Transportation Research Part E 216 (2026) 105134

Contents lists available at ScienceDirect

Transportation Research Part E

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

Integrating prompt-based zero-shot learning into Bayesian network for maritime cybersecurity risk analysis

Yunfeng Zhao [a], Huanhuan Li [b,*], Trung Thanh Nguyen [a], Christian Matthews [c], Zaili Yang [a,*]

[a] Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool, UK
[b] School of Engineering, University of Southampton, Southampton, UK
[c] Department of Maritime and Mechanical Engineering, Liverpool John Moores University, Liverpool, UK

ARTICLE INFO

Keywords:
Maritime cybersecurity
Large language model
Zero-shot learning
Prompt engineering
Bayesian network

ABSTRACT

The increasing digitalisation of maritime ecosystems, including vessels, ports, offshore facilities, and shore-based organisations, has expanded the attack surface of maritime operations and increased exposure to cyber threats. Cyber incidents affecting these interconnected systems may propagate across global logistics networks, highlighting the need for systematic cybersecurity risk analysis in maritime transport. However, existing studies often rely on expert judgement or limited datasets, which restrict scalability and hinder comprehensive identification of Risk Influential Factors (RIFs). This study develops a data-driven framework that integrates Large Language Model (LLM)-assisted text classification with Bayesian Network (BN) modelling to analyse maritime cybersecurity risks. A structured taxonomy of RIFs is defined to capture temporal, regional, operational, and technical attributes of cyber incidents. LLM-based prompt engineering is then employed to automatically extract these factors from unstructured incident reports. The reliability of the automated classification process is evaluated through manual annotation on a sampled subset of incident reports and comparative experiments across multiple LLMs. The BN model is further assessed using repeated cross-validation to evaluate predictive performance and robustness. The results reveal key dependencies among cyber incident characteristics and consequence types, providing probabilistic insights into maritime cyber risk patterns. The proposed framework offers a scalable approach for analysing cyber risks in maritime transport systems and supports risk-informed cybersecurity management.

1. Introduction

Maritime transport plays a vital role in global economic prosperity by facilitating both regional and international trade. Its significance has grown due to its comparatively lower costs and reduced emissions per unit of cargo transported, making it an increasingly attractive mode of transport (Trade, 2024). As the industry advances, technologies such as the Internet of Things (IoT), Big Data, Digital Twins (DTs), and Artificial Intelligence (AI) are accelerating the development of digital maritime infrastructure (Yang, 2019). However, this increasing digital connectivity also brings heightened cybersecurity risks, such as ransomware, Distributed Denial-of-Service (DDoS) attacks, and phishing, which have become more frequent and sophisticated in recent years (Wu et al., 2025). In

* Corresponding authors.
E-mail addresses: Huanhuan.Li@soton.ac.uk (H. Li), Z.Yang@ljmu.ac.uk (Z. Yang).

https://doi.org/10.1016/j.tre.2026.105134
Received 7 August 2025; Received in revised form 8 July 2026; Accepted 28 July 2026
Available online 10 August 2026

addition, maritime cyberattacks may disrupt shipboard operations, port logistics, offshore energy infrastructures and maritime governance, resulting in operational interruptions, economic losses, safety hazards, and regulatory challenges (Iaiani et al., 2021). Fig. 1 illustrates the distribution of maritime cyberattack incidents, based on data from the Maritime Cyber Attack Database (MCAD), an open-source platform developed by the Maritime IT Security Research Group in the Netherlands (Pijpker et al., 2024). These developments highlight the urgent need for more advanced and data-driven cybersecurity risk analysis methods for maritime transport systems.

To address these cybersecurity challenges, various risk assessment approaches have been developed to identify, evaluate, and manage cyber threats in maritime systems. Existing studies mainly rely on either conventional expert-driven methods or probabilistic data-driven approaches (Park et al., 2023; Yoo and Park, 2021). While these approaches provide useful support for risk assessment, many still depend heavily on expert knowledge or limited empirical data, reducing their scalability and reproducibility. More recently, Mohsendokht et al. (2024) developed a data-driven BN model based on the MCAD. However, several important Risk Influential Factors (RIFs), including threat source, attacker intent, targeted systems, and consequence types, were not incorporated into their model. As a result, valuable information remains embedded in unstructured cyber incident reports, limiting comprehensive risk analysis.

Recent advances in Natural Language Processing (NLP), particularly Large Language Models (LLMs), have provided new opportunities for extracting structured knowledge from such unstructured textual data (Zhang et al., 2026). The success of ChatGPT (Mao et al., 2024) has brought significant attention to prompt-based learning. Prompt-based learning enables LLMs to perform text classification tasks through carefully designed prompts without requiring large labelled datasets (Liu et al., 2023). This capability is particularly valuable in maritime cybersecurity research, where labelled datasets are often limited, facilitating data-driven cybersecurity risk analysis through effective knowledge extraction from cyber incident reports.

Based on the background and challenges discussed above, this study aims to develop a data-driven approach for analysing maritime cybersecurity risks using incident data. In particular, the study focuses on identifying RIFs from cyber incident reports and modelling their relationships with cybersecurity consequences. Accordingly, the following research questions are addressed:

RQ1: How can structured RIFs associated with maritime cyber incidents be systematically extracted from unstructured cyber incident reports?

RQ2: How can the relationships between these RIFs and different cybersecurity consequences be modelled to support data-driven maritime cybersecurity risk analysis?

RQ3: How can the resulting analytical insights support cybersecurity risk prioritisation and decision-making in maritime operations?

To address these questions, this study makes three main contributions.

(1) A structured taxonomy of maritime cybersecurity RIFs is established to capture temporal, regional, operational, and technical attributes of maritime cyber incidents.
(2) A data-driven analytical framework is developed by integrating LLM-assisted zero-shot classification with BN modelling. This framework enables automated extraction of structured RIFs from unstructured cyber incident reports and probabilistic analysis of their relationships.
(3) The proposed framework identifies key probabilistic dependencies between cyber incident characteristics and consequence types, providing data-driven insights that support cybersecurity risk prioritisation and risk-informed decision-making in maritime systems.

The remainder of this paper is organised as follows. Section 2 reviews the literature on maritime cybersecurity risk analysis and text

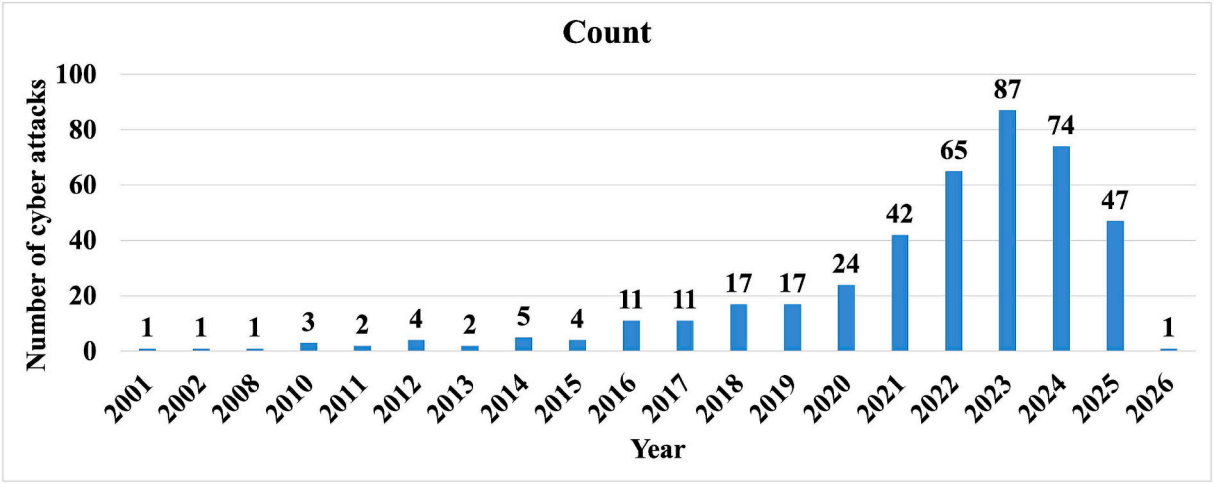

Fig. 1. The distribution of maritime cyber-attack incidents (up to January 2026).

classification methods. Section 3 presents the methodological framework, including the definition of RIFs, the LLM-assisted classification approach, and the BN modelling process. Section 4 evaluates the model performance and analyses the probabilistic relationships among RIFs. Section 5 discusses scenario-based insights and implications for maritime cybersecurity management. Finally, Section 6 concludes the study and outlines directions for future research.

2. Literature review

2.1. Maritime cybersecurity risk analysis methods

With the rapid digitalisation of maritime operations, cybersecurity has become an increasingly critical concern for ships, ports, and maritime supply chains. Consequently, a growing body of research has focused on identifying and analysing cybersecurity risks affecting maritime infrastructure. Existing literature reviews (Ashraf et al., 2023; Erbas et al., 2024; Peng et al., 2025) provide valuable overviews of maritime cyber threats and defence mechanisms. However, most existing studies remain largely descriptive and provide limited analytical capability for predictive cybersecurity risk assessment.

Conventional risk analysis techniques have long been used to support safety and security analysis in maritime and other industrial systems. Methods such as Hazard and Operability Study (HAZOP) (Mansoori et al., 2016), attack tree analysis (Komal, 2023), and Failure Modes and Effects Analysis (FMEA) (Söner et al., 2023) provide structured procedures for identifying potential hazards, analysing failure mechanisms, and evaluating operational risks. These approaches are typically applied through expert workshops or structured scenario analysis and are effective in systematically identifying potential failure modes or attack pathways. However, they rely heavily on expert judgement and predefined scenarios, which can limit their scalability and introduce unavoidable subjectivity when analysing large volumes of cybersecurity incident data or rapidly evolving cyber threats.

In recent years, several studies have proposed domain-specific cybersecurity risk assessment frameworks tailored to maritime systems. Representative examples include the Maritime Cyber Risk Assessment (MaCRA) framework (Tam and Jones 2019) and the MITIGATE Supply Chain Risk Assessment (SCRA) methodology (Schauer et al., 2019). These frameworks aim to support structured cyber risk analysis across maritime operational environments, including shipboard systems, port infrastructures, and maritime logistics networks. For instance, MaCRA facilitates the identification of vulnerability characteristics, exploitability, and potential cyberattack impacts within maritime information systems. Similarly, MITIGATE SCRA provides a multi-stage methodology for analysing cybersecurity risks within maritime supply chains. Although these frameworks offer systematic procedures for risk assessment, they typically rely on predefined attack scenarios and static knowledge structures, limiting their ability to adapt to emerging cyber threats.

Another stream of research focuses on expert-driven risk quantification approaches, which combine structured methodologies with expert knowledge to evaluate cybersecurity risks. For example, Gunes et al. (2021) proposed a 13-stage cybersecurity framework for seaport systems, while other studies have developed hybrid BN models incorporating expert judgement to analyse maritime cyber risks (Park et al., 2023; Uflaz et al., 2024). These approaches enable structured evaluation of cybersecurity risks across maritime infrastructure, including port terminal operations and vessel communication networks. However, their strong reliance on expert elicitation introduces potential biases and inconsistencies, particularly when historical cyber incident data are limited.

To address uncertainty and complex interdependencies among RIFs, previous studies have adopted data-driven BN-based models for maritime cybersecurity risk analysis. For example, Mohsendokht et al. (2024) employed BN modelling to analyse causal relationships among cybersecurity RIFs using data from the MCAD. BN-based approaches are well-suited for modelling probabilistic dependencies among cyber threats, vulnerabilities, and operational consequences. Nevertheless, the effectiveness of these models often depends on the availability of the chosen RIFs in secondary data sources. In many existing studies, the identification of RIFs and the associated data collection mainly serve the specific interests of the owners, which limits the scalability and generality of the data

Table 1
Comparison of maritime cybersecurity risk analysis approaches.

Method	Data requirement	Risk factor identification	Risk quantification	Automation level	Scalability
HAZOP (Mansoori et al., 2016)	Expert knowledge and process documentation	Manual expert identification	Semi-quantitative	Low	Limited
Attack tree analysis (Komal, 2023)	Expert-defined attack scenarios	Manual scenario construction	Semi-quantitative	Low	Limited
FMEA (Söner et al., 2023)	Expert evaluation of failure modes	Manual identification of failure causes	Semi-quantitative	Low	Limited
Domain-specific cyber risk frameworks (MaCRA, MITIGATE) (Schauer et al., 2019; Tam and Jones, 2019)	Structured system information and expert input	Scenario-based identification	Qualitative / semi-quantitative	Low	Limited
BN-based maritime cyber risk models (Mohsendokht et al., 2024; Park et al., 2023)	Incident data and expert-defined RIFs	Manual or expert-driven extraction	Probabilistic modelling	Medium	Moderate
Proposed framework (LLM + BN)	Cyber incident reports and textual data	Automated extraction using prompt-based zero-shot learning	Probabilistic modelling using BN	High	High

and may overlook emerging cyber threats.

Despite the progress achieved by existing approaches, several limitations remain. Traditional expert-driven methods have limited capability to process large-scale cybersecurity incident data and typically rely on manual identification of RIFs. Similarly, many BN-based maritime cybersecurity models depend on predefined or manually extracted RIFs, which may restrict their ability to capture emerging cyber threats embedded in unstructured cyber incident reports.

These limitations highlight the need for data-driven approaches capable of automatically extracting cybersecurity risk information from unstructured textual sources while supporting probabilistic modelling of risk dependencies. Such capabilities are particularly important for maritime transport systems, where cybersecurity incidents may disrupt vessel operations, port logistics activities, and global supply chains. To address these challenges, the present study proposes an integrated framework that combines prompt-based zero-shot learning with BN modelling, enabling automated identification of RIFs from cyber incident reports and scalable maritime cybersecurity risk analysis. Table 1 compares conventional maritime risk analysis methods with the proposed framework.

2.2. Text mining for cyber risk information extraction

In maritime cybersecurity research, a significant amount of relevant information, such as cyber incident reports, vulnerability disclosures, safety investigation reports, and operational documentation, is stored in unstructured textual formats. Extracting meaningful cybersecurity RIFs from these sources is essential for improving risk identification and assessment. However, traditional maritime cybersecurity studies often rely on manually curated datasets and RIFs, which limit the scope and scalability of risk analysis. Advances in NLP, particularly LLMs, offer new opportunities to automatically extract cybersecurity knowledge from textual data.

Text classification is one of the most widely used techniques in text mining and NLP (Minaee et al., 2021). It enables the categorisation of textual information into predefined classes, which can facilitate the identification of cybersecurity threats, vulnerabilities, and RIFs from large corpora of documents. Table 2 summarises several commonly used text classification approaches, including rule-based methods, Machine Learning (ML) models, Deep Learning (DL) techniques, pre-trained language models, and prompt-based learning methods.

Early text classification approaches primarily relied on rule-based methods, which classify documents based on manually defined rules, such as keyword matching or pattern recognition (Han et al., 2003). These methods are simple and interpretable, making them suitable for small-scale tasks with clearly defined classification criteria. However, rule-based approaches are difficult to scale and require substantial manual effort to construct and maintain rule sets. Moreover, they struggle to capture complex linguistic patterns and often fail to generalise to previously unseen textual data.

Subsequent research introduced ML-based text classification methods, including Support Vector Machines (SVM) (Keerthi et al., 2001), Random Forests (RF) (Xu et al., 2012), Naive Bayes (NB) (Rish, 2001), and Inductive Logic Programming (ILP) (Mooney, 1997). These approaches rely on feature extraction techniques such as bag-of-words and Term Frequency-Inverse Document Frequency (TF-IDF) to represent textual data numerically (Aggarwal and Zhai, 2012). While ML approaches provide improved classification performance compared with rule-based methods, they rely heavily on manual feature engineering and have limited ability to capture contextual semantics. These limitations reduce their effectiveness when analysing complex cybersecurity narratives and

Table 2
Detailed information on various text classification methods.

Method	Description	Strengths	Limitations
Rule-based methods	Early approaches relied on manually defined rules, such as keyword matching or pattern recognition, for text categorisation (Han et al., 2003).	Simple, interpretable.	Poor scalability, limited generalisation to unseen data, and time-consuming manual rule creation.
ML methods	The use of algorithms like ILP (Mooney, 1997), NB (Rish, 2001), SVM (Keerthi et al., 2001), and RF (Xu et al., 2012).	Better at capturing complex patterns than rule-based methods, automated learning from data.	Rely heavily on manual feature engineering, struggle to capture contextual semantics, have limited representational capacity, and exhibit weaker generalization.
DL methods	Introduction of CNNs (Gao et al., 2019), RNNs (Zulqarnain et al., 2020), and LSTM networks (Liu and Guo, 2019). These models learn hierarchical representations of text and capture sequential dependencies.	Automatic feature extraction from raw text, ability to handle long text sequences and complex patterns.	Requires large datasets, computationally expensive to train.
Pre-trained language models & fine-tuning	Models such as BERT (Koroteev, 2021), RoBERTa (Liu et al., 2019), T5 (Raffel et al., 2023), and LlaMA (Touvron et al., 2023), are trained on vast datasets using unsupervised learning techniques. These models generate contextualised word embeddings and can be fine-tuned for specific tasks.	Achieves state-of-the-art performance, handles context and word sense disambiguation, and can be fine-tuned for different tasks.	Requires fine-tuning, computationally intensive.
Prompt-based learning	An emerging technique that uses prompts to guide pre-trained language models to generate desired outputs without fine-tuning (Liu et al., 2023). Especially useful in few-shot or zero-shot learning scenarios.	Effective in low-data environments, reduce the need for extensive fine-tuning.	Performance is highly dependent on the design of prompts, which is still a nascent area.

domain-specific terminology.

More recently, DL models, including Convolutional Neural Networks (CNNs) (Gao et al., 2019), Recurrent Neural Networks (RNNs) (Zulqarnain et al., 2020), and Long Short-Term Memory (LSTM) (Liu and Guo, 2019), have demonstrated improved performance in text classification tasks by automatically learning hierarchical representations of textual data. These models are capable of capturing sequential dependencies and complex linguistic patterns without manual feature engineering. However, DL approaches typically require large labelled datasets for training and involve substantial computational costs. In domains such as maritime cybersecurity, where labelled datasets are scarce and cyber incidents are relatively infrequent, the practical application of DL methods remains challenging.

The emergence of LLMs has significantly transformed NLP research by enabling powerful language understanding capabilities through large-scale pre-training. Models such as Bidirectional Encoder Representations from Transformers (BERT) (Koroteev, 2021), Robustly Optimised BERT Pretraining Approach (RoBERTa) (Liu et al., 2019), Text-to-Text Transfer Transformer (T5) (Raffel et al., 2023), and Large Language Model Meta AI (LlaMA) (Touvron et al., 2023) have demonstrated state-of-the-art performance on various downstream NLP tasks using pre-trained language models. These models are pre-trained on extensive corpora using self-supervised learning and can be adapted to downstream tasks through fine-tuning or prompting strategies. More recently, advanced models such as GPT-4o (Mao et al., 2024) and Google Gemini (Imran and Almusharraf, 2024) demonstrate strong capabilities in contextual understanding, reasoning, and multi-step problem solving, enabling more accurate analysis of complex textual data. In addition, open-source models such as DeepSeek-V3.2 (Deng et al., 2025) further expand the LLM ecosystem by offering high-performance language models with improved efficiency and scalability.

Among these developments, prompt-based learning has emerged as an effective technique for leveraging pre-trained LLMs without requiring additional model training. By carefully designing prompts, LLMs can perform classification tasks in few-shot or zero-shot settings, which is particularly valuable in domains with limited labelled data (Liu et al., 2023). Various prompting strategies have been proposed to improve model performance. For example, zero-shot and few-shot prompting enable LLMs to perform classification tasks without extensive training data (Wei et al., 2022a). Additionally, Chain of Thought (CoT) prompting improves reasoning capabilities by guiding models to break down complex problems into intermediate reasoning steps (Wei et al., 2022b). Furthermore, ReAct Prompt further integrates reasoning and action to support multi-step decision-making tasks (Yao et al., 2023).

In particular, zero-shot prompting provides a promising solution for maritime cybersecurity text analysis, where labelled datasets are scarce, and new cyber threats emerge rapidly (Puri and Catanzaro, 2019; Wang et al., 2023). By leveraging the contextual understanding capabilities of LLMs, zero-shot prompting enables the automatic classification of cybersecurity-related information directly from unstructured textual data without requiring task-specific training datasets.

Although recent advances in LLMs have significantly improved text classification capabilities, their application in maritime cybersecurity risk analysis remains limited. In particular, few studies have explored the use of LLM-based methods to automatically extract cybersecurity RIFs from maritime cyber incident reports. Moreover, existing maritime cybersecurity studies often rely on manually constructed datasets rather than systematically mining unstructured textual data sources. This gap suggests significant potential for integrating prompt-based LLM techniques with cybersecurity risk modelling frameworks to enable scalable extraction and analysis of maritime cyber risk information.

2.3. Research gaps

Based on the literature review, several important gaps are revealed as follows:

(1) Limited representation of RIFs in maritime cybersecurity analysis.

Many existing studies rely on a relatively narrow set of predefined variables or expert-defined factors when modelling maritime cyber risks. Such approaches often overlook broader contextual attributes, including temporal dynamics, regional characteristics, targeted systems, attack methods, and operational consequences. As a result, current models may not fully capture the complex and multi-dimensional nature of maritime cyber incidents.

(2) Lack of scalable approaches for extracting structured cybersecurity knowledge from incident reports.

Maritime cyber incident reports contain valuable information describing attack scenarios, affected systems, and operational impacts. However, most existing studies rely on manual extraction or predefined classifications of RIFs, which limits scalability and may introduce subjective bias. Automated approaches capable of systematically extracting structured cybersecurity information from unstructured incident narratives remain underexplored in the maritime domain.

(3) Limited integration between incident factor extraction and probabilistic risk modelling.

Existing maritime cybersecurity studies often analyse incident characteristics descriptively or rely on expert-driven risk assessment frameworks. There remains a lack of integrated analytical approaches that combine systematic factor extraction with probabilistic modelling to quantify the relationships between cyber incident characteristics and their potential consequences. This limitation restricts the ability of current research to support comprehensive risk assessment and evidence-based cybersecurity management.

3. Methodology

The framework of this study combines automated extraction of RIFs from cyber incident reports with probabilistic modelling to examine the relationships between cyber incident characteristics and their potential consequences. Fig. 2 illustrates the overall analytical framework. The framework consists of three main components: dataset construction, automated RIF classification using LLMs, and probabilistic modelling using BNs.

First, maritime cyber incident reports are collected and organised into a structured dataset to support subsequent analysis. This dataset provides the empirical foundation for extracting cybersecurity-related attributes and modelling cyber risk patterns.

Second, a prompt-based zero-shot classification approach is applied to extract structured RIFs from unstructured incident reports. Based on a predefined taxonomy of RIFs, LLM-assisted prompt engineering is used to automatically classify key attributes describing maritime cyber incidents. This process enables scalable extraction of structured cybersecurity information from textual incident narratives without requiring extensive labelled training data.

Third, the extracted RIFs are incorporated into a data-driven BN model. The BN structure represents probabilistic dependencies among cyber incident characteristics and their potential consequences. Through this modelling process, the framework enables quantitative analysis of cyber risk patterns in maritime systems.

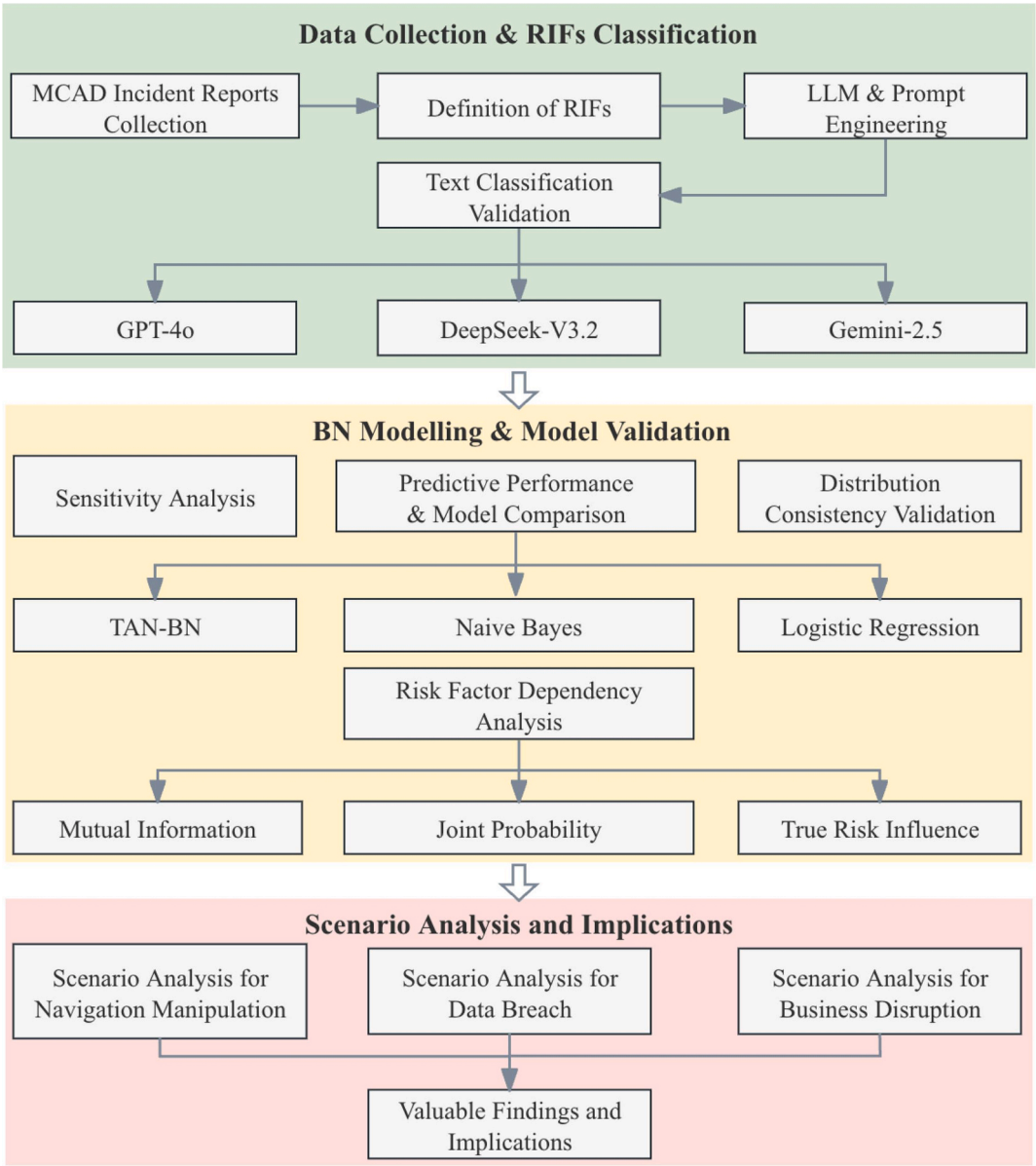

Fig. 2. The proposed framework.

The reliability of the automated classification process and the predictive capability of the BN model are evaluated in Section 4 using multiple validation approaches.

3.1. Data collection

To establish a reliable empirical foundation for maritime cybersecurity risk analysis, a systematic data collection and preparation workflow was implemented. The process consists of four stages: data source selection, data acquisition, data cleansing, and dataset consolidation. This structured pipeline ensures data quality, consistency, and suitability for subsequent risk factor extraction and probabilistic modelling.

(1) Data source selection

The primary data source used in this study is the MCAD, an open-access repository documenting cybersecurity incidents affecting the maritime sector. MCAD is currently one of the most comprehensive publicly available repositories dedicated to maritime cybersecurity incidents and has been widely referenced in recent maritime cybersecurity research (Mohsendokht et al., 2024). The database contains both structured attributes and narrative incident descriptions, which provide detailed contextual information about the attacks. This combination of structured and unstructured data makes MCAD particularly suitable for data-driven cybersecurity risk analysis and text mining applications.

(2) Data acquisition

To ensure accurate and efficient data extraction, the raw data were collected using a combination of Application Programming Interface (API) queries and automated web scraping (Demchenko et al., 2024). Structured attributes were retrieved through the MCAD list-data API. In parallel, web scraping techniques were used to extract narrative summaries from individual incident detail pages. These textual summaries provide detailed descriptions of cyberattack events and constitute the primary unstructured data source used for the subsequent prompt-based zero-shot learning process.

(3) Data cleansing.

Several data cleansing procedures were applied to improve the quality and consistency of the dataset prior to analysis. These procedures include:

a) Duplicate removal: Records containing identical incident descriptions and attributes were identified and removed.
b) Invalid entry filtering: Incidents with missing summaries or incomplete key attributes were excluded.
c) Text normalisation: Irregular formatting elements, including non-visible characters, inconsistent spacing, and encoding artefacts, were removed to ensure compatibility with subsequent LLM-based text analysis.

These steps help minimise potential noise in the textual data and improve the robustness of the downstream analytical process.

(4) Dataset consolidation

Following the cleansing process, all structured attributes and corresponding narrative descriptions were integrated into a unified dataset. The resulting dataset establishes a consistent linkage between structured variables and unstructured incident reports, enabling the systematic extraction of RIFs and supporting the subsequent BN modelling.

After data collection and preprocessing, a total of 419 maritime cyberattack incidents had been compiled by January 2026. These incidents cover a broad range of attack vectors, targeted systems, and operational impacts. The curated dataset provides a robust empirical foundation for analysing maritime cybersecurity risks and supports the integration of text mining techniques with probabilistic risk modelling.

3.2. Prompt engineering for RIF classification

3.2.1. Definition of RIFs

To enable systematic analysis of maritime cybersecurity incidents, a set of RIFs was defined to capture the temporal, geographical, attack method, and consequence impact-related characteristics of reported cyber incidents. The taxonomy was designed to reflect both general cybersecurity practices and the specific operational context of the maritime domain.

The Year Period factor represents the temporal context of each incident. Instead of using individual years, incidents were grouped into periods reflecting major technological and regulatory developments in maritime cybersecurity. The period of early digital maritime systems (2001–2016) corresponds to the initial adoption of digital navigation and communication technologies, during which cybersecurity governance remained limited. The period of cybersecurity regulatory awareness (2017–2020) reflects the growing recognition of cyber risks following the adoption of International Maritime Organization (IMO) Resolution MSC.428(98), which required cyber risk management to be incorporated into safety management systems under the international safety management

code (Hopcraft et al., 2023). The period of mandatory cyber risk management (2021–2023) corresponds to the implementation phase of these regulatory requirements. The most recent stage, cyber-secure ship design era (2024–2026), reflects the introduction of mandatory cybersecurity requirements for ship design under the International Association of Classification Societies' unified requirements, including UR E26 (cyber resilience of ships) and UR E27 (cyber resilience of onboard systems) (Progoulakis et al., 2024).

The Region factor identifies the geopolitical region where the incident occurred. Regions were defined according to maritime economic and operational groupings rather than strictly geographical continents. This approach reflects differences in maritime infrastructure, regulatory frameworks, and levels of digitalization across regions. For example, European maritime regions are characterised by advanced cybersecurity governance and highly automated port operations (Heikkilä et al., 2022), while East Asia represents a major hub of global maritime activity with significant shipbuilding capacity and container throughput (Yang et al., 2018). Other regions, such as Southeast Asia and the Middle East, represent important maritime corridors with distinct operational and infrastructural characteristics.

The Impact Area factor represents the operational environment in which the cyber incident occurred. The classification distinguishes between vessel-based systems, port infrastructure, offshore maritime facilities, and shore-based operational environments supporting maritime activities (Mohsendokht et al., 2024). This distinction reflects the distributed structure of maritime transport systems and the heterogeneous technological infrastructures involved in maritime operations.

The Targeted System factor identifies the primary technological or informational asset targeted during the incident. Categories include Information Technology (IT) systems, Operational Technology (OT) or control systems, navigation and positioning systems, and data or information assets (Li et al., 2024a). This classification reflects the layered architecture of modern maritime digital infrastructures, where enterprise IT systems, operational technologies, and navigation systems are increasingly interconnected.

The Attack Method factor captures the principal technique used in the cyber incident. The classification incorporates both conventional cyberattack mechanisms and techniques specific to maritime technologies and signal environments. This combined approach reflects the hybrid nature of maritime cyber threats, where attacks targeting IT infrastructures may coexist with signal manipulation affecting navigation or communication systems (Mohsendokht et al., 2024).

The Threat Source factor distinguishes whether the threat originates from external actors or from internal entities with authorised

Table 3
RIFs from the cyber-attack report.

Original RIFs	Description	States
Year Period	Temporal classification of maritime cyber incidents based on major technological, regulatory, and threat landscape developments in the maritime sector. Instead of individual years, incidents are grouped into historical phases to capture structural changes in maritime cybersecurity.	2001–2016 2017–2020 2021–2023 2024–2026
Region	The geopolitical region where the incident occurred. Regions are defined according to political and maritime economic groupings rather than purely geographic continental boundaries in order to capture regional differences in maritime infrastructure, governance, and cybersecurity environments.	North America Europe East Asia Southeast Asia Middle East Other
Impact Area	The operational maritime domain in which the incident occurred. This factor distinguishes different operational environments within the maritime ecosystem, reflecting where the affected activities or systems were located.	Vessel Port Offshore Onshore Operations
Targeted System	The primary digital system, technological infrastructure, or information asset that was targeted or exploited during the incident. This factor represents the category of asset within maritime digital environments that experienced the direct security compromise.	IT System OT System Navigation System Communication System Data/Information Other
Attack Method	The principal technique or mechanism used to conduct the incident. The classification integrates conventional cyberattack techniques with attack forms relevant to maritime systems and communication or navigation technologies.	DDoS Ransomware Malware Hacking Spoofing Jamming Phishing Going Dark
Threat Source	The origin of the threat actor relative to the affected organization. This factor distinguishes whether the threat originated from outside the organization or from individuals who possessed authorized access to internal systems or operational environments.	External Internal
Intent	The underlying intention associated with the incident. This factor differentiates deliberate malicious activities from accidental events resulting from human error, operational mistakes, or system misconfigurations.	Malicious Accidental
Consequence Type	The primary operational, informational, financial, geopolitical or safety-related impact resulting from the incident. This factor represents the main outcome affecting maritime operations, organizational assets, or navigation safety.	Business Disruption Data Breach Navigation Manipulation Geopolitical/Economic Impact

system access, such as employees or contractors. The Intent factor further differentiates between deliberate malicious actions and unintentional incidents resulting from operational errors, system misconfigurations, or accidental actions. Together, these factors capture both the origin of the threat actor and the underlying motivation behind the incident (Ebert et al., 2025).

Finally, the Consequence Type factor describes the primary outcome of the cyber incident. The categories capture impacts on operational continuity, data confidentiality, geopolitical tensions, financial assets, and maritime navigation safety. Including navigation-related impacts is particularly important in the maritime domain, where compromised navigation systems or signal manipulation may directly affect vessel safety and operational reliability (Yu et al., 2023).

Table 3 summarises the complete set of RIFs used in this study, the selected RIFs are derived from both the maritime cybersecurity literature and the structure of incident reports in the MCAD dataset. These factors enable a more detailed and structured representation of maritime cybersecurity incidents. By integrating temporal, operational, and technical dimensions, the proposed RIF framework provides a comprehensive basis for subsequent BN modelling and risk analysis.

3.2.2. Prompt engineering for automated text classification

Given the defined taxonomy of RIFs, automated text classification is required to systematically extract structured information from incident reports. Expert-based classification may introduce subjective bias and is difficult to scale when processing large incident datasets. Automated approaches can improve consistency and enable efficient processing of large volumes of incident reports. Therefore, this study adopts an automated classification approach based on prompt-based zero-shot learning. A key challenge is the limited availability of labelled maritime cybersecurity datasets, which restricts the applicability of conventional supervised ML or DL methods. To address this issue, LLMs were employed to perform classification without task-specific training.

Prompt engineering refers to the process of designing and structuring input prompts to effectively guide LLMs toward producing accurate and reliable outputs. It involves carefully specifying instructions, context, output formats, and constraints so that the model can better understand the task and generate responses aligned with the intended objective (Liu et al., 2023). In text classification tasks, prompt engineering is used to frame the classification problem as an instruction-based task for the model. The prompt typically defines the classification categories, provides clear task instructions, and may include examples or formatting requirements for the output. By structuring the prompt in this way, the model can analyse the input text and assign it to the appropriate category without additional model training. Effective prompt design can significantly improve classification accuracy, consistency, and interpretability, particularly when the prompt also requests supporting evidence or explanations for the predicted labels (Wang et al., 2023).

The prompt was designed to guide the LLM to classify maritime cybersecurity incident reports according to the predefined RIF taxonomy. The taxonomy includes contextual, technical, and outcome-related attributes, enabling structured extraction of information from unstructured incident narratives. First, the prompt defines the classification task and decision rules to ensure consistent outputs. The model is instructed to select labels strictly from predefined categories for each RIF, including Year Period, Region, Impact Area, Targeted System, Attack Method, Threat Source, Intent, and Consequence Type. Restricting the output space to predefined categories reduces ambiguity and ensures that the extracted labels align with the analytical framework of this study. Second, the prompt requires the model to return results in a structured JavaScript Object Notation (JSON) format. Structured outputs facilitate automatic parsing and integration into the dataset for subsequent statistical analysis and BN modelling. Enforcing a fixed output schema enables automated validation and eliminates the need for manual post-processing. To improve interpretability and support quality control, the prompt also requests an auxiliary output for each classification factor: evidence snippets. The evidence field contains short excerpts from the input report that support the assigned label, thereby enhancing traceability between the structured output and the original incident narrative. Furthermore, the model is instructed to return "Other" when an incident does not clearly match any predefined category. This constraint reduces the likelihood of unsupported assumptions or hallucinated classifications.

Overall, the prompt design balances structured output control, interpretability, and reproducibility, enabling LLMs to function as automated annotators for maritime cybersecurity incident datasets. The complete prompt templates are provided in Appendix A.

3.2.3. Text classification validation and comparison

To evaluate the reliability of the LLM-based classification process, a validation experiment was conducted using a manually annotated subset of the dataset. Specifically, 20% of the incident reports (84 records) were randomly selected and independently

Table 4
Performance comparison of LLMs in RIF classification.

RIF	DeepSeek-V3.2			Gemini-2.5-Flash			GPT-4o		
	Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Year Period	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Region	0.906	0.931	0.912	0.925	0.925	0.924	0.964	0.925	0.939
Impact Area	0.987	0.889	0.921	0.882	0.835	0.850	0.822	0.964	0.864
Targeted System	0.792	0.854	0.805	0.848	0.820	0.795	0.798	0.890	0.826
Attack Method	0.877	0.891	0.871	0.897	0.915	0.895	0.897	0.929	0.896
Threat Source	1.000	1.000	1.000	0.833	0.994	0.897	1.000	1.000	1.000
Intent	0.500	0.494	0.497	0.500	0.494	0.497	1.000	1.000	1.000
Consequence	0.743	0.653	0.683	0.756	0.726	0.729	0.855	0.744	0.746
Mean	0.851	0.839	0.836	0.830	0.839	0.823	0.917	0.932	0.909
Cohen's κ	0.787			0.758			0.922		

annotated to construct a reference dataset.

Three LLMs, GPT-4o, Gemini-2.5-Flash, and DeepSeek-V3.2, were evaluated using the same API configuration. These models were chosen because they represent advanced general-purpose LLMs with strong performance in natural language understanding and zero-shot classification tasks (Gao et al., 2026). To ensure reproducibility, the temperature parameter was set to 0, which minimises randomness and ensures deterministic classification outputs. For each RIF, model predictions were compared with the manually annotated labels. Standard classification metrics, including precision, recall, and F1-score, were calculated to quantify classification performance (Sokolova and Lapalme, 2009).

Table 4 summarises the classification performance of the evaluated LLMs across RIFs classification fields. The mean values represent the arithmetic average of precision, recall, and F1-score across the eight RIF classification fields, providing an overall comparison of classification consistency among different LLMs. GPT-4o achieves the best overall performance, with the highest mean precision (0.917), recall (0.932), and F1-score (0.909), indicating a more reliable structured classification of cybersecurity incident attributes. DeepSeek-V3.2 shows competitive performance, particularly with relatively high recall (0.851), indicating a strong ability to capture relevant entities, although with slightly lower precision. In contrast, Gemini-2.5-Flash exhibits comparatively lower performance across most metrics, with mean precision, recall, and F1-score values of 0.830, 0.839, and 0.823, respectively.

To further assess the agreement between model predictions and human annotations, Cohen's κ coefficient was calculated for each model (Landis and Koch, 1977). GPT-4o achieves the highest agreement (0.922), followed by DeepSeek-V3.2 (0.787) and Gemini-2.5-Flash (0.758). All values indicate substantial agreement with the manually annotated labels.

Considering both the classification metrics and agreement results, GPT-4o provides the most balanced performance across precision, recall, and F1-score. Therefore, GPT-4o was selected for the full-scale extraction of RIFs, and the resulting classifications were used as inputs for the subsequent BN modelling.

3.3. Data-driven BN modelling

BN provides a robust graphical framework for illustrating probabilistic dependencies between variables and applying structured reasoning and learning techniques. Various algorithms have been employed in data-driven BN modelling, each contributing unique strengths to different stages of model development and application. They include a K2 algorithm (Lerner and Malka, 2011), Markov Chain Monte Carlo (MCMC) (Magni et al., 1998), Naive Bayes Network (NBN) (Domingos and Pazzani, 1997), Augmented Bayesian Network (ABN), and Tree Augmented Naive Bayes (TAN) (Friedman et al., 1997). Unlike NBN, which assumes independence among attributes, TAN relaxes this constraint by incorporating dependencies between attributes, making the network structure a better representation of real-world conditions. In a TAN framework, each attribute is linked to both the class variable and another attribute, capturing complex interrelationships.

In the context of cybersecurity analysis, RIFs are likely to exhibit interdependencies rather than strict independence. Therefore, allowing dependencies among attributes can better reflect real-world relationships between cyberattack characteristics and their outcomes. For this reason, the TAN approach provides a suitable modelling framework for analysing the probabilistic relationships among the identified RIFs. Accordingly, this study adopts the TAN algorithm to construct the BN model. The variable "Consequence Type" is selected as the target node representing the outcome of a cyber incident. The model was implemented using Netica software, which supports structure learning and probabilistic inference. The resulting TAN network represents probabilistic dependencies among the identified RIFs and provides an interpretable framework for analysing how different cyberattack characteristics influence

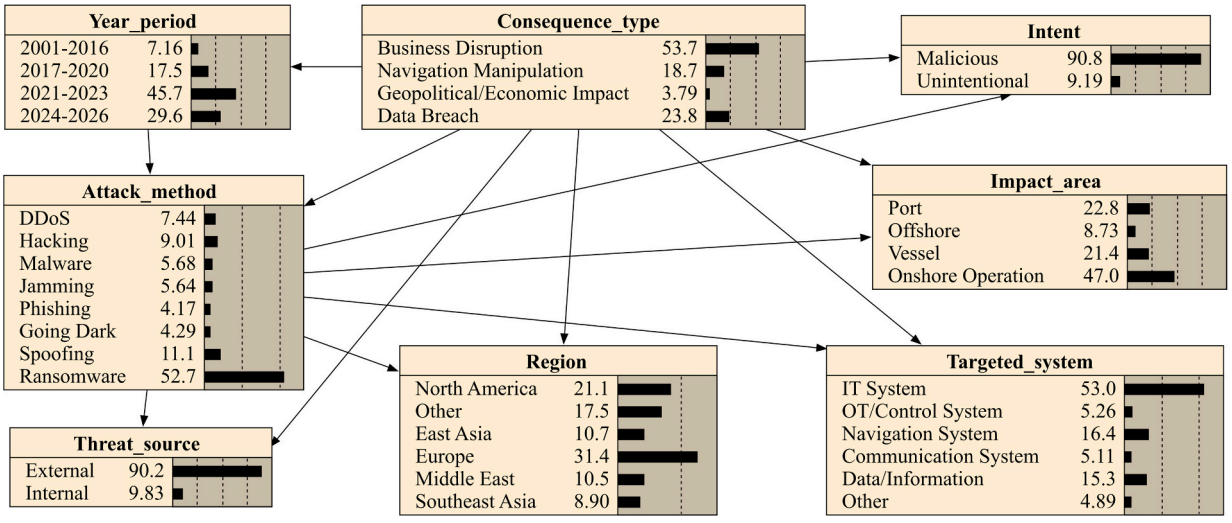

Fig. 3. TAN-BN constructed from the cybersecurity incident dataset. Nodes represent variables, directed edges indicate learned conditional dependencies rather than causal relationships, and bar charts show marginal probability distributions.

incident consequences. The learned network structure is illustrated in Fig. 3.

4. Model validation and risk analysis

To evaluate the reliability and analytical capability of the proposed BN framework for maritime cybersecurity analysis, a comprehensive validation and analysis procedure was conducted. The evaluation focuses on three aspects: predictive performance, robustness under class imbalance conditions, and comparative performance against baseline classifiers. These analyses provide evidence that the proposed model can both capture meaningful probabilistic relationships among RIFs and generate reliable predictions of cybersecurity consequences. In addition to predictive validation, further analyses, including Mutual Information (MI), joint probability, and True Risk Influence (TRI), are conducted in subsequent sections to examine the dependency structure among RIFs and to quantify their influence on maritime cybersecurity outcomes.

4.1. Predictive performance validation

The predictive capability of the proposed TAN-BN model was evaluated using a repeated cross-validation framework. Specifically, a 10-fold cross-validation repeated 10 times was applied, resulting in 100 evaluation folds (Kohavi, 1995). This strategy helps reduce the variance associated with a single train-test split and provides a more robust estimate of model performance. The classification task aims to predict the Consequence type, which represents the operational outcome of maritime cyber incidents. Model performance was evaluated using standard classification metrics, including precision, recall, F1-score, accuracy, and Cohen's κ coefficient (Landis and Koch, 1977; Sokolova and Lapalme, 2009).

The aggregated classification results obtained from all cross-validation folds are summarised in Table 5. Overall, the TAN-BN model achieved a mean accuracy of 0.824 ± 0.047 across the 100 evaluation folds. The model also achieved a Cohen's κ coefficient of 0.702 ± 0.079, indicating substantial agreement between predicted and observed consequence labels.

At the class level, the model performs well for the dominant consequence categories. In particular, Navigation Manipulation achieved the highest predictive performance, with a precision of 0.917, a recall of 0.963, and an F1-score of 0.939, indicating that the model can reliably identify incidents affecting maritime navigation systems. Business Disruption also achieved strong performance, with an F1-score of 0.875, reflecting balanced precision and recall. For the Data Breach category, the model achieved an F1-score of 0.683, suggesting moderate predictive capability. The relatively lower recall indicates that some data breach incidents are misclassified into other consequence categories.

In the dataset, the Geopolitical/Economic Impact category contains only a very limited number of samples, which leads to reduced predictive performance for this class. Such behaviour is expected in highly imbalanced datasets, as classifiers often struggle to learn reliable decision boundaries or probability distributions for extremely rare events (Krawczyk, 2016). In probabilistic models such as BNs, sparse observations may also result in unstable estimation of conditional probability tables, further limiting predictive capability for rare outcomes (Friedman et al., 1997). Nevertheless, this category was retained in the taxonomy to preserve the completeness of the maritime cybersecurity consequence framework. The objective of the proposed TAN-BN model is not solely to maximise predictive accuracy for rare events but rather to capture and analyse the probabilistic relationships among cybersecurity RIFs. Therefore, the rare category was maintained in the final model to support subsequent analysis of risk relationships within the maritime cybersecurity system.

4.2. Distribution consistency validation

Distribution consistency validation was conducted to evaluate whether the TAN model can accurately represent the probability distribution observed in historical maritime cybersecurity incident data. The marginal probability distribution of consequence types generated by the TAN model was compared with the empirical distribution derived from the original dataset. As shown in Table 6, the TAN-generated probabilities closely align with the historical distributions, with distribution consistency values exceeding 95% across all consequence categories. The small differences between the empirical and model-generated probabilities indicate that the TAN model effectively preserves the statistical characteristics of the original data and provides a reliable probabilistic representation of maritime cybersecurity consequences.

4.3. Sensitivity analysis

Table 7 presents the results of a cumulative sensitivity analysis examining the influence of incremental (+2%) perturbations across

Table 5
Overall classification performance.

Consequence	Precision	Recall	F1-score
Business Disruption	0.829	0.926	0.875
Data Breach	0.787	0.603	0.683
Geopolitical/Economic Impact	0.015	0.008	0.010
Navigation Manipulation	0.917	0.963	0.939

Table 6
Comparison between empirical and TAN-generated probability distributions.

Consequence type	Historical data (%)	TAN results (%)	Distribution consistency (%)
Business Disruption	54	53.7	99.4
Data Breach	23.8	23.8	100.0
Navigation Manipulation	18.6	18.7	99.5
Geopolitical/Economic Impact	3.6	3.79	95

Table 7
The combined influence of multiple variables.

Year period		+2%	+2%	+2%	+2%	+2%	+2%	+2%
Region			+2%	+2%	+2%	+2%	+2%	+2%
Impact Area				+2%	+2%	+2%	+2%	+2%
Targeted System					+2%	+2%	+2%	+2%
Attack Method						+2%	+2%	+2%
Threat Source							+2%	+2%
Intent								+2%
Business Disruption	53.659	54.428	54.89	55.661	57.098	58.49	58.49	58.49
Data Breach	23.848	23.868	24.117	24.511	24.643	24.672	24.672	24.672
Navigation Manipulation	18.699	18.183	17.527	16.421	14.906	13.504	13.504	13.504
Geopolitical/Economic Impact	3.794	3.521	3.466	3.407	3.353	3.334	3.334	3.334

seven model variables. The analysis evaluates how sequential parameter increases influence the proportional distribution of four consequence categories: Business Disruption, Data Breach, Navigation Manipulation, and Geopolitical/Economic Impact.

Overall, the results suggest that the model is more sensitive to operational and system-level variables than to broader contextual factors. Parameters directly related to system exposure and attack execution produce the most noticeable changes in outcome distribution, whereas higher-level contextual variables contribute comparatively limited variation.

Across all perturbation stages, Business Disruption remains the dominant risk outcome, increasing steadily from 53.659% to 58.490% as additional variables are adjusted. The most pronounced increase occurs after incorporating targeted system and attack method parameters, suggesting that system vulnerabilities and operational attack vectors strongly amplify disruption-related consequences. Once threat source and intent are introduced, the value stabilizes at 58.49%, indicating diminishing marginal influence from these final variables. This pattern suggests that operational disruption risks are primarily driven by technical exposure and attack execution characteristics rather than attacker motivation.

The Data Breach category exhibits relatively low sensitivity, increasing modestly from 23.848% to 24.672%. The gradual rise across successive perturbations indicates that data compromise risk is comparatively robust to moderate parameter changes. The limited variance implies that the data breach likelihood is more structurally embedded in the baseline model configuration rather than strongly dependent on contextual variables.

In contrast, Navigation Manipulation demonstrates a consistent downward trend, decreasing from 18.699% to 13.504% as variables are introduced. The most significant reductions occur when impact area, targeted system, and attack method parameters are incorporated. This indicates that more detailed operational parameters shift the risk distribution away from navigation manipulation toward disruption-oriented consequences. Such behaviour suggests that navigation manipulation scenarios may be overrepresented under simplified parameterisation but become less dominant when system-level attack dynamics are considered.

The Geopolitical/Economic Impact category remains relatively minor throughout the sensitivity analysis, decreasing slightly from 3.794% to 3.334%. The small magnitude and limited variation indicate that macro-level impacts are comparatively insensitive to incremental parameter perturbations within the current model structure. This suggests that geopolitical consequences are likely secondary or indirect outcomes rather than primary drivers of risk distribution in the modelled scenarios.

The sensitivity analysis indicates that the model demonstrates reasonable robustness under incremental parameter perturbations. Changes in the tested variables produce only moderate redistribution among consequence categories. While the share of Business Disruption increases and Navigation Manipulation decreases, Data Breach and Geopolitical/Economic Impact remain relatively stable. The most noticeable shifts occur when operational variables, particularly targeted system and attack method, are introduced, highlighting their stronger influence on risk outcomes compared with contextual variables such as threat source or intent. The limited magnitude of variation across perturbation stages suggests that the model structure is stable and not overly sensitive to small parameter adjustments.

4.4. Risk factor dependency analysis

4.4.1. Mutual information

MI was used to quantify the dependency between the consequence type and the identified RIFs (Li et al., 2023). It measures how much the uncertainty of the Consequence Type node is reduced when the state of another factor is known, thereby indicating the strength of probabilistic relationships within the BN.

Table 8 summarises the MI values and their relative contributions. Targeted System shows the strongest association with consequence type, accounting for 32.7% of the total MI. This indicates that the type of system targeted during a cyber incident plays a major role in determining the resulting operational consequences.

Attack Method is the second most influential factor (24.2%), suggesting that different attack techniques are closely associated with different consequence types. Impact Area contributes 14.2%, indicating that the operational environment in which the incident occurs also affects cybersecurity outcomes.

In contrast, contextual variables such as Year Period (6.07%) and Region (3.59%) have a relatively smaller influence. Threat Source (1.36%) and Intent (1.27%) contribute the least to the consequence prediction.

Overall, the results indicate that technical factors, particularly targeted systems and attack methods, play a more important role in determining maritime cyberattack consequences than contextual factors.

4.4.2. Joint probability

To examine how different factors influence the Consequence Type of maritime cybersecurity incidents, joint probabilities between the target node (Consequence Type) and related variables were calculated. Table 9 presents the joint probability distributions, illustrating how changes in the states of influencing nodes affect the likelihood of each consequence category. Higher joint probability values indicate stronger associations between specific variable states and particular consequences (Zhou et al., 2024). For clarity, the highest and lowest values for each consequence category are highlighted in bold.

The temporal analysis indicates clear shifts in dominant incident outcomes over time. During 2001–2016, Navigation Manipulation recorded the highest probability (32.639%), suggesting that early maritime cyber incidents were strongly associated with attacks affecting navigation-related systems. In 2017–2020, Business Disruption became the most probable outcome (39.717%), indicating a gradual shift towards cyber incidents affecting operational continuity and maritime service availability as digital technologies became increasingly integrated into maritime operations.

A notable change appears in 2021–2023, when Business Disruption increased significantly to 68.307%, while Navigation Manipulation and Geopolitical/Economic Impact declined to 6.816% and 1.952%, respectively. This shift indicates that recent incidents increasingly focus on disrupting operational processes rather than directly interfering with navigation systems. In the projected period 2024–2026, Business Disruption remains the most likely outcome (45.038%), although the probabilities of Data Breach (27.413%) and Navigation Manipulation (24.537%) increase, suggesting a broader distribution of potential consequences.

Regional differences are also evident. North America shows the highest probability of Business Disruption (62.533%), reflecting the high level of digitalisation in maritime logistics and port operations. East Asia records the highest probability of Navigation Manipulation (38.538%), likely related to the extensive use of advanced navigation and vessel traffic systems. In the Middle East, incidents are more likely to produce Geopolitical/Economic Impact (7.091%), which may be associated with the strategic importance of maritime routes and energy transportation in the region.

The operational context strongly influences incident outcomes. Incidents occurring in Onshore Operations and Ports are dominated by Business Disruption (64.89% and 60.797%, respectively), reflecting the reliance of port operations and logistics systems on digital infrastructure. In contrast, incidents involving Vessels show a different pattern, with Navigation Manipulation reaching 59.55%, indicating that shipborne navigation systems remain key targets. Offshore operations display a comparatively higher probability of Geopolitical/Economic Impact (8.702%), which may be linked to the strategic value of offshore energy infrastructure.

Different attack techniques are associated with distinct consequences. DDoS attacks primarily result in Business Disruption (75.989%), as they interrupt network availability. Ransomware similarly produces a high probability of operational disruption (70.149%). In contrast, Spoofing and Jamming are strongly associated with Navigation Manipulation (81.293% and 62.471%, respectively), due to their direct interference with positioning signals. Meanwhile, Phishing and Hacking significantly increase the probability of Data Breach (40.248% and 39.087%).

The type of targeted system has a substantial effect on incident outcomes. Attacks on IT Systems are most strongly associated with Business Disruption (81.771%), reflecting the central role of information systems in maritime operations. Attacks targeting Navigation Systems predominantly lead to Navigation Manipulation (79.271%). Similarly, attacks on Data/Information Systems are primarily linked to Data Breaches (75.996%), indicating the direct exposure of sensitive information when these systems are compromised.

The source and intent of cyber incidents also influence their consequences. External threats are most closely associated with Business Disruption (55.971%), whereas internal threats show relatively higher probabilities of Navigation Manipulation (32.268%) and Geopolitical/Economic Impact (11.147%). A similar pattern appears for attack intent: malicious incidents are primarily linked to

Table 8
MI between consequence type and other RIFs.

Node	Mutual Info	Percentage (%)	Variance of Beliefs
Consequence Type	1.607	100	0.412
Targeted System	0.526	32.7	0.122
Attack Method	0.389	24.2	0.064
Impact Area	0.229	14.2	0.032
Year Period	0.097	6.07	0.013
Region	0.058	3.59	0.005
Threat Source	0.022	1.36	0.002
Intent	0.02	1.27	0.002

Table 9
The joint probability.

	Business Disruption	Data Breach	Navigation Manipulation	Geopolitical/Economic Impact
Original	53.659	23.848	18.699	3.794
Year Period				
2001–2016	**29.822**	21.957	32.639	**15.582**
2017–2020	39.717	**20.995**	**34.18**	5.108
2021–2023	**68.307**	22.925	**6.816**	1.952
2024–2026	45.038	**27.413**	24.537	3.012
Region				
North America	**62.533**	**29.149**	**5.387**	2.931
Europe	55.856	25.916	16.024	**2.204**
East Asia	**39.012**	16.594	**38.538**	5.856
Southeast Asia	53.221	26.438	14.806	5.535
Middle East	48.897	**14.588**	29.424	**7.091**
Other	51.033	22.395	22.999	3.573
Impact Area				
Onshore Operation	**64.89**	**29.836**	**3.12**	**2.153**
Port	60.797	24.51	10.623	4.069
Offshore	43.595	24.394	23.309	**8.702**
Vessel	**25.551**	**9.799**	**59.55**	5.099
Attack Method				
DDOS	**75.989**	9.848	9.639	4.524
Hacking	42.91	39.087	12.469	5.534
Malware	54.985	21.175	16.28	7.56
Jamming	18.589	12.978	62.471	5.962
Phishing	26.147	**40.248**	17.189	**16.416**
Going Dark	24.985	17.057	46.068	11.89
Spoofing	**7.58**	**6.619**	**81.293**	4.507
Ransomware	70.149	27.544	**1.361**	**0.947**
Targeted System				
IT System	**81.771**	14.423	2.139	1.667
OT/Control System	47.227	21.846	21.562	**9.365**
Navigation System	**9.586**	**6.997**	**79.271**	4.147
Communication System	43.167	22.504	22.212	12.117
Data/Information	12.483	**75.996**	**7.435**	**4.086**
Threat Source				
External	**55.971**	**23.816**	**17.22**	**2.993**
Internal	**32.442**	**24.143**	**32.268**	**11.147**
Intent				
Malicious	**55.843**	**23.648**	**17.537**	**2.972**
Unintentional	**32.061**	**25.828**	**30.185**	**11.926**

Business Disruption (55.843%), while unintentional incidents show higher probabilities of Navigation Manipulation (30.185%) and Geopolitical/Economic Impact (11.926%), suggesting that operational errors or system misuse can also generate significant consequences.

The joint probability analysis indicates that operational and system-level factors, particularly attack methods and targeted systems, play a central role in shaping the consequences of maritime cybersecurity incidents.

4.4.3. True risk influence

To quantify the contribution of each RIF to variations in consequence types, the TRI metric was calculated (Li et al., 2024b). TRI measures how changes in a factor modify the posterior probability distribution of consequence types within the BN. Higher TRI values indicate a stronger influence on risk outcomes. Table 10 summarises the TRI values for all RIFs.

The results show that Targeted System, Attack Method, and Impact Area are the most influential variables. Targeted System has the highest average TRI value (28.596), followed by Attack Method (24.679) and Impact Area (15.294). These variables therefore account

Table 10
TRI of RIFs for "Consequence types".

	Business Disruption	Data Breach	Navigation Manipulation	Geopolitical/Economic Impact	Average
Year Period	19.243	3.209	13.682	6.815	10.737
Region	11.760	7.280	16.576	2.444	9.515
Impact Area	19.669	10.019	28.215	3.275	15.294
Attack Method	34.205	16.815	**39.966**	**7.735**	24.679
Targeted System	**36.093**	**34.499**	38.566	5.225	**28.596**
Threat Source	11.765	0.164	7.524	4.077	5.882
Intent	11.891	1.090	6.324	4.477	5.946

for most of the variation in consequence types.

The Targeted System variable shows consistently high TRI values across multiple consequences, including Business Disruption (36.093), Data Breach (34.499), and Navigation Manipulation (38.566). This indicates that the type of system affected largely determines the resulting consequence.

Attack Method also shows strong influence, particularly for Navigation Manipulation (39.966) and Business Disruption (34.205). This reflects the close relationship between specific attack techniques and operational impacts. For example, signal interference attacks mainly affect navigation systems, while attacks such as DDoS or ransomware disrupt operational processes.

The Impact Area variable has a notable effect on Navigation Manipulation (28.215), suggesting that the operational environment (e.g., vessels, ports, or offshore facilities) influences how cyber incidents propagate.

Year Period and Region show moderate influence, with average TRI values of 10.737 and 9.515, respectively. Year Period has relatively stronger effects on Business Disruption (19.243) and Navigation Manipulation (13.682), indicating temporal changes in dominant incident outcomes. Region shows its strongest influence on Navigation Manipulation (16.576), suggesting regional differences in infrastructure and operational conditions.

In contrast, Threat Source and Intent have relatively low TRI values, averaging 5.882 and 5.946. For example, the TRI value of Threat Source for Data Breach is only 0.164, indicating that whether an attack originates internally or externally has a limited impact on this outcome. Similarly, the low TRI values for Intent suggest that the nature of the incident (malicious or unintentional) has little direct effect on the type of consequence.

5. Scenario analysis and implications

5.1. Scenario analysis for navigation manipulation

To explore the risk patterns associated with navigation manipulation incidents, a scenario analysis was conducted by setting "Navigation Manipulation" as the target consequence in the BN. The posterior probability distributions of the related RIFs are shown in Fig. 4.

The results indicate that navigation manipulation incidents are strongly associated with navigation-related systems, which account for 69.7% of the targeted systems. In contrast, IT systems, OT/control systems, communication systems, and data assets each contribute only 6.07%, highlighting that attacks in this scenario primarily target maritime navigation infrastructures.

The operational environment of these incidents is also highly concentrated. Vessels account for 68.3% of the affected impact areas, indicating that navigation manipulation mainly occurs at the ship level rather than in ports, offshore platforms, or shore-based operations. This result reflects the vulnerability of onboard navigation technologies such as Global Navigation Satellite Systems (GNSS), Automatic Identification Systems (AIS), and other positioning systems, which are critical for safe vessel navigation.

From the perspective of attack techniques, spoofing dominates the attack methods (48.1%), followed by jamming (18.9%) and signal suppression techniques such as going dark (10.6%). These methods directly manipulate or disrupt navigation signals, enabling attackers to mislead vessels regarding their true location or operational status.

In terms of temporal and geographical characteristics, the probability of navigation manipulation incidents is highest in the 2024–2026 period (38.9%), suggesting that such threats are becoming more prominent in recent years as maritime digitalisation increases. Regionally, Europe (26.9%) and East Asia (22.0%) show relatively higher probabilities, reflecting the concentration of advanced maritime infrastructure and intensive shipping activities in these regions.

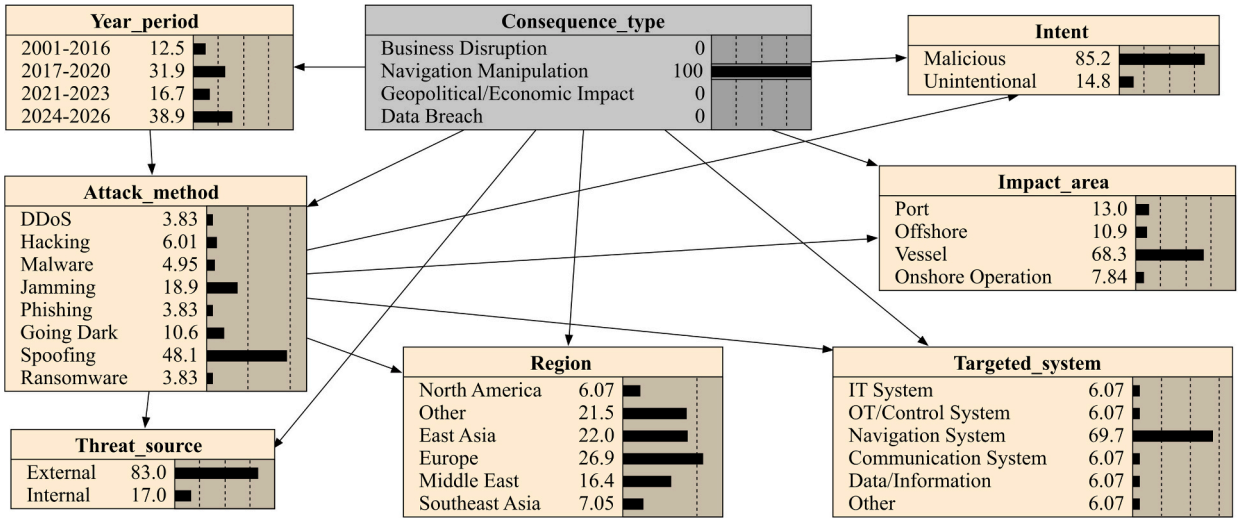

Fig. 4. Posterior probability distributions of RIFs under the Navigation Manipulation scenario.

Finally, the analysis indicates that external actors account for 83.0% of incidents, while malicious intent dominates (85.2%), suggesting that navigation manipulation attacks are largely deliberate cyber operations rather than accidental events.

Overall, the analysis reveals that navigation manipulation attacks are primarily characterised by external malicious actors targeting onboard navigation systems through signal manipulation techniques such as spoofing and jamming. These findings highlight the importance of strengthening the resilience of maritime navigation technologies, including multi-source positioning systems, signal authentication mechanisms, and enhanced monitoring of navigation anomalies.

5.2. Scenario analysis for data breach

The posterior probability distributions of the related RIFs about "Data Breach" as the consequence node are presented in Fig. 5.

The results indicate that data breach incidents are primarily associated with data and information assets, which account for 48.6% of the targeted systems. This represents a substantial increase compared with the baseline distribution and confirms that the theft or exposure of sensitive digital information is the defining characteristic of this scenario. In addition, IT systems (32.1%) also show a relatively high probability, suggesting that enterprise information infrastructures remain key entry points for data-oriented cyber incidents.

From the perspective of the operational environment, onshore operations account for 58.8% of incidents, indicating that data breaches mainly occur within shore-based digital infrastructures such as shipping company information systems, logistics platforms, and port management databases. Compared with vessel-based systems, shore-side information infrastructures typically store larger volumes of operational and commercial data, making them attractive targets for attackers.

The analysis of attack methods further highlights the dominant role of ransomware attacks, which account for 60.9% of the observed probability in this scenario. This result reflects the growing prevalence of ransomware campaigns targeting maritime organisations, where attackers encrypt or exfiltrate sensitive data to extort financial payments.

Temporal analysis indicates that the likelihood of data breach incidents increases significantly in the 2024–2026 period (34.1%), suggesting a continuing rise in cyber threats associated with digital information systems as maritime operations become increasingly data-driven.

Geographically, Europe (34.1%) and North America (25.7%) exhibit the highest probabilities in this scenario. These regions host highly digitalised maritime infrastructure and large shipping companies, which may increase both the exposure to cyber risks and the reporting of cybersecurity incidents.

Finally, the analysis shows that external actors account for 90.1% of incidents, while malicious intent dominates (90.1%), indicating that data breaches are overwhelmingly deliberate cyberattacks rather than accidental events.

This scenario analysis demonstrates that maritime data breach incidents are primarily characterised by external malicious actors targeting shore-based information infrastructures through ransomware attacks. These findings highlight the importance of strengthening data protection strategies in maritime organisations, including enhanced access control, encryption mechanisms, and proactive monitoring of corporate information systems.

5.3. Scenario analysis for business disruption

To examine the conditions under which business disruption becomes the dominant consequence, a scenario analysis was conducted using the BN. In this scenario, the following states were fixed: Targeted System = IT System, Attack Method = Ransomware, Region =

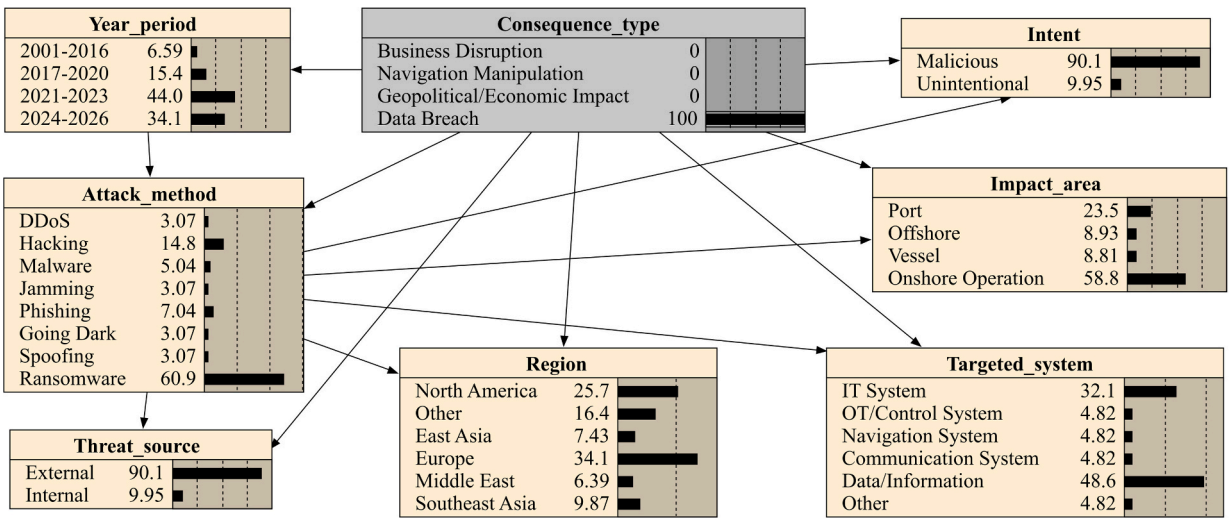

Fig. 5. Posterior probability distributions of RIFs under the Data Breach scenario.

Middle East, Impact Area=Onshore Operation, Threat Source = External, Year Period = 2021–2023, and Intent = Malicious. The resulting posterior distributions are shown in Fig. 6.

Under this configuration, the probability of Business Disruption increases to 95.0%, while the probabilities of other consequence types remain negligible. This result indicates that ransomware incidents targeting shore-based IT infrastructures are strongly associated with operational disruption in maritime organisations.

The scenario also shows that onshore operations account for the entire impact area, suggesting that incidents affecting corporate information systems are more likely to result in organisational disruptions than incidents targeting vessel or offshore systems. Shore-based IT systems support core functions such as scheduling, cargo management, logistics coordination, and administrative operations. Disruption of these systems may therefore interrupt multiple operational processes simultaneously.

The results further indicate that the incidents in this scenario originate from external actors with malicious intent, which is consistent with the typical characteristics of ransomware-related cyber incidents.

The scenario analysis suggests that ransomware attacks targeting shore-based IT infrastructures represent a critical pathway leading to business disruption in maritime organisations.

5.4. Implications for maritime cybersecurity enhancement

The proposed framework translates unstructured cyber incident reports into quantified risk relationships and scenario-based insights. Beyond the provision of general risk rankings, the analysis identifies combinations of targeted systems, attack methods, and operational contexts that are most strongly associated with specific consequence types. These results provide a basis for prioritising cybersecurity measures in maritime operations.

(1) Protection of navigation-related systems.

The scenario analysis for navigation manipulation indicates that incidents are strongly associated with attacks targeting navigation systems on vessels, particularly through signal manipulation techniques such as spoofing and jamming. These attacks exploit vulnerabilities in technologies such as the GNSS, GPS, and AIS.

Improving the resilience of navigation systems therefore represents an important priority. Potential measures include the use of multi-source positioning systems, signal authentication mechanisms, and continuous monitoring of navigation data anomalies. Strengthening the robustness of onboard navigation systems can reduce the risk of manipulation of vessel positioning and navigation information (Androjna et al., 2020).

(2) Protection of maritime data and information assets.

The data breach scenario highlights the role of data and information assets as primary targets of cyber incidents. These incidents are concentrated in shore-based operational environments, where corporate information systems store large volumes of operational and commercial data.

Protective measures should therefore focus on strengthening data security within maritime information infrastructures. Relevant approaches include improved access control mechanisms, encryption of sensitive data, and monitoring of abnormal data access patterns. In addition, incident detection and forensic analysis capabilities can support faster identification and investigation of

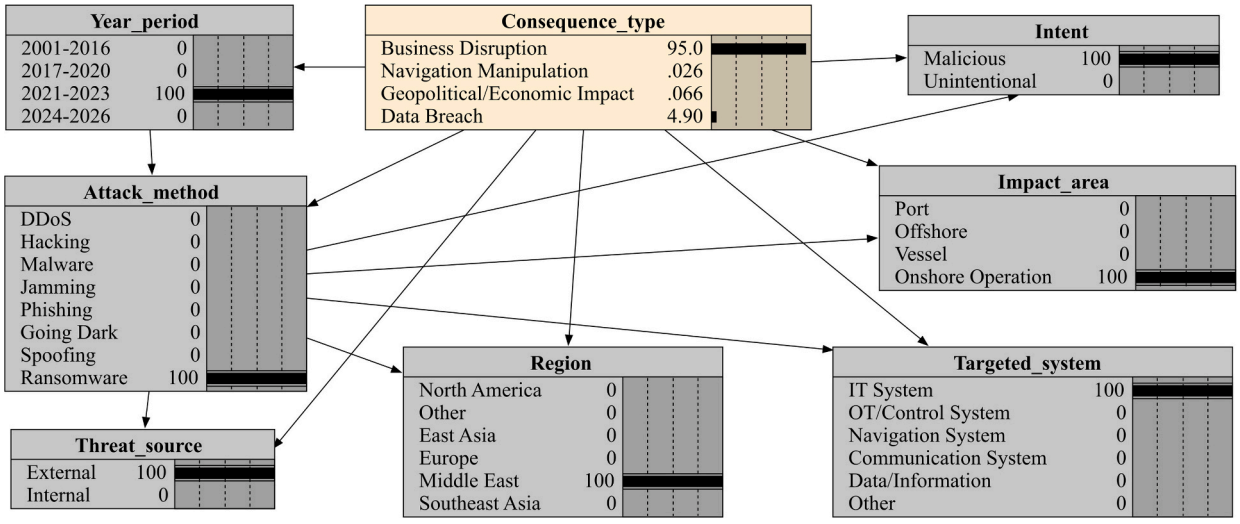

Fig. 6. Scenario analysis for the Business Disruption.

potential breaches (Costante et al., 2016; Salem et al., 2024).

(3) Strengthening shore-side IT infrastructure security.

The business disruption scenario indicates that ransomware attacks targeting shore-based IT systems are strongly associated with operational disruption. These systems support core functions such as fleet management, cargo scheduling, logistics coordination, and administrative operations. Disruption of these systems may therefore propagate across multiple operational processes.

Cybersecurity measures should therefore prioritise the protection of shore-side enterprise information systems. Relevant strategies include regular vulnerability assessment, network monitoring, and improved backup and recovery mechanisms to ensure operational continuity. In addition, employee awareness and basic cybersecurity practices remain important for reducing exposure to common attack vectors such as phishing and malware distribution (Alwashali et al., 2021).

The analysis also provides guidance for prioritising cybersecurity investments. The strong influence of targeted systems in the BN model suggests that protective efforts should focus primarily on critical asset classes, including navigation systems, corporate IT infrastructures, and data repositories. Scenario analyses further indicate that certain combinations of factors, such as ransomware targeting shore-based IT systems, are particularly likely to lead to severe operational consequences.

These findings may support the development of risk-based cybersecurity management strategies in maritime organisations and regulatory frameworks that prioritise high-impact systems and attack pathways.

6. Conclusions

This study develops an integrated framework that combines LLM-assisted text classification with BN modelling to analyse maritime cyber incidents and their potential consequences. By transforming unstructured cyber incident reports into structured analytical variables, the framework enables probabilistic modelling of relationships between cyberattack characteristics, operational contexts, and consequence types.

The analysis provides several insights into the structure of maritime cyber risks. Technical factors, particularly targeted systems and attack methods, exhibit stronger associations with incident consequences than broader contextual variables such as region or time period. Scenario analysis further illustrates that different consequence types are associated with distinct risk configurations. For example, navigation manipulation incidents are closely linked to attacks targeting navigation systems on vessels, while data breaches are primarily associated with shore-based information infrastructures. Business disruption is strongly connected to ransomware incidents affecting enterprise IT systems supporting maritime operations.

From a methodological perspective, the study demonstrates how LLM-assisted information extraction can support the construction of structured cybersecurity datasets from incident narratives, while BN modelling provides a transparent representation of probabilistic dependencies among cyber risk factors. This combination offers a practical approach for analysing cybersecurity risks in domains where structured incident data remains limited.

Several limitations should be noted. The classification relies on proprietary LLMs accessed via commercial APIs. Compared with locally deployed open-source models, these systems provide limited transparency and control over the underlying model architecture and training data. In addition, the analysis relies primarily on the MCAD, which represents the most comprehensive publicly available source but may not capture the full spectrum of maritime cyber incidents.

Future research may extend this work by incorporating additional incident datasets, exploring open-source LLMs for automated information extraction, and integrating heterogeneous cybersecurity data sources. These developments would further improve the empirical basis for maritime cybersecurity risk modelling and support more comprehensive analysis of cyber risks in maritime systems.

CRediT authorship contribution statement

Yunfeng Zhao: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Data curation, Conceptualization. **Huanhuan Li:** Writing – review & editing, Visualization, Validation, Supervision, Software, Project administration, Methodology, Conceptualization. **Trung Thanh Nguyen:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Christian Matthews:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Zaili Yang:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization.

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 Horizon 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.

Acknowledgement

This research was funded by a European Research Council project under the European Union's Horizon 2020 research and

innovation programme (TRUST CoG 2019 864724) and by LJMU Freeport Thematic Doctoral Pathways programme.

Appendix A

You are a maritime cybersecurity incident classifier. Read the incident report and classify it using the taxonomy below. Return ONLY valid JSON. Do not include explanations or markdown.

Rules:
1. Use only the allowed values exactly as written.
2. Classify strictly based on the information disclosed in the report, and do not make any assumptions or inferences beyond what is stated.
3. If an incident exists but does not match any category, output "Other".
4. Choose the primary category if multiple are mentioned.
5. Provide short evidence snippets copied from the text.

TAXONOMY
Year_period:
- 2001–2016
- 2017–2020
- 2021–2023
- 2024–2026

Region (incident location):
- North America
- Europe
- East Asia
- Southeast Asia
- Middle East
- Other

Impact_area:
- Onshore Operation
- Port
- Offshore
- Vessel
- Other

Targeted_system:
- IT System
- OT/Control System
- Navigation System
- Communication System
- Data/Information
- Other

Attack_method:
- DDoS
- Ransomware
- Malware
- Hacking
- Spoofing
- Jamming
- Phishing
- Going Dark
- Other

Threat_source:
- External
- Internal

Intent:
- Malicious
- Unintentional

Consequence_type:
- Business Disruption: Cyber incident that interrupts or degrades normal operations of ships, ports, logistics, or company systems.
- Data Breach: Unauthorized access, exposure, or theft of sensitive digital information such as operational, commercial, or personal data.
- Navigation Manipulation: Tampering with or falsifying navigation-related data or systems (GPS, AIS) that affects vessel positioning or tracking.
- Geopolitical/Economic Impact: Cybersecurity incidents that cause financial losses or cargo theft, or that escalate interstate tensions, carry military or national security implications, or contribute to wider geopolitical conflicts.
- Other

OUTPUT JSON FORMAT:
{
"Year_period": {"value": "", "evidence": [""]},
"Region": {"value": "", "evidence": [""]},
"Impact_area": {"value": "", "evidence": [""]},

(*continued on next page*)

(*continued*)

You are a maritime cybersecurity incident classifier. Read the incident report and classify it using the taxonomy below. Return ONLY valid JSON. Do not include explanations or markdown.

“Targeted_system”: {“value”: “”, “evidence”: [“”]},
“Attack_method”: {“value”: “”, “evidence”: [“”]},
“Threat_source”: {“value”: “”, “evidence”: [“”]},
“Intent”: {“value”: “”, “evidence”: [“”]},
“Consequence_type”: {“value”: “”, “evidence”: [“”]}
}
INCIDENT REPORT:
≪<≪
{{INCIDENT_REPORT}}
≫>≫

Data availability

Data will be made available on request.

References

Aggarwal, C.C., Zhai, C., 2012. A survey of text classification algorithms. In: Aggarwal, C.C., Zhai, C. (Eds.), Mining Text Data. Springer, US, Boston, MA, pp. 163–222. https://doi.org/10.1007/978-1-4614-3223-4_6.

Alwashali, A.A.M.A., Rahman, N.A.A., Ismail, N., 2021. A survey of ransomware as a service (RaaS) and methods to mitigate the attack. In: 2021 14th International Conference on Developments in eSystems Engineering (DeSE), pp. 92–96. https://doi.org/10.1109/DeSE54285.2021.9719456.

Androjna, A., Brcko, T., Pavic, I., Greidanus, H., 2020. Assessing cyber challenges of maritime navigation. J. Mar. Sci. Eng. 8, 776. https://doi.org/10.3390/jmse8100776.

Ashraf, I., Park, Y., Hur, S., Kim, S.W., Alroobaea, R., Zikria, Y.B., Nosheen, S., 2023. A survey on cyber security threats in IoT-enabled maritime industry. IEEE Trans. Intell. Transp. Syst. 24, 2677–2690. https://doi.org/10.1109/TITS.2022.3164678.

Costante, E., Fauri, D., Etalle, S., den Hartog, J., Zannone, N., 2016. A hybrid framework for data loss prevention and detection. In: 2016 IEEE Security and Privacy Workshops (SPW), pp. 324–333. https://doi.org/10.1109/SPW.2016.24.

Demchenko, Y., Cuadrado-Gallego, J.J., Chertov, O., Aleksandrova, M., 2024. Finding data on the web, data sets, web scraping, web API. In: Demchenko, Y., Cuadrado-Gallego, J.J., Chertov, O., Aleksandrova, M. (Eds.), Big Data Infrastructure Technologies for Data Analytics: Scaling Data Science Applications for Continuous Growth. Springer Nature Switzerland, Cham, pp. 417–446. https://doi.org/10.1007/978-3-031-69366-3_11.

Deng, Z., Ma, W., Han, Q.-L., Zhou, W., Zhu, X., Wen, S., Xiang, Y., 2025. Exploring deepseek: a survey on advances, applications, challenges and future directions. IEEE/CAA J. Autom. Sin. 12, 872–893. https://doi.org/10.1109/JAS.2025.125498.

Domingos, P., Pazzani, M., 1997. On the optimality of the simple bayesian classifier under zero-one loss. Mach. Learn. 29, 103–130. https://doi.org/10.1023/A:1007413511361.

Ebert, N., Schaltegger, T., Ambuehl, B., Geppert, T., Trammell, A., Knieps, M., Zimmermann, V., 2025. Learning from safety science: designing incident reporting systems in cybersecurity. J. Cyber Secur. 11, tyaf019. https://doi.org/10.1093/cybsec/tyaf019.

Erbas, M., Khalil, S.M., Tsiopoulos, L., 2024. Systematic literature review of threat modeling and risk assessment in ship cybersecurity. Ocean Eng. 306, 118059. https://doi.org/10.1016/j.oceaneng.2024.118059.

Friedman, N., Geiger, D., Goldszmidt, M., 1997. Bayesian network classifiers. Mach. Learn. 29, 131–163. https://doi.org/10.1023/A:1007465528199.

Gao, M., Li, T., Huang, P., 2019. Text classification research based on improved Word2vec and CNN. In: Liu, X., Mrissa, M., Zhang, L., Benslimane, D., Ghose, A., Wang, Z., Bucchiarone, A., Zhang, W., Zou, Y., Yu, Q. (Eds.), Service-Oriented Computing – ICSOC 2018 Workshops. Springer International Publishing, Cham, pp. 126–135. https://doi.org/10.1007/978-3-030-17642-6_11.

Gao, T., Jin, J., Ke, Z.T., Moryoussef, G., 2026. A comparison of deepseek and other LLMs. The American Statistician 80, 164–176. https://doi.org/10.1080/00031305.2025.2611010.

Gunes, B., Kayisoglu, G., Bolat, P., 2021. Cyber security risk assessment for seaports: a case study of a container port. Comput. Secur. 103, 102196. https://doi.org/10.1016/j.cose.2021.102196.

Han, H., Manavoglu, E., Giles, C.L., Zha, H., 2003. Rule-based word clustering for text classification. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval. pp. 445–446. https://doi.org/10.1145/860435.860543.

Heikkilä, M., Saarni, J., Saurama, A., 2022. Innovation in smart ports: future directions of digitalization in container ports. J. Mar. Sci. Eng. 10. https://doi.org/10.3390/jmse10121925.

Hopcraft, R., Tam, K., Misas, J.D.P., Moara-Nkwe, K., Jones, K., 2023. Developing a maritime cyber safety culture: improving safety of operations. Maritime Technol. Res. 5, 258750. https://doi.org/10.33175/mtr.2023.258750.

Iaiani, M., Tugnoli, A., Bonvicini, S., Cozzani, V., 2021. Analysis of cybersecurity-related incidents in the process industry. Reliab. Eng. Syst. Saf. 209, 107485. https://doi.org/10.1016/j.ress.2021.107485.

Imran, M., Almusharraf, N., 2024. Google Gemini as a next generation AI educational tool: a review of emerging educational technology. Smart Learn. Environ. 11, 22. https://doi.org/10.1186/s40561-024-00310-z.

Keerthi, S.S., Shevade, S.K., Bhattacharyya, C., Murthy, K.R.K., 2001. Improvements to Platt’s SMO algorithm for SVM classifier design. Neural Comput. 13, 637–649.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Ijcai. Montreal, Canada, pp. 1137–1145.

Komal, 2023. Fuzzy attack tree analysis of security threat assessment in an internet security system using algebraic *t*-norm and *t*-conorm, in: Garg, H., Ram, M. (Eds.), Engineering Reliability and Risk Assessment, Advances in Reliability Science. Elsevier, pp. 53–64. Doi: 10.1016/B978-0-323-91943-2.00003-4.

Koroteev, M.V., 2021. BERT: a review of applications in natural language processing and understanding. Doi: 10.48550/arXiv.2103.11943.

Krawczyk, B., 2016. Learning from imbalanced data: open challenges and future directions. Prog. Artif. Intell. 5, 221–232. https://doi.org/10.1007/s13748-016-0094-0.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. biometrics 159–174.

Lerner, B., Malka, R., 2011. Investigation of the K2 algorithm in learning bayesian network classifiers. Appl. Artif. Intell. 25, 74–96. https://doi.org/10.1080/08839514.2011.529265.

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

Li, H., Ren, X., Yang, Z., 2023. Data-driven Bayesian network for risk analysis of global maritime accidents. Reliab. Eng. Syst. Saf. 230, 108938. https://doi.org/10.1016/j.ress.2022.108938.
Li, M., Zhou, J., Chattopadhyay, S., Goh, M., 2024b. Maritime cybersecurity: a comprehensive. Review. https://doi.org/10.48550/arXiv.2409.11417.
Liu, G., Guo, J., 2019. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing 337, 325–338. https://doi.org/10.1016/j.neucom.2019.01.078.
Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., Neubig, G., 2023. Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. ACM Comput. Surv. 55, 1–35. https://doi.org/10.1145/3560815.
Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019. RoBERTa: A robustly optimized BERT pretraining approach. Doi: 10.48550/arXiv.1907.11692.
Magni, P., Bellazzi, R., De Nicolao, G., 1998. Bayesian function learning using MCMC methods. IEEE Trans. Pattern Anal. Mach. Intell. 20, 1319–1331. https://doi.org/10.1109/34.735805.
Mansoori, M., Welch, I., Choo, K.-K.-R., Maxion, R.A., 2016. Application of HAZOP to the design of cyber security experiments. In: 2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), pp. 790–799. https://doi.org/10.1109/AINA.2016.115.
Mao, R., Chen, G., Zhang, X., Guerin, F., Cambria, E., 2024. GPTEval: A survey on assessments of ChatGPT and GPT-4. Doi: 10.48550/arXiv.2308.12488.
Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., Gao, J., 2021. Deep Learning–based Text Classification: A Comprehensive Review. ACM Comput. Surv. 54, 62:1-62:40. Doi: 10.1145/3439726.
Mohsendokht, M., Li, H., Kontovas, C., Chang, C.-H., Qu, Z., Yang, Z., 2024. Decoding dependencies among the risk factors influencing maritime cybersecurity: lessons learned from historical incidents in the past two decades. Ocean Eng. 312, 119078. https://doi.org/10.1016/j.oceaneng.2024.119078.
Mooney, R.J., 1997. Inductive logic programming for natural language processing, in: Muggleton, S. (Ed.), Inductive Logic Programming. Springer, Berlin, Heidelberg, pp. 1–22. Doi: 10.1007/3-540-63494-0_45.
Park, C., Kontovas, C., Yang, Z., Chang, C.-H., 2023. A BN driven FMEA approach to assess maritime cybersecurity risks. Ocean Coast. Manag. 235, 106480. https://doi.org/10.1016/j.ocecoaman.2023.106480.
Peng, P., Xie, X., Claramunt, C., Lu, F., Gong, F., Yan, R., 2025. Bibliometric analysis of maritime cybersecurity: Research status, focus, and perspectives. Transp. Res. Part E: Logist. Transp. Rev. 195, 103971. https://doi.org/10.1016/j.tre.2025.103971.
Pijpker, J., McCombie, S., Johnson, S., Loves, R., Makrakis, G.M., 2024. an open-source database of cyberattacks on the maritime transportation system.
Progoulakis, I., Dagkinis, I.K., Dimakopoulou, A., Lilas, T., Nikitakos, N., Psomas, P.M., 2024. Cyber–physical security assessment for maritime vessels: study on drillship DP system using american petroleum institute security risk analysis and bow-tie analysis. J. Mar. Sci. Eng. 12. https://doi.org/10.3390/jmse12101757.
Puri, R., Catanzaro, B., 2019. Zero-shot text classification with generative language models. Doi: 10.48550/arXiv.1912.10165.
Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J., 2023. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Doi: 10.48550/arXiv.1910.10683.
Rish, I., 2001. An empirical study of the naive Bayes classifier, in: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence. Citeseer, pp. 41–46.
Salem, A.H., Azzam, S.M., Emam, O.E., Abohany, A.A., 2024. Advancing cybersecurity: a comprehensive review of AI-driven detection techniques. J. Big Data 11, 105. https://doi.org/10.1186/s40537-024-00957-y.
Schauer, S., Polemi, N., Mouratidis, H., 2019. MITIGATE: a dynamic supply chain cyber risk assessment methodology. J. Transp. Secur. 12, 1–35. https://doi.org/10.1007/s12198-018-0195-z.
Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. 45, 427–437. https://doi.org/10.1016/j.ipm.2009.03.002.
Söner, Ö., Kayisoglu, G., Bolat, P., Tam, K., 2023. Cybersecurity risk assessment of VDR. The Journal of Navigation 76, 20–37.
Tam, K., Jones, K., 2019. MaCRA: a model-based framework for maritime cyber-risk assessment. WMU J. Marit. Aff. 18, 129–163.
Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., Lample, G., 2023. LLaMA: Open and efficient foundation language models. Doi: 10.48550/arXiv.2302.13971.
Trade, U.N., 2024. Review of maritime transport 2024: navigating maritime chokepoints. United Nations.
Uflaz, E., Sezer, S.I., Tunçel, A.L., Aydin, M., Akyuz, E., Arslan, O., 2024. Quantifying potential cyber-attack risks in maritime transportation under Dempster–Shafer theory FMECA and rule-based Bayesian network modelling. Reliab. Eng. Syst. Saf. 243, 109825. https://doi.org/10.1016/j.ress.2023.109825.
Wang, Z., Pang, Y., Lin, Y., 2023. Large language models are zero-shot text classifiers. Doi: 10.48550/arXiv.2312.01044.
Wei, J., Bosma, M., Zhao, V.Y., Guu, K., Yu, A.W., Lester, B., Du, N., Dai, A.M., Le, Q.V., 2022a. Finetuned language models are zero-shot learners. Doi: 10.48550/arXiv.2109.01652.
Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q.V., Zhou, D., 2022b. Chain-of-thought prompting elicits reasoning in large language models. Adv. Neural Inf. Proces. Syst. 35, 24824–24837.
Wu, H., Li, M., Yu, C., Ouyang, Z., Lai, K., Zhao, Z., Pan, S., Wang, S., Zhong, R.Y., Kuo, Y.-H., Zhang, F., Huang, W., Shen, Z.-J.-M., Ballot, E., Huang, G.Q., 2025. Towards cyber-physical internet: a systematic review, fundamental model and future perspectives. Transp. Res. Part E: Logist. Transp. Rev. 197, 104051. https://doi.org/10.1016/j.tre.2025.104051.
Xu, B., Guo, X., Ye, Y., Cheng, J., 2012. An improved random forest classifier for text categorization. J. Comput. 7, 2913–2920.
Yang, C.-S., 2019. Maritime shipping digitalization: Blockchain-based technology applications, future improvements, and intention to use. Transp. Res. Part E: Logist. Transp. Rev. 131, 108–117. https://doi.org/10.1016/j.tre.2019.09.020.
Yang, Y., Zhong, M., Yao, H., Yu, F., Fu, X., Postolache, O., 2018. Internet of things for smart ports: Technologies and challenges. IEEE Instrum. Meas. Magaz. 21, 34–43. https://doi.org/10.1109/MIM.2018.8278808.
Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., Cao, Y., 2023. ReAct: synergizing reasoning and acting in language models. Doi: 10.48550/arXiv.2210.03629.
Yoo, Y., Park, H.-S., 2021. Qualitative Risk assessment of cybersecurity and development of vulnerability enhancement plans in consideration of digitalized ship. J. Mar. Sci. Eng. 9, 565. https://doi.org/10.3390/jmse9060565.
Yu, H., Meng, Q., Fang, Z., Liu, J., 2023. Literature review on maritime cybersecurity: state-of-the-art. The Journal of Navigation 76, 453–466. https://doi.org/10.1017/S0373463323000164.
Zhang, X., Gan, L., Cui, H., Shu, Y., Montewka, J., Yang, Z., 2026. A hybrid deep learning and large language models framework for ship collision accident analysis. Reliab. Eng. Syst. Saf. 273, 112333. https://doi.org/10.1016/j.ress.2026.112333.
Zhou, K., Xing, W., Wang, J., Li, H., Yang, Z., 2024. A data-driven risk model for maritime casualty analysis: a global perspective. Reliab. Eng. Syst. Saf. 244, 109925. https://doi.org/10.1016/j.ress.2023.109925.
Zulqarnain, M., Ghazali, R., Hassim, Y.M.M., Rehan, M., 2020. A comparative review on deep learning models for text classification. Indones. J. Electr. Eng. Comput. Sci 19, 325–335.