Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira, Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p27786-cd

A TOPSIS-Based Regression Fusion Model for Prediction in Maritime Evacuation

Xinjian Wang

Navigation College, Dalian Maritime University, China; Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, United Kingdom. E-mail: wangxinjian@dlmu.edu.cn

Zekai Cui

Navigation College, Dalian Maritime University, China. E-mail: kai7730@foxmail.com

Zhiwei Zhang

Navigation College, Dalian Maritime University, China. E-mail: dmuzhangzhiwei@dlmu.edu.cn

Yuhao Cao

Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, United Kingdom. E-mail: Y.Cao@ljmu.ac.uk

Zaili Yang

Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, United Kingdom. E-mail: Z.Yang@ljmu.ac.uk

Abstracts: In emergency evacuation scenarios, obtaining accurate evacuation time is crucial for developing effective evacuation plans and supporting emergency response. This study aims to construct a framework for evacuation time prediction in maritime emergencies. It uses a Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) based Regression Fusion (TRF) model to fuse predictions from a set of advanced Machine Learning (ML) models, and evaluates its performance using multi-criteria decision making with an emphasis on prediction accuracy and overall comprehensiveness. First, an agent-based modelling technique is used to simulate the evacuation process, and a multi-scenario dataset is generated from the simulation outputs. Second, seven ML models are trained for predicting evacuation time, and four complementary evaluation metrics are used to compare model performance. Third, TOPSIS is used to calculate integrated scores from the multi metric results, and these scores are converted into fusion weights for TRF. Finally, TRF fuses the predictive time from the seven ML models into a single integrated output, which can retain complementary strengths across models and improve overall predictive performance. The results reveal that TRF achieves lower prediction errors and more stable performance than the individual models across scenarios. This study provides an insightful scientific basis for using integrated ML prediction tools to support evacuation planning and safety assessment for maritime evacuation.

Keywords: Maritime safety, Emergency evacuation, Passenger ship, Evacuation simulation, Machine learning, TOPSIS.

1. Introduction

Unlike land-based buildings or vehicles, large passenger ships are characterized by complex internal spaces and a high level of human involvement in their operations, which creates substantial challenges for emergency response and evacuation. Once an accident occurs, it may result in large-scale casualties. Reported statistics indicate that only 2.64% and 10.25% of individuals, respectively, successfully evacuated to a safe area in maritime evacuation (Zhang et al., 2025). For example, the capsizing and sinking of the *Orient Star* in 2015 led to 442 deaths, and the *MV Nyerere* disaster in 2018 caused 228 deaths. In both tragic accidents, the situation escalated rapidly and comprehensive evacuation could not be carried out in time, resulting in severe loss of life. These facts reveal that accurate and effective evacuation planning is essential and requires significant improvement in maritime evacuation. Existing studies widely note that accurate prediction of evacuation time supports the development and effectiveness evaluation of evacuation plans (Zhang et al., 2025). In addition, the International Maritime Organization (IMO) treats evacuation time as one of the key indicators for assessing evacuation performance in its evacuation analysis guidance for passenger ships (MSC, 2016). Therefore, accurate prediction of human evacuation time in maritime evacuation is critical for developing effective evacuation plans.

To support research on human evacuation time in maritime evacuation, both academia and industry have carried out extensive work. Among these efforts, the European Union's SAFEGUARD project is a representative initiative, providing scenario-based data that underpin studies on maritime evacuation. To achieve accurate prediction in evacuation-time tasks, Machine Learning (ML) models have been increasingly applied to time prediction in maritime evacuation due to their ability to capture implicit relationships and handle interacting factors (Zhang et al., 2025). However, in multi-scenario prediction tasks, different ML models often have different strengths in accuracy, and the performance of a single model can also fluctuate with changes in the training split or the scenario composition. To reduce this uncertainty, existing studies commonly adopt ensemble learning to improve predictive performance, with the core idea of exploiting differences between models and combining them in a complementary way (Yang et al., 2025). In recent years, the continued development of ensemble learning, particularly its advantages in improving overall generalization and robustness, has made it feasible to obtain better prediction performance through ensemble learning (Wang et al., 2025).

Therefore, this study aims to design a Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) based Regression Fusion (TRF) model that can integrate multiple machine-learning models by employing TOPSIS-based multi-criteria decision-making to weight their individual contributions, resulting in a unified tool for predicting maritime evacuation times.

The remainder of this study is organized as follows. Section 2 presents the methodology. Section 3 indicates and discusses the results. Section 4 concludes this study and outlines future work.

2. Methodology

2.1. *Evacuation simulation setting*

AnyLogic is a simulation platform that supports multiple modelling paradigms, including agent-based modelling, and it can represent pedestrian evacuation processes with individual interactions and congestion dynamics. In this study, an agent-based evacuation model is developed in AnyLogic for maritime evacuation. The model set-up is based on the public SGVDS1 dataset released by the European SAFEGUARD project. Pedestrian agents evacuate from recorded initial areas to four muster stations via the ship's passageways and stairways. Individual attributes and walking speeds are assigned in accordance with IMO MSC.1/Circ.1533 (MSC, 2016).

Prior to model set-up and scenario construction, the key influencing factors defining the simulation scenarios were identified. The selection of these factors follows a three-step rationale. First, it aligns with the IMO evacuation analysis guidelines, ensuring that all inputs can be specified at the scenario set-up stage and remain consistent with regulatory practice (MSC, 2016). Second, ship-specific operational constraints were considered to ensure that the factor ranges are realistic for the target ship. For example, the evacuation population was set within the ship's feasible capacity, the heel and trim ranges were limited to operationally meaningful values, and stair width was treated as a ship-configuration parameter. Finally, additional factors drawn from the literature on human evacuation from passenger ships, particularly those related to crowd organization and onboard support, such as small-group behavior and the deployment of seafarers to assist evacuation. Based on this structured selection, the simulation scenarios were defined using the factors listed in Table 1.

Table 1. Influencing factors and scenario settings for evacuation simulation.

Influencing factors	Scene setting	Quantity	References
Total number of evacuees	400, 500, 600, 700, 800, 900, 1000, 1100, 1200	9	(Zhang et al., 2025)

Heel angle (°)	0, 5, 10, 15	4	(Fang et al., 2022)
Trim angle (°)	−20, −15, −10, −5, 0, 5, 10, 15, 20	9	(Fang et al., 2022)
Width of stairs	1.5, 2.0, 2.5, 3.0	4	(Zhang et al., 2025)
Group behaviour	Yes (with small groups), No (individual evacuation)	2	(Wang et al., 2020)
Number of guide seafarers	0, 15, 20	3	(Fang et al., 2022)

The influence of each selected factor on evacuation dynamics can be summarized as follows. Population size shapes crowding and queue formation, which directly affects corridor and stair throughput and therefore evacuation time (Fang et al., 2022). Heel and trim can reduce walking stability and speed, especially in narrow corridors and on stairs, and may alter overall evacuation dynamics (Fang et al., 2022). Stair width directly determines vertical bottleneck capacity; all else being equal, wider stairs generally allow higher flow rates and shorter evacuation times. In addition, small-group movement often involves waiting for companions and joint decisions, which changes arrivals at bottlenecks and local congestion patterns, and can influence evacuation completion time (Wang et al., 2020). Finally, seafarers can support wayfinding and crowd organization during the assembly and movement phases, helping to reduce hesitation and disorder and, in turn, improve evacuation efficiency.

2.2. *Baseline ML models and evaluation metrics*

To ensure diversity and complementarity among the base models for fusion, three representative families of ML models tare selected as baselines. Tree-based ensembles include Random Forest (RF), Gradient Boosting Decision Tree (GBDT), eXtreme Gradient Boosting (XGB), and Light Gradient Boosting Machine (LGBM). Kernel-based models include Support Vector Regression (SVR) and Kernel Ridge Regression (KRR). A Multi-layer Perceptron (MLP) is also included. These models embody different inductive preferences for capturing non-linear relationships, learning feature interactions, and representing complex functions, which provides complementary strengths and supports a more robust fusion framework. The core mechanism and main strengths of each model are summarized in Table 2.

Table 2. Baseline ML models used in this study.

Name	Core	Highlights	References
RF	Tree bagging	Robust generalization	Berhanu et al. (2024)
GBDT	Boosted trees	Strong interactions	Parisi et al. (2025)
XGB	Regularized boosting	Strong accuracy	Yang et al. (2025)
LGBM	Efficient boosting	Fast training	Li et al. (2024)
SVR	Kernel regression	Strong nonlinearity	Su et al. (2025)
KRR	Kernel ridge	Stable fitting	Guzman-Chavez et al. (2025)
MLP	Feed-forward network	Flexible representation	Su et al. (2025)

To comprehensively evaluate the performance of the selected ML models in predicting evacuation time, four key metrics are employed: Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (sMAPE). For all metrics, smaller values indicate better predictive performance. These metrics are chosen to ensure a comprehensive assessment of the model's prediction accuracy and stability in large-scale practical scenarios (Zhang et al., 2025).

After defining the baseline ML models and the evaluation metrics, this study specifies a unified training, tuning, and validation protocol to ensure fair comparison across models and a reproducible evaluation framework for the subsequent integrated scoring and fusion steps. This study uses the evacuation simulation dataset for maritime evacuation, comprising 7776 samples. The data are split into a training set (TRAIN), an internal validation set (VAL), and a test set (TEST) using a 6:2:2 ratio, resulting in $N_{\text{train}} = 4666$, $N_{\text{val}} = 1555$, and $N_{\text{test}} = 1555$ samples. The seven baseline ML models are fitted on TRAIN. Hyperparameters are selected by random search and evaluated via five-fold cross-validation within TRAIN to identify the best parameters setting. For models that require feature scaling (SVR, KRR, and MLP), scaling parameters are computed using the training data in each fold and then applied to the corresponding validation data to ensure a consistent and reproducible pre-processing procedure. The evaluation metrics are then computed on VAL and used as inputs for the integrated scoring and fusion procedures in Sections 2.3 and 2.4, while final performance is reported on TEST.

2.3. *TOPSIS-based integrated scoring*

TOPSIS is a widely used multi-criteria decision-making method (Wang et al., 2025). It constructs a positive-ideal reference point and a negative-ideal reference point, and then derives an integrated score based on each candidate's relative closeness to these two references, thereby enabling consistent comparison under multiple metrics. This study follows the standard TOPSIS framework and combines the multi-metric validation error evaluations on the VAL into a TOPSIS integrated score for each baseline ML model, which is then used as the input for TRF weight construction.

Specifically, this study assembles the four errors metrics of the seven baseline ML models on VAL into a normalized decision matrix. To avoid introducing subjective preferences, equal weights are assigned to all metrics, with the weight vector $w = (1/4, 1/4, 1/4, 1/4)$. Each entry of the normalized decision matrix is multiplied by its corresponding weight to obtain a weighted normalized matrix, ensuring that the contribution of each metric to the integrated evaluation is proportional to its weight. For each evaluation metric, this study takes the minimum and maximum values across the seven ML models to define Positive Ideal Solution (PIS) and the Negative Ideal Solution (NIS), as shown in Eq. (1) and (2):

$$\mathrm{PIS} = \left(\min_i \mathrm{MSE}_i\,,\ \min_i \mathrm{MAE}_i\,,\ \min_i \mathrm{MAPE}_i\,,\ \min_i \mathrm{sMAPE}_i\right) \tag{1}$$

$$\mathrm{NIS} = (\max_i \mathrm{MSE}_i\,,\ \max_i \mathrm{MAE}\,,\ \max_i \mathrm{MAPE}_i\,,\ \max_i \mathrm{sMAPE}_i) \tag{2}$$

where i denotes the i^{th} ML model ($i = 1,2,\dots,7$); MSE_i, MAE_i, MAPE_i, and sMAPE_i are the corresponding overall metric values of model i computed on VAL. The operators $\min_i$ and $\max_i$ take the minimum and maximum values over the seven ML models, respectively.

This study computes two distances for each model i, as defined in Eq. (3) and (4).

$$score_{i,min} = \sqrt{\begin{gathered}(MSE_i^w - MSE_{min})^2 + (MAE_i^w - MAE_{min})^2 + \\ (MAPE_i^w - MAPE_{min})^2 + (sMAPE_i^w - sMAPE_{min})^2\end{gathered}} \tag{3}$$

$$score_{i,max} = \sqrt{\begin{gathered}(MSE_i^w - MSE_{max})^2 + (MAE_i^w - MAE_{max})^2 + \\ (MAPE_i^w - MAPE_{max})^2 + (sMAPE_i^w - sMAPE_{max})^2\end{gathered}} \tag{4}$$

where, $score_{i,\min}$ is the Euclidean distance from model i to the PIS, and $score_{i,\max}$ is the Euclidean distance from model i to the NIS. The metric term with the superscript *w* denotes the weighted normalized value, obtained after normalization and subsequent weighting. The weighted normalized values used to calculate these distances.

This study distinguishes $score_{i,\min}$ and $score_{i,\max}$ because TOPSIS evaluates each model by considering both closeness to the ideal reference and separation from the worst reference. Based on these two distances, the TOPSIS integrated score is defined in Eq. (5) to quantify the model's relative closeness between the ideal and worst reference solutions.

$$\mathrm{TOPSIS\_score}_i = \frac{score_{i,\max}}{score_{i,\min} + score_{i,\max}} \tag{5}$$

where the definition yields a value in [0,1]. The model closer to the PIS has a smaller $score_{i,\min}$, and model farther from the NIS has a larger $score_{i,\max}$; therefore, a larger $\mathrm{TOPSIS\_score}_i$

indicates better overall performance under the selected metrics.

2.4. *TRF fusion weight construction and fused prediction*

After obtaining the integrated TOPSIS score for each baseline ML model in Section 2.3, this study constructs TRF to convert the integrated scores into fusion weights and to perform weighted regression fusion of the base-model predictions. This study normalizes the integrated scores of the seven ML models to obtain the fusion weights α_i, as defined in Eq. (6). Here, α_i denotes the fusion weight of model i, and M is the number of models.

$$\alpha_i = \frac{\text{TOPSIS\_score}_i}{\sum_{j=1}^{M} \text{TOPSIS\_score}_k}, i = 1,2,\dots,M \quad (6)$$

where $M = 7$. This normalization yields a non-negative weight vector whose elements sum to 1, ensuring that each model contributes to the fusion in proportion to the model's integrated score.

The TRF fused prediction is then defined in Eq. (7) as a weighted sum of the model predictions.

$$\hat{y}_j^{TRF} = \sum_{i=1}^{M} \alpha_i \ \hat{y}_{ij}, j = 1,2,\dots,N \quad (7)$$

where $\hat{y}_j^{TRF}$ is the TRF fused prediction of evacuation time for sample j, $\hat{y}_{ij}$ is the prediction for sample j produced by baseline ML model i, α_i is the fusion weight obtained by normalizing the TOPSIS integrated score, M is the number of baseline ML models, and N is the number of samples to be predicted.

2.5. *Model interpretability*

To make the TRF model's reasoning more transparent, SHapley Additive exPlanations (SHAP) are used for quantification. SHAP is a post-hoc, model-agnostic interpretability method that attributes each prediction to additive feature contributions. In addition, SHAP has been widely used to enhance the interpretability of machine learning models (Wang et al., 2025; Zhang et al., 2025). Therefore, SHAP is applied at the sample level to illustrate the direction and relative magnitude of each input factor's contribution to an individual prediction and to link the model output to interpretable driving factors.

3. Results

This section presents the key results derived from Section 2. Section 3.1 firstly reports the TOPSIS integrated scores obtained on the VAL and the resulting TRF fusion weights. Section 3.2 then evaluates the predictive performance of TRF on the TEST. Finally, an interpretability analysis is provided in Section 3.3 to clarify the influence patterns of the key factors.

3.1. TOPSIS integrated scores and TRF fusion weights

To compare candidate ML models in a stable manner without relying on a single error metric, this study computes the TOPSIS relative closeness on the internal VAL using metrics, and then maps the integrated performance to the fusion weights of TRF. The TOPSIS relative closeness on VAL and the corresponding TRF fusion weights are summarized in Fig. 1.

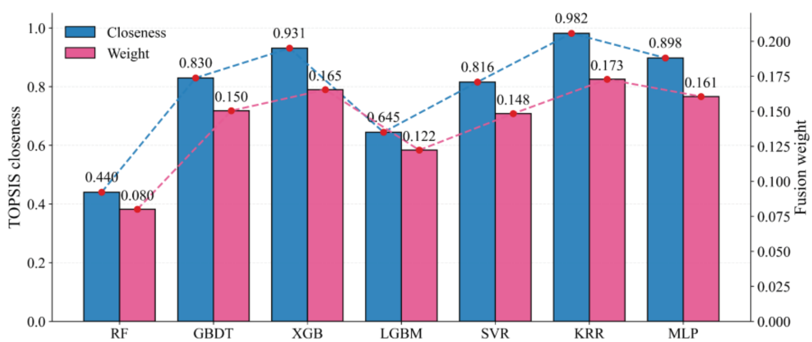

Fig. 1 TOPSIS closeness and derived TRF fusion weights.

In terms of integrated performance, KRR and XGB achieve the highest relative closeness (0.982 and 0.931), indicating the best overall trade-off across the error metrics. MLP also shows a high integrated score (0.898). GBDT and SVR form the second tier (0.830 and 0.816), whereas LGBM is lower (0.645) and RF is the lowest under the current VAL (0.440). Since the relative closeness is bounded within $[0,1]$, a larger value means being closer to the ideal solution and farther from the negative ideal solution, hence implying stronger integrated performance.

Based on these integrated scores in Fig. 1, the models' relative contributions in the fusion stage are represented by mapping the closeness results to TRF fusion weights, so that models with better integrated performance receive larger weights in the final prediction. Accordingly, KRR, XGB, and MLP obtain the largest weights (0.173, 0.165, and 0.161), followed by GBDT and SVR (0.150 and 0.148). LGBM receives a smaller weight (0.122), while RF is assigned the minimum weight (0.080) due to its lowest integrated score. Overall, the weight profile is moderately differentiated, which preserves performance-aware emphasis while preventing a single model from dominating the fused output, thereby providing a stable and reproducible basis for the subsequent TRF fused evaluation.

3.2. TRF performance evaluation on the TEST

Following the TOPSIS-based integrated scoring and the resulting fusion-weight assignment in Section 3.1, this section validates the predictive performance of TRF on the TEST. To highlight the predictive accuracy and prediction consistency of TRF in a clear manner, this study reports a quantitative comparison table, as shown in Table 3, together with a prediction–target scatter plot of TRF, as shown in Fig. 2.

Table 3. Performance comparison on the test dataset.

Model	MSE	MAE	MAPE	sMAPE
RF	614.1648	18.2882	5.3436	5.3516
GBDT	552.0523	17.5554	5.1290	5.1402
XGB	552.7589	17.6060	5.1666	5.1790
LGBM	586.6949	17.9907	5.2510	5.2618
SVR	556.9769	17.5654	5.1321	5.1652
KRR	**544.5417**	17.4559	5.1309	5.1435
MLP	552.8709	17.4424	5.1068	5.1332
TRF	548.7542	**17.4410**	**5.1032**	**5.1208**

As summarized in Table 3, TRF achieves the best performance on three integrated evaluation measures, namely MAE, MAPE, and sMAPE (17.441, 5.103, and 5.121, respectively), indicating its higher overall predictive accuracy on the TEST. Although TRF does not yield the minimum MSE, it is very close to the best single model KRR and outperforms most other baselines. These results suggest that TRF delivers a stronger overall predictive accuracy rate: by integrating the outputs of multiple base models with data-driven weights, TRF can leverage complementary strengths from different learners and hence better capture useful information across samples and error patterns, making the final prediction closer to the targets. As a result, TRF is more likely to achieve superior performance under multiple evaluation measures over that relying on a single model, highlighting the benefit of the proposed fusion scheme.

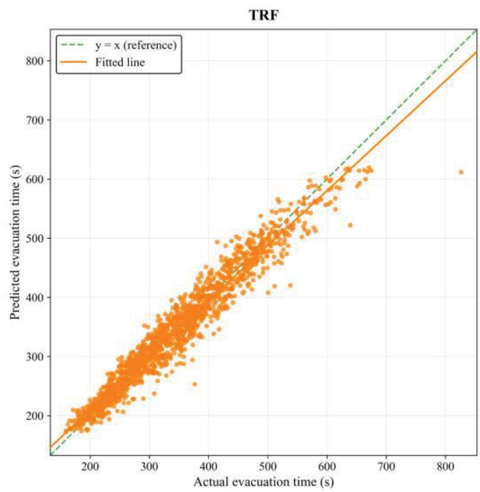

Fig. 2 Scatter plot of TRF predictions against the test targets.

As shown in Fig. 2, at the sample level, TRF shows strong agreement between predictions and targets. The scatter points generally follow the $y = x$ reference line, and the fitted line remains close to it, indicating limited overall bias and a stable prediction trend. Importantly, this alignment is maintained across different evacuation-time ranges. Therefore, the advantage of TRF is not only reflected in the aggregated performance results but is also clearly visible from the sample-wise comparison.

3.3. Model interpretability results

Since the TRF model is a fusion model, its feature importance reflects the combined contributions from the base models and their fusion weights. To make the model's reasoning clearer, SHAP is used to quantify, at the sample level, the direction and relative magnitude of each input factor's contribution to an individual prediction, thereby linking the output to interpretable driving factors. For illustration, the KRR sub-model is used as a representative case because it has the largest fusion weight and strong predictive performance, and the SHAP summary plot is reported in Fig. 3.

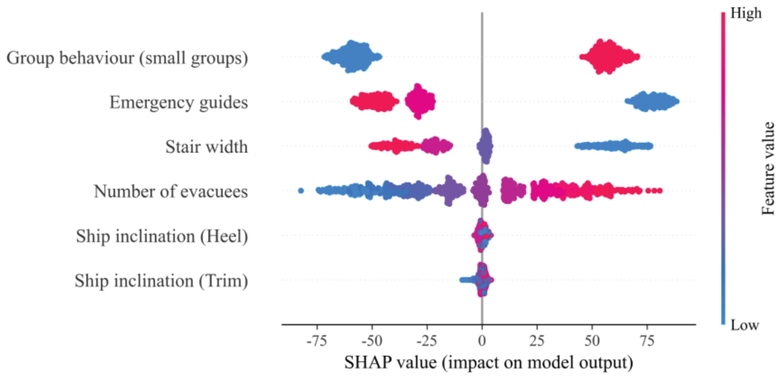

Fig. 3 Optimal sub-model (KRR) SHAP feature importance plot.

The results in Fig. 3 indicate that factors related to group behavior provide the strongest explanation for evacuation time prediction. The small group factor shows a consistent directional pattern: when small groups are absent, the model tends to predict a shorter evacuation time, whereas more pronounced small-group behavior more often contributes to longer evacuation times. This aligns with the mechanisms of maritime evacuation, where individuals moving alone or in smaller units are typically more efficient, while group movement can involve waiting for slower members and can increase local congestion at bottlenecks. Previous studies also indicate that small-group behavior reduces evacuation efficiency and that group initiation and wayfinding often take longer (Ren et al., 2022). Evacuee scale shows a similarly clear trend: as the number of evacuees increases, contributions more often increase the predicted evacuation time, reflecting crowding and competition for limited capacity (Li et al., 2025). Variables linked to organization and capacity also follow intuitive trends: a larger number of seafarers generally correspond to shorter predicted evacuation times (Lim et al., 2023), and wider stairs tend to shorten evacuation times by increasing bottleneck capacity. These widely accepted findings in the literature are largely supported by experiments and mechanism-based modelling, such as cellular automata, which further supports the effectiveness and credibility of TRF for predicting evacuation time for maritime evacuation.

4. Conclusions

This study develops a TRF model for time prediction in maritime evacuation. The model uses TOPSIS to derive fusion weights from multiple evaluation metrics and combines the outputs of seven baseline ML models into a single prediction. By assigning a larger contribution to models with stronger overall performance, it can integrate complementary strengths across models and provide more consistent overall predictions than relying on a single model under the adopted metrics. This work offers a practical and auditable approach for model selection and ensemble prediction in maritime evacuation, supporting more informed evacuation planning and safety assessment.

Despite these contributions, several limitations remain. The current fusion uses a limited set of evaluation metrics and may not reflect other performance aspects under different application needs, and the pool of candidate ML models can be expanded. Future work will extend both the model set and the weight-determination strategy, and will broaden the prediction targets and evaluation metrics, for example by considering evacuation time together with density-related risk. These extensions are expected to improve the practical value of the proposed approach by supporting model evaluation and fusion across a wider range of operational requirements, thereby strengthening the technical basis for evidence-informed evacuation planning and safety management in real maritime operations.

Acknowledgements

The authors gratefully acknowledge support from the National Natural Science Foundation of China [Grant No. 52571403, 52101399, 52501439], International Exchanges program of Royal Society, UK [Grant No. IEC\NSFC\233298], the MarRI-UK Core Project [Grant No. MarRI-UK_CORE_2024_02], and the European Union's Horizon 2020 research and innovation programme (TRUST CoG 2019 864724).

References

Berhanu, Y., D. Schröder, B.T. Wodajo, and E. Alemayehu (2024). Machine learning for predictions of road traffic accidents and spatial network analysis for safe routing on accident and congestion-prone road networks. *Results in Engineering* 23, 102737.

Fang, S., Z. Liu, X. Wang, J. Wang, and Z. Yang (2022). Simulation of evacuation in an inclined passenger vessel based on an improved social force model. *Safety Science* 148, 105675.

Guzman-Chavez, A.D., and E. Vargas-Rodriguez (2025). The application of kernel ridge regression for the improvement of a sensing interferometric system. *Sensors* 25(5), 1292.

Li, K., T. Yao, X. Lin, X. Lin, and X. You (2025). "Analysis of the Efficiency of Slope Evacuation for Dense Crowds in Urban Street-Type Public Spaces." *Applied Sciences* 15 (7), 3568.

Li, Q., D. Wang, F. Liu, J. Yu, and Z. Jia (2024). LightGBM hybrid model-based DEM correction for forested areas. *PLOS ONE* 19(10), e0309025.

Lim, H., H. Lee, and J.-H. Hwang (2023). "Multi-Agent Simulation on Staff Evacuation Behavior in Elderly Nursing Home Fire Emergencies." *Buildings* 13 (2), 400.

MSC, I. 1/Circ. 1533 Revised Guidelines on Evacuation Analysis for New and Existing Passenger Ships [R]. 2016.

Parisi, L., and M.L. Manaog (2025). "Optimal Machine Learning- and Deep Learning-driven Algorithms for Predicting the Future Value of Investments: A Systematic Review and Meta-analysis." *Engineering Applications of Artificial Intelligence* 142, 109924.

Ren, J., Z. Mao, D. Zhang, M. Gong, and S. Zuo (2022). Experimental study of crowd evacuation dynamics considering small group behavioral patterns. *International Journal of Disaster Risk Reduction* 80, 103228.

Su, S., L. Wang, S. Wu, Y. Zhao, C. Wang, and L. Shu (2025). "A Data-Driven Method Approach for Prediction of Coal Seam Gas Content Combining Feature Selection and Machine Learning." *Engineering Applications of Artificial Intelligence* 160, 112000.

Wang, S., X. Liu, S. Yuan, Y. Bian, H. Wu, and Q. Ye (2025). "Artificial Intelligence Based Multispecialty Mortality Prediction Models for Septic Shock in a Multicenter Retrospective Study." *npj Digital Medicine* 8 (Article 228).

Wang, X., G. Xia, J. Zhao, J. Wang, Z. Yang, S. Loughney, S. Fang, S. Zhang, Y. Xing, and Z. Liu (2023). A novel method for the risk assessment of human evacuation from cruise ships in maritime transportation. *Reliability Engineering & System Safety* 230, 108887.

Yang, K., D. Cui, C. Wang, Q. Tang, and L. Miao (2025). "Intelligent Assessment of Habitat Quality Based on Multiple Machine Learning Fusion Methods." *Engineering Applications of Artificial Intelligence* 162 (Part A), 112395.

Zhang, Z., Z. Liu, Z. Zhou, X. Wang, A. Majumdar, Y. Cao, and Z. Yang (2025). Time prediction of human evacuation from passenger ships based on machine learning methods. *Transportation Research Part A: Policy and Practice* 200, 104664.