Safety Science 174 (2024) 106485

Contents lists available at ScienceDirect

Safety Science

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

# Logistic regression vs machine learning to predict evacuation decisions in fire alarm situations

Adriana Balboa *, Arturo Cuesta, Javier González-Villa, Gemma Ortiz, Daniel Alvear

*Universidad de Cantabria, Los Castros s/n, Santander 39005, Spain*

A R T I C L E I N F O

*Keywords:*
Evacuation
Decision-making
Logistic regression
Machine learning
Fire alarm

A B S T R A C T

In this study we assessed logistic regression and machine learning models to explore their performance in predicting evacuation decisions and to provide readers with insights into the accuracy of these methods. We tested seven machine learning algorithms, including classification and regression tree, Naïve Bayes, K-nearest neighbours, support vector machine, random forest, extreme gradient boosting, and artificial neural network. We used data collected from 1,807 participants through web-based experiments to train and calibrate these models. The performance of each model was evaluated by area under the curve, accuracy, recall, specificity, precision, and F1-score. The results indicate that logistic regression had the highest area under the curve value (0.831), whereas extreme gradient boosting outperformed other machine learning models in terms of accuracy (0.780), specificity (0.810) and precision (0.820). K-nearest neighbours model had the greater recall (0.859) and artificial neural network the highest F1-score (0.785). The models identified that being with a close person was the most influential factor in the response to a fire alarm.

## 1. Introduction

Understanding decision making of people during an emergency is a main concern in safety science (Kuligowski, 2009; Kuligowski, 2011; Proulx, 2001; Wood, 1972). The “stay” or “go” decision of individuals is a crucial aspect of the pre-evacuation phase, especially when traditional warning signals such as alarm bells, horns, or sirens are used. These warning systems usually fail to create an immediate response (Proulx, 2001; Wood, 1972) as they only inform occupants about the potential danger without providing any further information of the emergency. Consequently, people may either ignore the warning and carry on with their activities or seek additional information (Proulx, 2001; Wood, 1972), resulting in a delayed response that can increase the risk to life safety (Fritz and Marks, 1954; Kuligowski, 2009).

To address this, researchers have focused on improving communication systems for warning individuals (D’Orazio et al., 2016; Ferraro and Settino, 2019; Kuligowski, 2011; Lin et al., 2023), evaluating the decision-making during pre-movement phase to support fire safety assessment (Liu and Lo, 2011; Lovreglio et al., 2015; Viswanathan and Lees, 2014) and analysing wayfinding behaviour (Lin et al., 2019; Vilar et al., 2018).

Enhancing our knowledge of the factors that influence human behaviour during emergencies is critical (Santos and Aguirre, 2004), and therefore, the analysis of these factors should be considered in evacuation studies and plans to avoid biased analysis (Song and Lovreglio, 2021). Previous research has already begun to analyse the impact of factors on human behaviour during building fires. Studies have examined the physical context, including the environment (Kinateder et al., 2018; Richardson et al., 2018; Cubukcu, 2003), and cues (Yamada and Akizuki, 2016; Fu et al., 2018; Saunders, 2001). Social influences such as the presence or absence of others (Fu et al., 2018; Song and Lovreglio, 2021; Lovreglio et al., 2015) have also been explored. Demographics such as age, education or gender have also been studied (Jeon et al., 2014; Song and Lovreglio, 2021; Saunders, 2001).

Furthermore, a significant body of literature has investigated decision-making in the context of wildfires. Kuligowski (2021) reviewed research and data collection on evacuation decision-making and behaviour during wildfires. Additionally, Kuligowski et al., 2022 conducted an online survey to measure pre-event and event factors influencing risk perception and evacuation decisions using linear and logistic regression models. McCaffrey et al., 2018 used the PADM model (Lindell and Perry, 2012) and surveyed wildfire survivors to identify factors

* Corresponding author.
*E-mail addresses:* balboaa@unican.es (A. Balboa), cuestaar@unican.es (A. Cuesta), javier.gonzalezvilla@unican.es (J. González-Villa), ortizg@unican.es (G. Ortiz), alveard@unican.es (D. Alvear).

https://doi.org/10.1016/j.ssci.2024.106485
Received 25 April 2023; Received in revised form 19 February 2024; Accepted 28 February 2024
Available online 10 March 2024

affecting evacuation decisions. They found that personal thoughts of actions efficacy, cues and risk attitudes were critical factors in determining responses. McLennan et al., 2012 analysed the reasons behind staying or evacuating in the face of wildfire threats through field interviews with survivors. They obtained that individuals who chose to stay did so to defend their properties, while those who recognized the intensity of the threat opted to evacuate. In a subsequent study, McLennan et al., 2019 examined evacuation policies, the behaviour of residents and factors influencing their actions. Toledo et al., 2018 provided data on the activities conducted by residents in a wildfire area, focusing on evacuation travel patterns and their association with household type, initial location and destination. Similarly, Wong et al., 2022 explored factors such as evacuation orders and transportation responses. However, there is still a lack of clarity regarding how people respond to building fires and the influence of multiple factors need to be analysed together.

Modelling has emerged as a promising approach for developing predictive pre-movement methods for use by evacuation models. These methods encompass traditional modelling based on assumptions (Kuligowski and Mileti, 2009; Reneke, 2013; Sherman et al., 2011), as well as new data-driven approaches that employ algorithms to generate insights, predictions, and decision-making models. By identifying patterns, correlations, and causal relationships, data-driven approaches can inform decision-making and provide more accurate predictions than traditional methods. The advantages of data-driven in evacuation modelling include the ability to handle complex nonlinear relationships.

Currently, machine learning algorithms and statistical modelling techniques are being used to analyse data and identify patterns and relationships. Lo and Zhang, 2009 proposed an artificial neural network for predicting pre-evacuation actions, such as fight the fire, inform others, collect belongings, escape or seek for information, in domestic building fires. Thakur et al., 2022 used logistic regression to investigate factors influencing immediate or delayed evacuation behaviour under a volcanic threat. The results indicated that demographics, destination, transport or household characteristics influenced decision-making. Xu et al., 2023 processed data from evacuation in wildfires using seven machine learning models compared to logistic regression. They used survey data from the Kincade Fire and founded that previous safety perception was a key factor in the binary decision of evacuate or stay. Zhu et al., 2023 trained machine learning and discrete choice models to understand directional choice evacuation decisions, emphasizing the importance of building size and familiarity for evacuation models. Lechner et al., 2019 used nested logistic regression to explore reasons behind households' decisions to stay in their houses instead of evacuate during a volcano eruption. They considered factors such as risk perception, preparedness, past experience, future intentions and household characteristics. Zhao et al., 2020 used random forest and machine learning interpretation tools to analyse unannounced evacuation drills in a cinema theatre, discovering the influence of surrounding behaviour, decision-maker location and group size on pre-evacuation decisions. Sun et al., 2023 proposed machine learning logistic regression approach to predict the decisions of households during hurricanes, revealing nonlinear patterns in demographic variables and the significance of household size after univariate threshold detection.

We plan to apply the methodology of previous machine learning (ML) and logistic regression (LR) studies to explore the factors that influence behaviour during fire alarm scenarios. The main contributions of this study are: 1) assessing the performance of LR and ML in predicting evacuation decisions during fire alarms in buildings, 2) identifying the influence of determinants of such decisions, 3) providing new data and insights into evacuation decision-making in building fires derived from a new approach using online experiments.

This study describes the development of a web-based simulation of conventional fire alarm scenarios that participants can remotely engage in without the need for laboratory settings. The experiment comprised six trials, each featuring a unique combination of physical and social factors, along with a post-questionnaire to gather individual characteristics. The evacuation decisions of 1,807 participants were recorded and analysed. To model the impact of various factors on decision-making, LR and ML techniques were employed. The purpose of this study was to: 1) explore the performance of LR and ML models in predicting evacuation decisions, 2) provide insights on their accuracy, and 3) identify influential factors in the response to a fire alarm in buildings.

## 2. Method

Web-based experiments were carried out to gather data on self-protective decision-making of individuals in building fire alarm scenarios. This research method was used due to its advantages: 1) the ability to easily access a diverse population of participants, including those who are typically inaccessible; 2) bringing the experiment to the participant, rather than requiring them to come to a specific location; 3) the potential for high statistical power by allowing access to large samples; and 4) the ability to emulate complex threats or stimuli in a virtual environment without risking harm to the participants. The experiments involve six trials and a questionnaire. The performance of machine learning (ML) and logistic regression (LR) predictive models was then assessed.

### 2.1. Data collection and procedure

*Participants.* Individuals were recruited via a survey company with the requirement of being representative of the Spanish population in terms of gender and age. Participants were given a monetary incentive for their participation managed by the survey company. A total of 1,807 naïve participants, 51 % of females and 49 % of males aged between 18 and 76 years (mean = 47.88, standard deviation = 16.13), were included in the study.

*Design.* The scope of the data collection was to investigate the influence of external and internal factors on the decisions of individuals during fire alarms in buildings. The study considered physical context and social influence as external factors. The physical context included open area (an indoor diaphanous environment such as a building hall) or room (an enclosed environment by walls such as a room of a building) options. The social influence could be alone, with people around evacuating or with people around remaining in the area. The personal information (such as sociodemographic characteristics) of participants was considered as internal factors. The dependent variable was the binary self-protective decision made by individuals between evacuate or stay. One of the main challenges in designing the experiment was to create a realistic environment. To achieve this, the physical context presented in the videos was based on a real building at the University of Cantabria. The 3D environment was created using the Pyrosim software (Pyrosim User Manual, 2022). Importing this file in the Pathfinder software (Pathfinder – Technical Reference, 2022) we added virtual people to the environment. The participants were presented with a subjective perspective and the option to choose between a generic male or female avatar, using the capabilities of the software. In addition, different 3D models of people were placed throughout the environment, walking or standing, to accomplish more realism. The social influence stimulus was implemented by having five virtual persons (virtual confederates) who either evacuate or stay when the fire alarm goes off. The sound used in the online study was the actual fire alarm of the building.

The experiment was pilot-tested allowing us the possibility to know whether the designed trials and the questionnaire fulfilled the purpose of the study. The pilot involved 41 individuals (24 females and 17 males, mean age = 41.9 years) who performed two out of the six trials. The social influence in the pilot involved people around the participants evacuating and staying in the same trial at the same time. This condition limited the ability to determine which actions of these virtual confederates influenced the participant decision. Hence, we modified the design to divide the virtual people actions into two trials (one where confederates evacuate and one where confederates stay in place). We also included an additional question to extend the analysis of familiar social influence in evacuation decision making. After completing the trial, and prior to the questionnaire, participants were asked to think about a close person (such as a relative or friend). Participants then decided again, this time imagining that they were in the fire alarm situation with that person. The final questionnaire contained 21 items divided into three blocks 1) sociodemographic factors, 2) experiment feeling factors, and 3) personal experience factors. Table 1 shows each question along with the corresponding response options.

*Procedure.* The experiment was conducted using the online PsyToolkit platform (Stoet, 2010; Stoet, 2017). This open-source platform was selected due to its flexibility, user-friendly interface, cost-effective, a wide range of features (e.g. randomization, counterbalancing, response time measurement, and data storage) for conducting our online study. Participants watched a pre-recorded video following their own avatar through the virtual building. They had no control over their movement, and there was no interactivity with the environment. After the fire alarm of the trial, participants made their decision by selecting one of the two options (evacuate or stay) presented on a screen.

A survey company distributed the experiment in May 2022. The experiment was designed to be accessible on both laptops and mobile phones, with a preference for laptop use due to the larger display. The participants were recommended to use headphones for a better experience/perception. 773 participants (42.78 %) reported using their own laptop, while 1,034 participants (57.22 %) reported using their own mobile phone to conduct the experiment. The order in which participants went through the screens of the experiment is presented below. Screen 2 (sound test) and screen 8 (write down) were used to check that questionnaire was not filled out by bots.

Screen 1) General information screen and consent,

Screen 2) Sound test (participants selected a sound to check the audio and their concentration),

Screen 3) Screen with the experiment structure (trial + questionnaire) and mandatory confirmation of having read it,

Screen 4) Situational context explanation (participants were situated in a health centre to pick up medical results),

Screen 5) Trial (3D video and alarm sounding),

Screen 6) Decision-making between evacuate or stay within a 10-second time limit (decisions after that time were also recorded),

Screen 7) Decision-making between evacuate or stay with a close person.

Screen 8) Participants write down a close person they have thought of (e.g. father/mother, grandfather/grandmother, brother/sister, partner).

Screen 9) Questionnaire: 21 items, 7 screens

The online experiment method can be checked through the following link: https://youtu.be/oZvLkJH7_a4.

*Experimental conditions.* Six trials were designed combining physical and social conditions (see Table 2). The physical context could be defined as the location the participant is when the fire alarm sounds. In trials 1, 3 and 5 participants were initially located in *Open area* whereas in trials 2, 4 and 6 they were located in *Room*. The social influence refers to the virtual confederates the participant could see (staying or evacuating) after the fire alarm sounded. Note that in trials 1 and 2 the participant was alone (*Individual*), in trials 3 and 4 the participant could see confederates evacuating (*Group Go*) after the alarm and in trials 5 and 6 the participant could see the confederates remaining in the area (*Group Stay*). The participants were randomly assigned to one of the six trials. Each participant only performed one trial to avoid learning behaviours. At least 300 individuals took part in each trial (in total 1,807 participants).

*Ecological validity.* The nature of web-based experiments presents challenges in terms of accurately presenting stimuli and virtual environments. Ecological validity, which refers to the degree findings from experiments hold true in real-life situations, is crucial for virtual studies (Nilsson and Kinateder, 2015). It should be noted that participants took part in three consecutive trials: fire alarm, explosion and shootings. This paper focuses on the first one. Note that the responses of the post-questionnaire of Experiment Feeling Factors (see Table 1) are an overall answer for all scenarios. In this study, the navigation solution and the 3D features of the online format were assessed with the realism question.

**Table 1**
Questions included in the questionnaire.

| Item | Item Question | Scale or Response Options |
|---|---|---|
| Sociodemographic Factors | | |
| *Age* | Indicate your age: | [18 – 90] |
| *Weight* | Indicate your weight in Kg: | [40 – 140] |
| *Height* | Indicate your height in cm: | [140 – 200] |
| *Gender* | What gender do you identify with? | Female; Male; Non-binary; Other |
| *Education level* | Select your completed education level: | Primary Education; Secondary Education; Vocational Training; High School; University |
| *Occupation* | Select your work situation: | Employed; Unemployed; Self-employed; Student; Retired |
| *Income* | Select the monthly income range of your family unit: | <1000€; 1000-1999€; 2000-2999€; >3000€ |
| *Residence* | Where do you live? | In a village (less than 5000 inhabitants); In a small city (from 5000 to 50,000 inhabitants); In a large city (more than 50,000 inhabitants) |
| *Politic* | Which political ideology is more in line with your ideas? | Left; Centre; Right; I am apolitical |
| *Religion* | Are you a religious person? | Yes; No |
| *People care* | Are you responsible for the care of any individuals? | Yes, minors; Yes, elderly; Yes, dependents; No, I do not have |
| *Sport* | How often do you do exercise or play a sport? | Never; Between 1 and 3 times per week; More than 3 times per week |
| *Fitness* | Compared to people your age, what is your fitness level? | Below average; On average; Above average |
| *Movement* | Can you walk/move fast? | No, I can't; I can with some limitations; Yes, I can |
| Experiment Feeling Factors | | |
| *Realism* | Rate the realism you perceived in the videos (0 very low, 10 very high): | [0 – 10] |
| *Stress* | Rate the stress you felt during the experiment (0 very low, 10 very high): | [0 – 10] |
| *Survival probability* | How likely are you to survive with the decisions you have made? (0 very low, 10 very high): | [0 – 10] |
| Personal Experience Factors | | |
| *Previous drill* | Have you participated in an evacuation drill before? | Yes; No |
| *Previous fire alarm* | Have you been involved in the situation in the experiment? | Yes; No |
| *Previous training* | Have you previously received safety and/or self-protection training? | Yes; No |
| *Previous First Responder (FR)* | Do you work or have you worked as a firefighter, police officer, sanitary, civil protective, etc.? | Yes; No |

**Table 2**
Characteristics of each trial.

| Trial | N Participants | Physical condition | Social condition |
|---|---|---|---|
| 1 | 301 | *Open area* | *Individual* |
| 2 | 301 | *Room* | *Individual* |
| 3 | 300 | *Open area* | *Group Go* |
| 4 | 304 | *Room* | *Group Go* |
| 5 | 300 | *Open area* | *Group Stay* |
| 6 | 301 | *Room* | *Group Stay* |

Therefore, participants were asked to rank their perceived realism of the videos. Here, we aimed to explore the impressions of participants regarding the graphics showing in the video, in terms of the 3D environment, 3D social influence and the auditory and visual cues. The other two questions included in the same block allowed participants to estimate their stress levels during the experiment and their perceived survival probability based on the decision made. We used these answers to assess the feelings of participants after the experiment. A 10-point Likert scale was used for these questions.

*Ethics.* The experimental protocol was approved by the Ethical Committee of the University of Cantabria (Ref. 14–06-2022.000062.CE. PI). The survey company had their own requirements for the fulfilment of the experiment. Participants provided informed consent by clicking the agreement section before taking part in the web-based experiment. The experiment was anonymous, and the privacy policy regarding the individual's posted information was clearly stated (including their personal information, data protection, and withdrawal rights).

### 2.2. Statistical analysis

All statistical analysis was conducted using RStudio software (version 4.2.0). The dependent variable was the decision made (evacuate or stay) by participants after each trial. The independent variables included the physical and social conditions, as well as responses from the post-questionnaire. All data were registered in an Excel spreadsheet. The datasets are available by contacting the corresponding author.

*Data preparation.* Upon reviewing the variables, we processed the data by defining new variables and addressing issues such as missing values and outliers. BMI was calculated using the height and weight of participants. The non-binary gender option was selected by only 0.17 % of individuals (3 out of 1,807), which was not a representative sample. Therefore, gender analysis focused on male–female. The overall fitness level of participants was determined combining the responses to sport, fitness and movement items to create a new variable called "fitness level". The 3-point individual scale of each item was ordinal categorized with value of 1, 2 or 3. Then, the new variable was generated by summing the scores of each individual scale. Composite scores were then categorized in seven categories as follows: very low (3), low (4), low-medium (5), medium (6), medium–high (7), high (8), and very high (9). The question related to whether participants have someone to care for, was recorded into binary (yes/no) to make the models easier to fit and interpret.

*Explanatory variables selection.* The data was categorized, except for the age variable that remains continuous. A total of 18 candidate explanatory variables were considered in this study (see Table 4). We confirmed no correlation between independent variables through an analysis excluding weight, height, sport, fitness and movement variables, used to obtain BMI and fitness level derived variables included in the dataset. The analysis was conducted using the *rcorr* function from the *Hmisc* package in RStudio. We ensured that the correlation threshold was below 0.8, as suggested in reference (Midi et al., 2010). A preliminary simple association was carried out to investigate which variables were correlated with the evacuation decision using one-way ANOVA test (F-value).

*Fit the model.* Following a supervised learning approach, we used logistic regression (LR) and seven machine learning (ML) algorithms to predict the probability of participants deciding to evacuate or to stay during the building fire alarm scenario and then we explored their performance. The overview of each model is below:

- LR: Logistic regression (Berkson, 1944) is a technique to make qualitative predictions. In this study, binary logistic regression was used to estimate the probability of a binary dependent outcome Y (either evacuate $Y = 1$ or stay $Y = 0$) based on a set of independent variables $X_s$. The logit link function was used, which combines the probability $p$ of the event of interest $P(Y = 1)$ with a linear combination of the independent variables. LR uses the odds ratio (Eq. (1) to represent the probability of the event occurring ($p$) to the probability of the event not occurring (*1-p*). The coefficients $b_0$ and $b_i$ are the estimated parameters for the intercept and the independent variables $X_s$, respectively. The odds ratio can be turn into probability $p$ using an exponential function (Eq. (2).

$$logit(p) = \ln\left(\frac{p}{1-p}\right) = b_0 + b_1{}^*X_1 + b_2{}^*X_2 + \ldots + b_n{}^*X_n \quad (1)$$

$$p = \frac{e^{b_0 + b_1{}^*X_1 + b_2{}^*X_2 + \ldots + b_n{}^*X_n}}{1 + e^{b_0 + b_1{}^*X_1 + b_2{}^*X_2 + \ldots + b_n{}^*X_n}} \quad (2)$$

- CART: The Classification And Regression Tree (CART) model, proposed by Breiman (1984), is a variation of the decision tree algorithm that can handle both classification (as used here) and regression tasks. This predictive classification algorithm builds decision trees to predict a categorical outcome using numeric and categorical inputs. The trees consist of nodes, starting with a root node at the top that takes the training set, which are split into sub-nodes by considering a threshold value of an attribute. The procedure in this study, automatically tests different complexity parameter (cp) values, selects the optimal cp that maximizes cross-validation performance, and fits the final best CART model. Each path from the root to a leaf represents a decision rule.

- NB: Naïve Bayes (NB) is an algorithm constructed using Bayes' theorem with the naïve assumption of conditional independence between every pair of features (Zhang, 2004). To make a prediction, NB calculates *P(data|class)* for each input variable separately and multiplies the results together. NB works well as a classifier for large datasets.

- KNN: K-Nearest Neighbour (KNN) is an algorithm based on a decision rule that assigns the classification of the nearest point of a set of previously classified points to an unclassified sample point (Cover and Hart, 1967). This classification algorithm locates some fixed number of nearest neighbours' points from the training set and uses them to determine the class of the test set.

- SVM: The Support Vector Machine (SVM) model is a binary classifier algorithm that constructs a hyperplane, or set of hyperplanes, based on labelled training data (Cortes and Vapnik, 1995). The hyperplane with the largest distance to the nearest training data points (functional margin) achieves good separation. For a multi-class classification problem, the one-against-one approach is used.

- RF: The Random Forest (RF) is an extension of the bagging algorithm, where several classifiers are created from different bootstrap samples of the training dataset, and is a widely applied ensemble method (Breiman, 2001). RF trains/fits several trees using a random subset of all variables at each split point to reduce the variance between correlated trees. This algorithm uses averaging to improve predictive accuracy and control overfitting.

- XGBoost: the extreme gradient boosting tree algorithm (XGBoost) is a scalable end-to-end tree boosting system (Chen and Guestrin, 2016), where each tree is studied from the prior one and impacts the following trees to improve model performance. Each new tree fits the error rate of the last prediction tree, up to a certain number of trees when the sum of the sample score is required to predict the score for each sample. The final prediction value is the sum of the scores for a sample in different trees.

- ANN: Artificial Neural Network (ANN), first proposed by McCulloch and Pitts (1943), analyses data through a network of decision functions and nodes. The nodes are organized into layers, and each one sends information to the next layer via edges with numeric weights. An activation function determines whether a neuron is activated based on the sum of the connected edges meeting a set threshold. The weights assigned to each edge are unique, preventing the nodes from producing the same solution. Supervised learning is used to reduce the cost value until the model's prediction matches the correct output. Backpropagation is a process of training the neural network that rolls in reverse from the output layer on the right to the input layer on the left, incrementally tweaking the network's weights until the lowest possible cost value is obtained.

The calculation included all 18 explanatory variables (see Table 4). The dataset was randomly split into a training set (80 % of data) and a testing set (20 % of data). The training set was used to fit the models, and the testing set was employed to assess their performance. The binomial LR model was calculated using the *glm* function from the *stats* package in RStudio. The significance level was set at $p$-values $< 0.05$. The ML algorithms were implemented using the framework of the *caret* package in RStudio (Kuhn, 2008). This package provides a uniform interface for model functions and standardizes common tasks, such as parameter tuning and variable importance, through the function *train.* To prevent overfitting, a 5-fold cross-validation procedure during the optimization of hyperparameters shown in Table 3 was applied. Pre-processing technique involving centering and scaling (specifically a min–max scaler) was applied to scale the data within the interval between zero and one. For models with a time-consuming training period (NB, KNN, SVM, ANN), the *tuneGrid* parameter code to compute a range of potential tuning values was employed (see Table 3). In order to assess the performance of the models, we select the optimal classification model based on the largest values of the AUC, accuracy, recall, specificity, precision and F1-score metrics.

*Model validation*. The performance of the models was mainly assessed by quantifying the area under the receiver operating curve (AUC), which measures the ability of a binary classifier to differentiate between classes. A value closer to 1 for the AUC indicated better model performance (note that an AUC value of 0.5 represents random guessing). Additionally, other metrics such as accuracy, recall, specificity, precision and F1-score were calculated for each model, as they are important metrics. Accuracy represents the percentage of correctly predicted classes among all classes. Recall measures the percentage of correctly predicted positive instances among all actual positive instances. Specificity calculates how often the model predicts non-events from the overall non-events. Precision denotes the percentage of correctly predicted positive values among all predicted positive values. F1-score considers both recall and precision through the Eq. (3). The best performing ML model was highlighted and contrasted with the binary LR model.

$$F1 - score = \frac{2*Precision*Recall}{Precision + Recall} \tag{3}$$

## 3. Results

Ecological validity was initially assessed to certify the nature and 3D elements of the web-based experiments through the realism perceived by participants in terms of the virtual graphics. The mean score and standard deviation obtained was $5.05 \pm 2.33$.

The feelings of participants during the experiment were evaluated as the stress perceived during the trial, as well as their survival probability based on the decision they made. The mean score and standard deviation of each item were: stress ($4.60 \pm 2.49$) and survival probability ($6.67 \pm 1.82$).

Table 4 presents the variables analysed in the study, including the profiles of participants, along with the scale ranges, percentages and ANOVA test results (F-value), indicating their association with the dependent variable. The results suggest that several factors, as physical and social conditions, being with a close person, age, education level, occupation, having people to care for, and previous drill experience, may impact evacuation decisions.

Table 5 displays the optimal hyperparameters that were finally selected by *caret* to optimize each metric during the model fitting process. The ability of models for predicting the evacuation decision (the predicted class was *evacuate*) was measured through the performances in Table 6 (confusion matrix results with a threshold of 0.5 for each model) and the AUC values can be compared in Fig. 1. LR and SVM achieved the highest AUC (classification performance) on the testing set, with values of 0.831 and 0.777, respectively. In terms of accuracy, XGBoost had the highest accuracy rate, correctly predicting 78.0 % of cases, followed by ANN (77.3 %) and SVM and CART (77.2 %) models. KNN obtained the highest recall value with 0.859, followed by RF with 0.829. XGBoost had the highest specificity and precision but lower F1-score and AUC compared to ANN. ANN achieved the best F1-score result and CART the second one. There is no consensus on the best metric to evaluate and select a machine learning method. Hence, we regarded the best model for each metric: LR for AUC (0.831), XGBoost for accuracy (0.780), specificity (0.810) and precision (0.820), KNN for recall (0.859) and ANN for F1-score (0.785).

Based on previous results, we contrasted XGBoost and LR. Fig. 2 illustrates the top 10 important variables generated by each model using the *varImp* function from the *caret* package in R. The results for the XGBoost model are the result of averaging the importance of the variables by optimizing each metric. Variables that were important for predicting the evacuation decision in both models included *close person*, *social influence*, *physical context*, *age*, *occupation*, and *previous drill* and *previous fire alarm*. Importantly, only considering the *close person* variable, we founded that 79 % of those participants who were without a close person decided to stay; meanwhile 70 % of participants who were with a close person decided to evacuate. However, considering all the significant variables the percentage of individuals deciding to stay was 54 %.

The results of the LR are included in Table 7. Eight factors are statistically significant ($p$-values $< 0.05$). In terms of odds ratio, we found that the presence of a *close person* increases people likelihood of start evacuation by 961.46 %. Being in a *room* and *see others evacuating*

**Table 3**
Characteristics of each model.

| ML Model | *Caret* function method (documentation) | *train* tuning Hyperparameters | Potential tuning values |
|---|---|---|---|
| CART | "rpart" (Therneau and Atkinson., 2022) | Complexity parameter (cp) | (0 to 0.1, by 0.01) |
| NB | "nb" (Roever et al., 2023) | Laplace Correction (fL) | (1 to 10, by 1) |
| | | Distribution Type (usekernel) | |
| | | Bandwidth Adjustment (adjust) | |
| KNN | "knn" (Ripley and Venables, 2023a) | #Neighbors (k) | (1,5,9,13,17,21,41,61,81, 101,151,201,251,301,351, 401,451) |
| SVM | "svmRadialCost" (Karatzoglou et al., 2023) | Cost (C) | (0.0625,0.125,0.25,0.5,1, 2,4,8,16) |
| RF | "rf" (Breiman et al., 2022) | #Randomly Selected Predictors (mtry) | (2,19,36) |
| XGBoost | "xgbTree" (Chen et al., 2023; Wickham, 2022) | #Boosting Iterations (nrounds) | (50,100,150) |
| | | Max Tree Depth (max_depth) | (1,2,3) |
| | | Shrinkage (eta) | (0.3,0.4) |
| | | Minimum Loss Reduction (gamma) | 0 |
| | | Subsample Ratio of Columns (colsample_bytree) | (0.6,0.8) |
| | | Minimum Sum of Instance Weight (min_child_weight) | 1 |
| | | Subsample Percentage (subsample) | (0.5,0.75,1) |
| ANN | "nnet" (Ripley and Venables, 2023b) | #Hidden Units (size) | (1 to 10, by 1) |
| | | Weight Decay (decay) | (0.2 to 1, by 0.2) |

**Table 4**
Code, descriptive analysis and ANOVA associations of variables.

| Variable | Scale options | Percentage (*n*) | One-way ANOVA F-value |
|---|---|---|---|
| Experimental Factors | | | |
| *Physical factor* | Open area | 49.86 % (901) | 24.08*** |
| | Room | 50.14 % (906) | |
| *Social factor* | Individual | 33.31 % (602) | 74.90*** |
| | Group Go | 33.43 % (604) | |
| | Group Stay | 33.26 % (601) | |
| Familiar Factor | | | |
| *Decision close person* | Evacuate | 70.89 % (1281) | 779.75*** |
| | Stay | 29.11 % (526) | |
| Sociodemographic Factors | | | |
| *Gender* | Male | 48.64 % (879) | 0.36 |
| | Female | 51.36 % (928) | |
| *Age* | [18 – 76] | | 4.84* |
| *BMI* | Normal | 45.82 % (828) | 0.36 |
| | Overweight | 26.80 % (665) | |
| | Obesity | 17.38 % (314) | |
| *Education level* | Primary | 3.93 % (71) | 9.47^ |
| | Secondary | 8.52 % (154) | |
| | Vocational Training (VT) | 22.14 % (400) | |
| | High School | 21.58 % (390) | |
| | University | 43.83 % (792) | |
| *Occupation* | Employed | 51.96 % (939) | 12.56* |
| | Unemployed | 10.96 % (198) | |
| | Self-employed | 6.14 % (111) | |
| | Student | 7.91 % (143) | |
| | Retired | 23.02 % (416) | |
| *Income* | < 1000€ | 17.04 % (308) | 0.64 |
| | 1000-1999€ | 38.13 % (689) | |
| | 2000-2999€ | 26.23 % (474) | |
| | > 3000€ | 18.59 % (336) | |
| *Residence* | Village | 13.50 % (244) | 0.27 |
| | Small city | 34.81 % (629) | |
| | Large city | 51.69 % (934) | |
| *Politic* | Left | 38.79 % (701) | 2.46 |
| | Centre | 20.86 % (377) | |
| | Right | 16.44 % (297) | |
| | Apolitical | 23.91 % (432) | |
| *Religion* | Yes | 33.59 % (607) | 0.23 |
| | No | 66.41 % (1200) | |
| *People care* | Yes | 35.36 % (639) | 4.26* |
| | No | 64.64 % (1168) | |
| *Fitness level* | Very Low | 1.05 % (19) | 3.92 |
| | Low | 4.43 % (80) | |
| | Low-Medium | 11.23 % (203) | |
| | Medium | 25.35 % (458) | |
| | Medium-High | 34.31 % (620) | |
| | High | 14.89 % (269) | |
| | Very-High | 8.74 % (158) | |
| Personal Experience Factors | | | |
| *Previous drill* | Yes | 47.48 % (858) | 3.88* |
| | No | 52.52 % (949) | |
| *Previous fire alarm* | Yes | 23.96 % (433) | 2.31 |
| | No | 76.04 % (1374) | |
| *Previous training* | Yes | 31.82 % (575) | 1.19 |
| | No | 68.18 % (1232) | |
| *Previous FR* | Yes | 12.89 % (233) | 1.68 |
| | No | 87.11 % (1574) | |

Significance codes: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$;^$p < 0.1$.

increase the decision to evacuate by 55.68 % and 89.11 % respectively. *Having people to care for* increases the decision to evacuate by 23.03 %, as well as have previous experience in evacuation drills (23.37 %). Nevertheless, *seeing others staying* reduce the likelihood of evacuation by 25.11 %. Also, *age* is a determinant factor in evacuation decision with retired participants being less willing to evacuate (-39.40 %) and a decrease in evacuation decision of 1.04 % each year.

## 4. Discussion

This study explores logistic regression (LR) and seven machine learning (ML) models to examine the factors that influence evacuation decisions using data from web-based experiments. A total of 1,807 participants took part in six trials, each involving a decision to evacuate or stay in a building after a fire alarm goes off. To prevent learning biases, each participant performed only one trial. The trials combined two physical contexts (open area and room) and three social conditions (alone, others around evacuating and others around staying) as external factors. Personal characteristics (internal factors) were collected through a post-experiment questionnaire.

We found that the LR model provided better prediction capability (AUC) than the ML models. However, in terms of accuracy, CART, SVM, XGBoost and ANN models outperformed LR, being the XGBoost the one with the highest result. The KNN model had higher recall (NB, RF, XGBoost models also outperformed LR), while the ANN model had better F1-score (CART model also obtained a higher result than LR). The XGBoost algorithm also outperformed the other models in terms of specificity and precision (CART, SVM, RF and ANN also obtained higher metrics than LR for these metrics). There is no single best metric for evaluating and selecting a predictive model. In our study, we contrasted XGBoost (higher performance in three metrics) and LR (higher AUC) considered the best candidate models to predict evacuation decision.

Our analysis showed that several variables, including *close person*, *social influence*, *physical context*, *age*, occupation, and previous drill and fire alarm were important in both models (XGBoost and LR). In addition,

**Table 5**
Hyperparameters optimization by *caret.*

| ML Model | Hyperparameter | Applied value | | | | | |
|---|---|---|---|---|---|---|---|
| | | AUC | Accuracy | Recall | Specificity | Precision | F1-score |
| **CART** | cp | 0.00 | 0.1 | 0.1 | 0.01 | 0.01 | 0.1 |
| **NB** | fL | 0 | 0 | 0 | 0 | 0 | 0 |
| | usekernel | TRUE | TRUE | TRUE | TRUE | TRUE | TRUE |
| | adjust | TRUE | TRUE | TRUE | TRUE | TRUE | TRUE |
| **KNN** | k | 451 | 401 | 451 | 13 | 81 | 401 |
| **SVM** | C | 2 | 0.125 | 8 | 0.125 | 0.125 | 4 |
| **RF** | mtry | 2 | 19 | 2 | 19 | 19 | 2 |
| **XGBoost** | nrounds | 50 | 50 | 50 | 50 | 50 | 50 |
| | max_depth | 1 | 2 | 1 | 2 | 2 | 1 |
| | eta | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 |
| | gamma | 0 | 0 | 0 | 0 | 0 | 0 |
| | colsample_bytree | 0.6 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 |
| | min_child_weight | 1 | 1 | 1 | 1 | 1 | 1 |
| | subsample | 1 | 1 | 0.5 | 1 | 1 | 0.5 |
| **ANN** | size | 1 | 1 | 9 | 1 | 1 | 1 |
| | decay | 0.2 | 1 | 1 | 0.4 | 0.4 | 1 |

**Table 6**
Performance of each model (six different trainings per row and per algorithm).

| Model | AUC [95 % CI] | Accuracy | Recall | Specificity | Precision | F1-score |
|---|---|---|---|---|---|---|
| **LR** | **0.831 [0.801–0.861]** | 0.756 | 0.790 | 0.716 | 0.767 | 0.778 |
| **CART** | 0.716 [0.683–0.749] | 0.772 | 0.762 | 0.783 | 0.805 | 0.783 |
| **NB** | 0.744 [0.712–0.776] | 0.748 | 0.793 | 0.695 | 0.753 | 0.773 |
| **KNN** | 0.711 [0.679–0.743] | 0.731 | **0.859** | 0.604 | 0.686 | 0.777 |
| **SVM** | 0.777 [0.747–0.808] | 0.772 | 0.760 | 0.783 | 0.805 | 0.776 |
| **RF** | 0.699 [0.666–0.731] | 0.754 | 0.829 | 0.770 | 0.791 | 0.757 |
| **XGBoost** | 0.761 [0.730–0.793] | **0.780** | 0.793 | **0.810** | **0.820** | 0.776 |
| **ANN** | 0.770 [0.739–0.800] | 0.773 | 0.732 | 0.783 | 0.805 | **0.785** |

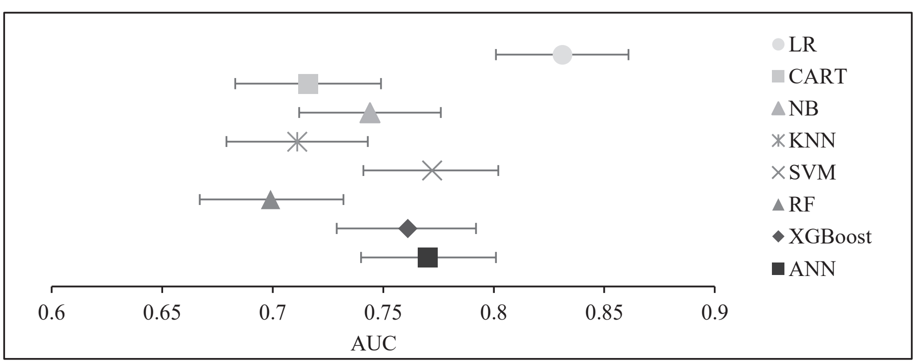


**Fig. 1.** AUC with 95 % CI.

*having people to care for* was significant predictor by LR. The results highlight the importance of considering the impact of various factors on the evacuation decision. As mentioned in the introduction section, people tend to disregard warnings in buildings. Our results confirmed this trend, as 54 % of participants chose to stay instead of evacuating. The lack of additional information besides the alarm sound, the absence of any perceived threat (such as smoke or flames), the virtual nature of the scenario, and the proposed activity of waiting for medical results may explain this behaviour. An interesting finding was observed when participants were asked to imagine being with a *close relative*. This time, when no other factors were considered, 70 % of individuals said they would evacuate, which contrasts with the 79 % who chose to stay when they were not with a close person. This confirms the impact of affiliative behaviour theory in evacuation decision (Mawson, 1978; Sime, 1983). In fact, the presence of a close person was the most important variable in predicting evacuation decision in both XGBoost and LR (%OR = 961). Furthermore, *seeing others* evacuating was the second most influential variable. The likelihood of deciding evacuation increases by 89 % when confederates left the area. This outcome is in line with studies such as (Cuesta et al., 2021; Haghani and Sarvi, 2017; Kinateder and Warren, 2016), which suggest that people tend to follow the actions of others. The next significant variable is the *physical context*. Our results showed that people in enclosures are more likely to start evacuation than people in open spaces (LR = 55 %), perhaps owning to a lack of visual access to remaining spaces within the building or the situation. This is an important issue for future research. Finally, each year reduced the likelihood of evacuating by 1.04 % in LR model. This could be because of the experimental design, as the older participants may have been more focused on receiving their medical results and therefore more likely to ignore the fire alarm.

Previous research has explored the performance of ML techniques compared to other statistical predictive methods in the field of human behaviour and decision-making. The results of these studies have been mixed, with some reporting better performance for ML techniques

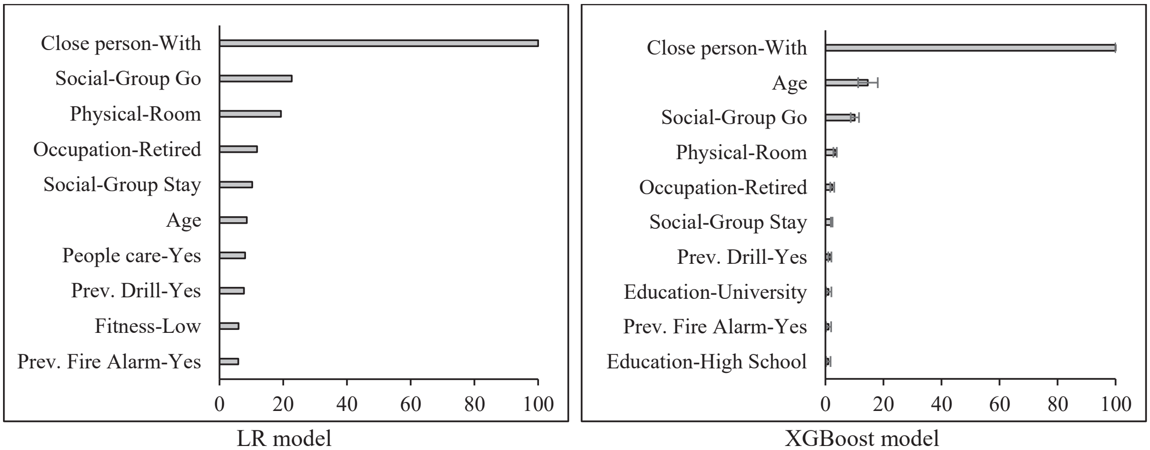


**Fig. 2.** Importance of the variables selected by LR and XGBoost models.

**Table 7**
Results of LR.

| Variable | b | SE | Wald | *p*-value | OR | %OR | 95 % CI for OR | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower | Upper |
| **Intercept** | **−1.903** | **0.564** | **−3.373** | **0.001****** | **0.149** | **−85.08 %** | **0.049** | **0.446** |
| **Physical-Room** | **0.443** | **0.091** | **4.882** | **0.000****** | **1.557** | **55.68 %** | **1.304** | **1.861** |
| **Social-Group Go** | **0.637** | **0.111** | **5.735** | **0.000****** | **1.891** | **89.11 %** | **1.522** | **2.353** |
| **Social-Group Stay** | **−0.289** | **0.110** | **−2.621** | **0.009**** | **0.749** | **−25.11 %** | **0.603** | **0.929** |
| **Close person-With** | **2.362** | **0.094** | **25.223** | **<2e-16****** | **10.615** | **961.46 %** | **8.850** | **12.778** |
| Gender-Male | −0.060 | 0.100 | −0.599 | 0.549 | 0.942 | −5.83 % | 0.774 | 1.146 |
| **Age** | **−0.010** | **0.005** | **−2.196** | **0.028*** | **0.990** | **−1.04 %** | **0.980** | **0.999** |
| BMI-Overweight | −0.018 | 0.106 | −0.167 | 0.868 | 0.982 | −1.75 % | 0.798 | 1.210 |
| BMI-Obesity | 0.064 | 0.140 | 0.456 | 0.649 | 1.066 | 6.61 % | 0.810 | 1.404 |
| Education-Secondary | −0.009 | 0.279 | −0.033 | 0.974 | 0.991 | −0.92 % | 0.574 | 1.716 |
| Education-VT | −0.051 | 0.253 | −0.203 | 0.839 | 0.950 | −5.02 % | 0.579 | 1.565 |
| Education-High School | 0.235 | 0.254 | 0.924 | 0.356 | 1.265 | 26.47 % | 0.770 | 2.089 |
| Education-University | 0.286 | 0.251 | 1.142 | 0.253 | 1.332 | 33.15 % | 0.817 | 2.184 |
| Occupation-Unemployed | 0.008 | 0.157 | 0.051 | 0.959 | 1.008 | 0.81 % | 0.742 | 1.370 |
| Occupation-Self-employed | −0.117 | 0.194 | −0.601 | 0.548 | 0.890 | −11.03 % | 0.607 | 1.302 |
| Occupation-Student | −0.264 | 0.205 | −1.292 | 0.196 | 0.768 | −23.23 % | 0.514 | 1.146 |
| **Occupation-Retired** | **−0.501** | **0.167** | **−2.995** | **0.003**** | **0.606** | **−39.40 %** | **0.436** | **0.841** |
| Incomes-(1000-1999€) | −0.101 | 0.138 | −0.729 | 0.466 | 0.904 | −9.59 % | 0.689 | 1.186 |
| Incomes-(2000-2999€) | −0.067 | 0.153 | −0.440 | 0.660 | 0.935 | −6.49 % | 0.693 | 1.261 |
| Incomes-(>3000€) | −0.036 | 0.168 | −0.214 | 0.830 | 0.965 | −3.54 % | 0.694 | 1.341 |
| Residence-Small city | 0.072 | 0.145 | 0.501 | 0.616 | 1.075 | 7.52 % | 0.810 | 1.428 |
| Residence-Large city | 0.067 | 0.141 | 0.473 | 0.636 | 1.069 | 6.88 % | 0.811 | 1.409 |
| Politic-Centre | 0.004 | 0.125 | 0.029 | 0.977 | 1.004 | 0.36 % | 0.786 | 1.282 |
| Politic-Right | −0.202 | 0.140 | −1.438 | 0.150 | 0.817 | −18.27 % | 0.620 | 1.076 |
| Politic-Apolitical | −0.065 | 0.122 | −0.533 | 0.594 | 0.937 | −6.29 % | 0.738 | 1.190 |
| Religion-Yes | −0.049 | 0.103 | −0.479 | 0.632 | 0.952 | −4.80 % | 0.779 | 1.164 |
| **People care-Yes** | **0.207** | **0.100** | **2.063** | **0.039*** | **1.230** | **23.03 %** | **1.011** | **1.498** |
| Fitness-Low | 0.755 | 0.493 | 1.533 | 0.125 | 2.128 | 112.82 % | 0.816 | 5.679 |
| Fitness-Low-Medium | 0.662 | 0.465 | 1.422 | 0.155 | 1.938 | 93.79 % | 0.785 | 4.910 |
| Fitness-Medium | 0.500 | 0.456 | 1.096 | 0.273 | 1.649 | 64.89 % | 0.680 | 4.107 |
| Fitness-Medium-High | 0.547 | 0.455 | 1.201 | 0.230 | 1.728 | 72.76 % | 0.714 | 4.294 |
| Fitness-High | 0.627 | 0.463 | 1.355 | 0.176 | 1.873 | 87.27 % | 0.761 | 4.725 |
| Fitness-Very High | 0.516 | 0.476 | 1.085 | 0.278 | 1.675 | 67.51 % | 0.664 | 4.326 |
| **Prev. Drill-Yes** | **0.210** | **0.107** | **1.969** | **0.049*** | **1.234** | **23.37 %** | **1.001** | **1.521** |
| Prev. Fire Alarm-Yes | 0.172 | 0.113 | 1.520 | 0.128 | 1.188 | 18.80 % | 0.951 | 1.484 |
| Prev. Training-Yes | −0.120 | 0.110 | −1.088 | 0.277 | 0.887 | −11.31 % | 0.714 | 1.101 |
| Prev. FR-Yes | −0.184 | 0.142 | −1.296 | 0.195 | 0.832 | −16.80 % | 0.630 | 1.098 |

*b - Coefficient; SE - Standard Error; OR - Odds Ratio (exp(b)); %OR - (exp(b)-1); CI - confidence interval.*
*Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05;^p < 0.1.*

(Lindner et al., 2017; Xu et al., 2023), others for mixed logit models (Zhu et al., 2023), and others finding no significant difference between methods (Wang and Ross, 2018). In this study, both the XGBoost ML model and traditional LR model produced similar results in terms of variable importance and their influence on evacuation decisions. The LR model had a better ability to differentiate between classes, as indicated by a higher AUC when predicting evacuation decisions during building fire alarm situations, whereas XGBoost exhibited superior accuracy in correctly predicting more classes. Both models offer advantages in data analysis, including applicability to large datasets. Furthermore, the XGBoost model excels in identifying complex patterns and handling missing data and outliers. However, limitations exist for both models. The LR model faces challenges in handling complex relationships, addressing class imbalance, and linearity between predictors and outcomes. The XGBoost model is a powerful yet complex algorithm, demanding a deeper understanding of its hyperparameters. It also lacks interpretability when compared to other models and is resource-intensive, requiring strong computational resources. Overall, the LR model was found to be accurate, pragmatic and well-calibrated. Incorporating a greater number of variables enhances calculation precision. Therefore, regression models should continue to play a key role in predicting evacuation decisions.

All models in this study are suitable for evacuation decision modelling application, but the choice should align with model interpretation, specific optimization goals, and the scenario. LR model provides faster calculations and greater interpretability, especially regarding the importance of included factors. Its coefficients (b) or %OR can be directly used in simulations to categorize the percentage of agents evacuating or staying. CART model, although slightly less performant, also outperforms LR in several categories but AUC and it is the easiest model to read. In cases of unbalanced data influencing the final decision, consider using an ANN, which achieved a higher F1-score (combining recall and precision). However, when rapid calculation is necessary and the data is balanced, the XGBoost model is a strong candidate, offering higher accuracy, specificity, and precision. The sample size of 1,807 individuals underscores the advantages of using a web-based method for data collection, such as reduced participants risk and the flexibility to explore various environmental conditions. Furthermore, this method is cost-effective compared to other types of data collection. Nevertheless, both the innovative approach to collect data from human behaviour experiments and the data analysis in this study have their limitations. Firstly, significant effort is required to design and built the environment to be as realistic as possible. The 3D environment in online experiments is unable to fully replicate an immersive space or the presence of participants, and it cannot faithfully reproduce the interactions and characteristics of people and objects. The perceived realism, rated by participants at 5.04 out of 10, reflects a moderate value. This underscores the importance of considering this limitation when using this methodology and interpreting its results. Furthermore, additional research, such as conducting real experiments, is needed in this area to fully validate this type of experiment. Secondly, considering the experimental procedure, it is unclear whether participants strictly followed the instructions, such as using headphones to enhance their experience or whether they performed the experiment under optimal conditions (i.e., being alone in a quiet place). Researchers must consider the autonomy of participants and may require additional confirmatory screens and information when conducting an online experiment. Moreover, it is worth noting that 57 % of participants conducted the experiment using mobile phones, which could introduce distractions and interruptions that might affect their level of attention and focus during the experiment. Furthermore, the use of smaller keyboards or touchscreens on mobile devices may pose challenges for participants when entering data or answering to complex experimental tasks. Thirdly, it is possible that participants did not interpret the acoustic alarm signal as intended or encountered difficulties with the virtual environment, potentially influencing their reactions. Lastly, since the study is limited to the Spanish population, its cultural perspective may be narrow, and restrict the generalization of the findings to other cultural contexts.

## 5. Conclusions

The contributions of this study can be summarized as follows: 1) supporting existing background knowledge that most occupants do not respond effectively to fire alarms; 2) identifying that evacuation decisions during building fire alarms are influenced by a combination of physical attributes of buildings, social influence, and personal conditions; 3) generating quantitative data that can be used to predict human behaviour in the context of fire evacuation scenarios; 4) using advanced technology to collect a large amount of behaviour data; and 5) finding that both logistic regression and machine learning models can accurately predict factors that influence evacuation decisions of individuals during fire alarms.

Further research is necessary to validate the findings of this study through non-immersive experiments and to gain additional insights. Additionally, future studies should investigate the impact of other factors that were not considered here. Gathering data from other countries and cultures would also be crucial to explore potential differences or similarities in evacuation decisions during building fire alarms. This could help to identify cultural factors that may influence behaviour and contribute to the development of more effective evacuation strategies in different contexts, ultimately ensuring the safety of building occupants during emergencies. Finally, machine learning algorithms could be further optimized by adjusting them with their specific codes rather than using the general *caret* package. This approach would allow their individual performance to be evaluated.

## CRediT authorship contribution statement

**Adriana Balboa:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Arturo Cuest:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Javier González-Villa:** Writing – review & editing, Methodology, Formal analysis, Conceptualization. **Gemma Ortiz:** Writing – review & editing, Project administration, Funding acquisition. **Daniel Alvear:** Writing – review & editing, Supervision, 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.

## Data availability

Data will be made available on request.

## Acknowledgements

The authors would like to acknowledge the Spanish Ministry of Science and Innovation and the Spanish State Research Agency funding the project “TRACES: UNDERSTANDING HUMAN BEHAVIOUR IN CASE OF TERRORISM ATTACKS IN MASS GATHERING BUILDINGS” PID2019-106025RB-100/AEI/10.13039/501100011033, within the 2019 call for R + D + I Projects in the framework of the State Program for Knowledge Generation and Scientific and Technological Strengthening for the R + D + I system and the Challenges of Society, under the State Plan for Scientific and Technical Research and Innovation 2017-2020 - “R + D + I Projects”.

## References

Berkson, J., 1944. Application of the logistic function to bio-assay. J. Am. Stat. Assoc. 9, 357–365. https://www.jstor.org/stable/2280041.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman, L., Cutler, A., Liaw, A., Wiener, M., 2022. Package ‘randomForest’ version 4.7-1.1.

Breiman, L., 1984. Classification and regression trees. (1st ed.), Routledge. https://doi.org/10.1201/9781315139470.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., Yuan, J., 2023. Package ‘xgboost’ version 1.7.5.1.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: In: Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. https://doi.org/10.1007/BF00994018.

Cover, T.M., Hart, P.E., 1967. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 13, 21–27. https://doi.org/10.1109/TIT.1967.1053964.
Cubukcu, E., 2003. Investigating wayfinding using virtual environments. Dissertation.
Cuesta, A., Abreu, O., Balboa, A., Alvear, D., 2021. Alone or with others: Experiments on evacuation decision making. Fire Saf. J. 120, 103018 https://doi.org/10.1016/j.firesaf.2020.103018.
D'Orazio, M., Occhialini, M., Bernardini, G., Ferracuti, F., Iarlori, S., Longhi, S., 2016. Fire exit signs: The use of neurological activity analysis for quantitative evaluations on their perceptiveness in a virtual environment. Fire Saf. J. 82, 63–75. https://doi.org/10.1016/j.firesaf.2016.03.003.
Ferraro, V., Settino, J., 2019. Evacuation and smart exit sign system. In: Cicirelli, F. et al. (eds.), The Internet of Things for Smart Urban Ecosystems, Springer International Publishing AG, pp. 363–383. https://doi.org/10.1007/978-3-319-96550-5_15.
Fritz, C., Marks, E., 1954. The NORC Studies of Human Behavior in Disaster.
Fu, L., Cao, S., Song, W., Fang, J., 2018. The influence of emergency signage on building evacuation behavior: An experimental study. Fire Mater. 43, 22–33. https://doi.org/10.1002/fam.2665.
Haghani, M., Sarvi, M., 2017. Following the crowd or avoiding it? Empirical investigation of imitative behaviour in emergency escape of human crows. Anim. Behav. 124, 47–56. https://doi.org/10.1016/j.anbehav.2016.11.024.
Jeon, G.-Y., Choi, J.-H., Hong, W.-h., Kim, J.-h., 2014. Study on relation between age and evacuation in case of underground evacuation. Contemp. Eng. Sci. 7, 1045–1052. https://doi.org/10.12988/ces.2014.49131.
Karatzoglou, A., Smola, A., Hornik, K., NICTA, Maniscalco, M.A., Teo, C., 2023. Package 'kernlab' version 0.9-32.
Kinateder, M., Warren, W.H., 2016. Social influence on evacuation behavior in real and virtual environments. Front. Robot. AI 3. https://doi.org/10.3389/frobt.2016.00043.
Kinateder, M., Comunale, B., Warren, W.H., 2018. Exit choice in an emergency evacuation scenario is influenced by exit familiarity and neighbor behaviour. Saf. Sci. 106, 170–175. https://doi.org/10.1016/j.ssci.2018.03.015.
Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28, 1–26. https://doi.org/10.18637/jss.v028.i05.
Kuligowski, E., 2021. Evacuation decision-making and behavior in wildfires: Past research, current challenges and a future research agenda. Fire Saf. J. 120, 103129 https://doi.org/10.1016/j.firesaf.2020.103129.
Kuligowski, E.D., Mileti, D.S., 2009. Modeling pre-evacuation delay by occupants in world trade center towers 1 and 2 on september 11, 2001. Fire Saf. J. 44, 487–496. https://doi.org/10.1016/j.firesaf.2008.10.001.
Kuligowski, E., Zhao, X., Lovreglio, R., Xu, N., Yang, K., Westbury, A., Nilsson, D., Brown, N., 2022. Modeling evacuation decisions in the 2019 kincade fire in California. Saf. Sci. 146, 105541 https://doi.org/10.1016/j.ssci.2021.105541.
Kuligowski, E.D., 2009. The Process of Human Behavior in Fires - NIST Technical Note 1632.
Kuligowski, E.D., 2011. Communicating the Emergency: Preliminary findings on the elements of an effective public warning message - NIST Technical Note 1689.
Lin, J., Cao, L., Li, N., 2019. Assessing the influence of repeated exposures and mental stress on human wayfinding performance in indoor environments using virtual reality technology. Adv. Eng. Inf. 39, 53–61. https://doi.org/10.1016/j.aei.2018.11.007.
Lin, W., Li, Z., Zhang, X., Gao, Y., Lin, J., 2023. Electrophysiological evidence for the effectiveness of images versus text in warnings. Sci. Rep. Nat. 13, 1278. https://doi.org/10.1038/s41598-023-28230-x.
Lindell, M., Perry, R., 2012. The protective action decision model: theoretical modifications and additional evidence. Risk Anal. 32, 616–632. https://doi.org/10.1111/j.1539-6924.2011.01647.x.
Lindner, A., Pitombo, C.S., Cunha, A.L., 2017. Estimating motorized travel mode choice using classifiers: An application for high-dimensional multicollinear data. Travel Behav. Soc. 6, 100–109. https://doi.org/10.1016/j.tbs.2016.08.003.
Liu, M., Lo, S.M., 2011. The quantitative investigation on people's pre-evacuation behavior under fire. Autom. Constr. 20, 620–628. https://doi.org/10.1016/j.autcon.2010.12.004.
Lo, S., Zhang, P., 2009. An artificial neural-network based predictive model for pre-evacuation human response in domestic building fire. Fire Technol. 45, 431–449. https://doi.org/10.1007/s10694-008-0064-6.
Lovreglio, R., Ronchi, E., Nilsson, D., 2015. A model of the decision-making process during pre-evacuation. Fire Saf. J. 78, 168–179. https://doi.org/10.1016/j.firesaf.2015.07.001.
Mawson, A.R., 1978. Panic behavior: A review and a new hypothesis. World Congress of Sociology, Uppsala, Sweden.
McCaffrey, S., Wilson, R., Konar, A., 2018. Should I stay or should I go now? or should I wait and see? Influences on wildfire evacuation decisions. Risk Anal. 38, 1390–1404. https://doi.org/10.1111/risa.12944.
McCulloch, W., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5, 115–133.
McLennan, J., Elliott, G., Omodei, M., 2012. Householder decision-making under imminent wildfire threat: stay and defend or leave? Int. J. Wildland Fire 21, 915–925. https://doi.org/10.1071/WF11061.
McLennan, J., Ryan, B., Bearman, C., Toh, K., 2019. Should we leave now? Behavioral factors in evacuation under wildfire threat. Fire Technol. 55, 487–516. https://doi.org/10.1007/s10694-018-0753-8.
Midi, H., Sarkar, S.K., Rana, S., 2010. Collinearity diagnostics of binary logistic regression model. J. Interdiscip. Math. 13, 253–267. https://doi.org/10.1080/09720502.2010.10700699.
Nilsson, D., Kinateder, M., 2015. Virtual reality experiments – The future or a dead end?. Conference: Human Behaviour in Fire Symposium 2015, Cambridge.
Pathfinder – Technical Reference, 2022. Thunderhead Engineering. https://support.thunderheadeng.com/docs/pathfinder/2022-1/technical-reference-manual/.
Proulx, G., 2001. Occupant behaviour and evacuation. In: In: Proceedings of the 9th International Fire Protection Symposium, pp. 219–232.
Pyrosim User Manual, 2022. Thunderhead Engineering. https://files.thunderheadeng.com/support/documents/pyrosim-user-manual-2022-1.pdf.
Reneke, P., 2013. Evacuation Decision Model - NISTIR 7914. https://doi.org/10.6028/NIST.IR.7914.
Richardson, O., Jalba, A., Muntean, A., 2018. Effects of environmental knowledge in evacuation scenarios involving fire and smoke: A multiscale modelling and simulation approach. Fire Technol. 55, 415–436. https://doi.org/10.1007/s10694-018-0743-x.
Ripley, B., Venables, W., 2023a. Package 'class' version 7.3-22.
Ripley, B., Venables, W., 2023b. Package 'nnet' version 7.3-19.
Roever, C., Raabe, N., Luebke, K., Ligges, U., Szepannek, G., Zentgraf, M., Meyer, D., 2023. Package 'klaR' version 1.7-2.
Santos, G., Aguirre, B., 2004. A critical review of emergency evacuation simulation models. Proceeding of Conference: Building Occupant Movement during Fire Emergencies. National Institute of Standards and Technology, Gaithersburg, Maryland.
Saunders, W., 2001. Decision making model of behaviour in office building fire evacuations. Thesis, Department of Psychology, Australia.
Sherman, M.F., Peyrot, M., Magda, L.A., Gershon, R.R.M., 2011. Modeling pre-evacuation delay by evacuees in world trade center towers 1 and 2 on september 11, 2001: A revisit using regression analysis. Fire Saf. J. 46, 414–424. https://doi.org/10.1016/j.firesaf.2011.07.001.
Sime, J.D., 1983. Affiliative behaviour during escape to building exits. J. Environ. Psychol. 3, 21–41. https://doi.org/10.1016/S0272-4944(83)80019-X.
Song, X., Lovreglio, R., 2021. Investigating personalized exit choice behavior in fire accidents using the hierarchical Bayes estimator of the random coefficient logit model. Anal. Methods Accident Res. 29 https://doi.org/10.1016/j.amar.2020.100140.
Stoet, G., 2010. PsyToolkit - A software package for programming psychological experiments using Linux. Behavl. Res. Methods 42, 1096–1104.
Stoet, G., 2017. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. Teach. Psychol. 44, 24–31.
Sun, Y., Huang, S.-K., Zhao, X., 2023. Predicting hurricane evacuation decisions with interpretable machine learning models. Environ. Sci. https://doi.org/10.48550/arXiv.2303.06557.
Thakur, S., Ranjitkar, P., Rashidi, S., 2022. Investigating evacuation behaviour under an imminent threat of volcanic eruption using a logistic regression-based approach. Saf. Sci. 149, 105688 https://doi.org/10.1016/j.ssci.2022.105688.
Therneau, T.M., Atkinson, E.J., 2022. An Introduction to Recursive Partitioning Using the RPART Routines.
Toledo, T., Marom, I., Grimberg, E., Bekhor, S., 2018. Analysis of evacuation behavior in a wildfire event. Int. J. Disaster Risk Reduct. 31, 1366–1373. https://doi.org/10.1016/j.ijdrr.2018.03.033.
Vilar, E., Rebelo, F., Noriega, P., 2018. Smart systems in emergency wayfinding: A literature review. In: Marcus, A., Wang, W. (Eds.), Design, User Experience, and Usability: Designing Interactions. Springer International Publishing AG, pp. 379–388. https://doi.org/10.1007/978-3-319-91803-7_28.
Viswanathan, V., Lees, M., 2014. An information processing based model of pre-evacuation behavior for agent based egress simulation. In: Weidmann, U., Kirsch, U., Schreckenberg, M. (Eds.), Pedestrian and Evacuation Dynamics 2012. Springer - Verlag, Germany, pp. 125–133.
Wang, F., Ross, C.L., 2018. Machine learning travel mode choices: Comparing the performance of an extreme gradient boosting model with a multinomial logit model. Transp. Res. Rec. 47, 35–45. https://doi.org/10.1177/0361198118773556.
Wickham, H., 2022. Package 'plyr' version 1.8.8.
Wong, S., Broader, J., Walker, J., Shaheen, S., 2022. Understanding California wildfire evacuee behavior and joint choice making. Transportation. https://doi.org/10.1007/s11116-022-10275-y.
Wood, P.G., 1972. The behaviour of people in fires. Fire Res. Station 953, 1–113.
Xu, N., Lovreglio, R., Kuligowski, E., Cova, T.J., Nilsson, D., Zhao, X., 2023. Predicting and assessing wildfire evacuation decision-making using machine learning: Findings from the 2019 kincade fire. Fire Technol. https://doi.org/10.1007/s10694-023-01363-1.
Yamada, T., Akizuki, Y., 2016. Visibility and human behavior in fire smoke. SFPE Handbook of Fire Protection Engineering, M.J. Hurley (ed.), Society of Fire Protection Engineers. DOI 10.1007/978-1-4939-2565-0_61.
Zhang, H., 2004. The optimality of Naïve Bayes. Am. Association Artif. Intell.
Zhao, X., Lovreglio, R., Nilsson, D., 2020. Modelling and interpreting pre-evacuation decision-making using machine learning. Autom. Constr. 113, 103140 https://doi.org/10.1016/j.autcon.2020.103140.
Zhu, R., Becerik-Gerber, B., Lin, J., Li, N., 2023. Behavioral, data-driven, agent-based evacuation simulation for building safety design using machine learning and discrete choice models. Adv. Eng. Inf. 55, 101827 https://doi.org/10.1016/j.aei.2022.101827.