space be S , and A be an event within S . If $E_1, E_2, E_3, \dots, E_n$ is a complete event group (i.e., they are mutually exclusive and exhaustive, and $E_1 \cup E_2 \cup E_3 \cup \dots \cup E_n = S$), and $P(A) > 0, P(E_i) > 0$ ($i = 1, 2, 3, \dots, n$), then Bayes' Theorem states:

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{\sum_{i=1}^n P(A|E_i)P(E_i)} \quad (4)$$

After the topological structure of the Bayesian network is determined through structural learning methods, the next crucial step is to solve for the network parameters. In this process, the Maximum Likelihood Estimation (MLE) method, as a commonly used statistical approach, is widely applied in parameter learning for Bayesian networks.

Based on the available accident data, MLE computes the likelihood function of the data and finds the parameters that maximize this function to estimate the Conditional Probability Table (CPT) for each node. The results of parameter estimation provide the CPTs for each node in the Bayesian network model, representing the probability distribution of each node given its parent node states. Further analysis of these parameters can reveal the dependency relationships between different factors and their degrees of influence.

3.5.2. Bayesian Network for Causes of Water Traffic Accidents

Based on the established Conditional Probability Table (CPT), we can get the Bayesian Network for causes of water traffic accidents using the GeNIe software. The network can be shown in Figure 5.

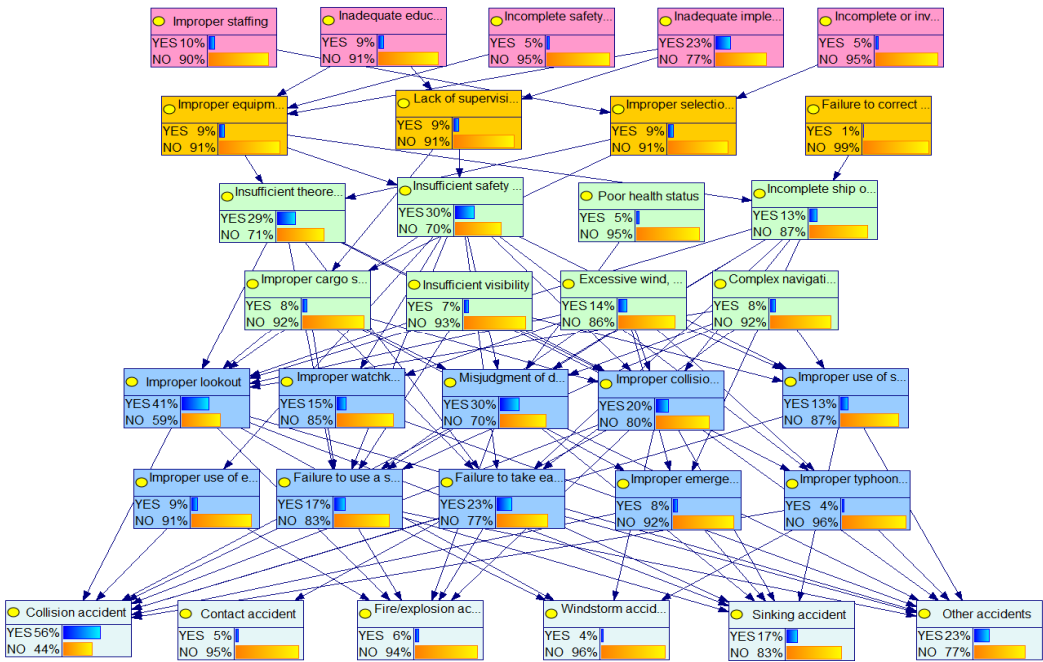


Figure 5. Bayesian Network Calculation Result Chart for Causes of Water Traffic Accidents.

4. Analysis of the Causal Chain of Waterborne Traffic Accidents

4.1. Identification of Key Factors in Waterborne Traffic Accidents

Identifying and analyzing the key factors is crucial for understanding the causes of waterborne traffic accidents, predicting accident risks, formulating effective preventive measures and emergency response strategies. Based on the node probabilities presented in Figure 5, the top three factors with higher occurrence probabilities in each level are extracted and compiled into a table of key factors, as shown in Figure 6.

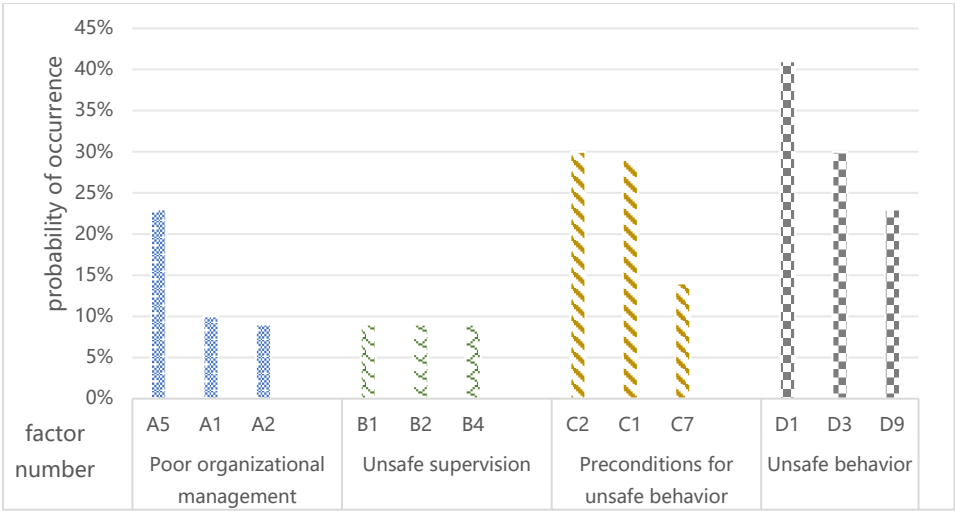


Figure 6. Top Three Causes with the Highest Occurrence Probabilities at Each Level.

Based on the analysis of Figure 6, at the organizational management level, inadequate safety management execution (A5), improper staffing (A1), and inadequate education and training (A2) are prominently identified as the three core factors affecting waterborne traffic safety. In particular, the

high incidence of inadequate safety management execution emphasizes a significant lack of safety management. Together, these three factors constitute potential root causes of waterborne traffic safety accidents, creating hidden dangers and increasing risks during ship operations.

At the unsafe supervision level, inadequate equipment provisioning and maintenance (B1), lack of supervision and guidance (B2), and improper navigation area selection and operational management (B4) are the three most problematic areas. These deficiencies not only have a direct impact on decision-making and execution at the next level, but also indirectly amplify safety risks in waterborne traffic.

At the preconditions for unsafe behaviors level, insufficient safety awareness (C2), inadequate theoretical knowledge, work experience, skill levels, and other competencies (C1), as well as external factors such as excessive wind, waves, and currents (C7), are recognized as the primary factors inducing waterborne traffic accidents. These preconditions provide conditions for unsafe behaviors to occur, increasing the likelihood of accidents.

At the unsafe behavior level, improper lookout (D1), misjudgment of hazards (D3), and failure to take early action (D9) are the most direct behaviors that can lead to accidents. These behaviors involve misjudgments and faulty decisions made by crew members during navigation, highlighting the importance of enhancing crew safety awareness and operational skills.

4.2. Sensitivity Analysis of Causes of Waterborne Traffic Accidents

Sensitive factors can be identified through sensitivity analysis of the Bayesian network. This analysis aims to determine the degree of influence of network parameters on the posterior probability of specific target nodes, thereby identifying the key factors contributing to various types of accidents. The core algorithm of Bayesian network sensitivity analysis involves calculating the differential of the target node's posterior probability with respect to network parameters, quantifying the impact of parameter changes on the posterior probability. In the GeNIe software, by setting the state of the target node and conducting sensitivity analysis, the importance of upper-level factors on lower-level factors can be visually displayed. Combined with the calculation results of sensitivity parameters, further quantitative analysis can be conducted.

Taking a windstorm accident as an example, we set E4 (windstorm accident) in the accident layer as the target node and conduct a sensitivity analysis. During this process, the colors of the Bayesian network nodes change to visually indicate the location of sensitive factors. Red nodes represent the locations of relevant parameters, with darker shades indicating higher sensitivity of the associated parameters. In Figure 7, it can be clearly seen that insufficient theoretical knowledge, work experience, skill levels, and other competencies (C1), excessive wind, waves, and currents (C7), and inadequate typhoon prevention measures (D12) are the most relevant factors to windstorm accidents. The causal strength of each factor needs to be combined with sensitivity parameters for quantitative assessment. The GeNIe software provides the maximum, minimum, and average values of the sensitivity of each node's relevant parameters. Among them, the maximum value is consistent with the shade of the associated node's color, indicating that the maximum value of the sensitivity analysis is the primary criterion for assessing the sensitivity of associated nodes.

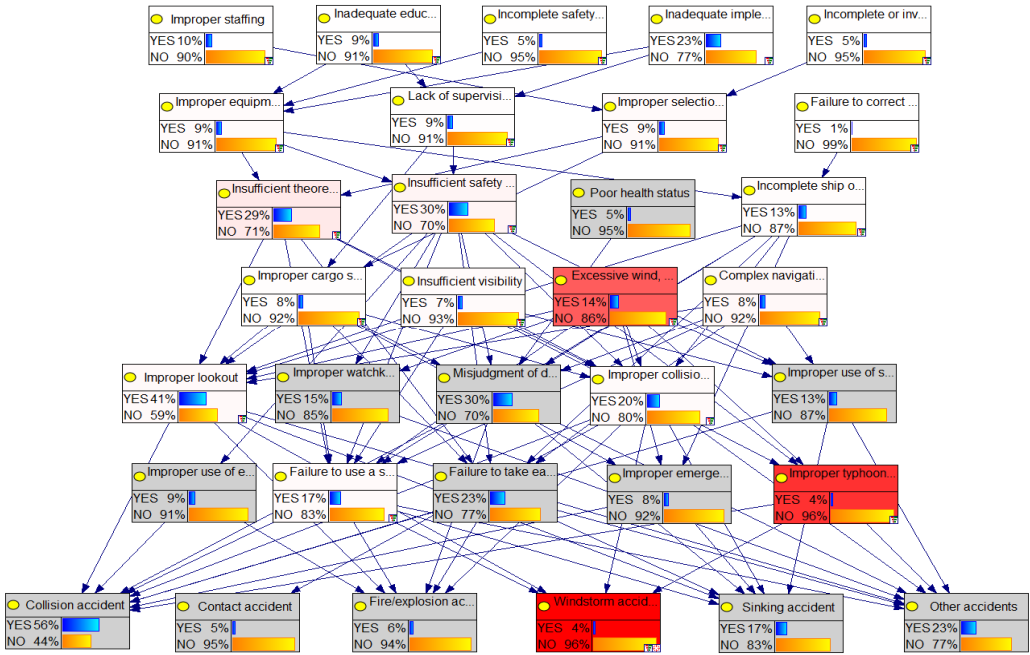


Figure 7. Sensitivity Analysis Network Diagram for Wind Disaster Accidents.

Through the sensitivity analysis of various factors contributing to waterborne traffic accidents, sensitivity parameter data are obtained in the Figures 8–14 lists top three sensitivity factors for different types of accidents. These parameters reveal the relative importance of causal factors in different types of accidents.

Based on Figures 8 and 14, for collision accidents, the analysis results show that insufficient safety awareness (C2) has the largest absolute value of sensitivity parameter, indicating that the level of crew's safety awareness has the most significant impact on the occurrence of collision accidents. Therefore, enhancing crew's safety awareness is of great significance in reducing collision accidents.

According to Figure 9 and Figure 14, for contact accidents, improper collision avoidance behavior (D4) has the largest absolute value of sensitivity parameter among the causal factors. This reflects that during navigation, proper collision avoidance behavior is crucial in preventing contact accidents.

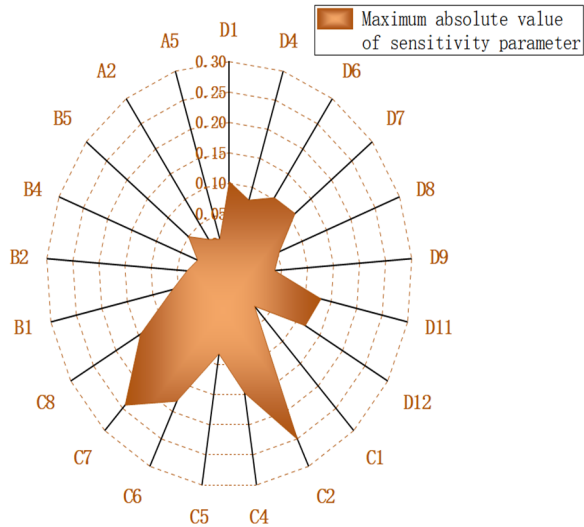


Figure 8. Sensitivity Analysis Results for Collision Accidents.

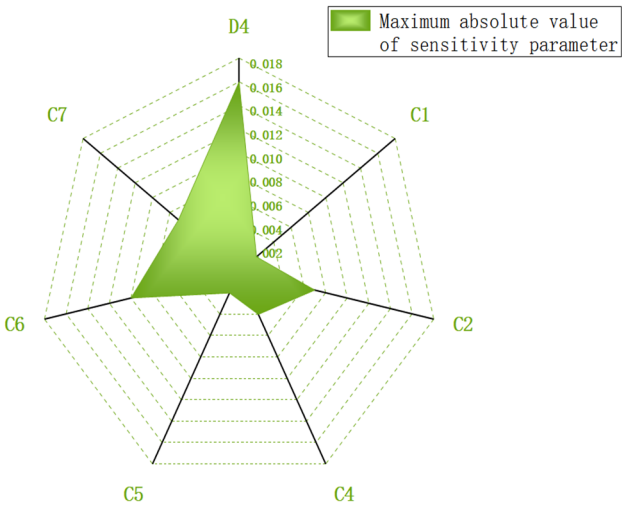


Figure 9. Sensitivity Analysis Results for Contact Accidents.

Based on Figures 10 and 14, for fire/explosion accidents, the absolute value of the sensitivity parameter for improper equipment use (D7) is the highest, indicating that improper use of equipment is one of the primary contributors to fire/explosion accidents. Therefore, enhancing the standardization and safety of equipment use is crucial in reducing fire/explosion accidents.

According to Figures 11 and 14, for windstorm accidents, the absolute value of the sensitivity parameter for inadequate typhoon prevention measures (D12) is the highest, indicating that the adequacy of typhoon prevention measures taken by ships directly relates to the incidence of windstorm accidents during severe weather conditions such as typhoons. Therefore, strengthening the formulation and implementation of typhoon prevention measures for ships is crucial in reducing the risk of windstorm accidents.

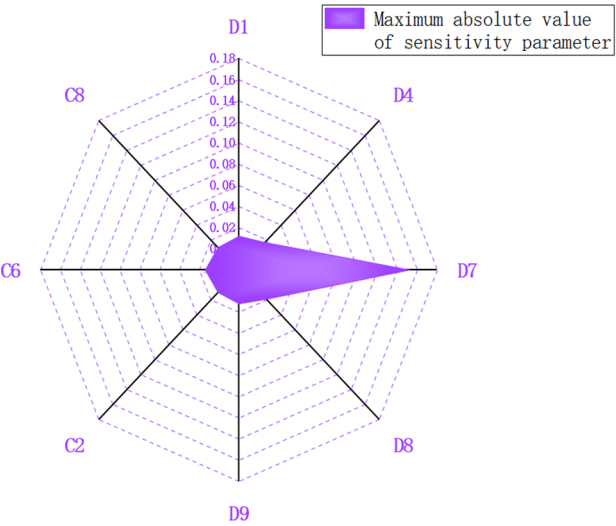


Figure 10. Sensitivity Analysis Results for Fire/Explosion Accidents.

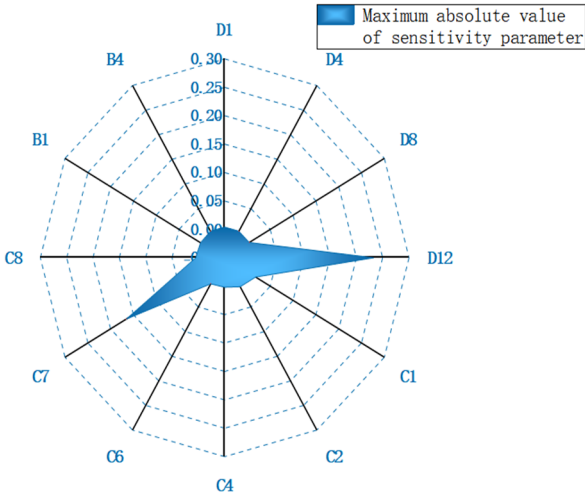


Figure 11. Sensitivity Analysis Results for Windstorm Accidents.

Through Figures 12 and 14, for sinking accidents, the absolute value of the sensitivity parameter for improper emergency measures (D11) is the highest, indicating that the adequacy of emergency measures taken by ships in emergency situations plays a decisive role in preventing sinking accidents. Therefore, improving crew's emergency response capabilities and formulating effective emergency measures are important ways to reduce sinking accidents.

Through Figures 13 and 14, for other types of accidents, insufficient safety awareness (C2) also has the largest absolute value of sensitivity parameter, which once again emphasizes the universal importance of enhancing crew's safety awareness in reducing waterborne traffic accidents.

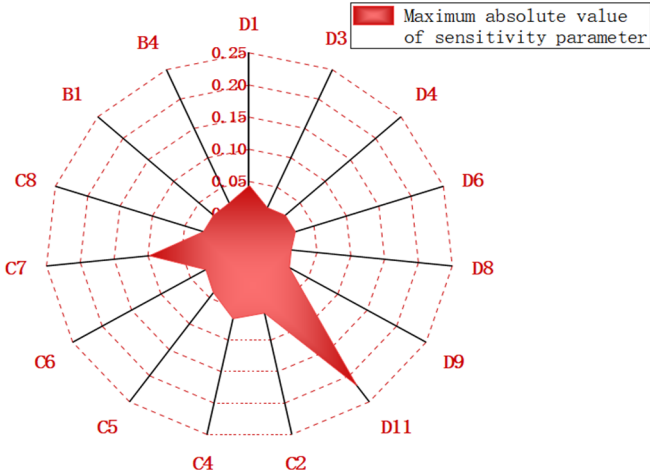


Figure 12. Sensitivity Analysis Results for Sinking Accidents.

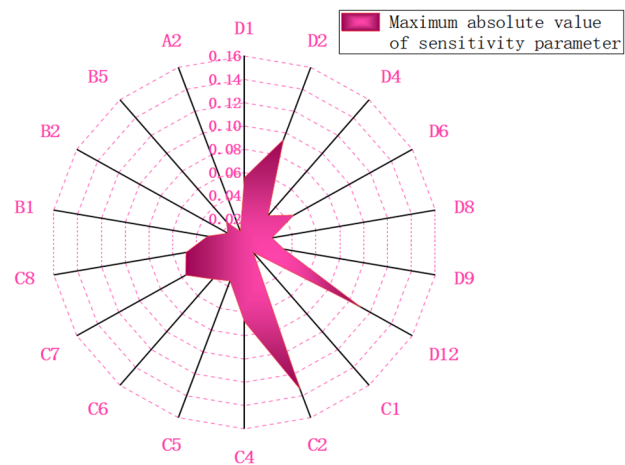


Figure 13. Sensitivity Analysis Results for Other Accidents.

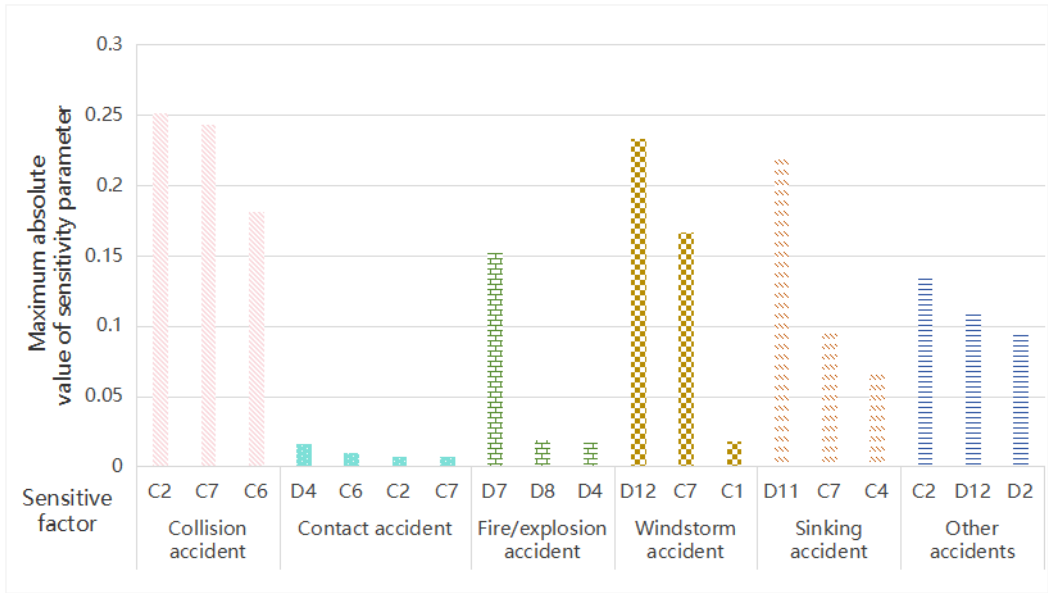


Figure 14. Top Three Sensitivity Factors for Different Types of Accidents.

4.3. Cause Paths of Waterborne Traffic Accidents

After deeply exploring the key and sensitive factors of waterborne traffic accidents, the next step is to construct the cause paths of waterborne traffic accidents. Based on the results of sensitivity analysis, a rigorous screening process is conducted for the causal factors in the waterborne traffic accident cause-chain analysis model. Specifically, the causal factors that have insignificant impacts on various accidents (i.e., with node sensitivity parameters less than 0.01) are eliminated.

However, in the analysis of causal factors for contact accidents, a special situation was identified: the sensitivity of most causal factors was less than 0.01. To comprehensively and thoroughly analyze the cause chain of contact accidents, a special strategy was adopted: when constructing the cause paths, the three nodes with relatively higher sensitivity at each level were selected for focused analysis. The results can be shown in Table 3 and Figure 15.

Table 3. Cause Paths for Different Types of Accidents.

Accident Type	Cause Paths
Collision accident (E1)	A2(A5)-B1-C2-D1-E1; A2(A5)-B1-C2-D6-E1; A2(A5)-B1-C2-D7-E1;
	A2(A5)-B1-C2-D12-E1; B5-C4-D1-E1; B5-C4-D11; C6-D1-E1;
	C6-D6-E1; C7-D1-E1; C7-D6-E1; C7-D11-E1; C7-D712-E1

Contact accident(E2)	C2-D4-E2; C6-D4-E2; C7-D4-E2
Fire/explosion accident (E3)	C2-D4-E3; C2-D7-E3; C2-D8-E3; C6-D4-E3; C8-D7-E3; C8-D8-E3
Windstorm accident (E4)	C1-D2-E4; C2-D2-E4
Sinking accident (E5)	B1-C2-D1-E5; B1-C4-D1-E5; B5-C4-D1-E5; C7-D11-E5 A2(A4)(A5)-B1-C2-D1-E6; A2(A4)(A5)-B1-C2-D2-E6; A2(A4)(A5)-B1-C2-D12-E6; A2(A4)(A5)-B1-C7-D12-E6; A2(A5)-B2-C2-D1-E6; A2(A5)-B2-C2-D2-E6; A2(A5)-B2-C2-D12-E6
Other accidents (E6)	

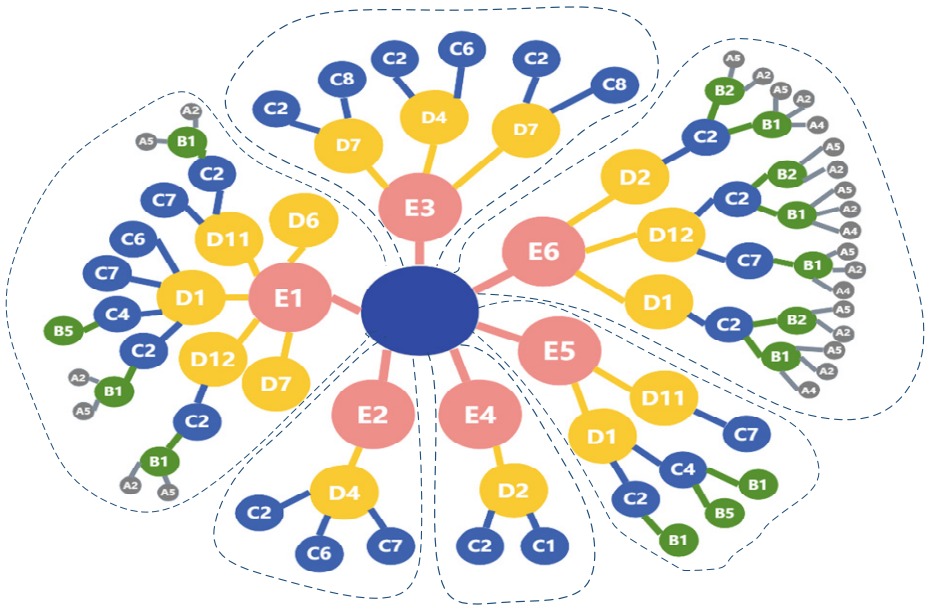


Figure 15. Cause paths of Different Types of Accidents.

As can be seen from Table 3 and Figure 15, the number of cause paths for various types of waterborne traffic accidents varies: there are 12 cause paths for collision accidents, 3 for contact accidents, 6 for fire/explosion accidents, 2 for windstorm accidents, 4 for sinking accidents, and 7 for other types of accidents.

The cause paths of collision accidents show complex diversity, encompassing the intertwined influences of human errors (such as improper lookout and operation) and environmental factors (such as insufficient visibility and complex current conditions). In contrast, the cause paths of contact accidents are relatively concise, centered on improper collision avoidance behavior, while also being influenced by certain environmental factors. The cause paths of fire/explosion accidents reveal potential risks across multiple links from equipment failures to improper human operations, demonstrating their unique complexity. The cause paths of windstorm accidents are closely related to extreme weather conditions, especially the insufficient wind resistance capabilities of ships under severe weather. The cause paths of sinking accidents involve improper cargo stowage and ship operational performance issues. The cause paths of other accident categories are more extensive and complex, reflecting the diversity and complexity of the causes of waterborne traffic accidents.

4.4. Global Causal Chain Analysis of Waterborne Traffic Accidents

Using the causal reasoning method of Bayesian networks, we analyze the global causal chains of the causal factors of waterborne traffic accidents. By the causal reasoning method, the target node

(i.e., the occurrence status of different accident types) is set as a known condition and the posterior probabilities of evidence nodes are calculated to infer the possible causes of the accident.

Taking windstorm accidents (E4) as an example, the occurrence status of the target node is set to 1, and the posterior probabilities of its four parent nodes - failure to use a safe speed, improper lookout, improper collision avoidance behavior, and inadequate typhoon prevention measures - are derived. Results show that the occurrence probability of inadequate typhoon prevention measures (D12) is the highest, reaching 69%. Further analysis on the three other parent nodes of inadequate typhoon prevention measures - lack of theoretical knowledge/work experience/skills/other abilities, insufficient safety awareness, and excessive wind/waves/currents - reveals that the posterior probability of excessive wind, waves, and currents (C7) is relatively high. Using the same method, the global causal chain analysis can be conducted for other types of accidents, and the global causal chains for all the accident types are obtained, as shown in Table 4.

Table 4. Global Causal Chains for Different Types of Accidents.

Accident Type	Global Causal Chain
Collision accident (E1)	A1-B4-C1-D1-E1
Contact accident(E2)	A5-B2-C2-D1-E2
Fire/explosion accident (E3)	A5-B1-C2-D7-E3
Windstorm accident (E4)	C7-D12-E4
Sinking accident (E5)	A1-B4-C1-D3-E5
Other accidents (E6)	A1/A5-B4/B2-C1/C2-D3

Based on the results obtained above, we can get some implications as follows.

- (1) Collision accidents occur frequently due to a complex array of causes. These include improper staffing, improper selection of navigation areas/improper operation and management, insufficient theoretical knowledge, work experience, skill level, and other abilities, and improper lookout, etc. Measures including enhancing crew safety training, optimizing navigation environments, improving crew management systems, and enhancing vessel lookout capabilities are essential to prevent collision accidents.
- (2) Contact accidents are mainly caused by inadequate implementation of safety management, lack of supervision and guidance, insufficient safety awareness, and improper lookout. Based on these results, strengthening crew training especially on the standardized process of safety management, enhancing supervision normalization, and enhancing vessel lookout capabilities are crucial to reduce the probability of contact accidents.
- (3) Fire/explosion accidents often result from inadequate implementation of safety management, improper equipment allocation and maintenance, insufficient safety awareness, and improper use of equipment etc. To prevent them, regular inspections and maintenance of fire-fighting and explosion-proof equipment are essential. Additionally, strict fire source management systems must be implemented.
- (4) Windstorm accidents frequently occur due to crew members' lack of theoretical knowledge, work experience, and skills, leading to ineffective supervision, guidance, and inadequate plans in severe weather. Hence, improving crew competence, enhancing decision-making under extreme conditions, and implementing scientific navigation plans are crucial for safe navigation during wind disasters.
- (5) Sinking accidents mainly stem from improper staffing, improper selection of navigation areas/improper operation and management, insufficient theoretical knowledge, work experience, skill level, and other abilities, and misjudgment of danger. To prevent these, enhancing crew safety training, optimizing navigation environments, supervision of ship maintenance should be strengthened. Other accidents often result from various combinations of factors, such as inadequate education/training, leading to ineffective safety management and improper equipment handling.

5. Conclusions

Water traffic accidents not only result in casualties and losses but also pollute waterways and damage ecosystems. This study focuses on investigating the causal chains of waterborne traffic accidents in a data-driven context, delving into their underlying factors and chain reactions. Based on China's nationwide investigation reports of waterborne traffic accidents, this research systematically extracts the causal factors leading to accidents, and constructs a hybrid framework based on the improved HFACS and Bayesian networks. Through quantitative analyses including chi-square tests and sensitivity analyses, the cause chains of different types of accidents are thoroughly explored, revealing the fundamental mechanisms behind accident occurrences. Results indicate that there are 12, 3, 6, 2, 4, and 7 causal chains leading to collisions, contacts, fire/explosion, windstorms, sinking, and other types of accidents respectively. The research results can provide a reference for the enhancement on the safe operation of waterborne transportation.

Future research will focus on the application of new technologies such as artificial intelligence to deepen the understanding of the cause chains of waterborne traffic accidents. We aim to explore the potential of these new technologies in accident prevention, monitoring, and emergency response. This will promote the development of water transportation safety management towards intelligence and automation.

Funding: This research was funded by Natural Science Foundation of China (52362055), Guangxi Science and Technology Major Program of China (Grant No. AA23062053).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Formela K, Weintrit A, Neumann T. Overview of definitions of maritime safety, safety at sea, navigational safety and safety in general[J]. *International Journal on Marine Navigation and Safety of Sea Transportation*, 2019, 13(2): 285-290.
2. Peng Z, Jiang Z, Chu X, et al. Spatiotemporal Distribution and Evolution Characteristics of Water Traffic Accidents in Asia since the 21st Century[J]. *Journal of Marine Science and Engineering*, 2023, 11(11): 2112.
3. Wu B, Kou L, Ma Q. Research on HFACS based on accident causality diagram[J]. *Open journal of safety science and technology*, 2017, 7(2): 77-85.
4. Zhang G, Thai V V. Expert elicitation and Bayesian Network modeling for shipping accidents: A literature review[J]. *Safety science*, 2016, 87: 53-62.
5. Wang H, Liu Z, Wang X, et al. An analysis of factors affecting the severity of marine accidents[J]. *Reliability Engineering & System Safety*, 2021, 210(3):107513.
6. Liu K, Yu Q, Yuan Z, et al. A systematic analysis for maritime accidents causation in Chinese coastal waters using machine learning approaches[J]. *Ocean & Coastal Management*, 2021, 213:105859.
7. Sun J, Li M, Xiu X. Analysis of the Causes of Water Transportation Accidents Based on Complex Network Theory [J]. *Journal of Dalian Maritime University*, 2023, 49(02): 80-90+160.
8. Wu Y, Jiang F, Yao H, et al. Analysis of Causal Factors and Risk Prediction for Inland River Ship Collision Accidents Based on Text Mining [J]. *Journal of Transport Information and Safety*, 2018.36(03): 8-18.
9. Wang H, Liu Z, Liu Z, et al. GIS-based analysis on the spatial patterns of global maritime accidents[J]. *Ocean engineering*, 2022, 245: 110569.
10. Kaptan M, Sarialioğlu S, Uğurlu Ö, et al. The evolution of the HFACS method used in analysis of marine accidents: A review[J]. *International Journal of Industrial Ergonomics*, 2021, 86: 103225
11. Yildiz S, Uğurlu Ö, Wang J, et al. Application of the HFACS-PV approach for identification of human and organizational factors (HOFs) influencing marine accidents[J]. *Reliability Engineering & System Safety*, 2021, 208: 107395.
12. Wang Q, Sha, Z, Zhang J, et al. Human Factor Analysis of Ship Grounding Accidents Based on the HFACS-FCMs Model [J]. *Journal of Shandong Jiaotong University*, 2024, 32(01): 103-109+123.
13. Yıldırım U, Başar E, Uğurlu Ö. Assessment of collisions and grounding accidents with human factors analysis and classification system (HFACS) and statistical methods[J]. *Safety Science*, 2019, 119: 412-425.

14. Huang H. Analysis and Application of Human Factors in Marine Traffic Based on the HFACS Method [J]. *Shandong Industrial Technology*, 2018, (05): 216-218+238.
15. Chen S, Wall A, Davies P, et al. A Human and Organizational Factors (HOFs) analysis method for marine casualties using HFACS-Maritime Accidents (HFACS-MA)[J]. *Safety science*, 2013, 60:105-114.
16. Chauvin C, Lardjane S, Morel G, et al. Human and organizational factors in maritime accidents: Analysis of collisions at sea using the HFACS[J]. *Accident Analysis & Prevention*, 2013, 59:26-37.
17. Fan S, Blanco-Davis E, Yan Z, et al. Incorporation of human factors into maritime accident analysis using a data-driven Bayesian network[J]. *Reliability Engineering & System Safety*, 2020, 203:107070.
18. Antão P, Soares C G. Analysis of the influence of human errors on the occurrence of coastal ship accidents in different wave conditions using Bayesian Belief Networks[J]. *Accident Analysis & Prevention*, 2019, 133: 105262.
19. Wang L, Yang Z. Bayesian network modelling and analysis of accident severity in waterborne transportation: A case study in China[J]. *Reliability Engineering & System Safety*, 2018, 180:277-289.
20. Meng X, Li H, Zhang W, et al. Analyzing risk influencing factors of ship collision accidents: A data-driven Bayesian network model integrating physical knowledge[J]. *Ocean & Coastal Management*, 2024, 256: 107311.
21. Fan S, Yang Z, Blanco-Davis E, et al. Analysis of maritime transport accidents using Bayesian networks[J]. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2020, 234(3): 439-454.
22. Tian Y, Qiao H, Hua L, et al. Bayesian Network Model for Maritime Ship Collisions and Its Application [J]. *Journal of Naval University of Engineering*, 2023, 35(06): 28-33.
23. Hänninen M, Kujala P. The effects of causation probability on the ship collision statistics in the Gulf of Finland[J]. *Marine Navigation and Safety of Sea Transportation*, 2010, 4(1):79-84.
24. Meng X, Li H, Zhang W, et al. Analyzing ship collision accidents in China: A framework based on the NK model and Bayesian networks[J]. *Ocean Engineering*, 2024, 309: 118619.
25. Wang H, Jin G. Cause analysis of marine accidents based on HFACS and Bayesian networks[C]. *International Conference on Smart Transportation and City Engineering*, 2024, 13018: 1022-1030.
26. Rostamabadi A, Jahangiri M, Zarei E, et al. A novel fuzzy bayesian network-HFACS (FBN-HFACS) model for analyzing human and organization factors (HOFs) in process accidents[J]. *Process Safety and Environmental Protection*, 2019, 132: 59-72.
27. Jiang Y, Wan Z, Chen J. Path Analysis of the Causes of Coastal Water Transportation Accidents in China [J]. *Journal of Dalian Maritime University*, 2024, 50(01): 76-84.
28. Li Y, Cheng Z, Yi T, et al. Use of HFACS and Bayesian network for human and organizational factors analysis of ship collision accidents in the Yangtze River[J]. *Maritime Policy & Management*, 2022, 49(8): 1169-1183.
29. Wang H, Chen N, Wu B, et al. Human and organizational factors analysis of collision accidents between merchant ships and fishing vessels based on HFACS-BN model[J]. *Reliability Engineering & System Safety*, 2024, 249: 110201.
30. Özkan U, Serdar Y, Sean L et al. Analyzing Collision, Grounding, and Sinking Accidents Occurring in the Black Sea Utilizing HFACS and Bayesian Networks[J]. *Risk analysis: an official publication of the Society for Risk Analysis*, 2020, 40(12):2610-2638.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.