

Maritime accident severity modeling with operationally realistic multi-factor data-driven bayesian networks

Rafi Ullah Khan , Zaili Yang ^{*}

Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, L3 3AF, UK

ARTICLE INFO

Keywords:

Maritime safety
Accident severity
Data-driven BN, Compliance signal
Operational locale
Accident geography

ABSTRACT

Maritime accident severity analysis supports safety management and emergency preparedness, yet operational realism remains limited in many quantitative models. Common limitations include coarse geographic representation, reduced-granularity met-ocean descriptors, and the omission of Port State Control (PSC) compliance signals. These constraints can mask actionable thresholds and hinder the identification of interpretable, decision-relevant severity pathways. To address this, the study develops Bayesian Network (BN) models to forecast accident severity using a globally distributed, multi-year (2015–2024) dataset of 1050 reports, with granular descriptors of route/region, operating locale, weather, sea-state, visibility, ship profile, human error, and accident type. It assembles a harmonized severity dataset unifying investigation narratives, PSC outcomes, ship registries, and met-ocean validation within a single standardized schema; embeds two-layer geography, multi-state environmental conditions, and compliance signals into one BN severity framework; and benchmarks three BN families, Tree-Augmented Naïve (TAN), Augmented Naïve Bayes (ANB), and a literature/expert topology, on the same dataset. Beyond structure comparison, a unified inference workflow combines Strength-of-Influence, state-wise tornado decomposition, True Risk Influence (TRI), and sequential and simultaneous scenario analysis. All three models were estimated via Expectation–Maximization (EM) and evaluated using stratified 10-fold ROC/AUC. The literature/expert BN achieved the strongest cross-validated discrimination across severity states and was selected for analysis, indicating that structure should be selected empirically rather than assumed a priori. Diagnostic, scenario, and sensitivity analysis show that severity escalates when open-sea operations on dense traffic corridors coincide with adverse environment (storm, poor visibility, rough sea, high winds), human error, and loss-of-control typologies (grounding/capsizing); severity attenuates under benign conditions and contact-type events. Strength-of-Influence results highlight the central roles of route/region→traffic density and weather/visibility→severity pathways. A targeted TRI assessment isolates Accident Type as the dominant lever shifting probability mass into the high-severity state. The findings enable consequence-aware routing and vessel traffic prioritization, type-conditioned emergency preparedness, and human-factor/equipment readiness.

1. Introduction

Global shipping has scaled in size, speed, and autonomy, intensifying traffic density and encounter complexity in constrained waterways and major sea lanes [1]. The result is a risk landscape marked by low-probability, high-consequence events, where a single grounding, collision, fire, or foundering can trigger cascading human, environmental, and supply-chain losses [2]. Recent BN-based and data-driven studies confirm that accident risks are multifactorial and time-evolving across fleets, routes, and operating contexts, and that model performance improves when interdependencies among factors

are explicitly represented rather than assumed away [3,4].

Recent accident experience underscores both the scale of consequences and the evolving mix of causal factors confronting modern shipping: the containership *Dali*'s 26 March 2024 allision that collapsed Baltimore's Francis Scott Key Bridge killed six workers, severed a critical logistics artery, and triggered multi-hundred-million-to-billion-dollar recovery and reconstruction costs; preliminary findings point to repeated power losses minutes before impact that disabled steering pumps, an archetypal machinery/electrical failure pathway with catastrophic infrastructure externalities [5]. The 2023–2024 period also saw a run of high-profile ro-ro/vehicle carrier and container-ship fires,

^{*} Corresponding author.

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

<https://doi.org/10.1016/j.ress.2026.112823>

Received 27 January 2026; Received in revised form 29 April 2026; Accepted 2 May 2026

Available online 6 May 2026

0951-8320/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

including the *Fremantle Highway* blaze off the Netherlands (one fatality, mass abandonment) that reignited scrutiny of lithium-ion battery cargo risks and onboard fire-fighting adequacy; while the exact ignition source remains contested, authorities and insurers note that EV/Li-ion cargoes can amplify fire severity, intensify heat release, and complicate suppression, driving urgent calls for detection and response upgrades [6,7]. At the same time, chemical/hazardous cargo incidents continue to produce outsized, long-tail harms: the *X-Press Pearl* (2021) fire and sinking off Sri Lanka released a complex mixture of chemicals and plastic nurdles, producing extensive shoreline contamination and fisheries disruption; subsequent court actions and scientific assessments highlight environmental and liability exposures an order of magnitude larger than traditional hull/cargo losses and have already prompted insurers to warn of pricing/coverage shifts for jurisdictions and cargoes lacking clear liability caps [8]. These case studies mirror broader statistics: machinery damage/failure remains the single largest incident category globally, while fire/explosion events have surged to a decade-high in 2024, and collisions/allisions continue to impose disproportionate human and economic costs, patterns that converge with congested approaches, constrained waterways, and high-energy cargoes [9]. Regional monitoring corroborates the spatial concentration of consequences in traffic-dense sea areas and port approaches (e.g., North Sea/English Channel, major Asian hubs), where geometry, met-ocean conditions (fog, wind, storms, ice), and operational complexity amplify both accident likelihood and severity; European surveillance further documents persistent fatalities and serious injuries despite long-term safety gains [10,11]. Collectively, these recent events and aggregates emphasize multi-factor causation, notably machinery/-power integrity, cargo fire hazards (incl. Li-ion), traffic/geometry constraints, and adverse weather/visibility, and reveal why severity-focused, geographically explicit, and weather-aware probabilistic modelling is needed to anticipate not only *if* an accident occurs, but *how bad* it is likely to be, and under which combinations of ship profile, operating locale, and compliance posture [9,11].

Why severity forecasting matters is because most operational decisions (routing, watchkeeping, resource pre-positioning, tug allocation, emergency readiness) are governed not only by the probability of an accident but by its likely severity. Data-driven BN studies show that severity is unevenly distributed across accident types (e.g., capsizing/foundering vs. minor contacts) and ship profiles (e.g., small fishing vessels vs. large commercial ships), implying that prevention and preparedness should be targeted rather than uniform [12]. Yet much of today's practice and a sizeable portion of the literature remain oriented to *occurrence* (accident/no accident) or to *binary detention* in compliance studies, leaving a gap for fine-grained severity modelling that incorporates weather, locale, and regulatory context in ways stakeholders can act on [12,13].

Over the past several years, several studies have advanced the state of maritime accident analysis using data-driven BNs and related ML approaches. In this regard, large-scale, multi-year accident databases have been assembled and modelled with BNs to identify key risk-influencing factors (RIFs) and their temporal evolution, showing performance gains and new managerial insights when factor interdependencies are learned from data [2,1]. From the perspective of Dynamic pattern change, comparative BN analyses across time windows reveal that the *salience* of RIFs shifts (e.g., port-area incidents rising with fleet and port complexity), underscoring the need for models that adapt to evolving traffic and operations [4]. In the domain of spatial evidence, independent spatial studies have mapped accident hotspots (e.g., North Sea/English Channel, Singapore/Shanghai approaches), attributing concentrations to traffic intensity, constrained geometry, and local met-ocean conditions; yet these insights are rarely embedded as explicit regional nodes in probabilistic causation models [14]. Similarly, in severity modelling, BN studies of severity identify accident type, tonnage, ship type, and (broad) location among the top drivers of "very serious" consequences, while recommending richer contextualization in

future models [12].

At the same time, important limitations persist: (i) Geographic information is often coarse, many models do not jointly encode named sea regions (e.g., North Sea, Baltic, South China Sea) and operating locale (port/coastal/open sea); (ii) weather granularity is limited (storms, ice, visibility, wind, sea state frequently enter as binary or proxy variables); (iii) regulatory/compliance signals, notably Port State Control (PSC) outcomes, are typically analyzed in separate detention-risk studies rather than integrated into accident severity models; (iv) many factor sets rely on binary states, which can obscure actionable thresholds; and (v) few studies compare a data-learned BN against an expert-crafted BN on the same dataset to quantify the trade-off between predictive performance and interpretability [12–14].

Against this state of the art, we identify a methodological gap that goes beyond factor coverage. Existing severity BNs typically (i) commit *ex ante* to a single BN family, either a data-learned TAN/ANB or an expert-elicited topology, without head-to-head, cross-validated benchmarking on the same dataset; (ii) rely predominantly on Strength-of-Influence or marginal sensitivity for inference, with limited use of more rigorous tools such as TRI [15] and only sequential evidence-setting in scenario analysis; and (iii) draw on single- or dual-source databases without a reproducible cross-jurisdictional fusion schema. The recent studies that come closest to a structural comparison [4] report that a data-learned TAN outperforms a layered expert structure for severity prediction; whether this ordering generalizes to other datasets and richer factor sets has not been independently tested.

This study addresses these gaps by making four explicit advances over prior severity-BN work, each demonstrable on the same 1050-record, 2015–2024, globally distributed dataset:

- (1) Head-to-head, cross-validated structural benchmark with empirical model selection. Three BN families, TAN, ANB, and a literature/expert topology, are estimated by EM on the same harmonized 22-variable dataset and evaluated by stratified 10-fold one-vs-rest ROC/AUC across all three severity states. Empirically, the literature/expert topology achieves the strongest discrimination, reversing the ordering reported by [4] on a different dataset and factor set. This is itself a methodological contribution to the field: it shows that the relative performance of data-learned versus expert BN structures is dataset- and factor-set-dependent, and that performance-led, structure-agnostic selection, rather than a default preference for either family, is the appropriate methodological posture for maritime severity modelling.
- (2) Unified, multi-tool inference workflow. On the selected model, we apply, in a single integrated workflow, predictive and diagnostic queries, sequential and simultaneous scenario analysis, full-graph Strength-of-Influence, state-wise tornado decompositions, and a targeted TRI analysis on the dominant severity lever. The combination of TRI with simultaneous evidence-setting and tornado-level state resolution has not, to our knowledge, been deployed jointly in any prior severity-BN study; existing work typically uses a strict subset of these tools. The richer inference apparatus is what allows the model to identify operational decision levers (route, locale, accident-type prevention, and weather-aware dispatch) at a granularity that prior severity BNs do not attempt.
- (3) A reproducible cross-jurisdictional data-fusion schema. The dataset is harmonised across eight independent source families, IMO GISIS, EMSA EMCIP, four national investigation bodies (UK MAIB, US NTSB, Canada TSB, Australia ATSB), four PSC regimes (Paris MoU, Tokyo MoU, USCG, AMSA), commercial fleet registers (S&P Global Sea-web, Clarksons World Fleet Register), met-ocean reanalysis (ECMWF ERA5, Copernicus Marine, NSIDC), and AIS corroboration (MarineTraffic, Spire), into a single 22-variable schema with conflict-resolution rules for divergent

fields and explicit reanalysis-based validation of the met-ocean variables. Section 3.1 details the schema. This level of source integration is a contribution beyond the single- or dual-source databases used in the closest prior work [3,12,16,13].

- (4) These three methodological advances are complementary to, and operationalize, the factor-coverage extensions already noted in the literature: explicit two-layer geography (sea region $\times$ operating locale), multi-state met-ocean variables (storm, ice, visibility, wind, sea state), separation of human error from reported cause, and integration of PSC inspection outcomes as a severity predictor. Table 3 quantifies the resulting coverage advance over 17 benchmark severity-BN studies. Together, these advances move the field from frequency- and binary-detection framings toward a unified, structure-benchmarked, multi-tool probabilistic framework for decision-grade severity inference, moving beyond frequency-only or binary-outcome framings toward decision-quality probabilistic insight [3,2,4,1]. The rest of the paper structure includes Literature review in Section 2, Methodology in Section 3, Results and Discussions in Section 4, Implications are presented in Section 5, while Section 6 concludes the findings of the study.

2. Literature review

2.1. Data-Driven bayesian network models in maritime accident analysis

Over the last decade, Bayesian networks have become a widely used technique for maritime risk analysis [17–19,16], providing a data-driven means to unravel the complex interplay of factors leading to accidents [20–22]. Early applications of BN models in this domain demonstrated their utility on relatively limited datasets. For example, a recent study applied a Naïve Bayesian Network to quantify the influence of different factors on various types of maritime accidents using 161 accident investigation reports [23]. Similarly, another study investigated accident causes in Hong Kong waters with a dataset of 331 incidents [24]. These studies, though modest in data size, validated that BNs can effectively capture causal factor relationships and provide probabilistic accident predictions. Wang and Yang [13] further showed the suitability of BN modeling for accident severity analysis, using a case study of shipping accidents in China to predict injury/fatality outcomes and demonstrating reasonable accuracy (ROC/AUC) in their BN model [13]. Collectively, such works established a foundation that data-driven BNs can support maritime safety assessments by combining historical data and expert knowledge in a coherent framework.

Building on these foundations, more recent studies have scaled up the data-driven BN approach to broader and larger accident datasets [25]. In an effort to gain a global perspective, Zhou et al. [1] compiled an updated worldwide casualty dataset and introduced new risk factors into a pure data-driven BN model to analyze maritime accident casualties [1]. Their study uniquely segmented the dataset into five temporal subsets, constructing separate BN models for different time periods to examine how the influence of risk factors on accident outcomes has evolved over time [1]. This dynamic analysis revealed that the roles of certain factors in contributing to casualties have changed, reflecting shifts in industry practice or technology over the years. In a related vein, Li et al. [4] investigated the dynamic evolution of maritime accidents from 2012–2021 using multiple BN structures [4]. They compared a layered BN structure versus a TAN network for modeling accident data in two half-decade phases and found notable changes in risk patterns over the past decade [4]. Notably, their findings indicated a significant rise in accidents involving large ships and in port areas in the later years, suggesting that as the global fleet grows in size, certain risks (like port-area incidents during anchoring or maneuvering) have increased [4]. They also reported that the TAN-based BN achieved better predictive performance for accident severity than the expert-defined layered model, highlighting the value of data-driven structure learning [4].

These studies underscore an emerging trend of using multiple BNs or time-evolving models to capture the changing landscape of maritime risk, an approach that adds a temporal dimension to what was traditionally a static analysis.

Concurrently, the scale of maritime accident data being analyzed with BNs has expanded dramatically in recent literature [26]. While the majority of earlier BN studies were limited to a few hundred accident cases (owing to data availability issues) [27–30,22], some recent works have assembled far larger databases. For instance, Cao et al. [12] utilized 1294 accident reports from 2000–2019 to build a TAN-based BN for marine accident severity analysis, which is among the largest datasets seen in this context [12]. They leveraged this comprehensive dataset to quantify how different Accident Influential Factors (AIFs) affect the severity outcome of incidents. Another remarkable effort by Jiang et al. [16] went even further, incorporating 55,469 maritime accidents (2002–2022) from global records to construct a BN model of risk factors [16]. This study, using a massive global accident database, identified that ship and voyage characteristics were dominant drivers of accident probability, for example, older and smaller vessels operating under flags of convenience in certain regions had markedly higher accident propensity [16]. The key factors highlighted by that analysis included ship age, size (tonnage), type, and location of operation [16]. The fact that “location” emerged as a top risk influence in a global model suggests that geographical context plays a significant role in maritime accident risk, an aspect sometimes overlooked in previous models [31,14]. Overall, the trajectory of recent research shows BNs being applied to ever more extensive and diverse datasets, moving from regional and small-scale analyses towards global, data-rich models that can capture broader risk trends [32,33,2,21,34–36].

2.2. Geographical and environmental factors

One clear knowledge gap in earlier maritime risk models is the integration of detailed geographic and environmental factors. Many BN-based studies treated the operating environment in a coarse way if at all, for instance, by indicating if an accident occurred in *open sea* vs *coastal waters*, or if weather was a factor (yes/no). Regional differences were often neglected: analyses either focused on a single region or treated the globe as homogeneous, without encoding the location of each incident as a variable in the model [12,32,33,2,4,13,1]. This is important, because accident rates and causes can vary significantly by region due to differences in traffic density, weather, and navigational challenges. A separate line of research using geospatial analysis has indeed shown that maritime accidents exhibit strong spatial patterns. In this regard, a recent study mapped 15 years of global accident data and found clear hotspots: accidents concentrate in busy and confined waterways such as around the UK and Denmark (North Sea/English Channel region), as well as near major hub ports like Singapore and Shanghai [14]. The high accident frequencies in these areas were attributed to intensive vessel traffic, challenging geographical features (e.g. narrow straits), and often harsh local weather or sea conditions [14]. Likewise, Choe et al. [37] applied spatial clustering techniques to maritime incident data and identified regional clusters of high accident probability. These geographic analyses highlight that location-specific factors, for example, the congested sea lanes of East Asia vs. the icy waters of the Baltic, influence both the likelihood and nature of accidents [37]. However, the insights from such GIS-based studies have not been fully incorporated into BN risk models. Most probabilistic models thus far have not explicitly modeled the *region* of occurrence as a risk node, with a few recent exceptions. For example, the global BN by Jiang et al. noted above included broad ocean zones and found the “North Atlantic” zone to correlate with higher accident probability (likely reflecting Atlantic storm conditions and trade volume) [16]. And in the context of high-seas accidents (far from coastal support), Qu et al. [38] demonstrated that the sea area of an accident is a significant factor affecting risk levels and accident types [38]. In their multi-source BN study of open-ocean

incidents, “sea area” (geographical region) notably impacted the accident outcome distribution, affirming that incorporating such spatial factors is both appropriate and necessary for risk analysis [38].

Beyond broad regions, the *immediate locale* of an accident, whether it happens in port waters, coastal approaches, or the open sea, also influences its characteristics and severity. Different water areas entail different operational conditions (port maneuvers vs. open-ocean transits) and emergency response capabilities, which can affect accident consequences. Some studies have included this aspect to a degree. For instance, Cao *et al.* [12] included a “location” factor in their BN model for severity, distinguishing the context of the accident; they found that location (presumably indicating whether the incident was in port, coastal, or offshore waters) was among the top five factors influencing accident severity levels [12]. Similarly, Li *et al.* [4] reported a recent increase in accidents in port areas, particularly during anchoring operations, correlating with certain accident types like collisions and groundings in those settings [4]. However, to the best of our knowledge, no prior study has simultaneously examined *both* the macro-region (e.g., specific sea or geographic region) *and* the local water type (port vs coastal vs open sea) within one unified accident causation model. In the literature, one typically finds either a spatial analysis at a coarse level or a binary distinction of waters, but not a combined, granular geographic representation. Our study addresses this gap by incorporating each accident’s region (e.g., North Sea, Baltic, South China Sea, etc.) alongside the location type (port/coastal/offshore) as distinct variables. By doing so, we can explore, for example, how a storm in the North Sea might impact accident outcomes differently than a storm in the South Pacific, or how accidents in busy harbor waters differ in severity from those on the high seas. Incorporating these geographic dimensions is expected to enrich the BN model and provide more context-sensitive insights that were absent from earlier models.

Environmental conditions are another critical facet. Adverse weather and sea state, fog, high winds, storms, heavy seas, ice conditions, etc., have long been known to precipitate accidents or worsen their outcomes [39–42]. Yet, many data-driven models either omit specific weather variables or simplify them drastically. Some recent research has started to account for these factors. For example, Qu *et al.* [38] note that about 20.6 % of high-seas accidents in their dataset occurred under strong wind conditions [38], indicating a non-negligible contribution of bad weather to incident frequency. In one of the BN scenario analyses by Li *et al.* [4], the authors qualitatively acknowledge that rapidly changing weather (fog, storms, ice, currents) can escalate risks and even cite how such factors can lead to loss of control or capsizing if not properly managed [4]. Despite this recognition, weather variables in many BN studies are represented only by proxies (e.g., season of year or general climate of region) or binary indicators (e.g., “weather was a contributing factor: yes/no”). There is a lack of granular inclusion of specific conditions, like icy waters or the presence of storms at the time of the accident. This is partly due to data limitations, detailed weather info is not always recorded in accident databases, but it represents a gap in capturing the full risk picture. A distinct and highly focused strand of literature has developed around the Baltic Sea winter navigation regime. In this regard, a study analyses four winter periods of Finnish accident data and shows that ship-independent navigation in ice is the operation with the highest accident rate [43]. Also, reporting that collisions are the most frequent winter navigation accident type, and that general cargo vessels are the most commonly involved ship class, accident severity in that dataset is mainly reported as ‘less serious’ under the EMSA classification. Similarly, another study extends this work into a Formal Safety Assessment-based BN model for winter navigation operations in the Gulf of Finland, identifying ship-independent navigation and convoy operations as the regimes with the highest probability of oil-spill outcomes [44]. The Aalto group has further assembled a multi-year AIS + ice-environment dataset dedicated to the Finnish–Swedish Winter Navigation System (FSWNS), supporting fine-grained analysis of ship performance in ice [45–47]. These studies provide an essential

Baltic-specific counterpart to the global severity modelling pursued here: they show that the accident population within the FSWNS is dominated by smaller general-cargo and feeder vessels (typically 1000–10,000 DWT) involved in ice-navigation collisions of mostly ‘less serious’ severity, and they establish the ice-, icebreaker- and convoy-specific risk mechanisms that shape Baltic winter consequences. Our global BN does not attempt to resolve this sub-population at the same granularity; rather, it incorporates the Baltic as one of six named sea areas and uses ice condition, season, and snow-weather evidence to trigger Baltic-consistent pathways in the joint distribution. Also, our study attempts to fill this gap by including explicit weather-related variables such as ice conditions and storm presence, along with other meteorological factors, as part of the risk factor set. By increasing the granularity of environmental data, the model can differentiate, for instance, between accidents in calm vs. rough seas or temperate vs. icy regions, thereby providing a more nuanced understanding of how weather extremes correlate with accident severity.

2.3. Human factors and port state control considerations

Human and organizational factors remain a leading cause of maritime accidents, often cited in both academic studies and industry reports (historically, human error is implicated in 75–90 % of incidents) [48,28, 49–51]. Many BN models include human elements, but typically in broad strokes. For example, “human error” might be one node in a network, or cause codes from reports (like “navigation error” or “maintenance failure”) are included among risk factors. Jiang *et al.* [33], in their BN analysis of Maritime Silk Road accidents, incorporated various causation factors which likely covered human errors (e.g., crew unsafe acts or rule violations) as part of the model input [33]. Likewise, Zhou *et al.* [1] enriched their global casualty analysis by adding new risk factors, which presumably included more detailed operational context and human-related factors, to improve prediction accuracy [1]. However, one limitation observed is that studies often combine or generalize human factors, rather than examining them in detail. In our approach, we distinguish the role of human error as a separate variable apart from the primary reported cause of the accident. This distinction allows us to analyze scenarios such as: an accident’s official cause might be “equipment failure”, but was there an underlying human error (e.g., in maintenance) also present? By separating these, we can assess the direct impact of human error on accident severity, controlling for other causes, an angle not explicitly explored in prior BN studies, which tended to either lump human factor as one category of cause or focus purely on technical factors. Our consideration of the crew size variable is another effort to capture and enhance the granularity of human factors in maritime safety.

Another novel aspect of our work is the integration of PSC inspection data into accident analysis. PSC regimes worldwide inspect ships for safety deficiencies, detain substandard vessels, and thus act as a front-line defense against accidents [52–58]. Intuitively, a vessel’s PSC inspection history (whether it has outstanding deficiencies or past detentions) could influence its likelihood of being involved in an accident or the severity of outcomes if an incident occurs (a poorly maintained ship might be less resilient in an accident). Despite this, the relationship between PSC records and accident severity has not been rigorously examined in the literature. Most studies dealing with PSC analyze it from a compliance perspective, for instance, identifying factors that lead to ship detentions or the patterns of deficiencies in certain regions [59]. Some empirical analyses have linked PSC *inspection performance* to accident risk on an aggregate level, suggesting that targeting vessels with poor inspection records could improve safety [60,61]. However, none of the current BN models of maritime accidents incorporate a vessel’s PSC status as a node in the causal model. Our study will be among the first to examine how PSC outcomes (categorized in our data as “No issues”, “With deficiencies”, or “Detained”) relate to accident occurrence and severity. By including the PSC inspection outcome

as a risk factor, we can investigate, for example, whether ships that had prior deficiencies or detentions are statistically more prone to severe accidents. This addition addresses a clear gap in prior research, which has so far treated accidents and PSC detentions mostly in isolation (with the exception of a few recent dynamic models focusing on how inspection regimes affect risk over time). We anticipate that integrating PSC data will shed light on the effectiveness of port state controls in mitigating serious accidents, a topic of high interest for regulators but not yet reflected in the modeling literature.

2.4. Accident severity modeling and remaining gaps

A number of studies have zeroed in on accident severity as the outcome of interest (as opposed to just accident frequency or type) [3, 62]. We have mentioned the pioneering work of Wang & Yang [13] in China and the large-scale analysis globally [12]. The latter, in particular, offers a benchmark for our research: Cao and colleagues established a BN to predict accident severity grades and identified “accident type,” “engine power,” “gross tonnage,” “ship type,” and “location” as the five most influential factors determining whether an accident becomes very serious [12]. They also observed that certain accident types like capsizing, sinking, or major collisions are most likely to escalate to the worst severity level, which aligns with intuition and historical incident analyses [12]. These findings are valuable and our work will build upon them. Notably, the “location” factor in their model, and then in a more recent study limited to regional waters [3], hints at the importance of where and in what context the accident occurs, reinforcing our earlier point about geography and local waters. Nonetheless, even that study did not differentiate *which* location or region (treating location somewhat generally), nor did it account for variables such as weather or PSC status. In fact, reviewing the recent literature reveals that it is hard to locate a study that has combined all the diverse elements we propose to examine. Each existing work covers part of the puzzle: some focus on technical ship factors (age, tonnage, power, etc.) [3,12,16], others on operational context like loading, voyage segment or port state [63,1], others on human/organizational causes, and a few on spatial distribution [14], but not all together. Moreover, many BN models simplify the problem by using mostly binary states for factors (e.g., “adverse weather: yes/no” or “crew error involved: yes/no”). While this yields a tractable model, it can oversimplify reality. Our dataset and model will introduce finer categorical states (for example, multiple levels of weather severity, several regions, ranges of crew size, etc.), increasing the granularity of analysis. By doing so, we expect to capture more nuanced relationships; for instance, the difference between a moderate storm and a violent storm might have different impacts on accident outcomes, which a binary variable would miss.

Another point of differentiation in our study is the methodological approach to model development and validation. Past studies typically develop one BN (or at most a couple of variants) and evaluate its performance [12,2,4]. In contrast, we plan to construct multiple BN models derived through different structure-learning algorithms and also include a model based on our previous *expert/literature-designed* BN structure for the same data. This approach will allow direct comparison between data-driven and expert-driven models in terms of predictive accuracy (e.g., using ROC/AUC as done previously [29,13]). The literature so far offers little on such comparisons, each study typically advocates its chosen model, but a systematic comparison of algorithms and a manually crafted network on the same problem is largely absent. We see this as an opportunity to evaluate the benefits of purely data-driven discovery of accident causation patterns versus relying on established expert knowledge structures. If our manual model shows higher prediction accuracy and richer sensitivity results, that contrast itself is informative for researchers deciding on modeling techniques for risk analysis.

Finally, when it comes to *utilizing* the BN models for insight, previous works have demonstrated various analysis techniques. Sensitivity

analysis is commonly used to rank the influence of different variables on the outcome of interest [2,4]. Scenario analysis (also known as evidence inference or “what-if” analysis) is also frequently performed: for example, Li et al. [2] and others have tested scenarios by setting certain evidence (e.g. a particular ship type or weather condition) to see how the probability of an accident or its severity changes [2]. However, these scenario analyses in literature usually consider one piece of evidence or one scenario at a time (or compare a few discrete scenarios). A novel angle we propose is to explore multi-factor scenarios, where evidence on several critical factors is set simultaneously to examine their combined effect on accident severity. This kind of *combined scenario analysis* can reveal interaction effects (for instance, how the conjunction of a storm, an inexperienced crew, and a small old ship might create a “perfect storm” for a severe accident). Such multi-factor stress tests are seldom reported in prior studies, which tend to vary one factor at a time. By conducting both sequential and simultaneous evidence setting in our analysis, we aim to provide a more holistic assessment of risk under complex real-world situations. This will complement the usual predictive (forward reasoning) and diagnostic (backward reasoning) analyses that BNs offer, as well as standard sensitivity and influence strength evaluations.

In summary, the existing literature on maritime accident risk analysis demonstrates substantial progress in applying BNs and other data-driven methods to understand accident causation and severity. Recent studies have grown in scope, from local case studies to global datasets, and have begun to incorporate dynamic and spatial perspectives. Key risk factors identified include ship-specific attributes (age, size, type), operational conditions (voyage segment, location type), and to a lesser extent environmental conditions and human factors [12,23,2,1]. Yet, there remains a clear gap in *comprehensiveness*. Very few prior studies have concurrently examined the full spectrum of factors, ship profile, human/organizational elements, environment/weather, and geography (both region and local waters), in a unified model for accident severity analysis. Most studies isolate subsets of these aspects, potentially missing their combined effects. Furthermore, innovative variables like PSC inspection results and crew composition, which could have significant safety implications, have been overlooked in data-driven models of maritime accidents. By addressing these gaps, our proposed research will add novel insights to the literature. It will offer a more complete picture of how and why severe maritime accidents occur across different regions and conditions, providing both a global overview and granular detail. This contribution is expected to be fruitful for the academic community and practitioners alike: academically, it extends the BN applications in maritime safety to incorporate new dimensions (regulatory and regional factors), and practically, it can inform more targeted risk mitigation strategies (e.g., highlighting if certain regions or vessel profiles under certain weather should receive more attention). In the following sections of the paper, we will detail our methodology and demonstrate how incorporating these diverse factors indeed improves the understanding and prediction of maritime accident severity.

3. Methodology

The study develops probabilistic severity models using BNs under three alternative structures: (i) a TAN BN learned from data, (ii) an ANB BN, and (iii) a literature-/expert-specified BN topology. For each structure, conditional probability tables are estimated via EM. Out-of-sample performance is assessed with stratified 10-fold cross-validation using one-vs-rest Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) for the three severity states. The highest-performing model is then used for inference: predictive queries, sequential and simultaneous scenario analysis, global and state-wise sensitivity (including tornado charts), strength-of-influence along directed edges, and a targeted TRI analysis for Accident Type. Together, these steps quantify both discriminative ability and causal leverage under operationally realistic, multi-factor conditions. Methodologically,

this workflow is among the first to systematically train and evaluate multiple mainstream BN formulations on the same severity dataset and then select the best-performing model as the basis for downstream inference. This comparative, performance-led model-selection strategy offers a replicable template for researchers to choose best-fit BN structures for consequence analysis under similar data constraints. The adopted method flowchart is depicted in Fig. 1.

Accident case data, PSC records, ship particulars, met-ocean information, and traffic-context sources used for data compilation, validation, and standardization are summarized in Table A1, in Appendix A. In practice, most accident reports already contained the operational, environmental, and consequence details required for the analysis; external sources were used selectively to validate or standardize fields such as route/region, weather bands, ship particulars, PSC history, and corridor-level traffic context across jurisdictions.

3.1. Bayesian network framework

Maritime accident severity datasets present a methodologically demanding combination of features that conventional statistical and discriminative machine-learning approaches handle only partially. The data are mid-sized (typically of the order of 10^2 – 10^3 records), heterogeneous (mixing categorical, ordinal and continuous fields drawn from investigation narratives, fleet registers, PSC regimes and met-ocean reanalysis), and characteristically incomplete: weather and ice fields are not always recorded in narrative reports, PSC histories are jurisdiction-dependent, and the most severe outcomes are sparse in the right tail of the consequence distribution [64,65]. Conventional approaches such as logistic and multinomial regression, ordered probit, and decision-tree or gradient-boosted ensembles can be fit to such data but require either complete-case deletion or external imputation of missing variables before estimation, both of which discard information

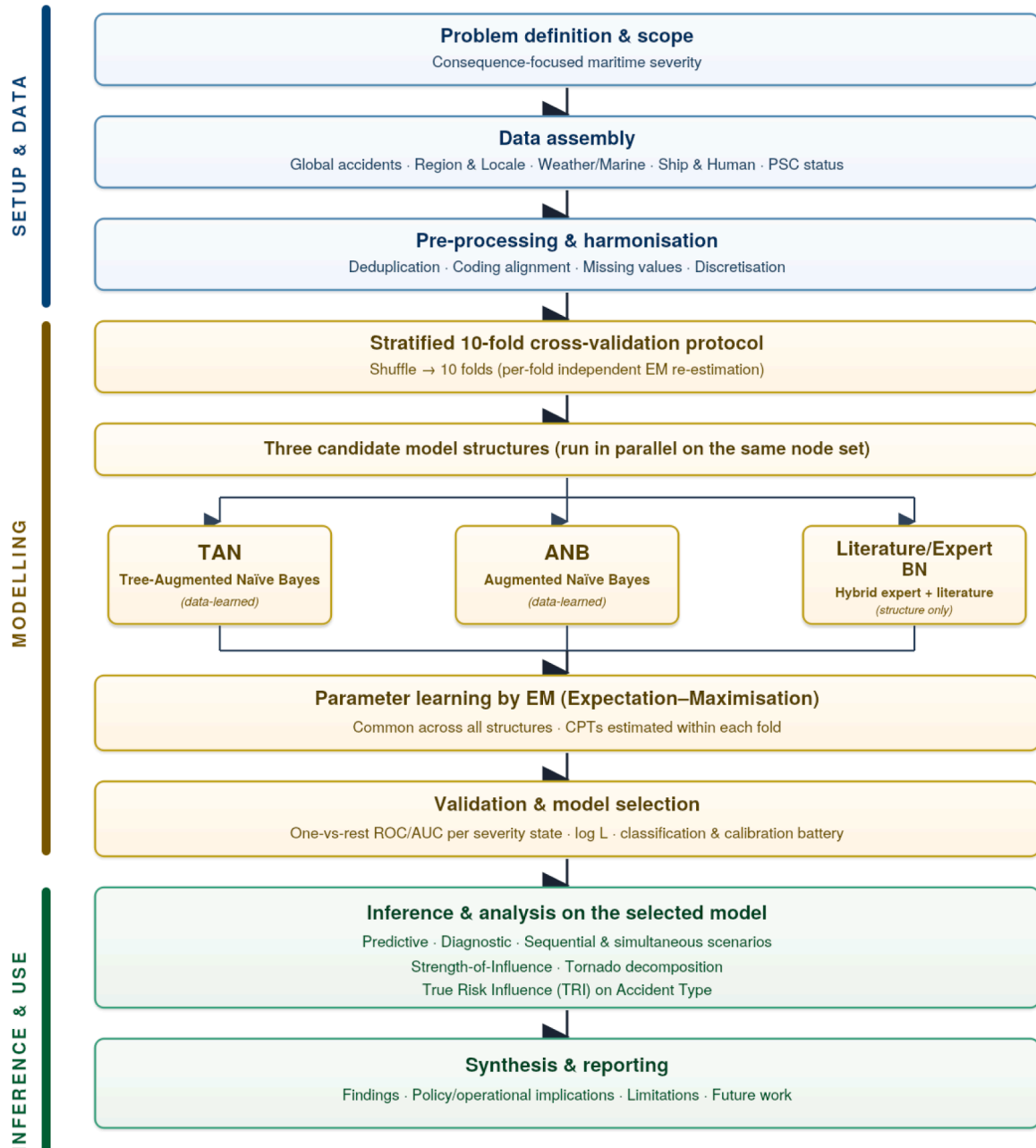


Fig. 1. A schematic flowchart of the adopted methodology.

and can bias the posterior of the rare High-severity class [65,66]. Modern neural-network classifiers add a further difficulty: they require sample sizes substantially larger than those typically available from official investigation repositories to avoid overfitting on partially correlated factors, and they yield point predictions or calibration-poor scores rather than auditable probability distributions over consequence states [64,65]. These limitations are well documented in recent maritime-safety reviews and were a primary reason why the data-driven BN family has emerged as the dominant tool for this class of problem [3, 12,16,13].

BNs address these limitations through four properties that are directly relevant to maritime severity inference. First, BNs handle missing or incomplete evidence natively: the inference engine marginalises unobserved nodes analytically through the network's conditional-independence structure, so the model produces a calibrated posterior over Severity Level even when several met-ocean, ship-profile, or compliance variables are absent for a given record [67,65,68]. The Expectation–Maximization algorithm used to estimate the conditional probability tables (Section 3.5) extends this property to the parameter-learning stage, allowing partial accident records to contribute information to CPT estimation rather than be discarded [65]. Second, BNs support both predictive (forward) and diagnostic (backward) inference from the same trained model, together with scenario reasoning under partial or simultaneous evidence; logistic/multinomial regression and modern classifier families are inherently unidirectional input→output mappings and cannot answer diagnostic queries such as "given a severe accident, which combinations of upstream conditions are most probable?" [64,12]. Third, the CPTs are interpretable and auditable, exposing the conditional probabilistic mechanism by which each parent influences each child node; this is essential for safety-critical decision support, where regulators and investigators must trace the reasoning behind a posterior [67,68,66]. Fourth, BNs allow data and domain knowledge to be combined explicitly: structure and CPT priors can encode expert knowledge or literature evidence in regions of the joint distribution that the dataset samples sparsely (e.g., the High-severity tail), reducing variance under intermediate sample sizes and improving stability of inference [67,65,68]. For these reasons, and consistent with the recent severity-BN literature converging on this methodology [69,19,70,20,34], BN is the methodologically appropriate framework for the multi-factor, partially-observed, decision-relevant maritime severity problem addressed in this study.

This study models maritime accident severity using BNs. A BN is a directed acyclic graph in which each node represents a random variable (e.g., ship type, weather, accident type, PSC status), and each directed edge encodes a conditional dependence between variables [71,72,29, 30]. Let $X = \{X_1, X_2, \dots, X_N\}$ denote the full set of variables in the network, including the target variable Accident_Severity. The joint probability distribution over all variables factorizes according to the BN structure as [12,69,73]

$$P(X) = \prod_{i=1}^N P(X_i | P_a(X_i)) \quad (1)$$

Where $P_a(X_i)$ is the set of parent nodes (direct causes or direct informational influencers) of node X_i . This factorization allows the BN to represent high-dimensional, multi-factor maritime risk using local conditional probability tables (CPTs) instead of one intractable full joint distribution.

In operational terms, this means the BN can:

- Predict forward (e.g., estimate the probability that an accident, under specific weather and ship conditions, will escalate to a severe outcome);
- Diagnose backward (e.g., given a severe accident, infer which combinations of factors, such as machinery failure under storm conditions in a confined approach, are most probable); and

- Support scenario reasoning, both sequential (one variable fixed at a time) and simultaneous (multiple conditions imposed together), even when some variables are missing.

All variables are discretized into interpretable, domain-relevant states before modelling (e.g., visibility bands, regional sea areas, storm presence, accident type, PSC status, "port waters / coastal waters / open sea", etc.). Accident_Severity is treated as an ordinal, multi-state node representing consequence level (Low / Medium / High).

The methodology proceeds in four main stages: (i) structure learning; (ii) parameter learning; (iii) validation and model selection; and (iv) inference analyses. We develop three BN variants, two data-driven networks and one literature and expert-structured network, and then select the most credible and best-performing model for downstream analysis.

3.2. Structure learning

(a) Tree-Augmented Naïve Bayes (TAN)

The first model is learned using a TAN procedure. TAN is a supervised BN structure learning approach that treats the target variable Accident_Severity as the class node and then allows each predictor to depend on Accident_Severity and, at most, one other predictor. Concretely:

1. A directed arc is created from Accident_Severity to every predictor $X_j | X_i$.
2. Pairwise dependencies among predictors are captured by adding a maximum-weight spanning tree connecting predictors, where edge weights reflect conditional mutual information between predictors given Accident_Severity.
3. The result is a sparse, directed acyclic structure in which every non-target variable has Accident_Severity as a parent and may additionally have one predictor parent.

The advantage of TAN in this context is that it balances interpretability and realism. Pure naïve Bayes would assume that all predictors are conditionally independent given Accident_Severity, which is not realistic in maritime safety (for example, "storm presence" and "reduced visibility" are often linked, as are "ship age" and "machinery failure"). TAN relaxes that assumption by explicitly learning the strongest conditional dependencies between predictors, while keeping the network sufficiently simple to interpret and interrogate through scenario analysis.

In our application, TAN is well suited for capturing coupled mechanisms such as:

- environmental exposure (storm, poor visibility, heavy sea state) interacting with operating locale (port waters vs. open sea),
- machinery/electrical failure co-occurring with loss of propulsion/steering,
- PSC compliance status co-occurring with reported human error or procedural failure, and relating these combinations to severity outcomes.

(b) Augmented Naïve Bayes (ANB)

The second data-driven model is an ANB network. ANB generalizes naïve Bayes by allowing limited additional arcs among predictors in predefined substructures (for example, grouping variables relating to weather and navigation; grouping crew-related and procedural/human-error variables). Compared to TAN, which builds a global tree over all predictors, ANB imposes a looser augmentation pattern: Accident_Severity is still set as the parent ("class parent") of all predictor nodes, but additional arcs are introduced among logically or operationally related predictors to capture known dependencies (e.g., "storm presence" → "visibility band", "PSC status" → "reported cause severity," etc.).

This produces a BN that is still compact and classification-oriented, but with explicit sub-clusters of interaction. ANB forms a useful comparator to TAN because both are data-driven, but they capture predictor dependencies differently: TAN via a single spanning tree optimized on conditional mutual information, ANB via clustered augmentation reflecting observed and operationally plausible linkages.

(c) Expert / Literature BN

The third network is built from domain knowledge and prior literature (discussed in introduction and literature review) rather than from automated structure learning. In this expert/literature BN:

- Nodes correspond to the same high-level factors as in the data-driven models (e.g., region, local water type, accident type, weather states, ship profile, PSC outcome, human error, etc.).
- Directed edges are assigned according to plausible causal ordering described in accident investigation practice and maritime safety research. For example, “storm presence” → “loss of manoeuvrability / collision-type accident,” “PSC status” → “baseline safety posture,” “operating region / local water type” → “traffic density / encounter complexity” → “accident severity.”
- Accident Severity is positioned as a consequence node receiving incoming arcs from immediate accident conditions (accident type, loss-of-propulsion, fire/explosion, etc.), exposure amplifiers (sea state, visibility, ice, confined geometry), and compliance/management posture.

This expert BN represents how practitioners and investigators reason causally about escalation: mechanical failure under adverse met-ocean conditions in a constrained traffic lane is more likely to become severe than the same failure in calm open sea, and poor compliance posture (e.g., detained PSC history) may correlate with degraded resilience. The purpose of this model is twofold: (i) to serve as an interpretable baseline consistent with current safety thinking; and (ii) to allow direct comparison between human-structured causality and data-learned causality on the *same* dataset.

The expert/literature topology was constructed through a hybrid, five-step procedure that combines literature-derived evidence with structured expert review, broadly consistent with established maritime-BN structure-elicitation practice [74,75,67,65,68,76,77]. (1) Variable identification: a systematic scan of the recent maritime accident-severity literature (Section 2) and of the investigation narratives in the harmonized dataset was used to identify the candidate factor set, and was finalized by the experts. The same factor set is shared with the TAN and ANB networks, so that downstream performance differences are attributable to structure rather than to inputs. (2) Candidate-arc specification from the literature and experts: for each pair of variables, the published evidence base was reviewed for documented causal or operational dependencies (e.g., region/locale → traffic density → severity; storm and ice presence → loss of maneuverability; PSC status → baseline safety posture; weather → human error via reduced visibility), and a candidate directed-arc set was assembled. Where the literature offered alternative orientations for the same dependence, the orientation that supports interpretive clarity and inference was selected, recognizing that arc orientation between non-collider neighbors does not, by itself, alter the joint distribution under standard BN semantics [67,65,68]. (3) Expert review and refinement: the candidate DAG was reviewed by the authors and experts as domain specialists in maritime safety and accident analysis; arcs not supported by both the literature and accepted investigation practice were removed, intermediate constructs were introduced where justified (e.g., the Ship Profile), and v-structures at synoptic nodes such as Weather Condition were retained to preserve known dependencies among met-ocean variables (e.g., storm-sea state, ice-sea state). (4) Consistency check against the harmonized dataset: the resulting DAG was checked for compatibility with the recorded fields and state definitions, ensuring that every parent-child dependency could be quantitatively learned by EM without requiring fields not

present in the source repositories. (5) Empirical validation and selection: the resulting structure was submitted to the same stratified 10-fold ROC/AUC evaluation as TAN and ANB, with no further modification once parameters were learned, and was selected as the primary inference model only because it achieved the strongest cross-validated discrimination across severity states. This procedure is hybrid by design and is intended to combine the interpretive transparency of expert/literature topologies with the empirical accountability of cross-validated performance comparison.

All three networks, TAN, ANB, and expert, use the same node set and state definitions so that differences in performance are attributable to structure, not to input content. A comparative overview of the said approaches is provided in Table 1.

3.3. Parameter learning via expectation-maximization

Once a BN structure is fixed (TAN, ANB, or literature/expert), its quantitative parameters are the conditional probability tables (CPTs) for each node given its parents. We estimate CPTs using the EM algorithm on the same data set for all the three models.

EM is an iterative maximum-likelihood procedure for probabilistic models with latent or missing values. Let θ denote the collection of BN parameters (the CPT entries and any priors). Let X denote observed variables in the accident database and Z denote the unobserved or missing components. At iteration t , EM proceeds in two steps:

E-step (Expectation step)

Compute the expected complete-data log-likelihood using the current parameter estimates $\theta^{(t)}$. Conceptually, this means inferring the posterior distribution over the missing entries Z given the observed data X and $\theta^{(t)}$, and then taking the expectation of the full log-likelihood $\log P(X, Z|\theta)$ under that posterior. This produces an expected log-likelihood function

$$Q((\theta|\theta^{(t)}) = E_{Z|X, \theta^{(t)}}[\log P(X, Z|\theta)] \quad (2)$$

Intuitively, the E-step says: “Given my current BN, how likely are the missing pieces of each case, and how do those implied completions contribute to the likelihood?”

M-step (Maximization step)

Update the parameters by maximizing this expected complete-data log-likelihood:

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}} Q(\theta|\theta^{(t)}) \quad (3)$$

In a discrete BN, this corresponds to re-estimating each CPT entry as an (expected) normalized count of how often each child state appears jointly with each configuration of its parent states, where “expected” accounts for uncertainty in missing data from the E-step.

The E-step and M-step are alternated until convergence, i.e., until changes in θ fall below a small threshold. At convergence, the resulting CPTs are maximum-likelihood estimates under the assumed BN structure, while consistently handling incomplete accident records. We apply this same EM procedure to all three candidate networks to ensure comparability. To improve reproducibility, we additionally report the dimensionality of representative conditional probability tables (CPTs) in the selected BN. For a child node X with r_X states and parent-state space $q_X = \prod_{j \in \text{Pa}(X)} r_j$, the CPT contains q_X parent configurations and $r_X q_X$ stored probabilities, of which $(r_X - 1)q_X$ are free parameters. In the final expert/literature BN, the Severity Level node has five direct parents (Accident Type, Navigational Zone, Human Error, Ship Profile, and Weather Condition), giving $q_X = 4 \times 3 \times 2 \times 3 \times 5 = 360$ parent configurations, i.e., 1080 stored probabilities (720 free parameters). The Accident Type node has six direct parents (Route/Region, Reported Cause, Weather Condition, Navigational Zone, Ship Profile, and Human Error), giving $q_X = 6 \times 3 \times 5 \times 3 \times 3 \times 2 = 1620$ parent configurations, i.e., 6480 stored probabilities (4860 free parameters). Representative exported CPTs for the key downstream nodes are provided in Appendix [A].

Table 1
Comparison of BN structures used for accident-severity modeling in this study (commonalities and differences).

BN variant	How the structure is obtained	Structural constraint	Strengths for severity modeling	Limitations / risks	Best Use	Key references
TAN	Data-learned structure (score-/mutual-information-based learning)	Class/target node is a parent of each predictor; predictors can have at most one additional predictor parent, forming a tree among predictors	Strong accuracy-interpretability balance; captures key interdependencies without a dense graph; computationally efficient; typically more robust than unrestricted BN in moderate data	Still a restriction: only a tree among predictors (cannot represent complex dependency motifs); can miss higher-order interactions; needs careful discretization/state design	When better predictive performance than Naïve Bayes is required, while keeping the graph readable and stable under limited data	[12,78,79, 2]
ANB	Data-learned (often greedy/hill-climbing or constrained search)	Class/target node links to predictors; additional arcs among predictors are allowed with weaker constraints than TAN (often a forest or limited in-degree, depending on implementation)	More flexible than TAN; can capture richer dependencies among predictors; often improves classification when the true dependencies are not tree-like	Higher risk of overfitting than TAN if constraints are loose; structure search can be less stable with sparse data; interpretability can degrade as arcs grow	When multiple interacting predictors are suspected beyond a tree and have enough data (or strong regularization/constraints)	[80,81,26, 82]
Expert/Literature BN	Expert-elicited / literature-driven topology (then parameters learned/elicited)	Edges reflect domain logic and evidence; dependencies are explicitly encoded to preserve causal/operational meaning	Highest face validity and interpretability; supports causal storytelling; stable under sparse data; easier to justify pathways (useful for reviewers/stakeholders)	Predictive performance can be lower than optimized data-learned structures; may reflect expert bias or outdated assumptions; topology is time-consuming to build and audit	When interpretability and defensible causal pathways are paramount, or when data sparsity makes unconstrained learning unreliable	[75,2,76, 77]

Parameter learning was performed in GeNIe/SMILE using its built-in discrete EM routine. In GeNIe, EM supports three initialization modes (Uniformize, Randomize, and Keep original), and the associated confidence parameter is interpreted as an equivalent sample size (ESS). When uniform or random initialization is used, the software assigns a very low confidence (ESS = 1), which acts as a weak numerical regularization rather than a strong informative prior. Accordingly, rare parent-state combinations were not manually corrected after learning; instead, their estimation was stabilized through the software's EM initialization/confidence settings together with prior discretization of variables into operationally meaningful state bands, which reduced unnecessary fragmentation of the CPT space. No hand-tuned post-hoc modification of CPT entries was applied.

3.4. Probabilistic inference and BN quantities

After learning parameters, the BN provides full joint and conditional probability queries. The factorization

$$P(X) = \prod_i P(X_i | P_a(X_i)) \quad (4)$$

allows direct computation of posterior probabilities using Bayes' theorem. For example, to estimate the probability that Accident_Severity takes a high-severity state s^* given evidence E (e.g., "storm present", "loss of propulsion", "region = North Sea / port waters", "PSC status = detained"), we evaluate

$$P(\text{Accident}_{\text{Severity}} = s^* | E) = \frac{P(\text{Accident}_{\text{Severity}} = s^*, E)}{\sum_s P(\text{Accident}_{\text{Severity}} = s, E)} \quad (5)$$

This supports multiple forms of inference:

- **Predictive analysis:** forward propagation from known precursors (ship, weather, geography, compliance posture) to severity.
- **Diagnostic analysis:** backward inference from an observed severe outcome to likely causal contributors.
- **Scenario analysis:** forcing specific variables/states to fixed values (100 % evidence) either one at a time ("sequential") or jointly ("simultaneous") to assess how combinations such as "storm + ice + confined approach," or "detained PSC status + machinery failure," reshape the severity distribution.
- **Sensitivity and strength-of-influence analysis:** quantifying which individual factors and combinations have the greatest leverage on the probability of severe outcomes.

These inference modes are executed on the final selected model (see below) and reported in the Results and Analysis section.

3.5. Model validation and selection

Each of the three BN structures (TAN, ANB, and expert) is treated as a candidate model and trained with EM as described above. To evaluate their reliability as severity predictors, we apply stratified k-fold cross-validation (with $k=10$):

1. The accident dataset is randomly partitioned into 10 approximately equal folds, stratifying on Accident_Severity so that all severity levels are represented in each fold.
2. For each fold in turn, nine folds are used to learn the CPTs via EM (training set), and the held-out fold is used for testing.
3. For each test instance, the BN is queried to obtain the posterior probabilities of each Accident_Severity state.

Predictive performance is summarized using ROC curves and the corresponding AUC for each Accident_Severity state. In multi-state problems like accident severity, this is typically done in a "one-vs-

rest" fashion: for each severity state, that state is treated as "positive" and all other severity states as "negative," and the BN's posterior for that state is scored. Higher AUC indicates stronger discriminative ability. Confusion matrices on the held-out fold are also examined to understand systematic confusion between adjacent severity bands, which is important because maritime accident severity is ordinal and escalation thresholds can blur in adverse conditions. Because all CPTs were re-estimated independently within each of the 10 training folds, the validation procedure also serves as a sample-sensitivity check at the parameter level: the reported discrimination metrics reflect repeated re-fitting under perturbed samples rather than dependence on a single sample-specific parameterization. Exact reruns in GeNIe additionally require reporting the initialization mode and random seed, since a seed value of 0 uses the system clock. As EM may, in general, be sensitive to initialization and converge to local optima, we report these settings explicitly for transparency; however, a systematic multi-start search over alternative seeds/initializations was not performed and is therefore acknowledged as a limitation.

After cross-validation, we compare discrimination (ROC/AUC by state) and the best-performing and most defensible model, i.e., the model that predicts severity most reliably, is then adopted as the main analytical engine for downstream interpretation.

3.6. Downstream analyses

The selected BN is subsequently used to perform the core analyses reported in the Results and Discussion section:

- **Predictive analysis:** to estimate the probability distribution of Accident Severity under specific ship / weather / region / PSC configurations.
- **Simultaneous scenario analysis:** to evaluate high-risk coupled exposures (e.g., "storm + loss of propulsion + confined port waters in the North Sea").
- **Sensitivity and TRI analysis:** to quantify which individual variables, when forced to their worst vs. best operational states, most strongly change the probability of high-severity outcomes.
- **Strength-of-influence analysis:** to rank the relative leverage of key variables and to distinguish baseline amplifiers (e.g., weather/geometry) from escalation triggers (e.g., machinery failure, cargo fire).

These analyses allow us not only to forecast risk but to explain it in causal terms that align with real-world decision levers: traffic management, routing and tug allocation in constrained approaches, cargo/fire preparedness, machinery reliability and redundancy, and PSC-driven compliance posture.

4. Results and discussions

4.1. Model development

Maritime accident severity is best treated as an ordered, multi-state outcome that reflects the joint effect of proximate failure mechanisms (e.g., collision, capsizing), the ship's technical and human profile, the operating locale (port waters, coastal approaches, open sea), the broader regional sea area (e.g., North Sea, South China Sea), and contemporaneous met-ocean exposure (storms, ice, visibility, wind, sea state). Severe consequences typically arise when a hazardous event coincides with constrained geometry or adverse weather, or when loss of propulsion/steering undermines recoverability, patterns repeatedly highlighted in recent data-driven studies [12,2,13,1]. Against this backdrop, we cast severity as a graded target with discrete bands and seek a model that (i) integrates heterogeneous evidence without collapsing it to binaries, (ii) represents interdependencies among predictors (e.g., weather $\leftrightarrow$ locale; ship type $\leftrightarrow$ accident type), and (iii) delivers transparent diagnostics for operational use. A BN learned/estimated from

data fulfills these requirements by mapping the evidence configuration to a full probability vector over severity bands while retaining an interpretable, causal-style graph that supports prediction, diagnosis, and scenario reasoning within one framework.

Concretely, we develop a TAN BN model, allowing each predictor to depend on the severity node and on at most one other predictor through a maximum-weight spanning tree built on conditional mutual information. This preserves parsimony while capturing the strongest conditional couplings (e.g., storm presence with poor visibility; regional sea area with local water type). In parallel, we estimate an ANB variant and a literature/expert BN that encodes widely reported causal relations. All three models share the same discretized factor set: multi-state weather and sea-state variables, explicit geography at two layers (named sea region and local waters), ship particulars (type, age, tonnage/DWT bands, crew size), human/organizational signals (human error, reported cause), and accident mechanics (accident type). This design operationalizes insights from spatial hotspot mapping and recent severity analyses, bringing region and locale inside the probabilistic graph rather than leaving them as external descriptors, and yields a structure in which high-severity pathways emerge naturally from the data [12,2,1].

4.1.1. Factor selection

This study selects factors that are (i) mechanistically proximate to casualty escalation, (ii) consistently observable in international accident records, and (iii) actionable for prevention and preparedness. Proximate event descriptors, Accident Type, Reported Cause, and Human Error, enter because they govern the physics and recoverability of incidents (e.g., collision/contact vs. capsizing/grounding) and repeatedly explain variance in consequence severity beyond occurrence alone. Ship Type, Age, DWT, Class (RO) and Crew Size are retained as upstream *capacity and vulnerability* determinants: ageing and class regime proxy maintenance/assurance quality; manning levels affect detection, containment, and on-scene response.

Geography and operating context are included because severity is highly conditional on *where* and *how* an event unfolds. Named Route/Region nodes capture distinct traffic geometry, infrastructure proximity, and regional met-ocean climatology; Navigational Zone (harbor/coastal/open sea) encodes maneuvering room and tug/salvage access; Traffic Density, Time (day/night) and Season reflect encounter complexity and resource availability. Weather exposure is modeled with granular states, Visibility, Wind Speed, Sea State, Storm Presence, Ice Condition, and a synoptic Weather Condition, because escalation pathways (e.g., loss of control, progressive flooding, fire spread, abandonment risk) depend on these forcings, and coarse binaries have been shown to understate effect sizes.

In the selected literature/expert BN, the synoptic node Weather Condition (Clear, Foggy, Rainy, Snow, Stormy) functions as an aggregator of the granular met-ocean indicators (Visibility, Wind Speed, Sea State, Storm Presence, Ice Condition, Season, Time, Route/Region) rather than as a forcing variable that physically causes them. The directed arcs from these indicators into Weather Condition, therefore, encode probabilistic information flow consistent with the way accident records are populated. Investigators report the granular indicators directly and assign the synoptic descriptor by integrating them, not a physical cause-and-effect direction. Under standard BN semantics, arc orientation between non-collider neighbors does not, in itself, alter the joint distribution or the conditional-independence structure for downstream inference [67,68]: the Markov-equivalent reverse orientations would yield the same posterior over Severity Level given any combination of evidence on the met-ocean nodes. Physical dependencies between these granular indicators (e.g., storm $\rightarrow$ high wind $\rightarrow$ rough sea, ice $\rightarrow$ wave attenuation) are preserved indirectly through the v-structure formed at Weather Condition, which becomes active whenever evidence is set on that node or on any common descendant on the inference path. Human Error is, accordingly, given the granular indicators (Visibility, Wind Speed) as parents, together with Time, Crew Size, Traffic Density,

and Inspection History, so that performance-degrading meteorological exposure feeds directly into the Human Error node without redundant routing through the synoptic Weather Condition descriptor. Adding direct arcs between these granular indicators would not change the joint distribution materially, by Markov equivalence, but would inflate the affected child CPTs and compound the fragmentation already present at Weather Condition (eight parents, 36,864 parent configurations); published evidence indicates that such redundant structure degrades rather than improves CPT estimation under moderate sample sizes [83,84].

Finally, Inspection History (PSC) is included as a concise compliance/condition signal that bridges safety management quality and technical integrity; it improves prior risk calibration without presupposing causality from enforcement alone. All variables are discretized into operational bands to balance statistical learnability with interpretability (e.g., harbor/coastal/open sea; calm/slight/moderate/rough), enabling scenario analysis that maps directly onto decisions (routing, watchkeeping policies, tug allocation, and readiness postures). Collectively, this set provides coverage of *initiating mechanism*, *inherent capacity/vulnerability*, and *environmental/operational stressors*, the three levers that jointly determine whether an accident remains minor or becomes severe.

A description of the selected factors, reference to the literature base for their selection, and their corresponding states are presented in Table 2.

Severity Level is the target node of the BN, it was operationally defined using explicit consequence thresholds. The coding scheme considers both life-loss/injury consequences and financial/property losses, and collapses an underlying consequence framework into the three BN: Low, Medium, and High states [85,86]. Specifically, the least severe original band was retained as Low, the next band as Medium, and the most severe bands were merged into High. This aggregation preserves the ordinal meaning of accident consequence while reducing sparsity in the most severe tail, which is important for stable parameter estimation and interpretable BN inference. The breakdown of the defined states is depicted in Table 2a.

The DWT bands (LT50K, 50K–100 K, 100K–150 K, GT150K) were selected for compatibility with the global accident dataset (2015–2024), in which large bulk carriers, tankers, and containerships contribute substantial mass to the sample. A consequence of this choice is that the LT50K band aggregates small-to-medium cargo vessels and feeders. This is adequate for discriminating severity-relevant mass/inertia effects at the global scale, but it does not resolve the 1000–10,000 DWT substructure that is characteristic of the Finnish–Swedish Winter Navigation System fleet [43]. Where our scenario analysis in Section 4.4 touches on Baltic winter operations, we therefore treat the DWT posterior as a qualitative signal rather than a quantitative size prediction, and defer to the Baltic-specific literature for vessel-level size characterization.

To position the proposed factor set within the state of the art, Table 3 summarizes the factor sets adopted in 17 representative BN-based maritime accident severity studies and benchmarks them against the variables considered in this work. For each factor, the Observed out of Total row reports how frequently it appears across the benchmark studies, providing a transparent basis for assessing coverage gaps and motivating the inclusion of underrepresented predictors in the present model.

After finalizing the factor set and discretization, three BN variants were estimated on the accident-severity dataset: a data-learned TAN and ANB model, and a literature/expert structure. The profiles of the experts who finalized the BN of the third model based on the consulted literature and their experience have been shown in Table 4. For the data-driven variants, structure learning was performed within each training split only to preserve out-of-sample validity; the learned graph was then applied to the corresponding hold-out fold. The expert network retained its directed topology across folds but, like the others, was re-parameterized per split. For all three models, conditional probability

Table 2

Selected factors with their description and corresponding states.

No	Factor (node)	State	Role	Brief description
1	Severity Level	Low; Medium; High	Outcome	States defined in Table 2A.
2	Accident Type	Collision; Grounding; Contact; Capsizing	Event mechanism	Proximate incident category shaping consequence pathways.
3	Reported Cause	Navigational; Technical; Environmental	Causal class	Investigator's primary causal grouping; complements Accident Type.
4	Human Error	Yes; No	Human/organizational	Presence of human performance contribution (watchkeeping, decision, etc.).
5	Ship Type	Container; Bulk Carrier; General Cargo; Tanker; Ro-Ro	Ship profile	Functional class influencing vulnerability and exposure.
6	Ship Age	New; Intermediate; Old	Ship profile	0-7, 8-15, ≥16 (years)
7	DWT (tonnage band)	LT50K; 50K–100K; 100K–150K; GT150K	Ship profile	Deadweight capacity bands capturing mass/inertia and damage stability effects.
8	Classification (RO)	DNV; BV; ABS; CCS; ClassNK; LR	Assurance	Recognized Organization overseeing compliance/construction standards.
9	Crew Size	Small; Medium; Large	Manning	≤12; 13-24; ≥25 (UK/IMO/MARAD safe manning).
10	Inspection History (PSC)	No Issues; Deficiencies; Detentions	Compliance signal	PSC outcome as an indicator of baseline control condition.
11	Route/Region	North Sea; South China Sea; English Channel; Baltic Sea; Gulf of M.; Malacca Strait	Geography (regional)	Named sea areas associated with accident and distinct traffic, geometry, and met-ocean regimes.
12	Navigational Zone	Open Sea; Coastal; Harbor	Geography (locale)	Local operating waters at time of event; constrains maneuvering/response.
13	Traffic Density	Low; Medium; High	Operating context	0-1; 2-3; ≥3 (ships per hour/2 nm).
14	Time	Day; Night	Operating context	Light/visibility and crew state proxy.
15	Season	Summer; Winter; Spring; Autumn	Operating context	Seasonal envelope for region based on time.
16	Weather Condition	Clear; Foggy; Rainy; Snow; Stormy	Met-ocean	Synoptic condition at/near event; complements granular variables.
17	Visibility	Good; Fair; Poor	Met-ocean	Local visual range, ≥5 nm; 5-2 nm; <2 nm.

(continued on next page)

Table 2 (continued)

No	Factor (node)	State	Role	Brief description
18	Wind Speed	Low; Medium; High; Extreme	Met-ocean	≤ 2 ; 3-5; 6-8; ≥ 9 (Beaufort).
19	Sea State	Calm; Slight; Moderate; Rough	Met-ocean	0-2; 3-4; 5-6; ≥ 7 (Based on Douglas Scale).
20	Storm Presence	Yes; No	Met-ocean	Binary indicator of storm conditions (combined wind/pressure systems).
21	Ice Condition	Yes; No	Met-ocean	Presence of ice affecting hull loads/navigability and emergency response.
22	Ship Profile (derived)	Low Risk; Medium Risk; High Risk	Composite prior	composite prior summarizing baseline risk from age/size/class/type, used as an intermediate node.

Table 2a

Operational definition of the Severity Level node used in the BN.

BN severity state	Losses in terms of life	Losses in terms of financial/property loss*
Low	Prominent/serious injury or disability; no death	Less than 10,000 RMB or approximately less than USD 1,460
Medium	1-2 deaths	10,000 to 100,000 RMB or approximately USD 1,460 to USD 1,4600
High	3 or more deaths	More than 100,000 RMB or approximately more than USD 1,4600

* Note: Approximate USD values are shown using an indicative conversion of 10 RMB $\approx$ 1.46 USD for reader orientation only. The original coding thresholds were retained in RMB [85,86]. Where the life-loss and property-loss consequences fell into different bands, the higher severity band was used for assignment.

tables were estimated by EM on the training portion of each split using GeNIe/SMILE's built-in parameter-learning routine. To avoid over-interpreting sparse parent-state configurations, we relied on the software's weak initialization/confidence settings and on the prior discretization of variables into operationally meaningful bands, rather than

on any manual post-estimation adjustment of CPT entries. No hand-crafted arcs or post-hoc tweaks were introduced after learning.

Ship Profile is a derived intermediate node of the BN, with three ordered states (Low / Medium / High Risk) summarizing the pre-event baseline condition and vulnerability of the vessel. It is not a field directly recorded in accident databases; instead, it integrates six indicators that are observed in the records and registries (Ship Type, Ship Age, DWT, Classification, Inspection History, Crew Size) into a graded latent construct. Though weather conditions serve as a parent node for the ship profile, however, it was not considered in the node quantification, as weather conditions serve as an external factor. The node is constructed in two steps. First, an expert finalized, literature-grounded scoring rule (Table 4a) assigns each indicator an ordinal score (1-3) using expert opinions through evidence-based bands drawn from recent BN severity studies [12,16,33,28,13,52,53,56,58,93]. The six structural indicator scores are summed with equal weights and three-binned into Low (6-9), Medium (10-13), and High (14-18) Risk to give an initialization for each accident record. Second, the conditional probability table P (Ship Profile | parents) is re-estimated by EM together with the rest of the network, treating Ship Profile as a partially observed latent node. This approach combines the interpretability and reproducibility of an explicit scoring rule with the data-driven refinement of EM, while keeping CPT dimensionality manageable and preserving stable parameter estimation under the available sample size.

The resulting models support fully probabilistic inference over the multi-state Severity Level node and are used consistently across the subsequent analyses: predictive, sequential and simultaneous scenario experiments (evidence setting on combined factors), and sensitivity studies (influence/ranking of variables). Comparative cross-validation and ROC/AUC assessments are reported next to identify the best-performing specification; that selected model is then carried forward for detailed interpretation and policy-oriented what-if analysis.

Table 4

Profiles of the experts who finalized the third BN Structure with reference to literature using 5 experts [92,63,28,93,22,94-96].

Expert No	Role	Qualification	Experience (years)
E1	Captain	M.Sc.	24
E2	Chief Officer/Engineer	M.Sc.	19
E3	Port Authority/Maritime Administration Officer	M.Sc.	22
E4	Professor in Maritime Safety	Ph.D.	15
E5	Associate Professor in Maritime Safety and Risk Assessment	Ph.D.	9

Table 3

Factor coverage across prior maritime severity/risk studies (Y=Considered, N=Not Considered).

Reference/Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
[12]	✓	✓		✓	✓	✓						✓	✓	✓			✓	✓	✓			
[13]	✓	✓		✓	✓	✓						✓	✓	✓	✓		✓	✓				
[33]	✓	✓			✓	✓						✓	✓	✓			✓	✓				
[2]		✓		✓	✓	✓			✓			✓	✓	✓		✓	✓	✓	✓			
[4]		✓		✓	✓	✓			✓			✓	✓	✓		✓	✓	✓	✓			
[1]	✓	✓		✓	✓	✓			✓			✓	✓	✓		✓	✓	✓	✓			
[34]	✓	✓			✓	✓				✓		✓	✓	✓		✓	✓	✓				
[87]	✓	✓			✓	✓						✓	✓	✓		✓	✓	✓				
[3]	✓	✓	✓		✓	✓					✓		✓	✓	✓	✓	✓	✓				
[14]	✓	✓			✓						✓		✓	✓	✓		✓	✓				
[88]		✓			✓									✓		✓	✓				✓	
[89]	✓	✓			✓	✓					✓					✓	✓				✓	
[90]					✓	✓				✓				✓			✓	✓				✓
[23]		✓			✓	✓						✓	✓	✓		✓			✓			✓
[24]	✓	✓		✓	✓	✓			✓			✓	✓	✓	✓		✓	✓				✓
[91]		✓		✓	✓	✓						✓	✓	✓		✓			✓			✓
[31]	✓	✓		✓	✓							✓	✓	✓			✓	✓	✓			
Observed out of Total=17	11	16	1	8	17	13	3	0	1	2	3	10	5	13	4	8	12	11	7	0	1	3

Table 4a
Indicator-level scoring rule used to initialize the Ship Profile prior.

Indicator (parent)	State → score	Rationale
Ship Age	New → 1; Intermediate → 2; Old → 3	Maintenance/structural-integrity decline with age; well-evidenced severity driver.
DWT (tonnage)	LT50K → 1; 50K–100K → 2; 100K–150K → 2; GT150K → 3	Mass/inertia and damage-stability effects; gross-tonnage among top-5 severity drivers.
Ship Type	Bulk Carrier/Tanker → 3; General Cargo/Ro-Ro → 2; Container → 2	Type-specific stability, free-surface, and cargo-hazard profiles drive consequence asymmetry.
Classification (RO)	IACS member → 1; Non-IACS recognized → 2; Other/none → 3	Class regime quality proxies assurance and survey rigor.
Inspection History (PSC)	No issues → 1; With deficiencies → 2; Detained → 3	Most direct compliance/condition signal; Paris MoU performance lists corroborate.
Crew Size	Adequate → 1; Marginal → 2; Below-typical → 3	Manning vulnerability affects detection, containment, and on-scene response.

4.2. Model validation

To assess the reliability of the accident-severity BNs, we used stratified 10-fold cross-validation across the full dataset, preserving the empirical proportions of the severity states within each fold. Records were first randomly permuted to remove any ordering effects, then partitioned into ten equal blocks; each block served once as the hold-out test set while the remaining nine formed the training set, cycling through all folds to obtain a variance-reduced estimate of generalization. For the data-driven models (TAN, ANB), structure learning was performed inside each training split only, and parameters for all three candidates (TAN, ANB, and the literature/expert network) were re-estimated on that split via EM before scoring the held-out fold. Predictive discrimination was quantified with Receiver Operating Characteristic (ROC) analysis using Area Under the Curve (AUC) in a one-vs-rest scheme over the severity classes [97], a standard approach for multi-category outcomes [98,99]. Higher AUC values, above 0.5, indicate better separation between the focal severity level and alternatives.

Because the severity classes are imbalanced, the validation results are interpreted primarily through the class-wise one-vs-rest ROC/AUC values rather than through raw overall accuracy alone. This choice is appropriate for the present BN setting because the models return posterior probabilities for each severity state, and AUC evaluates threshold-independent discrimination for each class-specific ranking problem. At the same time, several limitations should be noted. Raw accuracy may be optimistic in imbalanced datasets because acceptable overall correctness can still coexist with weak recognition of the minority high-severity class. In addition, for multiclass problems, one-vs-rest ROC/AUC does not provide a unique global summary, and aggregate measures can differ depending on the averaging scheme: micro-averaged indicators are influenced more strongly by the larger classes, whereas macro-averaged indicators assign equal weight to each class and therefore reflect minority-class performance more directly. For this reason, the present study supplements the ROC/AUC analysis with the confusion matrix and class-wise precision, recall, and F1-score, and also reports macro-recall (balanced accuracy) and macro-F1 as complementary indicators of performance under class imbalance, shown in Table 6.

To compare the predictive discrimination of the three BN formulations, ROC curves were generated for each severity class (low, medium, high) under a one-vs-rest setting, and the resulting AUC values were compiled in Table 5. As shown, the literature/expert BN clearly outperforms both data-driven alternatives (ANB and TAN) across all three severity levels. In particular, the expert model attains AUCs above 0.81

Table 5
ROC/AUC values for the three BN models (one-vs-rest by severity level).

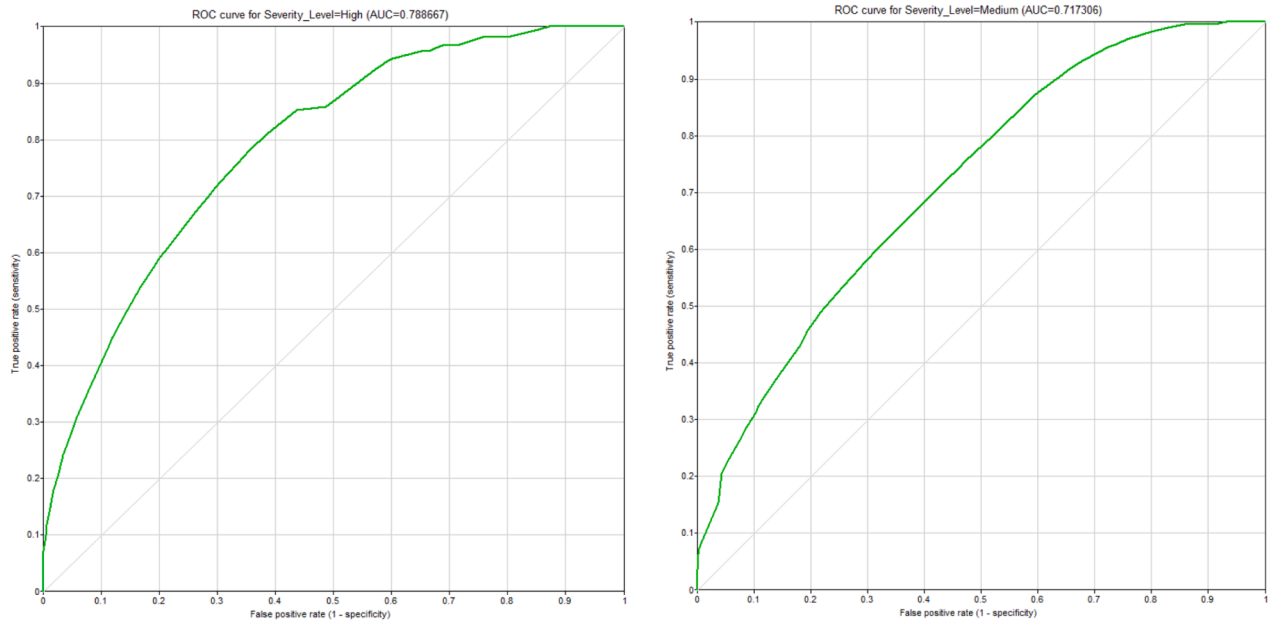
Model / Severity level	Low severity	Medium severity	High severity
ANB (data-driven)	0.6919	0.5530	0.6279
Literature / Expert BN	0.8161	0.7173	0.7887
TAN (data-driven)	0.6966	0.5596	0.6275

for low, 0.72 for medium, and 0.79 for high severity, whereas the ANB and TAN models remain around 0.69–0.70 for low, fall to about 0.55–0.56 for medium, and reach only about 0.63 for high. This pattern indicates that severity in the present dataset benefits from a structurally informed graph in which key causal channels are explicitly encoded rather than inferred only from data. On that basis, the literature/expert BN is selected as the reference model for the subsequent predictive, and scenario analyses, while the ANB and TAN results are retained as baselines. For completeness, representative ROC plots for the best-performing (literature/expert) model are provided in Fig. 2, while the ROCs for TAN and ANB models are provided in Appendix A.

To complement the cross-validated discrimination metrics in Table 5, we additionally report the EM-converged data log-likelihood ($\log L = \log P(D | G, \theta)$) for each candidate network, together with its free-parameter count $k = \sum (r_X - 1) \cdot q_X$ and the resulting sample-to-parameter ratio N/k . Standard information criteria such as $AIC = -2 \log L + 2k$ and $BIC = -2 \log L + k \log N$ are not reported as selection indicators in this study, because the present setting violates the regime under which their asymptotic theory is valid: Burnham & Anderson [100] and Hurvich & Tsai [101] recommend $N/k > 40$ as a minimum requirement for AIC/AICc validity [100–102], whereas the candidate networks here operate well below this threshold ($N/k = 1.27$ for TAN, 2.42 for ANB, and 0.012 for the literature/expert BN; see Table 5A). The literature/expert BN attains a low N/k because two domain-motivated synoptic/aggregator nodes, Weather_Condition (parents: Visibility, Wind_Speed, Sea_State, Storm_Presence, Ice_Condition, Season, Time, Route_Region) and Ship_Profile (parents: Ship_Type, Ship_Age, DWT, Classification, Inspection_History, Crew_Size, Weather_Condition), together account for approximately 80 % of its free parameters, by structural design and not by parameterization choice. Reducing these parent sets for the sake of a more favorable AIC/BIC would remove causally meaningful pathways that the BN encodes (Sections 3.2, 3.2.1) and would compromise the operational realism that motivates the modelling framework. In the small- N/k regime, accepted methodological practice is to base structure selection on out-of-sample, threshold-independent discrimination together with proper probabilistic scoring rules, rather than on penalty-based information criteria [103,67, 104]. We therefore use the cross-validated ROC/AUC results in Table 5 as the primary structure-selection indicator, supplemented by the classification battery in Table 6 and the Brier-score-based probabilistic-consistency assessment in Table 6A. Within this framework, $\log L$ itself is reported in Table 5a for transparency: it shows that TAN attains the highest in-sample log-likelihood while the literature/expert BN attains the strongest cross-validated discrimination, the classical pattern of a complex but generalizing structure outperforming a simpler over-fit alternative on held-out data.

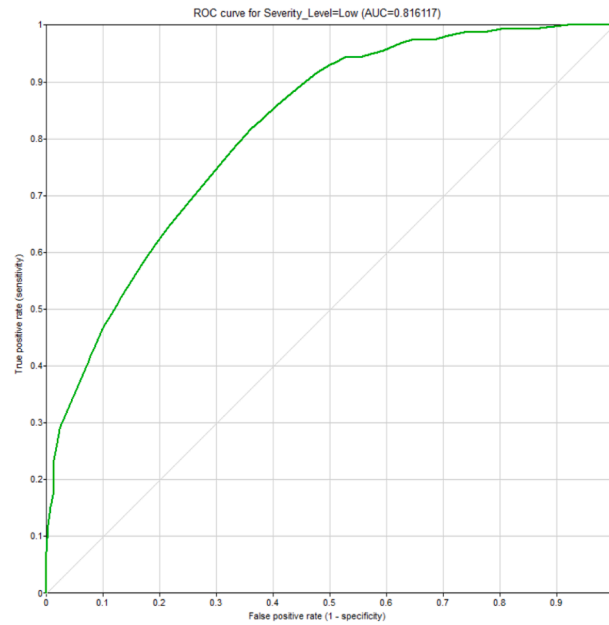
For the selected literature/expert BN, the confusion matrix and derived classification metrics show class-wise recall values of 0.506, 0.832, and 0.209 for Low, Medium, and High severity, respectively; the corresponding macro-recall (balanced accuracy) and macro-F1 are 0.516 and 0.524, confirming that model performance is more moderate when all severity classes are weighted equally, especially for the minority high-severity class.

Beyond discrimination (Table 5) and classification (Table 6), the consistency of the predicted posterior probabilities themselves was assessed for the selected literature/expert BN by computing the multi-class Brier score on the GeNIe 10-fold cross-validation output, $BS = (1/N) \sum_i \sum_k (p_{\{i,k\}} - o_{\{i,k\}})^2$, where $p_{\{i,k\}}$ is the model's posterior



(i) ROC curve for the severity level=High

(ii) ROC curve for the severity level=Medium



(iii) ROC curve for the severity level=Low

Fig. 2. ROCs with corresponding AUC values for Lit/Expert BN.

probability on Severity state k for record i and $o_{i,k} \in \{0,1\}$ is the indicator of the observed Severity state. The Brier score is a strictly proper scoring rule that admits an additive decomposition into reliability, resolution, and uncertainty components [105–108]. The corresponding Brier Skill Score (BSS) was computed against a non-informative baseline that predicts the marginal class frequencies for all records, for which $BS_{ref} = 1 - \|p\|^2$ evaluates to 0.621 under the empirical class distribution (Low: 30.1 %, Medium: 49.8 %, High: 20.1 %; see Table 6 supports). The selected literature/expert BN attains $BS = 0.494$ and $BSS = 0.204$, indicating a 20.4 % improvement in joint

calibration and resolution over the marginal-frequency baseline. The per-class Brier components (Table 6A) show the largest contribution from the Medium-severity class (0.211) and the smallest from the High-severity class (0.131), consistent with the higher recall but moderate precision of the Medium class observed in Table 6. Robustness with respect to data perturbation is implicit in the validation design: every CPT was independently re-estimated by EM within each of the 10 training folds (no parameter sharing across folds), so the metrics in Tables 5, 5a, 6, and 6a reflect repeated re-fitting under perturbed samples rather than a single sample-specific parameterization. Together

Table 5a

Sample-to-parameter ratios and EM-converged data log-likelihood for the three candidate BN structures, with the cross-validated AUC ranges from Table 5 reproduced for direct comparison.

BN structure	Free parameters, k	Data log-likelihood, log L	Sample-to-parameter ratio, N/k	Selection-relevant indicator
TAN	827	−20,984.85	1.27	AUC: 0.69–0.70 (Low), 0.55–0.56 (Med), 0.63 (High)
ANB	434	−21,253.43	2.42	AUC: 0.69–0.70 (Low), 0.55–0.56 (Med), 0.63 (High)
Literature/Expert (selected)	85,477	−21,410.30	0.012	AUC: > 0.81 (Low), > 0.72 (Med), > 0.79 (High)

with the cross-validated AUC values in Table 5 and the classification battery in Table 6, these results support the use of the predicted posteriors over Severity Level as probabilistic decision-support inputs for consequence-aware routing, traffic-prioritization, and emergency-preparedness decisions.

4.3. Prediction analysis

The prediction run of the literature/expert BN in Fig. 3 shows that severity is not being driven by a single variable, but by a layered chain that starts with where and how the ship is operating, passes through what type of accident actually materialized, and is then amplified or damped by environmental resistance (weather/sea/ice/visibility) and by the latent condition/compliance of the vessel. The prior on the target already leans toward *medium* severity ($\approx 40\%$), which is consistent with global statistics where a large share of reported events are neither trivial port scratches nor catastrophic foundering, but operationally disruptive contacts, groundings, and machinery-related casualties. That prior is being “fed” by several high-entropy parents: *Accident Type* is diversified (collisions and groundings together exceed half of the mass), *Ship Profile* is not dominated by the low-risk state (high-risk $\approx 1/3$), and *Weather Condition* and *Sea State* are both populated in the “bad” and “rough” states, so the model sees the database as genuinely mixed rather than benign. This is why the network keeps the medium band as the most likely outcome before any evidence is entered.

Another prominent aspect in this structure is the double gateway to severity. One gateway is the *navigational/exposure path*: *Route/Region* → *Traffic Density* → *Navigational Zone* → *Accident Type* → *Severity Level*. This is the spatial–operational aspect that recent AIS-linked and EMSA analyses have been emphasizing, dense routes, confined approaches, and port/coastal operations systematically push accidents toward costlier consequences because there is less room to maneuver, more targets to

hit, and more infrastructure to damage [10,1]. The second gateway is the *condition/compliance path*: *Inspection History* (no issues / deficiencies / past detentions) and *Classification* feed into *Ship Profile* and *Human Error*, which in turn connect to *Accident Type* and severity. In other words, the model is predicting that ships that have recently shown problems, or that are sailing with weaker class signals, are the same ships for which a “simple” contact is more likely to escalate into a serious event, exactly the PSC–accident linkage that several authors have called to integrate but usually treat in separate models [4,13].

A particularly valuable aspect of the observed results is the prominent role of Reported Cause. More than half of the sample is coded as navigational cause (54 %), with technical causes accounting for 22 % and environmental for 25 %. This trend represents situations that recent investigations have highlighted, navigational misjudgment in high traffic, machinery/power failures preceding allision, cargo-related fires intensified by weather, where the same accident label can have very different severity trajectories depending on the underlying cause [9,5]. The environmental block behaves credibly. *Storm presence*, *high wind*, *rough sea state*, *poor visibility*, and *ice condition* all point directly to *Weather Condition*, and that node has a direct arc to *Severity Level*. This means the BN is not forcing weather to act only through the choice of route or accident type; it is allowed to raise the consequence of whatever accident actually occurs. That is important because many of the recent high-profile events, fires on vehicle carriers, machinery/power losses leading to allision, cargo-related explosions, did not start as inherently “high-severity” types but became high severity because they happened in bad met-ocean settings or close to infrastructure [9,5]. The model mirrors that logic: a collision in an open sea, low-traffic, clear-weather setting will tend to stay in the low/medium bands; the same collision in a storm-affected, narrow, high-traffic region will move upward.

Another strong feature of this prediction structure is the centrality of *Human Error* (69 %). It collects information from *Time* (day/night), *Inspection History*, *Weather Condition*, and *Reported Cause*, and then feeds into *Accident Type* and *Severity Level*. That is a compact way of representing something that has been repeatedly observed in casualty investigations: human/organizational failings are not independent of environment and not independent of ship condition, they co-occur. By placing Human Error between these blocks, the BN can increase the probability of a more damaging accident whenever the environment is demanding and the vessel has a weaker safety/compliance record, which

Table 6a

Per-class Brier components and overall multiclass Brier score and Brier Skill Score for the selected literature/expert BN on the 10-fold cross-validation held-out predictions. The reference Brier score for a marginal-frequency classifier under the empirical class distribution is $BS_{ref} = 0.621$.

Severity state	Per-record mean ($p - o$) ²	Empirical share of records (Table 6)
Low	0.153	30.1 %
Medium	0.211	49.8 %
High	0.131	20.1 %
Multiclass Brier score (sum)	BS = 0.494	BSS vs baseline = 0.204

Table 6

Confusion matrix and class-wise classification performance for the selected literature/expert BN model.

Actual class	Predicted Low	Predicted Medium	Predicted High	Support	Precision	Recall	F1-score
Low	160	149	7	316	0.630	0.506	0.561
Medium	71	435	17	523	0.598	0.832	0.695
High	23	144	44	211	0.647	0.209	0.315
Column total	254	728	68	1050			
Overall / Macro summary					Macro precision = 0.625	Balanced accuracy = 0.516	Macro F1 = 0.524

Note: Precision, recall, and F1-score are reported on a one-vs-rest basis for each severity class. Balanced accuracy is the macro-average of class-wise recall.

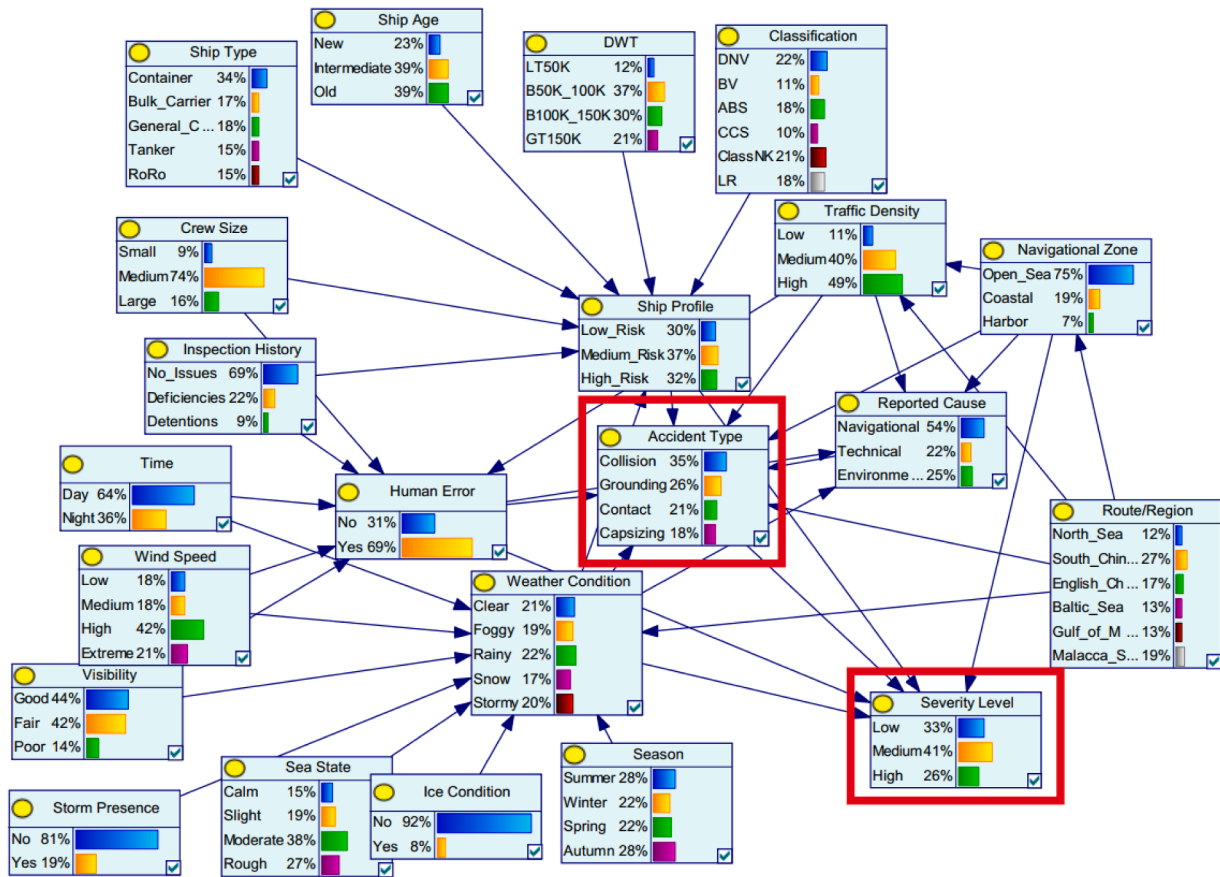


Fig. 3. Results for prediction analysis of the selected Expert/Lit model.

is fully consistent with the human-in-adverse-conditions narratives in relevant literature [12,2].

Technically, what this prediction run shows is that the expert-EM-style structure we adopted is well calibrated for multi-causal escalation: (i) exposure and geometry explain *why* an accident happens there; (ii) ship/human/PSC signals explain *how prepared* the ship was to prevent or contain it; and (iii) weather/sea/ice explain not only the occurrence, but *whether the same accident will stay small or become serious*. The prediction results confirm that the expert/literature BN is capturing the right *dynamics of severity formation* in contemporary shipping: it is spatially aware, operationally realistic, and sensitive to compliance signals, which is what the recent BN maritime safety literature has been advocating but seldom operationalizes in a single, integrated model.

4.4. Scenario analysis

1. Congested coastal collision and grounding case (exposure–geometry escalation).

- Evidence: *Traffic Density = High, Navigational Zone = Coastal, Accident Type = Collision and grounding, Reported Cause = Navigational.*

In the first scenario (collision in coastal waters under high traffic), as depicted in Fig. 4, the network settles on a severity profile of Low 31 % – Medium 49 % – High 20 %. That shape is important: almost half the probability mass sits in the medium band, signaling that, for the model, a coastal collision with a clear navigational reported cause (100 %) and human error active (≈ 71 %) is a serious but usually *manageable* event. Several upstream nodes explain why the model does not immediately escalate to high: the ship-profile node still has only 30 % in High_Risk (with Low_Risk 25 % and Medium_Risk 45 %), the classification and

DWT nodes stay broadly distributed, and the environmental layer, although unfavorable in parts (stormy 17 %, rough sea 27 %, poor visibility 14 %), does not force an extreme state. In other words, the model is reading this as “busy coastal waterway, human/navigational failure, collision, recoverable onsite,” which in accident statistics is exactly the regime that produces most *recorded* events but not most *catastrophic* outcomes. This is consistent with empirical BN studies that find collisions in port/coastal approaches to be frequent but predominantly medium-severity because tugs, VTS, and SAR assets are close, traffic is monitored, and energy levels are often partially attenuated in constrained waters [12,14].

The second scenario keeps almost everything constant, same coastal zone (100 %), same high traffic (100 %), same navigational reported cause (100 %), same human error mechanism active, but flips the accident type from collision to grounding, as shown in Fig. 5. That single change produces a visibly different posterior over the target: Low drops from 31 % to 18 %, Medium from 49 % to 42 %, and High jumps from 20 % to 40 %. Two things are happening simultaneously. First, grounding is treated by the network as a *persistent* accident state: unlike a collision, the ship remains in a hazardous configuration until it is refloated, lightered, or towed, so any adverse environmental node (rough sea, storm presence, poor visibility) now multiplies the exposure time. Second, the model responds to grounding by reclassifying the ship profile upward, High_Risk rises from 30 % $\rightarrow$ 35 % and Medium_Risk contracts from 45 % $\rightarrow$ 35 %, which is the model’s way of saying: *given this kind of accident in this kind of place, the underlying ship/operation was probably more stressed than the initial priors suggested*. That reclassification then pushes additional mass into the high-severity band. This is exactly what recent analyses of North Sea and Asian approach groundings report, when strandings happen in traffic-dense, geometry-constrained areas, the probability of hull damage, environmental release, or complex

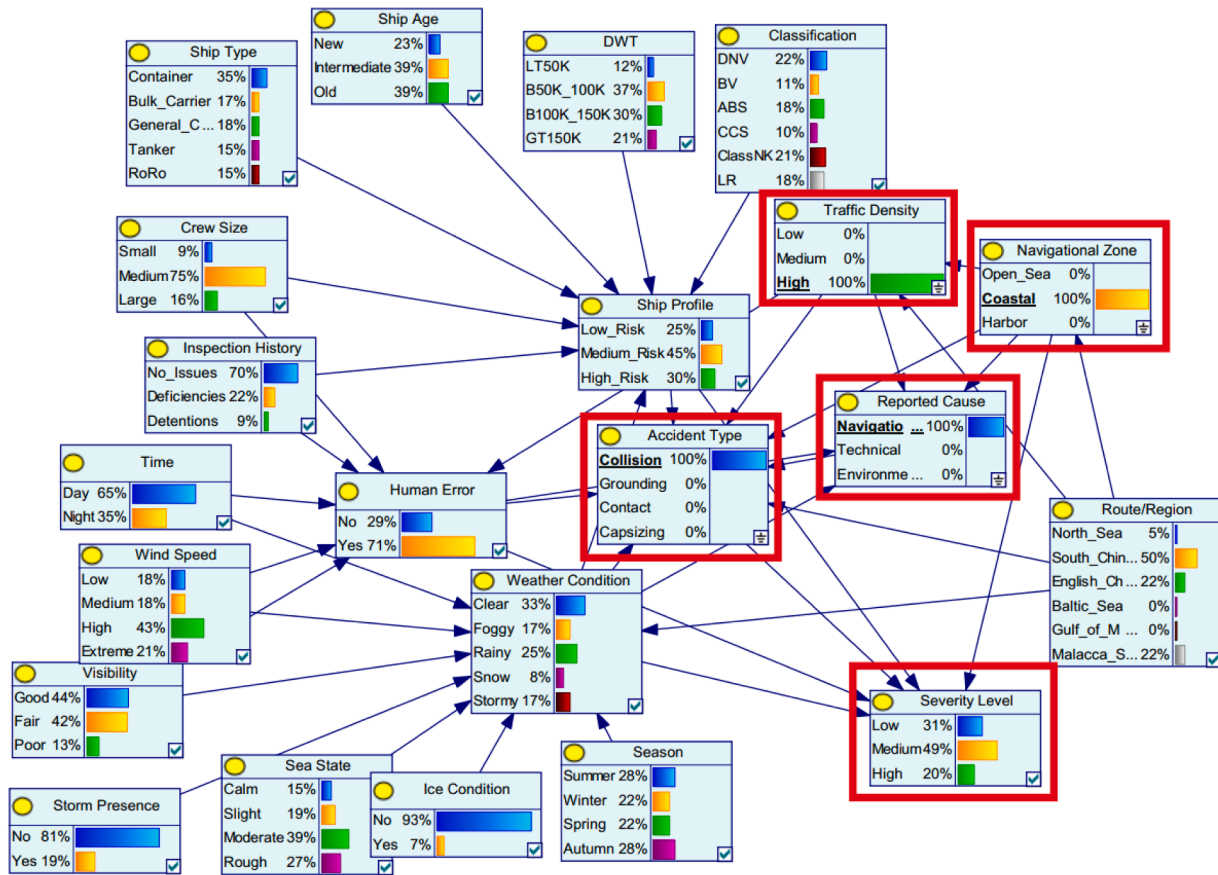


Fig. 4. Scenario 1 with evidence at collision.

salvage rises sharply compared with simple contacts, even when the initiating factor is the same (navigational/human) and even when weather is only moderately bad [10,109,1].

A second insight from the pair of scenarios is that cause alone does not drive severity. In both runs, the reported cause is 100 % navigational and human error is majority in “Yes,” yet the outcomes diverge strongly once the *physical expression* of that error is changed. That tells two things about the BN: (i) it has correctly separated *why* the event happened (navigational lapse) from *what physical state resulted* (collision vs. grounding); and (ii) it has learned that some accident types are structurally harder to terminate than others in coastal/high-traffic settings. This separation is important because several earlier maritime BN studies treated “navigational” or “human error” as near-sufficient precursors of high severity, which can overstate their effect when the consequent accident is actually a short, contact-type event. Here, the model behaves more realistically: navigation + human error *needs* to combine with a persistence-intensive accident type (grounding, capsizing) *and* an exposed locale to push the posterior toward the high band.

Finally, the comparison gives a useful operational reading. The collision scenario says: in crowded coastal water, medium is the design, expected, and most probable outcome, even with human error, because proximity of assistance, limited speed, and rapid detection cap escalation. The grounding scenario says: in the *same* water, with the *same* human and navigational precursors, the system’s buffer vanishes, the event sits almost in a 40/42 % high-or-medium window, because re-floating, cargo handling, and environmental protection are time- and resource-dependent. For practitioners this means that surveillance, tug allocation, and pilotage rules should be more tightly coupled to *accident type + locale* than to *reported cause* alone, and it corroborates the emphasis in recent incident reviews that stranding in coastal bottlenecks deserve higher default readiness and region-specific contingency

planning than routine collisions, even when both are “navigational” on paper.

2. Weather-driven loss of control

- Evidence: *Weather Condition = Stormy, Visibility = Poor, Sea State = Rough, Storm Presence = Yes, Reported Cause = Environmental*.

Setting simultaneous evidence on the “weather cluster” (stormy weather = 100 %, poor visibility = 100 %, rough sea state = 100 %, storm present = 100 %) and declaring the reported cause as *environmental*, as indicated in Fig. 6, pushes the network into a markedly different operating regime. The posterior over Severity Level becomes Low 26 % – Medium 41 % – High 33 %, i.e., the high-severity band gains more than ten percentage points compared with the earlier “coastal collision with navigational cause” case (High = 20 %), although it still remains slightly below the “coastal grounding” case (High = 40 %). This pattern indicates that the model treats extreme met-ocean conditions as a *broad escalator* of consequences: it does not deterministically force accidents into the worst class, but it raises the floor of expected severity across all plausible accident types because response, controllability, and survivability are simultaneously degraded.

The internal structure of the posterior explains this behavior. With weather fixed at stormy and the sea state at rough, the Accident Type node shifts away from purely contact-like events toward grounding (37 %) and capsizing (22 %), both of which are well known to be weather-exposed, high-energy outcomes in open/coastal waters. At the same time, the Navigational Zone remains dominated by open sea (74 %), which, in combination with poor visibility and storm presence, implies longer tow/salvage times, more difficult damage control, and higher exposure to secondary hazards (progressive flooding, cargo shift, loss of stability). Because the Reported Cause is fixed at *environmental* (100 %),

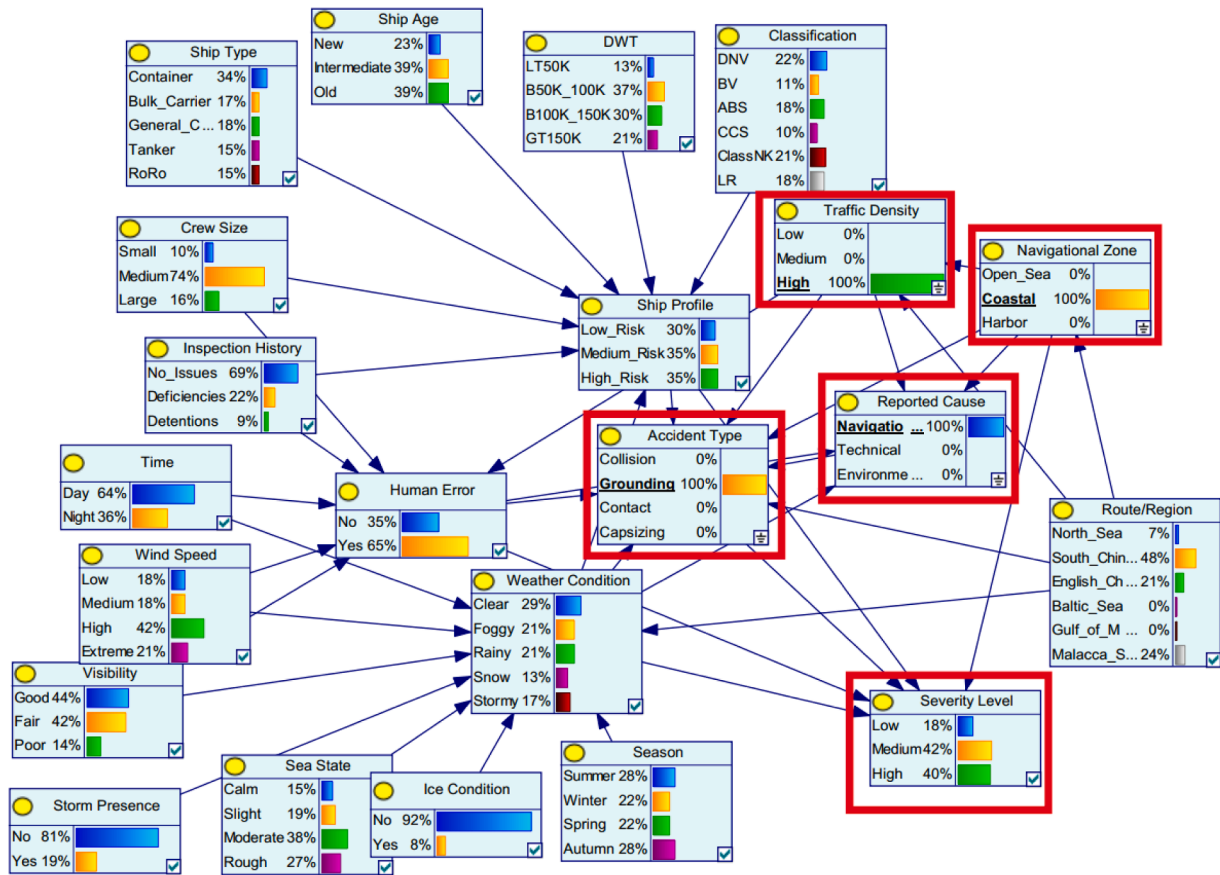


Fig. 5. Scenario 1 with evidence at grounding.

the model no longer attributes the event to a correctable human or navigational lapse; instead, it recognizes a driver that is exogenous to the ship and therefore harder to neutralize mid-voyage. This loss of controllability is what allows the high-severity probability to climb even though the ship-profile node stays fairly balanced (Low-risk 31 %, Medium-risk 37 %, High-risk 32 %). In other words, the model is saying: even a decently profiled ship can be pushed into severe outcomes when several environmental stressors co-occur.

Compared with the grounding scenario, the weather-driven case is *more diffuse*: high severity rises (33 %) but medium remains the modal class (41 %). That is consistent with casualty statistics in heavy weather, where many incidents are serious but not all become total losses because crews and systems can still partially mitigate (heave to, alter course, call for assistance) if stability is not yet critically compromised. What the network captures, however, is that multi-hazard weather (storm + rough sea + poor visibility) increases the range of serious accident types simultaneously, grounding on a lee shore, capsizing in beam seas, collision in poor visibility, so the tail of the distribution thickens even without changing the ship, crew, or traffic inputs. This reinforces a key managerial insight: environmental conditions act as *risk multipliers* on top of whatever technical and human vulnerabilities are already present; they should therefore be modelled and monitored as first-class variables, not as binary “bad-weather” flags. Operationally, such a scenario argues for earlier voyage adjustments (speed reduction, rerouting, delaying harbor approaches), dynamic tug/escort allocation in constrained waters under storms, and stricter thresholds for permitting high-wind or low-visibility movements, especially when the route/region node points to busy or geomorphologically constrained areas (e.g., English Channel, Malacca–Singapore).

3. Weak compliance and degraded ship profile scenario

- Evidence: *Inspection History* = *Detentions*, *Ship Age* = *Old*, *Ship Profile* = *High risk*, *Human error* = *Yes*, *Accident Type* = *Capsizing*, *Reported Cause* = *Vary*.

This scenario deliberately piles up the most degradation-prone attributes in the network, an old vessel, a high-risk ship profile, a history of detentions, and an ongoing human-error condition, with the accident type fixed at capsizing, depicted in Fig. 7. In the learned structure, each of these nodes feeds into severity through a different pathway: age and profile proxy long-term material/maintenance decline; detention history signals systemic compliance and safety-management weaknesses; human error opens the door to escalation during the emergency itself; and capsizing is one of the accident types with the steepest prior toward high consequences. Setting them simultaneously therefore creates a “stress-test” state of the BN: the model is asked what kind of outcome should be expected when the ship is old, organizational defenses have already failed in the past, the current event is human-factor dominated, and the physical mechanism is among the deadliest.

Against that stressed baseline, the reported cause acts as a *severity modulator* rather than a first-order driver. When the cause is set to navigational, the model returns about 45 % high-severity, with the remainder split between medium (≈ 29 %) and low (≈ 26 %). This outcome is consistent with capsizing triggered in constrained or traffic-dense waters (the scenario also holds traffic density at medium–high and navigation in open/coastal waters), where rapid external support and relatively benign met-ocean conditions keep a portion of cases in the medium band. Shifting the cause to technical slightly relaxes the tail: the high-severity share drops to ≈ 43 %, and both low and medium rise marginally (to ≈ 27 % and ≈ 30 %, respectively), as shown in Fig. 8. A technical explanation for the capsize signals a failure in machinery, steering, or stability systems, but it also implies that the triggering

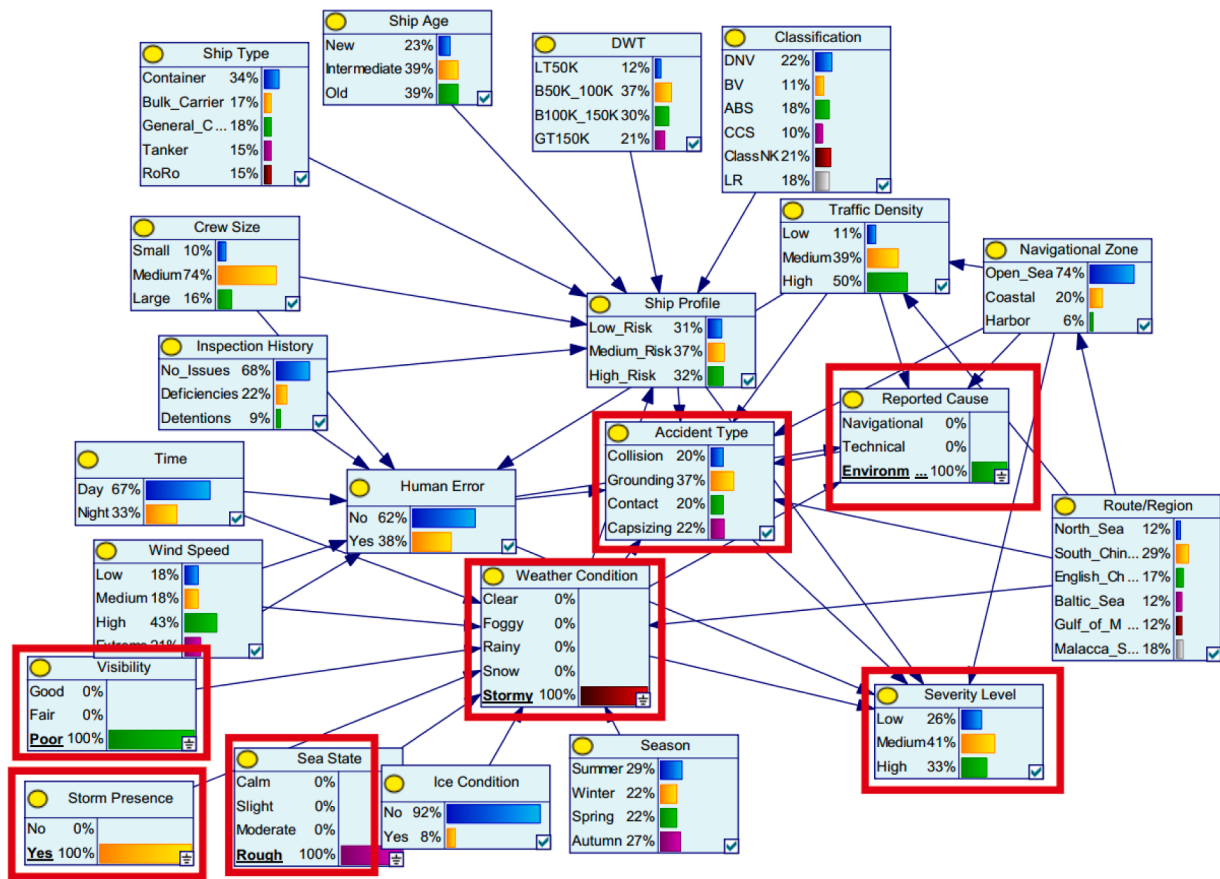


Fig. 6. Results for evidence setting of scenario 2.

condition may develop over minutes rather than seconds, allowing for partial corrective action and more effective abandonment; the network therefore reassigns probability from the extreme band to the intermediate ones.

The picture changes most clearly when the cause is set to environmental. Under otherwise identical evidence, as evident in Fig. 9, the high-severity probability rises to $\approx 49\%$, the highest of the three, and low severity falls to $\approx 23\%$. This is exactly the direction expected from the structure of the BN: an environmental cause links the capsizes to exogenous loads, storm, sea state, wind, that also degrade visibility and hamper rescue. In the environmental case, the network already “knows” that the ship is old, high-risk, and coming from a history of detentions; adding exogenous forcing tells the model that whatever residual safety margin remained is unlikely to be recovered by crew action. In other words, for navigational or technical failures, the human-error node can still play a *corrective* or *mitigating* role (e.g., right maneuver, ballast adjustment, timely distress), whereas under environmental forcing the same human-error state becomes *compounding*: the crew must execute more complex tasks in worse conditions, so the model pushes probability mass upward.

Taken together, these three runs show an important dynamic: once the ship/organization side is degraded (old + high-risk + prior detentions + human error), the type of immediate trigger mainly determines how much probability is left for medium outcomes. Navigational and technical explanations leave some space for recovery, and the BN fills that space with medium-severity cases; environmental explanations close that space and expand the high-severity tail. This also clarifies the operational signal: for legacy ships with poor inspection histories that operate in high-traffic sea areas, the largest payoff is not only in reducing human-error likelihood but in *de-exposing* those ships to severe environmental windows, because, as the third scenario shows,

once such a ship capsizes in weather-dominated conditions, the model’s posterior sits almost one-half in the worst band.

4. Scenario: Regional and Seasonal Stressors

- ice present, winter season, snowing, grounding with environmental cause

Setting evidence simultaneously on Ice_Condition = Yes, Season = Winter, Weather_Condition = Snow, Accident_Type = Grounding, and Reported_Cause = Environmental concentrates risk in a classic high-latitude winter profile. Under this evidence bundle, the model’s posterior places the event in the Baltic Sea ($\approx 100\%$), with Open-Sea operations ($\sim 81\%$) and Medium traffic density ($\sim 91\%$). The conditional severity distribution shifts toward the right tail (High $\approx 43\%$, Medium $\approx 41\%$, Low $\approx 17\%$; Fig. 10), indicating a materially greater propensity for serious consequences relative to the unconditional baseline (High $\approx 26\%$). We note explicitly that the values reported here for Wind_Speed (High/Extreme $\approx 63\%$) and Sea_State (Rough $\approx 25\%$) are posterior inferences rather than input evidence, i.e., they describe the model’s learned belief about the environmental conditions that co-occur with this class of event in the dataset, not a counterfactual claim about wave activity under ice cover.

The dynamics behind this shift are physically consistent and aligned with prior evidence. First, ice and snow degrade maneuverability and perception: added hull-ice resistance, reduced rudder/propeller effectiveness, and slush/pressure-ridge interactions shrink the safety margin for track-keeping; snow further impairs lookout and sensor performance, raising the likelihood that small navigational deviations culminate in seabed contact. The severity escalation in this scenario is not driven by wind-wave forcing under ice cover, which, consistent with Baltic oceanography, is materially attenuated by fast-ice cover [110,

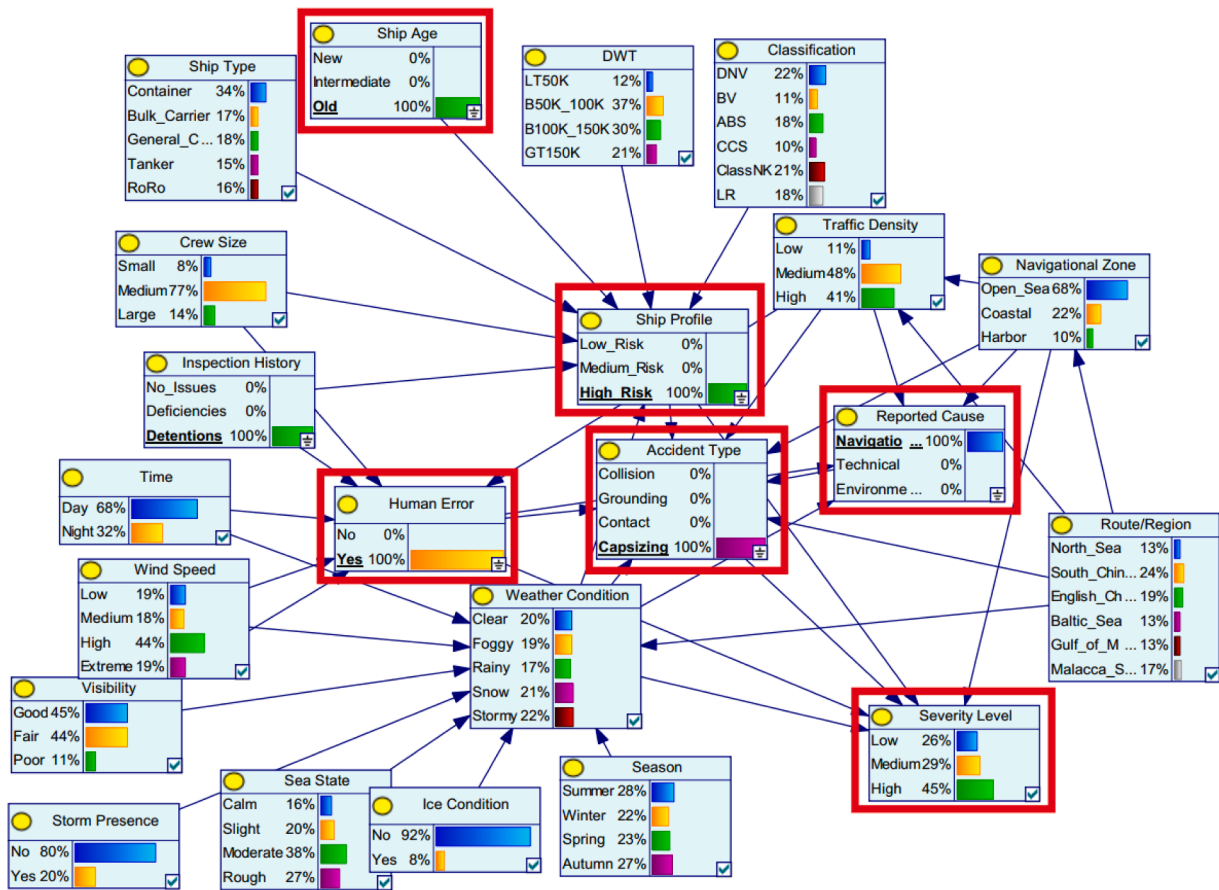


Fig. 7. Results for scenario 3 with reported cause Navigational.

[111], but by ice-related mechanisms that the operational and investigation literature has repeatedly identified as governing winter navigation consequences in the Baltic. These include: added hull-ice resistance and reduced rudder/propeller effectiveness degrading track-keeping [112–114]; wind-driven drift and ridging of the ice field, which in average and severe winters can move ice at 20–30 km/day under stormy conditions and generate pressure and ridges in key approach channels [115,112,113]; snow reducing perceptual acuity and sensor performance on the bridge; and the prolonged response, salvage and pollution-containment times that ice imposes on icebreaker, tug, and spill-response assets [10,109,116]. These mechanisms jointly explain why an environmental-cause grounding in ice carries a materially higher conditional severity than the equivalent grounding in temperate waters, and why the Baltic is the sea area that the BN selects ($\approx 100\%$) as the most probable region when this evidence bundle is entered. Second, the environmental reported cause, as opposed to machinery or purely navigational error, tilts attribution toward exogenous drivers (ice, visibility, sea state), which in this network co-vary with higher wind (High/Extreme $\approx 63\%$) and rough sea state ($\sim 25\%$). That combination increases leeway and drift, making off-track excursions harder to arrest once initiated, a mechanism echoed in European casualty syntheses that link winter weather and restricted maneuverability to serious outcomes in traffic-dense approaches [10].

The regional focus on the Baltic is not incidental: ice seasons, archipelagic fairways, and narrow approaches create a geometry where small control losses more readily convert into groundings, and oil- or cargo-handling consequences can be outsized. Recent hazard appraisals for the Baltic specifically single out harsh ice conditions as a primary accident trigger and highlight the persistence of winter-related risks despite overall safety improvements [10,109]. The posterior also retains Human Error at $\sim 51\%$, consistent with large reviews that, even in

adverse environments, human/organizational factors remain prominent contributors, suggesting that winter risk management must address both environmental exposure and human-system performance (procedures for convoy/ice-pilotage, speed/route adaptation, bridge resource management in low-visibility) [117,118].

This posterior should be read alongside the Baltic-specific winter navigation literature analyzing Finnish accident data from four winter periods and finds that winter navigation accidents in the FSWNS are most often collisions involving smaller general-cargo and feeder vessels (1000–10,000 DWT) under consolidated ice, with severity classified as 'less serious' under the EMSA scheme [43]. Another study proposes an FSA-Bayesian Network risk model for these same operations and identifies ship-independent navigation and convoys as the regimes with the highest oil-spill probability [44]. Our Scenario 4 is complementary to, rather than in conflict with, that evidence base, for three reasons. First, the severity scales are different: our states merge life and financial bands (see Table 2A), so a direct numerical comparison of percentages across schemes is not meaningful. Second, the figures in the Aalto-based studies are unconditional within the FSWNS, whereas our posterior is conditioned on a narrow, deliberately stressed bundle (environmental-cause grounding under ice and snow), which isolates a sub-population that the Aalto group itself identifies as especially consequence-laden because of hull, pollution, and salvage exposure. Third, their dominance of collision over grounding in the FSWNS is consistent with our own dataset when we condition on the same accident type: our BN places coastal collisions under winter conditions predominantly in the Medium band, in line with their findings [43]. The Scenario 4 elevation of High severity is therefore driven by the grounding evidence and the environmental-cause framing, not by a general claim that all Baltic winter accidents are high-severity. On vessel size, the LT50K DWT band in our model encompasses the Baltic winter fleet described by Valdez

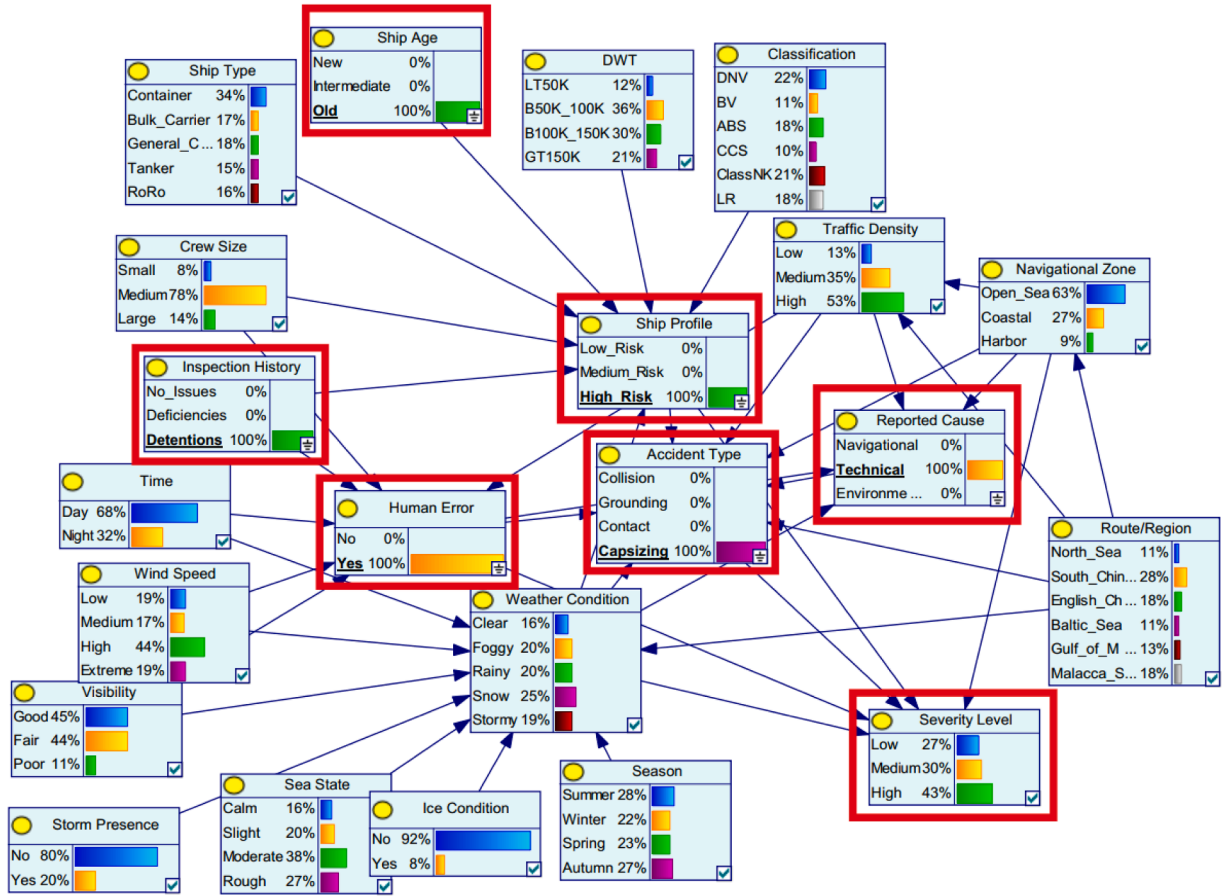


Fig. 8. Results for scenario 3 with reported cause Technical.

Banda et al. [43, Fig. 6], but without the intra-band resolution needed to reproduce their finer size distribution; we therefore do not use the DWT posterior in Scenario 4 as a quantitative claim about vessel size in the FSWNS, and defer to Valdez Banda et al. [43] for that characterization [43]. The FSWNS-specific mitigations identified that strengthened ship-independent navigation training, convoy/escort procedures, and ice-class adequacy remain the primary levers at the operational level [44]. Our global BN complements this evidence by situating the Baltic winter grounding pathway within a cross-regional severity framework that allows seasonal, environmental, and compliance signals to be compared across sea areas.

Comparatively, scenarios dominated by technical causes (e.g., machinery failure in temperate waters) in this study's network tend to produce less extreme right-tail shifts unless compounded by verification or access constraints. Here, by contrast, environment-driven grounding in ice pushes the posterior sharply toward High severity, reflecting (i) higher impact energies at drift groundings, (ii) slower assistance and salvage under ice, and (iii) escalation potential from hull breaches and pollution in sensitive waters, patterns consistent with European investigations and winter-navigation risk models [10,109].

For Baltic winter operations, the network underlines three levers: exposure reduction (seasonal routing and dynamic no-go areas tied to ice charts and visibility nowcasts), capability uplift (ice-class adequacy, escort/ice-pilotage, enhanced situational awareness when snowing), and response readiness (pre-arranged towage/salvage and spill resources sized for ice). These recommendations mirror the emphasis on ice-season specific procedures and targeted resource pre-positioning, and they explain why comparable environmental scenarios without ice (e.g., heavy rain but open water) in the model exhibit less severe posterior shifts [10,109]. Two interpretive caveats are important. First, the

43 %/41 %/17 % posterior is a conditional distribution over severity given the specific evidence bundle {Ice=Yes, Season=Winter, Weather=Snow, Grounding, Environmental}; it is not a statement about the unconditional severity distribution of all Baltic winter accidents, which is substantially closer to the global baseline. Second, in our coding High severity does not correspond to ship sinking or total loss; it is defined operationally by the consequence thresholds in Table 2A and therefore captures the broader population of serious-consequence casualties (multi-fatality events, major pollution and property loss, complex salvage, and structural damage) for which HELCOM and EMSA both document a persistent presence in Baltic winter operations.

Overall, the scenario demonstrates that winter environmental drivers, not merely traffic or profile, are decisive in pushing consequences into the high-severity regime in the Baltic, and it provides a quantitative, causally coherent explanation that accords with recent regional evidence.

4.5. Sensitivity analysis

4.5.1. Accident severity

Sensitivity analysis tests how much the posterior of a target node moves when its predictors are perturbed across their states. In a BN this provides two complementary payoffs: (i) prioritization, by ranking variables that exert the largest marginal pull on the target distribution; and (ii) explanation, by revealing which states (e.g., Open Sea, High risk profile, Stormy) drive those swings. The color-coded influence map Fig. 11 offers a global view, darker nodes are more sensitive, while tornado charts resolve the direction of change for specific states of the severity node (High/Medium/Low).

Across the global map, accident severity is most sensitive to the

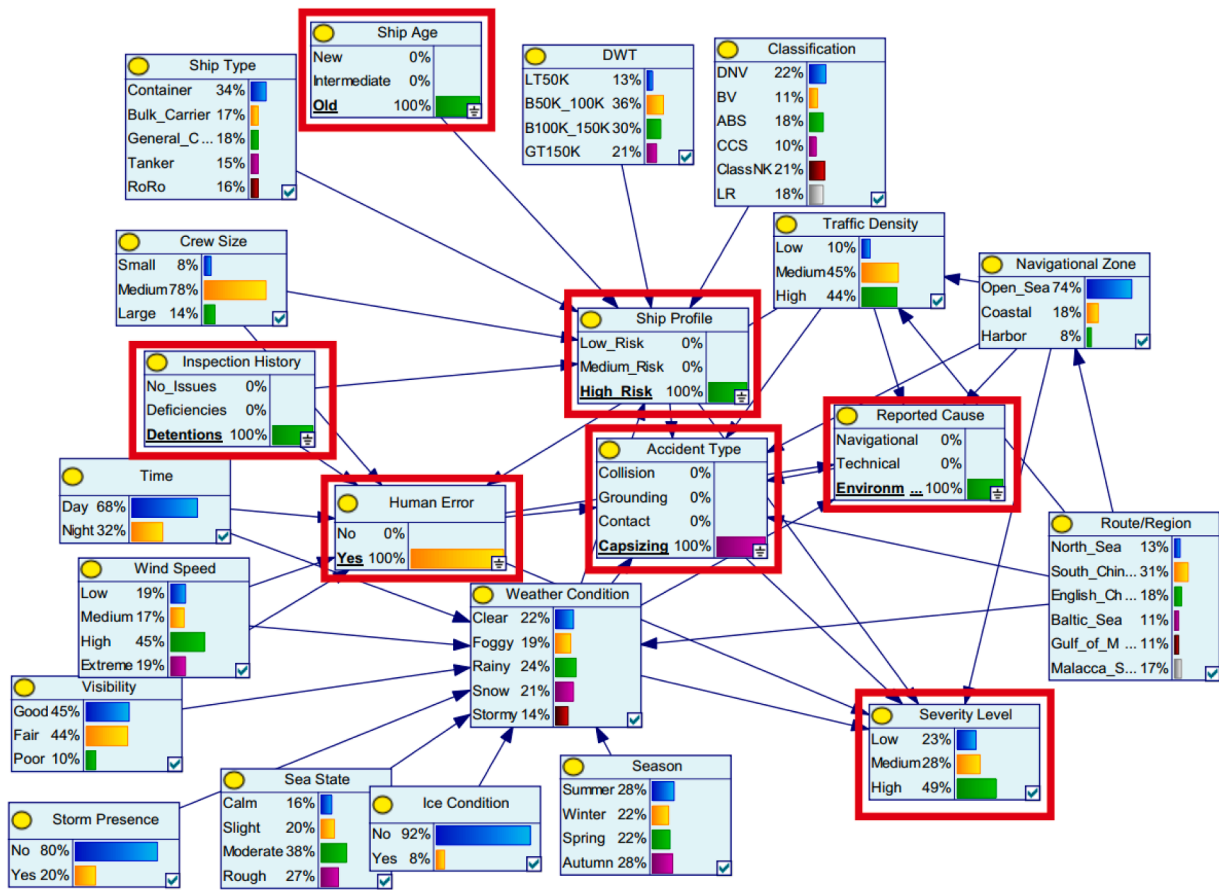


Fig. 9. Results for scenario 3 with reported cause Environmental.

navigation/route context (Navigational Zone; Route/Region), immediate causal framing (Accident Type; Reported Cause), and operational condition of the ship (Ship Profile; Human Error). Marine environment variables (Weather Condition, Wind Speed, Visibility, Sea State, Storm Presence) follow closely, while traffic density and inspection history provide additional modulation. This ordering is consistent with an operational narrative in which geometry and geography set exposure, proximate causes set the failure mode, and crew/ship condition and weather determine how far events escalate.

The tornado charts refine this picture, the multiple tornado charts are best summarized in Table 7 for better readability, while the tornado charts are presented in Appendix A. For High severity, the dominant upward shifts arise in open-water settings on high-exposure routes (e.g., South China Sea, Malacca Strait, English Channel), when *Grounding* or *Capsizing* is the accident type, the ship is assessed *High Risk*, *Human Error* = *Yes* is present, and the environment is adverse (Stormy/Rainy/Foggy, Poor visibility, Rough sea state, High-Extreme wind). Conversely, moving the locale to *Harbor/Coastal* and improving conditions (Clear weather, Good visibility, Calmer sea state) produce the largest downward swings. Medium severity is most elastic to *Collision* under deteriorated but not extreme conditions (Rainy/Foggy; Moderate-High sea state) combined with *Medium Risk* profile; both harsher weather and better visibility tend to push probability mass out of Medium, toward High or Low respectively. Low severity is maximized by *Collision* in clear/benign environments with *Low Risk* profile and no human error; it collapses under the same open-sea/adverse-weather combinations that inflate High severity.

These patterns have clear dynamics. Open sea and choke-point routes increase kinetic energy and maneuvering distance requirements, so identical initiating events (e.g., steering loss) more readily escalate, especially under compromised visibility or sea state.

Grounding and capsizing intrinsically couple to environmental forcings (bathymetry, lateral loads, stability margins), explaining their strong pull toward High severity when paired with rough seas or storms. The importance of *Ship Profile* and *Human Error* reflects accumulated latent conditions, maintenance, crew proficiency, bridge resource management, that shift the ease with which an incident can be contained once initiated. Finally, the directionality apparent in the tornado charts is policy-relevant: some levers (e.g., route/locale, weather windows, tug/escort allocation) systematically *decrease* High-severity propensity, while others (e.g., risky cargo/mission under poor visibility) systematically *increase* it.

In terms of positioning against recent empirical work, the dominance of geography and environmental stressors in the High-severity tail aligns with spatial hotspot and met-ocean studies that associate the South China Sea, Malacca Strait, and English Channel with elevated consequence potential under heavy traffic and adverse weather. Severity amplification by human-/organizational factors underlines findings that crew error and management quality remain cross-cutting accelerants even when proximate causes are “technical.” The pronounced sensitivity of Medium severity to *Collision* in marginal weather echoes observations that many contact/collision events cluster in intermediate conditions that are sufficiently challenging to induce error but not catastrophic in their own right; escalation beyond Medium then depends on whether secondary controls, maneuvering room, tugs, visibility, break favorably.

4.5.2. Accident type

Sensitivity analysis identifies which upstream conditions most change the posterior probability of each accident type (Collision, Grounding, Contact, Capsizing). The colored network view highlights global salience (deeper red = greater influence on accident type overall) is presented in Fig. 12, while the tornado plots decompose, for each focal

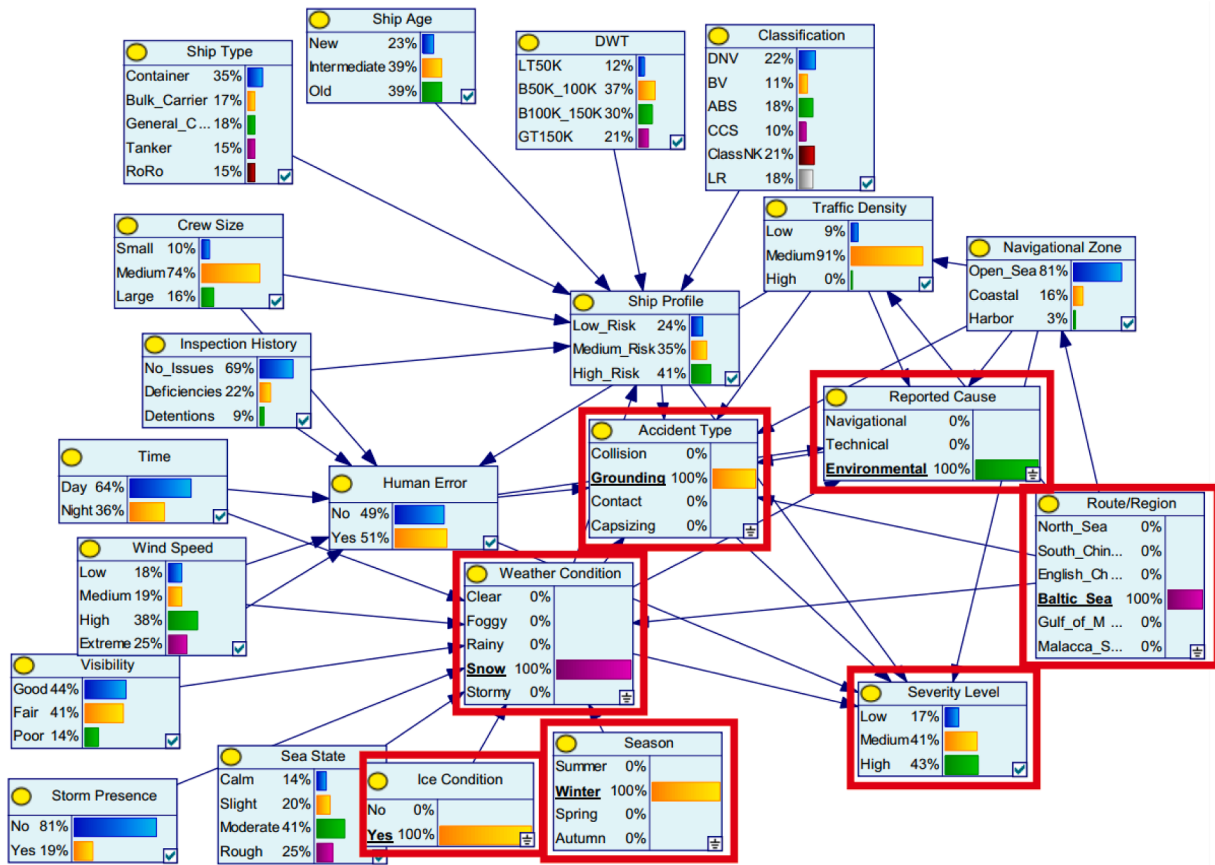


Fig. 10. Results for scenario 4 of seasonal and regional stressors.

accident type, the single and combined factor states that shift its probability the most, positively (red bars) or negatively (green bars). The tornado charts are included in Appendix, while their derived simple results are presented in Table 8. This analysis is complementary to predictive metrics: it explains *why* the model changes its belief and which controllable or contextual factors matter most.

Three robust themes emerge. (i) Spatial context dominates first order influence. Across accident types, the pair *Navigational_Zone* = *Open Sea* combined with specific *Route/Region* (notably South China Sea, Malacca Strait, and, for some rankings, Gulf of Mexico, English Channel, Baltic Sea) consistently appears among the top levers. This aligns with traffic geometry and encounter complexity documented for these corridors and reflects the model's ability to translate hotspot evidence into probabilistic influence.

(ii) Exposure amplifiers—traffic and weather—shift type propensities. High *Traffic_Density* repeatedly increases Collision and Contact likelihoods; adverse *Weather_Condition* (rainy, stormy, foggy) interacts with traffic and zone to push the model toward Collision or Grounding, whereas ice present/no-ice distinctions and season mainly modulate Grounding and Capsizing rankings.

(iii) Human/organizational factors shape pathways rather than acting alone. *Human_Error* = Yes and *Reported_Cause* = Navigational or Environmental seldom appear in isolation at the top of a tornado; instead, they co-occur with *Open Sea* + *high traffic* + *specific weather*, indicating that operator and cause signals are most consequential when coupled to demanding operating contexts.

The strongest movers in case of collision combine *Open Sea* with South China Sea or Malacca Strait, often under high traffic. When *Reported_Cause* = Navigational and *Human_Error* = Yes are present with rainy/clear conditions, the model's belief in Collision increases further, indicating that navigational misjudgment in dense, open-water

shipping lanes remains a primary pathway. The presence of ice = No ranks among influential negatives/positives depending on the bundle, suggesting Collision risk is less ice-driven than Grounding/Capsizing and more traffic-driven. For grounding the top entries are dominated by *Reported_Cause* = Navigational or Environmental paired with *Traffic_Density* = High, *Human_Error* = Yes, and *Open Sea* under rainy or clear weather. This pattern reflects classic loss-of-positional-awareness mechanisms (navigational error) and environmentally forced deviations (set/drift, hydrometeorological forcing) that lead to under-keel clearance exceedance, particularly in fast, congested corridors. Route effects remain material (South China Sea, Gulf of Mexico, North Sea), but the presence of Environmental cause elevates Grounding disproportionately relative to other types. For contact, the leading bundles again couple *Open Sea* + *Malacca Strait/South China Sea* with high traffic, pointing to close-quarters maneuvers and low-energy impacts in crowded fairways. Compared with Collision, Contact rankings include more cases with low/medium traffic as counter-levers, indicating a steeper sensitivity to density changes (i.e., Contact probability falls faster when density eases). Time-of-day and storm presence play secondary roles relative to spatial layout. In case of capsizing, *Open Sea* + *South China Sea/Malacca Strait* tops the list, with no ice frequently appearing, consistent with the dataset's mix where capsizing mechanisms are more tied to stability/exposure than ice accretion. Daytime appears among salient states in several items, suggesting that capsize events in the sample often occur during active operations rather than lay-by. Weather effects appear but are less dominant than for Grounding; capsizing propensity seems to hinge on hull/operations and sea-keeping exposure rather than visibility alone.

Recent empirical evidence is consistent with the dominance of traffic geometry, operating locale, and human/organizational factors observed in this model's sensitivity structure. European surveillance continues to

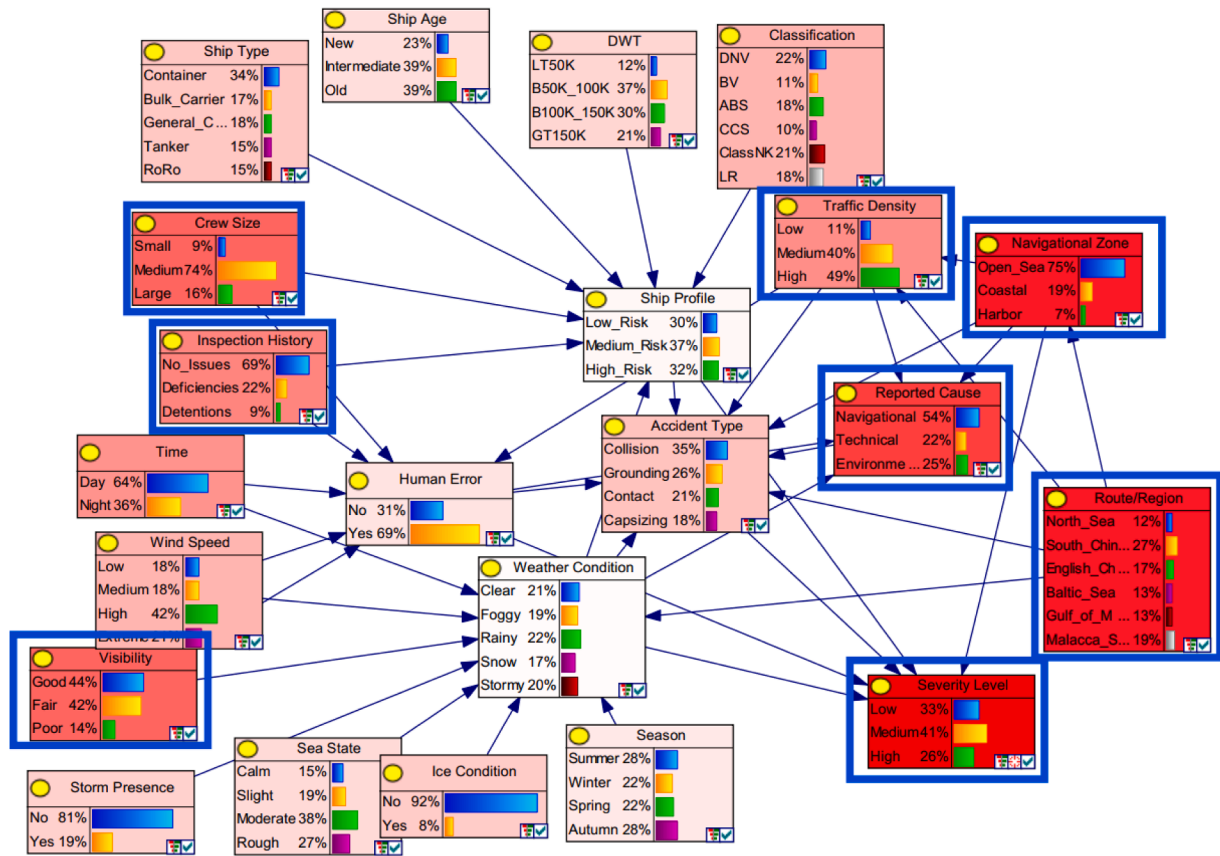


Fig. 11. Sensitivity analysis network view for the target node of Severity Level.

Table 7

Top drivers of accident severity by target state (derived from tornado charts).

Severity target	Rank 1 (↑ increases / ↓ decreases)	Rank 2	Rank 3	Rank 4	Rank 5
High	↑ Navigational Zone = Open Sea ∧ Route/Region = South China/Malacca/English Ch. ↓ Harbor/Coastal locale	↑ Accident Type = Grounding / Capsizing ↓ Accident Type = Collision/Contact	↑ Ship Profile = High Risk ↓ Ship Profile = Low Risk	↑ Human Error = Yes ↓ Human Error = No	↑ Adverse environment (Stormy/Rainy/Foggy; Poor visibility; Rough sea; High-Extreme wind) ↓ Clear weather; Good visibility; Calm-Slight sea; Low wind
Medium	↑ Accident Type = Collision under marginal weather (Rainy/Foggy)	↑ Ship Profile = Medium Risk	↑ Open Sea with Moderate sea/wind	↓ Clear weather & Good visibility	↓ Stormy/Rough pushing mass toward High
Low	↑ Collision in clear/benign conditions ↓ Open Sea on high-exposure routes	↑ Ship Profile = Low Risk ↓ Grounding/Capsizing	↑ Human Error = No ↓ High Risk profile	↑ Harbor/Coastal locale ↓ Human Error = Yes	↑ Good visibility; Calm-Slight sea; Low wind ↓ Stormy/rough/poor-visibility regimes

Notes: Entries summarize the top shifts visible in the tornado charts. “↑/↓” indicates whether the factor/state raises or lowers the probability of the specified severity outcome.

locate casualty frequency and consequence in traffic-dense sea areas and port approaches, e.g., the North Sea/English Channel and major Asian hubs, where constrained waterways and high encounter rates amplify collision/allision risk; EMSA’s latest Annual Overview documents these spatial patterns and the persistence of serious injuries and fatalities despite long-run safety gains [10,109]. In parallel, global insurer loss reviews report that machinery damage/failure remains the largest incident category worldwide while fire/explosion has surged to a decade high, trends that raise the baseline of severity conditional on an occurrence and that are mirrored by this model’s elevated sensitivity to propulsion/technical precursors and reported cause [119]. The state-wise influence of navigational zone (open sea vs. coastal/harbor) and regional routes (e.g., South China Sea, Malacca Strait) aligns with traffic-density findings in operational analytics: higher-density lanes

and chokepoints exhibit elevated collision proneness and narrower avoidance margins, which is precisely the pathway the tornado plots reflect when traffic density, open-sea navigation, and specific routes co-move with collision risk. Moreover, the model’s winter/ice and poor-visibility amplification for grounding is consistent with regional evidence that winter operations and ice accretion in northern waters materially increase the probability of loss of maneuvering margins and seabed contact, an effect highlighted in EMSA’s seasonal breakdowns and incident narratives.

On causation mix, the elevated contribution of human error and “navigational” reported causes to collision severity comports with earlier BN analyses that found human/organizational nodes to be proximal in collision chains, even when technical factors are present upstream [120]. At the same time, the model’s sensitivity to

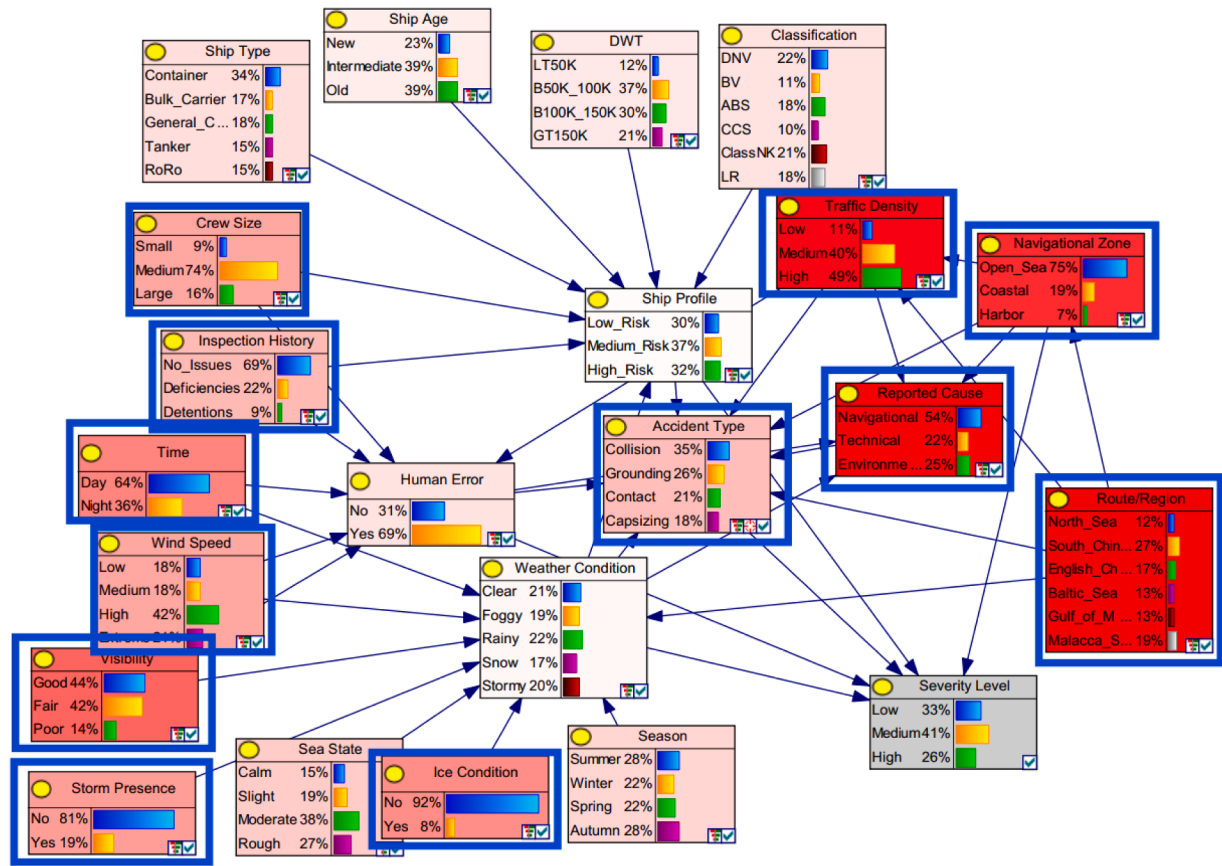


Fig. 12. Sensitivity Analysis network view for the target node of accident type.

Table 8

Top drivers from tornado charts (ranked 1–5 per accident type–States as per model).

Accident type	Rank 1 (↑ increases / ↓ decreases)	Rank 2	Rank 3	Rank 4	Rank 5
Collision	↑ Navigational Zone = Open Sea ∧ Route/Region = South China Sea	↑ Reported Cause = Navigational ∧ Traffic Density = High ∧ Human Error = Yes ∧ Weather = Rainy	↑ Navigational Zone = Open Sea ∧ Route/Region = Malacca Strait	↑ Reported Cause = Navigational ∧ Traffic Density = High ∧ Human Error = Yes ∧ Weather = Clear	↑ Navigational Zone = Open Sea ∧ Route/Region = South China Sea
	↓ Harbor/Coastal locale	↓ Traffic Density = Low / Medium ∧ Human Error = No ∧ Clear/Benign weather	↓ Route/Region = Gulf of Mexico / low-exposure routes	↓ Poor visibility / Stormy ∧ Rough sea (shifts mass toward Grounding/Capsizing)	↓ Route/Region = English Channel (relative)
Grounding	↑ Reported Cause = Navigational ∧ Traffic Density = High ∧ Human Error = Yes ∧ Weather = Rainy ∧ Open Sea	↑ Reported Cause = Environmental ∧ Traffic Density = High ∧ Human Error = Yes ∧ Weather = Rainy ∧ Open Sea	↑ Reported Cause = Navigational ∧ Traffic Density = High ∧ Human Error = Yes ∧ Weather = Clear ∧ Open Sea	↑ Reported Cause = Environmental ∧ Traffic Density = High ∧ Human Error = Yes ∧ Weather = Clear ∧ Open Sea	↑ Route/Region = South China Sea
	↓ Harbor/Coastal locale ∧ Traffic Density = Low ∧ Human Error = No ∧ Good visibility	↓ Calm/Slight sea ∧ Low wind	↓ Ice Conditions = Yes (ice-aware channels, pilotage)	↓ Route/Region = Gulf of Mexico / Baltic (relative)	↓ Navigational Zone = Coastal
Contact	↑ Navigational Zone = Open Sea ∧ Route/Region = Malacca Strait	↑ Navigational Zone = Open Sea ∧ Route/Region = South China Sea	↑ Traffic Density = High ∧ Navigational Zone = Open Sea ∧ Route/Region = Malacca Strait	↑ Route/Region = Malacca Strait (geometry + density)	↑ Navigational Zone = Open Sea ∧ Route/Region = Gulf of Mexico
	↓ Harbor/Coastal locale	↓ Traffic Density = Low	↓ Human Error = No	↓ Daylight/Good visibility (relative)	↓ Route/Region = English Channel (relative)
Capsizing	↑ Navigational Zone = Open Sea ∧ Route/Region = South China Sea	↑ Traffic Density = High	↑ Navigational Zone = Open Sea ∧ Route/Region = Malacca Strait	↑ Ice Conditions = No (open-water exposure); Time = Day (operational intensity)	↑ Navigational Zone = Open Sea ∧ Route/Region = Baltic Sea
	↓ Harbor/Coastal locale	↓ Traffic Density = Low / Medium	↓ Human Error = No	↓ Storm Presence = No (when combined with benign weather and pilotage)	↓ Route/Region = English Channel (relative)

environmental causes in grounding and capsizing tracks post-incident investigations of weather-exposed casualties and underscores the interaction between met-ocean stressors and ship profile that recent reviews recommend encoding explicitly in probabilistic frameworks [9, 64,120]. For prevention, sensitivity rankings suggest targeting high-leverage bundles rather than single variables: traffic management and speed control in the South China Sea/Malacca Strait corridors for Collision/Contact; navigational watch reinforcement, real-time environmental set-drift aids, and under-keel clearance management for Grounding in the same corridors and in Gulf of Mexico/North Sea approaches. For situational risk assessment, combining route, zone, and live traffic/meteorology will provide larger marginal gains than fleet-wide generic advisories. For training and audit, interventions that reduce navigational human error under high density seem to shift multiple accident types simultaneously in a favorable direction.

4.6. Strength of influence analysis

This study conducts strength of influence analysis using the GeNIe interface, and the results are presented in Table 9. The prominent trends observed across diverse drivers are discussed as follows.

Geography → exposure → outcomes. The largest influence in the network is the link from *Route/Region* to *Traffic Density* ($SoI_{wt} \approx 0.47$), confirming that where a ship operates strongly governs its background exposure to encounters and collision opportunities. This geographic channel then cascades: *Traffic Density* carries a substantial influence on *Reported Cause* (≈ 0.21), indicating that dense traffic elevates the likelihood that proximate causes are recorded as navigational/operational. In parallel, *Navigational Zone* (open sea vs. coastal/harbor) pushes both *Reported Cause* (≈ 0.21) and *Severity* (≈ 0.17), reflecting the higher kinetic energy, constrained maneuvering time, and reduced external support typical of exposed waters. Together, these edges articulate a coherent spatial narrative: sea area and corridor choice set the baseline risk climate, which then expresses through traffic, cause classification, and ultimately severity.

Weather as proximate escalator of consequences and causation. *Weather Condition* exerts a dual role: it directly shapes *Reported Cause* (≈ 0.23) and contributes to *Severity Level* (≈ 0.15), while also influencing *Accident Type* (≈ 0.05). This pattern is consistent with observed accident forensics where adverse weather not only precipitates events (e.g., loss of control, poor visibility) but also exacerbates outcomes (longer response times, higher energy impacts, secondary failures). Upstream, *Ice Condition*, *Season*, *Sea State*, *Wind Speed*, *Visibility*, *Time* (day/night), and even *Storm Presence* collectively feed *Weather Condition* with modest but coherent influences (≈ 0.04 – 0.05 each), indicating that the model captures meteorological granularity rather than treating weather as a simple binary stressor. Practically, this supports weather-aware routing and resource pre-positioning: small shifts across these sub-factors can move the system into substantially riskier regimes.

Human factors as a central mediator to both cause and severity. *Human Error* strongly shapes *Reported Cause* (≈ 0.196) and contributes materially to *Severity* (≈ 0.167), while also nudging *Accident Type* (≈ 0.045). Importantly, *Human Error* itself is sensitive to *Visibility* (≈ 0.149), *Time* (≈ 0.141), *Traffic Density* (≈ 0.107), *Wind Speed* (≈ 0.104), *Crew Size* (≈ 0.121), and *Inspection History* (≈ 0.108). This reveals a plausible operational mechanism: cognitively demanding contexts (poor visibility, nighttime, congestion, higher winds) and thinner crewing are associated with more human-performance lapses, which in turn tilt cause assignment and elevate severity. The non-trivial effect from *Inspection History* to *Human Error* suggests lingering organizational/maintenance deficits correlate with subsequent operational mistakes, consistent with the idea that safety culture and systems readiness shape attention, coordination, and workload.

Accident type and ship profile anchor the severity channel. *Accident Type* is a major direct driver of *Severity* (≈ 0.196), with *Ship Profile* (≈ 0.157) and *Weather Condition* (≈ 0.151) providing additional

Table 9

Strength of influence analysis with average, maximum and weighted.

Parent	→	Child	SoI (Average)	SoI (Maximum)	SoI (Weighted)
Route/Region	→	Traffic Density	0.471235	0.820216	0.471235
Weather Condition	→	Reported Cause	0.228995	0.655744	0.228995
Traffic Density	→	Reported Cause	0.210193	0.624897	0.210193
Navigational Zone	→	Reported Cause	0.210106	0.585314	0.210106
Accident Type	→	Severity Level	0.196488	0.775134	0.196488
Human Error	→	Reported Cause	0.195678	0.500000	0.195678
Navigational Zone	→	Severity Level	0.168924	0.734343	0.168924
Human Error	→	Severity Level	0.166509	0.568112	0.166509
Ship Profile	→	Severity Level	0.157378	0.750000	0.157378
Weather Condition	→	Severity Level	0.150635	0.792032	0.150635
Visibility	→	Human Error	0.148534	0.666667	0.148534
Time	→	Human Error	0.141420	0.625000	0.141420
Crew Size	→	Human Error	0.121366	0.704545	0.121366
Inspection History	→	Human Error	0.108120	0.678571	0.108120
Traffic Density	→	Human Error	0.106805	0.687500	0.106805
Wind Speed	→	Human Error	0.104328	0.687500	0.104328
Navigational Zone	→	Traffic Density	0.097311	0.289614	0.097311
Route/Region	→	Accident Type	0.056978	0.708946	0.056978
Reported Cause	→	Accident Type	0.055962	0.854294	0.055962
Weather Condition	→	Accident Type	0.054352	0.708946	0.054352
Navigational Zone	→	Accident Type	0.049886	0.708946	0.049886
Ice Condition	→	Weather Condition	0.049711	0.586894	0.049711
Time	→	Weather Condition	0.048635	0.632456	0.048635
Visibility	→	Weather Condition	0.047869	0.709068	0.047869
Ship Profile	→	Accident Type	0.047491	0.708946	0.047491
Route/Region	→	Weather Condition	0.046927	0.666667	0.046927
Storm Presence	→	Weather Condition	0.046893	0.709068	0.046893
Season	→	Weather Condition	0.044609	0.660303	0.044609
Sea State	→	Weather Condition	0.043907	0.666667	0.043907
Wind Speed	→	Weather Condition	0.043550	0.709068	0.043550
Route/Region	→	Navigational Zone	0.056309	0.092424	0.056309
Human Error	→	Accident Type	0.045202	0.708946	0.045202
DWT	→	Ship Profile	0.028331	0.666667	0.028331
Ship Age	→	Ship Profile	0.028345	0.629153	0.028345
Ship Type	→	Ship Profile	0.028390	0.629153	0.028390
Classification	→	Ship Profile	0.028736	0.666667	0.028736
Weather Condition	→	Ship Profile	0.028944	0.708742	0.028944
Inspection History	→	Ship Profile	0.030457	0.666667	0.030457
Crew Size	→	Ship Profile	0.030864	0.629153	0.030864

pull. This ordering matches maritime evidence that certain event classes (e.g., grounding, capsizing, high-energy collisions) and particular vessel profiles (size, configuration, stability margins) are predisposed to higher-consequence outcomes, especially in difficult weather. Notably, *Ship Profile* itself aggregates influences from *DWT*, *Ship Age*, *Ship Type*, *Classification*, *Weather Condition*, *Crew Size*, and *Inspection History*, each modest alone (≈ 0.028 – 0.031) but meaningful in aggregate. This argues for interventions that combine profile-aware prevention (e.g., targeted training, stability management) with context-aware operations (e.g., routing, pilotage, tugs).

Causal articulation among reported cause and accident type.

The edge *Reported Cause* $\rightarrow$ *Accident Type* (≈ 0.056) is smaller than *Weather/Zone/Traffic* routes into *Cause*, suggesting that the model sees *Reported Cause* as a downstream label shaped by exposure and conditions, with limited direct push back onto *Accident Type*. This is desirable: “cause” classification often codifies the antecedent conditions (environmental, navigational, human), and the stronger levers should sit upstream.

These trends in the SoI results indicates that (i) Routing and scheduling in high-density corridors dominate upstream risk; altering route selection and timing (e.g., avoiding peak congestion/night transits under poor visibility) attacks the largest lever in the graph. (ii) Weather-aware operations (ice, wind, sea state, storms, visibility) are not just about occurrence, they modulate severity; embedding fine-grained weather indices into dispatch and pilotage is warranted. (iii) Human-systems readiness remains central: crewing levels, fatigue-aware watchkeeping, and bridge resource management can reduce the human-error mediator that drives both cause and severity. (iv) Profile-targeted safeguards (for vessel classes/types and age bands) and organizational follow-through after inspections can push the system toward safer states across multiple downstream channels.

Overall, the SoI landscape is internally consistent and policy-relevant: geography sets exposure, weather and human performance escalate and shape consequences, and accident class and vessel profile determine how far events run once initiated. This decomposition provides a defensible basis for prioritizing prevention and preparedness investments where they yield the largest severity reductions.

4.7. True risk influence (TRI)

TRI is a one-at-a-time sensitivity metric that summarizes how much the probability of a specified severity state changes when a single predictor is forced to its most adverse versus its most protective state. Operationally, TRI is obtained by (i) taking the baseline probability of the target severity state; (ii) setting the predictor to the state that maximizes that probability (“high-risk push”) and recording the increase; (iii) setting the predictor to the state that minimizes it (“low-risk pull”) and recording the decrease; and (iv) averaging these two contrasts to yield a single, direction-agnostic magnitude. In this study the TRI values for the target node of severity level are presented in Table 10 for the accident type variable.

Accident Type exhibits material one-at-a-time sensitivity on consequence, with TRI indicating the largest swing for the Low severity state (≈ 18 pp), and still non-trivial shifts for Medium (≈ 11.5 pp) and High (≈ 11 pp). The conditional distributions align with operational

expectations and provide face validity. Collision concentrates probability in Medium severity (48 %) with a subdued High tail (16 %), reflecting typically controllable energy exchange and effective emergency response in many collision contexts. Grounding shifts mass toward Medium–High (50 % and 34 %), consistent with hull breach, bottom contact, and potential pollution risk, which often demand complex salvage or refloat operations that amplify consequences. Contact (minor allisions/berthing strikes) skews toward Low (52 %), matching its lower energy profile and typically localized damage footprint. Capsizing yields the highest High share (38 %), consistent with loss-of-stability cascades, rapid onset, and life-safety jeopardy.

Strategically, these results justify focusing TRI on Accident Type: it is the most proximal, standardized descriptor through which upstream conditions (weather, traffic density, human error, route/region) channel into outcomes. The baseline-to-posterior shifts are also decision-useful. For instance, a forecast dominated by grounding exposure (e.g., winter, ice-prone routes) should trigger pre-positioned salvage capability and boom assets, whereas capsizing exposure prioritizes life-saving readiness and rapid muster/abandon protocols. Conversely, an operational profile dominated by contact incidents suggests that berth approach management, fendering, and tug allocation can meaningfully contain consequences. Overall, Accident Type’s TRI and the associated conditional profiles provide a compact, policy-relevant lens for preparedness planning, scenario stress-testing, and targeted risk communication without confounding multi-factor mediation already captured elsewhere in the analysis.

4.8. Benchmarking of findings (factor-coverage evidence)

The findings reported in Sections 4.3–4.7 can be benchmarked against Table 3, which synthesizes factor usage across 17 closely related BN-based maritime safety and severity studies. This benchmarking strengthens interpretation in two complementary ways: (i) it shows where the dominant drivers identified in the current study are consistent with “high-consensus” factors commonly modeled in the literature, and (ii) it highlights where the current study’s prominent pathways rely on variables that remain underrepresented in prior BN severity modeling, thereby evidencing concrete limitations of the existing literature and clarifying the added value of the present results.

4.8.1. Alignment with high-consensus factors in prediction, sensitivity, and TRI

A central outcome of the prediction analysis (Section 4.3) and the downstream sensitivity/TRI analyses (Sections 4.5 and 4.7) is that Accident Type anchors the consequence channel. Table 3 shows Accident Type is included in 16/17 benchmark studies, demonstrating near-universal recognition that mechanism-level categorization (collision/grounding/contact/capsizing, etc.) is indispensable for severity modeling e.g., [91,23,33,2,4]. The present results advance this consensus by moving beyond inclusion to quantification of leverage: the TRI results (Section 4.7) isolate Accident Type as the dominant one-at-a-time lever shifting probability mass toward the high-severity state, while the tornado-based rankings for Severity (Table 6) and Accident Type (Table 7) show that the “type channel” is repeatedly among the top movers across target states. In other words, Table 3 supports that Accident Type is widely treated as fundamental; Sections 4.5–4.7 add a sharper operational message by demonstrating its state-wise influence under multi-factor conditions.

Similarly, ship descriptors that repeatedly appear in the present study’s dominant pathways (Prediction/Sensitivity/SoI) are also widely represented in Table 3. Ship Type (17/17) and Ship Age (13/17) are among the most consistently modeled factors e.g., [12,87,33,2,4,13,1]. This aligns directly with the current study’s inference landscape, where ship-related vulnerability information materially conditions severity pathways, particularly when interacting with accident mechanism and operating context. Notably, the current study further consolidates these

Table 10

TRI (by severity state) for *Accident Type* and posterior severity distributions when each accident type is set to 100%.

Metric / Conditioning	Low (%)	Medium (%)	High (%)
Baseline (no evidence)	33	41	26
TRI (percentage-point swing)	18.0	11.5	11.0
Collision = 100%	36	48	16
Grounding = 100%	16	50	34
Contact = 100%	52	27	21
Capsizing = 100%	29	33	38

attributes through a composite Ship Profile (derived) node, which Table 3 shows is only used in 3/17 studies [90,91,23]. The present results, therefore, both conform to the established literature focus on ship descriptors and extend it by showing how an interpretable composite prior can act as a structural amplifier in severity escalation channels.

4.8.2. Geography and operating context: prominent in results, under-modeled in the literature

One of the most important comparative insights emerges from the geography/context findings. Across prediction (Section 4.3), sensitivity (Section 4.5), and strength-of-influence (Section 4.6), geography and operating context repeatedly appear as first-order drivers, not merely background descriptors. However, Table 3 indicates that key geographic/context variables are not yet standard in BN severity modeling, particularly at the resolution used in the present work.

First, Table 3 shows that explicit Route/Region encoding appears in only 3/17 benchmark studies [3,89,14]. This scarcity is consequential when interpreted alongside the present results: Route/Region emerges as a dominant upstream driver for severity sensitivity (Section 4.5; Table 6) and also generates the largest influence edge in the network via the exposure pathway Route/Region → Traffic Density (Section 4.6; Table 8). The benchmarking, therefore, supports a strong and logical aspect: although regionalization is rarely operationalized in BN severity models, the present results show it is not optional. It structurally governs exposure and downstream consequence likelihood through corridor/geometry mechanisms.

Second, Table 3 shows Navigational Zone (port/coastal/open sea) appears in 10/17 studies [12,91,23,33,24,2,4,34,13,1]. This indicates the locale dimension is moderately represented, yet the present results demonstrate it is often the most sensitive single context variable for severity and accident-type shifts. The tornado summaries (Section 4.5; Tables 6–7) repeatedly place Navigational Zone at or near Rank 1 for multiple targets, and the scenario analysis (Section 4.4) shows that holding many factors constant while changing only mechanism/context combinations (e.g., coastal collision vs coastal grounding under high traffic) yields materially different severity distributions. In combination, the benchmarking suggests a refinement to the literature: locale is recognized by many studies, but the current results show that locale becomes most decision-relevant when jointly modeled with region and exposure, enabling the network to represent “where the accident happens” as both a severity driver and a mechanism shaper.

Third, the exposure proxy Traffic Density is included in only 5/17 benchmark studies [12,91,23,33,13], despite its operational salience. The present SoI results show Traffic Density is a major proximal driver on the severity channel and is structurally governed upstream by Route/Region. This creates a clear literature limitation supported by Table 3: even where traffic/exposure is acknowledged, it is not consistently integrated as a mediator linking geography to consequences, whereas the present study provides precisely that interpretable exposure mechanism.

4.8.3. Environmental and met-ocean factors: common in principle, but granularity gaps remain

Table 3 indicates that environmental factors are frequently included in some form, but with uneven granularity. Variables such as Visibility (12/17) and Wind Speed (11/17) are widely used e.g., [3,12,33,2,4,34,13,1], supporting the broader consensus that met-ocean stressors shape severity risk. However, Table 3 also reveals two important granularity gaps that become particularly visible when benchmarked against the present scenarios and sensitivity results.

First, Sea State appears in only 7/17 studies [12,91,23,2,4,31,1], and the explicit Storm Presence indicator appears in 0/17 benchmark studies. The present study’s scenario analysis (Section 4.4) demonstrates why this matters: simultaneous evidence-setting on stormy weather, poor visibility, and rough sea state produces distinct posterior shifts in accident mechanism propensity and severity distribution that cannot be

represented faithfully if storms are treated only as a coarse proxy or omitted entirely. Thus, even though “weather” is common in Table 3 (Weather Condition appears in 8/17 studies, including [3,91,23,24,2,4,34,1]), the present results provide logical and viable evidence that the number of environmental aspects and their state resolution materially change what the model can infer under compound conditions.

Second, Table 3 shows that Ice Condition appears in only 1/17 studies [88]. This rarity aligns with a major contribution of the current study’s seasonal/regional stressor scenario (Section 4.4), where ice/-winter conditions provide a physically coherent escalation pathway for grounding consequences. Benchmarking against Table 3 therefore, supports a precise statement: the literature rarely operationalizes ice despite its known relevance in certain sea areas; the present findings show that when included, ice-related evidence can decisively reshape severity posteriors in ways consistent with navigation constraints and emergency-response degradation in cold regions.

4.8.4. Compliance and assurance variables: prominent in current findings but largely absent in benchmark studies

A strong literature limitation supported directly by Table 3 concerns regulatory/compliance and assurance signals. Inspection History (PSC) is present in only 2/17 benchmark studies [90,34], and Classification (RO) is present in 0/17 studies. This benchmarking is especially powerful when interpreted with the present study’s “condition/compliance” gateway in prediction analysis (Section 4.3) and the stressed compliance/profile scenarios (Section 4.4). The current results show that compliance- and condition-related signals act as meaningful modifiers and amplifiers of severity pathways, particularly when they interact with ship vulnerability, human error, and accident mechanisms. Yet Table 3 demonstrates that such variables are largely outside the standard factor set used in BN severity models. Accordingly, the present findings can be framed as addressing a clear modeling blind spot: a substantial portion of the literature has focused on ship/environment/accident descriptors while leaving compliance and assurance signals under-integrated, limiting the ability to make policy-relevant inferences about inspection regimes and baseline control conditions.

4.8.5. Reported cause and human factors: confirming patterns while strengthening the evidence base

Table 3 indicates Human Error appears in 8/17 benchmark studies [12,91,24,2,4,31,13,1]. This supports the relevance of the present study’s strong human-factor mediation (Section 4.3) and SoI results (Section 4.6), which show human performance pathways contributing both to causal-class assignment and to downstream severity channels. In contrast, Reported Cause appears in only 1/17 benchmark studies [3], yet in the present results, it emerges as a prominent explanatory layer that helps discriminate how environmental vs navigational vs technical narratives modulate severity under otherwise similar evidence (Section 4.4, Scenario 3). The benchmarking, therefore, yields a crisp comparative conclusion: the literature often includes human error in some form, but rarely includes a separate reported-cause classification that can be used diagnostically to interpret severity pathways. Whereas the present results show that the reported cause operates as a meaningful modulator, especially under stressed ship/organization conditions.

Taken together, the benchmarking evidence in Table 3 strengthens the interpretation of Sections 4.3–4.7 in a logically consistent way. The current study’s dominant drivers, accident mechanism, and core ship descriptors align with the most commonly modeled factors in prior BN severity work, supporting external credibility of the main patterns e.g., [12,2,4,13,1]. At the same time, Table 3 provides clear and defensible support for novelty claims grounded in results: variables that prove structurally important in the present inference landscape (Route/Region; PSC outcomes; RO/classification; storm/ice resolution and compound met-ocean evidence) remain sparse or absent in benchmark models. This gap explains why the present study’s geography→exposure→outcome channel and compliance-conditioned

escalation pathways add substantive value: they are not merely “extra variables,” but underrepresented dimensions that materially change scenario inference, sensitivity rankings, and interpretable causal leverage for consequence-aware decision support.

4.8.6. Cause-and-effect interpretation of the severity escalation pathways

Taken together, the results indicate that severity escalation in the present BN follows a staged operational logic rather than a flat set of pairwise statistical dependencies. The first stage is exposure formation, in which Route/Region and Navigational Zone determine whether the vessel is operating in a corridor characterized by high encounter rates, constrained maneuvering margins, infrastructure proximity, or delayed access to external assistance. The second stage is loss of controllability and recoverability, where adverse weather conditions, poor Visibility, rough Sea State, stronger winds, and Human Error reduce the crew's ability to detect hazards early, maintain a safe position, and contain the initial disturbance. The third stage is consequence realization, where Accident Type converts these upstream conditions into different severity profiles. In this stage, Grounding and Capsizing are more escalation-prone because they typically prolong exposure, increase the likelihood of hull breach, flooding, stability loss, salvage complexity, and environmental damage, whereas Contact-type events are more likely to remain localized. Ship Profile and Inspection History act mainly as resilience modifiers in this sequence: they do not necessarily trigger the event themselves, but they influence how effectively the vessel can absorb disturbance once the accident has begun.

This interpretation also clarifies how the present findings extend the patterns reported in the existing literature. Previous severity-oriented BN studies have consistently identified accident Type, ship characteristics, and broad location as important predictors of severe consequences, while spatial studies have shown that dense traffic corridors and constrained approaches are recurrent maritime risk hotspots [12,4,13,14]. The present study goes beyond those broad factor rankings by resolving the escalation sequence more explicitly. In particular, the model suggests that regional geography affects severity primarily indirectly through Traffic Density and Navigational Zone; weather acts not only as an occurrence-related factor but also as a consequence amplifier; Human Error functions as a mediator between demanding operating conditions and accident realization; and compliance-related signals such as Inspection History condition the vessel's disturbance tolerance and recovery capacity. This explains why two cases with the same nominal reported cause often result in different consequence levels under different exposure, recoverability, and vessel-condition settings.

5. Implications

The results collectively point to a practical hierarchy for consequence reduction that can be operationalized. In consequence-informed planning, proximal drivers tied to accident type should anchor preparedness: operating profiles with higher exposure to grounding and capsizing merit pre-positioned salvage, towing, and life-saving capacity, while collision-dominated profiles benefit more from traffic management measures and escort/tug strategies near chokepoints. Using the BN in planning mode, setting credible evidence on route/region, navigational zone, weather/sea state, visibility, and traffic density, yields *state-specific playbooks* for equipment staging, muster plans, and notification thresholds. This moves risk management to consequence-conditioned readiness for the incidents the fleet is most likely to face on a given leg.

For voyage planning and routing, the model supports dynamic risk gating rather than blanket restrictions. When forecasts indicate storm presence, poor visibility, and rough sea on open-sea segments of high-density routes (e.g., South China Sea/Malacca/English Channel), the BN's scenario results justify advancing ETD/ETA windows, speed reductions, or controlled waiting to keep the ship within lower-severity regimes. Conversely, if transits must proceed, the outputs justify *graduated mitigations*, additional bridge resources, enhanced lookout,

pilotage/tug allocation near approaches, and pre-briefed contingency points, because these interventions specifically cut the pathways that shift probability mass into the medium/high severity states. At the port and VTS (Vessel Traffic Service) level, the sensitivity and influence patterns argue for adaptive traffic control where geometry and density amplify consequences. Practical levers include dynamic separation minima under marginal weather, targeted lane management for deep-draft or lower-maneuverability vessels, and priority pilotage for ships whose profile and recent inspection history suggest higher consequence potential. Publishing consequence-aware advisories (combining route/region, visibility/wind/sea state, and observed traffic density) provides a defensible basis for temporary restrictions that prioritize safety without unduly constraining throughput.

For fleet operations and training, the influence of human error and reported cause, conditioned by environment and route, supports context-specific drills (e.g., loss-of-propulsion in traffic, anchoring/grounding prevention under cross-winds, stability management under sea state) instead of generic, calendar-driven exercises. The model queries identify the combinations most associated with higher consequences; these can be translated into minimum voyage-specific briefings and bridge resource management (BRM) checklists that mirror the model's strongest pathways (e.g., navigation in open sea under poor visibility with high density). Crew sizing and experience distribution can also be tuned on legs whose configurations push the posterior toward higher severity.

The integration of Port State Control outcomes as a predictor, rather than a separate domain, has immediate compliance and audit implications. Where recent PSC history indicates detainable or clustered deficiencies, the model justifies incremental internal audits or class/RO engagement *before* entering traffic-dense sea areas or constrained approaches, because the benefit is realized not only in occurrence reduction but also in severity trimming if an incident happens. Administrations can likewise use the framework to target consequence-relevant inspections (navigation/propulsion/ISM focus) during seasons and on routes where the environment amplifies outcomes, improving the yield of limited inspection resources.

From an emergency preparedness perspective, the BN's conditional severity profiles support tiered readiness. When scenarios show a shift toward high severity (e.g., winter grounding exposure with ice/snow), ports and relevant authorities can elevate on-call salvage, medical, and pollution response status and pre-stage booms and dispersants. Where capsizing probability rises, drills and asset placement should emphasize life-saving response times and survivor recovery windows. Because the model quantifies the *marginal* severity effect of specific factor bundles, authorities can defend graduated posture levels linked to forecast and traffic intelligence rather than static seasonal assumptions.

Insurance and financing stakeholders can use the outputs to differentiate risk by leg and configuration, not only by asset class. Premium loadings or deductibles that reflect route/season/weather-conditioned consequence risk create incentives for shippers to adopt mitigating measures (e.g., pilotage in marginal weather, verified equipment status before high-exposure passages) that the BN shows to be most effective in shifting mass toward lower severity. Crucially, the analysis indicates that contextual amplifiers (route/region, environment, density) explain much of the variation in consequences, suggesting that behavioral credits for demonstrable mitigations are more defensible than coarse ratings based on flags or generic profiles alone.

For governance and data practice, the study underscores the value of standardized, multi-state environmental and locale descriptors in accident records (visibility bins, sea state categories, named sea regions and operating locale). This granularity materially improves consequence modelling and should be prioritized in reporting templates and AIS-meteorology data linkage. Likewise, maintaining PSC outcomes as structured fields alongside accident narratives enables joined-up safety analytics that bridge compliance and operational risk, rather than treating them as separate silos.

Finally, in deployment, the framework can be embedded in shore dashboards and voyage-planning tools to produce near-real-time consequence forecasts from routine feeds (AIS track, weather/sea state forecasts, route/region metadata, declared cargo/crew, recent PSC status). The same engine supports what-if exploration by masters, DPA/CSOs, pilots, and VTS (e.g., “If visibility drops to *poor* and wind rises to *high*, does delaying landfall two hours materially reduce expected severity?”). This encourages evidence-based choices that the model indicates will materially truncate consequence tails, the central objective for regulators, operators, and communities sharing congested waterways.

6. Conclusions

Maritime accident consequences remain unevenly distributed across ship types, routes, and operating contexts, yet most operational tools still emphasize occurrence rather than consequence. This study addressed that gap by developing and comparing consequence-focused BN models for accident severity, explicitly encoding geography (route/region and navigational zone), weather/sea state, traffic density, ship/human profile, inspection/compliance signals, and accident typologies. Three structures were estimated, TAN, ANB, and a literature/expert BN, with parameters learned by EM on a globally distributed, multi-year (2015–2024) dataset. Model reliability was assessed via stratified 10-fold cross-validation and Receiver Operating Characteristic/Area Under the Curve (ROC/AUC) analysis; the highest-performing model was then used for predictive, scenario, sensitivity, and strength-of-influence analyses.

Across diagnostic analysis, posterior pathways converged on a coherent causal narrative: open-sea operations on high-density corridors (South China Sea/Malacca/English Channel), under adverse weather (storm/poor visibility/rough sea/high winds), coupled with human error and high-risk ship profile, elevate the odds of *High* severity, particularly when the accident type is grounding or capsizing. Scenario experiments quantified these interactions: setting adverse environment + open sea + high traffic shifted mass from *Low/Medium* toward *High*, while benign conditions (clear weather, good visibility, calm-slight sea, low wind) and coastal/harbor locales reversed the shift. Adding navigational causes amplified *Medium/High* for collisions and grounding; environmental causes under winter/ice regimes particularly strengthened the grounding-high-severity channel. Sensitivity analysis (global color maps and state-wise tornado plots) ranked Accident Type as the dominant driver of severity, with environmental bundle (weather/visibility/sea state/wind), navigational zone/route-region, and human error as the next tier; ship profile moderated baselines but did not rival the top drivers once context was known. SoI results reinforced this ordering: route/region → traffic density and weather/visibility → severity links were strong, and human error exerted measurable influence both directly on severity and indirectly via reported cause and accident type. Finally, TRI isolated consequence leverage by state: grounding/capsizing drove the largest positive swings into *High*, collision concentrated mass in *Medium*, and contact redistributed toward *Low*.

These results matter operationally because they translate heterogeneous evidence into decision-quality, consequence-conditioned probabilities. The network’s structure and performance justify practical use in: (i) type-conditioned planning (stability/position-keeping and tow/

salvage readiness for grounding/capsizing legs; traffic-management for collision-prone legs); (ii) evidence-based routing/timing to avoid forecast combinations of open sea, high density, and marginal weather; (iii) VTS policies that adapt separation and sequencing under adverse forecasts; and (iv) crew resourcing and drills aligned to the highest-gain pathways (e.g., loss-of-control/grounding under poor visibility and rough sea). Because the BN supports posterior diagnostics and “what-if” queries, stakeholders can preview consequence shifts under controllable levers before committing to routing or scheduling decisions. More concretely, the framework can support at least three immediate operational uses. For regulators and PSC authorities, the identified high-severity pathways can be used to prioritize consequence-relevant inspections and audits, for example, by focusing navigation, propulsion, and ISM-related checks on vessels with poor inspection history, ageing/high-risk profiles, or planned transits through dense and weather-exposed routes. For VTS and port authorities, the same outputs can inform adaptive control measures such as temporary tightening of separation rules, prioritization of pilotage or tug assistance, and escalation of salvage, pollution, or medical readiness when the posterior shifts toward severe outcomes under poor visibility, rough sea state, winter/ice exposure, or high traffic density. For shipping companies, the scenario results can be translated into voyage-specific crew briefings, bridge resource management checklists, and targeted drills, for example, loss-of-propulsion response in traffic, grounding prevention under crosswinds or low visibility, and stability-focused drills for routes and seasons associated with elevated capsizing or grounding severity. In this sense, the model is not only descriptive, but can be used as a decision-support layer for consequence-aware inspection, traffic management, and crew-readiness planning. This study focuses on establishing and validating an interpretable, multi-factor BN framework for severity inference using harmonized accident-case records and discretized operational descriptors. As with any consequence model built from retrospective investigations, the level of detail that can be represented is bounded by the consistency and resolution of the underlying records across jurisdictions. Future work may explore incremental refinements as additional standardized data streams and deployment contexts become available.

CRedit authorship contribution statement

Rafi Ullah Khan: Writing – original draft, Visualization, Validation, Software, 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.

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant Agreement No. 864724).

Appendix A

Figures for Sensitivity Analysis Tornado Charts and Model Validation ROC/AUC of TAN and ANB

Figs. A1, A2, A3, A4, A5, A6, A7, A8, A9, A10, A11, A12 and Table A1

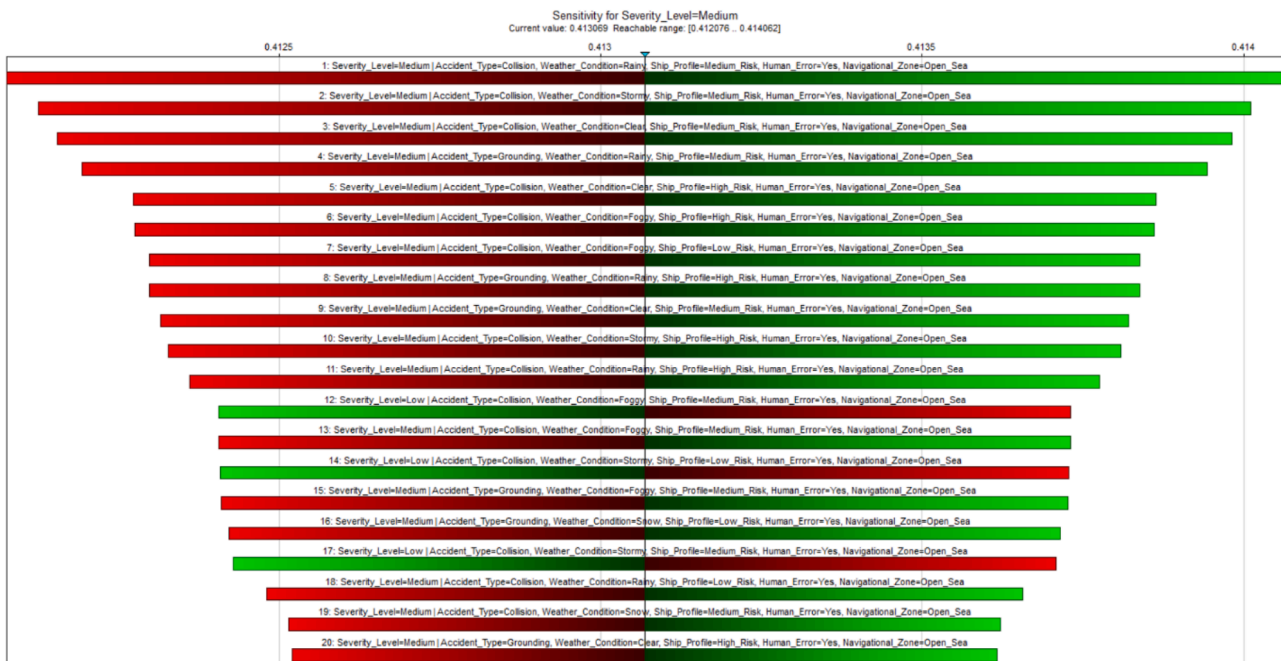


Fig. A1. Tornado chart for target node severity level-State= Medium.

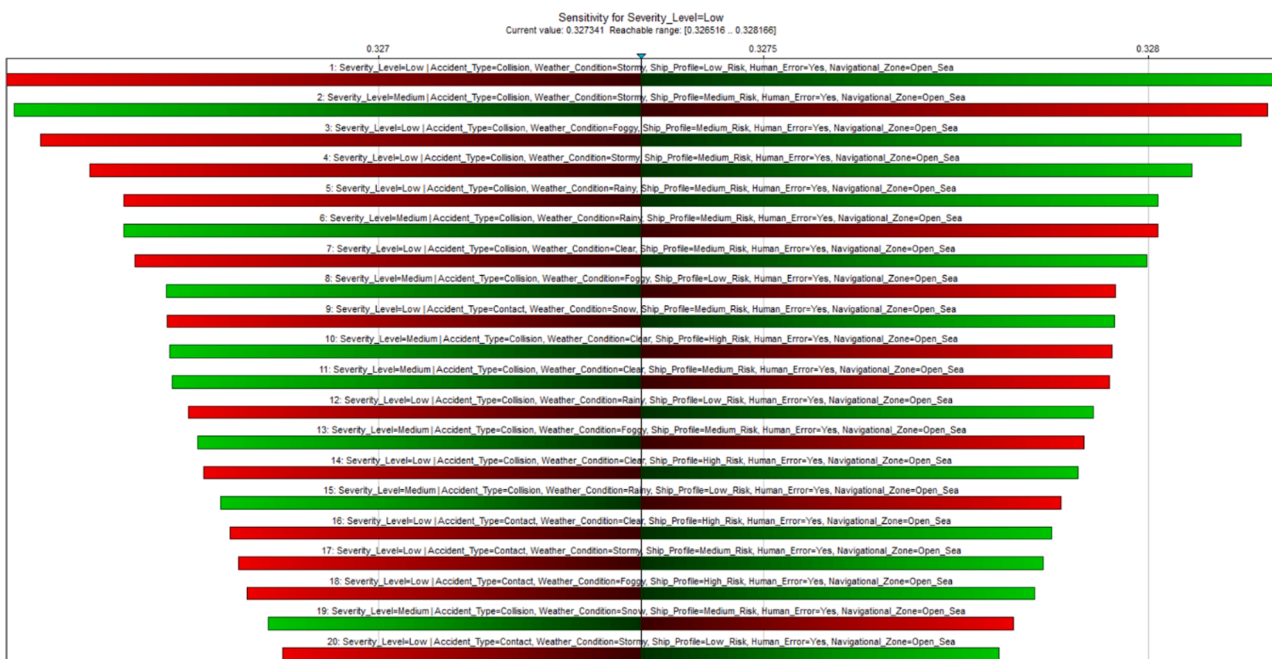


Fig. A2. Tornado chart for target node severity level-State= Low.

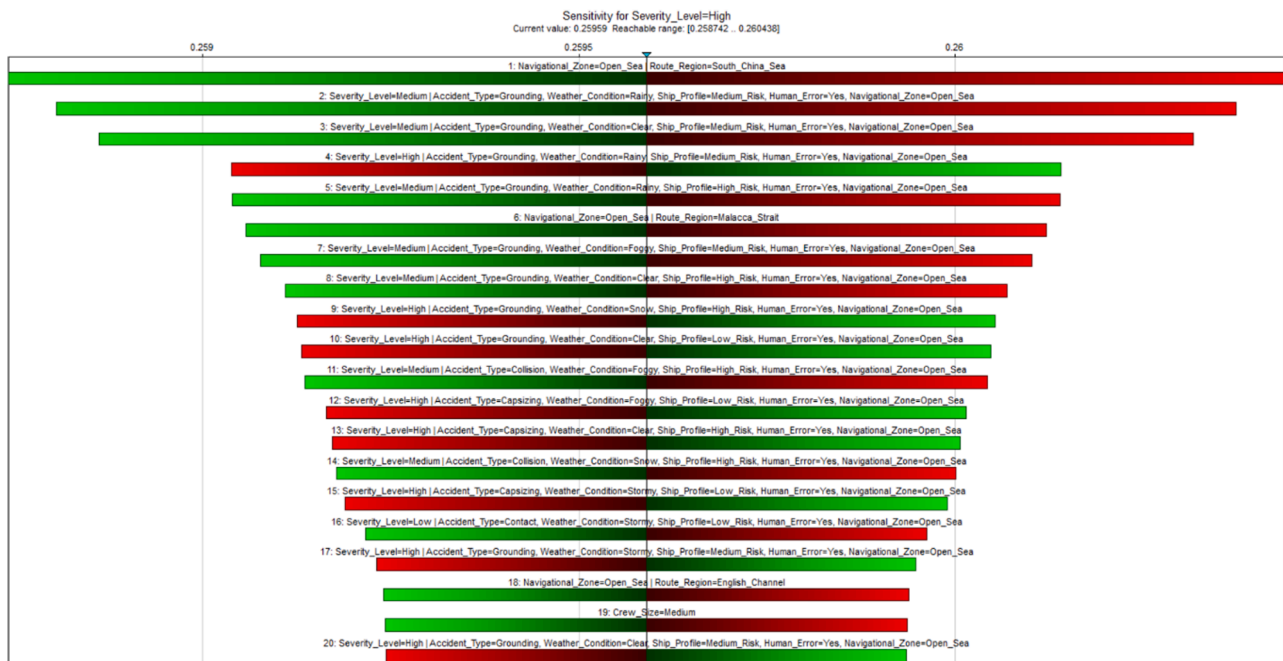


Fig. A3. Tornado chart for target node severity level-State= High.

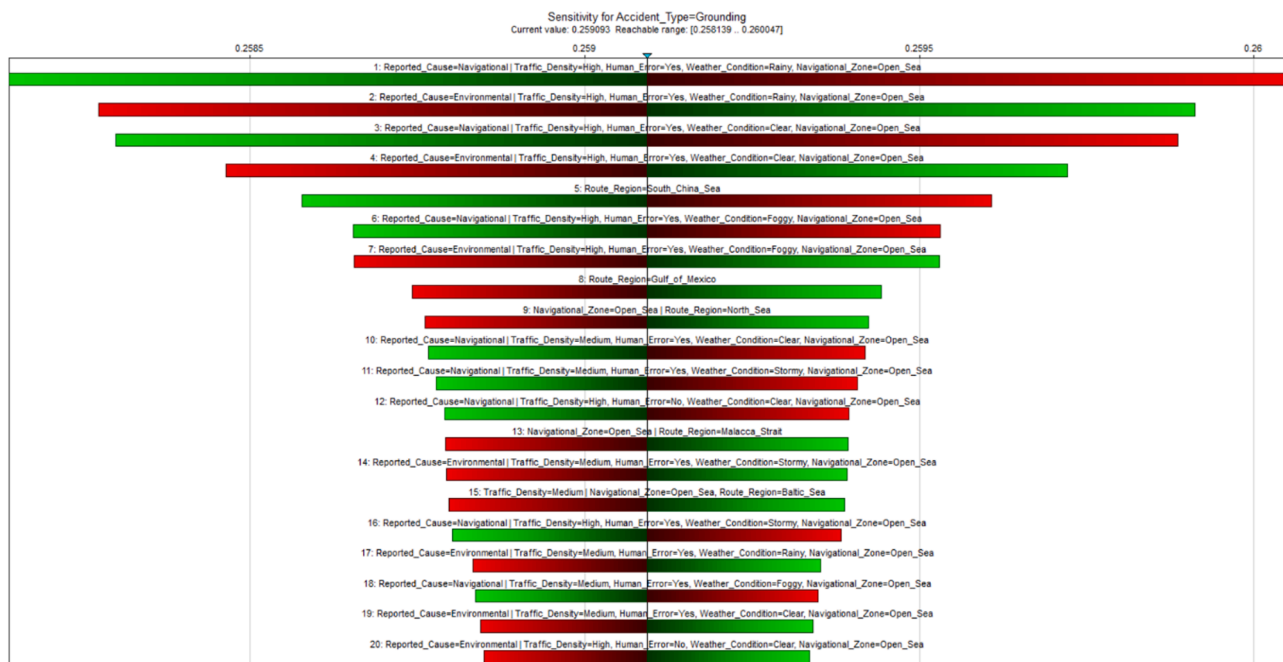


Fig. A4. Tornado chart for target node Accident Type-State= Grounding.

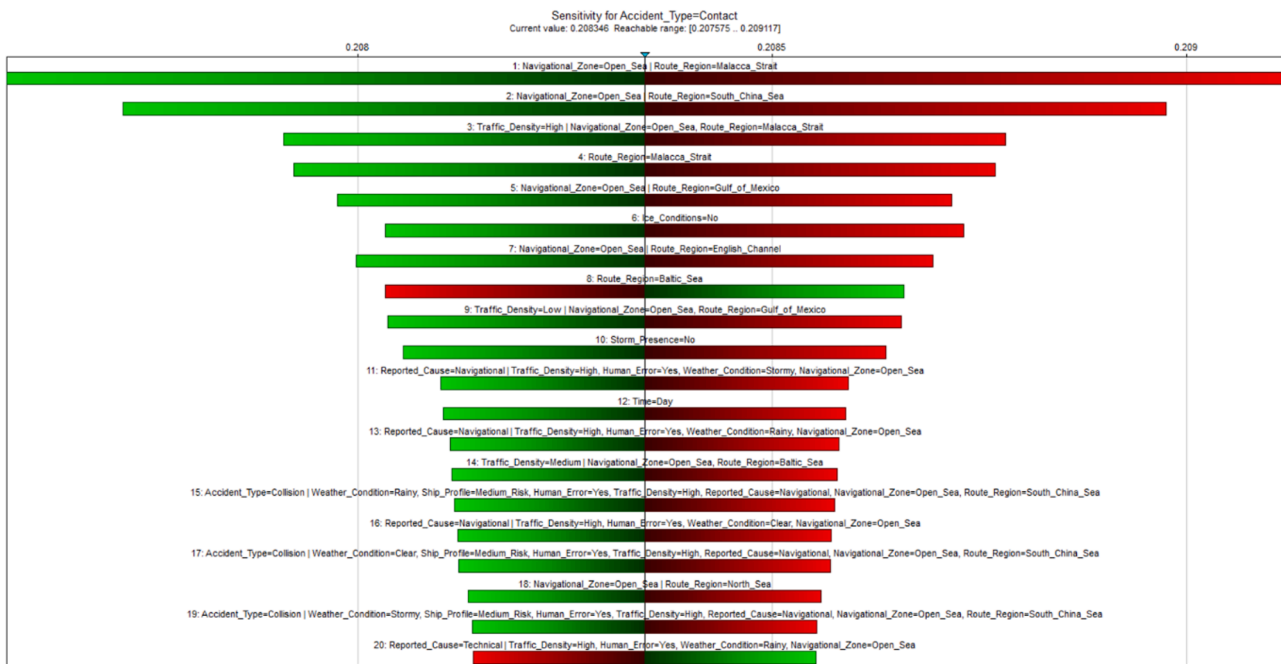


Fig. A5. Tornado chart for target node Accident Type-State= Contact.

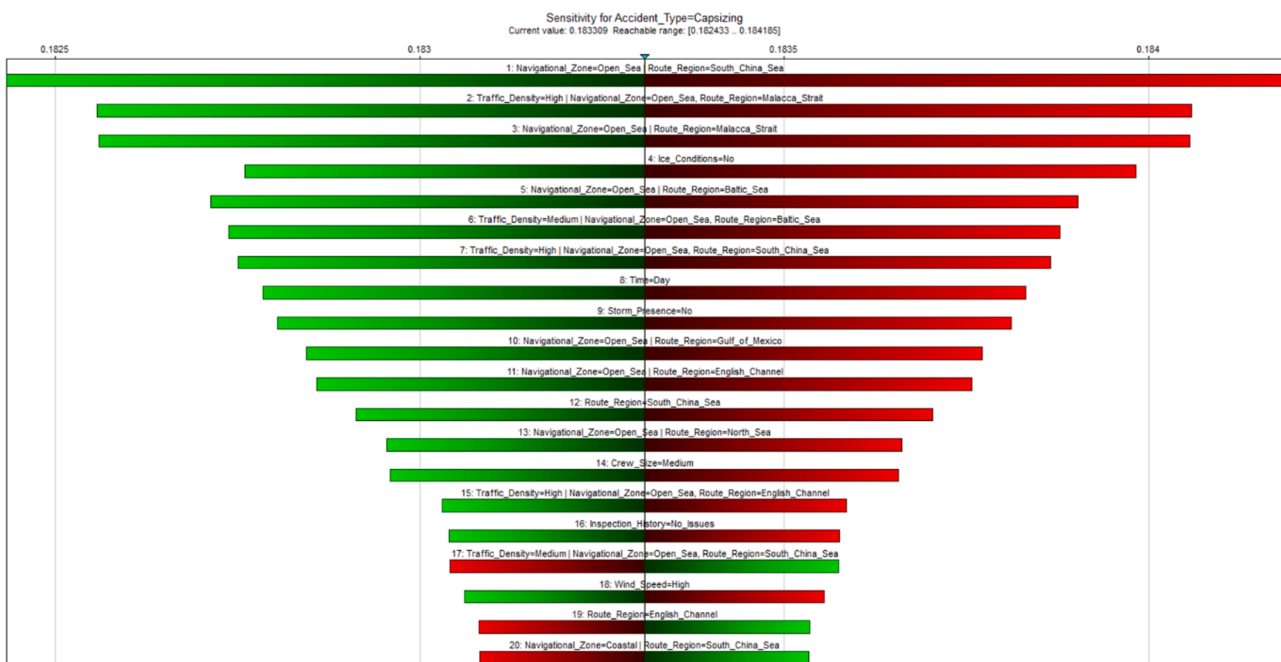


Fig. A6. Tornado chart for target node Accident Type-State= Capsizing.

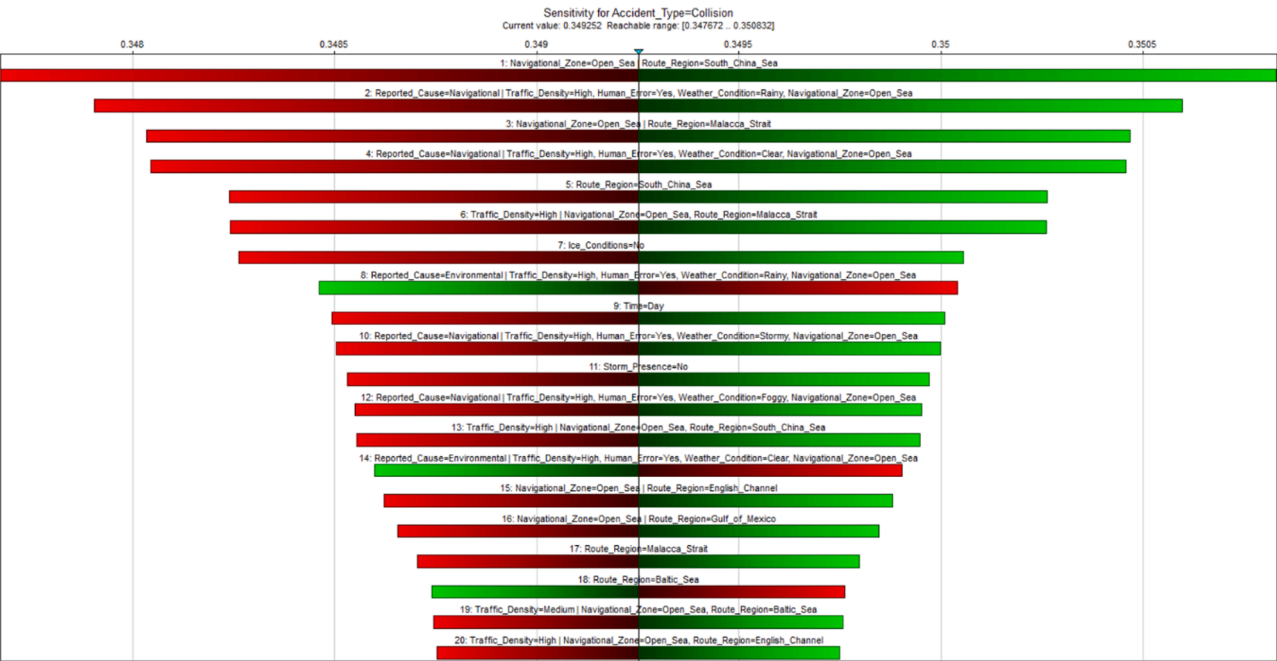


Fig. A7. Tornado chart for target node Accident Type-State= Collision.

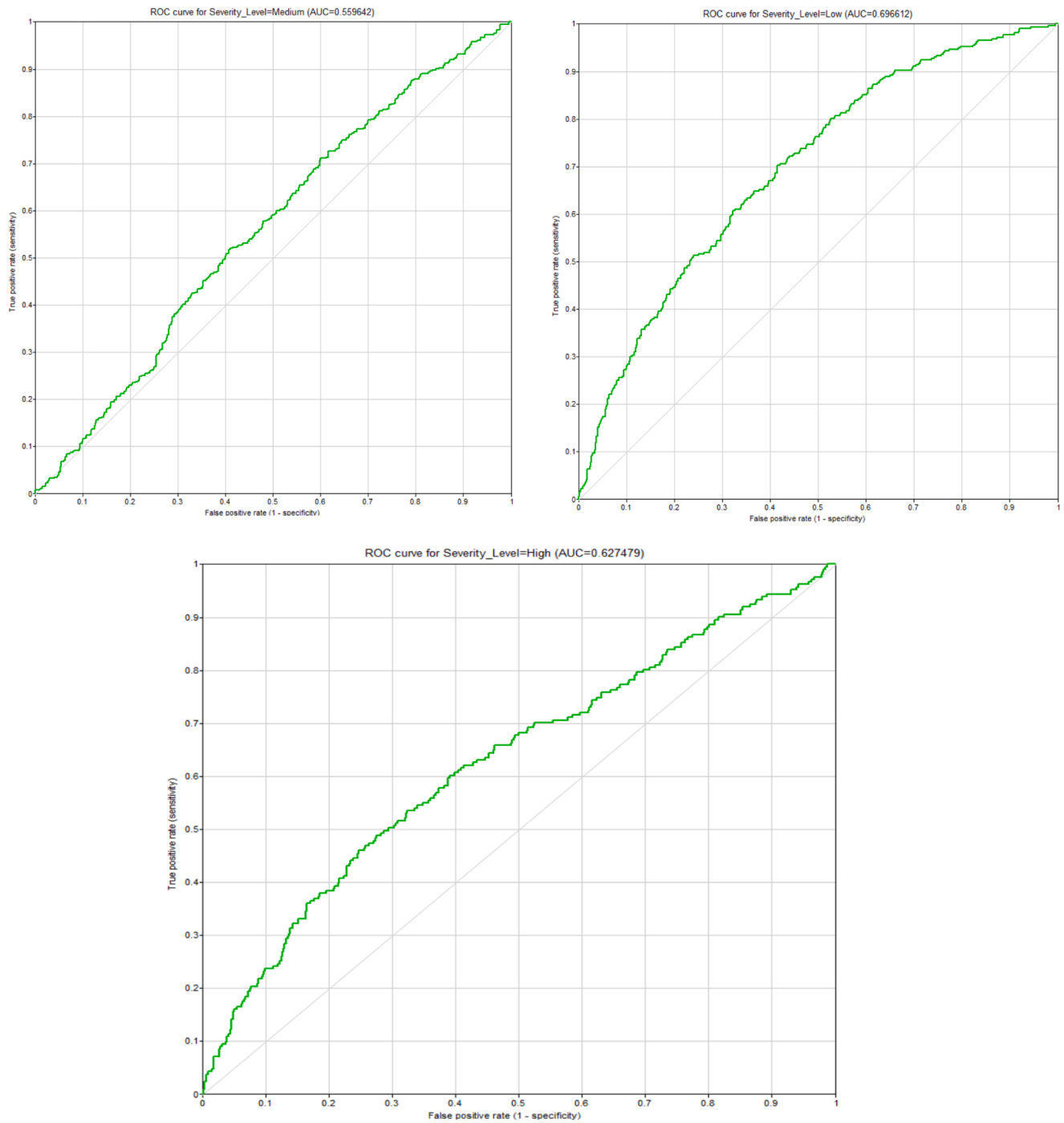


Fig. A8. ROC/AUC validation for TAN model.

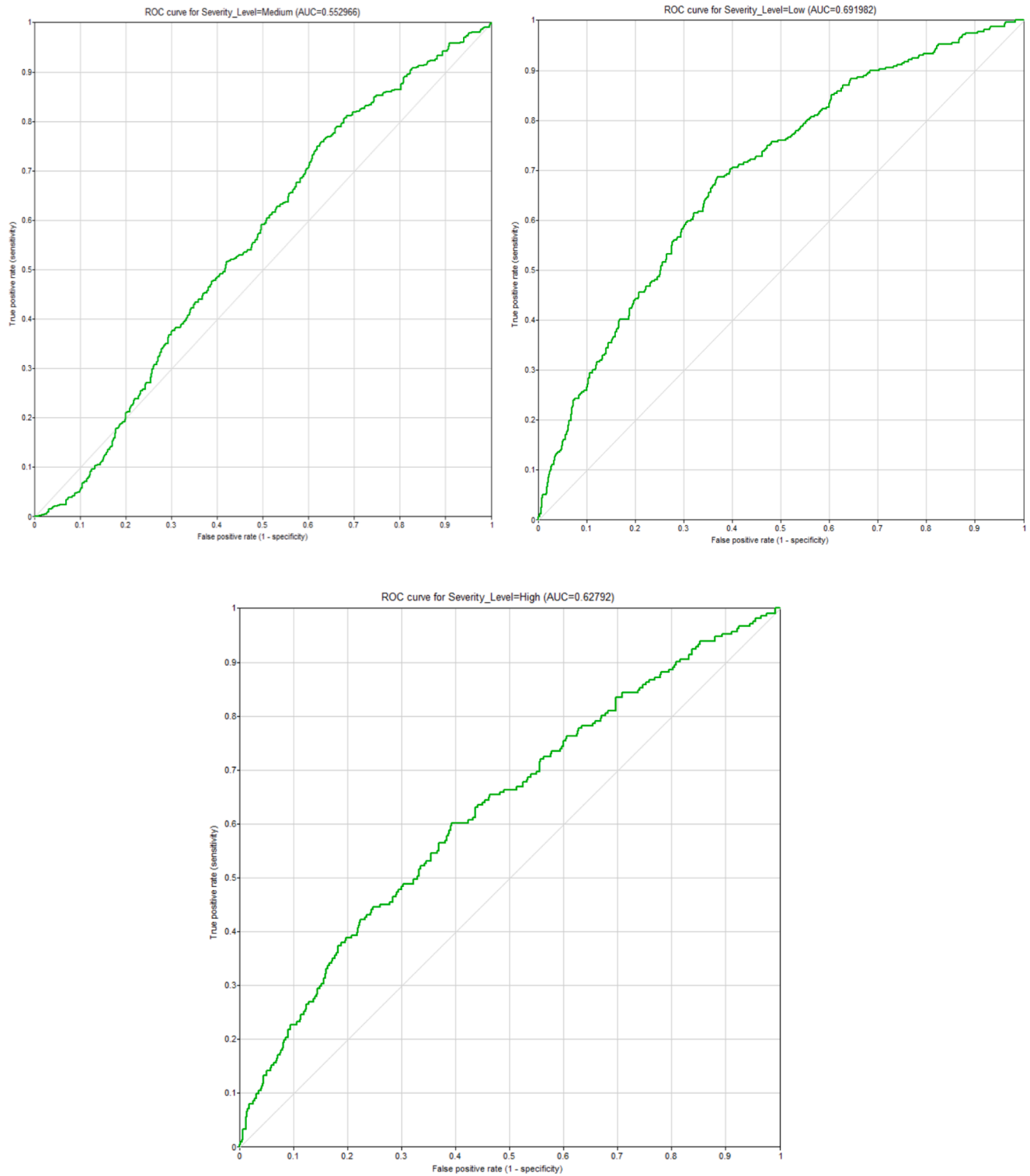


Fig. A9. ROC/AUC Validation curves for the ANB model.

Accident Type									
Weather Condition	Clear								
Ship Profile	Low_Risk			High_Risk					
Human Error	Yes			No			Yes		
Navigational Zone	Open_Sea	Coastal	Harbor	Open_Sea	Coastal	Harbor	Open_Sea	Coastal	
Low	0.34615385	0.25	0.375	0.5	0.083333333	0.33333333	0.37037037	0.5	
Medium	0.48076923	0.5	0.375	0.33333333	0.83333333	0.33333333	0.51851852	0.25	
High	0.17307692	0.25	0.25	0.16666667	0.08333333	0.33333333	0.11111111	0.25	

Fig. A10. CPT of Accident Severity from GeNIe Software 1.

Accident Type									
Weather Condition	Foggy								
Ship Profile	Medium_Risk								
Human Error	No			Yes			No		
Navigational Zone	Open_Sea	Coastal	Harbor	Open_Sea	Coastal	Harbor	Open_Sea	Coastal	
Low	0.27272727	0.06666667	0.77777778	0.5	0.41025641	0.33333333	0.5	0.44444444	
Medium	0.63636364	0.86666667	0.11111111	0.425	0.56410256	0.61904762	0.25	0.44444444	
High	0.09090909	0.06666667	0.11111111	0.075	0.025641026	0.047619048	0.25	0.11111111	

Fig. A11. CPT of Accident Severity from GeNIe Software 2.

Accident Type	Collision								
Weather Condition	Rainy								
Ship Profile	Risk			Medium_Risk					
Human Error	Yes			No			Yes		
Navigational Zone	Open_Sea	Coastal	Harbor	Open_Sea	Coastal	Harbor	Open_Sea	Coastal	
Low	0.48148148	0.44444444	0.16666667	0.17647059	0.055555556	0.33333333	0.33823529	0.27272727	
Medium	0.48148148	0.44444444	0.66666667	0.58823529	0.88888889	0.58333333	0.5	0.45454545	
High	0.037037037	0.11111111	0.16666667	0.23529412	0.055555556	0.08333333	0.16176471	0.27272727	

Fig. A12. CPT of Accident Severity from GeNIe Software 3.

Table A1

Data sources used for accident case compilation, validation, and standardization.

Data category	Source / institution	Coverage / type of information used	Web address
Accident investigation repository	IMO Global Integrated Shipping Information System (GISIS) casualty/incident records	Accident case data, including narratives, causal chains, operational conditions, and outcomes	https://gis.imo.org/
Accident investigation repository	EMSA European Marine Casualty Information Platform (EMCIP)	EU marine casualty and incident records	https://emsa.europa.eu/emcip.html
National investigation body	UK Marine Accident Investigation Branch (MAIB)	Full marine accident investigation reports	https://www.gov.uk/maib-reports
National investigation body	US National Transportation Safety Board (NTSB) Marine	Full marine accident investigation reports	https://www.nts.gov/investigations/AccidentReports/Pages/Reports.aspx
National investigation body	Transportation Safety Board of Canada (TSB) Marine	Marine accident statistics and reports	https://www.tsb.gc.ca/eng/stats/marine/data-6.html
National investigation body	Australian Transport Safety Bureau (ATSB) Marine	Marine investigation reports	https://www.atsb.gov.au/marine-investigation-reports
PSC regime portal	Paris MoU	PSC inspection status and inspection history	https://www.parismou.org/
PSC regime portal	Tokyo MoU	PSC inspection status and inspection history	https://www.tokyo-mou.org/
PSC regime portal	United States Coast Guard (USCG) Port State Control	PSC inspection status and inspection history	https://www.dco.uscg.mil/Port-State-Control/
PSC regime portal	Australian Maritime Safety Authority (AMSA) PSC	PSC inspection status and inspection history	https://www.amsa.gov.au/vessels-operators/port-state-control
Fleet register	S&P Global Maritime Intelligence / Sea-web	Ship particulars, including ship type, age, DWT, and class	https://www.spglobal.com/marketintelligence/en/solutions/maritime-intelligence
Fleet register	Clarksons Research World Fleet Register	Ship particulars, including ship type, age, DWT, and class	https://www.clarksons.com/services/research/products/world-fleet-register/
Met-ocean / reanalysis source	ECMWF ERA5	Environmental and met-ocean descriptors, including wind, visibility proxies, storms, and sea state	https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5
Met-ocean / ocean service	Copernicus Marine	Ocean and marine environmental conditions	https://marine.copernicus.eu/
Met-ocean / environmental archive	NOAA National Centers for Environmental Information (NCEI)	Historical environmental and weather information	https://www.ncei.noaa.gov/
Ice information source	National Snow and Ice Data Center (NSIDC)	Ice condition information	https://nsidc.org/
AIS / traffic corroboration source	MarineTraffic	Corridor-scale vessel traffic context, including density/exposure corroboration	https://www.marinetraffic.com/
AIS / traffic corroboration source	Spire / ExactEarth AIS	Corridor-scale vessel traffic context, including density/exposure corroboration	https://spire.com/maritime/ais/

Data availability

Data will be made available on request.

References

- [1] Zhou K, Xing W, Wang J, Li H, Yang Z. A data-driven risk model for maritime casualty analysis: A global perspective. *Reliab Eng Syst Saf* 2024;244:109925.
- [2] Li H, Ren X, Yang Z. Data-driven Bayesian network for risk analysis of global maritime accidents. *Reliab Eng Syst Saf* 2023;230:108938.
- [3] Bi H, Guo Z, Lu F, Wang S. Geospatial patterns and risk analysis of maritime accidents using GIS and association rules mining. *Ocean Eng* 2025;342:122975.
- [4] Li H, Zhou K, Zhang C, Bashir M, Yang Z. Dynamic evolution of maritime accidents: comparative analysis through data-driven Bayesian networks. *Ocean Eng* 2024;303:117736.
- [5] NTSB. Contact of containership dali with the Francis Scott Key bridge and subsequent bridge collapse. Marine investigation preliminary report, DCA24MM031. National Transportation Safety Board; 2024.
- [6] REUTERS. Fire on car carrier ablaze off dutch coast now less intense [Online]. Thomson Reuters; 2023. Available: https://www.reuters.com/world/europe/carrier-ablaze-off-dutch-coast-carrying-almost-500-electric-cars-2023-07-28/?utm_source=chatgpt.com [Accessed 18th October 2025].
- [7] Van CC, Jaiswal Rishabh. Ship carrying nearly 3000 cars ablaze off dutch coast, crew member dead. Reuters: Thomson Reuters; 2023.
- [8] Rist S, Ugwu K, Sampalo M, Karlsson TM, Rubesinghe CH, Acosta-Dacal A, Pérez-Luzardo O, Zumbado M, Almeda R. Impacts of spilled debris from the X-Press Pearl disaster in Sri Lanka on marine plankton. *Environ Res* 2025;274:121260.
- [9] ALLIANZ. Safety and shipping Review 2025. 2025. <https://commercial.allianz.com/news-and-insights/reports/shipping-safety.html>.
- [10] EMSA. Annual overview of marine casualties and incidents 2024. European Maritime Safety Agency; 2024.
- [11] Homendy J. In: HOMENDY J, editor. NTSB 2020 digest of maritime accident investigations. NTSB: U.S. Naval Institute Staff; 2021.
- [12] Cao Y, Wang X, Wang Y, Fan S, Wang H, Yang Z, Liu Z, Wang J, Shi R. Analysis of factors affecting the severity of marine accidents using a data-driven bayesian network. *Ocean Eng* 2023;269:113563.
- [13] Wang L, Yang Z. Bayesian network modelling and analysis of accident severity in waterborne transportation: A case study in China. *Reliab Eng Syst Saf* 2018;180:277–89.
- [14] Zhang Y, Sun X, Chen J, Cheng C. Spatial patterns and characteristics of global maritime accidents. *Reliab Eng Syst Saf* 2021;206:107310.
- [15] Alyami H, Yang Z, Riahi R, Bonsall S, Wang J. Advanced uncertainty modelling for container port risk analysis. *Acc Anal Prevent* 2019;123:411–21.
- [16] Jiang H, Zhang J, Wan C, Zhang M, Soares CG. A data-driven bayesian network model for risk influencing factors quantification based on global maritime accident database. *Ocean Coast Manage* 2024;259:107473.
- [17] 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.
- [18] Baksh A-A, Abbassi R, Garaniya V, Khan F. Marine transportation risk assessment using Bayesian Network: application to Arctic waters. *Ocean Eng* 2018;159:422–36.
- [19] 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.
- [20] Fan H, Wang J, Chang Z, Lyu J, Jia H. Embracing imperfect data: A novel data-driven bayesian network framework for maritime accidents severity risk assessment. *Ocean Eng* 2025;329:121212.
- [21] Wang H, Chen N, Jiang J, Guedes Soares C, Wu B. A data-driven AcciMap-BN model for risk factor analysis of collisions between merchant ships and fishing vessels. *Maritime Policy Manage* 2025:1–31.
- [22] Yin J, Khan RU, Wang X, Asad M. A data-centered multi-factor seaport disruption risk assessment using bayesian networks. *Ocean Eng* 2024;308:118338.
- [23] Fan S, Yang Z, Blanco-Davis E, Zhang J, Yan X. Analysis of maritime transport accidents using bayesian networks. *Proc Inst Mech Eng, Part O: J Risk Reliab* 2020;234:439–54.
- [24] Khan RU, Yin J, Mustafa FS, Liu H. Risk assessment and decision support for sustainable traffic safety in Hong Kong waters. *IEEE Access* 2020;8:72893–909.
- [25] Wang X, Fan L, Yin J, Yang Z. Dynamic effects of port State control on ship risk profiles. *ASCE-ASME J Risk Uncert Eng Syst, Part A: Civil Eng* 2025;11:04025079.
- [26] Liao S, Weng J, Zhang Z, Li Z, Li F. Probabilistic modeling of maritime accident scenarios leveraging bayesian network techniques. *J Mar Sci Eng* 2023;11 [Online].
- [27] Khan RU, Yin J, Mustafa FS. Accident and pollution risk assessment for hazardous cargo in a port environment. *PLoS One* 2021;16:e0252732.

- [28] Khan RU, Yin J, Mustafa FS, Wang S. Analyzing human factor involvement in sustainable hazardous cargo port operations. *Ocean Eng* 2022;250:111028.
- [29] Liu L, Ullah KR, Muhammad A, Asad M. A corpus-based quantitative risk assessment of language barriers in maritime safety. *Mar Policy Manage* 2025: 1–35.
- [30] 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.
- [31] Wang H, Liu Z, Wang X, Graham T, Wang J. An analysis of factors affecting the severity of marine accidents. *Reliab Eng Syst Saf* 2021;210:107513.
- [32] Jiang M, Lu J. Maritime accident risk estimation for sea lanes based on a dynamic bayesian network. *Mar Policy Manage* 2020;47:649–64.
- [33] Jiang M, Lu J, Yang Z, Li J. Risk analysis of maritime accidents along the main route of the Maritime Silk Road: a Bayesian network approach. *Mar Policy Manage* 2020;47:815–32.
- [34] Wang J, Fan H, Chang Z, Lyu J. Unleashing data power: driving maritime risk analysis with bayesian networks. *Reliab Eng Syst Saf* 2025;264:111310.
- [35] Wang T, Ma L, Ma X. Emergency scenario evolution of Arctic maritime accidents using data-driven dynamic Bayesian network. *Transp Res Part D: Transp Environ* 2025;146:104845.
- [36] Wang W, Zhao J, Chen Y, Shao P, Jia P. Process safety enhancement in maritime operations: A Bayesian network-based risk assessment framework for the RCEP region. *Process Saf Environ Prot* 2025;194:1235–56.
- [37] Choe C-W, Lim S, Kim DJ, Park H-C. Development of spatial clustering method and probabilistic prediction model for maritime accidents. *Appl Ocean Res* 2025; 154:104317.
- [38] Qu X, Wang C, Zhao R, Fang M, Xie X. Multi-source data-driven bayesian network for risk analysis of maritime accidents in the high sea. *Front Mar Sci* 2025;12: 1631650.
- [39] Finnis J, Reid-Musson E. Managing weather & fishing safety: marine meteorology and fishing decision-making from a governance and safety perspective. *Mar Policy* 2022;142:105120.
- [40] Kołowrocki K, Kuligowska E. Operation and climate-weather change impact on maritime ferry safety. Safety and reliability-safe societies in a changing world. CRC Press; 2018.
- [41] Perera LP, Soares CG. Weather routing and safe ship handling in the future of shipping. *Ocean Eng* 2017;130:684–95.
- [42] Rawson A, Brito M, Sabeur Z, Tran-Thanh L. A machine learning approach for monitoring ship safety in extreme weather events. *Saf Sci* 2021;141:105336.
- [43] Valdez Banda OA, Goerlandt F, Montewka J, Kujala P. A risk analysis of winter navigation in Finnish sea areas. *Acc Anal Prevent* 2015;79:100–16.
- [44] Valdez Banda OA, Goerlandt F, Kuzmin V, Kujala P, Montewka J. Risk management model of winter navigation operations. *Mar Pollut Bull* 2016;108: 242–62.
- [45] Goerlandt F, Goite H, Valdez Banda OA, Höglund A, Ahonen-Rainio P, Lensu M. An analysis of wintertime navigational accidents in the Northern Baltic Sea. *Saf Sci* 2017;92:66–84.
- [46] Goerlandt F, Montewka J, Zhang W, Kujala P. An analysis of ship escort and convoy operations in ice conditions. *Saf Sci* 2017;95:198–209.
- [47] Lensu M, Goerlandt F. Big maritime data for the Baltic Sea with a focus on the winter navigation system. *Mar Policy* 2019;104:53–65.
- [48] 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.
- [49] Maternová A, Materna M, Dávid A, Török A, Švábová L. Human error analysis and fatality prediction in maritime accidents. *J Mar Sci Eng* 2023;11 [Online].
- [50] Onyshchenko S, Bychkovsky Y, Melnyk O, Onishchenko O, Zaporozhets A, Bedrii D. Integrating Human factors in project-oriented risk management for maritime safety. In: Melnyk O, Onishchenko O, Zaporozhets A, editors. Maritime systems, transport and logistics I: safety and efficiency of operation. Cham: Springer Nature Switzerland; 2025.
- [51] Oraith H, Blanco-Davis E, Yang Z, Matellini B. An evaluation of the effects of Human factors on pilotage operations safety. *J Mar Sci Appl* 2021;20:393–409.
- [52] Yang Z, Lau Y-Y, Poo MC-P, Yin J, Yang Z. Port resilience to climate change in the Greater Bay Area. *Transp Res Part D: Transport Environ* 2025;142:104681.
- [53] 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.
- [54] Yang Z, Wan C, Yu Q, Yin J, Yang Z. A machine learning-based bayesian model for predicting the duration of ship detention in PSC inspection. *Transp Res Part E: Logist Transp Rev* 2023;180:103331.
- [55] Yang Z, Yang Z, Teixeira AP. Comparative analysis of the impact of new inspection regime on port state control inspection. *Transp Policy (Oxf)* 2020;92: 65–80.
- [56] Yang Z, Yang Z, Yin J. Realising advanced risk-based port state control inspection using data-driven bayesian networks. *Transp Res Part A: Policy Pract* 2018;110: 38–56.
- [57] Yang Z, Yang Z, Yin J, Qu Z. A risk-based game model for rational inspections in port state control. *Transp Res Part E: Logist Transp Rev* 2018;118:477–95.
- [58] Yang Z, Yu Q, Yang Z, Wan C. A data-driven Bayesian model for evaluating the duration of detention of ships in PSC inspections. *Transp Res Part E: Logist Transp Rev* 2024;181:103371.
- [59] Prieto JM, Almorza D, Amor-Esteban V, Endrina N. Review of ship risk analyses through deficiencies found in port state inspections. *J Mar Sci Eng* 2025;13 [Online].
- [60] Sheriff AM, Anantharaman M, Islam R, Nguyen H-O. An in-depth analysis of port state control inspections: A bibliometric analysis and systematic review. *J Int Mar Saf, Environ Affairs, Shipp* 2025;9:2454754.
- [61] Sheriff AM, Anantharaman M, Islam R, Nguyen H-O. A review of port state control inspections: critical issues and emerging trends in maritime oversight. *Austral J Mar Ocean Affairs* 2025;1–23.
- [62] Shi J, Liu Z. Accident data-driven consequence analysis in maritime industries. *J Mar Sci Eng* 2025;13(1):117.
- [63] 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 J Risk Uncert Eng Syst, Part A: Civil Eng* 2024;10: 04023065.
- [64] Animah I. Application of bayesian network in the maritime industry: comprehensive literature review. *Ocean Eng* 2024;302:117610.
- [65] 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.
- [66] Rawson A, Brito M. A survey of the opportunities and challenges of supervised machine learning in maritime risk analysis. *Transp Rev* 2023;43:108–30.
- [67] Koller D, Friedman N. Probabilistic graphical models: principles and techniques. MIT Press; 2009.
- [68] Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann; 2014.
- [69] Chang Z, He X, Fan H, Guan W, He L. Leverage Bayesian Network and Fault Tree method on risk assessment of LNG maritime transport shipping routes: application to the China–Australia route. *J Mar Sci Eng* 2023;11 [Online].
- [70] Fan H, Lyu J, He X, Li B, Ji Y, Chang Z. A novel object-oriented Bayesian network for risk assessment of sea lanes of communication. *Ocean Eng* 2024;300:117347.
- [71] Khan RU, Yin J, Wang S, Hussein M, Alqurashi Y, Al Sulaie S. A system theoretic quantitative risk assessment for port ammonia bunkering operations. *Int J Hydrogen Energy* 2025;172:151217.
- [72] Khan RU, Yin J, Wang X. Fuzzy Bayesian risk assessment of system and component failures in ammonia bunkering. *Ocean Eng* 2025;338:121960.
- [73] Fan H, Enshaei H, Jayasinghe SG. Human error probability assessment for LNG bunkering based on fuzzy Bayesian network-CREAM model. *J Mar Sci Eng* 2022; 10 [Online].
- [74] Hänninen M. Bayesian networks for maritime traffic accident prevention: benefits and challenges. *Acc Anal Prevent* 2014;73:305–12.
- [75] Hänninen M, Valdez Banda OA, Kujala P. Bayesian network model of maritime safety management. *Expert Syst Appl* 2014;41:7837–46.
- [76] Zhang G, Thai VV. Expert elicitation and Bayesian Network modeling for shipping accidents: A literature review. *Saf Sci* 2016;87:53–62.
- [77] Zhang G, Thai VV, Yuen KF, Loh HS, Zhou Q. Addressing the epistemic uncertainty in maritime accidents modelling using Bayesian network with interval probabilities. *Saf Sci* 2018;102:211–25.
- [78] De Campos CP, Corani G, Scanagatta M, Cuccu M, Zaffalon M. Learning extended tree augmented naive structures. *Int. J. Approx. Reason.* 2016;68:153–63.
- [79] Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. *Mach Learn* 1997;29:131–63.
- [80] Hamine, V. & Helman, P. 2005. Learning optimal augmented Bayes networks. arXiv preprint arXiv:cs/0509055.
- [81] Keogh EJ, Pazzani MJ. Learning augmented bayesian classifiers: a comparison of distribution-based and classification-based approaches. In *AIStats*. 1999.
- [82] Sugahara S, Ueno M. Exact learning augmented Naive Bayes classifier. *Entropy* 2013;23 [Online].
- [83] Ankan, A. & Panda, A. Pgmpy: probabilistic graphical models using Python. 2015 2015, 6–11.
- [84] Cooper, C., Webb, G.I. & Schmidt, D.F. 2025. Efficient parameter estimation for bayesian network classifiers using hierarchical linear smoothing. arXiv preprint arXiv:2505.23320.
- [85] Khan RU, Yin J, Mustafa FS, Anning N. Risk assessment for berthing of hazardous cargo vessels using Bayesian networks. *Ocean Coast Manag* 2021;210:105673.
- [86] Khan RU, Yin J, Mustafa FS, Shi W. Factor assessment of hazardous cargo ship berthing accidents using an ordered logit regression model. *Ocean Eng* 2023;284: 115211.
- [87] Fan L, Wang M, Yin J. The impacts of risk level based on PSC inspection deficiencies on ship accident consequences. *Res Transp Bus Manage* 2019;33: 100464.
- [88] Khan B, Khan F, Veitch B. A dynamic Bayesian Network model for ship-ice collision risk in the Arctic waters. *Saf Sci* 2020;130:104858.
- [89] Knapp S, Van De Velden M. Global ship risk profiles: safety and the marine environment. *Transp Res Part D: Transport Environ* 2011;16:595–603.
- [90] Ekici CV, Öztürk Ü, Şenol YE. A data-driven ship risk profile model for Turkish straits (TS-SRP) using Machine Learning. *Ocean Eng* 2024;311:119002.
- [91] 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.
- [92] Kaptan M. Analysis of accidents during vehicle stowage on RO-RO vessels by using Fuzzy Bayesian networks. *Ocean Eng* 2022;260:111997.
- [93] 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.
- [94] Yu Q, Liu K, Chang C-H, Yang Z. Realising advanced risk assessment of vessel traffic flows near offshore wind farms. *Reliab Eng Syst Saf* 2020;203:107086.

- [95] Zhang J, He A, Fan C, Yan X, Soares CG. Quantitative analysis on risk influencing factors in the Jiangsu segment of the Yangtze River. *Risk Analysis* 2021;41: 1560–78.
- [96] Zhang J, Teixeira ÂP, Guedes Soares C, Yan X, Liu K. Maritime transportation risk assessment of Tianjin port with bayesian belief networks. *Risk Anal* 2016;36: 1171–87.
- [97] 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. *Mar Policy Manage* 2024;1–20.
- [98] Marcot BG, Hanea AM. What is an optimal value of k in k-fold cross-validation in discrete bayesian network analysis? *Comput Stat* 2020.
- [99] Wong T-T, Yeh P-Y. Reliable accuracy estimates from k-fold cross validation. *IEEE Trans Knowl Data Eng* 2020;32:1586–94.
- [100] Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information-theoretic approach. Springer; 2002.
- [101] Hurvich CM, Tsai C-L. Regression and time series model selection in small samples. *Biometrika* 1989;76:297–307.
- [102] Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods* 2012;17:228.
- [103] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. New-York: Springer series in statistics; 2009.
- [104] Scutari M. Learning Bayesian networks with the bnlearn R package. *J Stat Softw* 2010;35:1–22.
- [105] Brier GW. Monthly weather review. War Department, Office of the Chief Signal Officer; 1950.
- [106] Bröcker J. Reliability, sufficiency, and the decomposition of proper scores. *Quart J R Meteorol Soc* 2009;135:1512–9.
- [107] Dimitriadis T, Gneiting T, Jordan AI, Vogel P. Evaluating probabilistic classifiers: the triptych. *Int J Forecast* 2024;40:1101–22.
- [108] Murphy AH. A new vector partition of the probability score. *J Appl Meteorol Climatol* 1973;12:595–600.
- [109] EMSA 2025. EMSA: Annual overview of Marine Casualties and Incidents 2025.
- [110] Björkqvist J-V, Lukas I, Alari V, Van Vledder GP, Hulst S, Pettersson H, Behrens A, Männik A. Comparing a 41-year model hindcast with decades of wave measurements from the Baltic Sea. *Ocean Eng* 2018;152:57–71.
- [111] Tuomi L, Kahma KK, Pettersson H. Wave hindcast statistics in the seasonally ice-covered Baltic sea. *Boreal Environ Res* 2011;16.
- [112] FINNISH METEOROLOGICAL INSTITUTE, F. *Ice movements* [Online]. a. Finnish Meteorological Institute; 2021. Available: <https://en.ilmatieteenlaitos.fi/ice-movements> [Accessed 12th April 2026].
- [113] FINNISH METEOROLOGICAL INSTITUTE, F. *Ice season in the Baltic Sea* [Online]. <https://en.ilmatieteenlaitos.fi/ice-season-in-the-baltic-sea>; 2021 [Accessed 10th April 2026].
- [114] HELCOM. HELCOM recommendation 25/7. helcom – HELSINKI commission katajanokanlaituri 6 b FI-00160 HELSINKI. Finland: HELCOM; 2016.
- [115] BIM. Baltic Sea Icebreaking Report 2020-2021. In: Management BI, editor. Baltic icebreaking management; 2022.
- [116] HELCOM. HELCOM recommendation 28E/11. HELCOM – HELSINKI commission katajanokanlaituri 6 b FI-00160 HELSINKI. Finland: HELCOM; 2007.
- [117] Shi X, Xu D, Zhuang H, Liu C. How unsafe acts occur: an automatic text mining study. *Mar Policy Manage* 2022;49:820–30.
- [118] Shi X, Zhuang H, Xu D. Structured survey of human factor-related maritime accident research. *Ocean Eng* 2021;237:109561.
- [119] Knapp S, Heij C. Evaluation of total risk exposure and insurance premiums in the maritime industry. *Transp Res Part D: Transport Environ* 2017;54:321–34.
- [120] Askari HR, Chowdhury MMH, Shaheen MMA, Islam R. Coastal and inland waterway transport accidents: a systematic literature review. *Mar Syst Ocean Technol* 2025;20:46.