Reliability Engineering and System Safety 273 (2026) 112358

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Quantitative assessment of language and cultural diversity as human reliability factors in maritime safety

Rafi Ullah Khan [a,*], Jingbo Yin [b], Muhammad Afzaal [c], Zaili Yang [a,*]

[a] Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, L33AF Liverpool, United Kingdom
[b] State Key Laboratory of Ocean Engineering, Department of Transportation Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, 200240 Shanghai, China
[c] Institute of Language Sciences, Shanghai International Studies University, Shanghai, China

ARTICLE INFO

Keywords:
Ship crew
Language diversity
Cultural diversity
Human reliability
Miscommunication
Risk assessment
Maritime safety

ABSTRACT

Reliable shipboard operations depend on effective communication and situational awareness among multinational crews. Language barriers and cultural diversity, if unmanaged, can degrade operational reliability and compromise safety-critical decision-making. Yet quantitative evidence on how language and cultural diversity shape accident risk remains limited. This study applies a Bayesian Network (BN) framework to assess the reliability and safety implications of language and cultural diversity using 550 maritime accident investigations. Natural Language Processing (NLP) extracted indicators, including crew nationality composition, language diversity (Shannon Index), cultural diversity (Hofstede Index), and adherence to Standard Maritime Communication Phrases (SMCP), which were integrated into the BN model. Parameters were estimated using the Expectation Maximization (EM) algorithm, allowing the model to quantify pathways to communication errors, conflict, and accident severity. Results show that high language diversity and low proficiency substantially increase communication failures: miscommunication and misinterpretation together account for 68% of observed errors (43% and 25%), with high language diversity the most frequent state (∼48%). Verbal exchanges remain the most failure-prone mode (47%), and container ships account for the largest share of communication-related accidents (42%), with collision and grounding comprising 41% and 24% of incidents. At the consequence level, major repairs are the modal outcome (∼40%), but scenario analysis shows that under highly adverse diversity and proficiency configurations, the combined probability of major repair and total loss can exceed 60%. Strength-of-influence results identify vessel size, nationality, and language diversity as structural drivers (0.54), while language training (0.26), SMCP use (0.24), and cultural familiarization (0.27) are the most effective levers to improve proficiency and reduce conflicts. Scenario analysis further demonstrates that aligning these levers can cut total loss probability on large container ships from about 23% under stressed human-factor conditions to 3% under best-practice communication settings, while shifting incident severity toward minor outcomes and containing casualties. These results provide a quantitative basis for setting language-training thresholds, enforcing SMCP, and designing crew-mix and cultural familiarization policies in maritime safety management.

1. Introduction

The safe operation of ships is a matter of critical importance given that around 90% of global trade by volume is carried by sea [42,43]. Despite improvements in ship design, technology, and regulation, maritime accidents continue to occur and can have catastrophic consequences, from loss of life and environmental damage to economic disruption [19,20]. Notably, analyses of maritime incidents consistently show that the majority are attributable to human factors or error [44–46,53]. In fact, human error has been identified as a contributing factor in more than 85% of maritime accidents [5,7,18,27]. From a reliability perspective, these statistics highlight that accident causation in shipping is not solely a function of mechanical or technical failure, but is strongly influenced by latent human-system interactions, particularly how crew members exchange information, coordinate actions, and respond under time pressure [15,72,73]. This interplay between technical reliability and human reliability represents a critical point of vulnerability in socio-technical maritime systems [41,54].

* Corresponding authors.
E-mail addresses: r.u.khan@ljmu.ac.uk (R.U. Khan), z.yang@ljmu.ac.uk (Z. Yang).

https://doi.org/10.1016/j.ress.2026.112358
Received 22 August 2025; Received in revised form 22 December 2025; Accepted 3 February 2026
Available online 4 February 2026
0951-8320/

Effective communication is fundamental to safe vessel operations, yet it is often hampered by the language and cultural diversity inherent in the modern shipping industry [26,29,53,73]. Shipping is a highly international enterprise; crews on board today's merchant ships are typically multicultural and multilingual, with seafarers frequently coming from a dozen or more different national backgrounds [53,73]. Approximately 70–80% of the world's merchant fleet is manned by multicultural crews [54]. This means that English (the de facto lingua franca in shipping) is rarely the mother tongue for most crew. In fact, about 80% of seafarers do not speak English as their first language [63]. Communication at sea thus often occurs between individuals who have varying levels of English proficiency and differing interpretations of words or phrases. Unsurprisingly, communication breakdowns in such settings are a well-recognized threat to safety. Industry reports indicate that communication deficiencies are implicated in up to 30–35% of ship accidents [62]. Indeed, the U.K. Maritime and Coastguard Agency has listed communication failure as one of the "Deadly Dozen" human factors that most frequently precipitate accidents at sea [62]. From a system safety standpoint, these breakdowns can be understood as failures in safety-critical information transfer within the socio-technical network of a ship, where degraded communication channels increase the probability of operational errors propagating into hazardous states. These statistics and assessments highlight why this study is needed: in reliability terms, language and cultural diversity constitute variability factors that influence human reliability parameters, and their impact on error likelihood within maritime operations remains poorly quantified.

Maritime accident case studies poignantly illustrate the link between language barriers and safety failures. A frequently cited example is the tragic Scandinavian Star ferry disaster of 1990, in which 161 people lost their lives to a fire. Investigators found that the multinational crew's inability to communicate in a common language severely hampered emergency response and passenger evacuation [63]. Simply put, crew members and passengers could not understand each other well enough during the crisis. Another stark example occurred on the bridge of the container ship Cosco Busan in 2007. The Cosco Busan allided with the San Francisco Bay Bridge, causing a major oil spill, and one of the root causes was a lack of effective communication between the American pilot and the vessel's Chinese master [54]. The pilot gave critical instructions and information in English that were misunderstood or not fully grasped by the captain, reflecting a dangerous language and cultural disconnect. From a human reliability analysis perspective, these incidents represent classic cases where communication failures eroded shared situational awareness, reduced decision quality, and compromised response effectiveness, demonstrating the cascading effect of human performance variability on overall system safety [41]. Such examples also underscore the systemic nature of the problem: miscommunication is rarely an isolated lapse, but rather the manifestation of a latent system-level vulnerability embedded in multicultural operational contexts.

The shipping industry has long recognized the importance of a common working language to mitigate such risks. Since 1988, the International Maritime Organization (IMO) has mandated English as the official language of the sea, adopting standardized protocols like Seaspeak and the Standard Marine Communication Phrases (SMCP) to facilitate clear communication [63]. These standardized phrases and vocabulary were specifically developed by linguistic and maritime experts to reduce ambiguity and misunderstanding in radio or face-to-face exchanges [14,39]. However, the existence of a "common language" framework does not automatically resolve the issue; its effectiveness depends on crew members' language proficiency and their willingness to actually use the standard protocols. Many maritime accidents still cite language issues despite the IMO standards [59,73]. For instance, an investigation into a collision between the bulk carrier *Huayang Endeavour* and the tanker *Seafrontier* in 2017 found that the officers on the two vessels attempted to communicate by Very High Frequency (VHF) radio, but their dialogue was imprecise and non-standard, leading to confusion. The post-accident report concluded that had the officers adhered to SMCP, using simple, standardized English phrases for bridge-to-bridge communication, the risk of misunderstanding would have been significantly reduced [63]. From a system reliability perspective, this illustrates that the presence of formal communication protocols represents only a *potential* safety barrier; the reliability of that barrier is contingent on correct and consistent human execution, which is susceptible to degradation under diverse crew compositions. Thus, even with English as a common language, miscommunication persists due to inconsistent language skills and non-use of prescribed communication protocols [13,15]. This gap in the literature, quantitatively linking language proficiency and cultural variability to human error probability in maritime operations, remains largely unexplored and represents the core motivation for the present study [4,26,29].

Complementing the language issue is the often subtler challenge of cultural diversity on ships. Culture shapes how people communicate, interpret others' behavior, follow rules, and approach authority, all of which can impact safety. Modern ship crews embody a mix of national cultures, each with its own norms and values. Research has shown that national culture can significantly influence individuals' safety attitudes and behaviors [55]. For example, power distance, the extent to which hierarchy is respected and subordinates are hesitant to challenge superiors, varies widely across cultures (as described by Hofstede's cultural dimensions theory) [30,31,55]. In a high power-distance culture, a junior officer might be reluctant to question a captain's decision even if they perceive a mistake, whereas someone from a low power-distance background might be more forthcoming with a warning. From a human reliability standpoint, high power-distance environments reduce the likelihood of error detection and correction, allowing unsafe decisions to propagate unchallenged through the system. This dynamic has clear implications for preventing errors: if crew members do not speak up about concerns or doubts, emerging problems can go uncorrected until an accident occurs [30,32]. In line with this, a study by Lu *et al.* examining seafarers in container shipping found that ships with crews from cultures characterized by lower power distance and more collectivist, uncertainty-avoiding values experienced fewer human-error-related incidents [55]. In essence, certain cultural traits (such as an egalitarian team approach and a strong emphasis on procedural caution) were associated with safer performance, while large gaps in hierarchy or more individualistic attitudes could elevate the risk of miscommunication or rule violations. In reliability terms, this can be interpreted as cultural variability altering the probability distribution of human performance outcomes. These findings reinforce the intuition that a mismatch of cultural expectations among crew can be a latent hazard. Cultural differences may also lead to misinterpretation of non-verbal cues and behaviors, failures in shared situational awareness mechanisms, or even body language or gestures can be misunderstood across cultures [63], or to the formation of stereotypes that impede teamwork. Overall, when seafarers do not share common cultural ground on matters like safety protocols or communication etiquette, it can create confusion and conflict on the bridge or on deck, adversely affecting coordination and decision-making, thereby lowering the overall reliability of shipboard operations.

Multiple maritime safety investigations and studies have indeed pointed to cultural factors as contributors to accidents [11,39,40]. For example, one report noted that poor command of English, wrong stereotyping of other nationalities, different understandings of safety practices, and diverging attitudes toward risk were observed factors in several marine accidents [64]. Such observations illustrate that language and culture intertwine in shaping safety outcomes: language proficiency is part of one's cultural background and training, and cultural attitudes influence how communication is carried out (or not carried out) in crucial moments [21,36]. Despite these recognized reliability hazards, maritime safety management systems have historically been biased toward engineering reliability, with less emphasis on human reliability and cross-cultural interaction risk modeling.

Cross-cultural communication skills have lagged behind in maritime education and training, which traditionally focused on hard skills like navigation, engineering, and compliance with regulations [59,73]. As noted by World Maritime University experts, the development of cross-cultural competency remains largely absent from standard seafarer training curricula, even though shipping is a highly globalized, multicultural field [64]. Similarly, a survey by the University of Rijeka found that shipping companies and academies need to broaden their focus beyond technical qualifications to include soft skills such as effective communication and intercultural understanding [64]. In reliability engineering terms, such training would function as a preventive barrier, reducing the probability of communication-induced human errors and increasing the resilience of shipboard operations. The literature suggests that addressing this gap is not just a matter of crew comfort or harmony, but a direct factor in operational safety. Poor interpersonal understanding on board can erode trust and efficiency, whereas culturally aware crews with good communication practices are better equipped to handle emergencies and complex operations, thereby increasing the probability of correct and timely safety-critical actions.

From a reliability and system-safety standpoint, current maritime human reliability analysis (HRA) and probabilistic modeling have advanced along several lines: BN for accident causation and severity, dependence modeling among performance-shaping factors (PSFs) [52], and dynamic risk propagation, yet they rarely operationalize language and cultural diversity as quantifiable reliability modifiers [23,28,48]. BN-based maritime models successfully integrate human/ship/environment nodes to explain accident severity and risk evolution, while broader reliability and system safety work has refined HRA with CREAM-style PSFs, mission-context dependence, and information-flow causation to capture how human error propagates in socio-technical systems [1–3]. However, the specific pathways by which multilingual crews and cross-cultural hierarchies degrade communication reliability, and hence increase incident likelihood/severity, remain largely unmodeled, typically treated qualitatively or folded into generic "human factor" placeholders. This leaves a measurable gap: diversity-driven communication reliability has not been parameterized and sensitivity-tested with real accident evidence within BN-HRA frameworks, despite decades of calls in relevant literature platforms to model latent organizational/human conditions with probabilistic rigor [2,6,38].

Our contribution is to close that gap by introducing language and cultural diversity as explicit latent variables in a BN-HRA tailored to shipboard communication reliability, estimated from a large set of real investigations. Concretely, we integrate crew language diversity (Shannon index), cultural profile variability (Hofstede-based proxies), proficiency, and protocol adherence (SMCP) as parent nodes of communication-failure mechanisms, with downstream links to incident likelihood and consequence severity. Parameters are learned via data-driven BN techniques (with EM), and influence/importance indices quantify which levers, proficiency uplift, cross-cultural familiarization, and standardized phraseology deliver the greatest reliability gains. In doing so, the model operationalizes what prior reliability studies suggested conceptually, turning qualitative "human factors" into measurable reliability inputs, and complements existing maritime BN work on severity/cause by exposing the causal bridge from diversity → communication reliability → accident outcomes [47,56,70]. This positions diversity as an engineerable performance shaping factor (PSF) with testable interventions rather than a descriptive backdrop.

Also, this study aligns with emerging reliability engineering directions on dynamic risk and system resilience by showing how diversity-driven communication reliability integrates with risk propagation, system resilience, and decision support [22,71]. Our BN outputs (posterior failure probabilities, sensitivity rankings) can feed dynamic BN or scenario-propagation layers to capture time-varying exposure (e.g., watch rotations, weather shifts), and interface with resilience models (e.g., port/route disruption) to prioritize training and manning policies with the highest marginal reliability return [50]. Methodologically, the current study echoes reliability and system safety advances in risk propagation and dependence handling, but adds a new class of PSFs, language/culture, with calibrated parameters and policy-ready levers. This extends maritime risk models beyond generic "human error" to quantified, intervention-oriented human reliability inputs that are transferable to other safety-critical domains employing multinational teams. This study rigorously examines the impact of language and cultural diversity on miscommunication and maritime safety, specifically analyzing whether vessels with more heterogeneous crews face higher rates of communication-related incidents. Motivated by persistent accident investigation findings, this research moves beyond anecdotal accounts toward a quantitative reliability assessment, applying a BN approach to capture how diversity-related variables influence the probability of safety-critical communication failures and subsequent incident likelihood.

In relation to existing maritime BN and human reliability studies, this work is original in three respects. First, it models language and cultural diversity as explicit, quantifiable performance-shaping factors (language diversity index, Hofstede-based cultural diversity, language proficiency, SMCP use, cultural familiarization, conflict type), rather than treating them as generic "human error" or "organizational" nodes. Second, these variables are parameterized from coded accident investigation reports, so that the full chain from crew composition → communication failures → conflicts → incident type, damage, casualties, and severity is empirically derived rather than assumed. Third, beyond standard BN prediction, the model applies mutual-information-based–based influence analysis, global sensitivity analysis, and multi-evidence scenario analysis to show how the importance of crew characteristics varies across conflict types and accident patterns, and to identify which interventions (e.g., proficiency uplift, SMCP enforcement, composition of multicultural crews) offer the greatest marginal reduction in communication-driven risk.

The significance of the study is threefold. First, it provides a quantitative, evidence-based foundation for understanding how language and cultural diversity affect maritime system safety, moving beyond anecdotal case descriptions and enabling concrete policy choices such as setting minimum language proficiency thresholds or designing crew-mix policies. Second, the use of BN modeling clarifies not just that diversity matters, but how it matters, by tracing the different communication- and conflict-mediated pathways through which language and cultural diversity influence incident outcomes (e.g., via communication error type, conflict type, and communication mode). This enhances theoretical understanding of causality and reliability in maritime human factors by making explicit the mechanisms through which diversity affects human error probability. Third, the findings have direct safety-management value: shipping companies, regulators, and training institutions can use the model outputs (influence rankings, sensitivities, scenarios) to prioritize language training, SMCP compliance, and cross-cultural team interventions, and the same modelling approach can be transferred to other safety-critical domains that rely on multinational teams [64]. In this way, the study helps bridge the gap between qualitative human-factors insights and quantitative reliability and risk-based decision support.

The rest of the study is arranged such that Section 2 encompasses the detailed methodology, Section 3 covers the detailed results and analysis, Section 4 provides a discussion, and Section 5 concludes the findings of the study.

2. Methodology

In terms of *how* the study is carried out, we employ a BN modeling approach to analyze data collected on maritime miscommunications and incidents. A BN is a probabilistic graphical model well-suited to representing complex cause-and-effect relationships under uncertainty, in this case, the interplay between language/cultural factors, communication failures, and accident occurrence [68]. BN models have recently been applied in maritime safety research to incorporate human factors into accident analysis and have proven effective in identifying critical risk contributors from historical data [23]. In our analysis, we quantify language diversity on board using the Shannon entropy index (also known as Shannon's H), which captures both the number of different languages present in a crew and the relative prevalence of each. This provides an objective measure of how linguistically heterogeneous a crew is; a higher Shannon H indicates a greater diversity of languages among crew members. Likewise, we quantify cultural diversity using indices derived from Hofstede's national culture dimensions. Each crew's cultural composition (based on the national

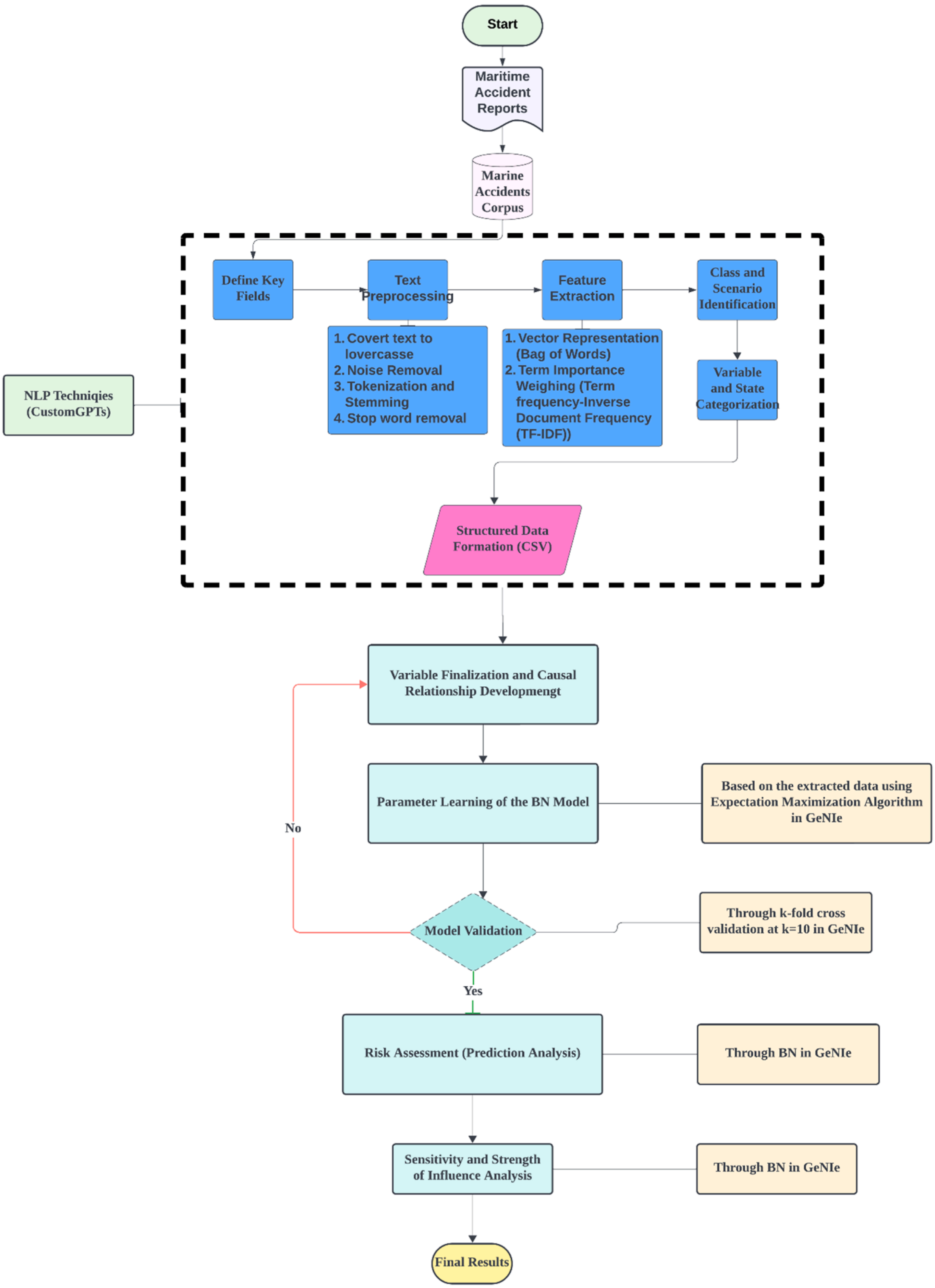


Fig. 1. A schematic flowchart of the employed data processing and analysis method.

cultures of its members) is mapped to an index that reflects diversity in key cultural traits (such as power distance, individualism vs. collectivism, uncertainty avoidance, etc.). By inputting these quantitative diversity measures into the BN alongside variables representing miscommunication frequency and safety incident rates, we can learn the probabilistic relationships from the data. The Conditional Probability Tables (CPT) in the network, essentially, the model's parameters, were learned from the structured dataset using expectation-maximization algorithms in the GeNIe BN software, ensuring that no prior bias skews the influence estimates. This data-driven learning process lets the model infer how strongly language and cultural diversity affect the likelihood of miscommunications, and in turn, how those communication failures influence accident risk.

This study systematically analyzes the influence of language and cultural diversity on miscommunications and maritime safety using a BN framework informed by a dataset of 550 maritime accident cases. The methodological process is outlined in four main steps: (1) data acquisition and structuring, (2) calculation of diversity indices, (3) BN model development, and (4) CPT quantification using EM. A data processing and analysis flowchart of the employed methodology has been depicted in Fig. 1.

2.1. Data acquisition and structuring

A total of 550 maritime accident cases (collision and contact events) were compiled from **reliable sources**, including global maritime safety databases such as EMSA (https://www.emsa.europa.eu/), MAIB (https://www.gov.uk/government/organisations/marine-accident-investigation-branch), IMO GISIS (https://gisis.imo.org/), and selected shipping company safety reports (COSCO Shipping). Each accident report included narrative incident descriptions, vessel particulars, and details on crew composition.

This study applies an existing GPT-based NLP framework, employed in our earlier work [53], to systematically convert narrative maritime accident investigation reports into a structured, analyzable dataset for probabilistic safety modeling. No new language model was developed for this research. Instead, an existing GPT-based tool was employed in a controlled extraction pipeline, leveraging its ability to perform contextual keyword recognition, text classification, and structured variable mapping from unstructured technical narratives. The objective was to identify and encode human factor indicators, particularly those relevant to communication reliability, such as *language diversity, cultural diversity, nationality diversity, vessel type, vessel size, training,* and *SMCP use*. These variables were selected through literature review, domain expertise, and alignment with recognized human reliability taxonomies. Detailed definitions for each variable were established prior to extraction to ensure repeatability and reduce subjectivity in coding.

2.1.1. Text pre-processing

The accident reports underwent a structured pre-processing sequence before feature extraction:

Standardization: All text was converted to lowercase, and formatting inconsistencies were corrected to ensure a consistent corpus.

Noise Removal: Non-informative characters, symbols, and formatting artifacts not contributing to causation analysis were removed.

Tokenization: Narratives were segmented into tokens and filtered to include only terms relevant to the identified variables.

Morphological Normalization: Stemming and lemmatization were applied to unify word variants without losing domain meaning.

Stop-word Filtering: A maritime-domain-specific stop-word list, refined over iterative trials, was applied to remove high-frequency, low-value words while preserving technical and causal terms.

2.1.2. Feature representation

The cleaned corpus was transformed into numerical vectors to support reliable and repeatable variable coding:

- **Bag-of-Words (BOW):** Constructed term-frequency matrices to capture prevalence of vocabulary relevant to the predefined nodes (e.g., vessel category terms for Vessel Type, incident descriptors for Incident Type, phrasing indicative of Communication Error Type states) [60].
- **TF–IDF Weighting:** Applied TF–IDF to attenuate high-frequency, low-discrimination terms and magnify less frequent but salient expressions aligned with the variables/states [60] in Table 1.
- **GPT-assisted Contextual Mapping:** Existing GPT tools were used only to assist with context disambiguation (e.g., differentiating Misinterpretation vs Miscommunication for Communication Error Type; identifying Primary Communication Mode as *Verbal/Radio, Written*, or *Digital/Electronic* from surrounding narrative context; distinguishing Cultural Conflict Type among *Hierarchy Issues, Risk Perception, Teamwork Issues, Behavioral Misinterpretation*). The tools returned suggested mappings, which were then applied systematically under the predefined schema.

2.1.3. Scenario classification and causation tagging

Each report was classified according to Incident Type (*Collision, Grounding, Near-miss, Equipment Failure*) and Incident Severity (*Minor, Moderate, Severe, Catastrophic*) using explicit labels or investigator

Table 1
Description of the Nodes and their states considered for the BN model.

Node Name	Description	States/Levels
Vessel Type	Type of vessel involved in the incident	Bulk Carrier, Container Ship, Tanker, General Cargo
Vessel Size	Size category based on gross tonnage	Small, Medium, Large, Very Large
Crew Nationality Diversity	Diversity of nationalities among crew (number of nationalities)	Low (1–5), Moderate (6–10), High (11–15)
Crew Language Proficiency (English)	Average English proficiency level onboard	High, Intermediate, Low
Crew Language Diversity Index	Shannon index for language diversity (binned for BN)	Low, Moderate, High
Crew Cultural Diversity Index	Hofstede's cultural diversity index (binned for BN)	Low, Moderate, High
Primary Communication Mode	Main communication channel at time of incident	Verbal/Radio, Written, Digital/Electronic
Communication Error Type	Type of communication error leading to incident	Misinterpretation, Miscommunication, Non-communication, Delayed communication
Standard Maritime English Use	Whether Standard Marine Communication Phrases (SMCP) used	Yes, No
English Language Training Level	Formality of crew's English language training	Certified, Informal, None
Cultural Misunderstanding Event	Whether cultural misunderstanding contributed	Yes, No
Cultural Conflict Type	Nature of cultural conflict involved	Hierarchy Issues, Risk Perception, Teamwork Issues, Behavioral Misinterpretation
Cultural Sensitivity Training Provided	Was crew given cultural sensitivity training?	Yes, No
Crew Cultural Familiarization Level	Crew's familiarity with onboard cultures	High, Medium, Low
Incident Type	Type of maritime incident/accident	Collision, Grounding, Near-miss, Equipment Failure
Incident Severity	Assessed severity of the incident	Minor, Moderate, Severe, Catastrophic
Number of Injuries/ Fatalities	Number of persons injured or killed	0, 1, 2, 3, 4, 5
Extent of Vessel Damage	Physical damage to the vessel	None, Minor Repairs, Major Repairs, Total Loss

descriptions. Outcome variables, including Number of Injuries/Fatalities (*0–5*) and Extent of Vessel Damage (*None, Minor Repairs, Major Repairs, Total Loss*) were coded directly from the textual evidence. Where multiple descriptors appeared, the most conservative (highest-severity) state consistent with the report was used.

2.1.4. Variable state and scenario encoding

For each accident report, all variables were coded strictly according to the predefined states. Vessel Type was classified as *Bulk Carrier, Container Ship, Tanker,* or *General Cargo*, while Vessel Size was assigned to *Small, Medium, Large,* or *Very Large* based on gross tonnage. Crew diversity measures, Crew Nationality Diversity, Crew Language Diversity Index (Shannon), and Crew Cultural Diversity Index (Hofstede), were binned into *Low, Moderate,* or *High* using pre-established thresholds. Language-related variables, including Crew Language Proficiency (English), Primary Communication Mode, Communication Error Type, Standard Maritime English Use, and English Language Training Level, were encoded directly from textual and documentary evidence. Cultural attributes such as Cultural Misunderstanding Event, Cultural Conflict Type, Cultural Sensitivity Training Provided, and Crew Cultural Familiarization Level were mapped to the relevant discrete states identified in Table 1. Incident descriptors, including Incident Type, Incident Severity, Number of Injuries/Fatalities, and Extent of Vessel Damage, were extracted from report narratives, applying a conservative coding rule where multiple severity indicators were present.

This enabled the integration of these variables into a BN model, capturing conditional dependencies relevant to communication reliability.

2.1.5. Dataset assembly for BN analysis

As all coded variables were complete, the dataset was compiled in tabular form (CSV), with each row representing an incident and each column corresponding to a BN node. The structured dataset was imported into GeNIe, where CPTs were generated using the EM algorithm, ensuring robust parameter estimation while maintaining the dependency relationships defined by the model structure.

This process produced a probabilistically compatible dataset, enabling the explicit modeling of language and cultural diversity as performance-shaping factors within the BN framework. The methodology mirrors our prior validated approach, ensuring methodological consistency while tailoring variable definitions and coding to the specific focus of this study.

Given that collision events inherently involve two ships, we focused on the vessel identified as "at fault" (i.e., causative vessel) in each accident. This approach aligns with practical accident analysis, as causal attribution and crew/cultural factors of the responsible vessel are more relevant to the risk modeling of miscommunication-induced accidents. Including both ships per incident would lead to non-independence and overrepresentation of accident features, especially as vessel size, type, and crew composition are tied to root-cause attribution.

2.2. Calculation of diversity indices

For each ship, two primary diversity indices were calculated:

(a) Shannon Language Diversity Index

The Shannon entropy (H) for onboard languages is calculated as:

$$H = -\sum_{i=1}^{N} p_i log p_i \tag{1}$$

Where p_i is the proportion of the crew speaking the ith language, and N is the total number of languages present among the crew. This index captures both the richness and evenness of language distribution and is widely used in maritime diversity research [30–32].

(b) Hofstede's Cultural Diversity Index

Cultural diversity was quantified using a composite Hofstede distance:

$$CDI = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} D_{ij} \tag{2}$$

Where D_{ij} is the Euclidean distance between crew members' countries on Hofstede's six cultural dimensions (power distance, individualism, masculinity, uncertainty avoidance, long-term orientation, indulgence), and n is the number of nationalities onboard. Country scores are obtained from Hofstede Insights (https://www.theculturefactor.com/country-comparison-tool).

After computing raw values, both indices were discretized into ordinal states (e.g., low, moderate, high, very high) to serve as input states in the BN model, based on established thresholds in maritime safety research.

2.3. BN model development

The BN model structure was developed by integrating data-driven analysis (model learning offered by GeNIe software) with structured expert judgment, as described in previous maritime BN studies [46,53,73,74]. However, it is noteworthy to mention that the final causal relationship and model structure were established and finalized by a panel of ten experts, the details of which have been provided in section 3.1. Nodes in the BN corresponded to all key factors: vessel characteristics, diversity indices, communication-related factors, and accident outcomes. The directed acyclic graph (DAG) was finalized by consensus, with edges reflecting hypothesized and empirically supported causal dependencies.

2.4. CPT quantification

Where sufficient case data existed, node CPTs were quantified directly from frequencies in the structured dataset. For nodes with high-dimensional parent sets or incomplete data, parameter estimation was conducted using the EM algorithm in GeNIe software [8].

(a) EM Algorithm in BN Parameter Estimation

To estimate the CPTs of the BN from incomplete or uncertain data, we employed the EM algorithm, as implemented in the GeNIe software. EM proceeds iteratively via:

E-step: For each parameter configuration, compute the expected counts of each variable and parent configuration using the current parameter estimates:

$$\widehat{N_{ijk}} = \sum_{n=1}^{N} P(X_i = k,\ P_{ai} = j | x^{(n)},\ \theta^{(t)}) \tag{3}$$

Where $x^{(n)}$ denotes the observed (possibly incomplete) data for case n.

M-step: Update the parameter estimates by normalizing the expected counts:

$$\theta_{ijk}^{(t+1)} = \frac{\widehat{N_{ijk}}}{\sum_{k'} \widehat{N_{ijk'}}} \tag{4}$$

X_i: The i^{th} variable (node) in the BN.

k: The possible state/index of the variable X_i. For example, if X_i is "Incident Severity" with 3 states (Low, Moderate, High), then $k \in \{1,2,3\}$.

P_{ai}: The parent nodes of X_i.

j: The specific configuration/state combination of the parents P_{ai}. If there are two parents with two states each, j runs over their combinations.

N_{ijk}: The expected (posterior) count of cases where variableX_i is in state k, and its parents P_{ai} are in configuration j, summed over all N data cases, taking into account missing data (hence "expected"). The "hat" (^) denotes that it is an *estimated* or *expected* count, not an observed (raw) count.

N: The total number of data cases (examples/accidents/rows) in your dataset.

n: The index of each data case.

$x^{(n)}$: The observed values for all variables in the n^{th} data case (the "evidence" for that row).

$\theta^{(t)}$: The current estimate of the BN parameters (CPTs) at iteration t of the EM algorithm.

$P\left(X_i = k,\ P_{ai} = j | x^{(n)},\ \theta^{(t)}\right)$: The probability, given the data in row n and the current parameter estimates, that X_i is in state k and its parents in configuration j.

$\theta_{ijk}^{(t+1)}$: The updated CPT entry: the probability that $X_i = k$ given $P_{ai} = j$, for the next EM iteration.

k': An index over all possible states of X_i (used in the denominator to ensure the probabilities sum to 1).

This yields the maximum-likelihood estimates for the CPTs given the (expected) sufficient statistics. These steps are repeated until convergence.

(b) BN Joint Probability Representation

The BN framework models the joint probability distribution of all variables as a product of local conditional probabilities:

$$P(X_1, X_2, \ldots,\ X_n) = \prod_{i=1}^{n} P(X_i | P_a(X_i)) \tag{5}$$

where $P_a(X_i)$ denotes the set of parent nodes for variable X_i. This structure enables efficient probabilistic inference and scenario analysis. When new evidence (such as the observation of a specific communication error) becomes available, the BN updates the posterior beliefs of all connected variables using Bayes' theorem:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \tag{6}$$

where H is a hypothesis (e.g., the presence of a high-severity incident), and E is the observed evidence.

3. Results

3.1. Model development

Effective communication and cultural understanding are fundamental to safe and reliable maritime operations, yet accidents and near-misses continue to occur due to persistent language barriers, cultural conflicts, and diverse crew backgrounds. As the global shipping industry becomes increasingly multicultural, the complexity of onboard interactions and their influence on safety-critical decisions has grown. However, quantifying the precise pathways through which language and cultural diversity contribute to incident causation remains a major challenge for risk analysts and safety regulators. In this study, we employ a BN approach to systematically model and evaluate the influence of linguistic and cultural variables on maritime safety outcomes, using structured data from 550 real-world accident cases. Table 1 provides an overview of the key variables (nodes) incorporated in our network, capturing technical, linguistic, cultural, and incident-specific dimensions relevant to accident causation. These variables were finalized by experts, and also, once the GeNIe software produced a data-centered causal model, it was presented to a panel of experts, who presented their recommendations, and the model's causal relationship was changed accordingly. Once all the experts agreed in the structured workshops on the causal relationships among variables, the model was finalized for analysis. The details of the experts who participated in the finalization of variables and model network structure have been provided in Table 2.

Besides that, the Hofstede cultural diversity index has been divided into three states as indicated in Table 1, and these states are determined based on the obtained numeric scores with ranges such that scores below 55 were classified as low. Those between 55 and 65 were classified as moderate, and scores above 65 were classified as high cultural diversity. For the ship repair levels classification, the detailed breakdown and followed protocols are provided in Table 3. While for the language diversity classification based on values achieved from the Shannon H' index calculations, the used classes and their division levels are given in Table 4.

3.2. Model validation

To ensure the robustness and credibility of our BN model, we employed ten-fold cross-validation, a standard technique for evaluating predictive models in complex, multivariate domains [57,69]. Our dataset was first randomly shuffled to remove any ordering effects, then partitioned into ten equal parts. For each fold, nine segments served as the training set and one as the testing set, cycling through all combinations. This process reduces variance and yields a reliable assessment of model performance [37]. Model predictive ability was further assessed using Receiver Operating Characteristic (ROC) curves and the corresponding Area Under the Curve (AUC) metric, as recommended for categorical outcome evaluation [9]. AUC values closer to 1.0 indicate stronger predictive accuracy. In our case, the "Accident Risk" node produced AUC scores well above 0.5, reflecting strong discrimination and reliable model output. Full ROC plots and AUC values for all

Table 2
A description of the roles and profiles of the experts participating in model development.

Designation	Qualification	Experience / Professional Background
Ship Captain	Master's Degree	Over 20 years of international navigation and leadership of multicultural crews
Chief Engineer	Master's Degree	18+ years of shipboard engineering operations, safety management, and crew coordination
Vessel Traffic Officer	Master's Degree	20+ years in vessel monitoring, navigation assistance, communication management, and compliance
Senior Safety Engineer	Master's Degree	Over 20 years of professional practice in maritime safety engineering, accident prevention, and risk management
Professor of Maritime Safety	PhD	15+ years of academic research in maritime accident causation, human reliability, and safety management systems
Professor of Linguistics	PhD	12+ years of research in communication, SMCP, and English for specific purposes in maritime operations
Professor of Human Factors (in Safety)	PhD	10+ years of academic research on human performance, risk perception, and maritime safety
Technical Superintendent	Master's Degree	15 years managing multinational crews, vessel operations, and compliance auditing in commercial shipping
Maritime Accident Investigator	Master's Degree	15+ years in accident investigation with emphasis on regulatory and industry investigation contexts
Maritime Safety & Risk Analyst	PhD	12+ years of applied expertise in probabilistic risk assessment and safety-critical system analysis, with direct engagement in industry safety projects

Table 3
Ship repair level classifications.

Category	Repair Cost as % of Vessel Value	Description
None	0%	No damage or so minor it doesn't require repair.
Minor Repairs	>0% – ≤5%	Cosmetic/superficial damage; does not impact vessel operation.
Major Repairs	>5% – <75%	Structural or significant system damage; repair is economically feasible; vessel is not a total loss.
Total Loss (CTL)	≥75% (varies by insurer, 75–80% typical)	Repair cost equals or exceeds the insured value; the vessel is declared a constructive total loss (CTL).

Table 4
Language diversity classification based on Shannon's H index.

Shannon Index (H')	Diversity Category	Maritime Industry Justification
0.00 – 0.60	Low	Usually homogeneous crews (Crews with shared working language, little multilingual interaction).
0.61 – 1.00	Moderate	Crews with different languages; for example, some officers speak one language, the crew speaks another, but still manageable; a frequent scenario in bulk carriers and container shipping.
1.01 – 1.30	High	Crews of diverse linguistic groups; typically mixed crews where shipping companies source seafarers globally (e.g., mixed Asian, Eastern European, and African crews).

relevant states are included in Fig. 2. These results confirm the validity of our model for analyzing the impact of language and cultural diversity on maritime safety outcomes.

In addition to ROC and AUC, we incorporated classification-based diagnostics for the *Incident Type* node. The confusion matrix in Table 5 compares predicted versus observed incident types (collision, grounding, equipment failure, near miss), yielding an overall accuracy of 62.5% (344 correctly classified out of 550 cases). Class-wise precision ranges from 50.0% (equipment failure) to 80.9% (near miss), with recalls between 40.9% and 79.1% and F1-scores from 45.0% to 80.0% (Table 6). Collision and near-miss incidents, which are both frequent and operationally critical, obtain F1-scores of 69.5% and 80.0%, respectively, indicating good discrimination between high-energy geometry-driven events and less severe near misses. For grounding and equipment failure, performance is lower, but still informative, which is consistent with their smaller sample sizes and partially overlapping causal patterns. Taken together with the ROC/AUC results, these metrics indicate that the BN captures the dominant structure of the data while leaving scope for further refinement as larger and more balanced datasets become available.

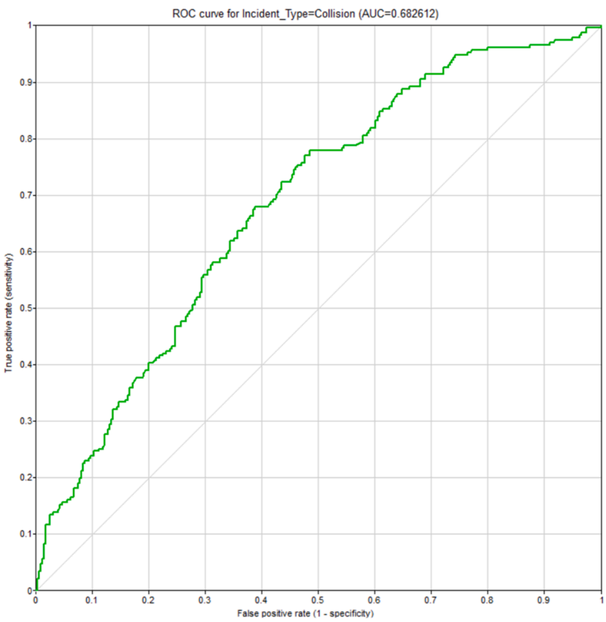
Following prior BN applications to workplace safety behavior, which use mutual information and belief updating to rank influential factors and compare intervention strategies [58], we compute strength-of-influence indices, perform sensitivity analysis, and perform scenario-based diagnostic and predictive analyses for key human-factor nodes.

3.3. Model prediction analysis

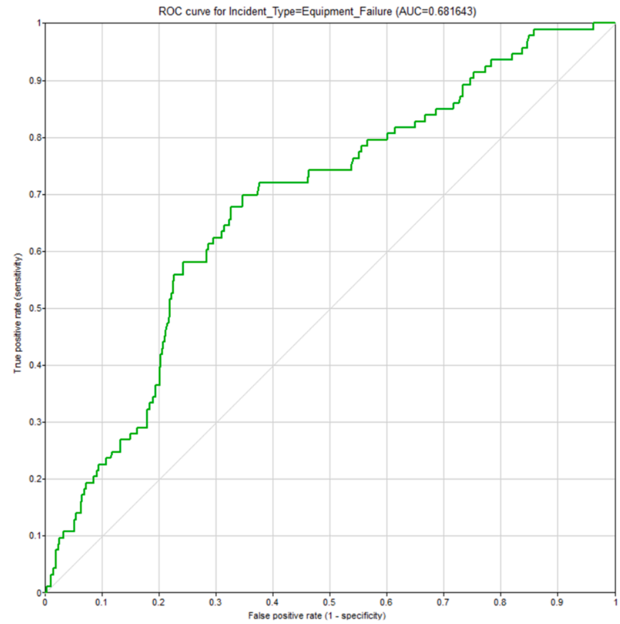
After validating the model, the next step is to run the prediction analysis and determine the impact of each node under consideration on the safety and outcomes. In this scenario, evaluating the communication error, the highest probability was exhibited by the states "Miscommunication" and "Misinterpretation", which had probabilities of 43% and 25% respectively, as depicted in Fig. 3. Among the upstream precursors, cultural diversity and language diversity exerted significant influence. Specifically, crews characterized by low cultural diversity showed higher accident involvement (55%). This counterintuitive finding can be explained by the absence of cross-checking behaviors in homogeneous crews, where assumed shared understanding leads to complacency and unchecked propagation of errors. In contrast, heterogeneous crews, although prone to miscommunication, may trigger more clarifications and corrective dialogue, which acts as a form of error detection. From a reliability standpoint, this dynamic resembles redundancy in safety systems, where diversity provides an additional defensive layer against common-mode failures. Similarly, high language diversity reached 48%, underscoring the criticality of heterogeneous communication environments. The role of Cultural Familiarization remained noticeable, with state "Medium" having the highest probability of 38%, reflecting the dominance of average familiarity onboard ships in shaping communication dynamics. From a reliability perspective, these results demonstrate that communication failures exhibit the characteristics of "frequent, low-consequence" initiating events that align with Reason's Swiss Cheese model [61]. Miscommunication and misinterpretation, while often non-catastrophic individually, become recurrent latent conditions that align and propagate through the system to trigger more severe outcomes. This framing is consistent with safety-critical industries such as aviation, where language errors rarely cause accidents in isolation but act as error-producing conditions that compromise system defenses [66].

Moving further into the analysis, the highest role in the accident causation was of verbal communication, which achieved a probability of 47%. The most prominent conflict types were the Teamwork Issues and the Hierarchy Levels, which caused the accidents. Moreover, the highest involvement was found to be of Container vessels, which had a probability of 42%, and the most prevalent type of accident as a result of communication issues was collision, followed by grounding with probabilities of 41% and 24%, respectively. Major repairs were the highest occurring consequences with a probability of 40%, while life losses were less frequent. This aligns with the incident severity distribution, where minor and moderate outcomes dominated (34% and 30%). The finding that verbal communication remains the most failure-prone mode is particularly important. In reliability terms, this mode constitutes a "single point of failure" in socio-technical systems, as its breakdown directly increases the likelihood of conflict and operational error. The persistence of hierarchy-driven conflicts reflects systemic "latent failures" embedded within organizational culture, similar to those observed in offshore oil and nuclear operations where rigid authority gradients impair safety-critical decision-making [24]. Thus, these quantitative results underline the systemic interaction of human, cultural, and organizational precursors, not just isolated human error.

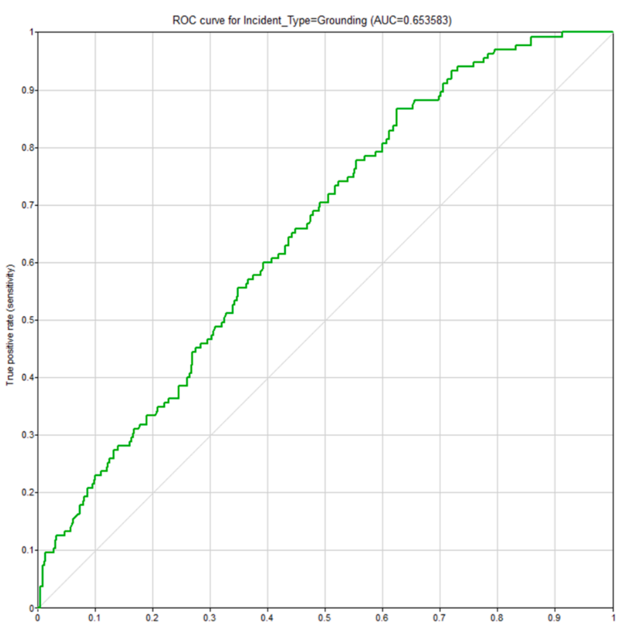
The distribution of accident types also provides insight into system-level dynamics. Collision (41%) and grounding (24%) dominate as communication-related outcomes, indicating that communication breakdowns disproportionately influence navigational safety barriers. This mirrors failure mode propagation observed in other domains (e.g., rail signaling or air traffic control), where misinterpretation cascades rapidly into high-consequence but preventable incidents [49]. Importantly, the prevalence of repair-related consequences (40%) but relatively lower fatality rates reflects the system's partial resilience: technical redundancy and recovery mechanisms often prevent catastrophic escalation, but do not eliminate frequent, costly disruptions. This aligns with reliability theory, where systems may appear "safe" in terms of fatalities but remain fragile in terms of recurrent operational losses. Taken together, the prediction analysis highlights that communication-related failures in shipping follow patterns well established in reliability engineering: frequent low-severity disruptions, rooted in latent organizational and cultural conditions, with occasional escalation to severe accidents. These findings emphasize the importance of not only language proficiency and cultural familiarization but also systemic interventions, such as authority de-biasing, standardized communication protocols, and organizational learning loops, to



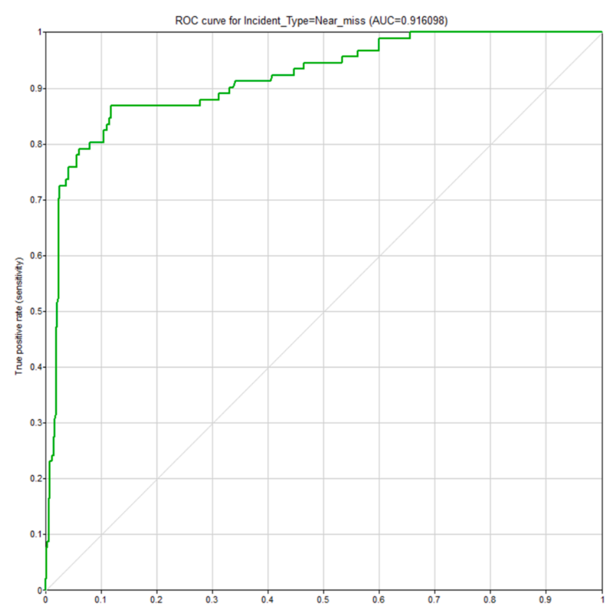
a. ROC and AUC for accident type collision



c. ROC and AUC for accident type equipment failure



b. ROC and AUC for accident type grounding



d. ROC and AUC for the near miss

Fig. 2. Validation of the accident type through ROC and AUC for each state of the node.

Table 5
Confusion Matrix for the node Conflict Type under model validation.

Actual \ Predicted	Collision	Grounding	Equipment_Failure	Near_miss
Collision	**173**	32	17	9
Grounding	56	**61**	15	3
Equipment_Failure	30	20	**38**	5
Near_miss	8	5	6	**72**

strengthen barriers and reduce alignment of latent failures.

From a category perspective, Table 7 reorganizes the prediction results into three categories, human factors, vessel features, and accident features, and clarifies how the network distributes baseline risk across the system before any evidence is entered. A first observation is the structural dominance of human-factor nodes: language diversity, nationality diversity, cultural diversity, cultural misunderstanding, language training, SMCP use, language proficiency, cultural familiarization, communication mode, communication error type, and conflict type. Together, these nodes define a dense “upper layer” of precursors that shape the likelihood of downstream accident states. Their priors show that the model is not calibrated around extreme or idealized

Table 6
Classification metrics (precision, recall, F1-score) for the incident type node.

Class	Precision	Recall	F1-score
Collision	64.8%	74.9%	69.5%
Grounding	51.7%	45.2%	48.2%
Equipment_Failure	50.0%	40.9%	45.0%
Near_miss	80.9%	79.1%	80.0%

conditions but around realistically mixed crews: for example, high language diversity (≈48%), moderate or high language proficiency (≈78% combined), and medium–high cultural familiarization (≈72% combined). This configuration represents a system where crews are generally competent but exposed to persistent variability in language, culture, and training quality.

Within this human-factor layer, several distributions highlight latent vulnerabilities. Certified language training and SMCP use are prevalent, yet miscommunication and misinterpretation still dominate the communication-error node. Similarly, conflict type is split across teamwork issues, hierarchy issues, and behavioral misinterpretation, indicating that no single conflict mechanism dominates; instead, the network encodes multiple interacting pathways through which human reliability can degrade. The Table 7, therefore, makes explicit that the model is not a simple “low proficiency → accidents” representation but a multidimensional human-factor landscape with several parallel routes to failure.

The vessel-feature category shows a different pattern. Very large vessels (≈45%) and container ships (≈42%) are structurally overrepresented, reflecting the traffic composition of global liner shipping and the expert judgement that these platforms concentrate both exposure and consequence. This means that any degradation in the human-factor layer is more likely to materialize in operating contexts with tight schedules, high traffic density, and complex port approaches, conditions typical for large container ships.

Accident-feature nodes translate these human and vessel conditions into outcomes. The priors for collision, grounding, and equipment failure, combined with the distribution of damage, injuries, and severity, show a system that experiences frequent, economically significant events (notably major repairs at ≈40%) but only occasionally escalates to catastrophic loss. Read together, Table 7 and Fig. 3 therefore convey a coherent story: a network in which human and cultural factors are central, vessel and traffic characteristics modulate exposure, and the accident layer reflects a safety regime that is partially resilient yet operationally fragile, with recurrent disruptions driven by communication- and culture-related mechanisms.

3.4. Sensitivity analysis

To further evaluate the explanatory power and practical relevance of the developed BN, a comprehensive sensitivity analysis was conducted using GeNIe software. This analysis systematically quantifies how changes in the input nodes influence the probability distribution of a selected target node, thereby identifying which variables exert the greatest impact on key safety outcomes. The GeNIe platform provides an intuitive visual output, with higher sensitivity indicated by deeper red coloration, highlighting those nodes where targeted interventions could significantly reduce miscommunication risk and augment maritime safety. Sensitivity analysis functions analogously to importance measures used in fault tree or probabilistic risk assessment (e.g., Fussell–Vesely or Birnbaum indices) [75]. Quantifying how uncertainty in one node propagates through the network highlights the “weakest links” in the socio-technical system, where small improvements in reliability can yield disproportionate safety benefits [12,76]. In this context, improving control strategies for sensitive nodes translates into targeted interventions with the highest leverage for risk reduction.

The sensitivity analysis was conducted in concurrence with the scope of the study by first setting the Conflict Type as the target node, for which the results are displayed in Fig. 4. For this scenario, communication mode and communication error type emerged as the most sensitive factors. Further, language proficiency, language training, vessel size, vessel type, and Hofstede’s cultural diversity index were also identified as highly sensitive nodes. These variables act as dominant failure precursors: their state changes significantly alter the probability

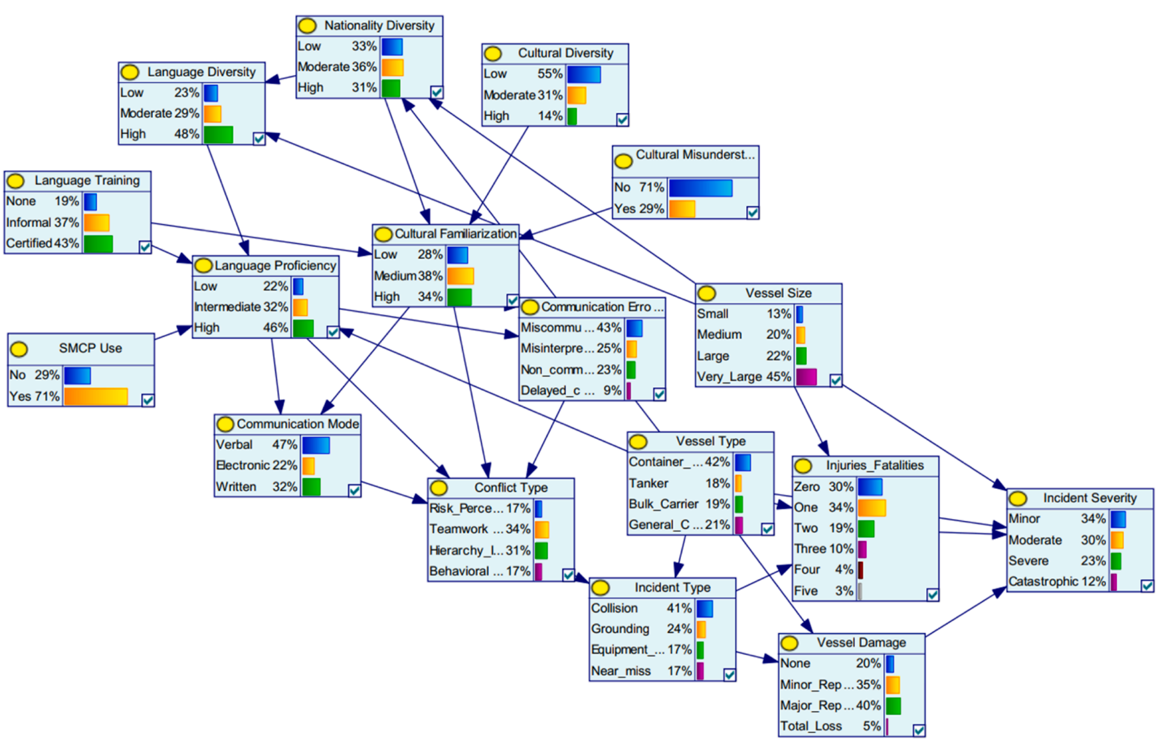


Fig. 3. Model prediction analysis.

Table 7
Node categorization and prior (no-evidence) probabilities for the prediction network in Fig. 3.

Category	Node	States and prior probabilities (no-evidence prediction)
Human factors	Language_Diversity	Low ≈ 23%; Moderate ≈ 29%; High ≈ 48%
	Nationality_Diversity	Low ≈ 33%; Moderate ≈ 36%; High ≈ 31%
	Cultural_Diversity	Low ≈ 55%; Moderate ≈ 31%; High ≈ 14%
	Cultural_Misunderstanding	No ≈ 71%; Yes ≈ 29%
	Language_Training	None ≈ 19%; Informal ≈ 37%; Certified ≈ 43%
	SMCP_Use	No ≈ 29%; Yes ≈ 71%
	Language_Proficiency	Low ≈ 22%; Intermediate ≈ 32%; High ≈ 46%
	Cultural_Familiarization	Low ≈ 28%; Medium ≈ 38%; High ≈ 34%
	Communication_Mode	Verbal ≈ 47%; Electronic ≈ 22%; Written ≈ 32%
	Communication_Error_Type	Miscommunication ≈ 43%; Misinterpretation ≈ 25%; Non_communication ≈ 23%; Delayed_communication ≈ 9%
	Conflict_Type	Risk_Perception ≈ 17%; Teamwork_Issues ≈ 34%; Hierarchy_Issues ≈ 31%; Behavioral_Misinterpretation ≈ 17%
Vessel features	Vessel_Size	Small ≈ 13%; Medium ≈ 20%; Large ≈ 22%; Very_Large ≈ 45%
	Vessel_Type	Container_Ship ≈ 42%; Tanker ≈ 18%; Bulk_Carrier ≈ 19%; General_Cargo ≈ 21%
Accident features	Incident_Type	Collision ≈ 41%; Grounding ≈ 24%; Equipment/Other failure ≈ 17%; Near_miss ≈ 17%
	Vessel_Damage	None ≈ 20%; Minor_Repair ≈ 35%; Major_Repair ≈ 40%; Total_Loss ≈ 5%
	Injuries_Fatalities	Zero ≈ 30%; One ≈ 34%; Two ≈ 19%; Three ≈ 10%; Four ≈ 4%; Five ≈ 3%
	Incident_Severity	Minor ≈ 34%; Moderate ≈ 30%; Severe ≈ 23%; Catastrophic ≈ 12%

distribution of downstream conflict, meaning that interventions at this stage could act as “preventive barriers” against the escalation of incidents.

For a further analysis, the top 5 ranking node combinations for each state of the Conflict Type, based on the tornado charts (provided in Appendix A), have been provided in Table 8 below, accompanied by an overview of the trends observed.

The sensitivity rankings for the four conflict types reveal a consistent picture of conflict as a *communication-system outcome* rather than a simple function of diversity alone. Across all states of *Conflict_Type*, the most influential configurations involve communication error type, communication mode, and the joint levels of language proficiency and cultural familiarization. In other words, the BN is telling us that it is not who is on board in abstract, but how they communicate under specific skill and familiarity conditions that drive conflict probabilities.

For Risk_Perception, the dominant configurations involve Miscommunication at high language proficiency and high cultural familiarization, and Verbal or Electronic modes with high proficiency and medium familiarization. This indicates that disagreements about risk are most sensitive to *how* capable, relatively integrated crews communicate hazard information. When risk cues are exchanged verbally or electronically among high-skill crews, small distortions or omissions in wording strongly shift the probability of risk-perception conflict. The presence of non-communication in the top five, again at high proficiency and medium familiarization, shows that failing to communicate risk at all in such settings is nearly as consequential as miscommunicating it.

For Teamwork_Issues, the top rankings are dominated by configurations that combine medium or high cultural familiarization and high proficiency with Verbal or Written modes and miscommunication or misinterpretation. In other words, teamwork tensions peak when well-qualified, reasonably familiarized crews are coordinating tasks through structured channels (verbal briefings, written or electronic instructions) and messages are still unclear or contested. Several of the highest-impact terms are complete conflict-type bundles where cultural familiarization, language proficiency, communication mode, and error type are simultaneously specified, confirming that the model sees teamwork conflict as arising from *packages* of conditions rather than isolated variables.

Hierarchy_Issues show a distinct signature. The single most influential configuration is Written communication with high proficiency and medium familiarization, followed by several bundles involving intermediate or high proficiency, medium or high familiarization, and Verbal mode with Miscommunication or Non-communication. This pattern suggests that hierarchy conflicts are especially sensitive to the *formality* of communication: written orders or records, and situations where superiors fail to communicate or do so ambiguously, are key triggers when crews are technically capable but not fully culturally aligned. Verbal interaction in highly familiarized, proficient crews appears less destabilizing, consistent with the idea that relational repair and negotiation are easier face-to-face.

Finally, Behavioral_Misinterpretation is driven predominantly by Miscommunication and Misinterpretation across a range of proficiency and familiarization levels, with one notable cross-effect: a hierarchy-issue configuration (medium familiarization, high proficiency, written mode, non-communication) is the single strongest influencer. This indicates that ambiguous or absent written communication in hierarchical exchanges not only produces explicit hierarchy conflict but also spills over into misreading of behavior and intent. Verbal mode with low or intermediate proficiency and high familiarization also features prominently, highlighting that even culturally close teams are vulnerable to misinterpreting tone and behavior when lexical resources are limited.

Overall, the trends show that the most effective leverage points for reducing conflict are design choices in communication mode and standardization, coupled with targeted investments in language proficiency and intercultural familiarization. Different conflict types exhibit distinct sensitivity profiles, risk perception to high-skill verbal/electronic exchanges, teamwork to written coordination, hierarchy to formal and non-communication, and behavioral misinterpretation to basic language and cue-reading errors, implying that interventions must be conflict-specific rather than generic “communication training.”

The next scenario examined factors influencing cultural familiarization onboard. The results for this scenario are depicted in Fig. 5. Here, Shannon’s language diversity index and vessel size were identified as the most critical variables. In addition, nationality diversity, cultural misunderstanding, and language training showed high impact. These findings underscore how system-level attributes (e.g., vessel size and crew mix) cascade into human factor vulnerabilities, shaping the resilience of communication pathways. Just as redundancy or diversity in engineered systems can either buffer or destabilize reliability depending on coupling, diversity in crew composition similarly modulates the effectiveness of cultural familiarization.

The sensitivity analysis for the Cultural Familiarization node shows that familiarization is governed primarily by the interplay between background composition (cultural and nationality diversity), the presence/absence of cultural misunderstandings, and the structure of language training, with vessel size emerging as a secondary but non-negligible driver. For a deeper overview at the state level, the top five parameters for each of the states of the target node for this sensitivity analysis have been summarized below in Table 9.

For the Low familiarization state, the single strongest lever is *Cultural_Diversity = Low*. This indicates that simply sailing with crews from relatively similar cultural backgrounds does not automatically generate high mutual understanding; in fact, the model sees low cultural diversity

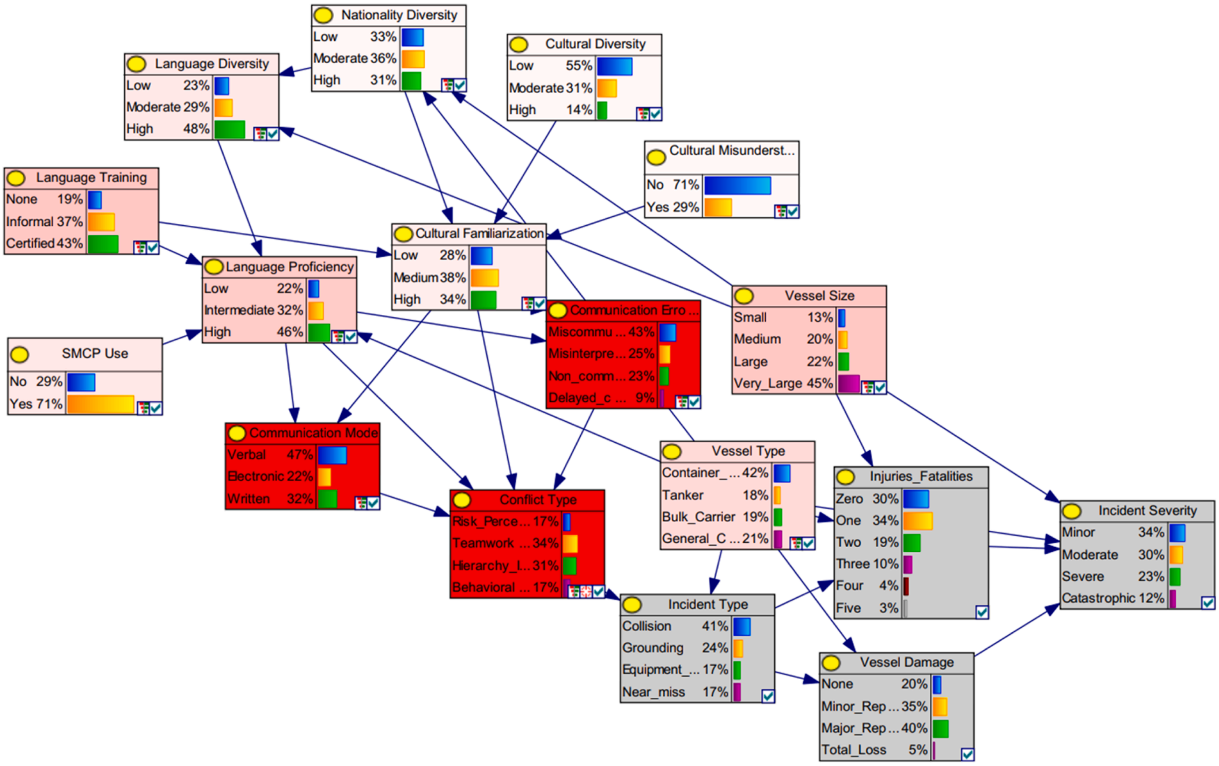


Fig. 4. Sensitivity analysis for the node Conflict Type.

as a key context in which low familiarization can persist. The remaining top-ranked configurations for the Low state are all detailed bundles where cultural diversity is low, cultural misunderstanding is not explicitly reported, and nationality diversity and training vary between low or moderate and certified or informal. These patterns suggest that low familiarization is sustained when crews from broadly similar cultural backgrounds are not systematically exposed to structured cross-cultural and language training, even if overt misunderstandings are not recorded. The presence of Vessel_Size = Very_Large among the strongest influences for the Low state points to an additional scale effect: on very large ships, organizational complexity and segmented work groups may make it easier for subcultures to remain siloed, limiting day-to-day familiarization despite nominal similarity in background.

For the Medium familiarization state, the top five configurations are all explicit Medium-familiarization bundles with low cultural diversity, absence of reported cultural misunderstanding, and a spread of nationality diversity and training regimes (certified and informal). This confirms that the Medium category behaves as a “pivot” state: small changes in training emphasis or crew mix within broadly low-diversity settings can push probability mass up toward High or down toward Low familiarization. The sensitivity of this state to whether training is certified versus informal, and to whether nationality diversity is low, moderate, or high, underlines that training design is a primary control knob for moving crews out of middling familiarization levels.

For the High familiarization state, the strongest configurations again involve low cultural diversity and absence of cultural misunderstanding, but now coupled with high or moderate nationality diversity and both informal and certified training. This pattern is important: high familiarization is most strongly associated with crews whose cultural distance is limited in aggregate (low cultural diversity index) but who still represent several nationalities, and who receive at least some systematic language training. In such settings, differences are sufficient to make training and mutual adjustment salient, yet not so large that stable shared norms cannot emerge. Vessel_size = Very_Large also appears as a major influence for High familiarization, suggesting that large ships can achieve high cross-cultural familiarity when formal training and deliberate integration mechanisms are in place, rather than being doomed to low familiarization by scale alone.

Overall, the dynamics indicate that cultural familiarization is not a simple monotonic function of diversity. Instead, it is shaped by how diversity is structured (low cultural distance but varying nationality mix), whether misunderstandings are actively managed, and how language training is deployed. For practice, this means that port-state control and company safety programs should focus less on reducing diversity per se and more on designing certified, context-specific language and cultural training schemes that convert diverse but culturally proximate crews into high-familiarization teams, particularly on very large vessels where the stakes and coordination demands are highest.

The final scenario set language proficiency as the target, as seen in Fig. 6. Here, language training, vessel size, and vessel type were the most critical determinants. This indicates that proficiency is not only an individual attribute but structurally linked to organizational choices in vessel deployment and crewing strategies. Notably, very large ships and container vessels showed the greatest sensitivity, confirming that operational complexity amplifies the need for standardized training. The use of SMCP was also identified as a critical mitigating factor, operating as a reliability control measure akin to procedural standardization in engineered safety systems. Hofstede’s cultural diversity index and nationality diversity further shaped proficiency outcomes, demonstrating how crew composition interacts with training effectiveness.

The sensitivity patterns for Language Proficiency show a very structured relationship between training, operational context, and the distribution of proficiency states in the crew. The detailed ranking-wise classification of the top five parameter configurations, for all the three states of the Language Proficiency has been provided in Table 10.

First, language training emerges as the dominant control lever across all three states. For Low proficiency, “Language_Training = Certified” is the strongest *negative* driver, substantially reducing the posterior probability of Low proficiency, while “Language_Training = None” pushes it upward. This symmetry confirms that formal training directly shifts the

Table 8
The top five most influential parameter configurations for each conflict type.

Conflict type state	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
Risk_Perception	Communication_Error_Type = Miscommunication (Language_Proficiency = High, Cultural_Familiarization = High)	Communication_Mode = Verbal (Language_Proficiency = High, Cultural_Familiarization = Medium)	Communication_Mode = Electronic (Language_Proficiency = High, Cultural_Familiarization = Medium)	(Cultural_Familiarization = Medium, Language_Proficiency = Low, Communication_Mode = Verbal, Communication_Error_Type = Miscommunication)	Communication_Error_Type = Non_communication (Language_Proficiency = High, Cultural_Familiarization = Medium)
Teamwork_Issues	(Cultural_Familiarization = Medium, Language_Proficiency = High, Communication_Mode = Verbal, Communication_Error_Type = Miscommunication)	Communication_Mode = Written (Language_Proficiency = High, Cultural_Familiarization = High)	Communication_Mode = Written (Language_Proficiency = High, Cultural_Familiarization = Medium)	(Cultural_Familiarization = High, Language_Proficiency = High, Communication_Mode = Written, Communication_Error_Type = Misinterpretation)	(Cultural_Familiarization = High, Language_Proficiency = High, Communication_Mode = Verbal, Communication_Error_Type = Miscommunication)
Hierarchy_Issues	Communication_Mode = Written (Language_Proficiency = High, Cultural_Familiarization = Medium)	(Cultural_Familiarization = Medium, Language_Proficiency = Intermediate, Communication_Mode = Verbal, Communication_Error_Type = Miscommunication)	(Cultural_Familiarization = Medium, Language_Proficiency = High, Communication_Mode = Written, Communication_Error_Type = Non_communication)	(Cultural_Familiarization = High, Language_Proficiency = Intermediate, Communication_Mode = Verbal, Communication_Error_Type = Miscommunication)	(Cultural_Familiarization = Medium, Language_Proficiency = High, Communication_Mode = Verbal, Communication_Error_Type = Miscommunication)
Behavioral_Misinterpretation	(Cultural_Familiarization = Medium, Language_Proficiency = High, Communication_Mode = Written, Communication_Error_Type = Non_communication)	Communication_Error_Type = Miscommunication (Language_Proficiency = High, Cultural_Familiarization = High)	Communication_Error_Type = Miscommunication (Language_Proficiency = Intermediate, Cultural_Familiarization = Low)	Communication_Error_Type = Miscommunication (Language_Proficiency = High, Cultural_Familiarization = Medium)	Communication_Mode = Verbal Miscommunication (Language_Proficiency = Low, Cultural_Familiarization = High)

skill distribution rather than acting only through intermediate nodes. For Intermediate and High proficiency, "Language_Training = Certified" appears again as a top influence: it decreases the probability of Intermediate but increases the probability of High proficiency. In other words, the model treats certification as a mechanism that upgrades crew from Intermediate toward High proficiency, compressing the lower tail of the distribution.

Second, the results highlight the role of *combined* training and standardized communication practices. Several of the strongest entries for Intermediate and High proficiency involve bundles where crew are on container ships, have SMCP use = Yes, and operate under Certified or Informal training with moderate or high language diversity. These bundles indicate that, conditional on being deployed on technically demanding and communication-intensive ship types, certification and routine SMCP use are associated with a shift from Intermediate toward High proficiency. Conversely, Low proficiency is amplified in scenarios with informal training, even when SMCP is reportedly used, suggesting that standardized phraseology alone cannot compensate for insufficient baseline competence.

Third, vessel type and size act as contextual amplifiers rather than primary causes. Vessel_Size = Very_Large and Vessel_Type = Container_Ship appears among the top movers for Low, Intermediate, and High states. Large container ships are typically operated by bigger, more internationally diverse crews and face tighter operational constraints; the BN therefore uses vessel type/size as proxies for organizational practices and manning standards. For Low proficiency, Very Large ships and specific language-diversity patterns can increase the probability of Low proficiency when combined with weak training, whereas under strong certification regimes the same platforms become associated with High proficiency.

Finally, the presence of Vessel_Type = General_Cargo as a positive driver of Intermediate proficiency and a negative driver of High proficiency suggests a segment of the fleet where language skills plateau at functional but sub-optimal levels. This aligns with an interpretation where smaller, less standardized operators invest enough to reach "good enough" proficiency but do not systematically push crews toward certification-driven excellence. From a management perspective, the sensitivity structure points to targeted levers: expanding certified training coverage, enforcing SMCP-based drills especially on high-risk vessel types, and prioritizing language-upskilling programs for general cargo operations, where the potential uplift from Intermediate to High proficiency is largest.

From the category's perspective, the nodes against all three sensitivity cases have been grouped into categories in Table 11. This summary table highlights, at a categorical level, how the influence structure of the BN reorganizes when different target nodes are used in sensitivity analysis. A consistent pattern is that *human and cultural factors* dominate whenever the target remains within the human–communication domain. For Conflict_Type, the most influential parents and upstream drivers are Cultural Familiarization, Language Proficiency and the diversity-related nodes, together with Communication_Mode and Communication_Error_Type. This confirms that conflicts are not just a function of immediate communication breakdowns but of deeper latent conditions in crew composition and cultural integration.

When Cultural Familiarization itself is the target, the sensitivity structure shifts further "upstream". Diversity and Language_Training become the primary levers, and conflict or communication nodes drop out of the top-influence set. This is consistent with a causal layering in which training interventions and crew mix shape cultural familiarization, which in turn modulates conflict patterns, rather than the reverse. The presence of Vessel_Size (Very_Large) in the low and high familiarization analyses indicates that large ships tend to be crewed and operated in ways that systematically co-vary with cultural dynamics (e.g., more multinational crews, different manning scales), but vessel attributes remain secondary.

For Language Proficiency, the sensitivity is even more tightly

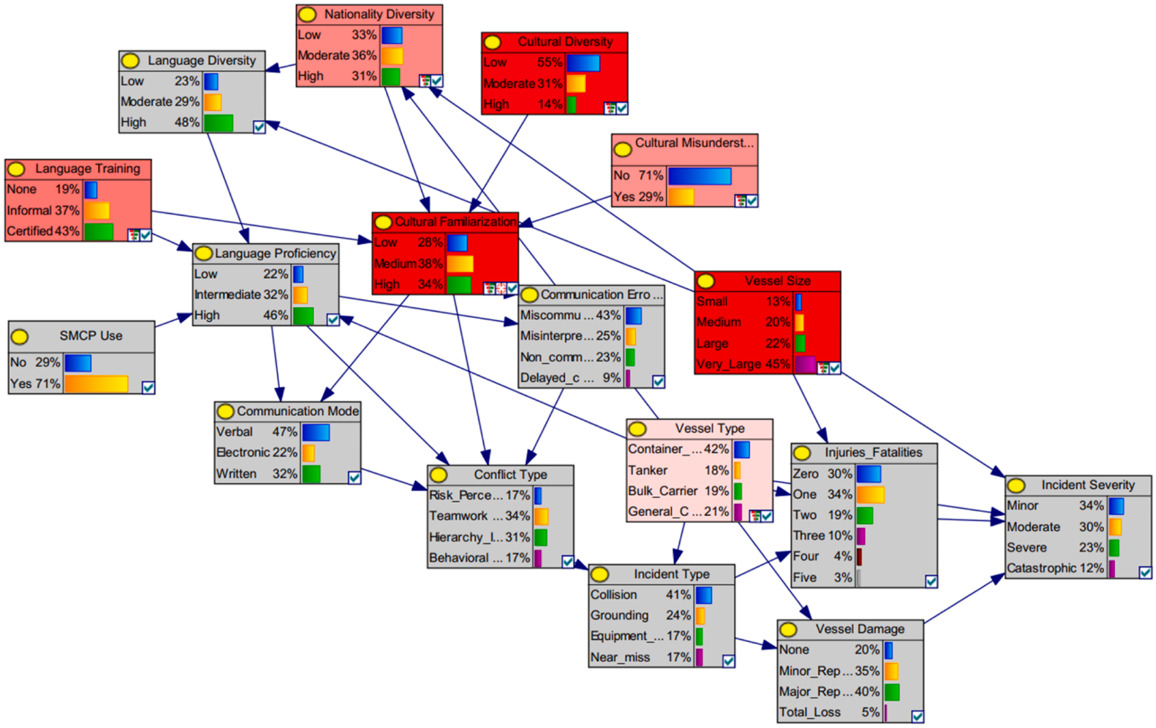


Fig. 5. Sensitivity Analysis for the node Cultural Familiarization.

Table 9
Top five most influential parameter configurations for each level of Cultural Familiarization.

Cultural familiarization state	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
Low	Cultural_Diversity = Low	Cultural_Familiarization = High (Cultural_Diversity = Low, Cultural_Misunderstanding = No, Nationality_Diversity = High, Language_Training = Informal)	Cultural_Familiarization = Low (Cultural_Diversity = Low, Cultural_Misunderstanding = No, Nationality_Diversity = Low, Language_Training = Certified)	Cultural_Familiarization = Medium (Cultural_Diversity = Low, Cultural_Misunderstanding = No, Nationality_Diversity = Low, Language_Training = Certified)	Vessel_Size = Very_Large
Medium	Cultural_Familiarization = Medium (Cultural_Diversity = Low, Cultural_Misunderstanding = No, Nationality_Diversity = Low, Language_Training = Certified)	Cultural_Familiarization = Medium (Cultural_Diversity = Low, Cultural_Misunderstanding = No, Nationality_Diversity = Moderate, Language_Training = Informal)	Cultural_Familiarization = Medium (Cultural_Diversity = Low, Cultural_Misunderstanding = No, Nationality_Diversity = High, Language_Training = Certified)	Cultural_Familiarization = Medium (Cultural_Diversity = Low, Cultural_Misunderstanding = No, Nationality_Diversity = Moderate, Language_Training = Certified)	Cultural_Familiarization = Medium (Cultural_Diversity = Low, Cultural_Misunderstanding = No, Nationality_Diversity = Low, Language_Training = Informal)
High	Cultural_Familiarization = High (Cultural_Diversity = Low, Cultural_Misunderstanding = No, Nationality_Diversity = High, Language_Training = Informal)	Vessel_Size = Very_Large	Cultural_Familiarization = High (Cultural_Diversity = Low, Cultural_Misunderstanding = No, Nationality_Diversity = High, Language_Training = Certified)	Cultural_Familiarization = High (Cultural_Diversity = Low, Cultural_Misunderstanding = No, Nationality_Diversity = Moderate, Language_Training = Certified)	Cultural_Familiarization = Medium (Cultural_Diversity = Low, Cultural_Misunderstanding = No, Nationality_Diversity = High, Language_Training = Certified)

concentrated in training and procedural controls: Language_Training, SMCP_Use and Language_Diversity are consistently dominant, while communication, conflict, and accident nodes are absent from the top-influence set. Vessel_Type (Container, General Cargo) and extreme vessel sizes appear as contextual modifiers rather than primary drivers. Importantly, across all three sensitivity configurations, *accident features* (injuries, damage, severity, incident type) never enter the top-five influencers. This reinforces the interpretation that they are downstream manifestations of a chain led by human, cultural, and communication factors, rather than direct control variables. Collectively, the table clarifies a coherent, layered influence architecture: structural crew composition and training → cultural familiarization and language proficiency → communication mode and error patterns → conflict type → accident consequences.

Taken together, these results demonstrate that sensitivity analysis does more than highlight statistical dependencies; it provides a risk-informed basis for prioritization. By identifying the nodes whose variability most strongly alters safety outcomes, the analysis highlights leverage points where interventions such as enhanced language training, rigorous SMCP enforcement, or tailored crewing policies for

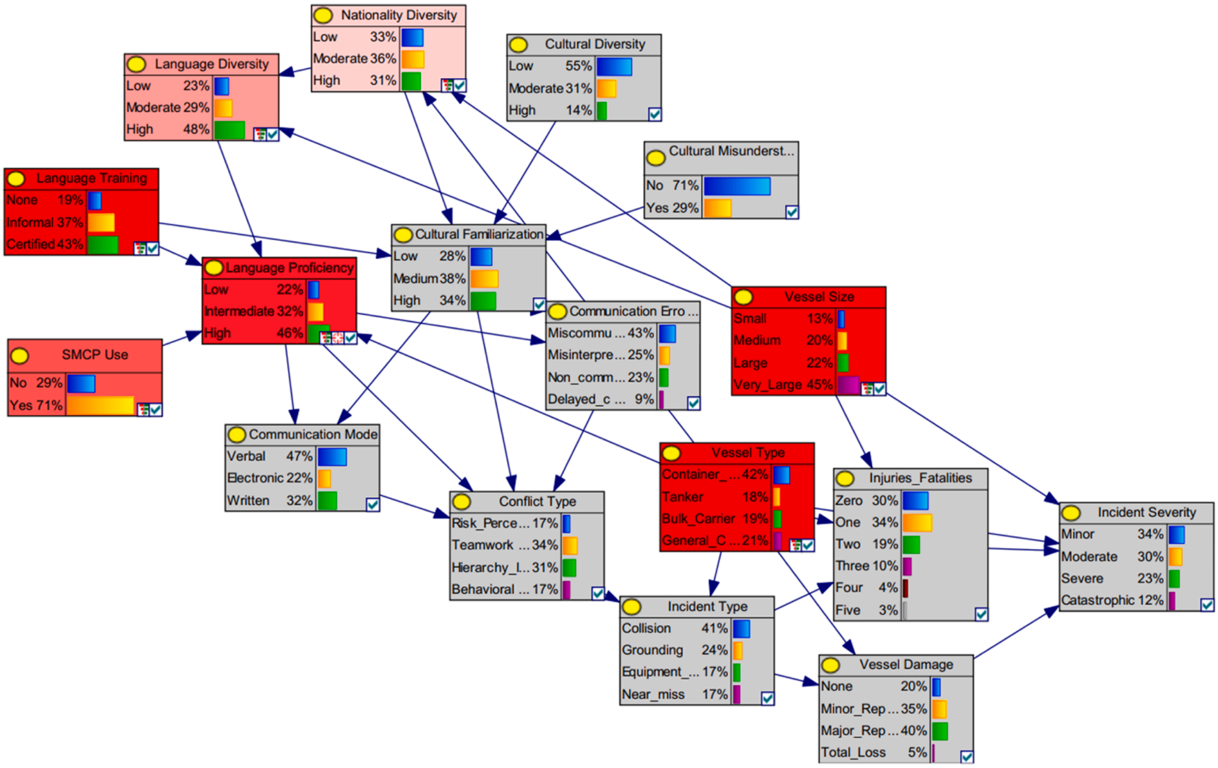


Fig. 6. Sensitivity Analysis for the Node Language Proficiency.

Table 10
Top five most influential parameter configurations for each level of Language Proficiency.

Language proficiency state	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
Low	Language_Training = Certified	Language_Proficiency = Low (Language_Training = Informal, SMCP_Use = Yes, Language_Diversity = High, Vessel_Type = Container_Ship)	Vessel_Size = Very_Large	Language_Training = None	Language_Diversity = Low (Nationality_Diversity = Low, Vessel_Size = Medium)
Intermediate	Language_Proficiency = High (Language_Training = Certified, SMCP_Use = Yes, Language_Diversity = High, Vessel_Type = Container_Ship)	Language_Proficiency = Intermediate (Language_Training = Certified, SMCP_Use = Yes, Language_Diversity = High, Vessel_Type = Container_Ship)	Vessel_Type = Container_Ship	Vessel_Type = General_Cargo	Language_Proficiency = High (Language_Training = Informal, SMCP_Use = Yes, Language_Diversity = Moderate, Vessel_Type = Container_Ship)
High	Language_Training = Certified	Language_Proficiency = High (Language_Training = Certified, SMCP_Use = Yes, Language_Diversity = High, Vessel_Type = Container_Ship)	Vessel_Type = Container_Ship	Language_Proficiency = High (Language_Training = Informal, SMCP_Use = Yes, Language_Diversity = Moderate, Vessel_Type = Container_Ship)	Vessel_Type = General_Cargo

large vessels can reduce accident risk most effectively. This approach, consistent with practices in other safety-critical domains such as nuclear and aviation, underscores how probabilistic modeling can directly inform resource allocation, policy design, and resilience building in maritime operations.

These results are further complemented by the quantitative display of the role of each factor for all three scenarios in the Appendix A.

TRI Analysis

In addition to the sensitivity analysis implemented in GeNIe, a True Risk Influence (TRI) check was performed to verify and refine the ranking of key nodes. Following [16,50,51], TRI measures how strongly a parent variable can shift a target outcome by forcing that parent into its most and least risk-favorable states. For each parent node, the state that produces the largest absolute increase in the probability of a given target state defines the high-risk influence (HRI), while the state that produces the smallest absolute change (either increase or decrease) defines the low-risk influence (LRI). TRI is then computed as the average of HRI and LRI, expressed in percentage points. This provides a compact measure of how much "room to move" a variable has over a target outcome, complementing mutual information by directly quantifying the magnitude of probability shifts under extreme but plausible configurations.

The TRI results for incident severity (Table 12) show that Incident Type is the dominant driver of outcome severity in the model. Switching the incident type between collision, grounding, equipment failure, and near miss produces sizeable changes in the probability of each severity band, with TRI values of 12.5 percentage points for Minor, 6.0 for Moderate, 4.5 for Severe, and 1.5 for Catastrophic outcomes. This means that, among the variables examined, the *physical expression* of the accident (e.g., collision vs. near miss) is the primary factor reshaping the

Table 11
Summary of dominant node categories across sensitivity analyses for Conflict Type, Cultural Familiarization, and Language Proficiency.

Target node (sensitivity)	Human/cultural factors	Communication/conflict factors	Vessel features	Accident features
Conflict_Type (all four states)	Dominant: Cultural_Familiarization, Language_Proficiency, background Cultural_Diversity	Dominant: Communication_Mode (Verbal, Written, Electronic), Communication_Error_Type (Miscommunication, Non-communication, Misinterpretation), Conflict_Type states	Minor / not in top-5 across states	Not present in top-5 for any state
Cultural_Familiarization (Low/Med/High)	Dominant: Cultural_Diversity, Nationality_Diversity, Language_Training, Cultural_Misunderstanding	Not present in top-5 (only linked indirectly via Conflict_Type elsewhere in the network)	Present: Vessel_Size = Very_Large (in Low and High familiarity runs)	Not present in top-5
Language_Proficiency (Low/Int/High)	Dominant: Language_Training, SMCP_Use, Language_Diversity (with some Nationality_Diversity context)	Not present in top-5 for any state	Present: Vessel_Type = Containership, General_Cargo, Vessel_Size (Small, Very_Large in Low proficiency run)	Not present in top-5

Table 12
TRI of Incident Type on Incident Severity.

Target node	Parent variable	TRI_Minor (%)	TRI_Moderate (%)	TRI_Severe (%)
Incident Severity	Incident_Type	12.5	6.0	4.5

severity profile once an incident has occurred. From a reliability perspective, this is consistent with the idea that the "failure mode" itself governs the escalation potential of an event, while other factors identified earlier in the BN mostly act upstream by altering the likelihood that a given incident type materializes in the first place.

At the conflict layer, Table 13 highlights that Conflict Type is most sensitive to the mechanism of communication failure and the crew's language proficiency. Communication Error Type shows the highest TRI for Teamwork Issues, with a shift of about 4 percentage points, indicating that changing the dominant error mechanism (e.g., from miscommunication to non-communication or delayed communication) substantially redistributes conflict from softer disagreement states toward more disruptive teamwork breakdowns. Language Proficiency exhibits similarly elevated TRI values (around 3.5 percentage points) for both Teamwork Issues and Hierarchy Issues, implying that even modest changes in proficiency meaningfully alter how often conflicts crystallize as coordination failures or authority-gradient problems. In reliability terms, these results confirm that, within the conflict sub-structure, *how well people can communicate* and *what kind of errors they make* are the key levers shaping the conflict pattern, while other factors play a comparatively secondary role.

Taken together, the TRI analysis complements the earlier prediction and sensitivity results by quantifying where small changes in key human-factor variables produce the largest shifts in risk-relevant states. At the outcome level, focusing on controls that prevent high-severity incident types (e.g., collision/grounding) offers the greatest leverage on severity. At the human-interaction level, targeted improvements in language proficiency and in preventing/repairing communication errors are the most efficient levers for reshaping the conflict landscape away from teamwork and hierarchy breakdowns and towards more manageable, lower-impact disagreement patterns.

Table 13
TRI of the Conflict Type under direct parent nodes.

Target node	Parent variable	Conflict state	TRI (%)
Conflict Type	Language_Proficiency	Teamwork_Issues	3.5
Conflict Type	Language_Proficiency	Hierarchy_Issues	3.5
Conflict Type	Communication_Error_Type	Risk_Perception	3.5
Conflict Type	Communication_Error_Type	Teamwork_Issues	3.5
Conflict Type	Communication_Error_Type	Hierarchy_Issues	4.0

3.5. *Scenario analysis*

For the scenario analysis, we interpret the BN under specific "what-if" crew profiles to see how the joint configuration of language and cultural factors reshapes the entire accident pathway. Scenario 1 represents a deliberately stressed communication setting that is still realistic for large commercial shipping, and it is used to explore how much risk is created *purely* by the interaction of human and communication factors, with the physical ship condition held broadly constant.

Scenario 1 – Highly diverse but poorly standardized bridge team (Fig. 7)

Evidence: Language Diversity = **High** (100%), Nationality Diversity = **High** (100%), Cultural Familiarization = **Low** (100%), Language Training = **Informal** (100%), SMCP Use = **No** (100%), Language Proficiency = **Low** (100%), Communication Mode = **Verbal** (100%), Vessel Size = **Very Large** (100%), Vessel Type = **Container** (100%).

In Scenario 1 we deliberately stack several adverse human-factor conditions to stress-test the model: high nationality and language diversity, low cultural familiarization, low language proficiency, only informal language training, no SMCP use, verbal communication as the dominant mode, and a very large container vessel (Fig. 7). This configuration represents a realistic but vulnerable bridge team on a deep-sea liner, where both linguistic and cultural safeguards are weak.

Given this evidence, the Communication Error Type node concentrates probability in miscommunication (45%) and misinterpretation (22%), with non-communication at 19% and delayed communication at 14%. Compared with the prior prediction (43%, 25%, 17%, 9%), the mass shifts away from "clean" misinterpretation toward failures to speak up or to repair errors once detected. In other words, the model does not simply say that messages are noisy; it infers a team that struggles to *correct* misunderstandings because crew members lack both the linguistic confidence and the shared phraseology (SMCP) to challenge unclear instructions.

This pattern propagates coherently into Conflict Type. Under Scenario 1, teamwork issues rise to 40% and hierarchy issues to 35%, while behavioral misinterpretation and risk-perception disputes fall to 14% and 10%. The model thus treats diversity under these conditions not as a source of redundancy but as a fragmenting force. High diversity with low familiarization encourages language and national sub-groups; in the absence of strong formal structures (certified training, SMCP), those sub-groups interact via steep authority gradients and ad-hoc translation. The BN, therefore, attributes most conflicts to coordination breakdowns and power distance, rather than to differences in individual risk perception.

At the accident layer, the Incident Type node moves toward a mixed collision/grounding regime: collision 37%, grounding 26%, equipment failure 21%, and near miss 16%. Relative to the baseline (41%, 24%, 17%, 17%), technical failures remain secondary, while geometry- and navigation-driven events gain weight, consistent with a system where hardware is intact but the navigational barrier is eroded by weak communication. Consequences follow the same direction: Incident

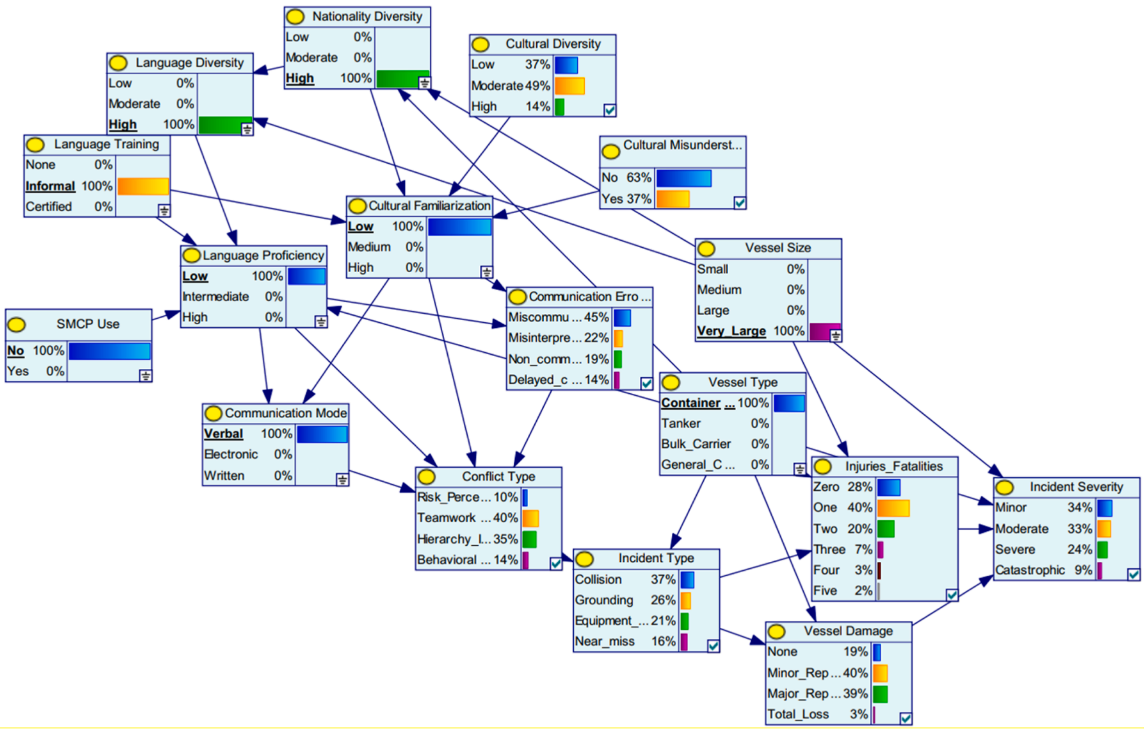


Fig. 7. Model results for scenario 1.

Severity is Minor 34%, Moderate 33%, Severe 24%, Catastrophic 9%; the probability of at least one fatality increases (injuries/fatalities "One" = 40%), and Vessel Damage is dominated by major repair (~39%) and total loss (~23%), with "no damage" and minor repair together below 40%. In reliability terms, Scenario 1 pushes the system into a regime of frequent, medium-to-high impact navigational accidents: not every event is catastrophic, but the combination of high diversity, low proficiency, and weak standardization systematically amplifies the cost and recoverability of errors. These trends point directly to targeted interventions, mandatory SMCP and certified language training on highly diverse, very large container ships, structured cultural familiarization, and closed-loop verbal protocols, as the most effective levers for risk reduction.

Scenario 2 – Standardized communication on a well-trained

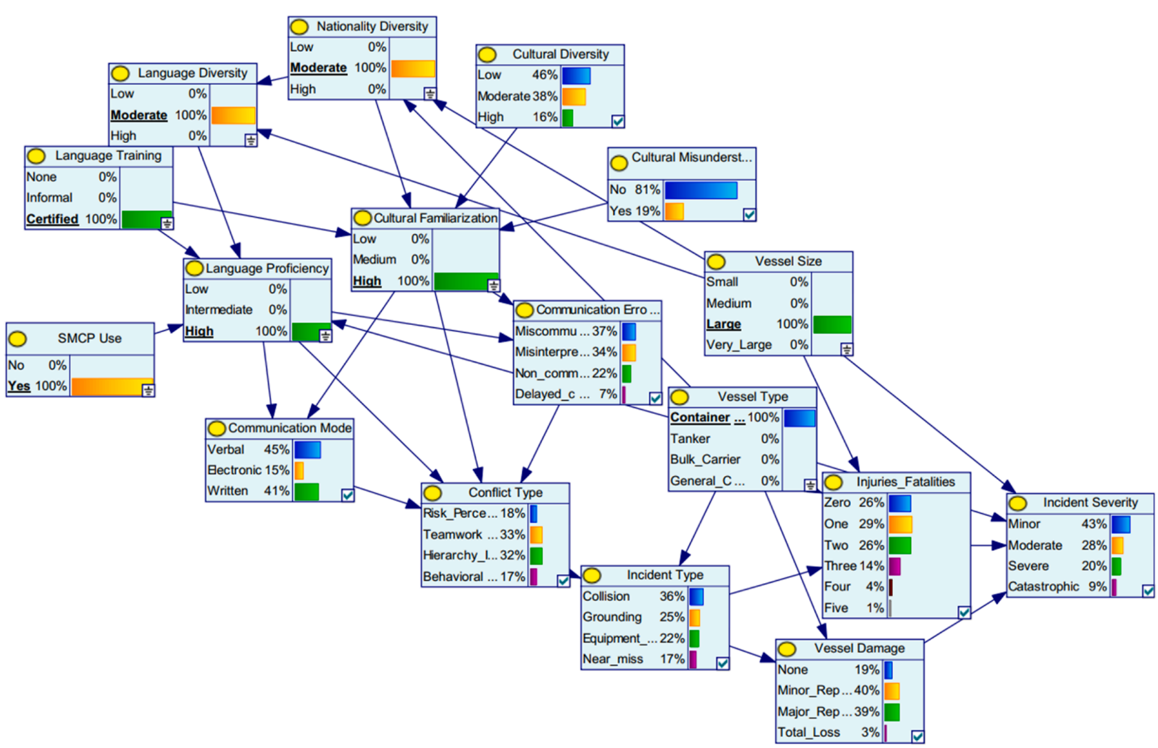


Fig. 8. Model results for scenario 2.

multicultural bridge (Fig. 8)

Evidence: Language Diversity = *Moderate* (100%), Nationality Diversity = *Moderate* (100%), Cultural Familiarization = *High* (100%), Language Training = *Certified* (100%), Language Proficiency = *High* (100%), SMCP Use = *Yes* (100%), Vessel Size = *Large* (100%), Vessel Type = *Container vessel* (100%) (Fig. 8).

In this scenario, the crew is still multinational and linguistically diverse, but all key "control knobs" on the human–language side are tuned to a best-practice setting: high proficiency, certified training, systematic SMCP use, and strong cultural familiarization. The network responds by reshaping both the error structure and the consequences. At the Communication Error node, miscommunication remains the single largest mode at 37%, but misinterpretation rises to 34%, while non-communication and delayed communication fall to 22% and 7%, respectively. Compared with Scenario 1 (non-communication 19%, delayed communication 14%), the share of *irreversible* breakdowns shrinks, and the share of *interpretable* errors grows. In other words, the BN is inferring that, under high proficiency and standardized phraseology, messages can still be imperfect, but they are more likely to be spoken, heard, and corrected within the operating cycle.

Conflict dynamics show a similar stabilizing pattern. Teamwork issues and hierarchy issues remain prominent at 33% and 32%, but they are no longer as dominant as in Scenario 1 (40% and 35%). Risk perception and behavioral misinterpretation together carry 35% (18% and 17%), indicating that disagreements increasingly concern *how risks are read* and *how behavior is interpreted*, rather than pure breakdowns in coordination or authority gradients. High cultural familiarization and moderate diversity reduce the tendency to fragment into national or language sub-groups; combined with certified training and SMCP, this allows more horizontal challenge and clarification, so that hierarchy conflicts no longer monopolize the conflict space.

At the accident layer, the Incident Type profile is a balanced mix of collision (36%), grounding (25%), equipment failure (22%), and near miss (17%). Relative to the prediction baseline, collision loses some dominance, and both equipment failures and near misses gain share, which is consistent with a system that still operates in demanding environments but manages to arrest part of the error chain before it becomes a full-blown navigational accident. This is reflected clearly in the outcome nodes. Incident Severity shifts towards less damaging states: Minor incidents rise to 43% (from 34% in the baseline and 34% in Scenario 1), while Severe cases drop to 20% and Catastrophic remains at 9%. Vessel Damage shows the strongest improvement: minor repair and major repair account for 40% and 39%, but total loss collapses to only 3%, compared with about 23% in Scenario 1. Injuries and fatalities remain present (e.g., "One" and "Two" each at 29–26%), but the overall profile is one of *contained* harm rather than systemic failure.

The key insight from Scenario 2 is that high diversity is not inherently risky; its impact depends on whether it is embedded in a structured communication regime. When diversity is coupled with certified training, SMCP, and strong cultural familiarization, the BN reinterprets many potential breakdowns as repairable misinterpretations rather than silent or unresolved failures. Operationally, this scenario supports policies that tie crew composition directly to language training and standardized communication protocols: large container ships with multinational crews can be made substantially more resilient not by reducing diversity, but by professionalizing how that diversity is managed on the bridge.

Scenario 3 – Diagnostic back-inference from a catastrophic container-ship collision (Fig. 9)

Evidence: Incident Type = *Collision* (100%), Incident Severity = *Catastrophic* (100%), Vessel Type = *Container vessel* (100%) (Fig. 9).

In Scenario 3, we reverse the logic of the previous cases: instead of asking "what happens if the crew profile is adverse?", we ask "given that a catastrophic collision of a container ship has occurred, what crew and conflict configuration does the network infer as most plausible?". Under this evidence (Fig. 9), the BN pushes the consequence space to the extreme: Incident Severity is fixed at 100% catastrophic, and Injuries/Fatalities shifts from a baseline with many zero-injury events to a profile dominated by multiple casualties (Two = 38%, One = 33%, Three = 12%, Four = 8%, Zero only 8%). Vessel Damage is likewise concentrated in high-impact states, with Major Repair at 58% and Total Loss at 3%, while "None" drops to 11%. The vessel context is implicitly one of large

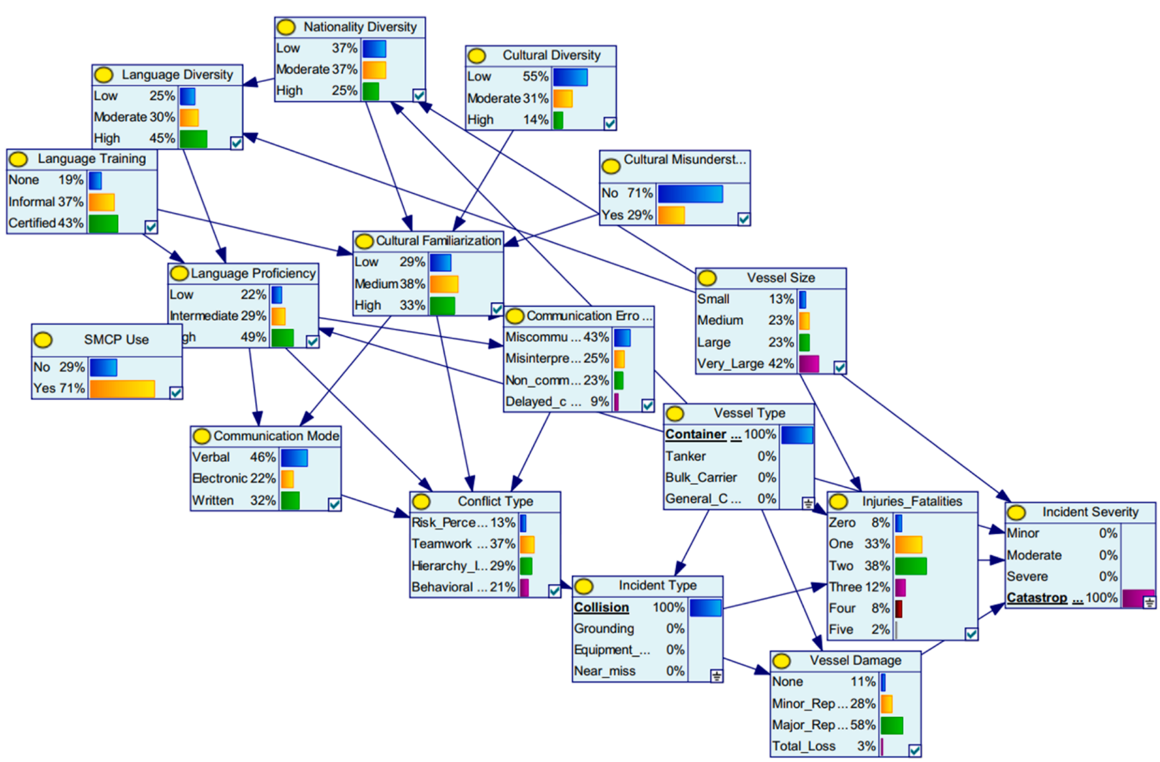


Fig. 9. Model results for scenario 3.

kinetic energy: the posterior for Vessel Size is skewed toward Very Large (42%) and Large (23%), with only 13–23% for small and medium ships.

What is more interesting, however, is what *does not* change dramatically. The human-factor nodes remain close to their priors: Language Proficiency still favors the High state (49%), Language Training keeps a balanced distribution with Certified at 43% and Informal at 37%, and SMCP Use remains mostly "Yes" (71%). Cultural Familiarization stays centered around Medium (38%) and High (33%), and Language and Nationality Diversity both sit in mixed, moderate–high regimes. In other words, the network does not diagnose catastrophic collisions as primarily a product of "worst-case" language or cultural profiles. Instead, it indicates that such events can arise even under relatively competent and formally trained crews, once certain exposure conditions (container ship, large/very-large size) and accident types (collision) are present.

The main structural shift appears at the Conflict Type node. Given a catastrophic collision, the model allocates 37% to Teamwork Issues and 29% to Hierarchy Issues, with Behavioral Misinterpretation and Risk-Perception disagreements at 21% and 13%. Compared with the prediction baseline, teamwork and behavioral conflicts both rise, suggesting that catastrophic collisions are associated less with pure technical failure and more with breakdowns in collective sense-making under time pressure: orders may be linguistically clear, but not accepted, coordinated, or executed consistently across the bridge team. This diagnostic reading has two important implications. First, it cautions against an overly simplistic narrative that "catastrophic outcomes equal low proficiency"; the BN instead highlights that once traffic density, ship size, and collision geometry align, even trained, high-proficiency crews remain vulnerable if team coordination and authority management fail. Second, it points to targeted interventions beyond training alone: bridge resource management that explicitly addresses cross-cultural teamwork, challenge-and-response across ranks, and decision-making under high-energy collision risk is essential, because these are the levers the model implicitly links to catastrophic impact when the physical accident type is fixed.

Scenario 4 – Equipment-failure incident on a tanker with moderate consequences (Fig. 10)

Evidence: Vessel_Type = *Tanker* (100%), Incident_Type = *Equipment failure* (100%), Incident_Severity = *Moderate* (100%) (Fig. 10).

In Scenario 4, the BN is deliberately steered toward a technically driven accident pathway, reflecting that not all incidents are primarily shaped by language or cultural diversity. Once we fix the vessel as a tanker and the accident as an equipment failure with moderate severity, the network produces a markedly different pattern from the collision-driven scenarios. At the outcome layer, the model locks the severity node into the Moderate band (100%), and Vessel Damage concentrates in Major Repair (61%) with only 1% in Total Loss and 19% in "No damage". Injuries and fatalities remain relatively dispersed but skewed toward lower harm: Zero and One together account for 61%, while "Two"–"Five" collectively occupy 39%. This profile is fully consistent with a controlled, technically contained failure that is costly in terms of repair and downtime but rarely escalates into mass-casualty or hull-loss events.

Upstream, the crew profile remains close to a "standard" configuration rather than drifting toward the extreme states seen in Scenario 1. Language Proficiency is slightly biased toward High (47%, Intermediate 31%, Low 23%), Language Training has a healthy share of Certified (44%) alongside Informal (37%), and SMCP use is predominantly "Yes" (71%). Cultural Familiarization stays centered around Medium (38%) and High (34%), and both Language and Nationality Diversity are broadly distributed. In other words, the BN does not infer that a moderate equipment-failure scenario requires an especially problematic linguistic or cultural configuration; normal levels of proficiency, diversity, and training are sufficient to generate such events.

The main shifts occur in the intermediate socio-cognitive layer. Communication Error Type still places most probability in miscommunication (44%) and misinterpretation (25%), but non-communication (22%) and delayed correction (9%) retain non-negligible mass.

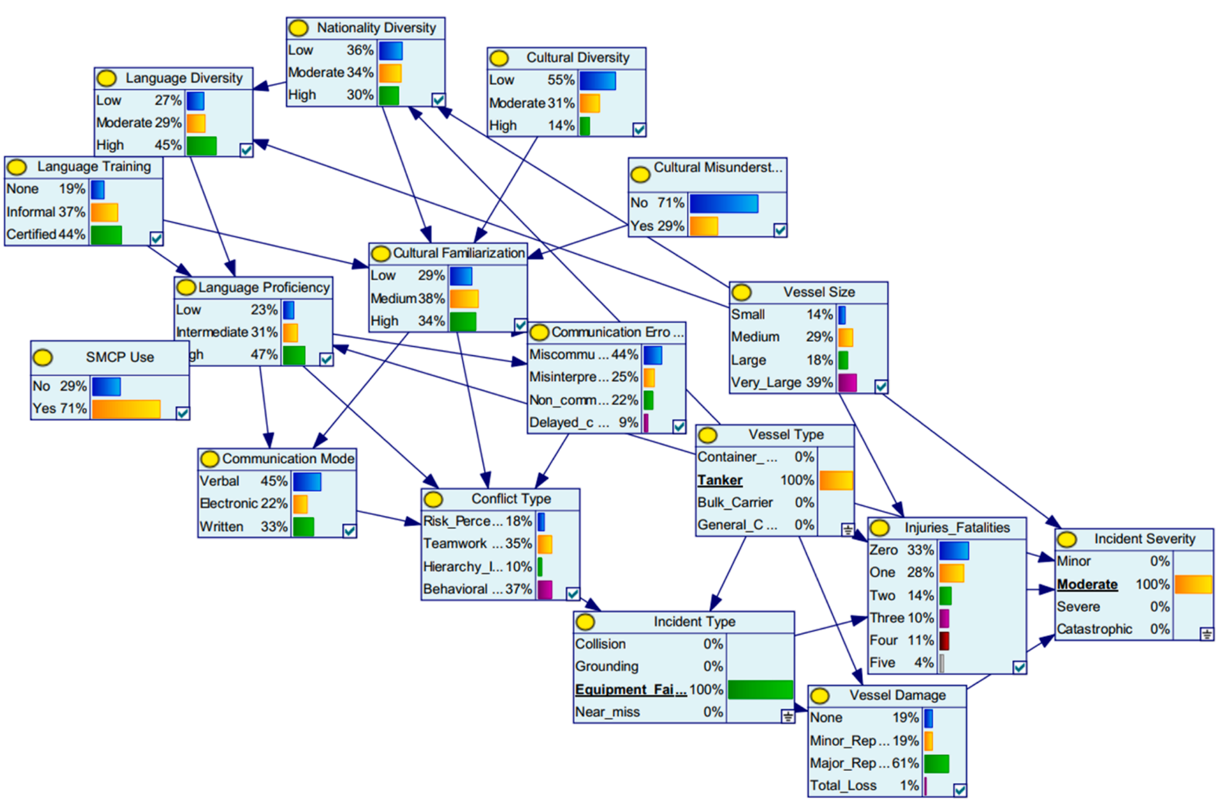


Fig. 10. Model results for scenario 4.

Conflict Type, however, tilts toward Behavioral Misinterpretation (37%) and Teamwork Issues (35%), with Hierarchy Issues smaller (10%) and Risk-Perception disputes at 18%. This combination suggests a qualitatively different dynamic from the collision scenarios. Here, the initiating trigger is technical, but the way the crew detects, interprets, and manages the failure is shaped by day-to-day behavioral and teamwork frictions, misread alarms, incomplete handovers, or inconsistent adherence to procedures, rather than by outright language breakdown or rigid authority gradients.

From a reliability and risk-management standpoint, Scenario 4 carries two important implications. First, it empirically supports that some incident types sit on a "technical majority" pathway: even with reasonably strong language, training, and familiarization, tankers experience moderate-severity equipment failures that primarily require engineering controls (maintenance quality, condition monitoring, redundancy) rather than radical changes to crew composition. Second, however, the elevated probabilities for behavioral and teamwork conflicts indicate that human factors still shape *how costly* these failures become. The network is effectively saying that, given an equipment failure on a tanker, the difference between minor disturbance and major repair lies less in language proficiency per se and more in the micro-coordination of the team: whether checklists are followed, small anomalies are communicated, and corrective actions are aligned across ranks.

Thus, Scenario 4 complements the earlier, human-factor-dominated scenarios by showing that the BN can distinguish between crew-sensitive and hardware-dominated accident pathways. It reinforces a nuanced intervention strategy: for tanker equipment failures, investment in technical reliability and maintenance regimes is primary. While human-factor efforts should focus on everyday teamwork discipline and behavioral consistency rather than on wholesale redesign of language and diversity policies.

3.6. Strength of influence analysis

The strength of influence analysis, computed using mutual information metrics from the developed BN, quantifies the interdependencies among linguistic, cultural, operational, and accident outcome variables. Unlike prediction and sensitivity analyses, which highlight probabilities and conditional vulnerabilities, this approach exposes structural couplings, that is, how uncertainty in one node propagates across the system and magnifies downstream risks. Such measures are widely used in reliability and system safety to identify dominant risk pathways and to assess whether failures cluster in particular subsystems or diffuse across multiple layers of interaction [12,76].

The analysis reveals several key dynamics. As evident in Table 10, vessel size exerts one of the strongest influences, shaping both nationality diversity (0.54) and language diversity (0.46). Larger ships tend to recruit more multinational crews, which increases language variation and thereby elevates communication complexity. This structural driver creates a higher baseline probability of miscommunication, a classic "latent condition" that aligns with how design parameters in engineered systems can embed systemic risk into operations. Nationality diversity, in turn, is the strongest driver of language diversity (0.52), reinforcing the link between crewing policies and communication outcomes. Both diversity indices cascade into language proficiency, primarily through their impact on training needs and cultural familiarization. Importantly, language training (0.26) and SMCP use (0.24) directly strengthen proficiency, acting as preventive safety barriers that can counterbalance structural risk introduced by vessel characteristics and crew heterogeneity.

Downstream effects are equally critical. Language proficiency strongly impacts communication error type (0.28) and conflict type (0.28), with cultural familiarization providing a mediating role (0.27). This pathway illustrates that failures in training or standardization do not remain isolated; instead, they propagate into operational conflicts that escalate into incident outcomes. At the accident level, incident type has a substantial influence on both vessel damage (0.40) and casualties (0.32), while injuries/fatalities (0.21) and vessel damage (0.22) jointly drive incident severity. This layered escalation mirrors accident progression models in safety-critical industries, where upstream human and organizational factors magnify into technical and physical losses if not intercepted by robust barriers [35,49,65].

A further insight emerges from the dual role of diversity. While cultural and nationality diversity introduce challenges in communication and familiarization, they also present opportunities for resilience when coupled with adequate training. For instance, cultural misunderstanding influences familiarization (0.16), but when addressed through structured cultural training, this pathway can be reoriented into a resilience-enhancing feature, much like redundancy or diversity in engineered systems that, when properly managed, improve overall robustness. Thus, the analysis highlights not just risk amplification but also potential leverage points for adaptive safety management.

Collectively, the results underscore that the most influential pathways are not evenly distributed but cluster around crew composition, training interventions, and vessel design parameters. This aligns with reliability engineering principles, where a small subset of high-impact variables typically dominates system risk. By quantifying these influence scores, the analysis provides a clear evidence base for prioritizing interventions such as enhanced language training, systematic SMCP enforcement, and tailored crewing strategies on large and container vessels. In practical terms, these measures serve the same role as engineered safeguards in technical systems: buffering high-risk pathways and preventing escalation into severe outcomes.

The detailed quantified results of the strength of influence analysis are presented in Table 14, which provides a comprehensive mapping of parent–child node relationships, influence magnitudes, and their implications for maritime safety management.

4. Discussion

This study quantitatively unpacks how language and cultural diversity, as measured by the Shannon Language Diversity Index and Hofstede Cultural Diversity Index, shape patterns of miscommunication and accident risk in global maritime operations. Several prominent insights emerge from the BN analysis, each with clear technical and organizational implications. These insights map to latent conditions, barrier performance, and escalation paths in socio-technical systems, providing a risk-informed basis for intervention prioritization [12,35,49,65,76]. To our knowledge, this is among the first studies to quantitatively link linguistic and cultural diversity with accident outcomes using BN strength-of-influence measures, and to interpret these findings explicitly through a reliability and safety lens.

First, the prediction and scenario analysis confirm that the BN provides credible forecasting of incident outcomes, with ROC–AUC and classification accuracy values demonstrating reliable model performance. This predictive validity ensures that the BN is not merely explanatory but can support proactive risk assessments in operational contexts. Within this validated framework, language diversity and proficiency are found to be the strongest upstream influencers of communication errors and subsequent safety incidents. The model's sensitivity and strength-of-influence results demonstrate that vessels with higher language diversity (i.e., crews composed of many equally represented languages) exhibit significantly elevated probabilities of communication error types such as misinterpretation and non-communication. This is consistent with prior findings by [10,26,32,34], who highlighted that not just the presence but the *balance* of languages matters: heterogeneous crews may be unable to establish an effective "working language," increasing ambiguity and the risk of error. Our results extend this by showing that language proficiency is an even more sensitive variable; crews with uniformly low English proficiency amplify the effect of language diversity, compounding the risk of communication

Table 14
Results of the strength of influence analysis.

Parent Node	Child Node	Average Influence	Maximum Influence	Weighted Influence
Communication Error Type	Conflict Type	0.289131	0.625694	0.289131
Communication Mode	Conflict Type	0.293234	0.525811	0.293234
Conflict Type	Incident Type	0.173725	0.305375	0.173725
Cultural Diversity	Cultural Familiarization	0.180643	0.643377	0.180643
Cultural Familiarization	Communication Mode	0.093637	0.163846	0.093637
Cultural Familiarization	Communication Error Type	0.082631	0.153073	0.082631
Cultural Familiarization	Conflict Type	0.266153	0.586302	0.266153
Cultural Misunderstanding	Cultural Familiarization	0.155901	0.487499	0.155901
Incident Type	Injuries_Fatalities	0.324413	0.661438	0.324413
Incident Type	Vessel Damage	0.396973	0.782164	0.396973
Injuries_Fatalities	Incident Severity	0.212406	0.661438	0.212406
Language Diversity	Language Proficiency	0.252814	0.666667	0.252814
Language Proficiency	Communication Mode	0.073577	0.155922	0.073577
Language Proficiency	Communication Error Type	0.060681	0.121294	0.060681
Language Proficiency	Conflict Type	0.280909	0.678924	0.280909
Language Training	Language Proficiency	0.256545	0.647502	0.256545
Language Training	Cultural Familiarization	0.185357	0.666745	0.185357
Nationality Diversity	Language Diversity	0.515012	0.918954	0.515012
Nationality Diversity	Cultural Familiarization	0.195867	0.527069	0.195867
SMCP Use	Language Proficiency	0.235041	0.585314	0.235041
Vessel Damage	Incident Severity	0.222466	0.739651	0.222466
Vessel Size	Nationality Diversity	0.539557	0.830673	0.539557
Vessel Size	Language Diversity	0.462059	0.917466	0.462059
Vessel Size	Injuries_Fatalities	0.271237	0.661438	0.271237
Vessel Size	Incident Severity	0.164398	0.764875	0.164398
Vessel Type	Nationality Diversity	0.076593	0.136615	0.076593
Vessel Type	Language Proficiency	0.234906	0.647502	0.234906
Vessel Type	Incident Type	0.165469	0.273378	0.165469
Vessel Type	Injuries_Fatalities	0.274277	0.745666	0.274277
Vessel Type	Vessel Damage	0.137686	0.298685	0.137686
Vessel Type	Incident Severity	0.169216	0.800000	0.169216

breakdown. This supports recommendations in recent industry reports [63], advocating for rigorous minimum language requirements and regular proficiency assessment, not only for officers, but for all bridge and engine room personnel. In reliability terms, language diversity and proficiency function as performance-shaping factors whose states modulate the likelihood of communication failure, suggesting that targeted proficiency uplift yields high leverage on risk reduction [49,75].

The data-driven network reveals a chain of influence: vessel type and size shape the crew composition, which in turn drives language and nationality diversity. Larger, more modern vessels (especially bulk carriers and tankers) tend to recruit from a broader international pool, creating higher diversity scores. This trend, previously noted by [67], is found in our dataset as well and confirms that ship operators' crewing strategies have a direct, quantifiable impact on communication risk. As vessel size and complexity increase, so too does the need for formal communication protocols and the use of SMCP. Notably, SMCP usage was found to mitigate some, but not all, of the risk posed by diversity; however, its effectiveness depends on the baseline proficiency and adherence among crew, echoing the findings in literature [14,73]. Viewed through sensitivity and barrier analysis, SMCP adherence reduces incident probability, but strength-of-influence analysis shows its effect is conditional on upstream performance shaping factors (proficiency, training), aligning with systems-theoretic views of layered defenses [35,49,65].

A critical insight from the BN analysis is that cultural diversity operates through multiple communication-related pathways. Higher Hofstede Cultural Diversity Index scores correlate with increased instances of cultural misunderstanding and conflict, which then influence communication mode selection (e.g., greater reliance on digital/written modes when verbal exchanges are strained) and, in turn, shape conflict type and ultimately incident type and severity. This multi-step representation reflects mechanisms described in intercultural communication and team performance research, and is consistent with the patterns observed in the coded accident narratives, rather than being imposed solely by the BN structure. The BN structure's ability to trace these multi-step dependencies highlights why interventions focused solely on language are insufficient: cross-cultural awareness and familiarity must be built into crew resource management (CRM) and safety training. The finding that cultural familiarization training lowers both the incidence and severity of accidents is strongly supported by empirical studies in other safety-critical domains, such as aviation [33], and by the MARCOM Project [25] recommendations in shipping. This pattern reflects escalation chains commonly observed in complex systems: latent variability in culture and coordination degrades information flow, which then manifests as error-prone communication modes and conflict, unless intercepted by effective organizational and training barriers [35].

The model's sensitivity analysis spotlights primary communication mode and error type as key mediators. For example, the reliance on radio or digital communication (rather than face-to-face or written) is associated with higher rates of miscommunication, especially on culturally diverse vessels. This finding is consistent with existing literature [17], which showed that mode-switching can exacerbate misunderstandings when protocols are not uniformly understood or followed. Further, the severity and outcome of incidents (e.g., injuries/fatalities and vessel damage) are not only shaped by the initial communication error but are *amplified or buffered* by downstream factors like crew training, cultural conflict type, and SMCP adherence. The BN reveals that ships with low training levels and weak cultural familiarization suffer not just more frequent but also more severe incidents. This resonates with previous studies [4,26,29,73], which argued that technical solutions must be paired with human factor interventions to achieve meaningful safety gains. Linking these results to importance-measure logic, the nodes flagged as most sensitive act analogously to high Birnbaum/Fussell–Vesely components in PRA; improving their reliability (e.g., training, protocol enforcement) should yield disproportionate system-level safety gains [12,76].

Building on these findings, we strongly recommend that maritime regulators and operators implement mandatory, multilevel language proficiency certification, not only at initial recruitment but also through periodic re-testing and onboard language drills, with particular focus on junior and support ranks where undetected errors can propagate. Integrated cultural awareness and communication training modules should be embedded in CRM, targeting practical conflict resolution and bridging power distance gaps, an approach that has demonstrably reduced incident rates in the aviation sector [33]. Crew mix management policies should move beyond cost-driven recruitment to consider operational "diversity windows," ensuring language and cultural heterogeneity remain within thresholds that promote effective communication (for instance, limiting the number of major language groups per vessel). Additionally, protocols such as the SMCP should be expanded and rigorously audited, extending beyond navigation to routine and emergency procedures. Finally, incident investigation processes must systematically identify language and cultural contributions to safety outcomes and integrate these insights into continuous improvement and targeted training, fostering a cycle of organizational learning and resilience. Operationally, these actions instantiate barrier management: define leading indicators (e.g., proficiency distributions, SMCP audit scores, familiarization coverage), monitor drift, and allocate resources to the highest-influence nodes identified by the BN to maximize safety return on investment [35,49,65].

5. Conclusion

Maritime safety remains a critical global concern, as human error, especially stemming from language and cultural diversity, continues to contribute significantly to maritime incidents, despite advances in technology and regulation. The increasing prevalence of multicultural and multilingual crews onboard modern vessels presents unique challenges to effective communication, decision-making, and risk management. However, rigorous quantitative assessments of how linguistic and cultural diversity propagate into operational risk have been scarce, leaving a gap in both theory and practice. This study addresses this gap by systematically quantifying the pathways through which diversity influences accident causation, thereby strengthening the empirical foundation for crew resource management and policy design.

To fill this gap, this study utilized a data-driven BN modeling framework, analyzing 550 real-world maritime accident cases sourced from established databases and industry reports. The methodology entailed extracting structured variables on vessel and crew characteristics, calculating standardized indices for language (Shannon index) and cultural diversity (Hofstede index), and parameterizing the BN via expectation-maximization. The model underwent rigorous validation through ten-fold cross-validation and ROC/AUC analysis to ensure reliability. Prediction and sensitivity analyses further elucidated the influence of diversity-related variables and communication processes on incident outcomes and severity. Prediction, sensitivity, and strength-of-influence analyses were then conducted to trace how diversity-related variables, training factors, and communication processes shape incident outcomes and severity.

The findings reveal several critical insights. High language and cultural diversity, especially when paired with low language proficiency and inadequate training, substantially increases the probability of communication errors, specifically, misinterpretation and miscommunication events, which then cascade to elevate the risk and severity of accidents. Vessel type and size influence crew composition, further shaping diversity dynamics and their impact on safety. Sensitivity and strength-of-influence analyses identified key leverage points for risk reduction: targeted interventions to improve language proficiency, the consistent implementation of SMCP, and structured cultural familiarization programs all demonstrably reduce accident likelihood and severity. Notably, the study also highlights that communication mode and error type act as primary mediators, while cultural familiarization and SMCP compliance buffer the adverse effects of diversity.

The significance of this research is multifold. By moving the conversation from anecdotal evidence to robust quantitative analysis, the study offers actionable recommendations for maritime operators, regulators, and training institutions. It provides a strong empirical rationale for multilingual proficiency certification, integrated cross-cultural training, and systematic auditing of communication protocols. At the same time, this study has several limitations. First, it relies on retrospective accident investigation reports, which may suffer from underreporting, variable depth of detail, and occasional misclassification of contributing factors; some language- and culture-related mechanisms are therefore only partially observed. Second, key constructs such as language and cultural diversity are represented through indices and discretized categories, which necessarily simplify complex realities of crew interaction and may introduce measurement uncertainty. Third, the Bayesian Network is event-based and static, capturing the state of the system around recorded incidents rather than the full temporal evolution of crew dynamics, learning, and organizational drift. Fourth, all conditional probabilities are estimated from the same finite accident database (550 investigation reports) using the EM algorithm. As a result, some conditional distributions rest on relatively sparse empirical support and should be interpreted as conditional on this specific dataset rather than as universal industry frequencies. The underlying database was compiled for this study and is not yet continuously maintained; periodic updating and re-estimation as further investigation reports become available would strengthen the external validity of the model. Future work should therefore incorporate longitudinal crewing and operational data, extend the analysis to routine (non-accident) operations, and test targeted interventions (e.g., enhanced training or SMCP enforcement) through field studies or simulation to better quantify their impact on communication reliability and safety outcomes.

CRediT authorship contribution statement

Rafi Ullah Khan: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jingbo Yin:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Muhammad Afzaal:** Writing – review & editing, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Zaili Yang:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Acknowledgment

This study was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 864724), National Natural Science Foundation of China [Grant No. 52472362], Ministry of Industry and Information Technology for research on the key technology of the high-tech ocean passenger ship construction logistics collection system [MC-202009-Z03], and the Open Research Fund of State Key Laboratory of Maritime Technology and Safety [Grant No. W24CG000037].

Appendix A

Tornado Charts for the Detailed Quantitative Overview of the Sensitivity of Different Target Nodes and their States

Fig. A1. Tornado chart of sensitivity analysis for target node conflict type, state Risk Perception.

Fig. A2. Tornado chart of sensitivity analysis for target node conflict type, state Teamwork Issues.

Fig. A3. Tornado chart of sensitivity analysis for target node conflict type, state Hierarchy Issues.

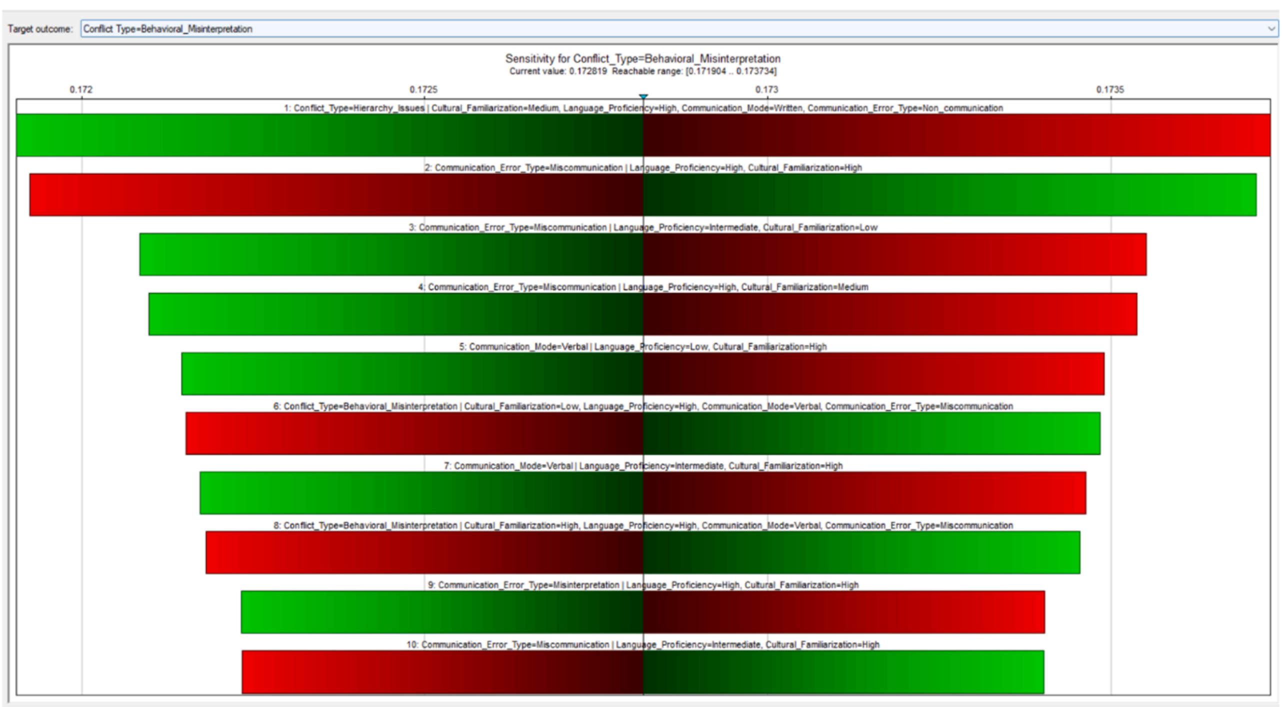


Fig. A4. Tornado chart of sensitivity analysis for target node conflict type, state Behavioral Misinterpretation.

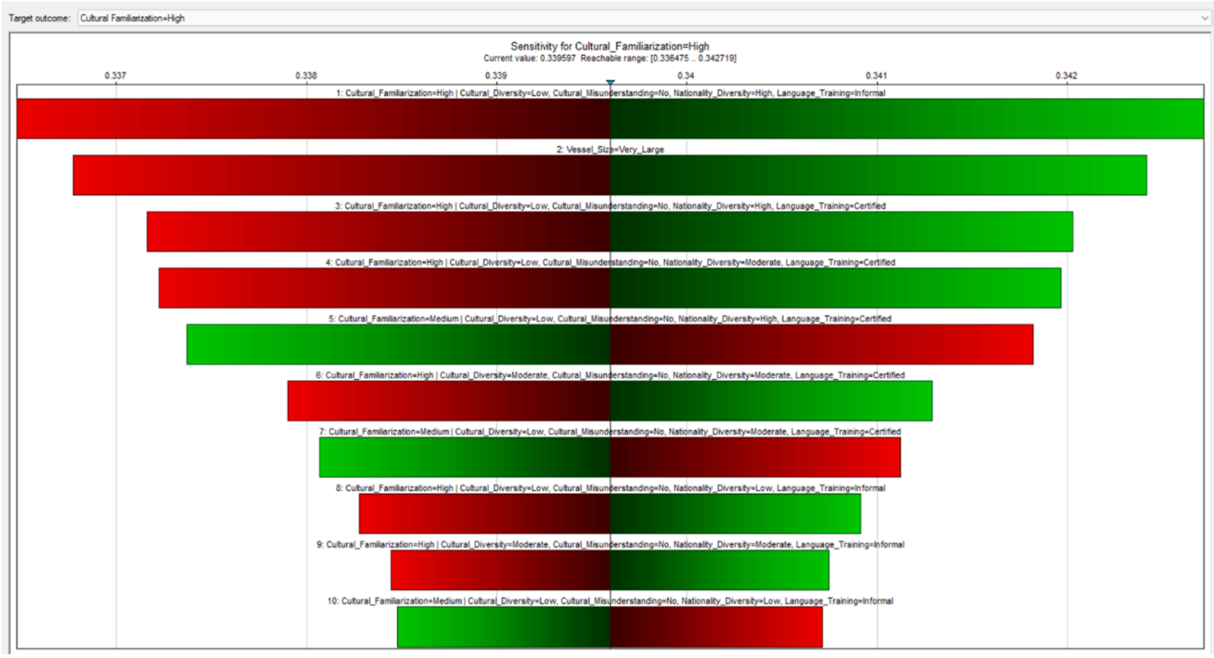


Fig. A5. Tornado chart of sensitivity analysis for target node Cultural Familiarization, state High.

Fig. A6. Tornado chart of sensitivity analysis for target node Cultural Familiarization, state Medium.

Fig. A7. Tornado chart of sensitivity analysis for target node Cultural Familiarization, state Low.

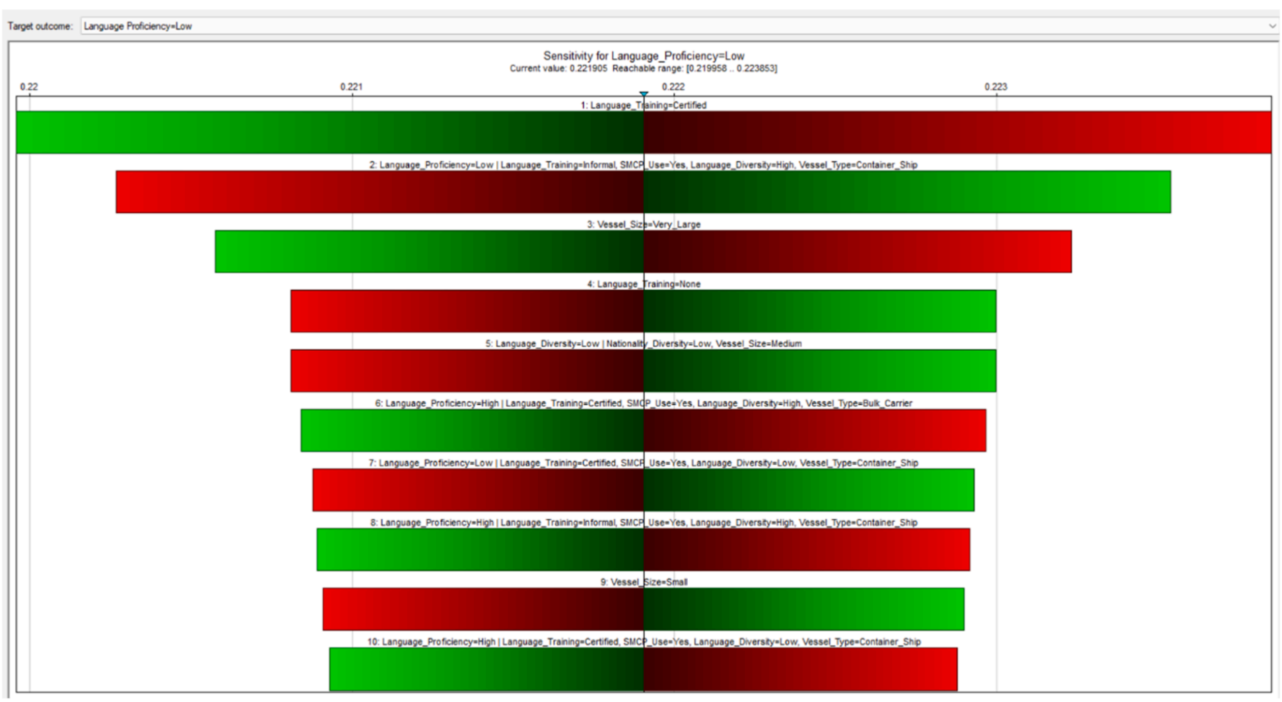


Fig. A8. Tornado chart of sensitivity analysis for target node Language Proficiency, state Low.

Fig. A9. Tornado chart of sensitivity analysis for target node Language Proficiency, state Medium.

Fig. A10. Tornado chart of sensitivity analysis for target node Language Proficiency, state High.

Data availability

The authors do not have permission to share data.

References

[1] ABREU DTMP, MATURANA MC, DROGUETT EL, MARTINS MR. Human reliability analysis of conventional maritime pilotage operations supported by a prospective model. Reliab Eng Syst Saf 2022;228:108763.

[2] ABRISHAMI S, KHAKZAD N, HOSSEINI SM. A data-based comparison of BN-HRA models in assessing human error probability: An offshore evacuation case study. Reliab Eng Syst Saf 2020;202:107043.

[3] ABRISHAMI S, KHAKZAD N, HOSSEINI SM, VAN GELDER P. BN-SLIM: A Bayesian Network methodology for human reliability assessment based on Success Likelihood Index Method (SLIM). Reliab Eng Syst Saf 2020;193:106647.

[4] AHMMED R. The discrepancy between standardised communication patterns and the real-life conversations of vessel traffic service: a case study in Chittagong Port, Bangladesh. WMU Journal of Maritime Affairs 2020;19:509–32.

[5] ANTÃO P, SOARES CG. Analysis of the influence of human errors on the occurrence of coastal ship accidents in different wave conditions using Bayesian Belief Networks. Accident Analysis & Prevention 2019;133:105262.

[6] ASADAYOOBI N, TAGHIPOUR S, JABER MY. Predicting human reliability based on probabilistic mission completion time using Bayesian Network. Reliab Eng Syst Saf 2022;221:108324.
[7] ASKARI HR, SHAHEEN MMA, CHOWDHURY MMH, ALSHAREEF MH, MUKTADIR DG. Analyzing factors contributing to inland waterway transport accidents in Bangladesh. Marine Systems & Ocean Technology 2025;20:27.
[8] BAYESFUSION, L. L. C. 2017. Genie modeler. User Manual. Available online: https://support.bayesfusion.com/docs/, 16, 30-32.
[9] BEI H, FANGJIAO Y, WENYANG W, TIANREN Y, MURCIO R. Maritime safety through multi-source data fusion: an AdaBoost-based approach for predictive ship detention by port state control. Maritime Policy & Management 2024:1–20.
[10] BERG N, STORGÅRD J, LAPPALAINEN J. The impact of ship crews on maritime safety. Publications of the Centre for Maritime Studies, University of Turku A 2013; 64:1–48.
[11] BOCANEGRA VALLE A. The Language of Seafaring: Standardized Conventions and Discursive Features in Speech Communications. International Journal of English Studies 2011;11:35–53.
[12] BORGONOVO E. A new uncertainty importance measure. Reliab Eng Syst Saf 2007; 92:771–84.
[13] BOSTRÖM, M. 2020a. Creating clarity and managing complexity through co-operation and communication: The case of Swedish icebreaker operations.
[14] BOSTRÖM M. Mind the Gap! A quantitative comparison between ship-to-ship communication and intended communication protocol. Saf Sci 2020;123:104567.
[15] BOSTRÖM M. Other-initiated repair as an indicator of critical communication in ship-to-ship interaction. J Pragmat 2021;174:78–92.
[16] CAO Y, WANG X, WANG Y, FAN S, WANG H, YANG Z, LIU Z, WANG J, SHI R. Analysis of factors affecting the severity of marine accidents using a data-driven Bayesian network. Ocean Engineering 2023;269:113563.
[17] CELIK M, CEBI S. Analytical HFACS for investigating human errors in shipping accidents. Accident Analysis & Prevention 2009;41:66–75.
[18] CHOWDHURY MN, SHAFI S, ARZAMAN M, FARHAN A, TEOH BA, KADHIM KA, SALAMUN H, ABDUL KADIR FK, SAID S, KADIR KA. Navigating Human Factors in Maritime Safety: A Review of Risks and Improvements in Engine Rooms of Ocean-Going Vessels. International Journal of Safety & Security Engineering 2024:14.
[19] DOMINGUEZ-PÉRY C, TASSABEHJI R, CORSET F, CHREIM Z. A holistic view of maritime navigation accidents and risk indicators: examining IMO reports from 2011 to 2021. Journal of Shipping and Trade 2023;8:11.
[20] DOMINGUEZ-PÉRY C, VUDDARAJU LNR, CORBETT-ETCHEVERS I, TASSABEHJI R. Reducing maritime accidents in ships by tackling human error: a bibliometric review and research agenda. Journal of Shipping and Trade 2021;6: 20.
[21] EVANGELOS T. Language barriers and miscommunication as a cause of maritime accidents. Macedonia: Merchant Marine Academy of Macedonia 2002.
[22] FAN H, JIA H, HE X, LYU J. Navigating uncertainty: A dynamic Bayesian network-based risk assessment framework for maritime trade routes. Reliab Eng Syst Saf 2024;250:110311.
[23] FAN S, BLANCO-DAVIS E, YANG Z, ZHANG J, YAN X. Incorporation of human factors into maritime accident analysis using a data-driven Bayesian network. Reliab Eng Syst Saf 2020;203:107070.
[24] FLIN R, O'CONNOR P. Safety at the sharp end: a guide to non-technical skills. CRC Press; 2017.
[25] FROHOLDT LL. I See You on My Radar': Displays of the Confirmatory Form in Maritime Technologically Mediated Interaction. Sociol Rev 2016;64:468–94.
[26] GABEDAVA G, HU Y. Enhancing maritime safety through linguistic analysis: a case study of communication failures in maritime accidents. WMU Journal of Maritime Affairs 2025.
[27] GALIERIKOVÁ A. The human factor and maritime safety. Transportation Research Procedia 2019;40:1319–26.
[28] GOERLANDT F, MONTEWKA J. Maritime transportation risk analysis: Review and analysis in light of some foundational issues. Reliab Eng Syst Saf 2015;138:115–34.
[29] HARYANI H, JANUARIUS M, RUDI H, ISSY Y. The Essential of Using Standard Marine Communication Phrases in Avoiding Accidents at Sea for Indonesian Seafarers. International Journal of Education, Vocational and Social Science 2023; 2:51–67.
[30] HAVOLD JI. Culture in maritime safety. Maritime Policy & Management 2000;27: 79–88.
[31] HÅVOLD JI. National cultures and safety orientation: A study of seafarers working for Norwegian shipping companies. Work & Stress 2007;21:173–95.
[32] HÅVOLD JI, OLTEDAL HA. Culture and maritime safety. Managing Maritime Safety 2018:53–70.
[33] HELMREICH RL, MERRITT AC. Culture at work in aviation and medicine: National, organizational and professional influences. Routledge; 2017.
[34] HETHERINGTON C, FLIN R, MEARNS K. Safety in shipping: The human element. J Safety Res 2006;37:401–11.
[35] HOLLNAGEL E. Safety-I and safety-II: the past and future of safety management. CRC press; 2018.
[36] HORCK J. An analysis of decision-making processes in multicultural maritime scenarios. Maritime Policy & Management 2004;31:15–29.
[37] JASON, L. B. 2023. A Gentle Introduction to k-fold Cross-Validation [Online]. Available: https://machinelearningmastery.com/k-fold-cross-validation/?utm_source=chatgpt.com [Accessed 17th April 2025].
[38] JI C, GAO F, LIU W. Dependence assessment in human reliability analysis based on cloud model and best-worst method. Reliab Eng Syst Saf 2024;242:109770.
[39] JOHN P, BROOKS B, SCHRIEVER U. Profiling maritime communication by non-native speakers: A quantitative comparison between the baseline and standard marine communication phraseology. English for Specific Purposes 2017;47:1–14.
[40] JOHN P, NOBLE A, BJÖRKROTH P. Low-fi simulation of bridge team communication: A study of the authenticity of language patterns observed in 'chat'messaging to facilitate Maritime English training. WMU Journal of Maritime Affairs 2016;15:337–51.
[41] KAPTAN M, UĞURLU Ö, WANG J. The effect of nonconformities encountered in the use of technology on the occurrence of collision, contact and grounding accidents. Reliab Eng Syst Saf 2021;215:107886.
[42] KHAN RAFI U, YIN J, MUSTAFA FALUK S, WANG S. Seaport Hazardous Cargo Loading and Unloading Risk Assessment Using Interval Type-2 Fuzzy Sets and Bayesian Networks. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 2024;10:04023065.
[43] KHAN RU, YIN J, AHANI E, NAWAZ R, YANG M. Seaport infrastructure risk assessment for hazardous cargo operations using Bayesian networks. Mar Pollut Bull 2024;208:116966.
[44] KHAN RU, YIN J, MUSTAFA FS, AHMAD FAREA AO. A data centered human factor analysis approach for hazardous cargo accidents in a port environment. J Loss Prev Process Ind 2021:104711.
[45] KHAN RU, YIN J, MUSTAFA FS, SHI W. Factor assessment of hazardous cargo ship berthing accidents using an ordered logit regression model. Ocean Engineering 2023;284:115211.
[46] KHAN RU, YIN J, MUSTAFA FS, WANG S. Analyzing human factor involvement in sustainable hazardous cargo port operations. Ocean Engineering 2022;250: 111028.
[47] KONG D, LIN Z, LI W, HE W. Development of an improved Bayesian network method for maritime accident safety assessment based on multiscale scenario analysis theory. Reliab Eng Syst Saf 2024;251:110344.
[48] LAN H, MA X, MA L, QIAO W. Pattern investigation of total loss maritime accidents based on association rule mining. Reliab Eng Syst Saf 2023;229:108893.
[49] LEVESON NG. Engineering a safer world: Systems thinking applied to safety. The MIT Press; 2016.
[50] LI H, REN X, YANG Z. Data-driven Bayesian network for risk analysis of global maritime accidents. Reliab Eng Syst Saf 2023;230:108938.
[51] LI H, ZHOU K, ZHANG C, BASHIR M, YANG Z. Dynamic evolution of maritime accidents: Comparative analysis through data-driven Bayesian Networks. Ocean Engineering 2024;303:117736.
[52] LIU J, ZOU Y, WANG W, ZIO E, YUAN C, WANG T, JIANG J. A Bayesian belief network framework for nuclear power plant human reliability analysis accounting for dependencies among performance shaping factors. Reliab Eng Syst Saf 2022; 228:108766.
[53] LIU L, ULLAH KR, MUHAMMAD A, ASAD M. A corpus-based quantitative risk assessment of language barriers in maritime safety. Maritime Policy & Management 2025:1–35.
[54] LU CS, HSU CN, LEE CH. The Impact of Seafarers' Perceptions of National Culture and Leadership on Safety Attitude and Safety Behavior in Dry Bulk Shipping. International Journal of e-Navigation and Maritime Economy 2016;4:75–87.
[55] LU CS, LAI KH, LUN YHV, CHENG TCE. Effects of national culture on human failures in container shipping: The moderating role of Confucian dynamism. Accident Analysis & Prevention 2012;49:457–69.
[56] MA X, DENG W, QIAO W, LAN H. A methodology to quantify the risk propagation of hazardous events for ship grounding accidents based on directed CN. Reliab Eng Syst Saf 2022;221:108334.
[57] MARCOT BG, HANEA AM. What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? Comput Stat 2020.
[58] MOHAMMADFAM I, GHASEMI F, KALATPOUR O, MOGHIMBEIGI A. Constructing a Bayesian network model for improving safety behavior of employees at workplaces. Appl Ergon 2017;58:35–47.
[59] NÆVESTAD TO, STØRKERSEN KV, LAIOU A, YANNIS G, MICHELARAKI E. Potential safety outcomes of communication difficulties in mixed nationality crews: A study of Greek and Norwegian vessels. Transportation Research Procedia 2023;72:2904–11.
[60] OFORI-BOATENG R, ACEVES-MARTINS M, WIRATUNGA N, MORENO-GARCIA CF. Towards the automation of systematic reviews using natural language processing, machine learning, and deep learning: a comprehensive review. Artif Intell Rev 2024;57:200.
[61] REASON J. Managing the Risks of Organizational Accidents. Taylor & Francis; 2016.
[62] SAFETY4SEA. 2020a. Language differences onboard: A barrier to ship safety? [Online]. Safety4Sea. Available: https://safety4sea.com/cm-language-differences-onboard-a-barrier-to-ship-safety/#:~:text=Did%20you%20know%3F [Accessed 10th March 2025].
[63] SAFETY4SEA. 2020b. Language differences onboard: A barrier to ship safety? [Online]. Available: https://safety4sea.com/cm-language-differences-onboard-a-barrier-to-ship-safety/#:~:text=%2A%20Today%2C%2080,%E2%80%98Deadly%20Dozen%E2%80%99%20factors%20at%20sea [Accessed 25th June 2025].
[64] SAFETY4SEA. 2021. Tackling the challenges of cultural differences onboard [Online]. Available: https://safety4sea.com/cm-tackling-the-challenges-of-cultural-differences-onboard/#:~:text=What%20is%20more%2C%20it%20is,impact%20a%20ship%E2%80%99s%20operational%20performance [Accessed 26th June 2025].
[65] SCHRÖDER-HINRICHS JU, HOLLNAGEL E, BALDAUF M, HOFMANN S, KATARIA A. Maritime human factors and IMO policy. Maritime Policy & Management 2013;40:243–60.
[66] SHAPPELL SA, WIEGMANN DA. Applying reason: The human factors analysis and classification system (HFACS). Human Factors and Aerospace Safety 2001;1:59–86.

[67] THEOTOKAS I, PROGOULAKI M. Cultural diversity, manning strategies and management practices in Greek shipping. Maritime Policy & Management 2007; 34:383–403.
[68] WANG N, WU M, YUEN KF. Assessment of port resilience using Bayesian network: A study of strategies to enhance readiness and response capacities. Reliab Eng Syst Saf 2023;237:109394.
[69] WONG TT, YEH PY. Reliable Accuracy Estimates from k-Fold Cross Validation. IEEE Trans Knowl Data Eng 2020;32:1586–94.
[70] WU C, HUANG L. A new accident causation model based on information flow and its application in Tianjin Port fire and explosion accident. Reliab Eng Syst Saf 2019; 182:73–85.
[71] YANG Z, WAN C, YANG Z, YU Q. Using Bayesian network-based TOPSIS to aid dynamic port state control detention risk control decision. Reliab Eng Syst Saf 2021;213:107784.
[72] YIN J, KHAN RU, AFZAAL M, ALMALKI HM, SALEEM KHASAWNEH MA, SULAIE SA. Quantitative risk assessment of speech acts and lexical factors in maritime communication failures and accidents. Saf Sci 2025;191:106968.
[73] YIN J, KHAN RU, AFZAAL M, NAWAZ R, SHANSHAN X, JAMAL A. A fuzzy bayesian quantitative risk assessment for language and communication induced accidents in maritime operations. Ocean Coast Manag 2024;259:107449.
[74] YIN J, KHAN RU, WANG X, ASAD M. A data-centered multi-factor seaport disruption risk assessment using Bayesian networks. Ocean Engineering 2024;308: 118338.
[75] ZIO E. An Introduction to the Basics of Reliability and Risk Analysis. World Scientific; 2007.
[76] ZIO E. Reliability engineering: Old problems and new challenges. Reliab Eng Syst Saf 2009;94:125–41.