Transportation Research Part A 200 (2025) 104664

Contents lists available at ScienceDirect

Transportation Research Part A

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

Time prediction of human evacuation from passenger ships based on machine learning methods

Zhiwei Zhang [a,b], Zhengjiang Liu [a,d], Zirui Zhou [a,d], Xinjian Wang [a,b,c,d,e,**], Arnab Majumdar [f], Yuhao Cao [b], Zaili Yang [b,*]

[a] Navigation College, Dalian Maritime University, Dalian 116026, China
[b] Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool L3 3AF, UK
[c] Seafarers Research Institute, Dalian Maritime University, Dalian 116026, China
[d] Key Laboratory of Navigation Safety Guarantee of Liaoning Province, Dalian 116026, China
[e] State Key Laboratory of Maritime Technology and Safety, Dalian 116026, China
[f] Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK

ARTICLE INFO

Keywords:
Maritime safety
Emergency evacuation
Passenger ship
AnyLogic
Machine learning
Model selection

ABSTRACT

During an emergency evacuation scenario, accurately and timely predicting human evacuation time beforehand is crucial for developing an efficient evacuation plan. This study aims to develop an innovative simulation-based framework in which series state-of-the-art Machine Learning (ML) models are applied to predict human evacuation time from passenger ships. It also develops a multi-dimensional decision-making approach to evaluate their performance from the perspectives of high prediction accuracy and timeliness to support rapid response during emergencies. Firstly, an agent-based modelling technique incorporating two objectives and seven influential factors specific to human evacuation scenarios onboard ships is used to simulate the evacuation process. Then, the evacuation model is validated using three indicators to ensure its accuracy and relevance. Secondly, nine state-of-the-art ML models are applied to predict and analyse human evacuation time. To further investigate the role of feature interactions and enhance predictive accuracy, an additional model called the Attention-enhanced Light Gradient Boosting Machine (Attention-LightGBM) is proposed. Additionally, four statistical indicators are utilised to monitor the performance of each model. Finally, a new weighted selection method based on analytic hierarchy process and entropy weight method is created to conduct a comprehensive assessment from the perspectives of accuracy and timeliness. The findings reveal that the Attention-LightGBM demonstrates significant advantages in prediction accuracy, while the LightGBM excels in prediction timeliness. This study not only provides theoretical and technical support for emergency management onboard ships but also suggests methodological advancements for future research on complex human evacuation scenarios from passenger ships. The source code is publicly available at: https://github.com/AdvMarTech/Eva_Predict_ML.

* Corresponding author at: Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool L3 3AF, UK.
** Corresponding author at: Navigation College, Dalian Maritime University, Dalian 116026, China.
E-mail addresses: dmuzhangzhiwei@dlmu.edu.cn (Z. Zhang), liuzhengjiang@dlmu.edu.cn (Z. Liu), zhouzirui997@163.com (Z. Zhou), wangxinjian@dlmu.edu.cn (X. Wang), a.majumdar@imperial.ac.uk (A. Majumdar), Y.Cao@ljmu.ac.uk (Y. Cao), Z.Yang@ljmu.ac.uk (Z. Yang).

https://doi.org/10.1016/j.tra.2025.104664
Received 18 February 2025; Received in revised form 27 August 2025; Accepted 28 August 2025
Available online 12 September 2025

1. Introduction

Transportation by passenger ships, especially for leisure, is increasing in popularity worldwide (Li et al., 2024b; Ma et al., 2024; Valcalda et al., 2023). A distinct feature of passenger ships is the high degree of human involvement in their operations, and hence safety and security operations are of paramount importance (Cao et al., 2025c; Liu et al., 2021, 2022a; Wang et al., 2021c). Despite significant progress in modern ship design, e.g., safety equipment, conventions, and legal norms, maritime accidents involving passenger ships continue to occur with major loss of life (Arshad et al., 2022; Cao et al., 2025a; Galea et al., 2013b; Ni et al., 2017). For instance, the passenger ship *Orient Star* capsized and sank in 2015, leading to the death of 442 people, while the MV *Nyerere* shipwreck in 2018 claimed 228 lives. During both shocking accidents, their rapid evolution coupled with the inability to evacuate people out in time contributed to the large loss of life. Statistical data indicates that, only 2.64 % and 10.25 % of individuals were successfully evacuated to safe area, respectively (Wang et al., 2023a). Such figures reveal with little doubt that effective and timely human evacuation strategies in the event of passenger ship emergencies are crucial and are in urgent need of improvement.

In contrast to scenarios involving fixed structures on land, e.g. buildings, any evacuation of individuals from passenger ships is highly susceptible to several factors, such the external environment (e.g., sea condition or deck weather), limited space (e.g., narrow corridors and stairs) and human behaviours (e.g., congestion and competitive behaviours) (Long et al., 2023; Shipman et al., 2024). These factors pose a considerable challenge for any successful evacuation from passenger ships. In recent decades, in response to these difficulties, significant efforts have made by both industry and academia. In particular, the Maritime Safety Committee (MSC) of the International Maritime Organisation (IMO) has issued a guideline of MSC.1/Circ.1533 for passenger ship evacuation analysis (MSC, 2016), which defines evacuation time as a key indicator for assessing human evacuation performance. Simultaneously, an increasing number of studies attempt to address evacuation from a passenger ship, in a diverse manner, e.g. i) human behavioural understanding (Arshad et al., 2024a; Arshad et al., 2024b; Chen et al., 2023; Fang et al., 2024a), ii) walking speed analysis (Fang et al., 2022b; Fang et al., 2023b; Sun et al., 2018b; Wang et al., 2021a), and iii) optimisation of evacuation paths (Liu et al., 2023a; Liu et al., 2022b; Lozowicka, 2021). These studies acknowledge that the accurate prediction of evacuation time from passenger ships is the key to developing an effective evacuation plan.

Typically, simulation modelling is used to develop evacuation plans. Recently, based on conventional simulation methods (Chen et al., 2023; Lovreglio et al., 2014; Xie et al., 2020b), various Machine Learning (ML) techniques have been applied to improve prediction and simulation performance (Ebrahimi et al., 2023; Li et al., 2023). For instance, Guo et al. (2023) employed the Light Gradient Boosting Machine (LightGBM) technique to predict people's evacuation processes at a metro station based on different scenarios constructed by simulation. Similarly, Zhao et al. (2020b) utilised Random Forests (RF) to model and predict human behaviour during the pre-evacuation phase of a building emergency. Despite these advances, the application of ML in human evacuation from passenger ships still requires several gaps to be addressed. Although ML's ability to identify implicit relationships gives it unique advantages with convenience and accuracy, its methods significantly differ in handling these complexities, with no universally agreed-upon model being robustly applicable for different scenarios and tasks. Furthermore, ML evaluation indicators are often non-uniform, lacking comprehensive considerations in their relevance and implications to any real evacuation.

To address these challenges, this study proposes an integrated analytical framework to innovatively employ a series of state-of-the-art ML methods and rigorously evaluate their individual prediction performance in a scenario of involving human evacuation from passenger ships. In addition, methodology advancements are underpinned by an agent-based simulation technique for modelling the unique layout and environment of passenger ships. Advanced ML models are utilised to predict evacuation time, quantitatively validated by four indicators. To offer a multi-dimensional decision support for model section, the Analytic Hierarchy Process (AHP) and Entropy Weight Method (EWM) −based standardised weighted model is proposed to reveal model accuracy and timeliness. This study aims to contribute significantly to enhancing the predictive efficiency and homogeneity of human evacuation from passenger ships, thereby facilitating a more robust response to evacuation challenges in the future.

The rest of the paper is organised as follows: Section 2 reviews the relevant literature, which is followed in Section 3 by the methodology. Section 4 analyses and discusses the results, and the paper concludes in Section 5.

2. Literature review

2.1. Research on human evacuation from passenger ships

Modelling the evacuation of passengers from a passenger ship requires various factors that affect the evacuation to be considered (Arshad et al., 2022). Studies in this domain have identified and evaluated several key factors including those related the humans: i) evacuees' attributes (e.g., age, gender of passengers), ii) crowd behaviour (e.g., small group, competitive behaviour) as well as iii) ship attributes (e.g. the layout, its inclination), etc. For instance, Wang et al. (2022) developed a simulation model by considering evacuee attributes including age, gender, and physical condition and optimized the ship's staircase layout. Sun et al. (2018a) conducted corridor-based experiments to examine the effect of ship inclination on evacuee movement, showing that trim and heel significantly reduce average walking speed. Inspired by these studies, Fang et al. (2022a) incorporated the ship's inclination angle and examined the effects of ship rolling on passengers' walking speed across different age groups and genders. To more systematically study the influential factors of human evacuation from passenger ships, Fang et al. (2023a) conducted an orthogonal experiment to quantify the combined effects of ship listing, stair unavailability, and evacuee priority. Their results indicated that ship inclination and blocked stairs had a greater impact on evacuation time than evacuee priority. Building on these studies, Fang et al. (2024b) developed a new two-layer social force model to simulate the intensified effects of increased rolling amplitude, which led to exponential delays in evacuation. Lee et al. (2022) also explored the coupling effect of transverse and longitudinal inclination on escape speed, enhancing the accuracy of ship evacuation simulations. To examine the influence of crowd behaviour on evacuation time, Ha et al. (2012) proposed a cellular automata model to effectively calculate the total evacuation time, which also integrates attributes like passenger characteristics, ship layouts, and group behaviours. In addition, Wang et al. (2020) conducted a questionnaire-based study on individual and group behaviour, identifying demographic differences in passenger behaviour during emergency evacuations based on age, gender, and mobility.

Under various uncertain sea conditions, modelling passenger evacuation requires reliable and robust approaches to account for factors such as ship motion, adverse weather, and fire scenarios. These scenarios pose unique challenges, which have been explored by previous studies across specific contexts. For example, Wang et al. (2021b) conducted human walking experiments on a real ship to study the effects of ship motion on human walking speed, including berthing and sailing scenarios. Balakhontceva et al. (2015) proposed a multi-agent simulation framework to study evacuation during storm-induced ship inclinations, emphasizing environmental factors like wind and wave intensity. Fire scenarios in particular have been extensively studied due to their critical impact on evacuation processes. Vanem and Skjong (2006) developed an evacuation modelling approach to consider the impact of fire location and passenger capacity. Similarly, Sarshar et al. (2013) applied a Bayesian Network (BN) model to assess fire-related evacuation time, highlighting fire location, conditions, passenger demographics, and ship inclination as key factors. Wu et al. (2018) utilised a fire dynamics simulator to model smoke and temperature spread and estimated casualties by comparing available and required egress time. Additionally, Liu et al. (2023b); Xie et al. (2020a) used fire simulation and agent-based modelling to explore key factors in fire evacuations and proposed strategies for optimal route planning.

2.2. Research on the time prediction of human evacuation

Accurate prediction of evacuation time is critical across domains including building fires, metro rail evacuations, and natural disaster response. Evacuation measures in different fields are often tailored to their specific environments, risks and crowd density. Through a systematic cross-domain literature review, this study identifies transferable modelling principles that can inform the development of robust evacuation frameworks for passenger ships. To explicitly review the methods that have been utilised, a literature search was conducted using the Web of Science (WoS) database, focusing on studies published between 2014 and 2024. The literature search combined keywords from two core themes: (1) evacuation-related terms (human evacuation, emergency evacuation) and (2) prediction-related terms (prediction, forecast). The filtering process involved title and abstract screening, excluding non-English/non-empirical studies, and unrelated studies (e.g. studies where the terms only appeared in the background). As a result, 277 studies on human evacuation prediction across various fields were retained. Based on the visualisation of key trends from the search, as shown in Fig. 1, it is evident that ML and Deep Learning (DL) methods have been increasingly applied in evacuation prediction, especially the past five year, with most applications focusing on the evacuation of buildings (Li et al., 2023; Sahebi et al., 2023;

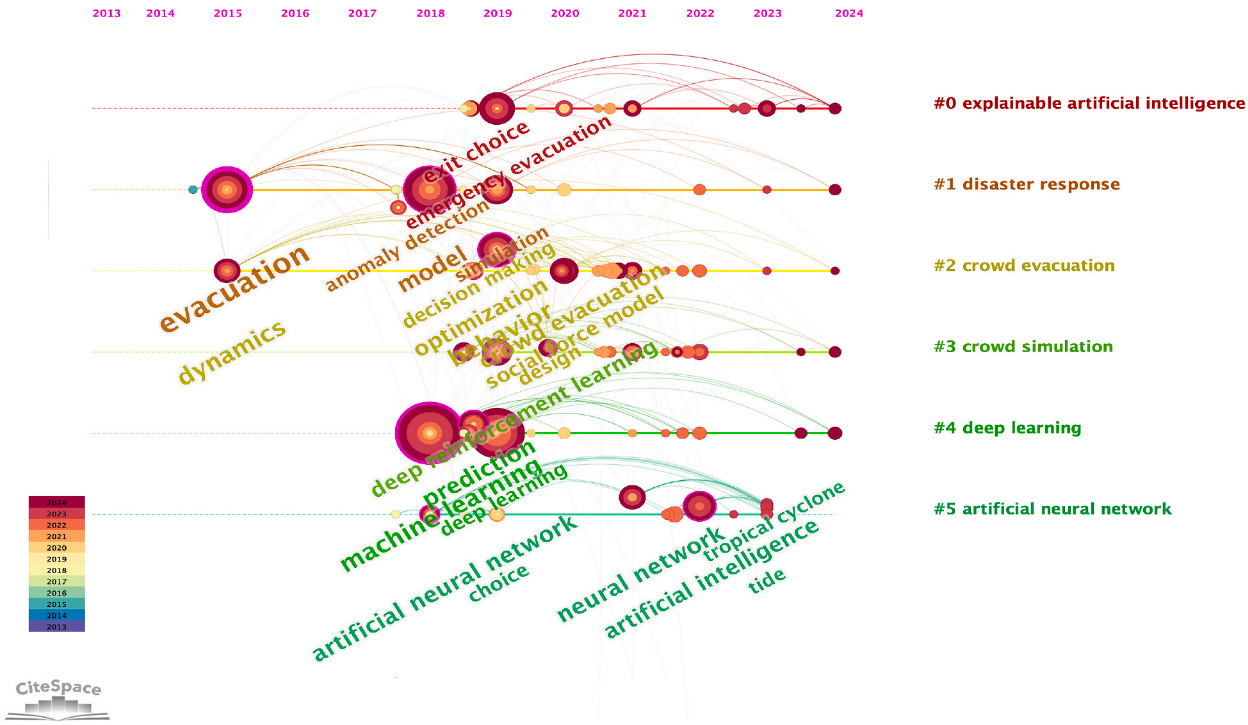


Fig. 1. The development trend of keywords in evacuation prediction research over time.

Wang et al., 2023b), metros (Guo et al., 2023; Yang et al., 2023), and natural disasters (Li et al., 2022; Sreejith and Sinimole, 2022; Sun et al., 2024).

Among these applications, DL models are frequently utilised for scenarios involving image-based data or large-scale datasets. However, DL models often face limitations, such as overfitting, when applied to smaller datasets, which can compromise their generalizability. In contrast, ML models have been widely employed in scenarios with relatively small datasets due to their interpretability, faster prediction speed, and suitability for understanding key factors influencing evacuation processes. This makes ML particularly well-suited for applications requiring real-time predictions or a clear analysis of influential factors in evacuation scenarios.

Currently, the application of ML and DL for building evacuation prediction is growing established. For instance, Li et al. (2023) applied the Extreme Learning Machine (ELM) with an improved seagull optimization algorithm for emergency risk assessment in a university library. The results showed that the improved ELM model enables efficient and rapid assessment of the emergency evacuation risk. Wang et al. (2023b) used simulated data to model evacuation time and exit selection, demonstrating the reliability of ML for planning evacuation strategies. Sahebi et al. (2023) assessed multiple ML models in a hospital fire scenario, showing that the RF model yielded the best predictive performance. These studies confirm that ML models can effectively model complex human behaviours and spatial constraints in static indoor environments.

In recent years, ML and DL have gradually been introduced into evacuation research in transportation, primarily in metro rail systems. Guo and Zhang (2022b) proposed a simulation-ML hybrid framework integrating nine input variables and three objectives, validated with real-world data from a metro station in Singapore. Guo et al. (2023) further combined simulation and ML to predict evacuation time and crowd density in congested metro areas, proposing an active guidance strategy to mitigate congestion. Yang et al. (2023) developed a multi-objective optimisation model for metro evacuation routes. By enhancing the backpropagation neural network model with the atomic orbit search algorithm, prediction accuracy was improved. These approaches demonstrate the growing potential of ML in modelling evacuation in large-scale and high-density transport systems.

Natural disasters are sudden events with often severe consequences. In the aftermath of such events, fast and accurate prediction of evacuation time is critical, where ML and DL approaches can effectively achieve this goal and address key challenges in such natural disaster evacuations. Based on a dataset with multiple scenarios relating to the evacuation from a real earthquake, Li et al. (2022) applied deep convolutional neural network to study the impact of environmental factors on crowd speed at various stages of evacuation. Sun et al. (2024) introduced an interpretable ML model to predict household evacuation decisions during hurricanes, outperforming traditional models in accuracy and explanatory power. Similarly, Sreejith and Sinimole (2022) developed a ML model to predict the household evacuation preparation time during a flood event. These works illustrate that data-driven prediction methods can be effectively adapted to dynamic and unpredictable evacuation scenarios.

2.3. Research gaps and contributions

The systematic analysis of factors that influence human evacuation from passenger ships, alongside studies of evacuation prediction in scenarios unrelated to shipping, highlights the need for the adaptation of more robust and accurate approaches to predict both the evacuation time and efficiency of evacuation in a greater depth. An extensive literature review reveals that the time for evacuation is highly affected by multiple influential factors. Furthermore, while advanced and intelligent methods have been applied broadly in several evacuation contexts, their use in passenger ship scenarios remain limited.

To address this gap, this study pioneers a simulation-based framework that applies nine state-of-the-art ML models and an improved LightGBM model to predict human evacuation time from passenger ships. The framework integrates two evacuation objectives and seven influential factors, developing a multidimensional decision-making approach to evaluate the prediction performance of ML models. This hybrid approach offers the potential for more accurate evacuation time prediction and highlights promising prospects for future research.

To more clearly exhibit the contributions of this study, the gaps (i.e., G1-G3) against the three contributions (i.e., C1-C3) is summarised below.

C1. An agent-based simulation modelling method integrating the unique characteristics of human evacuation from passenger ships

G1: Existing studies on human evacuation from passenger ships primarily rely on simulation models. However, the complex layout of passenger ships, lower personal familiarity, and the dynamic external marine environments make evacuation modelling in this domain significantly different from other domains. This environmental specificity of passenger ships necessitates highly tailored simulation input parameters to accurately reflect their evacuation process. Additionally, identifying relationships between the various influential factors is a major challenge for ML models, especially in capturing complex scenarios, such as group behaviour among passengers during evacuation.

Solution: This study develops an agent-based simulation model that comprehensively considers the effects of personnel attributes, group behaviour, ship layout, and external environmental factors in the scenario of passenger ships. By employing different combinations of variables, the simulation model can generate a wide range of evacuation scenarios reflecting diverse conditions. This approach helps establish the relationship between inputs (influencing factors) and outputs (evacuation time), thereby laying a solid foundation for ML models to identify relationships among various influencing factors.

C2. A comprehensive application of ML methods for accurate prediction of evacuation time in passenger ships

G2: The effectiveness of ML methods in land-based evacuations has been widely validated. For example, in crowd-dense metro systems and building evacuations, ML methods have significantly improved the accuracy and reliability of evacuation prediction. However, the application of ML in human evacuation from passenger ship scenarios remains limited. There have been few attempts to

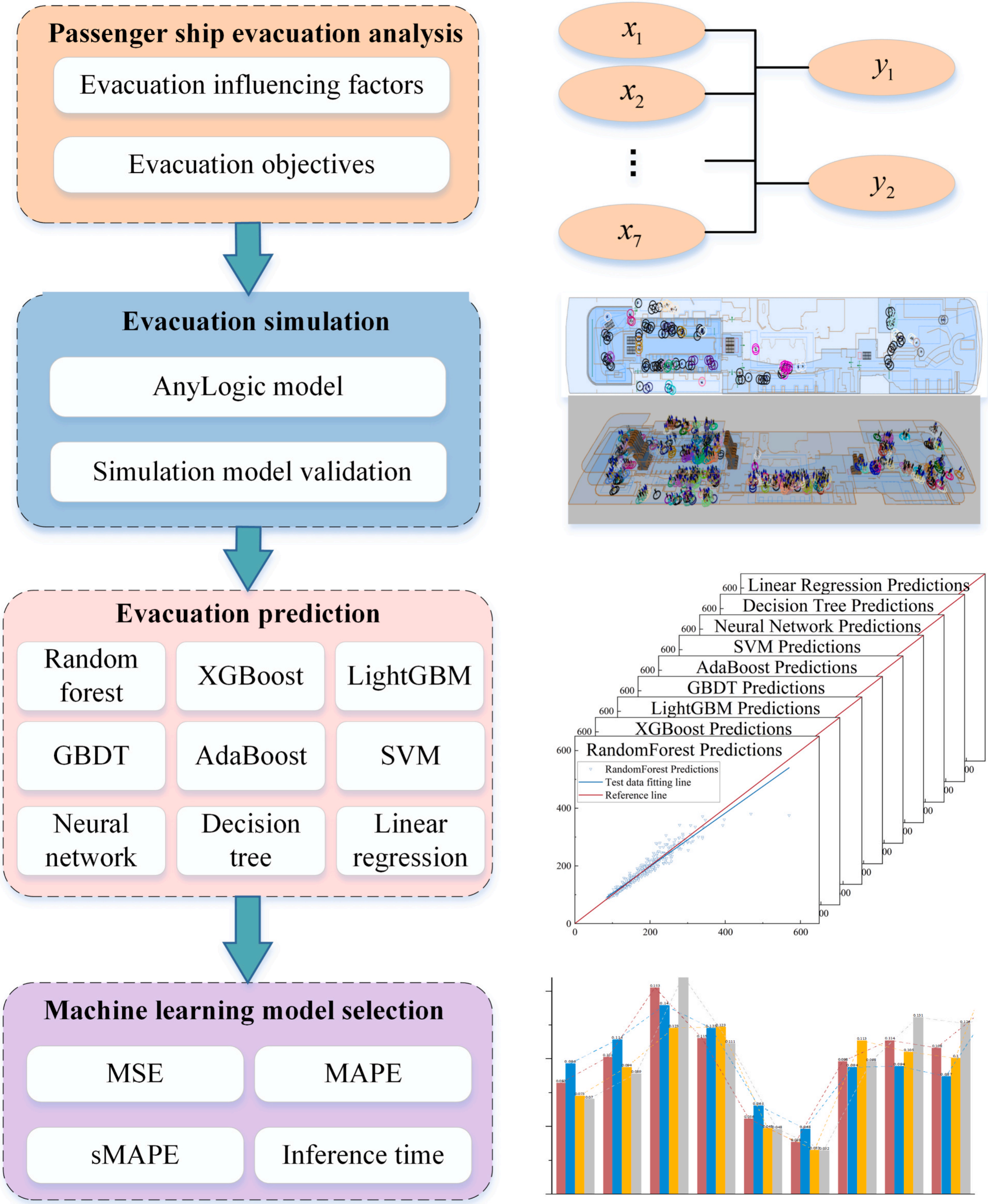


Fig. 2. The methodology framework for this study. *Note: XGBoost (Extreme Gradient Boosting), GBDT (Gradient Boosting Decision Tree), SVM (Support Vector Machine), MSE (Mean Squared Error), MAPE (Mean Absolute Percentage Error), sMAPE (Symmetric Mean Absolute Percentage Error).

explore this area, and the accuracy of predicting evacuation times for passenger ships is still questionable and needs further validation, particularly in large-scale, multi-scenario operations. This highlights the need for dedicated research to adapt and refine these ML models to better suit the unique complexities of maritime evacuation contexts.

Solution: In this study, nine state-of-the-art ML models are selected and introduced to enable the rapid prediction of human evacuation time from passenger ships. To further enhance the model diversity and capture complex feature dependencies, an additional attention-based LightGBM model is incorporated. This methodological extension offers a novel theoretical contribution by bridging advanced ML architectures with the domain of maritime evacuation prediction. The performance of these models is rigorously analysed by comparing their prediction results across various tasks, and the accuracy is verified by comparing with those simulation models. Furthermore, based on an analysis of practical challenges that may arise in passenger ship scenarios, four key indicators that reflect evacuation conditions are selected to evaluate the performance of the ML models from accuracy and timeliness perspectives.

C3. A selection method of standardised weighted model that balancing accuracy and timeliness effectively

G3: Various ML models, such as RF, Support Vector Machines (SVM), and LightGBM, have been applied to predict the evacuation time. However, the complexity of human evacuation from passenger ships requires models that not only achieve high prediction accuracy but also offer timeliness to support rapid response during emergencies. In such a time-critical context, the challenge of balancing prediction accuracy and response speed has become a pressing issue that needs to be addressed.

Solution: This study proposes two standardised and weighted model selection methods to balance both model accuracy and timeliness performance. Different scenarios are categorised according to the specific needs of passenger ships and the various tasks are set before, during, and after the evacuation process. By analysing model performance across different scenarios, this study identifies the most suitable ML model for each specific situation. Furthermore, this study discusses how various influential factors impact the predictive performance of these models and provide an in-depth explanation to ensure that the selected models perform optimally in real-world applications.

3. Methodology

Fig. 2 presents the methodological framework of this study. The first step involves defining the evacuation problem by identifying key influencing factors and formulating specific evacuation objectives. In this phase, a range of environmental and behavioural parameters are selected as critical variables. Next, a large-scale agent-based evacuation simulation model is developed and validated. This model systematically generates diverse sets of scenarios that represent various combinations of the identified variables. Each scenario captures the spatiotemporal dynamics of the evacuation process by simulating the continuous, individual-level movements of agents throughout the environment. Subsequently, several state-of-the-art ML models are trained using the scenario dataset to predict evacuation times. These models are designed to implicitly capture and learn from the complex spatiotemporal patterns encoded in the simulation data. Finally, the predictive performance of each model is evaluated using multiple quantitative indicators. Detailed descriptions of each step are provided in the following sections.

3.1. *Evacuation simulation parameters*

The proper selection of influential factors and indicators is a prerequisite for effectively evaluating the evacuation process. The ships selected for this study are sourced from publicly available datasets of the EU SAFEGUARD project (Brown et al., 2012; Deere

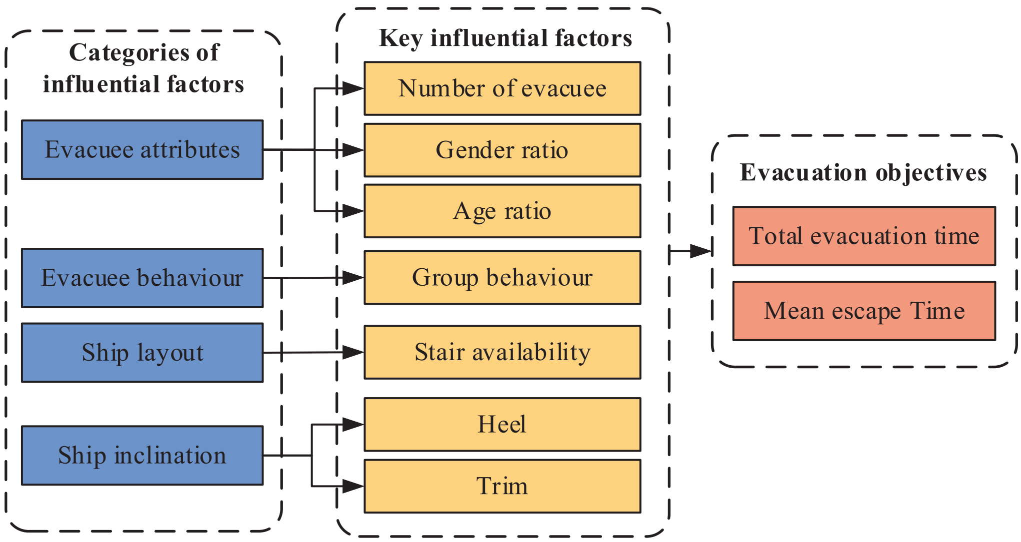


Fig. 3. Key influential factors and objectives of human evacuation from passenger ships (based on (Fang et al., 2023a)).

et al., 2012; Galea et al., 2013a). Based on an extensive literature review, four main categories of influential factors and two key evaluation objectives are identified, as shown in Fig. 3.

3.1.1. *Influential factors of human evacuation from passenger ships*

Table 1 presents seven influential factors across four categories related to human evacuation from passenger ships, highlighting their significant impacts on evacuation efficiency, time, safety, which pose challenges for emergency management onboard. One critical factor is the number of evacuees x_1, which directly affects evacuation time. With an increased number of evacuees, the density of people in evacuation routes and exits rises. Such crowds and congestions slow the movement of individuals, especially near stairs or narrow escape exits, often creating bottlenecks. Another factor is the gender ratio of passengers x_2, which impacts evacuation dynamics from the perspectives of movement speed and social roles. An experimental study showed the different movement speed during evacuation between genders, with males generally having greater physical strength and faster movement speed (Fang et al., 2023a). Additionally, in an emergency evacuation, the 'ladies first' principle may influence the evacuation order, reflecting gender-based social role differences (Mao et al., 2024). Similarly, passengers of varying ages x_3 have different reaction speeds, mobility, and needs during evacuation. Younger passengers typically move faster, while elder passengers may have mobility issues and slower reaction times. Children, due to their lower physical and cognitive abilities, often require assistance from adults or crew members. Therefore, the total number of evacuees, the gender ratio and the age ratio of passengers are identified as the key influential factors from the perspective of evacuee attributes.

Group behaviour x_4 is another key human-related characteristic that significantly influences the dynamics of evacuation. The presence of small groups directly impacts overall evacuation speed, as group actions are more likely to cause local congestion, particularly in narrow passageways and stairways. This is due to the tendency of groups to move together and make collective decisions, which can slow the evacuation process when navigating through constrained spaces.

Furthermore, in scenarios such as a ship fire, the chimney effect of the fire can impact stair availability x_5 as the smoke spreads up the stairs. This phenomenon may render some vertically connected stairs impassable, forcing evacuees to seek alternative routes. The need to find other escape paths not only complicates the evacuation process but also increases the time required to evacuate safely, thereby elevating the risk to all on board.

From the viewpoint of the ship inclination, both heel x_6 and trim x_7 significantly impact the evacuation process. Different directions of inclination alter the ship's centre of gravity, where such changes in the ship's centre of gravity typically affect both passenger balance and evacuation speed. The inclination of ship deck may also affect the use of escape routes and stairs, increasing the difficulty of evacuation path planning.

3.1.2. *Objectives of human evacuation from passenger ships*

(1) Total evacuation time (TET)

The TET refers to the duration required for all individuals on board to move from the initial location to the muster station (MSC, 2016). The IMO evacuation guidelines use TET as the primary evaluation indicator for evaluating ship evacuation efficacy. TET offers a global perspective, directly revealing the overall evacuation efficiency and comprehensively reflecting the impact of various influencing factors on evacuation performance as well as the allocation and utilisation of evacuation resources. In most cases, a shorter TET ensures that all evacuees can successfully reach a safe zone in the shortest possible time, thereby minimizing threats to personal safety. Furthermore, analysing TET across different evacuation scenarios is able to provide valuable insights for optimizing overall evacuation design and strategies.

(2) Mean escape time (MET)

The MET denotes the average time from the onset of an emergency on board a ship to the successful evacuation of all passengers to a designated safe area (Takimoto and Nagatani, 2003). In this study, the designated safe area refers specifically to the muster stations defined in the EU SAFEGUARD project dataset. These stations are standardized assembly points located on the ship, in accordance with IMO evacuation guidelines, and represent the official locations where passengers are considered to have completed the initial phase of

Table 1
Factors affecting human evacuation from passenger ships.

Categories	Factors	Descriptions	Reference
Evacuee attributes	Number of evacuees x_1	Total number of passengers onboard passenger ships at the time of evacuation.	(Fang et al., 2022b)
	Gender ratio of passengers x_2	The gender of passengers can influence the average speed of the evacuation process to some extent.	(Fang et al., 2024b)
	Age ratio of passengers x_3	The age of the passengers affects their movement speed.	(Fang et al., 2024b)
Evacuee behaviour	Group behaviour x_4	Passengers prefer evacuating together or in groups, rather than individually.	(Wang et al., 2020)
Ship layout	Stair availability x_5	A fire on a passenger ship may create a chimney effect, rendering all vertically connected stairs unusable.	(Liu et al., 2022c)
Ship inclination	Heel x_6	Left and right (lateral) listing caused by external conditions.	(Fang et al., 2023b)
	Trim x_7	Forward and aft (longitudinal) inclination of the ship caused by external conditions.	(Fang et al., 2023b)

evacuation.

The MET offers a more nuanced view of individual passenger evacuation efficiency, highlighting potential bottlenecks in the evacuation process, and resource utilisation compared to TET, offering insights beyond what TET can reveal. Specifically, a short MET implies that most passengers evacuate quickly, indicating a generally effective evacuation process that does not depend solely on the speed of the first few evacuees. Conversely, if the TET is long but MET is short, this suggests that while most passengers evacuated efficiently, the total evacuation time is extended by a few individuals due to special circumstances (e.g., mobility issues or blockages). In other words, MET helps to reveal early issues in the evacuation process or differences in evacuation times among different groups of passengers. It serves as a crucial metric for understanding how uniformly efficient an evacuation is and for identifying specific areas where improvements can be made to enhance overall safety.

3.2. Evacuation simulation modelling

3.2.1. Evacuation simulation model construction

As discussed in Section 2.1, simulation modelling has been widely applied in evacuation studies and offers a safer and more cost-effective alternative to real experiments. AnyLogic, a powerful simulation software used to model and simulate complex systems across a broad range of industries, such as logistics (Muravev et al., 2021), transportation (Guo and Zhang, 2022a), manufacturing (Farsi et al., 2020), and supply chain management (Bozdogan et al., 2023). It leverages agent-based modelling to simulate the interactions of various passenger behaviours during the evacuation onboard passenger ships. AnyLogic's advanced modelling capabilities also allow it to provide accurate human evacuation process information by considering various factors, such as the ship's layout, stair availability, and complex ship inclinations (e.g., heel and trim). To enhance the realism and replicability of simulated evacuation process, this study

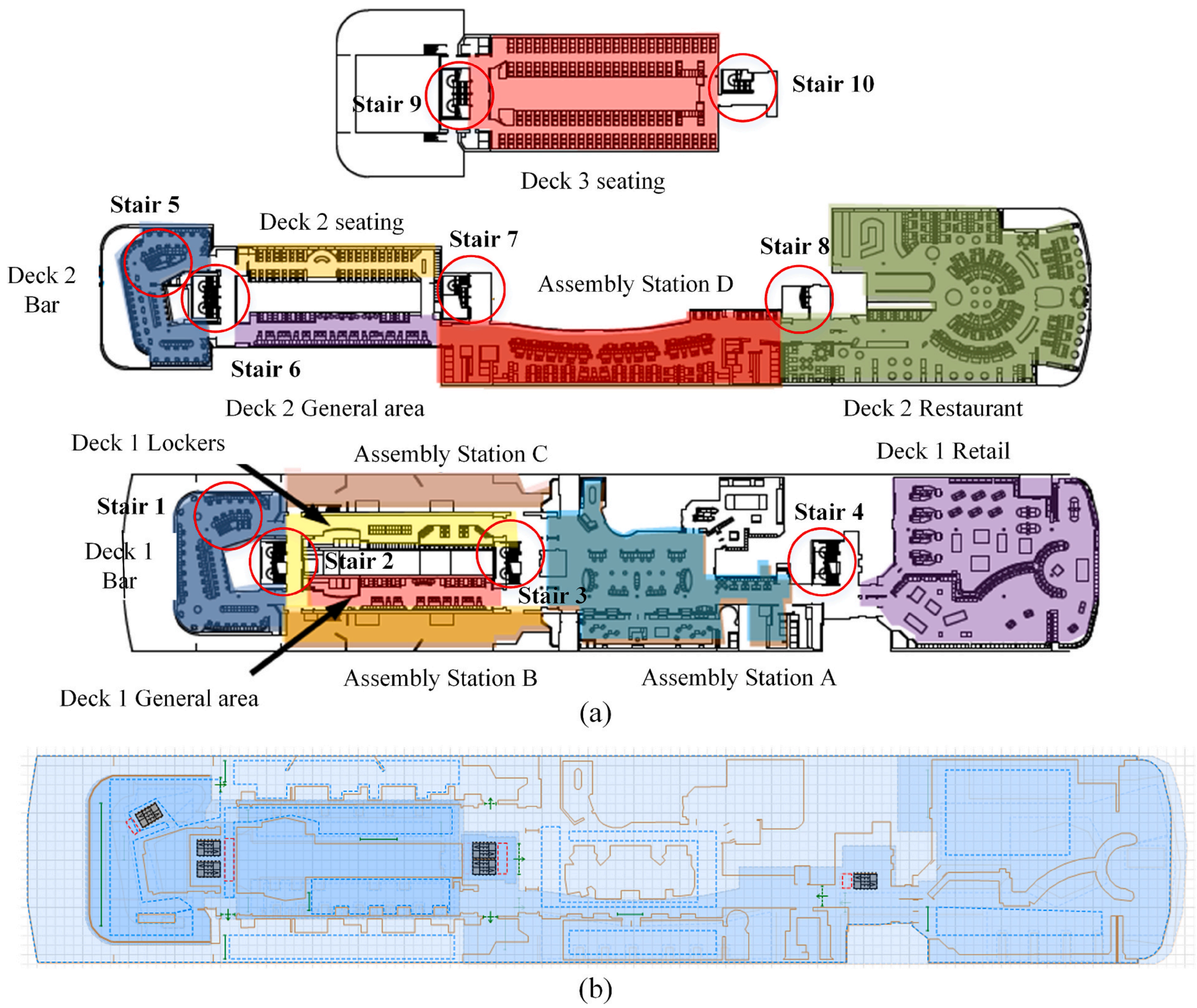


Fig. 4. Passenger ship layout and AnyLogic example: (a) CAD drawing of the ship; (b) AnyLogic 2D example.

incorporates the public SGVDS1 dataset from the European SAFEGUARD project alongside AnyLogic software in the simulation analysis. The CAD drawing of a ro-ro passenger ship and an example of AnyLogic software application in the study are shown in Fig. 4.

Fig. 4 illustrates the decks, stair configurations, and layout of the ro-ro passenger ship used in the simulation. This ship is selected for its representativeness in layout complexity and operational scenarios, making it an ideal case for evacuation studies. Moreover, it has been extensively analysed in previous research and is among the few ships with available experimental evacuation data, facilitating precise validation of the simulation model (Gayathri et al., 2022).

Overall, this ship comprises three decks, ten stairways, and four muster stations designed for evacuation purposes. The first deck includes the general area, bar area, storage area, retail area, and three muster stations labelled A, B, and C. The second deck includes the general area, bar area, seating area, dining area, and muster station D. The third deck consists of additional seating area. Notably, the ship does not include cabins or sleeping areas. In line with IMO guidelines, nighttime-specific evacuation scenarios are not considered, as the ship layout does not support overnight accommodation. Instead, response time under daytime operational conditions is the primary concern of this simulation, reflecting typical passenger distribution and activity patterns during service hours. The drawings also specify the dimensions of each deck. Both the first and second decks are identical in size: 138 m in length, 24 m in width, and 3 m in height. The third deck is smaller, with dimensions of 42 m in length, 15 m in width, and 2.8 m in height. Detailed information about the stair configurations, their connections, and widths is provided in Table 2.

3.2.2. Evacuation simulation model validation

After completing the simulation experiments, it is critical to quantify the degree of consistency between simulation results and real-ship experimental outcomes to ensure the simulation accuracy. In this study, this consistency is assessed by comparing the differences between simulation-predicted curves and actual measurement curves from experiments. Key indicator parameters such as the Euclidean Relative Difference (ERD), Euclidean Projection Coefficient (EPC), and Secant Cosine (SC) are employed.

Specifically, ERD measures the Euclidean distance between experimental and model data points. A smaller ERD value indicates a smaller difference. The acceptance criterion for ERD is generally set at ERD ≤ 0.45 (Galea et al., 2013b). EPC assesses the consistency between experimental and model data. The recommended acceptance criteria are: $0.6 \leq$ EPC ≤ 1.4 (Galea et al., 2013b). SC assesses the consistency of the shape of two datasets by comparing their rates of change. A SC value of 1.0 indicates that the shape of the model's curve perfectly matches the shape of the real curve, indicating identical gradient changes over time. The recommended acceptance criterion is SC > 0.6 (Galea et al., 2013b).

3.3. Evacuation prediction model

3.3.1. State-of-the-art ML models

Through a review of existing literatures, nine ML models have been identified as the state-of-the-art approach applied to provide more accurate prediction for emergency decision-making (Feng et al., 2024a; Feng et al., 2024b; Hatipoglu and Tosun, 2024; Ide et al., 2024; Lin et al., 2017; Ma et al., 2017), including: Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Gradient Boosting Decision Tree (GBDT), Adaptive Boosting (AdaBoost), Support Vector Machine (SVM), Neural Networks (NN), Decision Trees (DT), and Linear Regression (LR). In addition, this study proposes an enhanced model named Attention-LightGBM, which incorporates an attention mechanism into the LightGBM framework to more effectively capture complex non-linear interactions and dynamic feature importance during evacuation scenarios (Cao et al., 2025b).

These models encompass a range of types prevalent in the current ML landscape. From simple LR to sophisticated ensemble learning and deep learning techniques, they span a broad spectrum of application scenarios. For instance, LR is valued for its ability to handle large datasets quickly, facilitating rapid predictions in emergency situations to devise timely responses. DT and RF exhibit powerfulness in interpreting high-dimensional data and complex nonlinear relationships, effectively capturing multiple factors affecting human evacuation time in passenger ship scenarios. Gradient boosting models, such as XGBoost, LightGBM, and GBDT, are particularly suitable for managing complex evacuation scenarios due to their robust computational efficiency and nonlinear processing capabilities, enabling precise time predictions in a short period. Adaboost and SVM, as boosted learning and regression models, excel in handling uneven data, such as variations caused by ununiform stair layouts during evacuations. NN are notably proficient at handling complex unstructured data (Li et al., 2025). The diverse selection of models is tailored to accommodate the unique characteristics of evacuation data, including the inherent errors and biases within different data types. Although these models have been utilised across various scenarios in distinct studies, a systematic investigation that employs them in the same scenario for comparative analysis remains lacking. By leveraging and applying these models in this study, their respective strengths can be comparatively illustrated in

Table 2
The stair configurations, connections, and widths of the passenger ship (Galea et al., 2013a).

Stairs	Stair connections	Width of stairs (m)
Stair 1 and Stair 5	Decks 1 and 2	1
Stair 2 and Stair 6	Decks 1 and 2	1.35
Stairs 3 and Stair 7	Decks 1 and 2	1.35
Stair 4 and Stair 8	Decks 1 and 2	1.35
Stair 6 and Stair 9	Decks 2 and 3	1.35
Stair 7 and Stair 10	Decks 1 and 2	1.2

predicting human evacuation time in passenger ship scenarios, ensuring highly accurate and stable results.

Each model is subjected to hyperparameter tuning to optimize prediction performance. Tuning aimed to balance prediction accuracy and computational efficiency. Grid search is adopted as the tuning strategy due to its transparency and reproducibility. A set of key hyperparameters is defined for each model based on prior studies and implementation guidelines (Uhlík et al., 2024; Zhang et al., 2025). A detailed overview of hyperparameter ranges and selected configurations is provided in Table G of Appendix D. The final evaluation of model performance incorporated multiple criteria, including error indicators and inference time.

3.3.2. Four evaluation indicators

To comprehensively measure the performance of the selected ML models in predicting human evacuation time from passenger ships, four key indicators are employed: Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), symmetric Mean Absolute Percentage Error (sMAPE), and inference time. These indicators are chosen to ensure a comprehensive assessment of the model's prediction accuracy, stability, and efficiency in large-scale practical scenarios.

(1) Mean squared error (MSE)

MSE, also known as L2 loss, is a critical metric for measuring the prediction error of a model by calculating the average of the squared differences between the predicted and actual values. This squaring of the differences means that large errors are given disproportionately more weight than smaller ones, making MSE more sensitive to outliers. In the context of evacuation simulations, using MSE as an evaluation metric helps to emphasise and penalise such large errors more severely. This is especially important in ensuring that the model does not produce highly biased predictions that could result in costly delays in real-world scenarios. Thus, a lower MSE indicates a more accurate model, enhancing the reliability of the evacuation time predictions needed for effective emergency management (Chicco et al., 2021). The formula for MSE is expressed as Equation (1):

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(z_i - \widehat{z}_i)^2 \tag{1}$$

where z_i represents the true value for each sample i and $\widehat{z}_i$ represents the predicted value for each sample i. m is the total number of samples.

(2) Mean Absolute Percentage error (MAPE)

MAPE calculates the average of the absolute differences between the predicted and true values. This indicator gives equal proportional weighting to all error, contributing to its robustness as a measure of prediction accuracy. Compared to MSE, MAPE tends to be less sensitive to such extreme outliers. Another key advantage of MAPE is its ease of interpretation. Represented as a percentage, it allows decision-makers, who may not be experts in ML techniques, such as captains or crew members, to quickly understand the quality of predictions. This clarity can significantly enhance decision-making processes, allowing for more informed and effective safety strategies on board (Chicco et al., 2021). The formula for MAPE is shown as Equation (2):

$$MAPE = \frac{100\%}{m}\sum_{i=1}^{m}\left|\frac{z_i - \widehat{z}_i}{z_i}\right| \tag{2}$$

(3) Symmetric Mean Absolute Percentage error (sMAPE)

sMAPE is an enhanced version of MAPE designed to address its bias by treating the predicted and actual values symmetrically. In prediction of human evacuation time from passenger ship, there can be both overprediction (the model predicts too much time) and underprediction (the model predicts too little time). In safety management on passenger ships, accurate predictions are essential for planning effective emergency measures on board. Overpredictions might lead to unnecessary delays and resource allocations, whereas underpredictions can result in insufficient preparations, potentially affecting passenger safety. sMAPE provides a more balanced approach by equally weighting both types of prediction errors, which is crucial under the dual constraints that evacuation times neither be too long or too short. Furthermore, sMAPE keeps the strength in easy interpretation, improve its practical implications (Chicco et al., 2021). The formula for sMAPE is shown as Equation (3):

$$sMAPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|z_i - \widehat{z}_i|}{\left(\frac{|z_i|+|\widehat{z}_i|}{2}\right)} \tag{3}$$

(4) Inference time

Inference time refers to the time required for the model to make a single prediction, usually measured in seconds or milliseconds. It assesses the model's response speed when being applied in practical scenarios. A short inference time is particularly vital in situations that need rapid responses, such as predicting human evacuation times from passenger ships. Additionally, models with short inference time are better suited for real-time monitoring and dynamic adjustments. During evacuation, external conditions (e.g., fire, obstacles, or human behaviours) can change at any time. Models with short inference times can recalibrate predictions in response to changing conditions, thereby providing updated evacuation times that can help optimise evacuation routes and strategies (Uhlík et al., 2024).

3.4. New holistics weighted multi-indicator selection models

To ensure a fair and consistent comparison of ML models, the indicators identified in Section 3.3 need to be further synthesised. Denoting the values of MSE, MAPE, sMAPE, inference time as $\gamma_{Ai}, \gamma_{Bi}, \gamma_{Ci}, \gamma_{Di}$, the original values of the four indicators can be represented by a matrix, as shown in Equation (4):

$$\begin{bmatrix} \gamma_{A1} & \gamma_{B1} & \gamma_{C1} & \gamma_{D1} \\ \gamma_{A2} & \gamma_{B2} & \gamma_{C2} & \gamma_{D2} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{An} & \gamma_{Bn} & \gamma_{Cn} & \gamma_{Dn} \end{bmatrix} \tag{4}$$

where i denotes the $i-th$ ML model, and n denotes the number of ML models.

Since all these indicators are negatively correlated with model performance, where lower values indicate better model performance, the synthesis process involves transforming these indicators into a uniformed format where higher values indicate superior performance. Hence, the transforming process can be shown as Equation (5):

$$\begin{cases} \eta_{Ai} = \dfrac{\sum_{i=1}^{n}\gamma_{Ai}}{\gamma_{Ai}} \\ \eta_{Bi} = \dfrac{\sum_{i=1}^{n}\gamma_{Bi}}{\gamma_{Bi}} \\ \eta_{Ci} = \dfrac{\sum_{i=1}^{n}\gamma_{Ci}}{\gamma_{Ci}} \\ \eta_{Di} = \dfrac{\sum_{i=1}^{n}\gamma_{Di}}{\gamma_{Di}} \end{cases} \Rightarrow \begin{cases} \xi_{Ai} = \dfrac{\sum_{i=1}^{n}\eta_{Ai}}{\eta_{Ai}} \\ \xi_{Bi} = \dfrac{\sum_{i=1}^{n}\eta_{Bi}}{\eta_{Bi}} \\ \xi_{Ci} = \dfrac{\sum_{i=1}^{n}\eta_{Ci}}{\eta_{Ci}} \\ \xi_{Di} = \dfrac{\sum_{i=1}^{n}\eta_{Di}}{\eta_{Di}} \end{cases} \tag{5}$$

where $\eta_{Ai}, \eta_{Bi}, \eta_{Ci}, \eta_{Di}$ are the normalised performance indicators. $\xi_{Ai}, \xi_{Bi}, \xi_{Ci}, \xi_{Di}$ are the converted values transformed to ensure higher values uniformly represent superior performance.

Then, in this study, a novel comparison matrix is introduced to justify these adjusted indicators. The comparison matrix is attained as Equation (6):

$$\begin{bmatrix} \xi_{A1} & \xi_{B1} & \xi_{C1} & \xi_{D1} \\ \xi_{A2} & \xi_{B2} & \xi_{C2} & \xi_{D2} \\ \cdots & \cdots & \cdots & \cdots \\ \xi_{An} & \xi_{Bn} & \xi_{Cn} & \xi_{Dn} \end{bmatrix} \tag{6}$$

Subsequently, the standardised matrix is defined as Equation (7):

$$\begin{bmatrix} \varphi_{A1} & \varphi_{B1} & \varphi_{C1} & \varphi_{D1} \\ \varphi_{A2} & \varphi_{B2} & \varphi_{C2} & \varphi_{D2} \\ \cdots & \cdots & \cdots & \cdots \\ \varphi_{An} & \varphi_{Bn} & \varphi_{Cn} & \varphi_{Dn} \end{bmatrix} \tag{7}$$

where

$$\varphi_{ki} = \frac{\xi_{ki}}{\sum_{i=1}^{n}\xi_{ki}} \tag{8}$$

where, k denotes the index for four evaluation indicators, i.e., $k \in$ {A, B, C, D}.

Therefore, the standardised vector Ψ_i of the $i-th$ ML model is defined as:

$$\Psi_i = [\, \varphi_{Ai} \quad \varphi_{Bi} \quad \varphi_{Ci} \quad \varphi_{Di} \,] \tag{9}$$

Moreover, after the evacuation analysis by training different ML methods, an AHP method is employed to determine the weights of the indicators and highlight their importance in this study. AHP, as a commonly used decision analysis method, decomposes complex problems into multiple levels and determines the importance of each factor through pairwise comparisons. This method calculates the weights of the indicators based on expert opinion or empirical data, ensuring the model selection process reflects their true significance (Li et al., 2021). AHP is particularly suitable for this weighted operation due to its systematic structure, intuitive pairwise comparison process, and practical simplicity in deriving consistent weights. In this study, the weights determined by the AHP are used exclusively during the model selection phase to comprehensively evaluate the performance of different ML models across two key dimensions: accuracy and timeliness. To support reproducibility and ensure feasibility in future practical applications, a fixed-weight strategy is employed throughout the evaluation process.

Here, the weights of the indicators are determined through a pairwise comparison process, where the relative importance of each indicator is evaluated and aggregated using the geometric mean method. The final weights for the four indicators (A, B, C, D) are computed as shown in Equation (10):

$$\Delta_j = \left[\frac{\sum_{j=1}^{4}\lambda^{Aj}}{4} \; \frac{\sum_{j=1}^{4}\lambda^{Bj}}{4} \; \frac{\sum_{j=1}^{4}\lambda^{Cj}}{4} \; \frac{\sum_{j=1}^{4}\lambda^{Dj}}{4} \right]^{\top} \tag{10}$$

where $\lambda^{Aj}, \lambda^{Bj}, \lambda^{Cj}, \lambda^{Dj}$ represent the normalized weights of indicators A, B, C, and D respectively, derived from the pairwise comparisons. The weights are averaged over all comparisons to ensure consistency.

In summary, the final evaluation index is calculated by combining the weight vector Δ_j with the attribute vector Ψ_i, as shown in Equation (11):

$$Q_{ij} = \Psi_i \times \Delta_j = [\varphi_{Ai} \quad \varphi_{Bi} \quad \varphi_{Ci} \quad \varphi_{Di}] \left[\frac{\sum_{j=1}^{4} \lambda^{Aj}}{4} \ \frac{\sum_{j=1}^{4} \lambda^{Bj}}{4} \ \frac{\sum_{j=1}^{4} \lambda^{Cj}}{4} \ \frac{\sum_{j=1}^{4} \lambda^{Dj}}{4} \right]^{\top} \tag{11}$$

where Q_{ij} denotes the weighted composite score incorporating all indicators for the $i-th$ model.

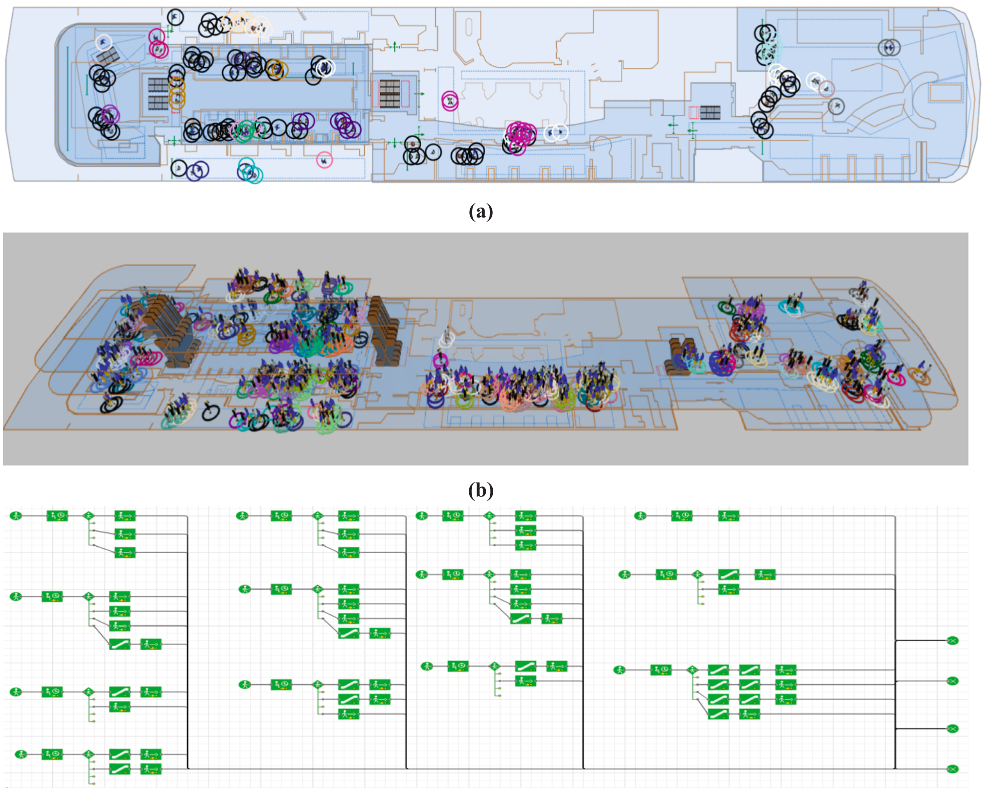
Further to this, to deliver a comparation with the aforementioned evaluation model, another objective data-based evaluation model is established. Specifically, replacing the AHP method, the EWM is incorporated to calculate the information entropy (i.e., the weight) of each indicator, offering an objective perspective to guarantee the fairness and diversity of the model selection process. This approach holds it capability when the expert knowledge and empirical data are missing. The weights of each model are fist recalculated to accurately represent the distribution of information across the indicators. To evaluate the degree of dispersion of each indicator, the concept of information entropy is employed. Based on the calculated information entropy, the entropy weight for each indicator is defined as shown in Equation (12):

$$\omega_k = \frac{1 - E_k}{\sum_{k=1}^{4} (1 - E_k)} \tag{12}$$

where E_k represents the information entropy of the $k-th$ indicator, and n is the number of models. The entropy value ranges from 0 to 1. A higher entropy value indicates greater uncertainty or a more even distribution of an indicator's information, making it less useful for evaluation and less impactful on decision-making. ω_k represents the weight of the $k-th$ indicator. A smaller entropy value indicates that the indicator is more informative, leading to a larger weight.

Therefore, the composite score G_{ki} can be obtained with the Equation (13):

$$G_{ki} = \Psi_i \times \omega_k = [\varphi_{Ai} \quad \varphi_{Bi} \quad \varphi_{Ci} \quad \varphi_{Di}][\omega_A \quad \omega_B \quad \omega_C \quad \omega_D]^{\top} \tag{13}$$



(a)

(b)

(c)

Fig. 5. The established evacuation simulation models of SGVDS1: (a) the 2D model (b) the 3D model (c) the behavioural logic of personnel.

4. Results and discussion

This section presents the key results obtained from section 3 and provides an integrated discussion of their theoretical and practical implications. It begins with the construction and validation of the simulation model used to generate evacuation data (Section 4.1), followed by the selection and evaluation of ML models for predicting evacuation time (Section 4.2 and 4.3). Section 4.4 compares different weighting strategies for model selection and discusses interpretability, while Section 4.5 reflects on the broader implications of the findings.

4.1. Simulation model results

Prior to implementing ML models for evacuation time prediction, it is crucial to establish a robust and realistic simulation framework that faithfully captures the dynamics of onboard passenger evacuation. This section outlines the systematic development of the simulation model, encompassing its construction (Section 4.1.1), validation against established benchmarks (Section 4.1.2), and the generation of synthetic data to support subsequent ML model training (Section 4.1.3). These preparatory steps are essential to ensure the integrity and reliability of the training data, thereby providing a solid foundation for assessing model performance across diverse evacuation scenarios of passenger ships.

4.1.1. Simulation model construction

As shown in Fig. 5, a 2D planar model, a 3D model, and pedestrian entities are constructed in AnyLogic for evacuation simulation. In this study, the number of evacuees is set between 400 and 900, increasing at a step of 100 individuals from 400 to 900. The initial locations of the evacuees are randomly distributed according to the areas recorded in the European 'SAFEGUARD' project dataset (SGVDS1), with the designated endpoints being four muster stations (Brown et al., 2012; Deere et al., 2012; Galea et al., 2013a). The age, gender composition, and speed of evacuees are determined based on the personal attributes outlined in the IMO MSC 1533 circular for advanced evacuation analysis (MSC, 2016). Further details are provided in Table A and Table B of Appendix A.

According to the stair combination and layout described in Section 3.2, there are four sets of vertically connected stairs. During a fire requiring emergency evacuation, the chimney effect may render stairs vertically connected unavailable. Therefore, stair availability is divided into five groups based on the vertical configuration, as shown in Table 3.

In emergencies, when the ship faces grounding or flooding, the ship may experience listing. Therefore, the impact of the list angle on passenger evacuation is a crucial consideration. Passenger ship inclination is categorized as either transverse or longitudinal, and the combined effect of both inclinations on individual walking speed has been considered (Sun et al., 2018a). Speed attenuation factors for individuals are listed in Table C of Appendix A.

Group behaviour is common during emergency evacuation, where evacuees often seek to escape with familiar people after receiving an alert. In this study, the number and proportion of small groups are derived from data observed on real ships. Details are provided in a previous research (Wang et al., 2020).

4.1.2. Validation of simulation models

To verify the applicability of AnyLogic, a set of simulation result is compared with real ship data after its construction. The validation process for the simulation results involves two key steps: First, three indicators (e.g., ERD, EPC, SC) of the total evacuation data are calculated. Second, three indicators for each of the four gathering stations are calculated separately, resulting in a total of 12 indicators. The validation criteria are as follows: 1) All three indicators of the total evacuation data must meet the acceptance criteria. 2) At least nine of the 12 indicators from the four gathering stations must meet the acceptance criteria, with no more than one parameter per station failing to meet these criteria. Satisfying these rules confirm the model's appropriateness for evacuation simulation studies (Galea et al., 2013a).

In the AnyLogic evacuation simulation model, parameters such as passengers' response time, evacuation speed, and spatial distribution identical to the validation file are predefined. Fifty evacuation simulations are conducted in a scenario to obtain averaged evacuation time data. Fig. 6 shows the arrival times of passengers at the four muster stations (A, B, C, and D), while Fig. 7 shows the collective arrival times of all personnel at these stations. The calculated values for the evaluation indicators are shown in Table 4.

As shown in Figs. 6 and 7, the blue solid lines represent the experimental results, and the red solid lines represent the AnyLogic simulation results. The curves of both methods exhibit similar trends with minimal deviation.

As shown in Table 4, all simulation results comply with the validation criteria, and the time curve in the simulation align with those in SGVDS1. Therefore, AnyLogic is validated as an effective tool for conducting evacuation simulation studies in passenger ship

Table 3
Grouping of unusable stairs due to chimney effect from fire.

Stair combinations	Stairs not available
Stairs group 1	None
Stairs group 2	Stair 1, Stair 5
Stairs group 3	Stair 2, Stair 6, Stair 9
Stairs group 4	Stair 3, Stair 7, Stair 10
Stairs group 5	Stair 4, Stair 8

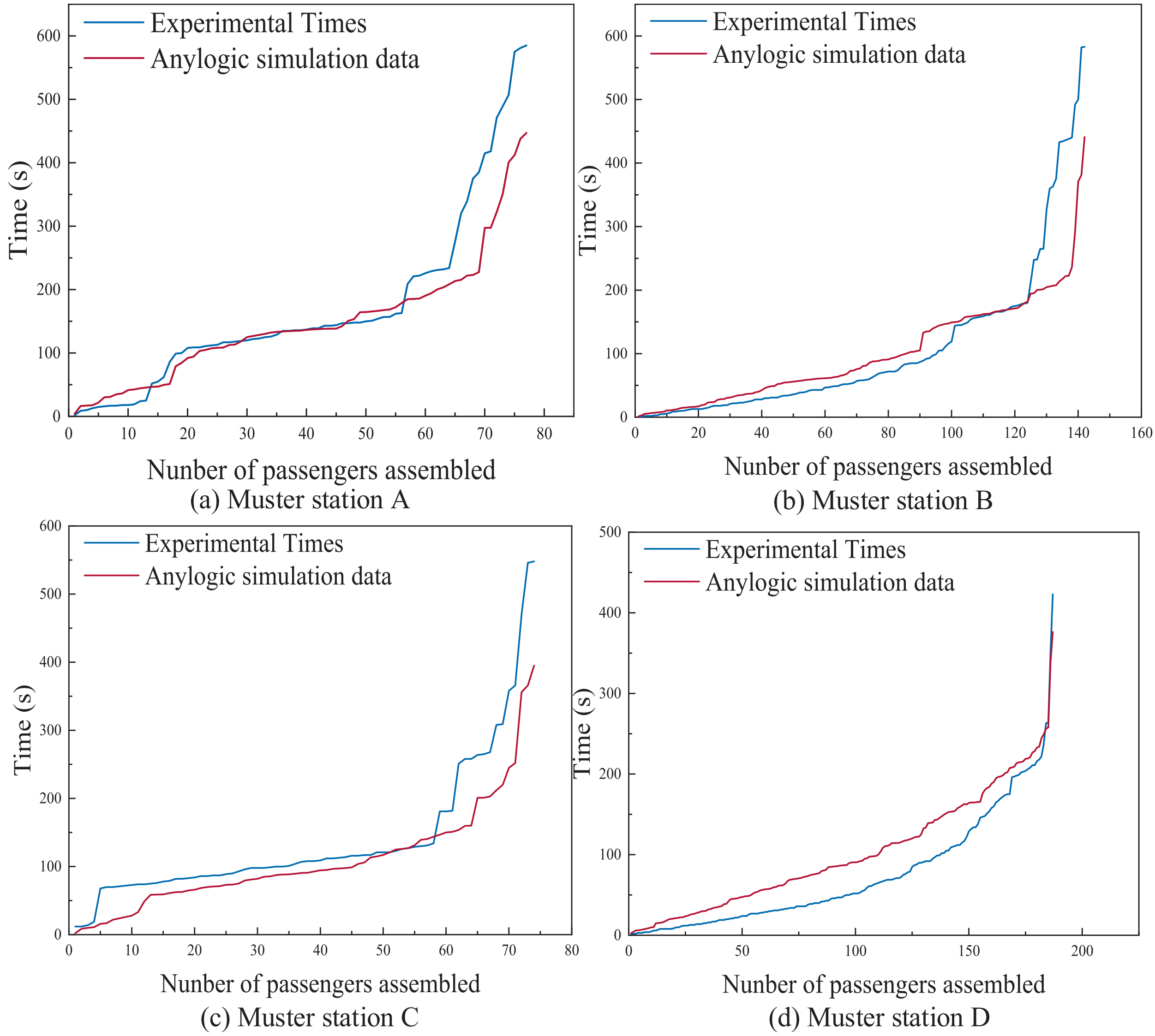


Fig. 6. Comparison of arrival curves of four muster stations.

scenarios.

4.1.3. *Collection of simulation data*

Based on the simulation model and parameter settings, different scenarios can be constructed using various combinations of variables, as detailed in Table 5. This study has developed 2,160 simulation scenarios, each numerically simulating the evacuation process to obtain both the TET and the individual MET of passengers. Following the IMO evacuation guide (IMO, 2016), in this study, the TET for 95 % of evacuees is used as simulation result to minimise the impact of extreme values and providing a more representative and reliable estimate of human evacuation time from passenger ships (Wang et al., 2022).

4.2. *ML model prediction results*

The simulation generated a dataset containing 2,160 evacuation scenarios under varying conditions. This dataset is randomly split into training (60 %), validation (20 %), and test (20 %) subsets. The training set is used to fit each ML model, while the validation set guided the hyperparameter tuning process. The test set is strictly reserved for evaluating generalization performance. Hyper-parameters tuning is performed via grid search within predefined parameter spaces tailored to each model. For instance, boosting models are tuned for parameters such as “n_estimators”, “learning_rate”, and “tree depth”. For models that support such optimization, such as NN, are additionally optimized for architecture settings like layer size and activation function.

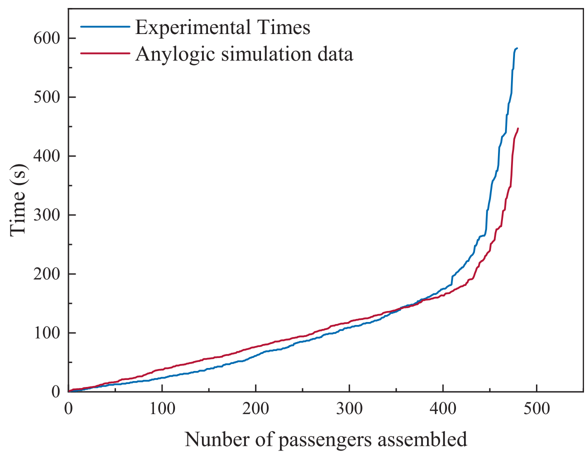


Fig. 7. Comparison of AnyLogic and SGVDS1 arrival curves.

Table 4
Computed values of simulation evaluation indicators.

Area	Evacuation time (s)		ERD	EPC	SC
	Experimental	Simulation			
Muster station A	585	447	0.251500	1.236361	0.787232
Muster station B	583	441	0.355621	1.221057	0.646337
Muster station C	548	395	0.279537	1.313032	0.859376
Muster station D	423	376	0.296948	0.816535	0.946288
Total	585	447	0.227910	1.162837	0.969737

Table 5
The evacuation simulation scenarios in this study.

Influential factors	Variable values	Number of variable values
Stair availability	Stair group 1, Stair group 2, Stair group 3, Stair group 4, Stair group 5	5
Ship heel angle (°)	0, 5, 10, 15	4
Ship trim angle (°)	−20, −15, −10, −5, 0, 5, 10, 15, 20	9
Small group situation	Yes, no	2
Total number of evacuees	400, 500, 600, 700, 800, 900	6

To prevent overfitting and enhance robustness, a 5-fold cross-validation scheme is applied during training. Optimal configurations and parameter values for each model are summarized in Table G of Appendix D. The prediction fitting results are displayed in Figs. 8 and 9. The evaluation indicator results for each model are shown in Tables 6 and 7.

As shown in Fig. 8, XGBoost, Attention-LightGBM, and GBDT present well-fitting curves for sparse time data, with adjusted R-squared values above 0.83. This indicates that these models effectively handle the task of predicting TET. Moreover, RF, AdaBoost, and DT perform moderately well but exhibit some errors, particularly in predicting extreme values, making them less suitable than the previous three models. The fitted line of the NN shows a noticeable bias, particularly in predicting low TET, leading to less accurate predictions. Although NN are typically adept at managing complex relationships, their performance in this specific task is inferior compared to the integrated methods. SVM and LR also perform poorly, with significant biases in their fitted lines and low adjusted R-squared values, suggesting difficulty in handling the complex non-linear relationships in this task.

Table 6 shows that models based on integrated learning generally exhibit better accuracy than other models. This superior performance can be attributed to the principle of integrated learning, which combines multiple base learners to reduce the instability of a single learner, improving prediction stability and accuracy (Feng et al., 2024a). Among them, the GBDT and Attention-LightGBM model performs relatively well across all four evaluation indicators, particularly in terms of accuracy indicators (i.e., MSE, MAPE, and sMAPE). Attention-LightGBM achieves the lowest MAPE and sMAPE, while GBDT attains the best MSE performance. Although Attention-LightGBM's inference time is slightly longer than DT's, it still offers a good balance between accuracy and timeliness, making it a strong candidate for predicting TET. Conversely, the SVM model has a higher error and the longest inference time, indicating it less suitable for applications that require timely predictions. Although DT does not achieve the lowest error rates, the shortest inference time among all models makes it particularly well-suited for timeliness scenarios where a rapid response is critical.

Fig. 9 shows that the models exhibit significantly better prediction accuracy for MET compared to TET, with better fitting curves and higher adjusted R-squared values across all models. Attention-LightGBM and RF exhibit similar predictive performance, making

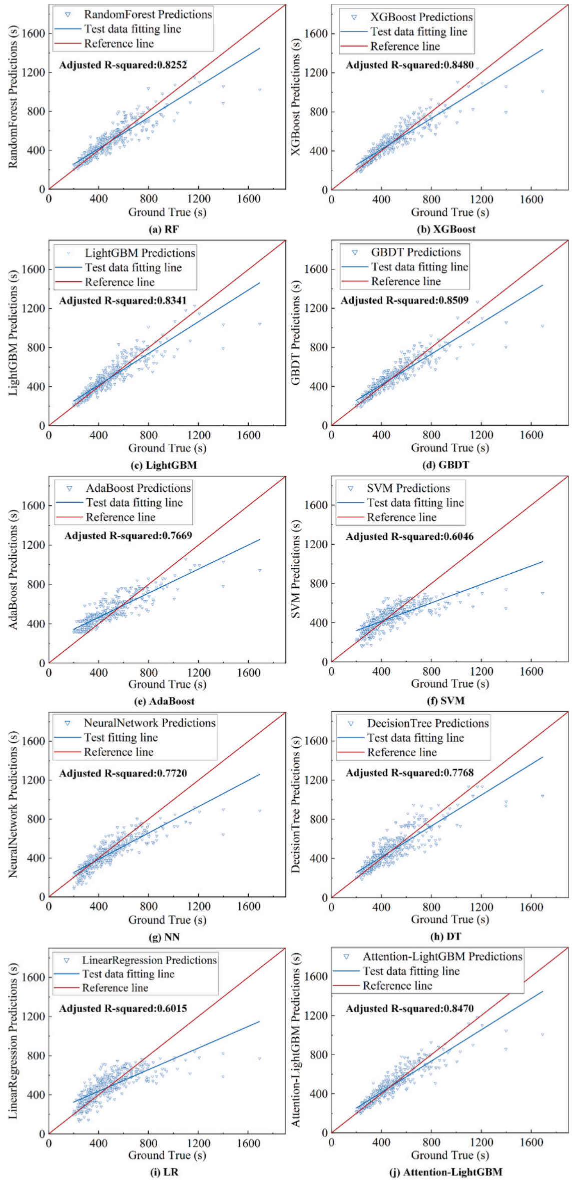


Fig. 8. Predictive fitting results of different models for TET.

them both reliable candidates for MET estimation in maritime evacuation applications. XGBoost, and GBDT emerge as the top performers, with their fitted lines aligning closely with the reference line and adjusted R-squared values reaching or exceeding 0.94. They demonstrate excellent predictive capability and are particularly suitable for handling complex nonlinear relationships. AdaBoost, and NN, although performing well, are not as well-fitted as the top three models when handling extreme values, and their adjusted R-squared values are slightly lower. Surprisingly, SVM performs robustly, with an adjusted R-squared value of 0.9455, explaining 94.55 % of the variation in the data. On the other hand, DT and LR show weaker performance: DT is prone to overfitting, and although LR has a higher adjusted R-squared value, the model faces challenges in accurately modelling complex relationships in this context.

As shown in Table 7, Attention-LightGBM and LightGBM are the best-performing models for MET, striking a good balance between accuracy and inference time. The comprehensive analysis of all model indicators shows only minor differences in their predictive performance. For instance, Attention-LightGBM achieves the lowest MAPE at 5.4952, while LightGBM follow closely with MAPE values of 5.5165, indicating that all three models perform comparably in terms of relative error. In terms of prediction speed, LightGBM and DT achieve the shortest inference time (0.001 s), while Attention-LightGBM's inference time is slightly longer at 0.0015 s, but still well within acceptable bounds for time-sensitive tasks.

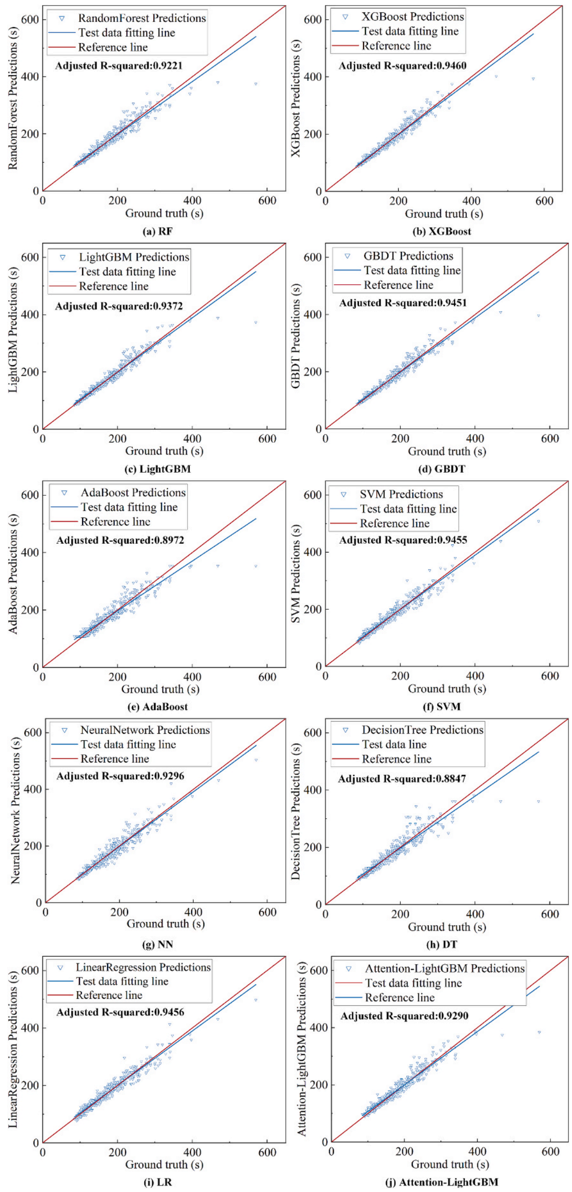


Fig. 9. Predictive fitting results of different models for MET.

Table 6
Predictive performance indicator values of different models for TET.

Model	MSE	MAPE	sMAPE	Inference Time (s)
RF	7232.9684	11.5533	11.2784	0.0367
XGBoost	6174.6646	9.9175	9.7854	0.0090
LightGBM	6332.0623	9.8171	9.6610	0.0120
GBDT	**6032.5218**	9.8383	9.6947	0.0060
AdaBoost	11472.0348	19.4777	17.3455	0.0668
SVM	37538.4488	27.9216	28.0450	0.1311
NN	10316.1942	13.2495	13.6140	0.0047
DT	8655.6277	13.4826	13.1297	**0.0032**
LR	17120.4600	19.7754	18.9459	0.0040
Attention-LightGBM	6224.0593	**9.7908**	**9.6580**	0.0035

Table 7
Predictive performance indicator values of different models for MET.

Model	MSE	MAPE	sMAPE	Inference Time (s)
RF	440.6888	7.1589	6.9675	0.0080
XGBoost	327.0342	6.3869	6.2624	0.0045
LightGBM	**287.3845**	5.5165	5.4273	**0.0010**
GBDT	327.5062	6.3213	6.1957	0.0020
AdaBoost	680.9783	10.6000	10.1130	0.0107
SVM	1195.4164	13.5496	13.199	0.0277
NN	412.9685	7.5967	7.7671	0.0020
DT	528.5615	8.2752	8.0935	**0.0010**
LR	543.9408	9.2967	9.0521	0.0010
Attention-LightGBM	290.0798	**5.4952**	**5.4175**	0.0015

4.3. Prediction model selection

4.3.1. Data collection

In Section 3.4, this study introduces the new weighted multi-indicator selection model based on AHP and EWM to select the optimal ML model for predicting human evacuation time from passenger ships. To facilitate this, five experts with expertise in emergency evacuation and ML are invited to form a panel, details are provided in Table D of Appendix B. The experts use the AHP method to conduct pairwise comparisons and determine the relative importance of four ML indicators. Meanwhile, EWM is employed as a complementary approach to calculate objective weights based on data variability. This ensures the robustness of weight assignment and achieves a balanced integration of subjective and objective perspectives.

4.3.2. Scenario analysis

Predicting the time of human evacuation from passenger ships necessitates both fast response and high accuracy to facilitate immediate decision-making during unexpected situations and ensure effective crowd management. Additionally, the extent of deviation in prediction results can impact the management of evacuation efforts. Therefore, determining the appropriate model evaluation indicators and their respective weights is essential for selecting the optimal prediction model. To achieve these goals, this study defines eight evaluation scenarios, grouped into two categories: Scenarios 1–5 form the control group, which reflect common single-objective or equal-weighted benchmarks, while Scenarios 6–8 constitute the experimental group, designed based on comprehensive weighting strategies that combine subjective and objective factors to better represent complex real-world demands, as show in Table 8.

S1- S5: Control groups

The control group consists of five baseline scenarios designed to provide reference models under different single-objective or balanced evaluation strategies. In Scenario 1, all four indicators (i.e., MSE, MAPE, sMAPE, and Inference time) are equally weighted, offering a general-purpose benchmark when no specific preference is assumed. In Scenarios 2–5, each scenario emphasizes a single indicator by assigning it a full weight of 1 while setting the weights of the other three indicators to 0. Specifically, Scenario 2 prioritizes MSE, Scenario 3 focuses on MAPE, Scenario 4 emphasizes sMAPE, and Scenario 5 highlights inference time.

The performance of each model in these scenarios is normalised and aggregated based on the assigned weights, enabling comparisons across different evaluation priorities. This group of scenarios is suitable for model screening under well-defined evaluation objectives or for understanding model performance trade-offs under distinct measurement focuses.

S6: Accurate prediction

This scenario applies when accurate prediction of human evacuation time from passenger ships is required, either before or after an emergency. Examples include developing evacuation plans or conducting post-evacuation analyses. Highly accurate models help stakeholders optimize evacuation route design, allocate resources, and improve emergency plans, among other critical tasks. Therefore, accuracy-related indicators (e.g., MSE, MAPE, sMAPE) are given higher weights in this scenario, while inference time is weighted lower.

S7: Timeliness prediction

In emergency evacuation scenarios, a fast response is similarly one of the most important requirements. This scenario applies when

Table 8
Weighting values between indicators in different scenarios.

Scenario	Group	MSE	MAPE	sMAPE	Inference Time
S1	Control Group 1	0.25	0.25	0.25	0.25
S2	Control Group 2	1	0	0	0
S3	Control Group 3	0	1	0	0
S4	Control Group 4	0	0	1	0
S5	Control Group 5	0	0	0	1
S6	Experimental Group 1	0.18123698	0.29293864	0.44729441	0.07852997
S7	Experimental Group 2	0.11543135	0.18113120	0.25333038	0.45010707
S8	Experimental Group 3	0.18153255	0.21383688	0.19825443	0.40637615

real-time feedback is needed during an emergency evacuation. For example, during on-site evacuation decisions, the timeliness prediction model can quickly estimate emergency situations with dynamically adjusted evacuation scenarios. Additionally, it can monitor evacuation progress, allocate resources, and prioritize evacuation of hazardous areas in time. Therefore, inference time is the most important indicator in the timeliness prediction scenario. A long inference time may delay evacuation instructions, impacting passenger safety. On the other hand, accuracy indicators still need be considered. Experts need to balance timeliness performance indicators with accuracy-related indicators.

S8: Objective prediction

In Section 3.4, the EWM-driven evaluation approach is proposed to compare the objective weight assignment with subjective interference. This scenario is particularly suited to situations where expert knowledge is limited or the importance of indicators is difficult to determine, providing a more transparent and equitable basis for model selection.

This multi-scenario control group not only complements the single equal-weighted benchmark, but also enhances the flexibility and adaptability of the proposed evaluation framework. It allows for clear identification of performance trade-offs and supports model selection under diverse real-world priorities. Generally, in the Control Group (S1–S5), different weighting strategies are applied to establish baselines for comparison. S1 assigns equal weights to all four indicators, while S2–S5 assign a weight of 1 to one individual indicator and 0 to the others, respectively. This setup establishes a baseline model for comparing the impact of different weightings and results in the other scenarios. The Accurate Prediction (S6) focuses on model accuracy, particularly its performance in terms of relative error (i.e., MAPE and sMAPE). The weight of inference time is comparatively reduced, reflecting a prioritization of accuracy over the model's timeliness performance. The Timeliness Prediction (S7) focuses on the model's timeliness performance, giving inference time the highest weight. While accuracy is still considered, quick response is prioritized in this scenario, which is designed for emergency evacuations requiring fast decision-making. The Objective Prediction (S8) emphasizes the informativeness of the indicators themselves, with weights assigned objectively. In this scenario, with fully considering model accuracy, greater weight is assigned to the model's inference time.

Tables E and F of Appendix C show the normalised results of the actual values for the evaluation metrics of TET and MET in the prediction model. Correspondingly, based on the method proposed in Section 3.4, the overall performance of the models in predicting TET and MET across different scenarios is evaluated and shown in Tables 9 and 10.

As shown in Table 9, significant performance differences exist across models and scenarios when predicting TET for human evacuation from passenger ships. Overall, LightGBM consistently outperforms the other models in S6 and S7, offering a balance of accuracy and inference time that makes them well-suited for handling multidimensional evacuation requirements. Whereas models like SVM and AdaBoost generally underperform, especially in S7. In S8, DT leads the performance ranking, suggesting that simple tree-based models may adapt well to purely data-driven weight settings without expert interference.

In the Control Groups (S1–S5), where each evaluation indicator is either equally or singly emphasized, Attention-LightGBM achieves the highest score in three of five scenarios (S1, S3, S4), demonstrating its strong adaptability in accuracy-prioritized contexts. DT ranks first in S5, highlighting its advantage in inference speed, while GBDT shows leading performance in S2.

In predicting the MET, the performance differences across models in various scenarios reflect their strengths and weaknesses in terms of accuracy and timeliness performance. In the control group (S1–S5), LightGBM and Attention-LightGBM consistently rank at the top. In contrast, SVM and AdaBoost exhibit the lowest scores across all five control scenarios, highlighting their limited suitability for either accurate or timely MET prediction.

Generally, LightGBM demonstrates superiority in most scenarios. Notably, its top performance in S8 underscores its robustness in objective, data-driven prediction. These findings emphasize LightGBM's adaptability to varying requirements, making it especially well-suited for tasks involving interference-free predictions. For S6 and S7, Attention-LightGBM delivers the top score, results are 20–45 % higher than models like RF and XGBoost, indicating that attention mechanisms can help retain accurate performance even under time-sensitive constraints.

4.4. *Discussions*

4.4.1. *Comparative analysis of subjective and objective weighting methods*

In multi-criteria decision-making issues, both subjective and objective weighting methods have their own advantages and

Table 9
Evaluation of different models' performance by predicting TET.

Model	S1	S2	S3	S4	S5	S6	S7	S8
RF	0.0897	0.1182	0.1109	0.1113	0.0185	0.1054	0.0705	0.0748
XGBoost	0.1178	**0.1385**	0.1292	0.1283	0.0754	0.1241	0.0977	0.1088
LightGBM	0.1130	0.1351	0.1305	**0.1299**	0.0567	**0.1354**	**0.1496**	0.1012
GBDT	0.1287	**0.1418**	0.1302	0.1295	0.1132	0.1261	0.1117	**0.1253**
AdaBoost	0.0557	0.0746	0.0658	0.0724	0.0102	0.0661	0.0452	0.0461
SVM	0.0297	0.0228	0.0459	0.0448	0.0052	0.0381	0.0252	0.0195
NN	0.1040	0.0829	0.0967	0.0922	0.1441	0.0996	0.1294	0.1102
DT	0.1258	0.0988	0.0950	0.0956	**0.2137**	0.1035	0.1303	**0.1393**
LR	0.0873	0.0500	0.0648	0.0662	0.1681	0.0728	0.1113	0.1154
Attention-LightGBM	**0.1483**	0.1374	**0.1309**	**0.1300**	0.1950	0.1289	0.1291	0.1244

Table 10
Evaluated values of MET predicted by different models.

Model	S1	S2	S3	S4	S5	S6	S7	S8
RF	0.0820	0.0955	0.1038	0.1046	0.0243	0.0964	0.0725	0.0701
XGBoost	0.1012	0.1287	0.1163	0.1164	0.0432	0.114	0.0937	0.0889
LightGBM	**0.1525**	**0.1464**	0.1347	0.1343	**0.1947**	0.1395	0.1226	**0.1611**
GBDT	0.1152	0.1285	0.1175	0.1176	0.0970	0.1225	0.1234	0.1112
AdaBoost	0.0555	0.0618	0.0701	0.0721	0.0182	0.0653	0.0484	0.0479
SVM	0.0383	0.0352	0.0548	0.0560	0.007	0.048	0.0324	0.0321
NN	0.0979	0.1019	0.0978	0.0938	0.0979	0.0936	0.1133	0.0978
DT	0.1135	0.0796	0.0898	0.0901	0.1946	0.0943	0.1049	0.1306
LR	0.1079	0.0774	0.0799	0.0805	0.1938	0.0868	0.1003	0.1259
Attention-LightGBM	0.1360	0.1451	**0.1352**	**0.1345**	0.1292	**0.1396**	**0.1886**	0.1344

disadvantages (Chen, 2020). Based on their characteristics and the applications in this study, this section aims to further discuss the following three aspects: 1) the rationality of different methods for assigning weights to various indicators, 2) the differences in the performance of models selected by each method, and 3) the potential for hybrid method applications. For a clearer comparative analysis, the weights assigned to each evaluation indicator under different scenarios and the corresponding model performance for each task are visualised for the control group (S1) and all experimental groups, as shown in Figs. 10 and 11.

As shown in Fig. 10, there is a clear difference in the indicator weights between the AHP-based scenarios (S6 and S7) and the EWM-based scenario (S8), although S7 and S8 exhibit a similar trend. Specifically, AHP incorporates the subjective knowledge of experts and adjusts weight assignments based on different application contexts (i.e., accurate prediction or real-time prediction). The flexibility and relevance of AHP can assess the importance of tailoring indicators to different application needs, demonstrated by the significant difference of the yellow line (S7) and the blue line (S6) within Fig. 10. However, one notable point is that this method could introduce some degree of subjective bias (Li et al., 2024a). On the other hand, EWM eliminates human subjectivity by assigning weights based on the objective data (i.e., the calculated information entropy). By considering data variability, EWM assigns greater weight to inference time, reflecting stronger differentiation. When viewed from the perspective of safety management in human evacuation from passenger ships, this data-driven weighting approach prefers timeliness requirements in real-world evacuation scenarios (Li et al., 2024b).

Fig. 11 shows the performance of different models across the two tasks. Overall, LightGBM and Attention-LightGBM consistently exhibit stable performance, while AdaBoost and SVM perform poorly. However, the differences between the methods warrant further attention. For predicting TET, DT and GBDT have the priority in the EWM-driven method (S8). Their simpler model structures exhibit the functional strength, satisfying well the higher weight of inference time. However, one concern here is that the EWM-driven approach seems to place less emphasis on accuracy. For predicting MET, Attention-LightGBM's performance under Scenario 7 significantly surpasses other models. This can be attributed to the higher subjective weight assigned to inference time by AHP, which potentially overemphasizes a single characteristic of the models, leading to a skewed evaluation that favours simpler models.

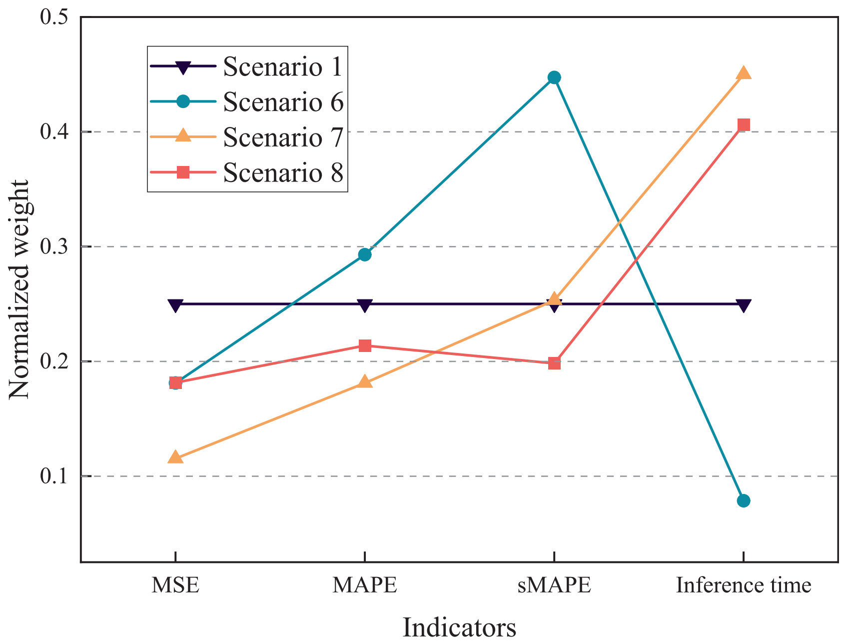


Fig. 10. Weighting values of indicators in different scenarios.

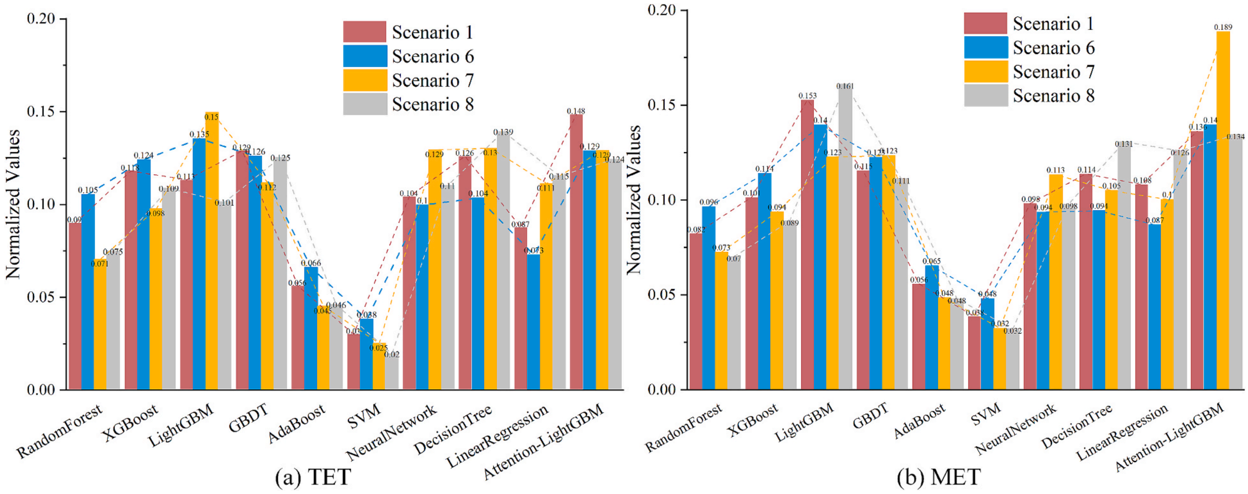


Fig. 11. Comparison of model evaluation values for predicting two different tasks.

Therefore, these disparities guide one potential reflection that is to integrate both subjective and objective perspectives in model selection.

In addressing multi-criteria decision-making issues, single methods may have limitations, while hybrid weighting methods offer

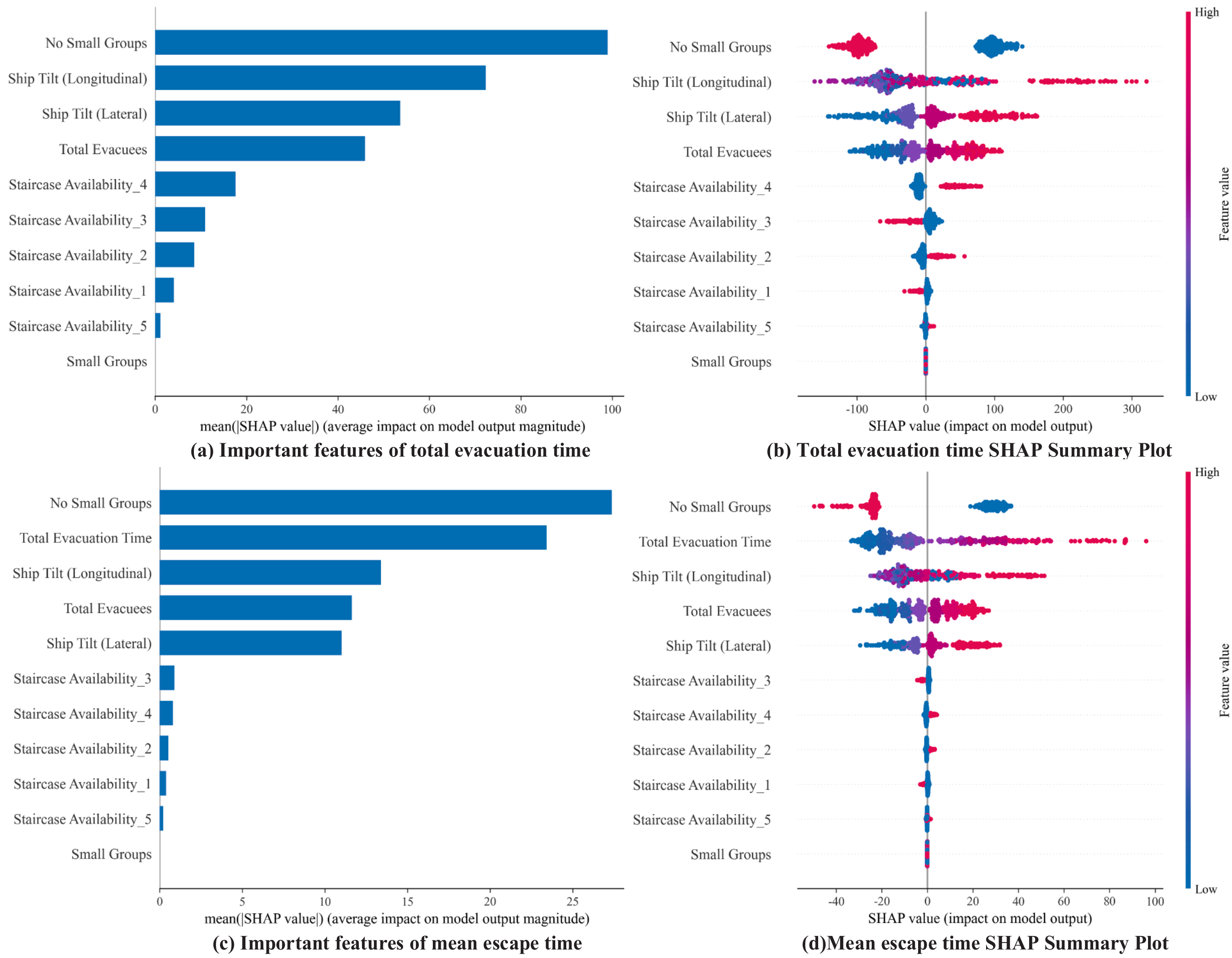


Fig. 12. LightGBM-based SHAP analysis of factors influencing evacuation.

significant potential for application. Increasingly, studies have recognised the complementary nature of AHP and EWM and applied both of them to weight assignment to enhance the flexibility and adaptability of model selection, thereby improving the robustness of multi-criteria decision-making (Liu et al., 2020). By incorporating EWM to adjust AHP results, the subjective weight assignment can better align with specific practical scenarios while more accurately reflecting data characteristics and mitigating the influence of subjective bias. However, most of these methods are used to optimize or adapt existing models. In the future, further exploration is needed to determine which ML models are best suited for such multi-criteria decision-making issues in real-world applications, especially in the context of hybrid weighting methods.

4.4.2. Interpretability of ML models

The interpretability of ML has always been a topic worth exploring as it offers a deeper understanding of models. In this study, the two best-performing ML models, LightGBM and GBDT, are selected for interpretability analysis. The accuracy of the ML models is explained by analysing the effect of each factor on the model's predictive performance (Zhao et al., 2020a). To achieve this, SHapley Additive exPlanations (SHAP) is employed— a model-agnostic interpretability approach grounded in cooperative game theory. Originally proposed by Lundberg and Lee (2017), SHAP quantifies the contribution of each input feature to a model's output by computing Shapley values, which represent the average marginal contribution of that feature across all possible feature combinations. By capturing both main effects (individual impacts) and interaction effects (joint impacts with other features), SHAP provides comprehensive local and global interpretability (Jiao et al., 2025). In this study, SHAP is employed to evaluate the influence of each input variable on the predicted evacuation times, thereby enhancing ML model transparency and supporting informed decision-making. The SHAP bar plot and SHAP summary plot for both models are shown in Figs. 12 and 13.

As shown in Fig. 12 and Fig. 13, the SHAP analysis provides clear insights into how different input features influence the predictions of the LightGBM and GBDT model. Panel (a) and (b) show the feature importance and SHAP summary plot for TET, while panels (c) and (d) present the corresponding results for MET.

Across both TET and MET predictions, the factor of small groups ("No Small Groups") is consistently the most influential factor. The SHAP values for this factor are predominantly negative, indicating that when the factor "small groups" is absent, the model predicts significantly shorter evacuation times. This aligns with real-world evacuation dynamics: individuals moving alone or in smaller units tend to move more efficiently than groups, which may wait for slower members and cause localized congestion. Comparatively, some studies can cross validate these findings. Specifically, a number of studies have shown that small group behaviour inhibits crowd evacuation efficiency (Haghani et al., 2019; Lu et al., 2017; Ren et al., 2022). People take longer to begin moving or orienting within a group compared to moving alone.

Ship inclination is another highly important factor. The SHAP summary plots show that higher angles of inclination (represented by red colour toward the right in Figs. 12 and 13) lead to positive SHAP values, thus contributing to longer predicted evacuation times. This is likely due to balance issues and slower movement in inclined conditions, particularly on stairways and narrow corridors. Additionally, some studies also show that ship inclination is an important factor affecting evacuation efficiency (Fang et al., 2022a; Fang et al., 2024b; Fang et al., 2023a; Sun et al., 2018a). An increase in the ship's list angle significantly decreases the average individual movement speed. These peer-recognised conclusions widely come from the result of actual experiments and reasoning from mathematical models such as cellular automata. This further demonstrates that ML models are useful for predicting human evacuation time from passenger ships reliably.

4.5. Implications

4.5.1. Theoretical implications

(1) Selection of key factors for building simulation models

This study creates an advanced and diverse simulation scenario by scientifically selecting and analysing key factors that influence evacuation efficiency. The accuracy of simulation outputs is highly dependent on the precise modelling of these critical factors (i.e., inputs) that affect human evacuation from passenger ships. A particularly complex and challenging task involves carefully considering the diverse passenger attributes defined in the IMO evacuation guidelines, while comprehensively integrating the ship's internal layout and external environmental factors into the simulation model. Furthermore, this study takes into account a range of complex scenarios following passenger ship accidents, such as deck inclination due to flooding, fires, and potential behaviour responses from passengers. By comprehensively incorporating these factors, this study significantly enhances the realism of simulation data and the multidimensionality of simulation scenarios, providing a more robust theoretical foundation for in-depth investigations into evacuation efficiency.

(2) Application of ML models in the field of human evacuation from passenger ship

The application of ML in the field of evacuation analysis is emerging, and to the best of our knowledge, this study is the first to utilise ML for predicting the human evacuation time from passenger ships. Traditional methods often require substantial time to simulate and analyse various evacuation scenarios, whereas ML models can process and analyse multi-dimensional data rapidly, generating near real-time evacuation predictions once the model is trained. This speed advantage is crucial for emergency response and real-time decision-making, especially during critical situation. Further to this, ML models also have the ability to continuously enhance the accuracy of prediction models by training on large volumes of historical, simulated, and actual evacuation data. This breakthrough not only opens new prospects for research in evacuation modelling but also encourages the broader application of Artificial Intelligence (AI) technology in emergency management.

(3) New decision making tools to guide the selection of ML models

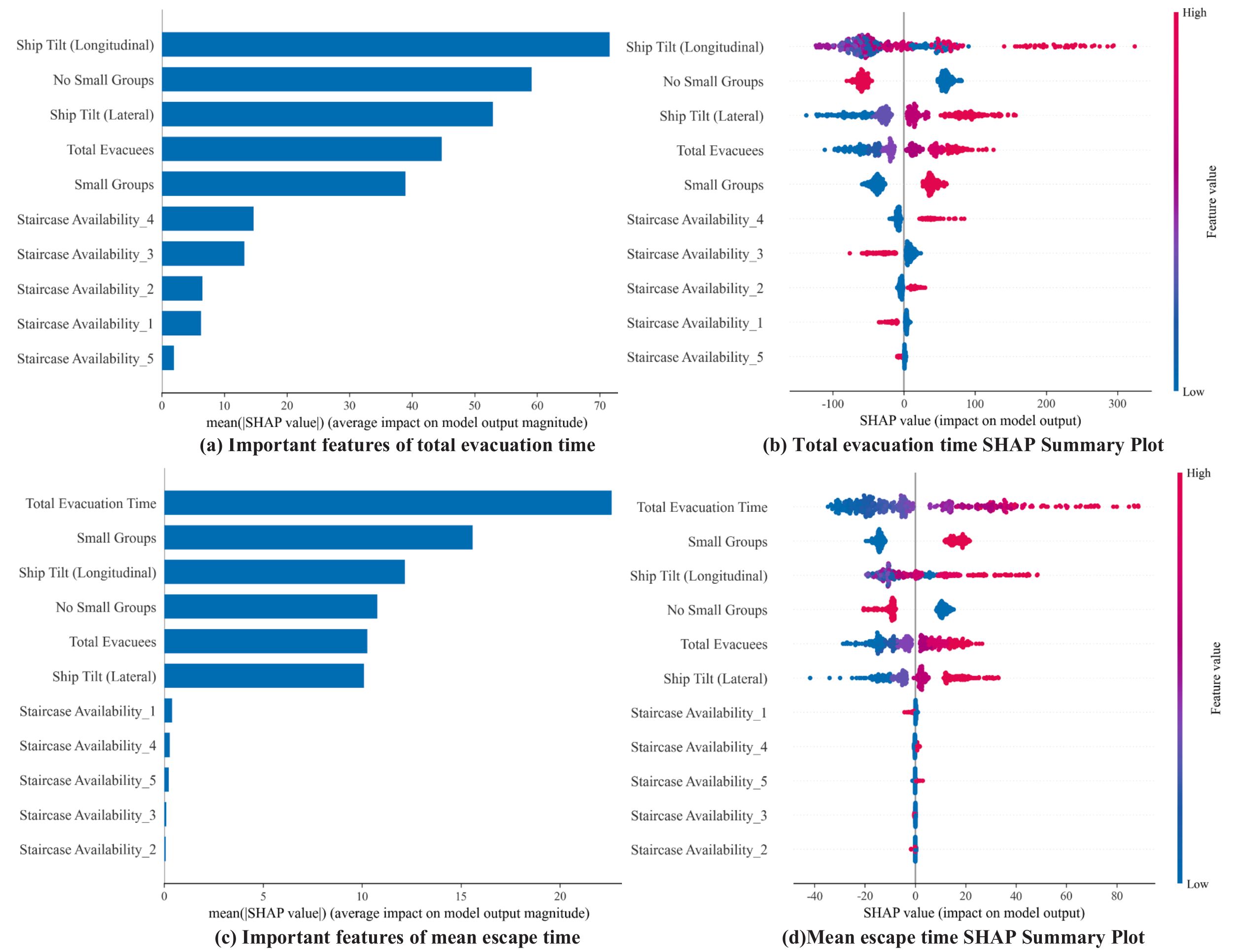


Fig. 13. GBDT-based SHAP analysis of factors influencing evacuation.

This study proposes two weighted ML model selection methods using AHP and EWM. In ML applications, selecting appropriate evaluation metrics is often a challenge. Through the development of reasonable weighting systems, the proposed methods provide a solid theoretical foundation for selecting the best ML models across various scenarios. Specifically, in the context of human evacuation from passenger ships, this method can aid in selecting ML evaluation indicators that best align with evacuation safety requirements and in identifying the most effective ML model for predicting evacuation time. Further, this method provides innovative ideas for researchers to consider combining AI techniques and human knowledge in emergency evacuation prediction modelling.

4.5.2. Practical implications

(1) Implications for maritime safety policy and regulatory frameworks

This study proposes a holistic, multi-indicator evaluation framework for predictive modelling in human evacuation from passenger ships, offering substantial insights for enhancing maritime safety policies and regulatory systems. Current international maritime standards, such as IMO MSC.1/Circ.1533, predominantly rely on static thresholds and pre-defined clearance benchmarks, without accounting for scenario-specific dynamics or real-time adaptation (Wang et al., 2023a). The proposed framework addresses this gap by enabling comparative and weighted evaluation of multiple ML models based on both accuracy and operational suitability.

In policy terms, these findings support the revision of international safety codes to include adaptive, data-driven model integration mechanisms. For instance, regulatory authorities such as the IMO and national maritime administrations may consider establishing formal guidelines for validating and certifying predictive evacuation models under different vessel layouts and passenger densities (Cao et al., 2024). Furthermore, the results provide a technical basis for mandating the inclusion of intelligent predictive systems as part of Safety Management Systems (SMS), especially for large cruise ships or ro-ro passenger ships. This transition from experience-based planning to model-supported decision-making can systematically reduce human error, improve response coordination, and enhance resilience under emergency conditions.

In addition, policymakers should consider developing standardised evaluation metrics for model performance within regulatory frameworks, ensuring that algorithms deployed in real-world settings meet minimum safety and robustness thresholds. These standards can be integrated into the design approval process for new vessels and periodically repaired during safety audits.

(2) Implications for onboard evacuation planning and bridge team operations

The practical implications for onboard operations covering both macro-level strategy formulation and micro-level operational refinement. By leveraging high-fidelity TET predictions generated by ML models, bridge teams and safety officers can make pre-emptive and evidence-based decisions on route designation, staging area allocation, and initial evacuation order. This data-driven approach is particularly valuable in complex spatial layouts or scenarios with partial facility failures (e.g., blocked stairs or unserviceable exits).

At the micro level, the individualised MET predictions enable fine-grained density management, helping reduce localised congestion and mitigate the risk of bottlenecks near lifeboat stations or narrow corridors. This is particularly critical for high-density risk zones, such as the entrances to mustering stations, and aligns with IMO's push for proactive congestion management strategies.

Additionally, the analysis of small-group behaviours enriches the understanding of crowd dynamics at sea, which differ significantly from land-based evacuations due to enclosed spaces and limited exit points (Fang et al., 2023b). As this study reveals, groups with pre-existing social ties (e.g., families, tour groups) tend to exhibit collective behaviour and emotional interdependence, which can either facilitate or hinder evacuation. This calls for bridge teams to identify such groups early during mustering and assign targeted guidance crew members.

(3) Implications for safety management systems and personnel training

The predictive insights generated by this study offer tangible guidance for improving the SMS of passenger ships and the training practices of both maritime authorities and shipping companies. In current industry practice, emergency preparedness often focuses on general compliance with SOLAS protocols and routine drills, but lacks refined, data-informed adaptation to scenario-specific evacuation risks (Wang et al., 2025). The integration of ML-based evacuation time predictions presents a new opportunity to elevate the precision and responsiveness of onboard safety protocols.

First, the research findings support the stratified design of training programs that reflect different risk levels across ship zones. For instance, areas with historically higher MET—such as narrow corridor intersections—can be designated as “priority zones” in drills, where additional crew deployment or rehearsed rerouting procedures should be enhanced. These data-driven strategies should be incorporated into each vessel's SMS documentation and used to guide minimum personnel requirements and zone-specific responsibilities under emergency conditions.

Second, this study highlights the need for modular, scenario-based training curricula that simulate complex but plausible emergencies, such as obstructed escape routes, passenger gathering, or stair unavailability. Instead of relying solely on full-ship drills, shipping companies could employ small-group simulations calibrated with model-derived congestion predictions, allowing crew members to practice crowd management in high-density zones. This aligns with international calls for competency-based training in emergency operations.

Lastly, the predictive framework developed in this study can also inform the safety auditing processes conducted by maritime authorities and classification societies. They may consider integrating predictive performance assessments into SMS evaluations and operational audits, ensuring that safety management practices are no longer merely reactive, but also anticipatory and data-supported. By embedding ML-informed scenario planning into routine safety management, shipping companies can move towards more adaptive and evidence-based risk governance.

5. Conclusions

In the event of an emergency of a passenger ship, rapid evacuation of personnel is crucial to ensure passenger safety. Accurate and timely prediction of human evacuation time from passenger ships is essential to develop efficient evacuation plans in time. In this study, seven key factors affecting human evacuation from passenger ships are identified and incorporated, leading to the construction and validation of a new evacuation simulation model. Following this, ten different ML models are selected for a comprehensive analysis, focusing on the prediction of TET and MET. To systematically assess the performance of different models, this study introduced two new weighted selection models based on AHP and EWM. This approach compares the ML models from two perspectives—accuracy prediction and timeliness prediction—and selects the optimal model. By examining interpretability and a comparative discussion, the robustness of ML models in predicting human evacuation time from passenger ships is further assured. The results indicate that XGBoost, LightGBM, and GBDT models exhibit more balanced (between accuracy and timeliness) and superior performance in prediction of human evacuation time from passenger ships.

Although this study reveals the promising applications of ML models in improving emergency plans and enhancing evacuation efficiency, several areas warrant further refinement. These include how to balance subjective and objective data for holistic evaluation, and how to address the reliance on simulation-generated data, which may affect the generalizability of the results to real-world ship types. Nevertheless, the proposed framework is adaptable and can be retrained for different vessel configurations. Furthermore, integrating multiple models or algorithms into hybrid AI solutions presents a promising direction for enhancing prediction accuracy and supporting more complex, real-time decision-making. Future research could focus on the development and implementation of such hybrid AI approaches to enhance the robustness and accuracy of decision-making processes.

CRediT authorship contribution statement

Zhiwei Zhang: Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Zhengjiang Liu:** Writing – review & editing, Validation, Supervision, Funding acquisition. **Zirui Zhou:** Software, Methodology, Conceptualization. **Xinjian**

Wang: Writing – original draft, Validation, Supervision, Methodology, Funding acquisition, Conceptualization. **Arnab Majumdar:** Writing – review & editing, Validation, Investigation, Formal analysis. **Yuhao Cao:** Writing – review & editing, Validation, Investigation, Formal analysis. **Zaili Yang:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

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

Acknowledgements

The authors gratefully acknowledge support from the National Natural Science Foundation of China [Grant No. 52571403, 52101399], International Exchanges program of Royal Society, UK [Grant No. IEC\NSFC\233298] and the MarRI-UK Core Project [Grant No. MarRI-UK_CORE_2024_02]. This research is also funded by a European Research Council project under the European Union's Horizon 2020 research and innovation programme [Grant No. TRUST CoG 2019 864724].

Appendix A

Table A
Gender and age ratio of passengers.

Gender and age of the personnel	Percentage (%)
Women aged under 30	7
Women aged 30 to 50	7
Women aged over 50	16
Females aged 50 or above – mobility impaired (1)	10
Females aged 50 or above – mobility impaired (2)	10
Males aged below 30	7
Males aged 30 to 50	7
Males aged 50 or above	16
Male aged 50 or above – mobility impaired (1)	10
Males aged over 50 – mobility impaired (2)	10

Table B
Gender and age of passengers in relation to speed.

Genders	Age (years)	Speed (m/s)
Male	2–5	0.3 + 0.16* age
	5–12.5	0.8 + 0.06* age
	12.5–18.8	1.45 + 0.008* age
	18.8–39.2	1.78–0.01* age
	39.2–70	1.75–0.009* age
Female	2–8.3	0.5 + 0.06* age
	8.3–13.3	0.67 + 0.04* age
	13.3–22.25	0.94 + 0.02* age
	22.25–37.5	1.78–0.018* age
	37.5–70	1.45–0.01* age

Table C
The effect of the compounding effect of the angle of inclination on the attenuation of passenger velocities.

Heel	Trim	Velocity attenuation factor	Heel	Trim	Velocity attenuation factor
0	−20	0.962894217	10	−20	0.758963293
0	−15	1.080224104	10	−15	0.905809273
0	−10	1.092938115	10	−10	0.945272244
0	−5	1.059641411	10	−5	0.951296169
0	0	1	10	0	0.977726282
0	5	0.970990645	10	5	0.881941321
0	10	0.906680407	10	10	0.828745167
0	15	0.879709394	10	15	0.724641725
0	20	0.762464913	10	20	0.649786771
5	−20	0.820278649	15	−20	0.671855517
5	−15	0.990964113	15	−15	0.754180592

(continued on next page)

Table C (*continued*)

Heel	Trim	Velocity attenuation factor	Heel	Trim	Velocity attenuation factor
5	−10	0.972021203	15	−10	0.839500549
5	−5	1.02172143	15	−5	0.834689381
5	0	0.997904722	15	0	0.961567587
5	5	0.933321187	15	5	0.793199456
5	10	0.877671053	15	10	0.745913353
5	15	0.811937392	15	15	0.655907489
5	20	0.664522043	15	20	0.551217596

Appendix B

Table D
Specifics of experts in the field.

Expert	Employer	Major	Working Period	Reason of Employment
1	Shipping company	Passenger ship evacuation	10	Specializing in the field of passenger ship evacuation.
2	Research institute	Passenger ship emergency response	8	Developed emergency response systems for passenger ships.
3	Industrial sector	Passenger ship design	10	Design engineers specializing in the overall structure and layout of passenger ships.
4	Maritime administration	Passenger ship safety	8	Experienced in passenger ship safety, with a focus on evacuation emergency management.
5	University	Passenger ship evacuation	21	Specialized in passenger ship evacuation prediction, with a deep understanding of artificial intelligence methods.

Appendix C

Table E
Results of the normalisation of model indicators for predicting TET.

Model	MSE	MAPE	sMAPE	Inference time (s)
RandomForest	0.1182428	0.1109113	0.11128127	0.01852925
XGBoost	0.13850897	0.12920589	0.12825989	0.07535306
LightGBM	0.13506602	0.13052765	0.12991207	0.05668095
GBDT	**0.14177262**	0.13024515	0.12945985	0.11320898
AdaBoost	0.07455055	0.06578775	0.07235727	0.0101793
SVM	0.02278321	0.04589268	0.04475217	0.00518794
NeuralNetwork	0.08290329	0.09671298	0.09219007	0.14410702
DecisionTree	0.09880814	0.09504084	0.09559038	**0.21365175**
LinearRegression	0.04995464	0.06479767	0.06624537	0.16806969
Attention-LightGBM	0.13740975	**0.13087811**	**0.12995167**	0.19503206

Table F
Results of the normalisation of model indicators for predicting MET.

Model	MSE	MAPE	sMAPE	Inference time (s)
RandomForest	0.09548739	0.10378709	0.10461605	0.02428539
XGBoost	0.12867197	0.11633115	0.11639428	0.04320658
LightGBM	**0.14642448**	0.13468654	0.13430567	**0.1947233**
GBDT	0.1284865	0.11753994	0.11764654	0.09698002
AdaBoost	0.06179364	0.07009784	0.07207631	0.01819321
SVM	0.03520123	0.05483582	0.05598452	0.007031
NeuralNetwork	0.1018967	0.09780598	0.09384529	0.09787518
DecisionTree	0.07961254	0.08978641	0.09006098	0.19463046
LinearRegression	0.0773616	0.07992143	0.08052343	0.19384492
Attention-LightGBM	0.14506394	**0.1352078**	**0.13454693**	0.12922994

Appendix D

Table G
Hyperparameter search ranges and optimal values.

Model	Hyperparameters Tuned	Search Range / Setting	Optimal Configuration	Search Method
RF	n_estimators	[100, 200]	100	Grid Search
	max_depth	[10, 20]	10	
	min_samples_split	[2, 5]	2	
	min_samples_leaf	[1, 4]	1	
XGBoost	n_estimators	[100, 200]	200	Grid Search
	max_depth	[3, 4]	3	
	learning_rate	[0.05, 0.1]	0.05	
LightGBM	n_estimators	[100, 200]	200	Grid Search
	num_leaves	[20, 31]	20	
	learning_rate	[0.05, 0.1]	0.05	
GBDT	n_estimatorsmax_depthlearning_rate	[100, 200]	100	Grid Search
		[3, 4]	3	
		[0.05, 0.1]	0.1	
AdaBoost	n_estimators	[100, 200]	200	Grid Search
	learning_rate	[0.05, 0.1]	0.1	
SVM	C	[1, 10]	10	Grid Search
	epsilon	[0.1]	0.1	
NN	hidden_layer_sizes	[(50, 50), (100,)]	(100,)	Grid Search
	activation	['relu', 'tanh']	relu	
	learning_rate_init	[0.001, 0.01]	0.01	
DT	max_depth, min_samples_split, min_samples_leaf	[10, 20]	10	Grid Search
		[2, 5]	2	
		[1, 4]	4	
LR	None (default parameters)	Default	Default	Default
Attention-LightGBM	max_depth	[6, 8]	7	Grid Search
	learning_rate	[0.05, 0.1]	0.1	
	n_estimators	[100, 200]	100	

Data availability

Data will be made available on request.

References

Arshad, H., Emblemsvåg, J., Li, G., Ostnes, R., 2022. Determinants, methods, and solutions of evacuation models for passenger ships: a systematic literature review. Ocean Eng. 263, 112371. https://doi.org/10.1016/j.oceaneng.2022.112371.

Arshad, H., Emblemsvåg, J., Li, G.Y., 2024a. Multi-period human evacuation model for passenger ships under walking speed uncertainty. Ocean Eng. 309, 118590. https://doi.org/10.1016/j.oceaneng.2024.118590.

Arshad, H., Emblemsvåg, J., Zhao, X.L., 2024b. A data-driven, scenario-based human evacuation model for passenger ships addressing hybrid uncertainty. Int. J. Disaster Risk Reduct. 100. https://doi.org/10.1016/j.ijdrr.2023.104213.

Balakhontceva, M., Karbovskii, V., Rybokonenko, D., Boukhanovsky, A., 2015. Multi-agent simulation of Passenger Evacuation considering Ship Motions. Procedia Comput. Sci. 66, 140–149. https://doi.org/10.1016/j.procs.2015.11.017.

Bozdogan, A., Aykut, L.G., Demirel, N., 2023. An agent-based modeling framework for the design of a dynamic closed-loop supply chain network. Complex Intell. Syst. 9 (1), 247–265. https://doi.org/10.1007/s40747-022-00780-z.

Brown, R., Galea, E., Deere, S., Filippidis, L., 2012. Passenger response time data-sets for large passenger ferries and cruise ships derived from sea trials and recommendations to IMO to update MSC Circ 1238. The Royal Institution of Naval Architects: London, UK. https://doi.org/10.5750/ijme.v155iA1.894.

Cao, Y., Iulia, M., Majumdar, A., Feng, Y., Xin, X., Wang, X., Wang, H., Yang, Z., 2025a. Investigation of the risk influential factors of maritime accidents: a novel topology and robustness analytical framework. Reliab. Eng. Syst. Saf. 254, 110636. https://doi.org/10.1016/j.ress.2024.110636.

Cao, W., Wang, X., Feng, Y., Zhou, J., Yang, Z., 2025b. Improving maritime accident severity prediction accuracy: A holistic machine learning framework with data balancing and explainability techniques. Reliability Engineering & System Safety 266, 111648. https://doi.org/10.1016/j.ress.2025.111648.

Cao, W., Wang, X., Li, J., Zhang, Z., Cao, Y., Feng, Y., 2024. A novel integrated method for heterogeneity analysis of marine accidents involving different ship types. Ocean Eng 312, 119295. https://doi.org/10.1016/j.oceaneng.2024.119295.

Cao, Y., Xin, X., Wang, X., Wang, J., Yang, Z., 2025c. Multi-objective resilience-oriented optimisation for the global container shipping network against cascading failures. Transportation Research Part A: Policy and Practice 200, 104659. https://doi.org/10.1016/j.tra.2025.104659.

Chen, C.H., 2020. A Novel Multi-Criteria Decision-making Model for Building Material Supplier selection based on Entropy-AHP Weighted TOPSIS. Entropy 22 (2). https://doi.org/10.3390/e22020259.

Chen, M., Wu, K.G., Zhang, H.P., Han, D.F., Guo, M.Y., 2023. A ship evacuation model considering the interaction between pedestrians based on cellular automata. Ocean Eng. 281. https://doi.org/10.1016/j.oceaneng.2023.114644.

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput. Sci. 7, e623. https://doi.org/10.7717/peerj-cs.623.

Deere, S.J., Galea, E.R., Filippidis, L., Brown, R., 2012. Data collection methodologies used in the SAFEGUARD project to collect human factors data. In: RINA SAFEGUARD Passenger Evacuation Seminar, Vol. 30.

Ebrahimi, H., Sattari, F., Lefsrud, L., Macciotta, R., 2023. A machine learning and data analytics approach for predicting evacuation and identifying contributing factors during hazardous materials incidents on railways. Saf. Sci. 164. https://doi.org/10.1016/j.ssci.2023.106180.

Fang, S., Liu, Z., Zhang, S., Wang, X., Wang, Y., Ni, S., 2022a. Evacuation simulation of an Ro-Ro passenger ship considering the effects of inclination and crew's guidance. Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment 237 (1), 192–205. https://doi.org/10.1177/14750902221106566.

Fang, S.M., Liu, Z.J., Jiang, Y.T., Cao, Y.H., Wang, X.J., Wang, H.X., 2024a. Analysis of Human Competitive Behavior on an Inclined Ship. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 10 (1). https://doi.org/10.1061/ajrua6.Rueng-1198.

Fang, S.M., Liu, Z.J., Wang, X.J., Cao, Y.H., Yang, Z.L., 2024b. Dynamic analysis of emergency evacuation in a rolling passenger ship using a two-layer social force model. Expert Syst. Appl. 247. https://doi.org/10.1016/j.eswa.2024.123310.

Fang, S.M., Liu, Z.J., Wang, X.J., Wang, J., Yang, Z.L., 2022b. Simulation of evacuation in an inclined passenger vessel based on an improved social force model. Saf. Sci. 148. https://doi.org/10.1016/j.ssci.2022.105675.

Fang, S.M., Liu, Z.J., Yang, X.H., Wang, X.J., Wang, J., Yang, Z.L., 2023a. A quantitative study of the factors influencing human evacuation from ships. Ocean Eng. 285. https://doi.org/10.1016/j.oceaneng.2023.115156.

Fang, S.M., Liu, Z.J., Zhang, S.Q., Wang, X.J., Wang, Y.H., Ni, S.K., 2023b. Evacuation simulation of an Ro-Ro passenger ship considering the effects of inclination and crew's guidance. Proceedings of the Institution of Mechanical Engineers Part M-Journal of Engineering for the Maritime Environment 237 (1), 192–205. https://doi.org/10.1177/14750902221106566.

Farsi, M., Latsou, C., Erkoyuncu, J.A., Morris, G., 2020. RFID Application in a Multi-Agent Cyber Physical Manufacturing System. Journal of Manufacturing and Materials Processing 4 (4). https://doi.org/10.3390/jmmp4040103.

Feng, Y., Wang, X., Chen, Q., Yang, Z., Wang, J., Li, H., Xia, G., Liu, Z., 2024a. Prediction of the severity of marine accidents using improved machine learning. Transportation Research Part E: Logistics and Transportation Review 188, 103647. https://doi.org/10.1016/j.tre.2024.103647.

Feng, Y., Wang, X., Luan, J., Wang, H., Li, H., Li, H., Liu, Z., Yang, Z., 2024b. A novel method for ship carbon emissions prediction under the influence of emergency events. Transp. Res. Part C Emerging Technol. 165, 104749. https://doi.org/10.1016/j.trc.2024.104749.

Galea, E.R., Deere, S., Brown, R., Filippidis, L., 2013a. An Experimental Validation of an Evacuation Model using Data Sets Generated from two Large Passenger Ships. J. Ship Res. 57 (3). https://doi.org/10.5957/JOSR.57.3.120037.

Galea, E.R., Deere, S., Brown, R., Filippidis, L., 2013b. An Experimental Validation of an Evacuation Model using Data Sets Generated from two Large Passenger Ships. J. Ship Res. 57 (3), 155–170. https://doi.org/10.5957/josr.57.3.120037.

Gayathri, R., Rani, S.U., Cepová, L., Rajesh, M., Kalita, K., 2022. A Comparative Analysis of Machine Learning Models in Prediction of Mortar Compressive Strength. Processes 10 (7). https://doi.org/10.3390/pr10071387.

Guo, K., Zhang, L., 2022a. Adaptive multi-objective optimization for emergency evacuation at metro stations. Reliab. Eng. Syst. Saf. 219, 108210. https://doi.org/10.1016/j.ress.2021.108210.

Guo, K., Zhang, L., 2022b. Simulation-based passenger evacuation optimization in metro stations considering multi-objectives. Autom. Constr. 133, 104010. https://doi.org/10.1016/j.autcon.2021.104010.

Guo, K., Zhang, L., Wu, M., 2023. Simulation-based multi-objective optimization towards proactive evacuation planning at metro stations. Eng. Appl. Artif. Intel. 120, 105858. https://doi.org/10.1016/j.engappai.2023.105858.

Ha, S., Ku, N.-K., Roh, M.-I., Lee, K.-Y., 2012. Cell-based evacuation simulation considering human behavior in a passenger ship. Ocean Eng. 53, 138–152. https://doi.org/10.1016/j.oceaneng.2012.05.019.

Haghani, M., Sarvi, M., Shahhoseini, Z., Boltes, M., 2019. Dynamics of social groups' decision-making in evacuations. Transportation Research Part C-Emerging Technologies 104, 135–157. https://doi.org/10.1016/j.trc.2019.04.029.

Hatipoglu, I., Tosun, O., 2024. Predictive Modeling of Flight Delays at an Airport using Machine Learning Methods. Appl. Sci.-Basel 14 (13). https://doi.org/10.3390/app14135472.

Ide, Y., Ozaki, S., Izutsu, S., Kotoura, T., Yamashiro, M., Kodama, M., 2024. Hybrid real-time wave forecasting model combining Gaussian process regression and neural networks. Ocean Eng. 312. https://doi.org/10.1016/j.oceaneng.2024.119028.

Jiao, J., An, R., Du, D., Zhu, M., 2025. Non-linear and heterogeneous relationship between proximity to high-speed rail stations and land value in China: Analysis using XGBoost-SHAP modelling. Transp. Res. A Policy Pract. 196, 104486. https://doi.org/10.1016/j.tra.2025.104486.

Lee, J., Kim, H., Kwon, S., 2022. Evacuation analysis of a passenger ship with an inclined passage considering the coupled effect of trim and heel. Int. J. Nav. Archit. Ocean Eng. 14, 100450. https://doi.org/10.1016/j.ijnaoe.2022.100450.

Li, H., Díaz, H., Guedes Soares, C., 2021. A failure analysis of floating offshore wind turbines using AHP-FMEA methodology. Ocean Eng. 234, 109261. https://doi.org/10.1016/j.oceaneng.2021.109261.

Li, P.Y., Bao, X.H., Hong, C.Y., Wang, D.C., Xie, X.F., Fan, J.H., Li, H., Liao, L.H., 2024a. Evacuation simulation and Fire-Risk Assessment on Underground Space of Guangzhou International Financial City. Fire-Switzerland 7 (9). https://doi.org/10.3390/fire7090307.

Li, S., Tong, L., Zhai, C.H., 2022. Extraction and modelling application of evacuation movement characteristic parameters in real earthquake evacuation video based on deep learning. Int. J. Disaster Risk Reduct. 80. https://doi.org/10.1016/j.ijdrr.2022.103213.

Li, S.Y., Ma, H.Y., Zhang, Y.D., Wang, S., Guo, R., He, W., Xu, J.C., Xie, Z.Y., 2023. Emergency evacuation risk assessment method for educational buildings based on improved extreme learning machine. Reliab. Eng. Syst. Saf. 238. https://doi.org/10.1016/j.ress.2023.109454.

Li, T., Wang, X., Zhang, Z., Feng, Y., 2025. A novel feature engineering method for severity prediction of marine accidents. J Mar Eng Technol 24, 1–16. https://doi.org/10.1080/20464177.2025.2469336.

Li, Y.P., Xiao, Q., Gu, J.Y., Cai, W., Hu, M., 2024b. Modeling and solving passenger ship evacuation arrangement problem. Reliab. Eng. Syst. Saf. 246. https://doi.org/10.1016/j.ress.2024.110075.

Lin, L., Wang, F., Xie, X.L., Zhong, S.S., 2017. Random forests-based extreme learning machine ensemble for multi-regime time series prediction. Expert Syst. Appl. 83, 164–176. https://doi.org/10.1016/j.eswa.2017.04.013.

Liu, C., Yang, S., Cui, Y., Yang, Y., 2020. An improved risk assessment method based on a comprehensive weighting algorithm in railway signaling safety analysis. Saf. Sci. 128, 104768. https://doi.org/10.1016/j.ssci.2020.104768.

Liu, K.Z., Ma, Y.T., Chen, M.Z., Wang, K.H., Zheng, K., 2022a. A survey of crowd evacuation on passenger ships: recent advances and future challenges. Ocean Eng. 263. https://doi.org/10.1016/j.oceaneng.2022.112403.

Liu, L.F., Zhang, H.J., Bo, Y.S., Chen, B.Y., Du, J.F., Su, Y.X., 2023a. Dynamic Evacuation Routes Navigation for Passenger Ship Fire based on Intelligence Algorithms. Adv. Theor. Simul. 6 (11). https://doi.org/10.1002/adts.202300351.

Liu, L.F., Zhang, H.J., Shi, J., Geng, J.N., 2023b. Evacuation simulation of a Passenger Ship Fire based on a Modified Cellular Automaton. Adv. Theor. Simul. 6 (7). https://doi.org/10.1002/adts.202300141.

Liu, L.F., Zhang, H.J., Xie, J.P., Zhao, Q., 2021. Dynamic Evacuation Planning on Cruise Ships based on an improved Ant Colony System (IACS). Journal of Marine Science and Engineering 9 (2). https://doi.org/10.3390/jmse9020220.

Liu, L.F., Zhang, H.J., Zhan, Y., Su, Y.X., Zhang, C.F., 2022b. Intelligent optimization method for the evacuation routes of dense crowds on cruise ships. Simul. Model. Pract. Theory 117. https://doi.org/10.1016/j.simpat.2022.102496.

Liu, Y., Guo, Z., Jia, X.Y., Li, C.F., Kang, J.C., 2022c. Space temperature field and structure response of multi-chamber cabin subjected to fire. Ocean Eng. 266. https://doi.org/10.1016/j.oceaneng.2022.112904.

Long, F.C., Majumdar, A., Carter, H., 2023. Understanding levels of compliance with emergency responder instructions for members of the Public involved in emergencies: evidence from the Grenfell Tower fire. Int. J. Disaster Risk Reduct. 84. https://doi.org/10.1016/j.ijdrr.2022.103374.

Lovreglio, R., Ronchi, E., Borri, D., 2014. The validation of evacuation simulation models through the analysis of behavioural uncertainty. Reliab. Eng. Syst. Saf. 131, 166–174. https://doi.org/10.1016/j.ress.2014.07.007.

Lozowicka, D., 2021. The design of the arrangement of evacuation routes on a passenger ship using the method of genetic algorithms. PLoS One 16 (8). https://doi.org/10.1371/journal.pone.0255993.

Lu, L.L., Chan, C.Y., Wang, J., Wang, W., 2017. A study of pedestrian group behaviors in crowd evacuation based on an extended floor field cellular automaton model. Transportation Research Part C-Emerging Technologies 81, 317–329. https://doi.org/10.1016/j.trc.2016.08.018.

Lundberg, S.M., Lee, S.-I., 2017. Consistent feature attribution for tree ensembles. arXiv preprint arXiv:1706.06060. https://arxiv.org/abs/1802.03888.

Ma, X.L., Ding, C., Luan, S., Wang, Y., Wang, Y.P., 2017. Prioritizing influential Factors for Freeway Incident Clearance Time Prediction using the Gradient Boosting Decision Trees Method. IEEE Trans. Intell. Transp. Syst. 18 (9), 2303–2310. https://doi.org/10.1109/tits.2016.2635719.

Ma, Y.T., Liu, K.Z., Chen, M.Z., Li, Y.H., Sun, R., Ranjan, R., 2024. Rapid Crowd Evacuation for Passenger Ships using LPWAN. IEEE Trans. Intell. Transp. Syst. 25 (2), 1720–1735. https://doi.org/10.1109/tits.2023.3314081.

Mao, Y., Wang, X., He, W., Pan, G.F., 2024. An investigation into the influence of gender on crowd exit selection in indoor evacuation. Int. J. Disaster Risk Reduct. 108. https://doi.org/10.1016/j.ijdrr.2024.104563.

MSC, I. 1/Circ. 1533 Revised Guidelines on Evacuation Analysis for New and Existing Passenger Ships [R]. 2016. https://imorules.com/MSCCIRC_1533.html.

Muravev, D., Hu, H., Rakhmangulov, A., Mishkurov, P., 2021. Multi-agent optimization of the intermodal terminal main parameters by using AnyLogic simulation platform: Case study on the Ningbo-Zhoushan Port. Int. J. Inf. Manag. 57. https://doi.org/10.1016/j.ijinfomgt.2020.102133.

Ni, B.C., Li, Z., Zhang, P., Li, X., 2017. An Evacuation Model for Passenger Ships that includes the Influence of Obstacles in Cabins. Math. Probl. Eng. 2017. https://doi.org/10.1155/2017/5907876.

Ren, J.X., Mao, Z.L., Zhang, D., Gong, M.L., Zuo, S.T., 2022. Experimental study of crowd evacuation dynamics considering small group behavioral patterns. Int. J. Disaster Risk Reduct. 80. https://doi.org/10.1016/j.ijdrr.2022.103228.

Sahebi, A., Jahangiri, K., Alibabaei, A., Khorasani-Zavareh, D., 2023. Using Artificial Intelligence for predicting the Duration of Emergency Evacuation during Hospital Fire. Disaster Med. Public Health Prep. 17, e229. https://doi.org/10.1017/dmp.2022.187.

Sarshar, P., Granmo, O.C., Radianti, J., Gonzalez, J.J., 2013. A Bayesian network model for evacuation time analysis during a ship fire. In: 2013 IEEE Symposium on Computational Intelligence in Dynamic and Uncertain Environments (CIDUE). IEEE, pp. 100–107. https://doi.org/10.1109/CIDUE.2013.6595778.

Shipman, A., Majumdar, A., Boyce, N., Lovreglio, R., 2024. Movement behaviour of pedestrians in knife-based terrorist attacks: an experimental approach. Transp. Res. Part C Emerging Technol. 166, 104790. https://doi.org/10.1016/j.trc.2024.104790.

Sreejith, R., Sinimole, K.R., 2022. Modelling evacuation preparation time prior to floods: a machine learning approach. Sustain. Cities Soc. 87. https://doi.org/10.1016/j.scs.2022.104257.

Sun, J., Guo, Y., Li, C., Lo, S., Lu, S., 2018a. An experimental study on individual walking speed during ship evacuation with the combined effect of heeling and trim. Ocean Eng. 166, 396–403. https://doi.org/10.1016/j.oceaneng.2017.10.008.

Sun, J., Lu, S., Lo, S., Ma, J., Xie, Q., 2018b. Moving characteristics of single file passengers considering the effect of ship trim and heeling. Physica A 490, 476–487. https://doi.org/10.1016/j.physa.2017.08.031.

Sun, Y.R., Huang, S.K., Zhao, X.L., 2024. Predicting Hurricane Evacuation Decisions with Interpretable Machine Learning Methods. International Journal of Disaster Risk Science 15 (1), 134–148. https://doi.org/10.1007/s13753-024-00541-1.

Takimoto, K., Nagatani, T., 2003. Spatio-temporal distribution of escape time in evacuation process. Physica A 320, 611–621. https://doi.org/10.1016/S0378-4371(02)01540-6.

Uhlík, O., Okřinová, P., Tokarevskikh, A., Apeltauer, T., Apeltauer, J., 2024. Real-time RSET prediction across three types of geometries and simulation training dataset: a comparative study of machine learning models. Dev. Built Environ. 18, 100461. https://doi.org/10.1016/j.dibe.2024.100461.

Valcalda, A., de Koningh, D., Kana, A., 2023. A method to assess the impact of safe return to port regulatory framework on passenger ships concept design. Journal of Marine Engineering & Technology 22 (3), 111–122. https://doi.org/10.1080/20464177.2022.2031557.

Vanem, E., Skjong, R., 2006. Designing for safety in passenger ships utilizing advanced evacuation analyses—A risk based approach. Saf. Sci. 44 (2), 111–135. https://doi.org/10.1016/j.ssci.2005.06.007.

Wang, X., Liu, Z., Zhao, Z., Wang, J., Loughney, S., Wang, H., 2020. Passengers' likely behaviour based on demographic difference during an emergency evacuation in a Ro-Ro passenger ship. Saf. Sci. 129, 104803. https://doi.org/10.1016/j.ssci.2020.104803.

Wang, X.J., Liu, Z.J., Loughney, S., Yang, Z.L., Wang, Y.F., Wang, J., 2021a. An experimental analysis of evacuees' walking speeds under different rolling conditions of a ship. Ocean Eng. 233. https://doi.org/10.1016/j.oceaneng.2021.108997.

Wang, X.J., Liu, Z.J., Loughney, S., Yang, Z.L., Wang, Y.F., Wang, J., 2022. Numerical analysis and staircase layout optimisation for a Ro-Ro passenger ship during emergency evacuation. Reliab. Eng. Syst. Saf. 217. https://doi.org/10.1016/j.ress.2021.108056.

Wang, X.J., Liu, Z.J., Wang, J., Loughney, S., Yang, Z.L., Gao, X.W., 2021b. Experimental study on individual walking speed during emergency evacuation with the influence of ship motion. Physica a-Statistical Mechanics and Its Applications 562. https://doi.org/10.1016/j.physa.2020.125369.

Wang, X.J., Liu, Z.J., Wang, J., Loughney, S., Zhao, Z.W., Cao, L., 2021c. Passengers' safety awareness and perception of wayfinding tools in a Ro-Ro passenger ship during an emergency evacuation. Saf. Sci. 137. https://doi.org/10.1016/j.ssci.2021.105189.

Wang, X.J., Xia, G.Q., Zhao, J., Wang, J., Yang, Z.L., Loughney, S., Fang, S.M., Zhang, S.K., Xing, Y.H., Liu, Z.J., 2023a. A novel method for the risk assessment of human evacuation from cruise ships in maritime transportation. Reliab. Eng. Syst. Saf. 230. https://doi.org/10.1016/j.ress.2022.108887.

Wang, X., Cao, W., Li, T., Feng, Y., Uğurlu, Ö., Wang, J., 2025. An integrated multidimensional model for heterogeneity analysis of maritime accidents during different watchkeeping periods. Ocean & Coastal Management 264, 107625. https://doi.org/10.1016/j.ocecoaman.2025.107625.

Wang, Y., Chen, J., Hu, Y., Weng, X., 2023b. Modelling and interpreting evacuation time and exit choice for large-scale ancient architectural complex using machine learning. Journal of Building Engineering 80, 108133. https://doi.org/10.1016/j.jobe.2023.108133.

Wu, B., Zong, L., Yip, T.L., Wang, Y., 2018. A probabilistic model for fatality estimation of ship fire accidents. Ocean Eng. 170, 266–275. https://doi.org/10.1016/j.oceaneng.2018.10.056.

Xie, Q.M., Li, S.S., Ma, C., Wang, J.H., Liu, J.H., Wang, Y., 2020a. Uncertainty analysis of passenger evacuation time for ships' safe return to port in fires using polynomial chaos expansion with Gauss quadrature. Appl. Ocean Res. 101. https://doi.org/10.1016/j.apor.2020.102190.

Xie, Q.M., Wang, P.C., Li, S.S., Wang, J.H., Lo, S.M., Wang, W.L., 2020b. An uncertainty analysis method for passenger travel time under ship fires: a coupling technique of nested sampling and polynomial chaos expansion method. Ocean Eng. 195. https://doi.org/10.1016/j.oceaneng.2019.106604.

Yang, X., Yang, Y., Qu, D., Chen, X., Li, Y., 2023. Multi-Objective Optimization of Evacuation Route for Heterogeneous Passengers in the Metro Station considering Node Efficiency. IEEE Trans. Intell. Transp. Syst. 24 (11), 12448–12461. https://doi.org/10.1109/TITS.2023.3292912.

Zhang, Y.X., Kinateder, M., Huang, X.Y., Warren, W.H., 2025. Modeling competing guidance on evacuation choices under time pressure using virtual reality and machine learning. Expert Syst. Appl. 262. https://doi.org/10.1016/j.eswa.2024.125582.

Zhao, X., Lovreglio, R., Nilsson, D., 2020a. Modelling and interpreting pre-evacuation decision-making using machine learning. Autom. Constr. 113, 103140. https://doi.org/10.1016/j.autcon.2020.103140.

Zhao, X.L., Lovreglio, R., Nilsson, D., 2020b. Modelling and interpreting pre-evacuation decision-making using machine learning. Autom. Constr. 113. https://doi.org/10.1016/j.autcon.2020.103140.