ARTICLE

Predicting Hurricane Evacuation Decisions with Interpretable Machine Learning Methods

Yuran Sun[1] · Shih-Kai Huang[2] · Xilei Zhao[1]

Abstract
Facing the escalating effects of climate change, it is critical to improve the prediction and understanding of the hurricane evacuation decisions made by households in order to enhance emergency management. Current studies in this area often have relied on psychology-driven linear models, which frequently exhibited limitations in practice. The present study proposed a novel interpretable machine learning approach to predict household-level evacuation decisions by leveraging easily accessible demographic and resource-related predictors, compared to existing models that mainly rely on psychological factors. An enhanced logistic regression model (that is, an interpretable machine learning approach) was developed for accurate predictions by automatically accounting for nonlinearities and interactions (that is, univariate and bivariate threshold effects). Specifically, nonlinearity and interaction detection were enabled by low-depth decision trees, which offer transparent model structure and robustness. A survey dataset collected in the aftermath of Hurricanes Katrina and Rita, two of the most intense tropical storms of the last two decades, was employed to test the new methodology. The findings show that, when predicting the households' evacuation decisions, the enhanced logistic regression model outperformed previous linear models in terms of both model fit and predictive capability. This outcome suggests that our proposed methodology could provide a new tool and framework for emergency management authorities to improve the prediction of evacuation traffic demands in a timely and accurate manner.

Keywords Artificial Intelligence (AI) · Decision-making modeling · Hurricane evacuation · Interpretable machine learning · Nonlinearity and interaction detection

1 Introduction

The population in the U.S. coastline regions reached 94.7 million people (29.1% of the total U.S. population) in 2017, up from 47.4 million people six decades ago, an increase of 99.8% (Census Bureau 2019). Meanwhile, over 50% of households live within 50 miles of the shoreline (Kildow et al. 2016). Furthermore, the major population growth is within the Gulf of Mexico regions and the adjacent Atlantic regions, which are most vulnerable to hurricane risks (Census Bureau 2019). Therefore, establishing planning-based evacuation strategies becomes urgent and essential for most coastal counties (Lindell 2019). Disaster researchers have been closely monitoring this issue and striving to improve knowledge and technology (Yang et al. 2019; Davidson et al. 2020). As Sorensen (2000) has noted, one of the primary challenges in the modern era is enhancing prediction to enable timely and effective responses.

Research on evacuation decision predictions and derivative traffic demand estimations has been booming since the systematic review of Baker (1991) of hurricane evacuation case studies. Major follow-up studies have focused on three related areas: (1) exploring factors facilitating or inhibiting evacuation (Alawadi et al. 2020; Ersing et al. 2020); (2) examining causal relationships between predictors and evacuation decisions (Horney et al. 2010; Ling et al. 2022); and (3) developing comprehensive cause-effect models (Huang et al. 2012; Huang et al. 2017). While these studies utilizing bivariate or linear analysis methods (Lindell and Perry 2012; Huang et al. 2016; Huang et al. 2017; Lindell 2018) have

✉ Xilei Zhao
  xilei.zhao@essie.ufl.edu

[1] Department of Civil and Coastal Engineering, University of Florida, Gainesville, FL 32611, USA

[2] Department of Emergency Management and Public Administration, Jacksonville State University, Jacksonville, AL 36265, USA

contributed to evacuation demand model development (Lindell 2013; Tanim et al. 2022), they tend to overemphasize the role of individuals' psychological processes in mediating the effects of information seeking/receiving and other contextual circumstances (Lindell and Perry 2012; Lindell 2018), resulting in certain limitations.

First, previous linear or bivariate-based analyses primarily test causality rather than optimize predictions (Huang et al. 2012). Therefore, these models, like other social science models, may have lower model fit and accuracy in forecasting evacuation demand, probably due to their limited ability to capture interactions and nonlinearities (Zhao et al. 2020). Next, psychological models face limitations related to availability of psychological data and consistency of measuring instruments (Baker 1991). Additionally, they risk overlooking crucial impacts of social and household contexts, such as resource availability (Schorr 2015; Metaxa-Kakavouli et al. 2018), social vulnerability (Metaxa-Kakavouli et al. 2018; Meyer et al. 2018), and social capital (Metaxa-Kakavouli et al. 2018; Fraser 2022), when estimating evacuation demands.

Possible measures can be adopted to address these limitations. Recent advancements in Artificial Intelligence (AI) and machine learning (ML) offer opportunities to capture complex relationships within data and improve prediction accuracy (Zhao et al. 2021). To tackle the potential issues of the black-box nature of machine learning models (for example, lacking interpretability and trust (Rudin 2019)), a recent effort is to adopt interpretable machine learning models for high-stakes decision making (Rudin 2019; Zhao et al. 2021). A prominent example is the penalized logit tree regression model developed by Dumitrescu et al. (2022), which utilizes low-depth decision trees to improve logistic regression. The approach involves building one-layer decision trees with binary splits for each predictor, creating binary variables to represent each resulting leaf node. Subsequently, two-layer decision trees are constructed for every possible pair of predictors, partitioning the data into three leaf nodes. Similarly, binary variables are generated for each leaf node. The final model is a penalized logistic regression that includes binary variables associated with the leftmost leaf node of each one-layer decision tree, and binary variables associated with the two leftmost leaf nodes of each two-layer decision tree. The model attains satisfactory prediction performance, exhibiting significant improvement over the original logistic regression model and comparable results to black-box machine learning models like random forest, while maintaining high transparency.

Given the advantages of the method by Dumitrescu et al. (2022), we decided to apply it to predict hurricane evacuation decisions to assist life-and-death decision making in emergencies. With its ability to reveal underlying relationships between social and household contextual variables and evacuation decisions, we can place more emphasis on these variables and the dependence on psychological variables can be reduced. However, before its application in this context, necessary improvements for Dumitrescu et al. (2022) are needed due to: (1) the model proposed by Dumitrescu et al. (2022) solely consists of an extensive set of binary variables, limiting its interpretability for decision making; (2) the model is unable to track trends in predictor contributions; and (3) the statistical significance of the nonlinearities (binary variables derived from decision trees) is untested, potentially leading to overfitting.

The primary goal of this study was to improve the predictive accuracy of the logistic regression models with machine learning techniques, and to provide novel behavioral insights to enhance understanding of hurricane evacuation decision making. Specifically, the study set the following objectives: (1) Extend the model developed by Dumitrescu et al. (2022) to improve its interpretability and enable the model to better track predictor contributions. Additionally, leverage statistical testing to only retain the statistically significant nonlinearities. (2) Empirically develop an evacuation decision prediction model considering social and household contextual variables and then examine the model fit and accuracy against traditional models. (3) Derive new behavioral insights on how these variables affect evacuation decisions through the model interpretation.

The following sections are organized as follows: Section 2 provides a thorough literature review of key predictors for evacuation decisions and related modeling techniques. Section 3 details the development of an enhanced logistic regression (ELR) framework and the model performance comparison metrics. Section 4 describes the Hurricanes Katrina and Rita evacuation survey data utilized in the ELR model. Section 5 compares the ELR model's performance with baseline models (the model without nonlinear effects and the model with psychological variables) and discusses the results of the ELR model. Section 6 encapsulates the key findings, addresses the study's limitations, and outlines directions for future research.

2 Literature Review

This section offers a comprehensive review of relevant studies within the current literature, detailing their findings, the characteristics of these studies, as well as identifying research gaps and limitations. It is organized into subsections that discuss: the predictors used in models analyzing household evacuation decisions; significant predictors identified by existing literature as influential in evacuation decisions; linear models developed for predicting and understanding these decisions; and the application of AI and ML models to detect nonlinear

effects and their use in studying hurricane evacuation behaviors beyond decision making. A concluding summary subsection synthesizes the identified research gaps and limitations, providing insights derived from these observations.

2.1 Predictors for Households' Evacuation Decisions

Previous review studies (Baker 1991; Huang et al. 2016) categorized predictors for household hurricane evacuation decisions into social and household contexts, geographic characteristics, information-related items, and psychological variables. In general, social and household contexts include gender, age, household size, race, income, education level, previous experience, home structure, as well as available resources and preparedness levels; geographic variables are the proximity to hazardous areas; information-related variables measure information sources and warning dissemination; psychological variables include individuals' perceptions and expectations to the resource-related concerns, threat conditions and exposures, and possible consequences (Baker 1991; Gladwin 1997).

Most hurricane evacuation studies (Huang et al. 2012; Lazo et al. 2015; Huang et al. 2017) have used this framework or a subset thereof to develop their statistical or mathematical models. While social and household contexts have been discussed most, they often yielded none or marginal significance (Huang et al. 2017; Sadri et al. 2017; Sarwar et al. 2018). Among them, females and mobile homeowners tended to evacuate more, whereas minorities and homeowners usually hesitated to leave (Huang et al. 2016; Huang et al. 2017). Few studies have investigated resource requirements and availabilities, though some (Lazo et al. 2015; Sadri et al. 2017; Sarwar et al. 2018) found vehicle ownership significantly affected evacuation decisions. In fact, evacuation logistic studies often discussed resource-related variables (Sarwar et al. 2018). Proximity to risk sources, regardless of measurement, was one of the strongest predictors (Huang et al. 2017; Sadri et al. 2017; Sarwar et al. 2018) as it mirrored evacuees' immediate (or intuitive) responses to the threat exposures (Huang et al. 2016; Huang et al. 2017).

Similarly, information-related variables, including official warnings and social or environmental cues, were vital (Huang et al. 2012) as they were direct stimuli (Huang et al. 2016; Huang et al. 2017). For psychological variables, perceived threat exposures and expected consequences were key in evacuation decision models in most studies (Lazo et al. 2015; Huang et al. 2016; Huang et al. 2017), as the process whereby mental cognitions determined the behavioral decisions was consistent with the elaboration likelihood model of persuasion (Petty and Cacioppo 1986).

2.2 Linear Modeling Approaches to Analyzing Hurricane Evacuation Decisions

The aforementioned predictors have long been incorporated into advanced data analyses to reach the results with comprehensive concerns. Among diverse modeling techniques, descriptive and regression analyses are the most common methods. One prototypical example of hurricane evacuation decisions can be traced back to the study of Hurricane Andrew by Peacock and Gladwin in 1997 (Gladwin 1997). The study compared evacuation percentages in different residential zones and used logistic regressions to analyze predictors' contributions to evacuation decisions.

Since then, mixed logistic regression models have frequently been employed to explain the correlations and effects between predictors and evacuation decisions, including studies of Hurricanes Lili (Lindell et al. 2005), Katrina and Rita (Huang et al. 2017), Ike (Huang et al. 2012), Sandy (Sadri et al. 2017), and Ivan (Hasan et al. 2011; Sarwar et al. 2018). Furthermore, two studies by Huang and his colleagues (Huang et al. 2012; Huang et al. 2017) and Morss et al. (2016) developed multistage regression analyses to clarify the path of effects among predictors.

In addition, ordered probit models were utilized in Cahyanto et al.'s study on tourists' hurricane evacuation decisions (Cahyanto et al. 2014) and Lazo et al.'s hypothetical study (Lazo et al. 2015). Recently, Ahmed et al. (2020) stepped forward by modeling the influence of social networks on evacuation decisions using zero-inflated Poisson, Tobit, linear regression, and the multinomial logistic regression models.

However, hurricane evacuation decision making can be nonlinear, multilayered, and much more complex (Gladwin et al. 2007). Many efforts have attempted to capture this complexity using crowd modeling methods (Vreugdenhil et al. 2015) or tree-based frameworks (Zhao et al. 2020) to detect nonlinearities, but these are still in their infancy as the findings from those pioneering studies are in need of further verification.

2.3 Artificial Intelligence (AI) and Machine Learning (ML) for Evacuation Modeling

Evacuation decisions can be nonlinear, multilayered, and complex, which may not be fully captured by linear models (Gladwin et al. 2007). Although research specifically targeting the complexities of hurricane evacuation decisions is limited, progress has been made through studies in other contexts (Xu et al. 2023; Zhu et al. 2023). Many of these

studies have utilized AI and ML techniques to effectively capture the multifaceted nature of evacuation behaviors.

Zhao et al. (2020) predicted the pre-evacuation behavior of building occupants, utilizing a machine learning algorithm—specifically, a random forest—to grasp the nonlinearities. They also employed partial dependence plots (PDPs) to illustrate the complex interplay between contributing factors and evacuation behavior. Nara et al. (2021) identified the dependency relationships between contributing factors and evacuation decisions using the Bayesian network method. This approach captures nonlinearity through the network's structure and its conditional probability distributions. Lo et al. (2009) aimed to model the stochastic and ambiguous aspects of human behavior during pre-evacuation in domestic building fires by integrating fuzzy logic into an artificial neural network (ANN) for prediction.

Furthermore, AI and ML techniques have been applied to various aspects of hurricane response beyond evacuation decisions. For instance, these techniques play an important role in the planning of contraflow, which is an essential component of hurricane evacuation strategies. Burris et al. (2015) identified the optimal timing for the activation of the contraflow lane with traffic condition information, employing decision tree methods in their approach. Bagloee et al. (2019) implemented a hybrid supervised learning approach within an optimization framework for the design of contraflow. Besides, AI and ML techniques have also been applied for destination location prediction (Anyidoho et al. 2023), routing (Li et al. 2023), and traffic demand prediction (Roy et al. 2021).

2.4 Summary of Research Gaps

Upon reviewing the literature, we identified existing research gaps as follows: (1) Most existing studies have employed linear models like logistic regression models to examine the factors influencing evacuation decisions. However, such linear models may not adequately represent the complexity inherent in evacuation decision making, thus leading to poor predictive accuracy and simplistic behavioral understanding. (2) Artificial Intelligence and ML have been applied in some recent hurricane evacuation studies to improve model performance and/or identify nonlinear effects. Nevertheless, most of these applications rely on complex, black-box models that lack transparency and interpretability. (3) For most existing models of hurricane evacuation decisions, a majority of significant variables are psychological variables. In contrast, social and household contextual variables are either often deemed insignificant or insufficiently studied. Consequently, the contributions of these variables to evacuation decisions remain less understood, although such variables may be equally important as psychological variables and have greater accessibility and more consistent measurements.

3 Methods

From the literature review, logistic regression models are found effective in capturing the relation between predictors and evacuation decisions (Hasan et al. 2011; Sadri et al. 2017; Sarwar et al. 2018). However, these models might over-rely on psychological variables, which are subjective and vulnerable to expense, timeliness in gathering, and consistency issues in data collection (Huang et al. 2017). Meanwhile, many objective and easily-accessible social and household contextual variables (for example, demographic and resource-related variables) reportedly have minor or indirect effects on evacuation decisions.

This study began by proposing a framework for establishing the interpretable machine learning model, hypothesizing that the inconsistent contributions of the objective predictors may be attributable to the oversimple linear structure of logistic regression models. In essence, the study posited that such predictors might have crucial nonlinear effects on evacuation decisions. Thereby, as illustrated in Fig. 1, this study developed a new methodology to (1) build a model with only demographic, geographic, and resource-related predictors that could attain similar (or even better) predictive power to the logistic regression model with psychological variables; and (2) detect and model the nonlinearities with the interpretable machine learning approach (Rudin 2019), offering more flexibility and intelligence than classic statistical methods (for example, subjectively identifying nonlinear relationships from scatter plots) while maintaining model transparency.

Our study focused on two common nonlinearities: **univariate** and **bivariate threshold effects**. Univariate threshold effects pertain to the varying impacts of a single predictor on a response variable across different ranges. Bivariate threshold effects refer to the interaction effect of two predictors on a response variable, which exists only when the values of these two predictors fall within specific ranges. We used an approach similar to Dumitrescu et al.'s (2022) for detecting threshold effects with low-depth decision trees but with innovative modifications: (1) instead of using extensive binary variables in the logistic regression model like Dumitrescu et al. (2022), incorporating original predictors along with their nonlinearities, cutting points (the thresholds for different intervals where a predictor may have varying impacts on the response variable) and interaction terms identified by one-layer and two-layer decision trees, into the logistic regression model, making it easier to identify and interpret the causal effects of specific variables
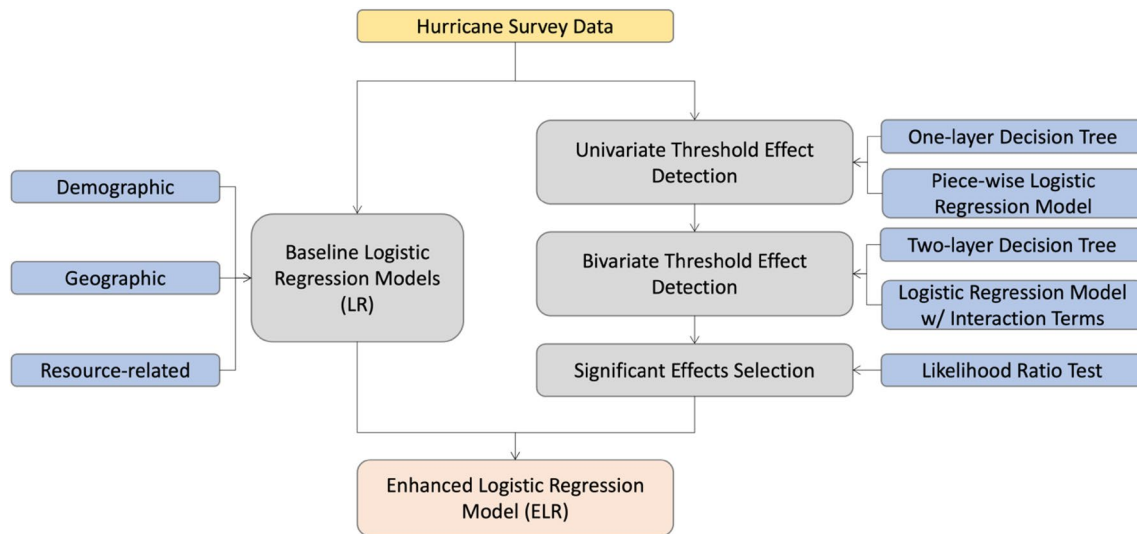
Fig. 1 Overview of the hurricane evacuation decisions study methodology

by examining the resulting coefficients; (2) conducting statistical tests to examine the significance of nonlinearities; and (3) retaining only significant nonlinearities to reduce dimensionality and prevent overfitting. Decision tree structures are ideal for capturing interactions between predictors, and they can visually display thresholds and classification results, thus offering good explanations (Molnar 2020). Particularly, low-depth decision trees are the most robust and provide straightforward explanations (since for each split, the observation falls into either one or the other leaf, and binary decisions are easy for humans to understand) (Molnar 2020); therefore, low-depth decision trees are suitable for threshold effects detection.

Figure 1 outlines the overall framework. We first built baseline logistic regression models (LR) based on demographic, geographic, and resource-related variables. Next, we detected **univariate** and **bivariate threshold effects** using low-depth decision trees. We then included those significant threshold effects that passed the likelihood ratio tests in the logistic regression model to form our enhanced logistic regression model (ELR). The following subsections provide details of threshold effects, baseline and enhanced logistic regression models, threshold effect detecting using low-depth decision trees, statistical testing process, and the comparison of the ELR model with other baseline models.

3.1 Threshold Effects

This study focused on two common nonlinear relationships between predictors and evacuation decisions: univariate and bivariate threshold effects. The univariate threshold effect is the threshold effect of a single predictor. Denote the predictor as x_i, where $i \in \{1, 2, \cdots, p\}$ and p is the number of

predictors. A univariate threshold effect exists if x_i's relationship with evacuation decisions varies depending on whether x_i is below or above a certain constant threshold a_i. This can be observed via different coefficients in the logistic regression model for $x_i \leq a_i$ and $x_i > a_i$ (details of the logistic function are provided in Sect. 3.2).

The bivariate threshold effect concerns the interaction term between two predictors. Denote two predictors as x_i and x_j where $i, j \in \{1, 2, \cdots, p\}$. x_i and x_j have the bivariate threshold effect if their interaction term (that is, $x_i \cdot x_j$) affects the evacuation decision when the values of x_i and x_j fall in one of the following intervals: (1) $x_i \leq b_i$ and $x_j \leq b_j$; (2) $x_i \leq b_i$ and $x_j > b_j$; (3) $x_i > b_i$ and $x_j \leq b_j$; and (4) $x_i > b_i$ and $x_j > b_j$ (the logistic function is shown in Sect. 3.2).

While within-category interactions might occur, they are typically minor. For example, a household's socioeconomic status might potentially moderate the effect of household size on evacuation decisions. However, resource availability and needs are generally more influential than socioeconomic contexts when it comes to interpreting such effects. Hence, our study assumed that bivariate threshold effects occurred between different category predictors and did not consider within-in category interactions.

3.2 Incorporating Threshold Effects into Logistic Regression Models

Denote the evacuation decision as a binary variable $y_k \in \{0, 1\}$ and the input data for the k_{th} observation as a p-dimensional vector x_k, where p is the number of predictors, $k = 1, 2, \cdots, n$, and n is the number of observations. The logistic regression model estimates the probability $P(y_k = 1 | x_k)$

that the k_{th} respondent decides to evacuate during the hurricane given his/her features x_k. A household is deemed to evacuate if this probability is larger than a threshold π, and to stay at home otherwise. The logistic function of the baseline logistic regression model (LR) is:

$$P(y_k = 1|x_k, \beta) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{p} \beta_i x_{k,i})}}, \qquad (1)$$

where $\beta = (\beta_0, \beta_1, \cdots \beta_p)^T$ is the set of parameters to be estimated. The optimum estimator $\hat{\beta}$ is obtained by maximizing the log-likelihood function:

$$L(\beta) = \sum_{k=1}^{n} y_k \ln(P(y_k = 1|x_k, \beta)) + \sum_{k=1}^{n} (1 - y_k) \ln(1 - P(y_k = 1|x_k, \beta)). \qquad (2)$$

Our modified logistic regression models incorporate threshold effects for predicting evacuation probability. In particular, for each predictor x_i, after its threshold a_i is detected (details in Sect. 3.3), x_i is regarded to have a univariate threshold effect on the response variable and the coefficients of x_i in the logistic model vary depending on whether x_i is above or below the threshold a_i. The model's logistic function with the univariate effect of x_i is expressed as:

$$P(y_k = 1|x_k, \beta^{(i)}) = \frac{1}{1 + e^{-(\beta_0 + (\sum_{j=1}^{p} \beta_j x_{k,j}) + \phi_i x_{k,i} I(x_{k,i} > a_i))}}, \qquad (3)$$

where $\beta^{(i)} = (\beta, \phi_i)^T$ is the set of parameters to be estimated.

In the case of two predictors, x_i and x_j, after their bivariate thresholds b_i and b_j are detected (details in Sect. 3.3), their interaction is considered influential for evacuation decision when they are in one of the corresponding ranges, that is, $x_i \leq b_i$ and $x_j \leq b_j$, $x_i \leq b_i$ and $x_j > b_j$, $x_i > b_i$ and $x_j \leq b_j$, $x_i > b_i$ and $x_j > b_j$. The model's logistic function with the bivariate effect of x_i and x_j is written as:

$$P(y_k = 1|x_k, \beta^{(i,j)}) = \frac{1}{1 + e^{-(\beta_0 + (\sum_{m=1}^{p} \beta_m x_{k,m}) + \phi_{ij} \gamma_{k,i} \cdot \gamma_{k,j})}}, \qquad (4)$$

where $\gamma_i \cdot \gamma_j$ is the bivariate threshold effect that fulfills:

and $\beta^{(ij)} = (\beta, \phi_{ij})^T$ is the set of parameters to be estimated.

It is worth noting that not every threshold effect has statistical significance. Only significant ones are meaningful and should be included in the final model to avoid misinterpretation from data fluctuations. The criteria and process for selecting significant effects are thoroughly outlined in Sect. 3.4. The logistic function of the ELR model, incorporating all significant threshold effects is represented as:

$$P(y_k = 1|x_k, \beta_{ELR}) = \frac{1}{1 + e^{-\left(\beta_0 + \sum_{i=1}^{p} \beta_i x_{k,i} + \sum_{j=1}^{l} \phi_{uj} x_{k,uj} I(x_{k,uj} > a_{uj}) + \sum_{m=1}^{q} \phi_{bm1bm2} \gamma_{k,bm1} \cdot \gamma_{k,bm2}\right)}}, \qquad (5)$$

where $(x_{u1}, x_{u2}, \cdots, x_{ul})^T$ is the set of predictors having significant univariate threshold effects, with l representing their count. $((\gamma_{b11}, \gamma_{b12}), \cdots, (\gamma_{bq1}, \gamma_{bq2}))^T$ is the set of pairs of predictors having significant bivariate threshold effects, with q indicating their count, $\beta_{ELR} = (\beta, \phi_{u1}, \phi_{u2}, \cdots, \phi_{ul}, \phi_{b11b12}, \cdots, \phi_{bq1bq2})^T$ is the set of parameters to be estimated.

3.3 Detecting Threshold Effects with Low-Depth Decision Trees

Decision Tree is a tree-structured model utilized for regression and classification, with the classification and regression tree (CART) algorithm being especially popular. In each iteration of the CART algorithm, the decision node and its splitting value minimize the Gini Index (the impurity level of the observations). The dataset is split into two subsets according to the splitting value, making the observations within the resulting partitions more homogeneous. This procedure is repeatedly carried out until a pre-specified condition is met (for example, the maximal depth of the tree is reached), or the Gini Index can no longer decrease. The algorithm detects optimal threshold values and facilitates capturing interactions through binary splits, making it suitable for automatically uncovering nonlinear relationships. Moreover, compared with non-pruned trees (multi-layer trees), low-depth decision trees are more robust with fewer splits. Hence, we adopted an approach similar to that of Dumitrescu et al. (2022) by leveraging the low-depth CART, as it is well-suited for threshold effects detection.

For detecting univariate threshold effects of p variables, p one-layer decision trees were built, each using one of the p

$$\gamma_i \cdot \gamma_j = \begin{cases} x_i \cdot x_j, \gamma_i \leq b_i, \gamma_j \leq b_j \text{ or } \gamma_i > b_i, \gamma_j \leq b_j \text{ or } \gamma_i \leq b_i, \gamma_j > b_j \text{ or } \gamma_i > b_i, \gamma_j > b_j \\ 0, \text{ otherwise} \end{cases}$$
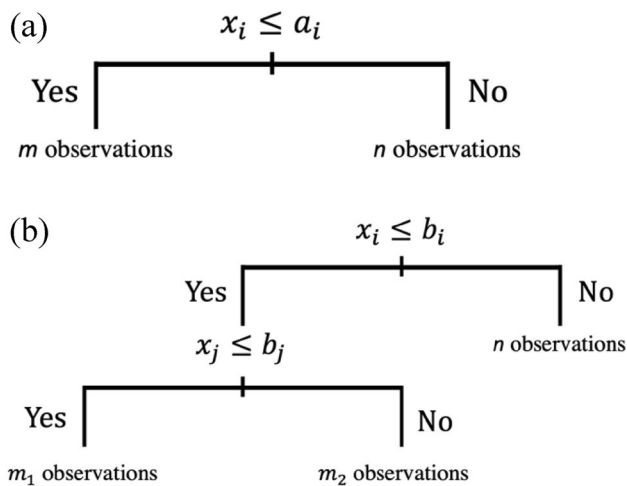
Fig. 2 Examples of decision tree for threshold effects detection. **a** Univariate threshold effect detection; **b** Bivariate threshold effect detection.

variables as the root node. The splitting value of the root node of each tree was selected as the threshold. The i_{th} decision tree with the variable x_i as the candidate node is taken as an example to explain the one-layer decision tree in detail. As depicted in Fig. 2a, the tree comprises one root node (x_i) and two leaf nodes, each containing m and n observations. With the threshold observed as a_i, the contribution of x_i to the response variable is different when x_i is below and above a_i.

Bivariate threshold effects were detected by running decision trees with only two from the p variables as the candidate variables. Shapes of decision trees may differ depending on the way of splitting. The decision tree with variables x_i and x_j is taken as an example for illustration. As shown in Fig. 2b, the tree is a two-layer tree with three leaf nodes, each with $m_1, m_2,$ and n observations. x_i is the decision node of the first iteration, and x_j participates in the second splitting iteration. The two leftmost leaf node observations are regarded to have bivariate threshold effects. That is, when $x_i \leq b_i$ and $x_j \leq b_j$, or $x_i \leq b_i$ and $x_j > b_j$, the interaction term, $x_i \cdot x_j$, affects evacuation decisions.

Some variables x_i may have high feature importance and hence participate in the first two splitting iterations. For detecting their bivariate threshold effects with other variable x_j, three-layer decision trees were also applied where x_i was the decision node for the first two splits and x_j participated in the third split. We limited the trees to three layers to avoid sparse leaf nodes.

3.4 Significant Threshold Effects Selection

To determine the significance of detected threshold effects, we conducted the likelihood ratio test (LRT). The LRT tests the goodness-of-fit between nested models. In this study,

LRT was applied to assess if the models incorporating a univariate effect of x_i, or a bivariate effect of x_i and x_j, fitted the dataset significantly better than the original model. For each model that includes the univariate threshold effect, the likelihood ratio is:

$$\text{LR}^{(i)} = -2\left[L\left(\widehat{\beta}\right) - L\left(\widehat{\beta}^{(i)}\right)\right]. \tag{6}$$

For each model that includes the bivariate threshold effect, the likelihood ratio is:

$$\text{LR}^{(ij)} = -2\left[L\left(\widehat{\beta}\right) - L\left(\widehat{\beta}^{(ij)}\right)\right]. \tag{7}$$

The likelihood ratio approximately followed a χ^2 distribution with a single degree of freedom (only one additional parameter was added to the new model). If the p-value of the LRT was less than 0.01, the model with either a univariate or bivariate threshold effect was considered to be superior to the original model, and the two models were deemed statistically different. If the model with a univariate threshold effect of x_i was statistically better than the original model, and the p-values of both x_i and $x_i I(x_i > a_i)$ were less than 0.01, the univariate effect of x_i was considered to be significant. If the model with a bivariate threshold effect of x_i and x_j had the p-value of LRT smaller than 0.01, and the p-value of $\gamma_i \cdot \gamma_j$ was smaller than 0.01, the bivariate effect of x_i and x_j was deemed significant.

3.5 Model Performance Comparison

To evaluate the model performance, we split the dataset into a 90% training set for threshold detection and model fitting, and a 10% test set for model performance assessment (Bhavan et al. 2019; Yang et al. 2021). This split mitigates overfitting, secures ample training data, and still maintains a sufficiently large test set for robust evaluation. We then examined if the ELR model with all significant threshold effects outperformed the two baseline models: two standard logistic models with and without psychological variables (Huang et al. 2017). We compared model performance using measures including R^2 and adjusted R^2 for fitting ability; and accuracy, precision, recall, F1 score, and AUC for prediction ability.

R^2 and adjusted R^2, ranging from 0 to 1, measure the proportion of the variance in the response variable explained by the predictors. Higher values of them indicate better explanations of the data by the regression model. For the response variable $y_i, i = 1, 2, \cdots, n$, n is the number of observations, R^2 is defined as:

$$R^2 = \frac{SS_{Regression}}{SS_{Total}} = \frac{\sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}, \tag{8}$$

where $SS_{Regression}$ is the sum of squares due to the regression and SS_{Total} is the total sum of squares, $\hat{y}_i$ is the predicted value of y_i and $\bar{y}$ is the mean of the response values for all observations. Adjusted R^2 is an adjusted version of R^2, which considers the degree of freedom. It is defined as:

$$\text{Adjusted } R^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1}, \tag{9}$$

where p is the number of predictors.

To evaluate our ELR model's prediction results, we also compared the models' accuracy, precision, recall, F1 score, and AUC. Accuracy measures the proportion of correctly predicted observations over the total observations. It is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{10}$$

where TP (true positive) is the number of correctly predicted "to evacuate," TN (true negative) is the number of correctly predicted "to stay at home," FP (false positive) is the number of incorrectly predicted "to stay at home," and FN (false negative) is the number of incorrectly predicted "to evacuate." Precision measures the proportion of correctly identified positive responses over the positive predictions made. It is defined as:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{11}$$

Recall measures the proportion of correctly identified positive responses out of all actual positive responses. It is defined as:

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{12}$$

F1 score is the harmonic mean of precision and recall. It is defined as:

$$F1 = \frac{2\frac{TP}{TP+FN}\frac{TP}{TP+FP}}{\frac{TP}{TP+FN} + \frac{TP}{TP+FP}}. \tag{13}$$

AUC, or the area under the ROC (receiver operating characteristics) curve—a plot of the true positive rate against false positive rate—also ranges from 0 to 1. When AUC is close to 1, the model is better at distinguishing between respondents choosing "to evacuate" and "to stay at home."

4 Data

In our study, we used a post-Hurricanes Katrina and Rita survey (Huang et al. 2017) for nonlinearity detection and model construction.

4.1 Hurricanes Katrina and Rita

Hurricane Katrina is known for its severe destruction and significant loss of life. On 27 August 2005, a hurricane watch was declared by the National Hurricane Center (NHC) at 10:00 a.m. Central Daylight Time (CDT). A hurricane warning followed 13 hours later when the intensity of Katrina reached Category 3 and was still strengthening. The following day, 28 August, it upgraded to a Category 5 hurricane upon reaching the Gulf of Mexico. Over the next three days, it gradually shifted from the southwest to the north (Knabb et al. 2023). On 29 August, with the intensity of Category 3, Katrina made landfall near Buras, Louisiana; the storm surge was 2.4 to 6.7 m (8 to 22 ft).

In September 2005, Hurricane Rita swiped Texas (Knabb et al. 2006). A hurricane watch was issued by NHC at 4:00 p.m. CDT on 21 September, which was upgraded to a warning 18 hours later as Rita reached Category 5. The storm first threatened Corpus Christi, Texas, then turned eastward to Galveston, Texas. At 2:38 a.m. CDT on 24 September, with the intensity of Category 3, it made landfall close to Johnson Bayou, Louisiana, and Sabine Pass, Texas. The storm surge was between 1.2 and 2.1 m (4 to 7 ft).

4.2 Data Collection Procedures

Four months following Hurricane Katrina, two mail surveys were initiated by the Hazard Reduction and Recovery Center (HRRC) of Texas A&M University. The Katrina survey included two parishes in Louisiana—St. Charles and Jefferson. The Rita survey covered seven Texas counties—three inland counties (Hardin, Jasper, and Newton) and two coastal counties (Orange and Jefferson) from the Sabine study area (SSA), along with one inland county (Harris) and one coastal county (Galveston) from the Houston-Galveston study area (GSA).

A disproportionate stratified sampling procedure was adopted to select participants. Anticipating a 50% response rate and 200 respondents from each parish and county, 800 and 2800 questionnaires were mailed to households in Louisiana and Texas, respectively. The distribution and collection of questionnaires were guided by Dillman's guidelines (Dillman 2011). Each chosen household received a packet containing a cover letter, a questionnaire, and a prepaid, stamped return envelope. Households that did not return a completed questionnaire within two weeks received a reminder postcard. If the questionnaire remained uncompleted, the household would receive replacement packets every two weeks until the

questionnaire was filled out or three replacement packets and one reminder postcard had already been sent.

Of the 800 households that received the Katrina survey, 270 returned valid questionnaires, marking a response rate of 39.9% (37% in Jefferson and 43% in St. Charles). For the Rita survey sent to 2800 households, 1007 completed questionnaires were received, yielding a response rate of 41.8% (the response rates across all counties were similar). Cases from the two surveys were combined after a homogeneity test (Huang et al. 2017).

4.3 Treatment of Missing Data

Of the 1277 observations, 558 (42.7%) had missing data. Although the missing rates of most variables were lower than 5%, these missing values may significantly affect the results. Therefore, before further analyses, the missing values were substituted using the expectation-maximization (EM) algorithm.

4.4 Variable Description

Our study only focused on demographic, geographic, and resource-related variables. Psychological variables were omitted from the model development. The updated dataset comprises 1277 observations and 15 variables, with one response variable **EvaDec** (Evacuation Decision) and 14

predictors. For the values of **EvaDec**, 0 is for "To Stay at Home," the proportion is 17.07%, 1 is for "To Evacuate," and the proportion is 82.93%. All the predictors, except for the binary predictors (Female, White, Married, and HmOwn), are continuous variables. The predictor name, description, type, value or scale, and the meaning of each value are illustrated in Table 1.

For the values of the predictor **RiskArea**, 0 represents Barrier Islands, 1 is assigned to areas prone to Category 1 or 2 hurricanes (Zip-Zone A for GSA, Risk Areas 1 and 2 for SSA, and Phase I for Louisiana State), 2 pertains to locations susceptible to Category 3 hurricanes (Zip-Zone B for GSA, Risk Area 3 for SSA, and Phase II for Louisiana State), 3 is for locations that could be affected by Category 4 or 5 hurricanes (Zip-Zone C for GSA, Risk Areas 4 and 5 for SSA, and Phase III for Louisiana), and 4 corresponds to locations farther inland. The larger the value, the farther the respondent was from the coast.

5 Results

This section presents a comparative analysis of the model performance between the ELR models and the baseline models. Additionally, it includes the outcomes from all models and offers a detailed discussion and interpretation of these results.

Table 1 Predictors included in the hurricane evacuation decisions study

Variable	Description	Mean	Std.	Min	Max	Proportion	Category
Female	Gender	0: 48.63% 1: 51.37%				0: Male (48.63%) 1: Female (51.37%)	Demographic
White	Race	0: 22.47% 1: 77.53%				0: Others (22.47%) 1: White (77.53%)	
Married	Marital status	0: 30.54% 1: 69.46%				0: Others (30.62%) 1: Married (69.38%)	
HmOwn	House ownership	0: 12.61% 1: 87.39%				0: Others (12.61%) 1: Own house (87.39%)	
Age	Age	53.53	15.08	19.00	94.00	–	
HHSize	Household size	2.84	1.57	1.00	17.00	–	
Edu	Education years	13.98	2.42	9.00	18.00	–	
Income	Annual income (USD)	38,088.00	12,660.93	15,000.00	53,398.00	–	
RiskArea	Risk area	0: 2.35% 1: 9.48% 2: 24.67% 3: 23.72% 4: 39.78%				0: Barrier Island 1: Risk Area A or 1 and 2 2: Risk Area B or 3 3: Risk Area C or 4 and 5 4: Inland	Geographic
RegVeh	Registered vehicle number	2.15	0.93	0.00	9.00	–	Resource-related
EvaVeh	Estimated number of vehicles to take in the evacuation	1.41	0.70	0.00	5.00	–	
EvaTrail	Estimated number of trailers to take in the evacuation	0.12	0.32	0.00	2.00	–	
EvaCost	Estimated cost for evacuation	1178.00	1875.15	0.00	41,150.00	–	

Table 2 Model performance comparison in the hurricane evacuation decisions study

Model	In-sample performance		Out-of-sample performance				
	R^2	Adj R^2	Accuracy	Precision	Recall	F1 score	AUC
Baseline LR	0.11	0.10	0.77	0.77	0.99	0.87	0.50
Baseline LR w/psychological variables	0.30	0.29	0.85	0.86	0.97	0.91	0.71
ELR w/significant univariate threshold effects	0.54	0.53	0.88	0.89	0.96	0.92	0.77
ELR w/all significant threshold effects	0.83	0.83	0.94	0.93	0.99	0.96	0.87

LR logistic regression model, *ELR* enhanced logistic regression model.

5.1 Model Performance Comparison Results

Table 2 presents performance metrics for four models: the baseline logistic regression (LR) model with demographic, geographic, and resource-related variables; the baseline LR model incorporating psychological variables selected by Huang et al. (2017); the ELR model with significant univariate threshold effects; and the ELR model with all significant threshold effects.

As shown in Table 2, the model without psychological variables performs the worst due to insignificant linear contributions. However, the ELR model with the significant univariate threshold effects has already outperformed the model with psychological variables from both model fit and prediction ability: R^2 and adjusted R^2 increase from 0.3 to 0.5; Accuracy, precision, and F1 score increase about 3.0%, 3.5%, and 1.5%, respectively; AUC increases about 18%. This indicates that correctly detecting nonlinear effects of demographic, geographic, and resource-related factors can yield a model as predictive as one with psychological variables. The ELR model with all significant threshold effects outperforms all other models: R^2 and adjusted R^2 increase to 0.8. Accuracy, precision, F1 score, and AUC are larger than those of ELR with significant univariate effects. It indicates that the model performance is further improved with interaction term inclusion.

5.2 Significant Variables and Interpretations

Our study focused on the univariate threshold effects of all continuous predictors, and the bivariate threshold effects of variables from different categories, that is, one demographic/geographic variable and one resource-related variable, as discussed in the Method section. The model comparison results demonstrate the superior performance of the ELR model, making it suitable for evacuation decision prediction and analysis. We utilized the ELR model, along with other baseline models, on the complete dataset to compare the significant factors that influence evacuation decisions during Hurricanes Katrina and Rita across different models. Table 3 displays the results from several models: the

baseline LR, the baseline LR with psychological factors, the ELR model with only the univariate threshold effects, and the ELR model encompassing all significant thresholds (4 univariate and 4 bivariate threshold effects). Based on the findings in Table 3, we made the following observations and interpretations.

In the baseline model, it can be observed that only two social and household contextual variables have a significant influence on evacuation decisions. Specifically, gender exhibits a generalized linear effect on evacuation decisions, where women are more inclined towards evacuation ($\beta = 0.657$). Additionally, the distance from the coast (RiskArea) emerges as another significant factor impacting evacuation decisions, as residents farther from the coast are less prone to evacuate ($\beta = -0.882$).

The model with psychological factors shows that variables with significant influence on evacuation decisions are predominantly psychological variables. These include receiving official warnings ($\beta = 0.796$), expected rapid onset ($\beta = -0.368$), expected wind impacts ($\beta = 0.667$), and expected evacuation impediments ($\beta = -0.448$). In contrast, aside from RiskArea ($\beta = -0.719$), other social and household contextual variables have no significant effects on the evacuation decisions.

The ELR model including univariate effects has already yielded some new insights. It reveals that marital status is a significant variable, with married individuals being more inclined to evacuate ($\beta = 0.601$). Besides, the nonlinear effects can be interpreted. When the household size is smaller than 2.39, it negatively contributes to the evacuation decision ($\beta = -0.687$). However, as it exceeds 2.39, its negative impact decreases ($\beta\prime = -0.117$). While larger households are more likely to contain elderly individuals who have more evacuation impediments (for example, disability and massive medical equipment), they are also more likely to include children. Previous research by Gladwin (1997) suggests that families with children are more inclined to evacuate, which may explain the diminishing effect of household size when it exceeds 2.39. When the number of registered vehicles is smaller than 2.01, it has a positive contribution to the decision making ($\beta = 1.503$).

Table 3 Results of different logistic regression models used in the hurricane evacuation decisions study

Variable	Baseline LR	Baseline LR w/ psychological variables	ELR w/significant univariate threshold effects	ELR w/all significant threshold effects
Intercept	5.113	0.701	− 1.713	1.383
Age	− 0.009	− 0.007	0.011	− 0.009
Female	0.657***	0.237	1.130***	1.535***
White	0.080	0.179	− 0.052	− 0.190
Married	0.167	− 0.062	0.601*	1.480***
HHSize	0.046	0.060	− 0.687**	− 0.984**
Edu	− 0.005	0.022	0.063	0.045
Income	0.000	0.000	0.000	0.000
HmOwn	− 0.280	− 0.174	− 0.616	− 0.312
RiskArea	− 0.882***	− 0.719***	− 0.750***	− 0.329
RegVeh	0.080	–	1.503***	0.681~
EvaVeh	− 0.099	–	3.900***	3.121***
EvaTrail	0.030	–	0.195	0.229
EvaCost	0.000	–	0.007***	0.003**
HrrExp	–	− 0.016	–	–
UnnecEvac	–	0.116	–	–
NewsMedia	–	0.079	–	–
OffWarn	–	0.796***	–	–
SocCues	–	0.032	–	–
EnvCues	–	0.048	–	–
ExRapOnset	–	− 0.368***	–	–
ExStmThreat	–	− 0.039	–	–
ExHydroImp	–	0.171	–	–
ExWindImp	–	0.667***	–	–
ExEvacImp	–	− 0.448***	–	–
HHSize (> 2.39)	–	–	0.570***	0.238
RegVeh (> 2.01)	–	–	− 0.912***	− 0.626**
EvaVeh (> 1.00)	–	–	− 4.081***	− 4.815***
EvaCost (> 704.03)	–	–	− 0.007***	− 0.003**
RiskArea (> 3.45)· EvaCost (> 704.03)	–	–	–	0.000
HHSize (≤ 15.00)· RegVeh (> 2.99)	–	–	–	0.157*
HHSize (≤ 13.50)· EvaVeh (> 2.00)	–	–	–	0.969***
Edu (> 10.33)· EvaCost (≤ 511.22)	–	–	–	0.000**

LR logistic regression model, *ELR* enhanced logistic regression model

*** < 0.001, ** < 0.01, * < 0.05–< 0.1

However, as the number exceeds 2.01, its influence becomes more negligible ($\beta\prime = 0.591$). This can be explained by the concept of diminishing marginal utility, where households only drive a necessary number of cars for evacuation, despite owning additional vehicles that could potentially provide extra resources for evacuation. The estimated number of vehicles required for the evacuation positively affects the evacuation decision when less than 1.00 ($\beta = 3.900$). When it exceeds 1.00, it negatively contributes to the decision making ($\beta\prime = -0.181$). This suggests that people who are able to take a car for evacuation are more likely to evacuate, whereas additional vehicle requirements could pose a burden on household evacuation. When the estimated evacuation cost is smaller than USD 704.03, it positively contributes to the evacuation decision ($\beta = 0.007$). When it is above USD 704.03, its influence becomes negligible ($\beta\prime = 0.000$). This implies that households have a plan regarding the expense scale of their evacuation trip. Overall, the results indicate that increased resource demands and costs could hesitate households' ultimate evacuation decisions.

The ELR model including all significant threshold effects offers enhanced insights. It identifies not only nonlinearity

within individual variables but also significant interaction effects. These interaction effects play an equally important role in influencing evacuation decisions. The interaction between household size (≤ 15.00) and the number of registered vehicles (> 2.99) positively affects the evacuation decision. This means that as household size increases, the coefficient of registered vehicles also increases. For larger families, the number of registered vehicles contributes more positively to decision making. This is logical because a larger family typically requires more means to transport people; thus, having multiple vehicles is more advantageous for these households to evacuate. Similarly, the interaction term of household size (≤ 13.5) and the number of vehicles to take in the evacuation (> 2.00) positively influences the decision. This indicates that as the household size increases, the coefficient of vehicles to take also increases. For larger families, the number of cars to take positively affects decision making. Besides, the interaction between education (> 10.33) and evacuation cost (≤ 511.22) has a positive contribution to decision making. This demonstrates that as education year increases, the coefficient of estimated evacuation cost also increases.

6 Discussion

The enhanced logistic regression (ELR) model, as illustrated in Table 3, incorporates only significant threshold effects selected by the likelihood ratio test, thereby avoiding overfitting issues. Each coefficient is related to a specific variable or interaction effect, making it highly interpretable. Besides, Table 2 demonstrates that the ELR model substantially surpasses traditional logistic regression approaches, both with and without psychological variables, in model fit and predictive accuracy. This advancement suggests that the ELR model not only enhances the predictive performance of logistic regression, consistent with results in Dumitrescu et al. (2022), but also addresses the limitations of their study (see the discussion in Introduction). The outstanding performance of ELR also suggests that once the complex relationships between social and household contextual variables and evacuation decisions can be captured by the model, accurate predictions can be achieved based on these objective and easily-accessible variables, thereby reducing the reliance on psychological variables. Therefore, we can conclude that ELR has improved the timeliness and accuracy of the evacuation decision model without collecting householders' psychological variables.

High-stakes decision making, such as evacuation planning, necessitates accurate predictions to provide emergency managers with reliable evidence for evacuation zone division, risk assessment, and other relevant goals (Dash and Gladwin 2007). Additionally, transparency is crucial for

models or systems to gain trust and acceptance (Kim 2015). Consequently, in such applications, accurate interpretable machine learning models are probably preferred over the black-box counterparts, making ELR a promising technique. For example, ELR maintains high prediction capability and retains the transparency and intrinsic interpretability of the LR model. Thus, emergency managers can directly use ELR to forecast evacuation traffic demand and identify households requiring assistance during evacuations.

Simulation is also essential for high-stakes decision making, particularly in providing precautions for disaster-prone areas that lack experience of a disaster or are remote (Ramaguru and Pasupuleti 2019). To effectively generate and analyze simulations, supporting technologies must consider human behavior (Kim et al. 2004; Lee et al. 2004; Zhao et al. 2021). Artificial Intelligence has shown promise in enhancing traffic simulators by producing more realistic outputs (Zhao et al. 2021) and rapidly extracting useful knowledge (Sun et al. 2020). Therefore, leveraging ELR to assist the development of the next-generation AI-based evacuation simulation tools hold significant potential due to its high prediction accuracy, straightforward model formulation (Eq. 5), minimal need for model specification, and reliance only on readily simulated social and household contextual data.

This study offers new insights into the underlying relationships between social and household contextual variables and evacuation decisions. Specifically, we discovered nonlinear patterns in demographic variables and the diminishing marginal effects in resource-related variables. The effect of HHSize (household size) is statistically insignificant when treated as linear but attains significance after its univariate effect is detected. When the value of each resource-related variable (that is, RegVeh, EvaVeh, and EvaCost) exceeds the specific threshold, its marginal effect decreases. This implies that when available resources reach certain limits, their impacts on people's evacuation decisions become minimal.

7 Conclusion

In this study, we developed an interpretable machine learning approach—the enhanced logistic regression model (ELR)—to predict hurricane evacuation decisions. Empirical results confirm that ELR can capture and incorporate nonlinearities of demographic and resource-related variables and can better predict households' evacuation decisions than do psychological models. This study thus contributes to the protective action behavioral theories by showing that the predictors of household and social contexts including resource availability, social vulnerability, and social capital have significant effects (both linear and nonlinear) in predicting householders' evacuation decisions.

Admittedly, this study has limitations and future work can further explore the application of interpretable machine learning models in predicting hurricane evacuation decisions. For example, in our study, we used only one dataset to compare the performance of multiple models and test the validity of the proposed ELR. In future research, additional datasets should be employed to investigate the generalizability of the ELR model across various hurricane events. There is also a potential for further adapting the ELR model to apply to a broader range of disaster types, including tornadoes, floods, avalanches, wildfires, dam collapses, and other events that require rapid and massive evacuation to escape imminent threats. Additionally, the study assumed that vehicles and trailers are the primary means of evacuation for residents. However, other travel modes such as walking, biking, or boats may also be employed for evacuation in some countries and/or areas. Future research should aim to ensure that the ELR model is adaptable to these varied evacuation scenarios as well.

Moreover, the scope of this study is restricted to analyzing univariate and bivariate effects, with the latter being limited to distinct-category effects only. Future research could expand to include within-category bivariate threshold effects and multivariate threshold effects and non-pruned decision trees, like the classification and regression tree (CART) algorithm, could be employed to evaluate their impact on the model's predictive power and to investigate potential overfitting issues.

References

Ahmed, M.A., A.M. Sadri, and M. Hadi. 2020. Modeling social network influence on hurricane evacuation decision consistency and sharing capacity. *Transportation Research Interdisciplinary Perspectives* 7: Article 100180.

Alawadi, R., P. Murray-Tuite, D. Marasco, S. Ukkusuri, and Y. Ge. 2020. Determinants of full and partial household evacuation decision making in Hurricane Matthew. *Transportation Research Part D: Transport and Environment* 83: Article 102313.

Anyidoho, P.K., X. Ju, R.A. Davidson, and L.K. Nozick. 2023. A machine learning approach for predicting hurricane evacuee destination location using smartphone location data. *Computational Urban Science* 3(1): Article 30.

Bagloee, S.A., K.H. Johansson, and M. Asadi. 2019. A hybrid machine-learning and optimization method for contraflow design in post-disaster cases and traffic management scenarios. *Expert Systems with Applications* 124: 67–81.

Baker, E.J. 1991. Hurricane evacuation behavior. *International Journal of Mass Emergencies & Disasters* 9(2): 287–310.

Bhavan, A., P. Chauhan, and R.R. Shah. 2019. Bagged support vector machines for emotion recognition from speech. *Knowledge-Based Systems* 184: Article 104886.

Burris, J.W., R. Shrestha, B. Gautam, and B. Bista. 2015. Machine learning for the activation of contraflows during hurricane evacuation. In *Proceedings of the 2015 IEEE global humanitarian technology conference (GHTC)*, 8–11 October 2015, Seattle, WA, USA, 254–258.

Cahyanto, I., L. Pennington-Gray, B. Thapa, S. Srinivasan, J. Villegas, C. Matyas, and S. Kiousis. 2014. An empirical evaluation of the determinants of tourist's hurricane evacuation decision making. *Journal of Destination Marketing & Management* 2(4): 253–265.

Census Bureau, U.S. 2019. Coastline America. http://www.ncdc.noaa. gov/billions. Accessed 8 Dec 2023.

Dash, N., and H. Gladwin. 2007. Evacuation decision making and behavioral responses: Individual and household. *Natural Hazards Review* 8(3): 69–77.

Davidson, R.A., L.K. Nozick, T. Wachtendorf, B. Blanton, B. Colle, R.L. Kolar, S. DeYoung, and K.M. Dresback et al. 2020. An integrated scenario ensemble-based framework for hurricane evacuation modeling: Part 1 – Decision support system. *Risk Analysis* 40(1): 97–116.

Dillman, D.A. 2011. *Mail and Internet surveys: The tailored design method – 2007 update with new Internet, visual, and mixed-mode guide*. Hoboken, NJ: Wiley.

Dumitrescu, E., S. Hué, C. Hurlin, and S. Tokpavi. 2022. Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research* 297(3): 1178–1192.

Ersing, R.L., C. Pearce, J. Collins, M.E. Saunders, and A. Polen. 2020. Geophysical and social influences on evacuation decision-making: The case of hurricane Irma. *Atmosphere* 11(8): Article 851.

Fraser, T. 2022. Fleeing the unsustainable city: Soft policy and the dual effect of social capital in hurricane evacuation. *Sustainability Science* 17(5): 1995–2011.

Gladwin, H. 1997. Warning and evacuation: A night for hard houses. In *Hurricane Andrew: Ethnicity, gender and the sociology of disasters*, ed. W. Gillis Peacock, H. Gladwin, and B. Hearn Morrow, 52–74. London: Taylor & Francis.

Gladwin, H., J.K. Lazo, B.H. Morrow, W.G. Peacock, and H.E. Willoughby. 2007. Social science research needs for the hurricane forecast and warning system. *Natural Hazards Review* 8(3): 87–95.

Hasan, S., S. Ukkusuri, H. Gladwin, and P. Murray-Tuite. 2011. Behavioral model to understand household-level hurricane evacuation

decision making. *Journal of Transportation Engineering* 137(5): 341–348.

Horney, J.A., P.D.M. MacDonald, M. Van Willigen, P.R. Berke, and J.S. Kaufman. 2010. Factors associated with evacuation from Hurricane Isabel in North Carolina, 2003. *International Journal of Mass Emergencies & Disasters* 28(1): 33–58.

Huang, S.-K., M.K. Lindell, C.S. Prater, H.-C. Wu, and L.K. Siebeneck. 2012. Household evacuation decision making in response to Hurricane Ike. *Natural Hazards Review* 13(4): 283–296.

Huang, S.-K., M.K. Lindell, and C.S. Prater. 2016. Who leaves and who stays? A review and statistical meta-analysis of hurricane evacuation studies. *Environment and Behavior* 48(8): 991–1029.

Huang, S.-K., M.K. Lindell, and C.S. Prater. 2017. Multistage model of hurricane evacuation decision: Empirical study of Hurricanes Katrina and Rita. *Natural Hazards Review* 18(3): Article 05016008.

Kildow, J.T., C.S. Colgan, and J. Pat. 2016. State of the U.S. ocean and coastal economies 2016 update. https://cbe.miis.edu/noep_publications/18. Accessed 30 Dec 2023.

Kim, P.B. 2015. Interactive and interpretable machine learning models for human machine collaboration. Ph.D. thesis. Massachusetts Institute of Technology, Cambridge, MA, USA.

Kim, H., J.-H. Park, D. Lee, and Y.-S. Yang. 2004. Establishing the methodologies for human evacuation simulation in marine accidents. *Computers & Industrial Engineering* 46(4): 725–740.

Knabb, R.D., D.P. Brown, and J.R. Rhome. 2006. Tropical cyclone report: Hurricane Rita 18–26 September 2005. https://www.nhc.noaa.gov/data/tcr/AL182005_Rita.pdf. Accessed 30 Dec 2023.

Knabb, R.D., J.R. Rhome, and D.P. Brown. 2023. Tropical cyclone report: Hurricane Katrina, 23–30 August 2005. https://www.nhc.noaa.gov/data/tcr/AL122005_Katrina.pdf. Accessed 30 Dec 2023.

Lazo, J.K., A. Bostrom, R.E. Morss, J.L. Demuth, and H. Lazrus. 2015. Factors affecting hurricane evacuation intentions. *Risk Analysis* 35(10): 1837–1857.

Lee, D., J.-H. Park, and H. Kim. 2004. A study on experiment of human behavior for evacuation simulation. *Ocean Engineering* 31(8–9): 931–941.

Li, D., Z. Zhang, B. Alizadeh, Z. Zhang, N. Duffield, M.A. Meyer, C.M. Thompson, H. Gao, and A.H. Behzadan. 2023. A reinforcement learning-based routing algorithm for large street networks. *International Journal of Geographical Information Science*. https://doi.org/10.1080/13658816.2023.2279975.

Lindell, M.K. 2013. Evacuation planning, analysis, and management. In *Handbook of emergency response: A human factors and systems engineering approach*, ed. A.B. Badiru, and L. Racz, 121–149. Boca Raton, FL: CRC Press.

Lindell, M.K. 2018. Communicating imminent risk. In *Handbook of disaster research*, ed. H. Rodríguez, W. Donner, and J.E. Trainor, 449–477. Cham: Springer.

Lindell, M.K. 2019. *The Routledge handbook of urban disaster resilience: Integrating mitigation, preparedness, and recovery planning*. London: Routledge.

Lindell, M.K., and R.W. Perry. 2012. The protective action decision model: Theoretical modifications and additional evidence. *Risk Analysis* 32(4): 616–632.

Lindell, M.K., J.-C. Lu, and C.S. Prater. 2005. Household decision making and evacuation in response to Hurricane Lili. *Natural Hazards Review* 6(4): 171–179.

Ling, L., S.V. Ukkusuri, P. Murray-Tuite, S. Lee, and Y. Ge. 2022. Modeling the causality of received information, certainty, and decision making in hurricane evacuations. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4127811.

Lo, S.M., M. Liu, P.H. Zhang, and R.K. Yuen. 2009. An artificial neural-network based predictive model for pre-evacuation human response in domestic building fire. *Fire Technology* 45: 431–449.

Metaxa-Kakavouli, D., P. Maas, and D.P. Aldrich. 2018. How social ties influence hurricane evacuation behavior. In *Proceedings of the ACM on Human-Computer Interaction* 2(CSCW): Article 122.

Meyer, M.A., B. Mitchell, J.C. Purdum, K. Breen, and R.L. Iles. 2018. Previous hurricane evacuation decisions and future evacuation intentions among residents of southeast Louisiana. *International Journal of Disaster Risk Reduction* 31: 1231–1244.

Molnar, C. 2020. *Interpretable machine learning*. https://christophm.github.io/interpretable-ml-book/. Accessed 5 Feb 2024.

Morss, R.E., J.L. Demuth, J.K. Lazo, K. Dickinson, H. Lazrus, and B.H. Morrow. 2016. Understanding public hurricane evacuation decisions and responses to forecast and warning messages. *Weather and Forecasting* 31(2): 395–417.

Nara, A., S.G. Machiani, N. Luo, A. Ahmadi, K. Robinett, K. Tominaga, A. Ahmadi, and K. Robinett et al. 2021. Learning dependence relationships of evacuation decision making factors from Tweets. In *Empowering human dynamics research with social media and geospatial data analytics*, ed. A. Nara, and M.-H. Tsou, 113–138. Cham: Springer.

Petty, R.E., and J.T. Cacioppo. 1986. *The elaboration likelihood model of persuasion*. Cham: Springer.

Ramaguru, G.P., and V.D.K. Pasupuleti. 2019. Human evacuation simulation during disaster: A web tool. In *Innovative research in transportation infrastructure: Proceedings of ICIIF 2018*, ed. D. Deb, V.E. Balas, R. Dey, and J. Shah, 33–43. Singapore: Springer.

Roy, K.C., S. Hasan, A. Culotta, and N. Eluru. 2021. Predicting traffic demand during hurricane evacuation using real-time data from transportation systems and social media. *Transportation Research Part C: Emerging Technologies* 131: Article 103339.

Rudin, C. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1(5): 206–215.

Sadri, A.M., S.V. Ukkusuri, and H. Gladwin. 2017. The role of social networks and information sources on hurricane evacuation decision making. *Natural Hazards Review* 18(3): Article 04017005.

Sarwar, M.T., PCh. Anastasopoulos, S.V. Ukkusuri, P. Murray-Tuite, and F.L. Mannering. 2018. A statistical analysis of the dynamics of household hurricane-evacuation decisions. *Transportation* 45: 51–70.

Schorr, J.P. 2015. *Bridging the gap between social and transportation networks: An integrated, dynamic evacuation decision making model*. https://scholarspace.library.gwu.edu/concern/gw_etds/c534fn99d. Accessed 30 Dec 2023.

Sorensen, J.H. 2000. Hazard warning systems: Review of 20 years of progress. *Natural Hazards Review* 1(2): 119–125.

Sun, W., P. Bocchini, and B.D. Davison. 2020. Applications of artificial intelligence for disaster management. *Natural Hazards* 103(3): 2631–2689.

Tanim, S.H., B.M. Wiernik, S. Reader, and Y. Hu. 2022. Predictors of hurricane evacuation decisions: A meta-analysis. *Journal of Environmental Psychology* 79: Article 101742.

Vreugdenhil, B.J., N. Bellomo, and P.S. Townsend. 2015. Using crowd modelling in evacuation decision making. In *Proceedings of the ISCRAM 2015 conference*. https://idl.iscram.org/files/bjvreugdenhil/2015/1288_B.J.Vreugdenhil_etal2015.pdf. Accessed 2 Feb 2024.

Xu, N., R. Lovreglio, E.D. Kuligowski, T.J. Cova, D. Nilsson, and X. Zhao. 2023. Predicting and assessing wildfire evacuation decision-making using machine learning: Findings from the 2019 Kincade Fire. *Fire Technology* 59(2): 793–825.

Yang, K., R.A. Davidson, B. Blanton, B. Colle, K. Dresback, R. Kolar, L.K. Nozick, J. Trivedi, and T. Wachtendorf. 2019. Hurricane evacuations in the face of uncertainty: Use of integrated models to support robust, adaptive, and repeated decision-making. *International Journal of Disaster Risk Reduction* 36: Article 101093.

Yang, J., R. Shi, and B. Ni. 2021. MedMNIST classification decathlon: A lightweight AutoML benchmark for medical image analysis. In *Proceedings of the 2021 IEEE 18th International symposium*

on biomedical imaging (ISBI), 13–16 April 2021, Nice, France, 191–195.

Zhao, X., R. Lovreglio, E. Kuligowski, and D. Nilsson. 2021. Using artificial intelligence for safe and effective wildfire evacuations. *Fire Technology* 57: 483–485.

Zhao, X., R. Lovreglio, and D. Nilsson. 2020. Modelling and interpreting pre-evacuation decision-making using machine learning. *Automation in Construction* 113: Article 103140.

Zhu, R., B. Becerik-Gerber, J. Lin, and N. Li. 2023. Behavioral, data-driven, agent-based evacuation simulation for building safety design using machine learning and discrete choice models. *Advanced Engineering Informatics* 55: Article 101827.