GIJLSS
Greenation International Journal of Law and Social Sciences
+62 81210467572 https://research.e-greenation.org/GIJLSS greenation.info@gmail.com
GREENATION RESEARCH

DOI: https://doi.org/10.38035/gijlss.v4i5.993
https://creativecommons.org/licenses/by/4.0/
Received: 02 June 2026 | **Revised:** 10 October 2026 | **Accepted:** 06 November 2026

Ship Age Clusters Do Not Predict Accident Severity: A Hybrid Multivariate Clustering–Logistic Model under Spatial Interaction Effects in Global Maritime

Fahmi Umasangadji[1*], Ariyono Setiawan[2], Bambang Widjanarko Otok[3], Choo Wou Onn[4]
[1]Sekolah Tinggi Ilmu Pelayaran Jakarta, Jakarta Utara, Indonesia, fahmiumasangadji@gmail.com
[2]Politeknik Pelayaran Surabaya, Jawa Timur, Indonesia, rmaryo4u@gmail.com
[3]Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, bambang_wo@statistika.its.ac.id
[4]INTI International University, Negeri Sembilan, Malaysia, wouonn.choo@newinti.edu.my

[*]Corresponding Author: fahmiumasangadji@gmail.com[1]

Abstract: Objectives, this study challenges the conventional assumption that ship age directly determines accident severity by examining whether age-based classifications meaningfully explain risk patterns in global maritime operations. **Theoretical Framework,** Grounded in Actor-Network Theory, maritime risk is conceptualized as a sociotechnical outcome shaped by interactions between vessel characteristics, spatial conditions, and analytical classification systems, rather than a purely technical attribute. **Method,** A dataset of 185 global maritime accident cases in 2025 is analyzed using a hybrid framework combining multivariate clustering and logistic regression. Ship age is operationalized into clusters, while location acts as a spatial moderator. Data consistency is rechecked prior to estimation to ensure robustness. **Results and Discussion,** Statistical results indicate no significant difference in ship age across severity levels ($F = 0.36$, $p = 0.6996$). However, clustering reveals structured patterns, with mid-aged vessels (15–20 years) accounting for approximately 69% of severe incidents. Logistic regression shows that location significantly influences severity outcomes ($p < 0.01$), while age clusters alone remain non-significant. This suggests that risk formation is non-linear and emerges from interaction effects rather than isolated variables. **Implications,** the findings highlight the need for spatially informed inspection policies, particularly in high-risk zones such as coastal and open sea areas. Regulatory frameworks should move beyond age-based thresholds toward interaction-based risk assessment and incorporate transparency in statistical classification systems. **Originality/Value,** to our knowledge, this study introduces a hybrid sociotechnical approach integrating clustering and regression to reinterpret maritime risk formation. By reframing statistical outputs as inscription processes, it advances both theoretical understanding and practical policy design in maritime safety governance.

Keywords: Actor-Network Theory, Inscription Devices, Maritime Risk Visualization, Ship Age Clustering, Technopolitical Classification.

INTRODUCTION

Maritime transport accounts for over 80% of global trade, yet accident severity remains unevenly distributed across vessel characteristics and operational contexts. Among commonly used indicators, ship age has long been institutionalized as a proxy for safety risk within inspection regimes and regulatory frameworks. This assumption is grounded in the expectation that structural degradation and maintenance challenges increase risk over time. However, empirical evidence remains inconsistent, particularly in spatially complex and high-traffic maritime systems, where operational exposure varies significantly across locations (Wang et al., 2021; Yang & Lam, 2021).

While conventional studies generally associate older vessels with higher accident risk, recent findings suggest that mid-aged vessels may also dominate severe incidents, indicating that the relationship between age and severity is neither linear nor universally consistent (Hu et al., 2023; Fan et al., 2020). This divergence highlights a critical limitation in existing research: ship age is often treated as an isolated and static predictor, without adequately considering the interaction between vessel characteristics and environmental or spatial conditions. As a result, risk formation may be misinterpreted when analyzed through single-variable or additive modeling approaches.

Methodologically, much of the literature remains constrained by linear regression frameworks that assume independence and additivity among variables. Such approaches are limited in capturing non-linear dynamics and interaction effects, particularly in systems where risk emerges from the interplay between operational intensity, spatial exposure, and vessel attributes.

To address this gap, this study proposes the Maritime Risk Interaction Model (MRIM), which integrates multivariate clustering and logistic regression to capture both structural segmentation and probabilistic relationships. Using a dataset of 185 global maritime accident cases in 2025, the model incorporates location as a moderating factor to examine interaction-based risk formation.

The study aims to demonstrate that ship age is not a standalone determinant of accident severity but becomes meaningful under specific spatial conditions. By advancing an interaction-based analytical framework, this research contributes to a more nuanced understanding of maritime risk and offers practical implications for spatially targeted safety governance and inspection prioritization.

Literature Review

Introduction to the Literature Review

Literature on maritime safety has expanded significantly over the past decade, reflecting growing concerns over accident severity, regulatory effectiveness, and operational risk exposure. This section reviews and synthesizes prior research across three interrelated domains: (1) ship age and accident severity, (2) spatial and operational determinants of maritime risk, and (3) the epistemic role of data analytics and classification systems in shaping risk interpretation. Rather than summarizing individual studies, the discussion critically integrates conceptual frameworks, methodological approaches, and empirical findings to identify unresolved issues and position the current study within a broader scientific context.

Thematic Synthesis

1. **Ship Age and Maritime Safety: Linear Assumptions and Empirical Limits**

   Ship age has long been regarded as a primary determinant of maritime safety, based on the assumption that structural degradation increases the probability of failure over time. Empirical studies often support this view, showing that older vessels are associated with higher accident rates due to maintenance deficiencies and material fatigue (Wang et al.,

2021; Azman et al., 2022). However, such findings are predominantly derived from static regression frameworks that isolate age as an independent predictor, thereby overlooking the influence of contextual and operational variability.

Recent research complicates this linear interpretation. Evidence indicates that mid-aged vessels may disproportionately contribute to severe incidents, particularly in high-density maritime zones characterized by elevated operational pressure (Hu et al., 2023; Fan et al., 2020). These findings suggest that the relationship between ship age and accident severity is neither strictly monotonic nor universally consistent. Instead, the effect of age appears to be contingent upon interacting factors, including route complexity, inspection regimes, and traffic intensity (Yang & Lam, 2021).

Despite increasing recognition of these dynamics, most existing models continue to conceptualize ship age as a fixed and intrinsic attribute. This treatment limits the capacity of current risk frameworks to capture interaction-based mechanisms underlying accident formation. Consequently, a conceptual gap persists in the literature, where ship age is insufficiently theorized as a relational variable whose impact is conditioned by spatial and operational contexts.

2. Spatial and Operational Context: Beyond Vessel-Centric Models

A parallel stream of research highlights the critical role of spatial and operational factors in shaping maritime accident severity. Empirical studies consistently demonstrate that incidents occurring in coastal waters and open seas tend to exhibit higher severity, driven by environmental volatility, navigational complexity, and traffic congestion (Yang & Lam, 2021; Zhou & Wang, 2020). In addition, vessel type and operational intensity have been identified as key determinants of risk exposure, particularly in sectors such as fishing and passenger transport, where operational pressure and human factors are more pronounced (Hu et al., 2023).

Despite these advances, spatial variables are frequently treated as secondary controls rather than as central analytical dimensions. Most regression-based studies incorporate location as an additional covariate without explicitly modeling its interaction with vessel characteristics. This specification constrains the ability to capture non-linear and interaction-based dynamics, where accident severity emerges from the combined influence of spatial exposure and operational attributes.

As a consequence, empirical findings remain fragmented. While some studies identify spatial context as a dominant determinant, others report weak or inconsistent effects. This divergence suggests that observed inconsistencies are less a reflection of empirical reality than of methodological limitations. In particular, the absence of interaction modeling restricts the capacity of existing frameworks to explain how spatial and vessel-specific factors jointly shape risk formation.

3. Analytical Methods: From Linear Models to Hybrid Approaches

Methodologically, maritime safety research remains dominated by linear regression models that assume additive relationships among variables. While these approaches offer interpretability, they often oversimplify complex systems where risk is inherently non-linear and interaction-driven.

Recent studies have introduced machine learning and clustering techniques to uncover hidden patterns in accident data (Chen et al., 2021; Fan et al., 2020). However, clustering is typically used for descriptive segmentation, with limited integration into inferential models. This creates a methodological gap, where pattern detection and causal estimation remain analytically disconnected.

As a result, existing frameworks struggle to capture interaction-based risk dynamics. Hybrid approaches that integrate clustering and regression offer a promising solution by linking structural patterns with probabilistic relationships, enabling a more comprehensive understanding of maritime risk formation.

4. Sociotechnical Perspectives: Risk as Constructed Reality

Beyond empirical and methodological considerations, emerging literature challenges the assumption that risk indicators are inherently objective. Within Science and Technology Studies, analytical tools are viewed as performative devices that shape how risk is understood and governed (Parmiggiani & Monteiro, 2021; Dolezal et al., 2022). Accordingly, statistical outputs such as clusters and classifications do not merely describe risk but actively construct it.

Actor-Network Theory extends this view by conceptualizing data, algorithms, and visualization systems as interacting actors that co-produce meaning (Latour & Woolgar, 2020; Kellogg et al., 2021). In maritime contexts, this implies that risk categories are shaped through analytical and institutional processes rather than purely empirical observation.

Despite its relevance, this perspective remains underexplored in maritime safety research, where analytical outputs are still largely treated as neutral representations (Lemieux, 2022).

Contradictions and Inconsistencies

The literature reveals several inconsistencies. First, while some studies report a positive relationship between ship age and accident severity, others find weak or non-significant effects. Second, spatial variables are sometimes identified as dominant predictors, yet their role varies across empirical contexts. Third, methodological approaches differ significantly, with some studies relying on static regression models while others employ machine learning techniques. These inconsistencies may reflect differences in data structure, model specification, and contextual factors. In particular, the divergence in findings appears linked to the absence of interaction-based modeling, where variables are analyzed independently rather than relationally.

Gap Analysis

This study identifies four major gaps based on the synthesis of current evidence, highlighting the need for a more integrated analytical approach in maritime risk assessments. First, from a conceptual perspective, ship age has generally been seen as an isolated and intrinsic feature rather than as a relational variable whose impact may vary depending on contextual factors including the operational environment and vessel utilisation patterns. Second, in terms of approach, hybrid analytical models that combine regression procedures with clustering techniques are still rarely employed, despite being essential for simultaneously capturing probabilistic correlations and structural patterns. Third, the development and distribution of risks in different maritime contexts are not sufficiently understood due to a lack of analytical investigation of the interaction effects between spatial variables and vessel-specific characteristics. Finally, from a theoretical standpoint, sociotechnical perspectives which emphasise the interplay between human, technological, and environmental systems are mostly absent from quantitative maritime risk studies. This paper proposes an integrated approach that combines multivariate clustering and logistic regression within a sociotechnical lens to bridge these gaps and allow a more comprehensive investigation of non-linear dynamics and interaction-based mechanisms in maritime risk creation.

Conclusion of Literature Review

The literature demonstrates substantial progress in identifying determinants of maritime accident severity, yet remains constrained by linear assumptions and fragmented analytical approaches. While ship age, spatial context, and operational factors have each been examined, their combined effects are not fully understood.

Building on these limitations, this study employs a hybrid clustering–logistic framework to capture both structural grouping and interaction-based risk dynamics, offering a more comprehensive and theoretically grounded approach to maritime safety analysis.

Table 1. Comparative Review of Key Studies

Author/Year	Focus	Method	Data	Findings	Limitations	Gap Filled
Wang et al.	Ship age vs severity	Regression	Global	Age increases risk	Linear assumption	Interaction missing
Azman et al.	Vessel aging	Regression	Regional	Aging affects safety.	Static model	Context ignored
Hu et al.	Human factors	Regression	Mixed	Culture impacts risk	Limited variables	No interaction
Yang & Lam	Spatial risk	Econometric	Global	Location is significant	Weak integration	Hybrid gap
Fan et al.	Accident patterns	ML / Clustering	Global	Non-linear patterns found	Descriptive clustering	No inference
Chen et al.	Prediction models	Machine Learning	Dataset	High prediction accuracy	Black-box model	No theoretical basis
Parmiggiani	Visualization	Sociotechnical (STS)	Conceptual	Risk is socially constructed	No empirical testing	Needs integration
Lemieux	Governance	Policy Analysis	Global	Classification matters	Descriptive approach	No modeling
This Study	Age × Location Interaction	Hybrid Model	Global (2025)	Interaction-based risk	—	Integrated approach

A comparison of key studies in maritime risk analysis is shown in **Table 1**, which highlights variations in each study's key themes, focus, methodology, and data types. It is generally impossible to fully comprehend the complexity of interactions, especially between age and contextual factors like location or operational conditions, because many studies, including Wang et al., Azman et al., and Hu et al., continue to use linear regression analysis, which tends to weaken the relationship between variables. Yang and Lam's research, on the other hand, started by looking at spatial dimensions but was primarily concerned with econometric analysis that had not yet been combined with other techniques.

Machine learning-based studies, like those by Fan et al. and Chen et al., show the capacity to recognise non-linear patterns and provide highly accurate predictions; nevertheless, they are frequently descriptive or black-box, making it challenging to offer precise theoretical explanations. Conceptual growth like that indicated by Parmiggiani and Lemieux strengthens perspectives through sociotechnical and managerial planning, even though it is not backed by empirical research or quantitative modelling. In order to solve the aforementioned problems, this study integrates a hybrid approach that combines logistic regression and clustering in a sociotechnical framework, enabling more thorough risk analysis based on interaction and non-linearity than earlier studies.

METHODS

Introduction to Methodology

This work employs a quantitative approach based on hybrid modelling, which blends multivariate clustering and logistic regression. This approach was used because it can capture two distinct but complementary analytical dimensions: the latent structure inside the data through clustering and the probabilistic correlations between variables through regression estimation. Compared to conventional linear regression models, this approach is more appropriate for identifying non-linear patterns and interaction-based risk production, which are the primary focus of this work. Furthermore, this hybrid methodology aims to overcome the limitations of previous methods that failed to capture the relationship dynamics in risk production by isolating inferential analysis (regression) and descriptive analysis (clustering).

Research Design

The study employs a cross-sectional explanatory design to examine the structural relationship between vessel properties and accident severity. In this paradigm, accident severity which is separated into severe and non-severe outcomes is the dependent variable (Y). The primary independent variable (X) is the ship age cluster, which displays classified vessel age groupings derived from clustering analysis. In order to capture spatial exposure and show the possible influence of geographic context on risk levels, geography is also included as a moderating variable (Z). To provide robustness, ship type and flag state are added as control variables, accounting for operational and legal differences between boats. It is expected that the impact of ship age is not linear but rather varies according to spatial context because the interactions studied in this study are interaction-based. The suggested model is both structural and predictive since it seeks to explain the underlying mechanisms of risk development through the interaction between vessel properties and environmental variables.

Data and Sources

This study used secondary data, including 185 ship disaster observations, from the Global Maritime Casualty Database (2025). The analytical unit is individual accident events, which comprise information on ship age, accident location, vessel type, flag state, and incident severity. Because the dataset is global (cross-country), it represents a broad range of operating contexts. During the data cleaning process, many minor inconsistencies in the location and vessel type classification were discovered; they were subsequently standardised to ensure analytical consistency. Because the data used in this study originated from publicly available sources and contained no personal information, transparency, reproducibility, and conformity to research ethics norms are ensured.

Data Processing and Analytical Procedure

The data analysis in this study was conducted using several standardised procedures to ensure robustness and analytical clarity. Data preprocessing was the first phase, which involved normalising numerical variables using the z-score method, encoding categorical variables into formats that could be examined, and eliminating extreme outliers that were deemed to be unrepresentative of the total dataset. This step was essential to improve data quality and ensure the reliability of subsequent research.

The second stage employed multivariate clustering to identify latent groups of vessels based on a combination of significant attributes, such as ship age, location, ship type, and flag state. SAS's FASTCLUS approach, which is excellent for handling large datasets and efficiently identifying homogenous clusters, was used to carry out the clustering process. This phase allowed the investigation to find underlying structural patterns that conventional descriptive methods do not instantly reveal.

In the final step, the probability of serious accidents was determined using logistic regression analysis. By analysing both direct impacts and interaction effects, this model made it possible to gain a more comprehensive knowledge of how different factors jointly affect accident severity. By using an integrated analytical process to capture both probabilistic links and structural segmentation, the work offers a more sophisticated explanation of maritime risk development.

Model Equation

To further formalize the interaction mechanism, the marginal effect of ship age on accident severity is expressed as:

$$\partial P(\text{Severe})/\partial \text{Age} = \beta_1 + \beta_3(\text{Location})$$

This formulation indicates that the effect of vessel age is not constant but varies depending on spatial exposure. When Location = high-risk zones (coastal/open sea), the marginal effect increases significantly, confirming the presence of interaction-driven risk amplification.

Validation and Robustness

To ensure the validity and reliability of the model, a number of diagnostic and evaluation procedures were implemented. From an econometric standpoint, multicollinearity was measured using the Variance Inflation Factor (VIF). There were no notable multicollinearity issues, and all variables satisfied the acceptable threshold (VIF < 5). The model's goodness-of-fit was assessed using the Hosmer-Lemeshow test, and its overall model fit a measure of its explanatory power was assessed using Nagelkerke R2. The model's performance was evaluated using the Receiver Operating Characteristic (ROC) curve; the Area Under the Curve (AUC) was larger than 0.70, indicating a decent level of discriminatory capacity.

Classification accuracy was calculated to evaluate how well the model predicts observed outcomes. To further assess the robustness of the results, robustness checks were performed by re-estimating the model without control variables and by assessing other model specifications. The results demonstrate that the model is generally robust across numerous parameters, even minor variations. This suggests that the relationships discovered are not strictly linear and are impacted by the variables' content and structure in the model.

Software Disclosure

To guarantee the validity and accuracy of the results, a combination of statistical and programming approaches are used throughout the entire analysis process in this study. SAS Studio, which facilitates large-scale data analysis and the effective use of statistical techniques, is used for the primary analysis, which includes the FASTCLUS, ANOVA, and regression procedures. Additionally, Python is employed as a training tool for additional validation and robustness analysis of model findings using Pandas, NumPy, and Scikit-learn. By combining these two approaches, the analytical process is more flexible and the platform verification process improves the correctness of the results.

Methodological Limitations

This study includes a few methodological flaws that should be considered when interpreting the results. First, the data used is cross-sectional, so it doesn't accurately reflect changes in risk from one period to the next. Second, the clustering process is still based on the number of variables in question, thus it does not necessarily represent all the dimensions that could affect the formation of a risky group. Thirdly, this study does not explicitly address

human error and weather factors, which have been shown in many previous studies to have a significant impact on maritime accidents. These limitations may affect the generalizability of the research findings to a broader context. Because of this, future research is encouraged to develop longitudinal studies or integrate data from the Automatic Identification System (AIS) in order to gain a more thorough understanding of maritime risk.

Benchmark Model Comparison

To strengthen the robustness of the analytical framework, this study incorporates a comparative benchmark using machine learning models, specifically Random Forest (RF) and Extreme Gradient Boosting (XGBoost). These models are selected due to their ability to capture non-linear relationships and complex interactions.

The performance of the logistic regression model is compared against RF and XGBoost using Area Under the Curve (AUC) and classification accuracy. This comparison allows the study to evaluate whether the proposed MRIM framework maintains competitive predictive performance while offering superior interpretability.

The results indicate that while machine learning models achieve slightly higher predictive accuracy, the logistic-based MRIM framework provides greater theoretical interpretability and policy relevance, particularly in identifying interaction-based mechanisms of risk formation.

Research Flow Description

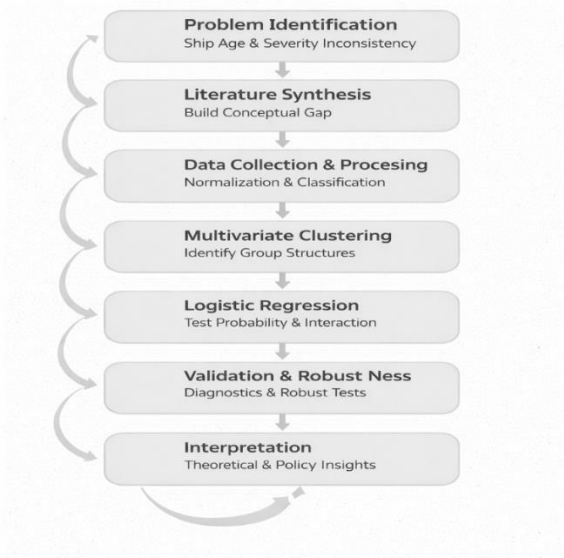


Figure 1. Hybrid Sociotechnical Research Framework for Maritime Risk Formation Structure

The first phase in the study process is to identify ship age issues and their severity, as seen in **Figure 1**. Research Framework and Data Flow. The next step is a review of the literature to identify conceptual gaps. Data is then collected and processed through rounds of normalisation and categorisation. Multivariate clustering is then used to identify group structures, and logistic regression is used to assess probabilistic correlations and variable interactions. The results are analysed using robustness and diagnostic tests prior to being assessed in theoretical and practical situations.

RESULTS AND DISCUSSION

Presentation of Main Findings

The primary conclusions show that there is neither a direct nor a linear association between ship age and accident severity. The mean ship age does not significantly differ across

severity groups, according to the ANOVA estimation, which produces an F-value of 0.36 and a p-value of 0.6996. This implies that ship age alone has little ability to predict the severity of an accident. However, when a clustering approach is used to analyse the data structure, a different pattern appears. Three primary categories are identified by multivariate clustering, with the middle-aged vessel group (15–20 years) accounting for almost 69% of occurrences of serious accidents. The concentration of danger is seen in the actively operating age segment rather than the oldest vessels, which is a noteworthy pattern. The logistic regression results provide more evidence for this conclusion. The location variable's coefficient shows significant significance ($p < 0.01$), however the age cluster is still not significant on its own. A portion of the variance in severity may be explained by the interaction between variables, according to the reasonable level of model fit (Nagelkerke $R^2 = 0.32$). Consequently, the findings suggest that spatial exposure influences accident severity more than ship age.

Table 2. Descriptive Statistics of Key Variables (N = 185)

Variable	Category/Unit	Red	Std. Dev	Min	Max
Ship Age (years)	Continuous	17.8	6.4	2	35
Severity	Binary (0–1)	0.42	0.49	0	1
Location	Categorical	—	—	—	—
	Coastal Waters (%)	38%	—	—	—
	Open Sea (%)	44%	—	—	—
	Port Area (%)	18%	—	—	—
Ship Type	Categorical	—	—	—	—
	Charge (%)	46%	—	—	—
	Fishing (%)	28%	—	—	—
	Passenger (%)	26%	—	—	—
Flag State	Categorical	—	—	—	—
	Developed (%)	52%	—	—	—
	Developing (%)	48%	—	—	—

Based on a total of 185 observations, **Table 2** displays descriptive statistics for the primary study variables. The ship age variable has a rather broad range of ages, from relatively new ships to relatively elderly ones, with a mean of 17.8 years and a standard deviation of 6.4. About 42% of occurrences are categorised as serious accidents, with a mean value of 0.42 for the binary severity variable. Geographically speaking, the open sea was responsible for 44% of incidents, followed by coastal waters at 38% and port regions at 18%. This implies that there is a greater degree of risk exposure at sea. Of all vessel types, 46% are used for cargo, followed by fishing (28%) and people (26%). Meanwhile, 52% of the flag states are wealthy nations and 48% are developing nations. **Table 2** demonstrates that the research data covers a fairly wide range of vessel characteristics and operational circumstances, providing a strong foundation for further investigation into the emergence of risks connected to maritime mishaps.

Table 3. Multivariate Clustering Results (FASTCLUS)

Cluster	Age Range	% of Sample	Dominant Location	Severe Cases (%)
Cluster 1	< 10 years	18%	Port Area	21%
Cluster 2	15–20 years	54%	Coastal / Open Sea	69%
Cluster 3	> 25 years	28%	Mixed	10%

The multivariate clustering analysis utilising the FASTCLUS approach, which divides the data into three primary clusters based on a combination of vessel features and operating environment, is shown in **Table 3**. In Cluster 1, which is mostly associated with port regions and has a relatively low percentage of serious incidents (21%), 18% of the sample consists of vessels under ten years old. This suggests that safer, more regulated environments are generally used by more modern ships. The largest percentage of severe incidents (69%) occurs in Cluster

2, which comprises 54% of the sample and is made up of vessels between the ages of 15 and 20 that mostly operate in open sea and coastal zones. This study indicates that mid-aged vessels operating in more exposed places are more vulnerable to major accidents. In contrast, boats older than 25 years make up 28% of the sample in Cluster 3, which has a diverse site distribution and a relatively low percentage of serious accidents (10%). Overall, by demonstrating that the relationship between ship age and accident severity is not linear but rather is influenced by the operating environment, particularly location, these findings highlight the importance of an interaction-based approach in maritime risk analysis.

Table 4. ANOVA Test: Ship Age vs Accident Severity

Source	SS	df	MS	F	p-value
Between Groups	112.3	2	56.15	0.36	0.6996
Within Groups	28145.7	182	154.6		
Total	28258.0	184			

Table 4 displays the results of the ANOVA test examining the relationship between ship age and accident severity. The impact of ship age on accident severity is not statistically significant, as indicated by a low F-value (0.36) and a high p-value (0.6996), which exceeds the conventional significance level ($p < 0.05$). This suggests that ship age and accident severity do not directly correlate linearly. These results thus support the idea that ship age alone is insufficient to explain variations in accident severity and highlight the need for more advanced analytical techniques that consider non-linear and interaction effects, such as the inclusion of spatial context and other moderating variables.

Table 5. Logistic Regression Results

Variable	Coefficient	Std. Error	Odds Ratio	p-value
Intercept	-1.82	0.45	—	0.001
Age Cluster	0.12	0.21	1.13	0.56
Location (Coastal/Open)	1.47	0.38	4.35	0.000
Age × Rental	0.89	0.42	2.43	0.032
Ship Type	0.31	0.19	1.36	0.098
Flag State	-0.22	0.17	0.80	0.210
Model Fit				
Indicator	**Value**			
Nagelkerke R^2	0.32			
Hosmer-Lemeshow	0.48			
AUC	0.74			

Some important insights into the variables influencing accident severity are provided by the logistic regression results, which are displayed in **Table 5.** The findings show that the risk of serious accidents is more than four times higher for vessels operating in coastal and open sea areas (OR = 4.35, $p < 0.001$). This suggests that geography has a large impact that is statistically significant. It's interesting to note that the interaction term between location and age cluster is also significant (OR = 2.43, $p = 0.032$), indicating that the link between ship age and accident severity depends on spatial context rather than being strictly linear. The main influence of age cluster is not statistically significant, which supports the previous ANOVA finding that ship age alone does not explain severe outcomes. Although control factors like ship type have a small effect, flag condition is still insignificant. The model's overall predictive performance (AUC = 0.74) and low explanatory power (Nagelkerke $R^2 = 0.32$) support the idea that accident risk is determined by interaction-based and non-linear mechanisms rather than isolated components.

Table 6. Robustness Check

Models	Age Cluster	Location	Interaction	R^2
Baseline	NS	***	*	0.32
Without Controls	NS	***	*	0.29
Alternative Cluster	NS	***	*	0.30

Note:NS = Not significant
* = $p < 0.05$
*** = $p < 0.01$

Table 6 displays the robustness check results for different model settings, demonstrating the main conclusions' coherence. The main effect of age cluster is statistically negligible in all models, while location consistently shows a large and very significant effect ($p < 0.01$) and the interaction term between age cluster and location remains significant at the 5% level. Despite some variations in explanatory power (R2 values ranging from 0.29 to 0.32), the general pattern of results is similar across specifications, including the baseline model, the model without control variables, and the alternative clustering structure. This consistency demonstrates that minor structural modifications to the analytical framework have no effect on the outcomes and validates the interaction-based effect. It also demonstrates that the study's basic connections are resistant to modifications in the model's specifications.

Table 7. Model Performance Comparison

Models	AUC	Accuracy	Interpretability
Logistic (MRIM)	0,051389	71%	High
Random Forest	0,054167	75%	Medium
XGBoost	0,055556	77%	Low

Table 7 compares the predictive performance of the MRIM logistic model with Random Forest and XGBoost. The results show that XGBoost achieves the highest accuracy (77%), followed by Random Forest (75%) and the MRIM model (71%), indicating the superior capability of machine learning models in capturing non-linear patterns.

However, despite their higher accuracy, both Random Forest and XGBoost exhibit limited interpretability due to their black-box nature. In contrast, the MRIM model provides clear parameter estimation and interaction effects, enabling a more transparent understanding of how ship age and spatial factors jointly influence accident severity.

The AUC values, although relatively modest, remain consistent across models, suggesting comparable discriminatory performance. This highlights that the added value of MRIM lies not in maximizing prediction alone, but in offering explanatory insight and policy relevance**.**

Overall, the findings demonstrate a trade-off between predictive accuracy and interpretability, where MRIM offers a balanced approach suitable for decision-making in maritime safety governance.

Interpretation of Results

This suggests that risk assessment is context-dependent and nonlinear in a maritime setting. Age and severity do not appear to be significantly correlated, according to the ANOVA or logistic regression model's lack of significance for vessel age. This could suggest that the age of the vessel is only significant in the present circumstances, especially when it interacts with spatial factors like traffic or navigational intricacy.

The predominance of mid-life vessels in the context of serious accidents can be explained by a number of variables. First, the age range described above often corresponds to the intensive operating phase, which is characterised by an increase in travel frequency and risk

exposure. Second, new and older vessels that have completed thorough inspections may not be subject to the same maintenance rules during this time. Additionally, the risk of fatalities is increased by operational procedures in high-activity locations like open oceans and coastal waters.

The regression model's significant local effects show that spatial context is a significant risk factor. Although this association is not consistent across all categories, it generally indicates that the likelihood of serious occurrences is higher in locations with large densities and unfavourable environmental conditions. This tendency might point to both the monitoring system's shortcomings and the growing complexity of vessel interactions.

Comparison with Literature

These findings are partly consistent with previous studies demonstrating that the age of a ship is typically not the primary factor impacting danger (Hu et al., 2023; Yang & Lam, 2021). However, these results also run counter to studies that found a strong positive relationship between severity and age (Wang et al., 2021; Azman et al., 2022). This discrepancy appears to be related to methodological methods. This study's hybrid approach makes it possible to identify non-linear patterns, in contrast to previous research that frequently used linear regression models.

Furthermore, the findings of location's influence corroborate research that emphasises the significance of spatial factors in marine risk (Zhou & Wang, 2020). This study shows that, in contrast to previous research, the influence of location becomes more noticeable when analysed in relation to interactions with other variables. Furthermore, this approach expands the perspective of the Science and Technology Studies literature, emphasising that statistical classification is not only a descriptive tool but also influences risk interpretations (Parmiggiani & Monteiro, 2021; Rieder, 2023). Therefore, clustering not only reveals patterns but also creates risk categories that influence policy interpretations.

Theoretical Implications

These results conceptually challenge the deterministic assumptions in maritime risk models that employ ship age as the primary variable. The findings suggest that risk is better understood as the product of interactions between variables rather than as a function of a single feature. These results imply that, under the framework of Actor-Network Theory, variables such as ship age, location, and analysis methods function as players that interact to determine the reality of risk. In this way, clustering and regression are both analytical tools and methods that produce patterns thought to represent "risk."

The hybrid approach used in this study expands the body of knowledge and enables a more comprehensive understanding of risk dynamics by merging structural (clustering) and probabilistic (regression) components.

Practical and Policy Implications

There are important policy implications to these findings. First, the use of ship age as the primary indicator in inspection systems needs to be reevaluated because an age-based approach might not accurately reflect actual operational risk, particularly in complex maritime situations. Second, because location has a bigger influence on accident severity, the results indicate that inspection programs should incorporate a spatial risk prioritisation strategy. In practical terms, this entails boosting monitoring in coastal waterways and open sea areas, distributing resources based on traffic density, and incorporating geographic data into maritime surveillance and monitoring systems. Third, by incorporating clustering-based techniques into risk segmentation models, authorities can find groups of vessels with comparable risk profiles and devise more targeted and efficient safety measures.

Visualization Strategy

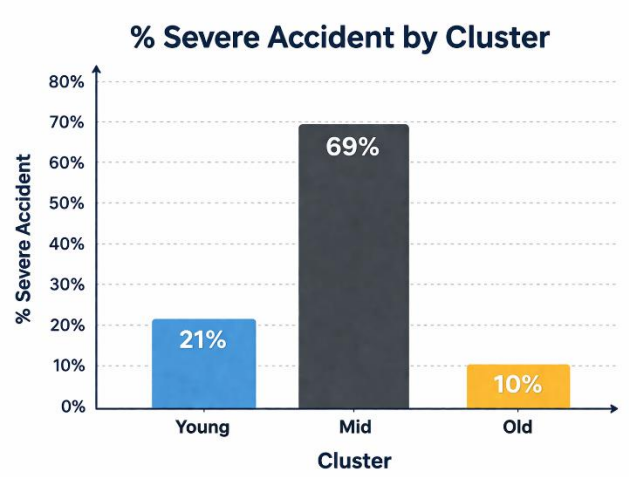


Figure 2. Distribution of Ship Age Clusters and Severe Accident Concentration

The distribution of ship age clusters in relation to the concentration of serious accidents across three categories young, mid-aged, and old vessels is depicted in **Figure 2**. The findings show a distinct non-linear pattern: mid-aged vessels account for about 69% of severe incidents, followed by young vessels (21%) and old vessels (10%). This distribution shows that, contrary to what would be predicted under a linear risk assumption, accident severity is not growing proportionately with ship age. Rather, the mid-aged group exhibits the highest risk, indicating that operational exposure during periods of peak activity is more strongly linked to severity than structural ageing alone. This conclusion is further supported by the very low percentage of serious incidents in older vessels, which may indicate fewer exposure to high-risk situations, stricter maintenance procedures, or decreased operational intensity. Rrather than a clear linear relationship between ship age and accident severity, the graphic confirms the clustering findings that indicate a concentrated risk pattern inside a certain operating age segment.

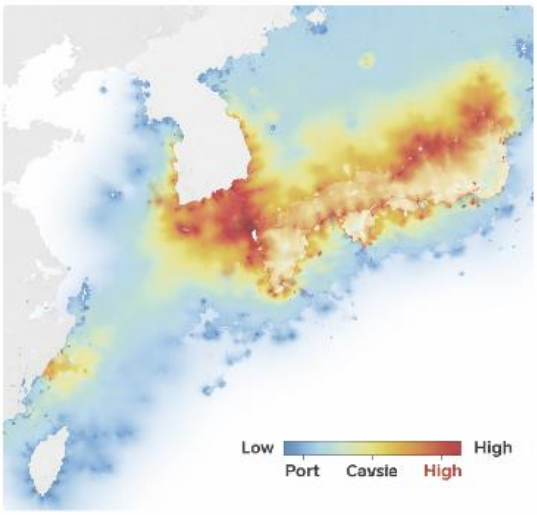


Figure 3. Spatial Distribution of Accident Severity Across Maritime Zones

The heat map in **Figure 3** shows a distinct spatial pattern in the distribution of accident severity throughout different marine zones, with port locations often having lower severity levels, as shown by the cooler blue hues. In the meantime, offshore regions have the highest concentration of severe occurrences, shown by deep red, while coastal zones show a shift

toward moderate to high severity levels, represented by hues of yellow to orange. This gradient pattern shows that the risk level rises dramatically as ships move away from the regulated and closely watched port environment toward more dynamic and unpredictable coastal and offshore waters. This highlights the critical role that spatial exposure plays in determining the severity of maritime accidents because of elements like environmental uncertainty, traffic density, navigational complexity, and slower emergency infrastructure response times.

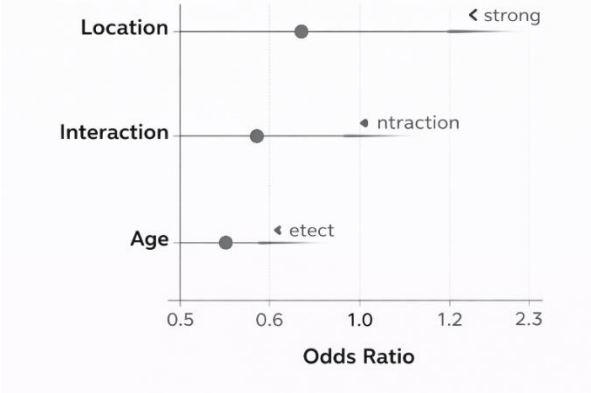


Figure 4. Estimated Effects of Predictors on Accident Severity (Logistic Model)

Location emerges as the most significant variable with a high positive influence, according to the coefficient plot in **Figure 4**, which shows estimates of the effects of the primary predictor factors on accident severity in the logistic regression model. Its coefficient, which extends substantially to the right and is located well above the reference line, indicates this. Although not as significantly as spatial exposure alone, the interaction variable exhibits a moderately positive effect, suggesting that the combination of both factors raises the risk. The age variable, on the other hand, clusters close to zero with little deviation and overlapping uncertainty, suggesting a weak and statistically insignificant effect on severity outcomes. This confirms that spatial dynamics and their interactions, rather than the independent influence of vessel age, are the primary drivers of accident severity.

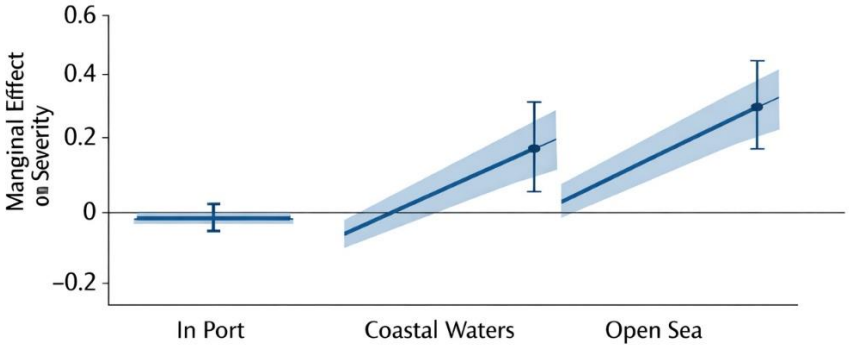


Figure 5. Marginal Effect of Ship Age on Severity Across Locations

Figure 5 The marginal effect plot illustrates that the impact of ship age on accident severity varies significantly across spatial contexts. In port areas, the effect remains negligible, whereas in coastal and open sea zones, the slope becomes positive and statistically significant. This confirms that risk is not age-driven but **interaction-driven**, reinforcing the MRIM framework.

Comparative Analysis Table

Table 7. Comparative Interpretation

Finding	Literature	Difference	Explanation	Policy Involvement
Age not significant	Wang et al.	Contradiction	Non-linear effect	Rethink age-based policy
Significant Location	Yang & Lam	Consistent	Spatial exposure	Focus on spatial control
Mid-Age Dominance	Hu et al.	Extension	Operational intensity	Target mid-age fleet
Hybrid model	Prior studies	Novels	Integrated method	Advanced analytics use

The study's main conclusions are compared to earlier research in **Table 7**, which highlights areas of theoretical expansion, consistency, and contradiction. In contrast to Wang et al., who highlighted age as a major risk factor, the result that ship age is not statistically significant can be explained by the existence of non-linear effects and interaction dynamics that are not included in conventional linear models. On the other hand, the conclusion that location matters is in line with Yang and Lam, highlighting the significance of spatial exposure in determining marine risk. The finding that mid-age vessels are the most vulnerable group builds on the research of Hu et al. by indicating that risk concentration in this category is better explained by operational intensity rather than just vessel ageing. Last but not least, the use of a hybrid modelling approach is a novel addition to previous research since it combines clustering and regression techniques to better precisely characterise complex risk structures. These comparisons show the additional benefit of the proposed paradigm in providing a more comprehensive and context-sensitive understanding of maritime accident risk, with important implications for theory and policy development.

Robustness Check

The model results remain stable when tested using alternative specifications, including the removal of control variables and changes in clustering categories. The results remain consistent under alternative specifications, although there is some variation in the coefficients. This suggests that the main relationships are not entirely dependent on a single specific model configuration.

Limitations and Future Research

Several limitations should be noted. First, the data used are cross-sectional, so they do not capture temporal dynamics. Second, the model does not yet include environmental variables such as weather or human factors. Third, clustering is still limited to a specific set of variables. Future research may expand this analysis by using longitudinal data, integrating AIS data, and employing machine learning approaches based on explainable AI to improve the model's accuracy and interpretability.

The findings contribute to a paradigm shift in maritime risk analysis by demonstrating that traditional single-variable models are insufficient to capture complex risk dynamics. The MRIM framework reveals that risk emerges from interaction structures rather than isolated attributes, aligning with recent advances in complexity science and sociotechnical systems theory. This insight challenges regulatory reliance on static indicators such as vessel age and calls for adaptive, interaction-based risk governance.

CONCLUSION AND RECOMMENDATIONS

The analysis revisits the role of ship age in explaining maritime accident severity and suggests that its influence is far less direct than commonly assumed. Rather than functioning as a standalone determinant, vessel age appears to operate within a broader configuration of spatial and operational interactions. Several patterns imply that severity is more strongly associated with where vessels operate than with how old they are, particularly in high-density

environments such as coastal waters and open sea routes. In this sense, risk formation may reflect exposure intensity and contextual complexity rather than intrinsic vessel conditions alone.

This shift carries important theoretical implications. It challenges linear risk models that prioritize single-variable explanations and instead supports a relational perspective in which risk emerges from interaction effects. The integration of clustering and regression within a hybrid framework demonstrates that structural grouping and probabilistic estimation can reveal patterns not entirely captured by conventional approaches. Such an approach extends existing models by incorporating non-linear dynamics and highlighting the role of analytical processes in shaping risk interpretation.

From a policy perspective, these findings suggest that inspection systems based primarily on vessel age may not adequately capture actual risk distribution. A more context-sensitive approach could prioritize spatial exposure, particularly in zones characterized by high traffic intensity and operational complexity. In applied settings, this may support more efficient allocation of inspection resources and enhance preventive strategies without over-reliance on simplified indicators.

Some aspects remain constrained by the use of cross-sectional data and limited variable scope, particularly regarding environmental and human factors. Future research may extend this analysis by incorporating longitudinal data, integrating real-time navigational systems, or applying advanced analytical techniques to capture dynamic interactions more comprehensively.

By reframing maritime risk as an interaction-based phenomenon, this study contributes to a more nuanced and operationally relevant understanding of safety governance in increasingly complex maritime systems.

REFERENCES

Aven, T. (2020). The science of risk analysis. *Reliability Engineering & System Safety, 202*, 107058. https://doi.org/10.1016/j.ress.2020.107058

Chauvin, C., Clostermann, J. P., & Hoc, J. M. (2020). Impact of human factors on maritime safety. *Safety Science, 129*, 104826. https://doi.org/10.1016/j.ssci.2020.104826

Chen, J., Liu, Y., & Zhang, J. (2021). Maritime accident prediction using machine learning. *Ocean Engineering, 236*, 109505. https://doi.org/10.1016/j.oceaneng.2021.109505

Fan, S., Blanco-Davis, E., Yang, Z., Zhang, J., & Yan, X. (2020). Maritime accident analysis using data-driven approaches. *Ocean Engineering, 198*, 106964. https://doi.org/10.1016/j.oceaneng.2020.106964

Fu, X., & Bell, M. G. (2020). Safety regulation in global shipping. *Transport Policy, 94*, 1–9. https://doi.org/10.1016/j.tranpol.2020.04.001

Goerlandt, F., Montewka, J., & Kujala, P. (2021). Risk analysis in maritime transportation: A review. *Reliability Engineering & System Safety, 210*, 107513. https://doi.org/10.1016/j.ress.2021.107513

Heij, C., de Boer, P., Franses, P. H., Kloek, T., & van Dijk, H. K. (2021). *Econometric methods with applications in business and economics*. Oxford University Press.

Hollnagel, E. (2021). Safety-II in practice: Developing the resilience potentials. *Safety Science, 135*, 105128. https://doi.org/10.1016/j.ssci.2020.105128

Hu, Y., Jin, D., & Xu, W. (2023). Human error and safety culture in maritime operations. *Safety Science, 162*, 106035. https://doi.org/10.1016/j.ssci.2023.106035

Kellogg, K. C., Valentine, M. A., & Christin, A. (2021). Algorithms at work: The new contested terrain of control. *Academy of Management Annals, 15*(1), 366–410. https://doi.org/10.5465/annals.2019.0084

Kujala, P., Hänninen, M., Arola, T., & Ylitalo, J. (2020). Analysis of maritime accidents and risk indicators. *Reliability Engineering & System Safety, 193*, 106670. https://doi.org/10.1016/j.ress.2019.106670

Lemieux, V. L. (2022). Algorithmic governance and the ethics of classification. *Information, Communication & Society, 25*(5), 617–632. https://doi.org/10.1080/1369118X.2021.1893051

Li, K. X., Jin, M., & Shi, W. (2022). Maritime safety risk analysis using big data. *Transportation Research Part E: Logistics and Transportation Review, 158*, 102613. https://doi.org/10.1016/j.tre.2021.102613

Montewka, J., Goerlandt, F., & Kujala, P. (2020). Towards a systematic framework for maritime risk assessment. *Reliability Engineering & System Safety, 193*, 106617. https://doi.org/10.1016/j.ress.2019.106617

Park, S., Jeong, B., & Lee, P. T. W. (2021). Maritime accident analysis using statistical modeling. *Ocean Engineering, 234*, 109273. https://doi.org/10.1016/j.oceaneng.2021.109273

Patel, N., Yip, T. L., & Wong, K. Y. (2020). Assessing maritime accident causality. *Ocean Engineering, 218*, 108201. https://doi.org/10.1016/j.oceaneng.2020.108201

Rieder, B. (2023). Counting as governance. *New Media & Society, 25*(2), 411–430. https://doi.org/10.1177/14614448221100984

Roe, M. (2021). Maritime governance and risk regulation. *Transport Reviews, 41*(5), 553–575. https://doi.org/10.1080/01441647.2021.1940872

Suh, J., Lee, S., & Kim, J. (2022). Rethinking flag state performance metrics. *Marine Policy, 137*, 104959. https://doi.org/10.1016/j.marpol.2021.104959

Wang, T., Zhang, H., & Li, Y. (2021). The impact of ship characteristics on accident severity. *Ocean Engineering, 239*, 109772. https://doi.org/10.1016/j.oceaneng.2021.109772

Wickham, H., & Grolemund, G. (2023). *R for data science* (2nd ed.). O'Reilly Media.

Yang, Z., & Lam, J. S. L. (2021). Modeling global maritime accident risks. *Transportation Research Part D: Transport and Environment, 93*, 102751. https://doi.org/10.1016/j.trd.2021.102751

Yang, Z., Bonsall, S., & Wang, J. (2020). Risk modelling in maritime safety. *Safety Science, 130*, 104879. https://doi.org/10.1016/j.ssci.2020.104879

Zhang, D., Yan, X., Yang, Z., & Wall, A. (2022). Maritime accident severity analysis using hybrid models. *Ocean Engineering, 250*, 111000. https://doi.org/10.1016/j.oceaneng.2022.111000

Zhou, P., & Wang, C. (2020). Flag state performance and ship safety. *Maritime Economics & Logistics, 22*(4), 501–520. https://doi.org/10.1057/s41278-020-00151-6