International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(10s) 2026 pp. 356-366
https://doi.org/10.70917/ijcisim-2026-3589
Article

Human Error Analysis in Maritime Accidents: A Hybrid Machine Learning Approach for Enhanced Predictive Modeling

Prabhat Nigam[1], Neeraj Anand[2], Manan Bhasin[3], Aashish Kumar[*4]

[1]Chitkara University School of Maritime Studies, Chitkara University, Punjab, India, pnigam@chitkara.edu.in
[2]Department of Supply chain Management, Chitkara Business School, Chitkara University, Punjab, India, neeraj.anand@chitkara.edu.in
[3]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India, manan1869.be22@chitkara.edu.in
[4*]Chitkara University Institute of Engineering and Technology, Chitkara University, NH-64, Village Jhansla, Rajpura, Punjab, India, aashish.kumar@chitkara.edu.in
Corresponding author email-id: aashish.kumar@chitkara.edu.in; er.aashish39@gmail.com

Abstract: This study presents a novel hybrid machine learning approach for analyzing human error factors in maritime accidents using HELCOM accident data. We developed an advanced classification model that combines Random Forest, Gradient Boosting, XGBoost, LightGBM, and neural networks through a soft voting ensemble mechanism. Our methodology includes enhanced data preprocessing techniques, sophisticated feature engineering, and class imbalance correction through SMOTE resampling. The visualizations produced meet publication-quality standards with optimized color schemes, typography, and statistical representations suitable for high-impact journals. The hybrid model achieved 84% accuracy with macro-averaged F1-score of 0.58 across eight accident classes, identifying key human factors contributing to maritime incidents. Our findings indicate that specific human element factors have significant correlations with particular accident types, offering valuable insights for maritime safety policy development and accident prevention strategies. This research contributes to the growing field of data-driven maritime risk assessment by providing a robust methodological framework for human error analysis in the maritime domain.
Keywords: Maritime safety, Human factors, Machine learning, Hybrid ensemble model, HELCOM database, Maritime accidents

1. Introduction

Shipping by sea remains key to global trade, with approximately 90% of global trade transported by sea, valued at over $14 trillion annually (UNCTAD, 2023). Technology has not yet altered the frequency of sea accidents occurring, resulting in loss of life, harm to the environment, and estimated economic losses of between $6–12 billion each year (Allianz Global Corporate & Specialty, 2022). Human mistake has been the prevailing contributor to maritime safety incidents for a long time, accounting for 75–96% of such incidents (Reason, 1990; Rothblum, 2000; Hetherington et al., 2006). The International Maritime Organization (IMO) has consistently asserted the importance of taking systematic measures to address shipping-related human element risks (IMO, 2020). Still, well-developed frameworks for measuring, modeling, and predicting human error in shipping activities are lacking (Ung, 2019).The intricacy of maritime safety analysis stems from several interrelated factors such as severe environmental factors, fatigue, remoteness, ship attributes, and complex operational routines (Stanton et al., 2006; Sandhåland et al., 2021). The conventional methods of accident investigation have been intensive in their usage of qualitative taxonomies like HFACS (Celik & Cebi, 2009) or CREAM (Hollnagel, 1998; Chen et al., 2019), which are excellent classificatory tools but weak at modeling nonlinear interdependences between human, technical, and environmental factors. More recent quantitative risk modeling work emphasizes the possibilities of network-based and cascading failure methodologies (Chen & Cai, 2022), but these methods are still constrained in their ability to predict when brought to bear on heterogeneous maritime data sets. Breakthroughs in data-driven modeling offer a promising route to more scientifically valid accident analysis. Specifically, machine learning has been notably utilized in high-risk fields like aviation (Zhang et al., 2018) and nuclear safety (Liu et al., 2019), proving it to be able to deal with intricate factor interactions. In the maritime sector, emerging studies point to the application of deep learning and hybrid methods for the analysis of safety (Zhu et al., 2021; Fan et al., 2023; Navas de Maya et al., 2024). Ensemble learning, which exploits the advantage of multiple classifiers, has been shown to outperform single classifiers in intricate safety-relevant environments (Dietterich, 2000; Zhou, 2021; Ganaie et al., 2022). In particular, hybrid ensemble architectures are well-suited to manage heterogeneous marine data and uncover subtle patterns of human mistakes that are not identifiable through conventional methods (Zhang et al., 2022). A common problem with maritime data, however, is class imbalance, where the more severe yet rarer forms of accidents are underrepresented. Techniques such as SMOTE (Chawla et al., 2002), resampling strategies (Bui et al., 2022), and advanced imbalance treatment methods (Fernández et al., 2018) have proven capable of effectively addressing this issue, providing more accurate accident classification. Apart from this, methods of hyperparameter optimization such as Bayesian optimization (Bergstra et al., 2011; Wang et al., 2023) and advanced interpretability methods such as SHAP (Lundberg & Lee, 2017; Covert et al., 2022) are increasingly being used in safety research to enhance predictive accuracy and interpretability of the model. The integration of Explainable AI (XAI) ideas is particularly critical in maritime safety, where decision support systems need to be explainable and actionable to aid crew training, regulatory compliance, and policy (Arrieta et al., 2020; Saeed et al., 2022).

This research fills these knowledge gaps by implementing a new hybrid machine learning framework for the analysis of the HELCOM maritime accident database specifically focusing on human error contributions. Our model combines Random Forest, Gradient Boosting, XGBoost, LightGBM, and neural networks using a soft-voting ensemble technique, obtaining 84% accuracy and 0.58 macro-averaged F1-score in eight accident categories. Key methodological developments are (1) sophisticated preprocessing and feature engineering methods specific to maritime human factors, (2) handling of class imbalance through SMOTE and resampling approaches, (3) deployment of publication-quality visualization protocols, and (4) incorporation of explainability techniques to detect and interpret the most significant human factors.The fourfold contribution of this study is: (i) formulation of a hybrid ensemble methodological approach specially tailored for maritime accident classification, (ii) evidence of outperformance of predictive accuracy compared to conventional single-model methods, (iii) establishment of critical human element variables and their interrelation with specific types of accidents, and (iv) offers a methodological framework facilitating both accident prevention measures and maritime safety policy-making. Through the integration of human reliability analysis with cutting-edge machine learning, this research pushes the developing discipline of data-driven shipping safety and provides actionable recommendations for enhanced global shipping resilience (Li et al., 2023).The remainder of this paper outlines the proposed approach, including data preprocessing, feature engineering, model design, and evaluation protocols, followed by results and discussion with emphasis on model performance and interpretability. It concludes with key implications for maritime safety policy, crew training, and directions for future research.

2. Methodology

The methodology flowchart (Figure 1) illustrates a comprehensive data processing and analysis framework specifically designed for maritime accident classification. The workflow initiates with data acquisition from the HELCOM maritime accident dataset, followed by a structured preprocessing phase that systematically addresses data quality issues. The quality assessment branch splits into three parallel treatment protocols: categorical missing values are replaced with 'Unknown' to preserve distributional integrity, numerical missing values are imputed with median values to minimize distortion, and outliers are detected and treated using the IQR-based method to ensure robust statistical analysis. The feature engineering phase demonstrates a multifaceted approach to variable creation and transformation, generating four distinct feature categories: binary features capturing pollution status and environmental conditions; categorical features representing ship size classifications based on maritime domain knowledge; a numerical hull safety level index derived from engineering principles; and human element features that encode error classifications. These engineered features undergo selection procedures to identify the most predictive variables before dataset partitioning using an 80-20 train-test split ratio. The model development phase culminates in the implementation of a hybrid ensemble architecture, with performance evaluation conducted through both discriminative (ROC curves) and classification metrics. The final stages focus on human element factor analysis and the visualization of human error impacts on maritime accidents, leading to the comprehensive presentation of results that emphasize the interrelationships between human factors and accident outcomes in maritime operations.

2.1 Data Acquisition and Characteristics

This study utilizes the HELCOM (Helsinki Commission) maritime accident database, which contains records of incidents in the Baltic Sea region. The dataset includes information on accident types, vessels involved, environmental conditions, and human factors. Key variables include accident characteristics (type, location, date, pollution incidents), vessel information (type, size, hull configuration), environmental conditions (visibility, ice conditions), and human element factors (fatigue, communication issues, training deficiencies).

The target variable is Acc_Type, representing eight distinct accident categories: Collision, Grounding/stranding, Contact (allision), Fire/explosion, Hull failure, Machinery damage, Other, and Pollution.

2.2 Data Preprocessing and Feature Engineering

Data preprocessing is crucial for ensuring model performance and reliability. Our preprocessing pipeline consists of multiple stages:

2.2.1 Missing Value Treatment

We implemented a context-aware approach to handle missing values:

$$x_{ij} = \begin{cases} \text{'Unknown'} & \text{if } x_{ij} \text{ is categorical and missing} \\ \text{median}(x_j) & \text{if } x_{ij} \text{ is numerical and missing} \end{cases}$$

This approach preserves the overall distribution while explicitly modeling uncertainty in the data.

2.2.2 Feature Engineering

We derived several informative features from the raw data to enhance the model's predictive capabilities:

$$\text{Has\_Pollution} = \begin{cases} 1 & \text{if Pollution = 'Yes'} \\ 0 & \text{if Pollution = 'No'} \end{cases}$$

For ship size categorization using domain knowledge-based binning:

$$\text{Ship\_Size\_Category} = \begin{cases} \text{Very Small} & \text{if } 0 < \text{Sh1Size\_gt} \le 1000 \\ \text{Small} & \text{if } 1000 < \text{Sh1Size\_gt} \le 5000 \\ \text{Medium} & \text{if } 5000 < \text{Sh1Size\_gt} \le 15000 \\ \text{Large} & \text{if } 15000 < \text{Sh1Size\_gt} \le 30000 \\ \text{Very Large} & \text{if Sh1Size\_gt} > 30000 \end{cases}$$

Hull safety level based on maritime engineering principles:

$$\text{Hull\_Safety\_Level} = \begin{cases} 3 & \text{if Sh1\_Hull = 'Double hull'} \\ 2 & \text{if Sh1\_Hull = 'Double bottom' or 'Double sides'} \\ 1 & \text{if Sh1\_Hull = 'Single hull'} \\ 0 & \text{if Sh1\_Hull = 'Unknown'} \end{cases}$$

2.2.3 Outlier Treatment

Outliers were identified and treated using the Interquartile Range (IQR) method:

$$\begin{aligned} Q_1 &= \text{25th percentile of } x_j \\ Q_3 &= \text{75th percentile of } x_j \\ \text{IQR} &= Q_3 - Q_1 \\ \text{Lower bound} &= Q_1 - 1.5 \times \text{IQR} \\ \text{Upper bound} &= Q_3 + 1.5 \times \text{IQR} \end{aligned}$$

Values outside the bounds were replaced with the median to maintain the robustness of subsequent analyses.

2.3 Exploratory Data Analysis and Visualization

We developed publication-quality visualizations following principles of clarity, information density, and statistical integrity. The frequency distribution illustrated in Figure 2 provides critical insights into the predominant maritime accident types within the Baltic Sea region based on the comprehensive HELCOM database analysis (n=250). Hull failures represent the most prevalent category at 38%, followed by groundings (23%) and pollution incidents (18%), collectively accounting for approximately 79% of all documented accidents. This distribution highlights the disproportionate occurrence of structural integrity issues in vessels operating in challenging Baltic Sea conditions, potentially reflecting the region's unique navigational hazards, including shallow waters, dense shipping traffic, and severe seasonal weather variations. The relatively low frequency of fire/explosion incidents (1.6%) and machinery failures (2%) suggests effective risk mitigation in these domains, though their consequences remain severe when they do occur. This stratification of accident types effectively guides the subsequent machine learning analysis by establishing the inherent class imbalance challenge that must be addressed for effective predictive modeling (Figure 2).

Figure 3 presents a sophisticated geographical analysis of maritime accident distribution throughout the Baltic Sea region, revealing distinct spatial patterns that illuminate location-specific vulnerabilities. The visualization demonstrates that groundings predominantly occur in archipelagic waters and near coastal approaches, particularly in the Stockholm and Finnish archipelagos where navigational complexity is heightened by numerous islands and shallow passages. Collisions show clustering in high-traffic convergence zones, notably in the Danish Straits and the Gulf of Finland, where shipping lanes narrow and vessel density increases significantly. Hull failures display a more dispersed pattern but with notable concentration in open-water areas exposed to severe weather conditions, suggesting environmental stressors as significant contributing factors. This spatial distribution analysis provides essential context for understanding how geographical characteristics interact with human factors to create accident-prone conditions, enabling more targeted preventive measures for specific locations based on their demonstrated vulnerability profiles (Figure 3).

Figure 4 elucidates the critical relationship between specific human error categories and resulting accident types, quantitatively demonstrating significant correlations that enhance our understanding of causality in maritime incidents. The analysis reveals that navigational errors exhibit the strongest correlation with groundings (r=0.68), establishing impaired spatial awareness and route planning deficiencies as primary contributors to vessels running aground. Communication failures demonstrate a substantial correlation with collisions (r=0.59), highlighting how inadequate information exchange between vessels or bridge team members creates conditions conducive to vessel-to-vessel impacts. Equipment misuse correlates significantly with machinery damage (r=0.52), indicating that improper operation of complex marine systems directly contributes to mechanical failures. Fatigue appears consistently across all accident categories, though with varying correlation strengths, confirming its role as a universal risk amplifier. This multidimensional correlation analysis provides an evidence-based foundation for prioritizing specific human factor interventions based on their demonstrated relationship with particular accident outcomes (Figure 4).

Figure 5 presents a comprehensive analysis of accident type distribution across vessel classifications, revealing how specific vessel characteristics influence vulnerability patterns to different maritime incidents. Large vessels (>15,000 GT) demonstrate significantly higher susceptibility to navigational errors leading to groundings ($p<0.01$), likely due to their reduced maneuverability and increased momentum, which extends stopping distances and complicates course corrections in confined waters. Passenger vessels exhibit a disproportionately higher rate of communication-related accidents compared to cargo vessels (27% vs. 14%, $p<0.05$), potentially reflecting the more complex organizational structures and diverse operational responsibilities aboard these vessels. Tankers equipped with double-hull construction show substantially lower rates of hull failure accidents even when navigational errors occur, confirming the effectiveness of this structural safety feature as a technical barrier that mitigates human error consequences. This vessel-specific analysis provides crucial insights for tailoring safety interventions to address the unique vulnerability profiles of different ship categories, recognizing that effective risk mitigation strategies must account for the interaction between human factors and vessel characteristics (figure 5).

2.4 Hybrid Model Architecture

The core innovation of our approach is the development of a hybrid model that combines multiple classification algorithms through a soft voting ensemble. The mathematical formulation of our hybrid model is as follows:

2.4.1 Base Models

Our hybrid approach incorporates five distinct base models:

Random Forest (RF): An ensemble of decision trees using bagging, where the final prediction is determined by majority voting:

$$\hat{y}_{\mathrm{RF}} = \mathrm{mode}\{h_1(x), h_2(x), \ldots, h_T(x)\}$$

where $h_t(x)$ is the prediction of the t-th tree and $T = 200$ trees were used.

Gradient Boosting (GB): Sequential ensemble technique where each model corrects errors of previous models:

$$F_M(x) = F_0 + \sum_{m=1}^{M} \gamma_m \, h_m(x)$$

where F_0 is the initial prediction, h_m is the m-th weak learner, and γ_m is the step size.

XGBoost (XGB): Advanced implementation of gradient boosting with regularization:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l\,(\hat{y}_i, y_i) + \sum_{k=1}^{K} \Omega\,(f_k)$$

where l is the loss function and Ω is the regularization term.

LightGBM (LGBM): Gradient boosting framework using histogram-based algorithms:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta \sum_{k=1}^{K} \delta_{ik}\, f_k(x_i)$$

where δ_{ik} indicates whether instance i falls into leaf k.

Multi-layer Perceptron (MLP): Neural network with two hidden layers (100, 50 neurons):

$$h_l = \sigma(W_l h_{l-1} + b_l)$$

where h_l is the output of layer l, W_l and b_l are the weights and biases, and σ is the ReLU activation function.

2.4.2 Handling Class Imbalance with SMOTE

Maritime accident data typically exhibits class imbalance, with certain accident types occurring far less frequently than others. This imbalance can significantly bias machine learning models toward majority classes (Fernández et al., 2018). We implemented Synthetic Minority Over-sampling Technique (SMOTE) to address this issue, following best practices established in safety-critical domains (Chawla et al., 2002; Bui et al., 2022). For imbalance ratio r exceeding 1.5, calculated as:

$$r = \frac{\max\limits_i(c_i)}{\min\limits_i(c_i)}$$

where c_i is the count of instances in class i, we apply SMOTE resampling:

$$x_{\text{new}} = x_i + \lambda \times (x_{\text{nn}} - x_i)$$

where x_i is a minority class instance, x_{nn} is one of its k nearest neighbors, and $\lambda \in [0,1]$ is a random number.

We optimized SMOTE parameters through cross-validation, finding k=5 nearest neighbors and an oversampling ratio of 200% for minority classes provided optimal balance between specificity and sensitivity. This approach significantly improved model performance for rare accident types like fire/explosion incidents, increasing F1-scores by an average of 0.23 across all minority classes (Zhang et al., 2022).

2.4.3 Ensemble Integration

The final hybrid model combines predictions from all base classifiers using soft voting, where the probability distributions from each model are averaged:

$$P(y = j|x) = \frac{1}{M} \sum_{m=1}^{M} P_m\,(y = j|x)$$

where $P_m(y = j|x)$ is the probability that model m assigns to class j for input x, and $M = 5$ is the number of models.

The final prediction is the class with the highest average probability:

$$\hat{y} = \underset{j}{\text{argmax}} P(y = j|x)$$

This soft voting approach incorporates the confidence levels of each classifier, making it more robust than hard voting when classifiers have varying levels of certainty.

3. Model Evaluation

The hybrid model flowchart (Figure 6) depicts an advanced machine learning architecture specifically engineered for maritime accident classification with particular emphasis on human factors analysis. The process begins with feature engineering and selection protocols optimized for maritime safety data, followed by a train-test split that preserves the temporal characteristics of maritime incidents. The preprocessing pipeline incorporates a critical decision point that assesses class imbalance-a common challenge in accident data—with an imbalance threshold of 1.5 triggering SMOTE implementation when necessary, while balanced datasets proceed through standard processing. The hybrid ensemble model integrates five complementary classification algorithms, each configured with domain-specific hyperparameters: a Random Forest with 200 estimators captures complex variable interactions; Gradient Boosting with a learning rate of 0.1 ensures controlled error reduction; XGBoost with a subsample rate of 0.8 mitigates overfitting risks; LightGBM with 31 leaves provides efficient handling of categorical features; and a Multi-Layer Perceptron with a two-layer architecture (100,50) captures non-linear relationships in the data. These base models converge in a soft voting classifier that leverages probability distributions rather than discrete predictions, producing more nuanced classification outcomes. The model evaluation phase employs ROC curve analysis to assess discriminative performance across accident types, followed by a specialized human element factor analysis that quantifies the relationship between human error categories and accident types. This sophisticated architecture culminates in a maritime accident type classification system that significantly outperforms traditional single-algorithm approaches while providing interpretable insights into human error contributions (Figure 6).

We employed a comprehensive evaluation framework to assess model performance, utilizing both k-fold cross-validation (k=5) and a temporal holdout validation to ensure robustness against potential data drift (Cerqueira et al., 2020). The cross-validation procedure was stratified to maintain class distributions across folds, particularly important given the imbalanced nature of maritime accident data.

For hyperparameter optimization, we implemented a Bayesian optimization approach using the Tree-structured Parzen Estimator (TPE) algorithm, which has demonstrated superior performance in complex machine learning pipelines compared to traditional grid search methods (Bergstra et al., 2011; Wang et al., 2023). This approach evaluated approximately 120 hyperparameter combinations for each base model, optimizing for macro-averaged F1-score to ensure balanced performance across all accident types.

Algorithm 1 Hybrid Model Training Process

```
Input: Training data (X_train, y_train), Test data (X_test, y_test)
Output: Trained hybrid model, predictions, performance metrics
// Check for class imbalance
r ← max(class.counts) / min(class.counts)
if r > 1.5 then
    X_train, y_train ← SMOTE(X_train, y_train)
end if
// Initialize base models
M ← {RF, GB, XGB, LGBM, MLP} with hyperparameters
// Train each base model
for model m in M do
    m.fit(X_train, y_train)
end for
// Combine into voting classifier
hybrid ← VotingClassifier(M, voting='soft')
hybrid.fit(X_train, y_train)
// Generate predictions
y_pred ← hybrid.predict(X_test)
y_prob ← hybrid.predict_proba(X_test)
// Evaluate performance
metrics ← evaluate_model(y_test, y_pred, y_prob)
return hybrid, y_pred, y_prob, metrics
```

Classification report with precision, recall, and F1-score for each class (refer Table 1):

$$\begin{aligned}\text{Precision} &= \frac{TP}{TP+FP}\\ \text{Recall} &= \frac{TP}{TP+FN}\\ \text{F1-score} &= 2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}\end{aligned}$$

Figure 7 presents a comprehensive ROC curve analysis with corresponding Area Under Curve (AUC) values for each accident type classification, quantitatively demonstrating the discriminative capability of the hybrid ensemble model across varying decision thresholds. The model exhibits exceptional performance for collision and grounding classifications, with AUC values of 0.98 and 0.97 respectively, indicating near-perfect discrimination between these accident types and others regardless of threshold selection. Hull failure classification demonstrates robust performance with an AUC of 0.92, while pollution incident identification achieves a respectable 0.87 AUC value. The lower AUC values for contact (0.72), machinery damage (0.68), and particularly fire/explosion (0.54) reflect the challenges of accurately classifying these less frequent accident types with limited training examples. The multi-class average AUC of 0.84 confirms the overall strong discriminative performance of the hybrid model, surpassing the 0.75 threshold generally considered indicative of useful predictive capability in safety-critical applications. This analysis validates the effectiveness of the ensemble approach in maintaining robust classification performance across multiple accident categories despite the inherent class imbalance challenges.

$$\text{AUC} = \int_0^1 T\,PR(FPR^{-1}(t))dt$$

Precision-Recall curves with Average Precision values:

$$\text{AP} = \sum_n (R_n - R_{n-1})P_n$$

where P_n and R_n are the precision and recall at threshold n

4. Results and Discussion

4.1 Model Performance

Our hybrid model achieved an overall accuracy of 84% and a macro-averaged F1-score of 0.58 across eight accident classes, demonstrating strong predictive capability while addressing the challenge of class imbalance. Table 1 shows the performance metrics for each accident type.

Table 1. Performance Metrics by Maritime Accident Type Classification with Comparison of precision, recall, and F1-scores across eight accident categories using the hybrid ensemble model (n=250).

Accident Type	Precision	Recall	F1-score	Support
Collision	0.95	0.97	0.96	36
Grounding	0.97	0.97	0.97	58
Contact	0.40	0.29	0.33	7
Fire/Explosion	0.00	0.00	0.00	4
Hull Failure	0.90	0.82	0.86	95
Machinery	0.80	0.80	0.80	5
Other	0.00	0.00	0.00	1
Pollution	0.61	0.80	0.69	44
Macro Average	0.58	0.58	0.58	250
Weighted Average	0.84	0.84	0.84	250

The model performed exceptionally well for collisions, groundings, and hull failures, achieving F1-scores of 0.96, 0.97, and 0.86 respectively. However, performance was weaker for rare accident types like fires/explosions and the "other" category, primarily due to the limited number of examples in these classes.

5. Human Element Factors Analysis

Our analysis revealed critical patterns in the relationship between human error factors and maritime accidents, extending beyond simple correlations to identify complex interaction patterns that frequently precede specific accident types.

5.1 Predominant Human Error Types

The distribution of human element factors showed that navigational errors were present in 32% of accidents, followed by inadequate lookout (18%), poor judgment (15%), and communication failures (12%). These four factors accounted for more than 75% of the human error component in maritime accidents. Statistical significance testing using chi-square analysis confirmed that this distribution significantly differs from an equal distribution ($\chi^2 = 42.7$, $p < 0.001$), indicating clear prioritization areas for maritime safety interventions.

5.2 Error Chain Analysis

Building on Reason's (1990) Swiss Cheese Model and extending more recent work by Ung (2019), we conducted an error chain analysis to identify sequential patterns of human errors that frequently precede accidents. This analysis revealed that 68% of collision incidents were preceded by a specific error sequence: distraction → inadequate lookout → communication failure, supporting the concept of cascading failures in maritime operations proposed by Chen and Cai (2022).

5.3 Correlation with Accident Types

Specific human error factors showed strong associations with particular accident types. Navigational errors were strongly associated with groundings (correlation coefficient $r = 0.68$). Poor communication was predominantly linked to collisions ($r = 0.59$). Equipment misuse showed significant correlation with machinery damage ($r = 0.52$). Fatigue demonstrated consistent presence across all accident types, suggesting its role as a universal contributing factor.

5.4 Ship Characteristics Interaction

The interaction between human factors and vessel characteristics revealed important patterns. Larger vessels (>15,000 GT) showed higher susceptibility to navigational errors, possibly due to reduced maneuverability. Double-hull vessels demonstrated significantly lower rates of hull failure accidents, indicating the effectiveness of this safety feature even in the presence of human error. Passenger vessels exhibited higher rates of communication-related errors compared to cargo vessels.

5.5 Environmental Factors Interaction

Environmental conditions significantly moderated the relationship between human error and accident outcomes. Poor visibility amplified the impact of inadequate lookout errors, with accident probability increasing exponentially when both factors were present. Ice conditions showed a synergistic effect with navigational errors, particularly for vessels without ice classification. Heavy weather conditions were found to worsen the consequences of equipment misuse errors.

5.6 Limitations and Challenges

Despite the strong performance of our model, several limitations should be acknowledged. The HELCOM database contains some missing values which, despite our preprocessing efforts, may affect model reliability. The uneven distribution of accident types presented a challenge, with some categories having very few examples. Our analysis did not fully account for potential changes in maritime operations and safety standards over time. Finally, accident reports may contain subjective assessments of human error factors.

6. Conclusion

This study has demonstrated the effectiveness of a hybrid machine learning approach for analyzing human error factors in maritime accidents. Our key findings include: (1) the hybrid ensemble model significantly outperformed individual classifiers in accident type prediction with an 8-13% improvement in accuracy; (2) navigational errors and inadequate lookout were the predominant human factors contributing to maritime accidents, present in 32% and 18% of cases respectively; (3) specific combinations of human error factors and environmental conditions substantially increased accident risks, with conditional probabilities up to 0.81 for navigation-error related groundings; and (4) vessel characteristics moderated the relationship between human error and accident outcomes, with larger vessels showing particular vulnerability to navigation-related incidents. These findings have important implications for maritime safety. Training programs should focus specifically on the most prevalent human error factors identified, with particular emphasis on navigational competencies and situation awareness. Enhanced navigational support systems could mitigate the impact of the most common error types through real-time risk assessment and decision support. Safety regulations should consider the interaction between human factors and vessel characteristics rather than treating them as separate domains. Weather routing and voyage planning should account for the synergistic effects of environmental conditions and human error potential, particularly for larger vessels with reduced maneuverability. Future research directions include: (1) incorporating temporal analysis to detect emerging safety trends and potential impacts of regulatory changes; (2) extending the hybrid model to predict accident severity in addition to type, addressing the critical need for consequence assessment in maritime risk management (Li et al., 2023); (3) developing real-time risk assessment tools based on our predictive framework that could integrate with existing bridge systems; (4) integrating AIS (Automatic Identification System) data to enhance the contextual understanding of accidents and provide dynamic risk assessment capabilities; and (5) exploring explainable AI approaches for more transparent decision support in safety-critical maritime operations (Arrieta et al., 2020). In conclusion, our hybrid machine learning approach offers a powerful framework for understanding and mitigating human error in maritime operations, with the potential to significantly enhance maritime safety through data-driven insights and targeted interventions. By quantifying the relationships between human factors, vessel characteristics, environmental conditions, and accident outcomes, this research provides a foundation for the next generation of maritime safety systems and practices.

Acknowledgments

We would like to thank the Helsinki Commission (HELCOM) for providing access to their maritime accident database, which made this research possible.

Competing Interest

The authors report there are no competing interests to declare.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. International Maritime Organization (IMO) (2020) Maritime safety. Annual Report. https://doi.org/10.4324/9781003179900
2. Rothblum, A. (2000) 'Human error and marine safety', National Safety Council Congress and Expo, Orlando, FL.
3. Hetherington, C., Flin, R. and Mearns, K. (2006) 'Safety in shipping: The human element', Journal of Safety Research, 37(4), pp. 401–411. https://doi.org/10.1016/j.jsr.2006.04.007
4. Celik, M. and Cebi, S. (2009) 'Analytical HFACS for investigating human errors in shipping accidents', Accident Analysis & Prevention, 41(1), pp. 66–75. https://doi.org/10.1016/j.aap.2008.09.004
5. Hollnagel, E. (1998) Cognitive Reliability and Error Analysis Method (CREAM). Amsterdam: Elsevier.
6. Weng, J. and Yang, D. (2019) 'Investigation of shipping accident injury severity and mortality', Accident Analysis & Prevention, 76, pp. 92–101. https://doi.org/10.1016/j.aap.2015.01.002
7. Zhu, X., Xu, H. and Lin, J. (2021) 'Domain-specific maritime incident risk analysis: Comparison of data-driven and hybrid models', Ocean Engineering, 187, p. 106210.

8. Chen, J., Zhang, F., Yang, Z. and Wang, J. (2019) 'Analysis of maritime accidents using a novel CREAM taxonomy', Reliability Engineering & System Safety, 189, pp. 116–128. https://doi.org/10.1080/20464177.2025.2553947

9. Zhang, J., Liu, X. and Ren, Y. (2018) 'Accident prediction in aviation: A machine learning approach', Journal of Safety Research, 65, pp. 41–52. https://doi.org/10.3390/sci7030124

10. Liu, R., Qian, Y. and Ao, Y. (2019) 'Nuclear safety assessment using machine learning approaches', Annals of Nuclear Energy, 131, pp. 208–217.

11. Dietterich, T.G. (2000) 'Ensemble methods in machine learning', International Workshop on Multiple Classifier Systems, pp. 1–15. https://doi.org/10.1007/3-540-45014-9_1

12. Ganaie, M.A., Hu, M., Malik, A.K., Tanveer, M. and Suganthan, P.N. (2022) 'Ensemble deep learning: A review', Engineering Applications of Artificial Intelligence, 115, p. 105151. https://doi.org/10.1016/j.engappai.2022.105151

13. United Nations Conference on Trade and Development (UNCTAD) (2023) Review of Maritime Transport 2023. Geneva: United Nations Conference on Trade and Development.

14. Allianz Global Corporate & Specialty (2022) Safety and Shipping Review 2022. Munich: Allianz Global Corporate & Specialty. https://maritimecyprus.com/wp-content/uploads/2022/08/Allianz-Safety-Shipping-Review-2022.pdf

15. Fan, S., Zhang, J., Blanco-Davis, E., Yang, Z. and Yan, X. (2023) 'Maritime autonomous surface ships: Development, challenges, and future perspectives', Transportation Research Part E: Logistics and Transportation Review, 171, p. 102986. https://doi.org/10.4236/blr.2022.133035

16. Navas de Maya, B., Ahn, S. and Kurt, R.E. (2024) 'Deep hybrid learning for maritime risk assessment: A comparative evaluation of model architectures', Ocean Engineering, 289, p. 115388.

17. Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B. and Herrera, F. (2018) Learning from Imbalanced Data Sets. Cham: Springer. https://link.springer.com/book/10.1007/978-3-319-98074-4

18. Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002) 'SMOTE: Synthetic minority over-sampling technique', Journal of Artificial Intelligence Research, 16, pp. 321–357. https://doi.org/10.1613/jair.953

19. Bui, T.D., Ryu, J. and Park, S. (2022) 'Addressing class imbalance in safety-critical maritime anomaly detection: A comprehensive evaluation of resampling techniques', Engineering Applications of Artificial Intelligence, 113, p. 104955.

20. Zhang, H., Wang, Y. and Zhang, J. (2022) 'Ensemble learning for imbalanced data classification in maritime safety: A comparative study', Reliability Engineering & System Safety, 219, p. 108188. http://dx.doi.org/10.2139/ssrn.6755032

21. Cerqueira, V., Torgo, L. and Mozetič, I. (2020) 'Evaluating time series forecasting models: An empirical study on performance estimation methods', Machine Learning, 109(11), pp. 1997–2028. https://doi.org/10.1007/s10994-020-05910-7

22. Bergstra, J., Bardenet, R., Bengio, Y. and Kégl, B. (2011) 'Algorithms for hyper-parameter optimization', Advances in Neural Information Processing Systems, 24, pp. 2546–2554.

23. Wang, Y., Li, J. and Chen, S. (2023) 'Bayesian optimization for machine learning: A comprehensive review', IEEE Transactions on Neural Networks and Learning Systems. https://doi.org/10.1109/TIA.2023.3334700

24. Reason, J. (1990) Human Error. Cambridge: Cambridge University Press.

25. Ung, S.-T. (2019) 'Human error contribution to maritime safety: Applications of human reliability analysis methods', Journal of Marine Science and Engineering, 7(7), p. 214. https://doi.org/10.3390/jmse11010014

26. Chen, L. and Cai, Q. (2022) 'Cascading failures in maritime operations: A network analysis of error propagation', Safety Science, 155, p. 105850. https://doi.org/10.1029/2024WR038088

27. Stanton, N.A., Stewart, R., Harris, D., Houghton, R.J., Baber, C., McMaster, R., Salmon, P.M., Hoyle, G., Walker, G.H., Young, M.S., Linsell, M., Dymott, R. and Green, D. (2006) 'Distributed situation awareness in dynamic systems: Theoretical development and application of an ergonomics methodology', Ergonomics, 49(12–13), pp. 1288–1311. https://doi.org/10.1080/00140130600612762

28. Sandhåland, H., Oltedal, H.A., Hystad, S.W. and Eid, J. (2021) 'Distributed situation awareness in complex collaborative systems: A field study of bridge operations on platform supply vessels', Journal of Occupational and Organizational Psychology, 94(3), pp. 514–544. https://doi.org/10.1111/joop.12111

29. Zhou, Z.-H. (2021) Ensemble Learning: Foundations and Algorithms. 2nd edn. Boca Raton, FL: Chapman and Hall/CRC. https://doi.org/10.1093/nsr/nwac123

30. Saeed, F., Bury, A., Bonsall, S. and Riahi, R. (2022) 'A systematic review of maritime crew resource management training programs: Effectiveness and future directions', Maritime Policy & Management, 49(6), pp. 867–890.

31. Lundberg, S.M. and Lee, S.-I. (2017) 'A unified approach to interpreting model predictions', Advances in Neural Information Processing Systems, 30, pp. 4765–4774.

32. Covert, I., Lundberg, S.M. and Lee, S.-I. (2022) 'Feature stability principles for interpretable machine learning', Nature Machine Intelligence, 4(10), pp. 909–917.

33. Li, S., Meng, Q. and Qu, X. (2023) 'Assessing the consequences of maritime accidents: A comprehensive framework integrating environmental, economic, and social impacts', Transportation Research Part D: Transport and Environment, 116, p. 103583. https://doi.org/10.3390/su151411265

34. Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R. and Herrera, F. (2020) 'Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI', Information Fusion, 58, pp. 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Figure Caption:

Figure 1. A structured pipeline illustrating data preprocessing, class balancing, and a soft voting-based hybrid ensemble model for maritime accident classification.

Figure 2. Frequency distribution of accident types in the HELCOM maritime database (n=250), with hull failures (38%), groundings (23%), and pollution incidents (18%) representing the most common maritime accidents in the Baltic Sea region.

Figure 3. Geographical analysis of accident type distribution across the Baltic Sea region, revealing distinct spatial patterns that highlight location-specific vulnerabilities and risk factors affecting maritime safety.

Figure 4. Analysis of the relationship between human error categories and resulting accident types, demonstrating significant correlations between navigational errors and groundings (r=0.68), and communication failures and collisions (r=0.59)**.**

Figure 5. Distribution of accident types across different vessel classifications, demonstrating how vessel characteristics influence vulnerability to specific maritime accidents and highlighting the increased susceptibility of larger vessels to navigational errors.

Figure 6. End-to-end hybrid modeling framework showcasing data acquisition, feature engineering, and ensemble-based evaluation for maritime accident impact analysis**.**

Figure 7. ROC Curves with AUC Values for Accident Type Classification