International Journal of Digital Earth

ISSN: 1753-8947 (Print) 1753-8955 (Online) Journal homepage: www.tandfonline.com/journals/tjde20

Breaking the black box: an interpretable machine learning model for global terrorism forecasting

Xiang Zhang , Xin Qiu , Chenglong Yu , Ziheng Chen , Geyuan Zhu , Liangyang Dai , Qingfeng Guan & Yao Yao

To cite this article: Xiang Zhang , Xin Qiu , Chenglong Yu , Ziheng Chen , Geyuan Zhu , Liangyang Dai , Qingfeng Guan & Yao Yao (2026) Breaking the black box: an interpretable machine learning model for global terrorism forecasting, International Journal of Digital Earth, 19:1, 2687356, DOI: 10.1080/17538947.2026.2687356

To link to this article: https://doi.org/10.1080/17538947.2026.2687356

Published online: 12 Jun 2026.

INTERNATIONAL JOURNAL OF DIGITAL EARTH
2026, VOL. 19, NO. 1, 2687356
https://doi.org/10.1080/17538947.2026.2687356

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Breaking the black box: an interpretable machine learning model for global terrorism forecasting

Xiang Zhang[a,b,1], Xin Qiu[a,1], Chenglong Yu[a,b], Ziheng Chen[c], Geyuan Zhu[a,b], Liangyang Dai[d], Qingfeng Guan[a] and Yao Yao[a,b,e,f]

[a]UrbanComp Lab, School of Geography and Information Engineering, China University of Geosciences, Wuhan, Hubei, People's Republic of China; [b]LocationMind Institution, LocationMind Inc., Chiyoda, Tokyo, Japan; [c]Institute of Seismology, China Earthquake Administration, Wuhan, People's Republic of China; [d]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, Hubei, People's Republic of China; [e]Hitotsubashi Institute for Advanced Study, Hitotsubashi University, Kunitachi, Tokyo, Japan; [f]Faculty of Engineering, Reitaku University, Kashiwa, Chiba, Japan

ABSTRACT

Terrorist attacks significantly threaten a nation's stability, prosperity, and social cohesion. Therefore, predicting terrorist attacks and identifying their underlying drivers are crucial for formulating effective counterterrorism strategies. Existing studies often prioritize either temporal or spatial dimensions, while their interplay and specific socioeconomic drivers are less explored. In this study, global news data are leveraged to construct a novel global conflict index (GCI), which integrates multisource datasets to comprehensively characterize the key drivers of terrorist attacks. TerrorXG is proposed to predict terrorist attacks, and SHAP analysis is applied to quantitatively interpret the importance and contributions of the driving factors. TerrorXG demonstrated superior performance (RMSE: 0.319; PCC: 0.777) and high computational efficiency. Compared with the second most influential factor (population size), the proposed GCI has a 42.4% greater impact on terrorist attacks. The interpretability analysis of the model highlights socioeconomic inequality as a primary determinant: the impacts of child malnutrition and infant mortality are 38.4% to 108.5% greater than the effect of urbanization. The influence of ethnicity represents only 9.7% of the impact of the GCI, providing empirical evidence that challenges traditional theoretical perspectives on ethnic conflict in terrorism research. This study provides valuable insights for optimizing the allocation of counterterrorism resources.

ARTICLE HISTORY
Received 3 February 2026
Accepted 7 June 2026

KEYWORDS
Terrorist attack prediction; spatiotemporal data mining; XGBoost; driving factors

1. Introduction

Terrorist attacks pose major threats to global political security, social stability, and economic development (Cai et al. 2024; Meng and Ghafoori 2024; Nabin et al. 2022). However, challenges such as the lack of a unified counterterrorism strategy (Asongu, Le Roux, and Singh 2021; Karlsrud 2017) and geopolitical instability (Ackerman and Peterson 2020) have exacerbated global terrorist threats. The global counterterrorism situation remains challenging. Terrorist attacks in several regions have precipitated military conflicts and humanitarian crises (Kiyani et al. 2025). Specifically, the Israel–Hamas conflict has resulted in more than 69,185 reported fatalities and estimated reconstruction costs of $53 billion (Haghani et al. 2022), while the 2025 New Orleans incident resulted in in the death of 15 people and contributed to a 280% increase in terrorism-related deaths in the West. Furthermore, East Turkestan forces orchestrated thousands of attacks between 1990 and 2016. Therefore, advancing research on terrorism is essential for developing effective prevention strategies.

Existing studies predominantly employ qualitative analysis methods owing to, the lack of terrorist attack datasets and the limitations of current quantitative methods. However, qualitative approaches struggle to reveal the underlying causes of terrorist attacks and the associated patterns (Schuurman 2020). A

CONTACT Prof. Dr. Yao Yao yaoy@cug.edu.cn UrbanComp Lab, School of Geography and Information Engineering, China University of Geosciences, Wuhan, Hubei, People's Republic of China
[1]These authors contributed equally to this work.

quantitative assessment of terrorist attack risk levels and occurrence trends is crucial for national security (Zhao, Liu, and Wang 2024). The advent of open-source databases, particularly the Global Terrorism Database (GTD) (LaFree and Dugan 2007), has enabled a transition toward the use of quantitative approaches in terrorism research. Open-source databases are essential for extracting and analysing terrorist attack information. Scholars predominantly use the GTD for geographical statistical information and analyses of terrorist attacks (Wang et al. 2021). For instance, Chen et al. (2020) utilised the GTD to analyse the spatiotemporal characteristics of terrorist attacks in the Middle East, demonstrating the potential of the GTD for information mining in quantitative studies of terrorist attacks. Despite the considerable contributions of open-source databases to terrorism research, their reliance on single-source data limits the comprehensive identification and analysis of underlying driving factors.

Terrorist attacks are significantly influenced by social, ethnic, and natural factors. For instance, income inequality has been shown to increase elevate the likelihood of terrorist attacks (Krieger and Meierrieks 2019). In terms of ethnic factors, characteristics such as ethnic diversity and imbalanced power distributions among ethnic groups have been linked to armed conflict (Weidmann, Rød, and Cederman 2010). In terms of natural factors, extreme rainfall and climate change increase the probability of terrorist attacks (Ge et al. 2022). News data, which encompass a broad spectrum of information, can partially reflect the multiple factors underlying terrorist attacks (Koutidis et al. 2024). The Global Database of Events, Language, and Tone (GDELT) (Leetaru and Schrodt 2013), which is characterised by its extensive temporal coverage and real-time updates, has considerable potential for investigating the determinants of terrorist attacks (Galla and Burke 2018; Shen et al. 2020). For example, Qin, Luo, and Yao (2019) utilised the GDELT to construct an interactive grid of national conflict events, revealing that variations in the grid data correspond to the occurrence of national conflicts. The observed correlation suggests that national conflict indicators extracted from news data can be leveraged for event prediction. Similarly, Voukelatou et al. (2022) used the GDELT to predict the global peace index for six months, demonstrating the feasibility of using news data in terrorism-related studies. Therefore, extracting quantifiable, relevant features from the GDELT and integrating them with multisource data to construct an enhanced terrorist attack dataset can provide critical support for increasing the accuracy of terrorist attack prediction.

The prediction of terrorist attacks is critical for developing effective prevention strategies. Terrorist attacks are driven by various factors, with highly complex interactions. Predicting such human-driven phenomena presents inherent challenges because of their profound complexity and dynamic nature across both spatial and temporal dimensions (Al-Sabbagh et al. 2026). Unlike natural disasters, which are governed by physical laws, terrorist activities are the product of strategic human agency and are influenced by a volatile interplay of political grievances, socioeconomic disparities, and shifting ideologies. Spatially, these incidents exhibit high nonstationarity (Zhang et al. 2026) and often cluster on the basis of porous borders or localised geopolitical instability. Temporally, the patterns are nonlinear and adaptive, as terrorist organisations frequently evolve their tactics in response to counterterrorism interventions.

Conventional mathematical and statistical methods often struggle to accurately predict the dynamic evolution of terrorist attacks (Alonso 2023). Machine learning techniques have been increasingly applied in terrorism studies (Murphy, Sharpe, and Huang 2024). Machine learning enables the rapid processing of large datasets, facilitates the assessment and quantification of multiple influencing factors, and has the potential to increase the accuracy of terrorist attack predictions. For instance, Mo et al. (2017) conducted a comparative analysis using various machine learning methods and verified the feasibility of applying machine learning approaches in terrorism research. Additionally, previous studies have demonstrated the feasibility of predicting terrorist attacks from either a temporal or spatial perspective. From a temporal standpoint, Luo et al. (2020) proposed a long short-term memory (LSTM) neural network model for the short-term prediction of terrorist event timing in Iraq. From a spatial perspective, Huamaní, Alicia, and Roman-Gonzalez (2020) employed various artificial intelligence techniques to identify regions prone to terrorist attacks worldwide. However, existing studies on terrorist attack prediction are predominantly limited to either temporal analysis or spatial modelling. This gap reflects a limited understanding of terrorist attack mechanisms, highlighting the need for a spatiotemporal approach (Al-Sabbagh et al. 2026). Consequently, effectively integrating both temporal and spatial information remains a critical challenge and requires urgent attention.

Although machine learning models have yielded promising results in terrorist attack studies (He et al. 2023; Liu et al. 2018), they often require substantial amounts of memory and exhibit slow training times (Shwartz-Ziv and Armon 2022). The XGBoost model, a decision tree-based ensemble learning algorithm, is computationally efficient and enables assessment of feature importance, thereby facilitating explanatory analysis. This model is particularly well-suited for processing large-scale data (Sagi and Rokach 2021). For instance, Atoum et al. (2025) compared multiple datasets and machine learning models for terrorist attack prediction and reported that the XGBoost model outperformed other models, demonstrating superior adaptability and robustness across diverse feature sets. Understanding the driving factors of terrorist attacks is crucial for developing targeted prevention strategies. However, machine learning models often act as 'black boxes,' obscuring their decision-making processes (Li 2022). Further explanation requires explanatory models. The Shapley additive explanations (SHAP) model (Lundberg and Lee 2017) increases interpretability by quantifying each factor's contribution to individual predictions. For example, Wang et al. (2022) applied the SHAP model to analyse the impact of sewage treatment processes, demonstrating its effectiveness in increasing the interpretability of the results of machine learning models. Moreover, studies utilising the SHAP method to explain the results of machine learning models have revealed spatial effects, such as spatial autocorrelations and heterogeneity, among various factors (Li 2022). These findings underscore the ability of the SHAP method capacity to significantly improve model interpretability (Van den Broeck et al. 2022). By providing clearer insights into model predictions, SHAP analysis enhances the ability of policy-makers to address complex issues such as crime and terrorism.

In summary, existing studies on terrorist attack prediction have predominantly focused on either temporal or spatial scales separately. Furthermore, the reliance on single data sources limits the comprehensive identification and assessment of the factors underlying terrorist attacks. Thus, two key challenges remain in terrorist attack research: effectively coupling temporal and spatial scales for prediction and analysing the underlying factors using multisource data. To address these issues, in this study, open-source databases (GDELT and GTD) are integrated with multisource data to construct a novel composite dataset. We developed TerrorXG, a spatial-temporal terrorist attack prediction model based on XGBoost, which forecasts the global distribution of terrorism incidents and facilitates international terrorism risk assessment. Then, the SHAP was utilised to analyse the interpretability of the results of the TerrorXG model, revealing the influence of various driving factors on the likelihood of a predicted attack.

2. Data description and preprocessing

Multidimensional data encompassing social, ethnic, and environmental factors were collected to construct a dataset supporting terrorist attack prediction and the identification of driving factors. Table 1 provides an overview of these data sources and their descriptions.

Terrorist attack events serve as the foundation of the study of terrorism. The data were obtained from the Global Terrorism Database (GTD), which records detailed information on global terrorist attacks from

Table 1. Description of the data.

Data name	Datasets	Data source
News events	Global Terrorism Database (GTD)	https://www.start.umd.edu/gtd/
Event geographic coordinates		
Country/region of news events		
Goldstein scale		
Number of event mentions		
Spatial distribution of ethnic groups	Ethnic Power Relations Dataset Geo-EPR 2021	https://icr.ethz.ch/data/epr/geoepr/
Terrorist attack dates	Global Database of Events, Language, and Tone (GDELT)	https://www.gdeltproject.org/
Latitude and longitude of terrorist attacks		
…		
Number of people kidnapped		
Gross domestic product	Peace Research Institute Oslo (PRIO-GRID)	https://grid.prio.org/#/
Nighttime lights intensity		
Population		
Annual precipitation		
Percentage of mountainous areas		
…		
Presence of drug cultivation		

1970 to the present. To ensure a representative temporal distribution, data were collected at five-year intervals from 1990 to 2010, covering five time points. The decision to utilise a five-year temporal resolution is a crucial methodological choice. This interval helps mitigate the inherent spatiotemporal sparsity and class imbalance present in terrorism data by aggregating events. Additionally, it ensures that adequate variance is captured from slow-moving drivers, such as socioeconomic indicators. This approach effectively prevents model overfitting from low-variance predictors and yields a more robust feature set for both model training and the subsequent SHAP analysis.

Extracting features from multisource data is essential for comprehensive and accurate terrorist attack prediction. In this study, multidimensional features were extracted from the PRIO-GRID global grid system (Tollefsen et al. 2016) ($0.5° \times 0.5°$ decimal degree resolution), including economic indicators such as the gross cell product measured on the basis of market exchange rates and differences in purchasing power, and subnational poverty proxies such as infant mortality rates and child malnutrition rates. The social and urbanisation variables include population size from the GPW and HYDE and urbanisation levels derived from urban land cover percentages. Environmental and resource factors include annual precipitation, temperature, drought severity, mountainous terrain proportions, and the presence of petroleum, diamond, gold, gemstones, or large-scale drug cultivation. Furthermore, spatial accessibility metrics, such as distance to the national capital and distances to the nearest land borders, were utilised to increase the model's predictive capacity.

National conflict feature data were sourced from the GDELT, a global news database, covering the same years as the GTD. Entries without event location attributes were removed from this dataset. Ethnic data were obtained from the Georeferencing Ethnic Power Relations (Geo-EPR) global database. Since the collection of terrorist attack data began in 1990, ethnic data from before 1990 were removed. Additionally, ethnic regions that fully encompassed other regions were excluded. The PRIO-GRID geographic grid system was then used to partition the filtered ethnic data and compute the number of ethnic groups in each grid cell. The number of ethnic groups in each grid cell served as the ethnic mixing index.

3. Methodology

The methodology of this study comprises three steps: (1) Dataset construction. In this step, the terrorist attack index and national conflict intensity were quantified using data from the GTD and GDELT, after which the ethnic mixing index and PRIO-GRID geographic grid data were integrated to form the final dataset. (2) Terrorist attack prediction model development. In this step, the TerrorXG spatiotemporal prediction model for terrorist attacks was constructed on the basis of the XGBoost algorithm. Model accuracy was increased by optimising the sample ratio and applying regularisation techniques. The model's prediction performance was then analysed and evaluated. (3) Analysis of the driving factors. The SHAP model was employed to interpret the predictions of TerrorXG, enabling a quantitative analysis of the key driving factors. This analysis provides a scientific basis for counterterrorism policy decisions. The technical approach is shown in Figure 1.

3.1. Dataset construction

3.1.1. Derivation of the terrorism index from the GTD

The severity of terrorist attacks events varies significantly, yet previous research has often overlooked a classification system based on this factor. To address this gap, the calculation methodology of the global terrorism index (GTI) proposed by the Institute for Economics & Peace (2024) was adopted and modified. Recognising the heterogeneity of terrorist events, we calculate the severity of each event and propose the GTDIndex as a quantitative measure of attack severity. The modified calculation method is presented in Equation (1).

$$GTDindex = Nkill + Nwound \times 0.5 + Nhostkid \times 0.3 + Property + AttackType \tag{1}$$

Nkill represents the number of victims of the event; *Nwound* denotes the number of individuals injured; *Nhostkid* indicates the number of hostages taken; *Property* is scored on the basis of whether property

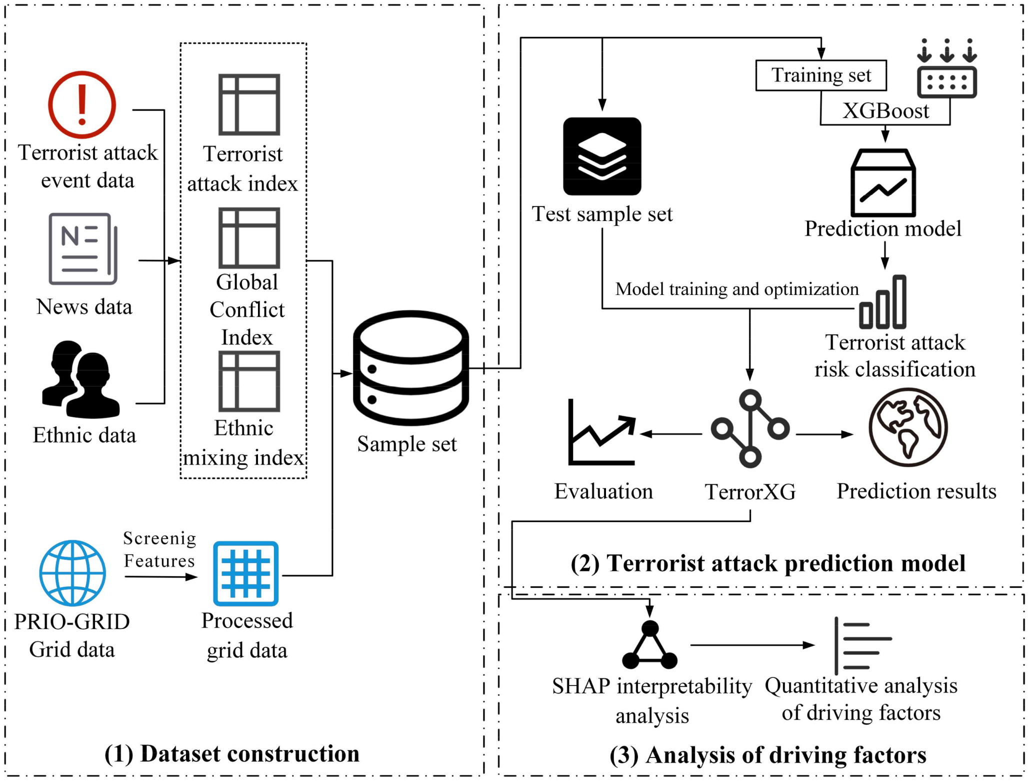

Figure 1. Technology roadmap: (1) Dataset construction. (2) Development of the terrorist attack prediction model. (3) Analysis of driving factors.

Table 2. Scoring table for types of terrorist attacks.

Types of terrorist attacks	Score
Assassination -1, Hijacking -2, Kidnapping -3	1.5
Barricade Incident -4, Bombing/Explosion -5, Armed Assault -6	1
Unarmed Assault -7, Facility/Infrastructure Attack -8, Unknown -9	0.5

damage occurred, with 'yes' assigned a score of 1 and 'no' assigned a score of 0; and *AttackType* is assigned on basis of the type of terrorist event, as shown in Table 2, where types 1, 2, and 3 are assigned scores of 1.5; types 4, 5, and 6 are assigned scores of 1; and types 7, 8, and 9 are assigned scores of 0.5.

3.1.2. Derivation of the national conflict index from the GDELT

Given the uneven spatial distribution of global conflict news and significant regional disparities, Qi et al. (2021) proposed the global conflict index (GCI) as a quantitative measure of conflict intensity on the basis of news data. In this study, the GCI calculation methodology is adopted as a reference. The calculation process involves assigning values to relevant news events to account for inherent differences. The GCI calculation method, as proposed by Qi et al. (2021), is provided in Equation (2).

$$GCI = \frac{\sum_{i=1}^{N} ABS(GS_i)}{N} \times \frac{\sum_{i=1}^{N} MT_i}{\sum_{i=1}^{M} MT_i} \tag{2}$$

In this equation, $ABS(GS_i)$ represents the absolute value of the Goldstein scale score for each conflict event; N represents the total number of conflict news events in a country/region; MT_i represents the number times that event i was mentioned; and M represents the total number of all news events in a country/region.

3.1.3. Geographical gridding and model training

Geographical gridding is an effective strategy for reducing spatial uncertainty within a specified scale range (Wan and Cao 2016). Furthermore, this approach enables the allocation of statistical values to geographic grids, thereby facilitating the integration and analysis of diverse datasets (Zhang, Sun, and Wang 2015). To address the challenge of integrating spatiotemporal data, we constructed a panel dataset by structuring the information for each $0.5° \times 0.5°$ grid cell across multiple time points. The temporal resolution was set at five-year intervals from 1990 to 2010. This interval was chosen to capture significant long-term shifts in terrorism trends and their underlying drivers, while mitigating the influence of short-term, transient fluctuations that might obscure more stable, structural relationships. The $0.5° \times 0.5°$ spatial grid resolution was selected to align directly with the PRIO-GRID dataset, which serves as a primary source for social and environmental variables, thus ensuring seamless data integration without the need for complex spatial resampling.

The TerrorXG model learns from this spatiotemporal structure by treating each grid-time unit as a distinct sample. As shown in Figure 2, for each grid cell at a specific time point, a feature vector is compiled, containing the GCI, ethnic heterogeneity metrics (EPR), and various PRIO-GRID variables for that precise location and period. The model is trained to predict the corresponding terrorism index (GTDIndex). This framework allows the model to implicitly learn the spatiotemporal interactions. For example, the model can identify how a change in a predictor variable such as the GCI in a specific grid cell at one time point influences the risk of a terrorist attack in the same cell at a subsequent time point. The corresponding formula is presented in Equation (3).

$$GRID(x, y, t) = (GCI(x, y, t), GTDindex(x, y, t), EPR(x, y, t), PRIO(x, y, t)) \tag{3}$$

$GRID(x, y, t)$ denotes each grid cell; $GCI(x, y, t)$, $GTDindex(x, y, t)$, $EPR(x, y, t)$ and $PRIO(x, y, t)$ represent the national conflict index, terrorism index, ethnic heterogeneity index, and PRIO-GRID data variables for that grid cell, respectively. Additionally, (x, y) are the coordinates within the grid, and t denotes the time of the data.

3.2. XGBoost-based terrorist attack prediction model

3.2.1. Construction of the TerrorXG model based on XGBoost

The TerrorXG terrorist attack prediction model is based on XGBoost, a widely utilised ensemble learning method known for its high predictive accuracy, strong interpretability, and computational efficiency

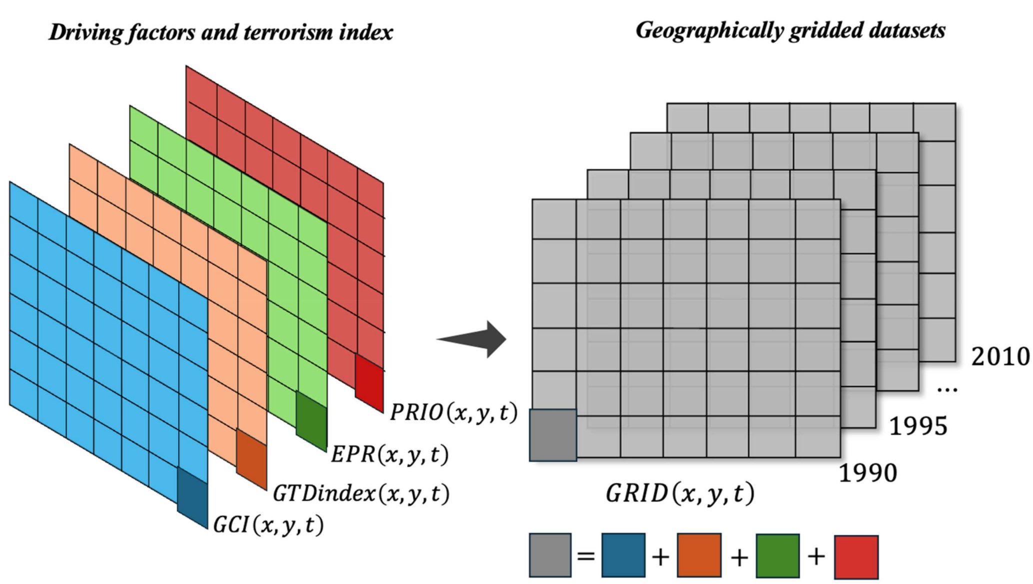

Figure 2. Geospatial data gridding.

(Chen and Guestrin 2016). In the context of terrorist attack prediction, the model's accuracy is increased by minimising the objective function (Ma et al. 2021). Importantly, the prediction task in this study is not to forecast the specific time and location of future terrorist incidents. Instead, the task involves a spatiotemporal risk assessment aimed at evaluating the relative risk levels of terrorist attacks occurring in various regions globally within a specific year. The objective is to identify geographical areas with a greater threat of terrorism, thereby providing a basis for resource optimisation and strategic development. The specific steps are as follows:

The formulation of the objective function follows the standard XGBoost framework proposed by Chen and Guestrin (2016). The objective function consists of a loss function and a regularisation term, as shown in Equation (4).

$$l(E) = \sum_{k=1}^{K} l(A_i, \hat{A}_i^{\,k-1} + f_k(B_i)) + \Omega(f_k) \tag{4}$$

In Equation (4), $l(E)$ denotes the performance of the terrorist attack prediction model, while $\Omega(f_k)$ represents the regularisation term. The expression $\sum_{k=1}^{K} l(A_i, \hat{A}_i^{\,k-1} + f_k(B_i))$ serves as the loss function, which is used to evaluate the accuracy of the prediction model. The loss function for the (k-1)th iteration is provided in Equation (5).

$$l(A_i, \hat{A}_i^{\,k-1} + f_k(B_i)) \tag{5}$$

$\hat{A}_i^{\,k-1}$ is the predicted value from the (k-1)th iteration; $f_k(B_i)$ represents the prediction result of the k-th tree for different influencing factors B_i. $\Omega(f_k)$ represents the regularisation term, which is used to control the complexity of the model, as detailed in Equation (6).

$$\Omega(f_k) = \mu S + \frac{1}{2}\lambda \sum_{j=1}^{S} \omega_j^2 + \gamma \sum_{j=1}^{S} \omega_j \tag{6}$$

In Equation (6), S represents the number of leaves in the tree; ω_j is the weight of the j-th leaf in the current tree, corresponding to the weight of the influencing factor; and μ, λ and γ are regularisation parameters.

The final predicted value for the i-th sample, $\hat{y}_i$, is obtained by iteratively minimising the objective function through multiple stages, as defined in Equation (7).

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in \Theta \tag{7}$$

In Equation (7), x represents the factors influencing terrorist attacks; f_k denotes the k-th regression tree; K is the number of regression trees; and Θ is the space of all possible regression trees.

3.2.2. TerrorXG model training and optimisation

To train and evaluate the TerrorXG model, we constructed distinct training and test sets. The 2010 terrorist attack data from the Global Terrorism Database (GTD) served as the test set to assess the model's generalisability. Data imbalance represented a significant challenge, stemming from the scarcity of nonzero GTDIndex samples, which mainly affected the training data. To address this issue, we selected a random subset of zero-value samples and combined them with all available nonzero samples to create the final training dataset. Furthermore, regularisation techniques were employed during model training to mitigate overfitting and increase generalisability and accuracy.

3.3. Model evaluation

The coefficient of determination (R^2), root mean square error (RMSE), and Pearson correlation coefficient (PCC) were employed to assess the predictive performance of the TerrorXG model. The R^2 value reflects the proportion of variance in the dependent variable explained by the independent variables. The RMSE is commonly used to measure the deviation between the predicted values and the actual observed values. The PCC measures the strength of the linear correlation between the predicted and actual values. The evaluation metrics follow the standard definitions widely used in regression analysis (Hastie, Friedman, and Tibshirani 2001). The evaluation metrics for assessing model performance are summarised in Equations (8) to (10).

$$R^2 = \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{N} (\hat{y}_i - \overline{y}_i)^2} \tag{8}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2} \tag{9}$$

$$r_{xy} = \frac{Cov(X, Y)}{\sqrt{Std(X)Std(Y)}} \tag{10}$$

In these equations, N represents the total number of target value samples in the test set; y_i denotes the target value of the test set; $\hat{y}_i$ is the predicted target value of the test set; $\overline{y}_i$ is the mean target value of the test set; $Cov(X, Y)$ refers to the covariance between variables X and Y; and $Std(X)$ and $Std(Y)$ are the standard deviations of the variables X and Y, respectively.

3.4. Analysing the drivers of terrorist attacks

The SHAP method was employed to interpret the results of the TerrorXG model and identify the driving factors of terrorist attacks. SHAP analysis is a framework for interpreting the outputs of machine learning models (Lundberg and Lee 2017); this approach involves, calculating Shapley values for each prediction to quantify the influence of each feature. The calculation of Shapley values is rooted in game theory and is typically implemented using the Monte Carlo sampling method (Rubinstein and Kroese 2016). The Shapley value formulation follows the SHAP framework proposed by Lundberg and Lee (2017), as detailed in Equation (11).

$$\varepsilon_i = \sum_{s \in S_i} \frac{(n - |s|)!(|s| - 1)!}{n!} [v(s) - v(s \setminus i)] \tag{11}$$

Here, ε_i is the Shapley value for member i; n is the number of influencing factors for a terrorist attack; S_i is the set of all subsets containing member i; S denotes a specific coalition of factors belonging to S_i; $v(s)$ is the output of the model given the i-th variable and all other spatiotemporal data variables; and $v(s \setminus i)$ is the model's output after removing member i from set S and rearranging the other variables.

The model unifies various additive feature attribution methods (Lundberg and Lee 2017). With the SHAP method, the predicted value of the model is the sum of the SHAP values for each input feature, as shown in Equation (12).

$$y_i = y_{base} + f(x_{i,1}) + f(x_{i,2}) + \cdots + f(x_{i,j}) + \cdots + f(x_{i,k}) \tag{12}$$

In this equation, y_i represents the probability of a terrorist attack occurring for the i-th sample; y_{base} is the average prediction across all samples; x_i denotes the i-th sample; $f(x_{i,j})$ is the SHAP value of the j-th feature of the i-th sample; and k represents the number of input features.

4. Results

4.1. Prediction performance analysis

The constructed dataset of terrorist attack features was used for model training and validation. The dataset was divided into training (80%) and validation (20%) sets. The dataset contained numerous samples with a GTDIndex of zero. On the basis of the experimental results, with the aim of achieving the optimal model

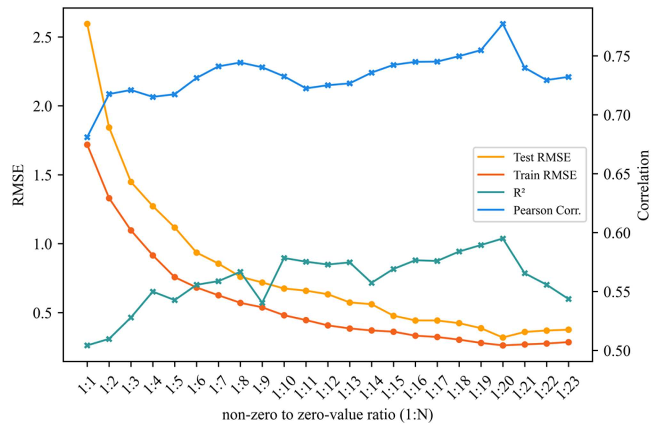

Figure 3. Model performance under different proportions of nonzero to zero values.

Table 3. Performance of common regression algorithms.

Model	RMSE(test set)	RMSE (training set)	R^2 (test set)	PCC (test set)	RMSE (test set 2010)	PCC (test set 2010)	Training time(s)
Linear	0.644	0.671	0.182	0.426	0.346	0.179	1.854
GAM	0.590	0.556	0.305	0.554	0.541	0.237	92.342
KNN	0.419	0.283	0.468	0.691	1.079	0.252	292.895
SVM	0.750	0.777	0.048	0.247	0.207	0.132	1228.571
BSM	0.357	0.384	0.521	0.701	0.557	0.216	1534.154
LightGBM	0.390	0.370	0.505	0.714	0.301	0.388	4.510
Random Forest	**0.224**	**0.043**	**0.716**	**0.847**	1.433	0.217	102.781
AdaBoost	3.031	3.039	-2.846	0.374	3.243	0.181	18.977
TerrorXG	0.319	0.262	0.595	0.777	**0.489**	**0.272**	**5.382**

accuracy, the final training data were constructed with a 1:20 ratio of nonzero to zerovalue GTDIndex samples, and the results are shown in Figure 3.

To objectively evaluate the performance of the TerrorXG model in predicting terrorist attacks, eight representative regression algorithms were selected as baseline models, including traditional statistical models (e.g. Linear, GAM), nonparametric models (KNN, SVM), and contemporary ensemble learning algorithms (e.g. Random Forest, LightGBM, AdaBoost). By comparing the predictive accuracy, generalisability, and computational efficiency of these models on a standardised dataset, we aim to verify the superiority of the TerrorXG model in processing complex global spatiotemporal data. The comparative results are summarised in Table 3.

The results show that the TerrorXG model achieved an R^2 value greater than 0.5, an RMSE value lower than 0.3, and a PCC value exceeding 0.7. These metrics collectively suggest that the TerrorXG model exhibits strong predictive performance. Furthermore, the strong agreement between the model predictions and the actual observed values underscores the model's potential usefulness in predicting terrorist attacks.

Among the evaluated machine learning models, the TerrorXG model generally outperformed all the other models except the random forest model. However, in the generalisability validation conducted with the 2010 dataset, the random forest model yielded an RMSE value that was 193% greater than that of the TerrorXG model. However, compared with the TerrorXG model, the random forest model yielded a 20.2% lower PCC and required 95% more training time. The results suggest that the random forest model had comparatively weaker generalisability than the TerrorXG model, demonstrating the superiority of TerrorXG in this context. Although the LightGBM model demonstrated competitive performance on the 2010 test set, its overall accuracy remained lower than that of TerrorXG.

The terrorist attacks predicted by the TerrorXG model, which are primarily concentrated in Central America, Southern Europe, Central and Southern Africa, the Middle East, Southwest China, and India, are shown in Figure 4. A comparative visualisation of the prediction outcomes among the different models is provided in Figure 5. Panels A to E present comparisons between the actual data and model prediction results for Central America, Europe, the Middle East and the Arabian Peninsula, South Asia, and southern

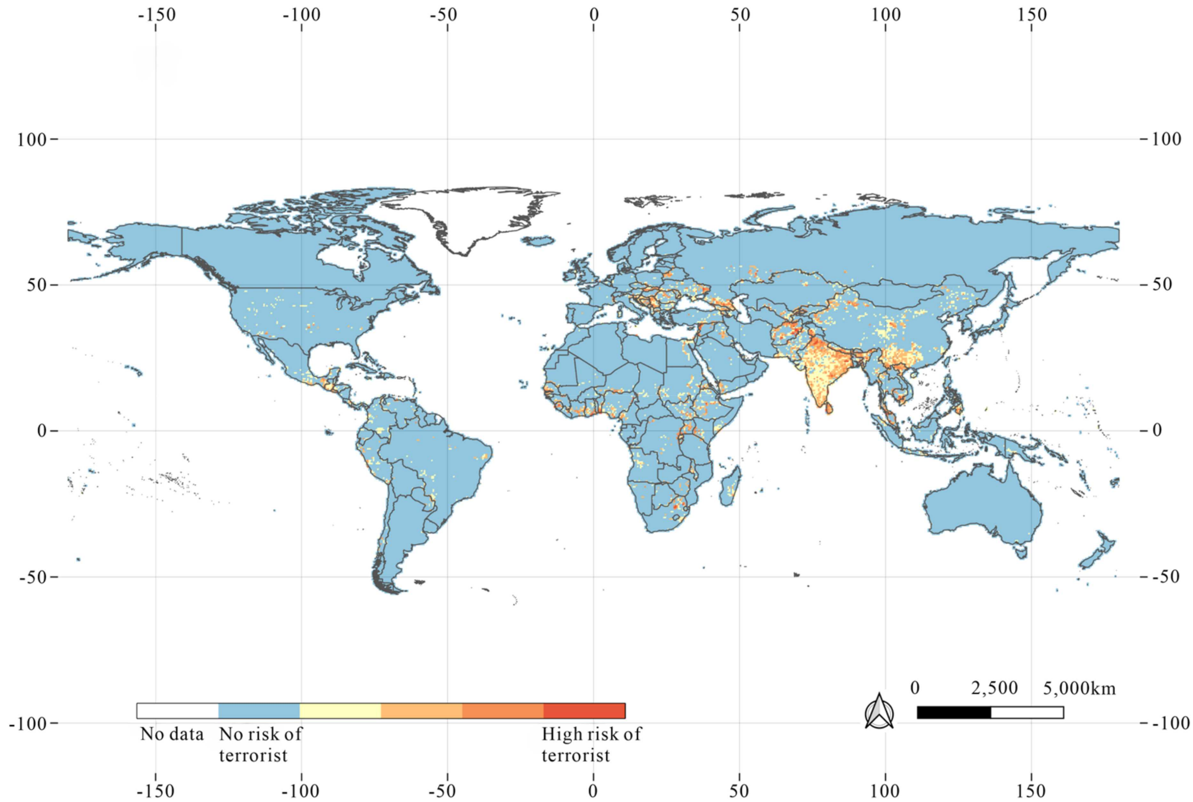

Figure 4. Prediction results of the TerrorXG model.

China combined with the Indo-China Peninsula. A qualitative comparison suggests that compared with the predictions of the other models, the predictions of the TerrorXG model align more closely with the actual observations. In comparison, the random forest model tends to predict more extensive high-risk areas in the Indian subcontinent and Southwest China than those reflected by actual observations, which might indicate potential overfitting. These observations support the strong generalisability of the TerrorXG model.

To further analyse the prediction accuracy, the model's predictions of major terrorist attacks were compared with actual incidents that were recorded in the original 2010 data (Figure 6). The predicted hotspots closely match the spatial distribution of incidents observed in the original data. In the Panama Canal region, the model overestimates the extent of low-risk areas compared with the actual observations. Similarly, in southern European countries bordering the Adriatic Sea and in northwestern Belarus, the model identifies additional medium-risk areas that differ from the actual data. Along Ukraine's eastern border adjacent to Russia, the medium-risk regions predicted by the model are not clearly supported by the observed incident patterns. Additionally, in the border region between Russia and the Middle East, the prevalence of low- and medium-risk areas is overestimated relative to the observed data. For the northern border of India and southern Sri Lanka, the model predicts additional medium-risk regions that are not reflected clearly in the actual data. A similar discrepancy is observed for central India, where the model identifies more low-risk areas than indicated by the observed incidents. Similarly, in Southwest China, compared with the actual observations, the model predicts a greater presence of low- and medium-risk areas.

Overall, the terrorist attack risk areas predicted by the model tend to align with national borders or coastlines, suggesting possible artifacts arising from geopolitical boundary influences or regional aggregation effects. Considering these discrepancies, we recommend prioritising future monitoring efforts and targeted counterterrorism strategies in regions such as the Panama Canal area, southern Europe, the Indian border area, and Southwest China.

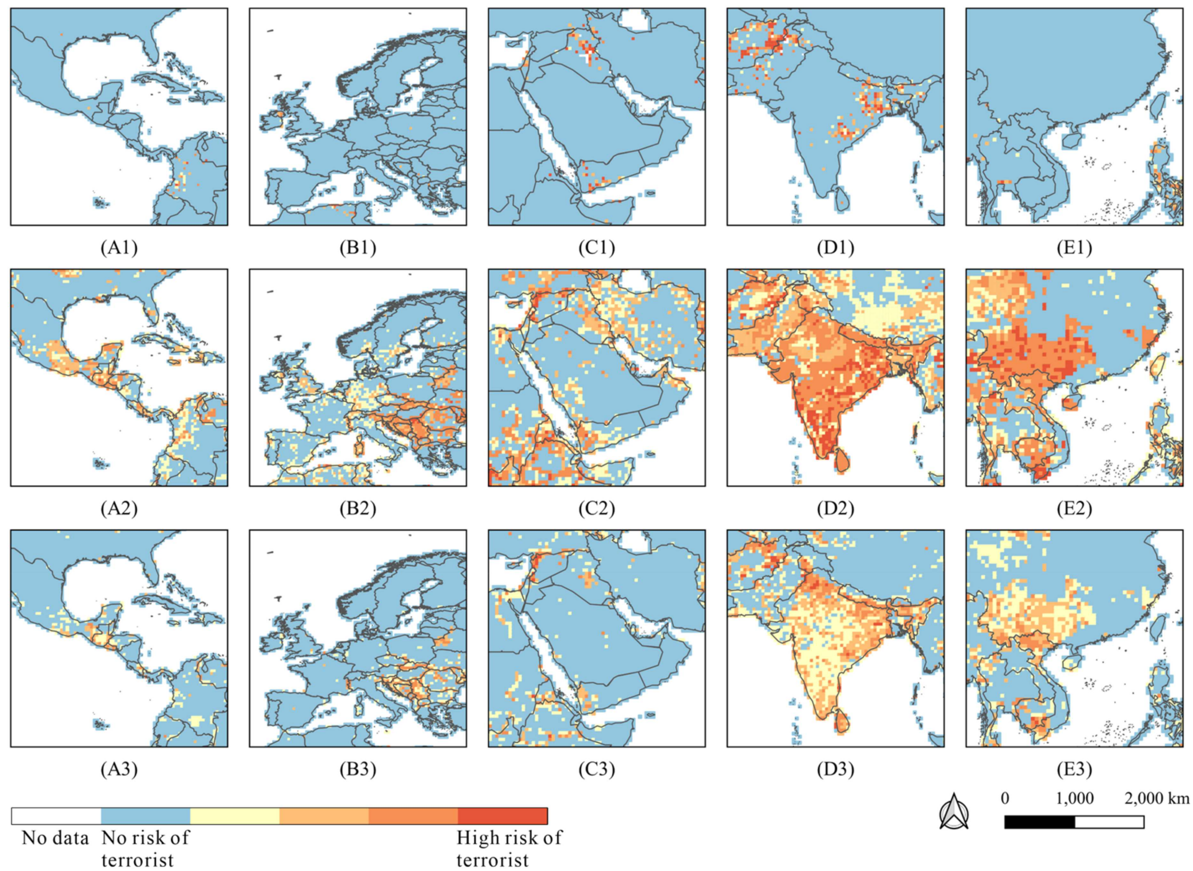

Figure 5. Comparisons of the results of the different methods: (A1)–(E1) show the actual data; (A2)–(E2) show the predictions of the random forest model; (A3)–(E3) show the predictions of the TerrorXG model.

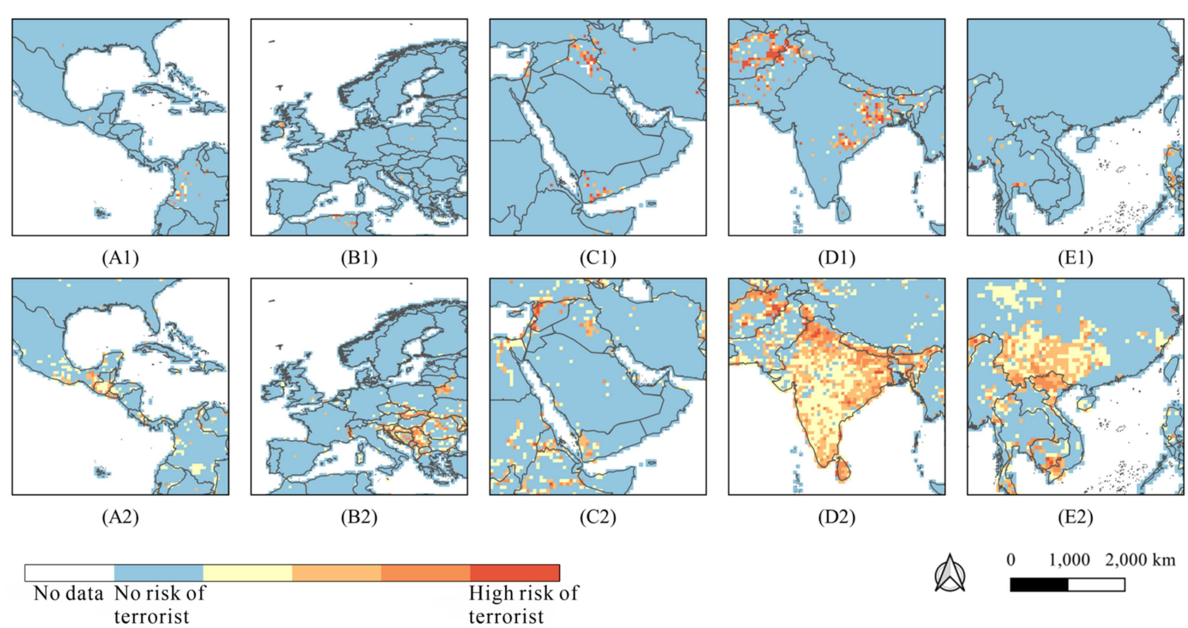

Figure 6. Comparisons between the TerrorXG prediction results and actual data: (A1)–(E1) show the actual data; (A2)–(E2) show the prediction results of the TerrorXG model.

4.2. Analysis of the driving factors of terrorist attacks

SHAP analysis was employed to perform an interpretability analysis of the TerrorXG model, with the aim of identifying key factors influencing the model outputs. For each driving factor input into the model, the average SHAP values were calculated. An analysis of the average SHAP values enabled the assessment of the relative importance of each factor. The relative importance of each driving factor derived from the average SHAP values is shown in Figure 7. Here, higher average SHAP values reflect a greater influence of the corresponding factor on the model's terrorism predictions. To better reveal the influence of each feature and detect potential prediction anomalies, the SHAP values for each factor were plotted across all the samples, as shown in Figure 8.

The driving factors were categorised into six groups: social, demographic, ethnic, land- use structure, natural, and economic. As illustrated in Figure 7, social and demographic factors, characterised by higher average absolute SHAP values, have substantially greater influences on terrorist attack events than natural and ethnic factors do, revealing a clear hierarchical pattern in factor importance.

The results in Figures 7 and 8 consistently indicate that the GCI is the most influential driver in the model. Its average absolute SHAP value is 42.4% greater than that of the population, which ranked second in terms of feature importance. Moreover, the results in Figure 8 clearly reveal a positive relationship between the GCI and the SHAP values, indicating that increasing conflict intensity significantly increases the predicted risk of terrorist attacks. Population size has a similar positive relationship with the SHAP values, highlighting the role of concentrated demographic regions in amplifying potential targets and attack likelihood.

Socioeconomic inequality indicators, including child malnutrition and infant mortality rates, ranked third and fourth, respectively. As shown in Figure 7, their contributions exceed those of urbanisation (ranked sixth) by 38.4% and 108.5%, respectively. The data in Figure 8 further reveal that higher values of these variables are generally associated with increased SHAP values, suggesting that structural social vulnerability plays a critical role in shaping terrorism risk.

The proportion of urban areas has a moderate influence on terrorism risk. The SHAP distribution in Figure 8 indicates a generally positive effect, implying that higher levels of urbanisation are associated with an increased likelihood of terrorist attacks, although the magnitude of this effect remains secondary to key social and demographic factors.

Ethnic factors contribute relatively little to the model. Regional ethnic mixing is associated with low average SHAP values, which are only 13.0% and 9.7% of those for population and the GCI, respectively.

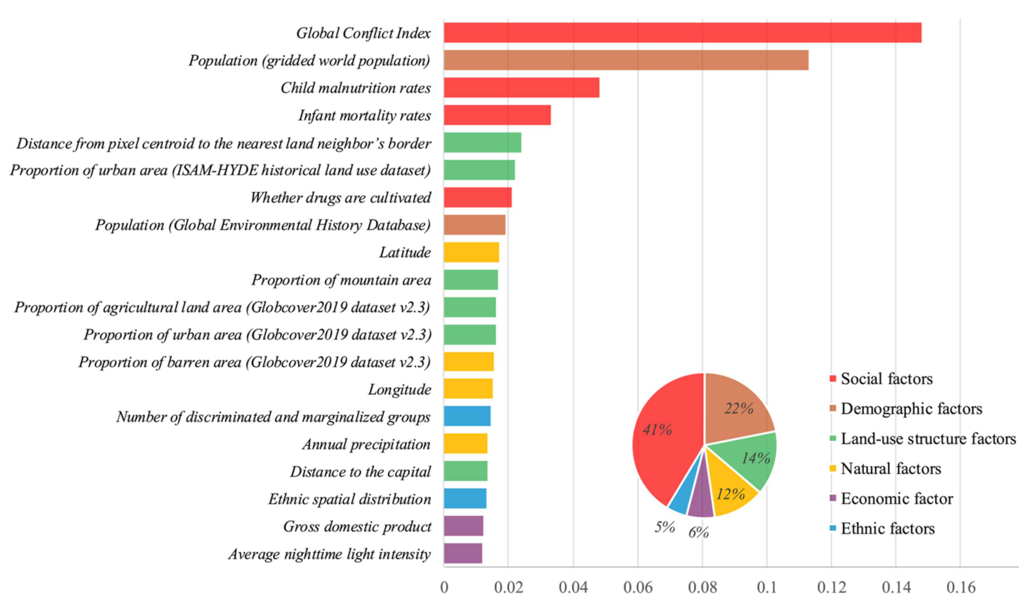

Figure 7. Histogram of the mean SHAP values for different terrorist attack factors.

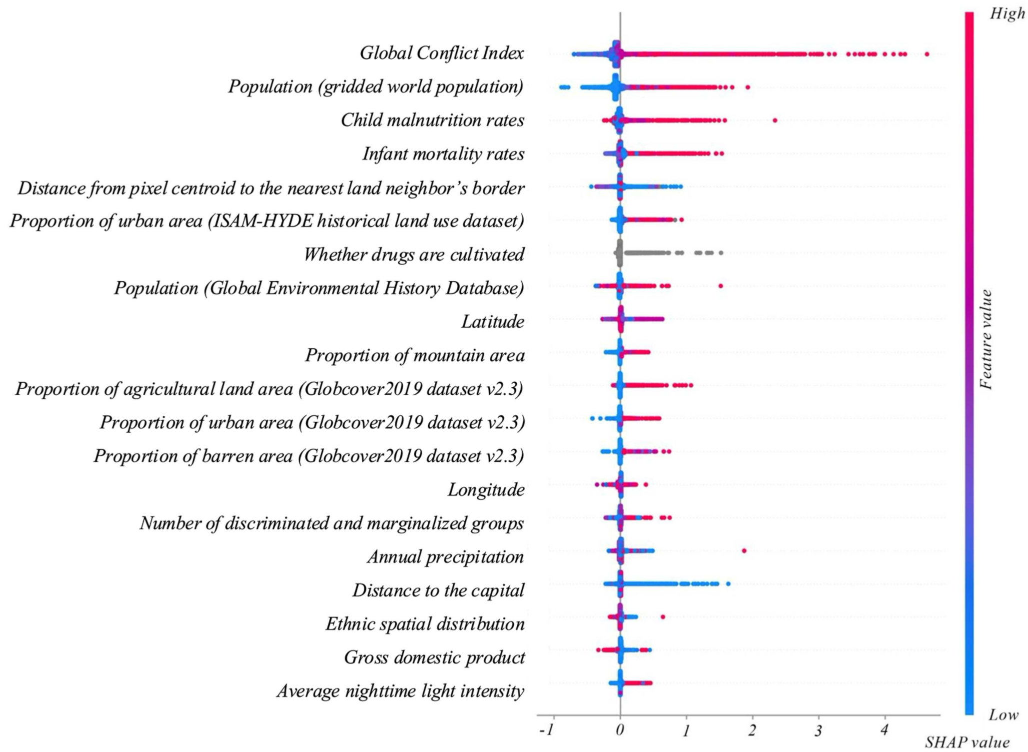

Figure 8. Dependency plot of the SHAP values for different terrorist attack factors.

This limited contribution is further reflected in Figure 8, as the SHAP values vary only weakly across different levels of ethnic mixing, suggesting the relatively minor and less consistent effect of this factor on terrorism prediction.

5. Discussion

Counterterrorism is crucial for national security and strategic decision-making worldwide. However, effectively predicting terrorist attacks through integrated spatiotemporal modelling and the systematic identification of driving factors on the basis of multisource data remains challenging. In this study, a news-based global conflict index was applied for the first time, and a novel dataset integrating multisource spatiotemporal data for predicting terrorist attacks was constructed. An XGBoost-based spatiotemporal model (TerrorXG) was developed to predict the likelihood of terrorist attacks, and the SHAP model was employed to analyse the underlying drivers of terrorist attacks. The model effectively predicted the global spatiotemporal distribution of terrorist attacks while revealing and quantifying the degree of influence of various driving factors. This study provides evidence to support the development of more targeted counterterrorism strategies.

The TerrorXG model addresses the limitations of previous methods, which typically focused on either temporal or spatial aspects separately. Compared with conventional prediction approaches, the TerrorXG model provides global-scale spatiotemporal terrorist attack predictions with significantly higher accuracy. TerrorXG outperformed all the other machine learning models and significantly outperformed existing models from previous studies (Matereke, Nyirenda, and Ghaziasgar 2021) in terms of generalisability and training efficiency. In social science prediction models, an $R^2 \geq 0.1$ indicates theoretical explanatory value if the core variables are statistically significant (Ozili 2023). The R^2 value of our model exceeded this benchmark by nearly six times ($R^2 = 0.595$), confirming its effectiveness in capturing the spatiotemporal dynamics of terrorist attacks. Overall, TerrorXG demonstrated high predictive accuracy, offering actionable insights for the allocation of counterterrorism resources. Regions identified by the model as having higher terrorism risk than historically observed trends warrant targeted attention from relevant authorities.

The clandestine nature of terrorist activities often leads to incomplete public event records, introducing bias and uncertainty in prediction outcomes. This discrepancy reflects the model's success in capturing latent risk. In regions where the predicted and observed terrorist events differ, an analysis of the multidimensional drivers reveals the complex underlying mechanisms. The results of the SHAP analysis (Figures 7 and 8) quantitatively demonstrated that the distance from the pixel centroid to the border of the nearest land neighbour was the fifth most significant factor affecting the model predictions, while the proportion of mountainous area also played a critical role.

The densely populated, predominantly mountainous terrain of Southwest China enhances terrorists' ability to disseminate information and evade authorities postattack (McClanahan and Linnemann 2018). Additionally, the proximity to international borders, such as with India, provides routes for cross-border escape and mobility. Furthermore, the location within the Golden Triangle increases the likelihood of illicit activities, including drug trafficking (Omelicheva and Markowitz 2019; Yang et al. 2019) and telecommunication fraud (China Academy of Information and Communications Technology 2020). This confluence of challenging terrain, porous borders, and transnational crime increases the risk of terrorism in Southwest China. With respect to India, existing research indicates that political elections may foster terrorism (Dmello, Perliger, and Sweeney 2022). These factors collectively indicate that although the recorded values in the GTD suggest a lower risk of terrorism in these regions, the model highlights persistent terrorism risk in these regions, which warrants further attention.

By employing quantitative SHAP analysis with the TerrorXG model, socioeconomic inequality and urbanisation were identified as key drivers influencing global terrorist attacks. Furthermore, compared with natural and environmental factors, social factors generally have stronger effects on terrorist attack patterns. Given the close interconnections between social issues and governmental policies (Van Der Does et al. 2021), as well as the predominantly political motives underlying terrorism, the findings emphasise the pivotal role of governance and social policy interventions in addressing terrorism.

In this study, the GCI is proposed as a novel, news-based metric to quantify the intensity of national conflicts within the context of social factors. The quantitative SHAP analysis revealed that the GCI is the predominant factor influencing terrorist attacks. The impact of the GCI was 42.4% greater than that of population, which was the second most influential factor. A higher GCI was associated with an increased terrorist attack index in the corresponding regions. Tense international relations undermine regional peace and stability (Broeders, Cristiano, and Weggemans 2023), as areas prone to violent disputes attract terrorist organisations seeking to infiltrate and recruit members (Python, Brandsch, and Tskhay 2017). These findings suggest that terrorism risk is structurally embedded in broader geopolitical instability, emphasising the importance of incorporating conflict intensity into prediction frameworks.

Additionally, socioeconomic inequality was identified as a major determinant of terrorist attacks. The child malnutrition rate and infant mortality rate, which reflect gaps between the wealthy and the poor (Van de Poel et al. 2008), ranked third and fourth, respectively, in terms of predictive importance, surpassing the urbanisation rate. The higher ranking of the child malnutrition rate and infant mortality rate compared with the urbanisation rate suggests that areas characterised by socioeconomic inequality are more susceptible to terrorist attacks than regions with high levels of urban development are. This may reflect the role of socioeconomic deprivation in fostering grievances, reducing opportunity costs for participating in extremist activities, and increasing susceptibility to radicalisation.

Furthermore, densely populated urban areas are commonly targets of terrorist organisations. The influence of the population size on the risk of terrorist attacks ranked second, which was consistent with existing views (Stottlemyre 2014). Terrorist attacks frequently occur in densely populated areas to maximise impact and facilitate terrorists' objectives (Al-Sabbagh et al. 2026). In addition to urbanisation itself, factors such as a high population density, developed transportation infrastructure, and proximity to major urban centres facilitate rapid dissemination of information, all of which increase the vulnerability of these locations to terrorist attacks.

In contrast to the conclusions of prior research, our findings revealed that ethnic and racial factors exerted relatively minimal and statistically insignificant effects on the risk of terrorist attacks. Ethnic variables explained only 9.7% of the global conflict, challenging established views in the terrorism literature (Choi 2022) and highlighting the need to revisit previously emphasised theories and systematically reexamine the influence of ethnic conflicts on terrorist activities. These findings challenge conventional assumptions in terrorism studies and suggest that structural and governance-related factors may play more fundamental roles than the previously emphasised ethnic explanations.

On the basis of these findings, we propose targeted recommendations for mitigating the risk of terrorist attacks. Policy interventions should shift from the traditional focus on ethnic and racial factors to address the more significant drivers, namely, governance and social dynamic factors. First, given that the GCI is the predominant predictor of terrorist attacks, prioritising diplomatic engagement and conflict resolution mechanisms to de-escalate regional and international tensions is critical. A stable geopolitical environment systematically reduces the operational capacity and recruitment appeal of terrorist organisations. Second, addressing socioeconomic disparities, as evidenced by indicators such as the child malnutrition and infant mortality rates, is crucial. Implementing targeted social welfare policies to reduce the gap between the wealthy and the poor can effectively undermine the grievances that fuel radicalisation. Finally, while urbanisation itself is a weaker causal factor, the inherent vulnerability of densely populated centres necessitates enhanced security governance, including robust infrastructure protection and advanced emergency response capabilities, to mitigate the potential impact of terrorist attacks.

Although we developed the TerrorXG prediction model on the basis of the XGBoost algorithm and identified key factors driving terrorist attacks, several limitations should be acknowledged. For instance, the impact of political systems was not explicitly examined, and the limited data availability restricted the comprehensive analysis of some relevant factors, such as governmental stability and the regime type. Additionally, the definition of national conflict intensity did not differentiate among specific types of conflicts (such as diplomatic conflicts). Furthermore, the data source for assessing the intensity of national conflicts is relatively singular, which may introduce certain biases in the analysis. Owing to data limitations, it was also challenging to comprehensively analyse the evolving forms of terrorism, such as cyberterrorism. In future studies, these limitations should be addressed by carefully examining political systems, utilising improved datasets, and distinguishing among different types of national conflicts.

6. Conclusion

In this study, a quantitative global conflict index was constructed on the basis of news data, and the index was integrated with diverse multisource datasets to support global-scale terrorism prediction. Leveraging this enriched dataset, we developed TerrorXG, an XGBoost-based model for the spatio-temporal forecasting of terrorist incidents. SHAP analysis identified key factors influencing terrorist incidents. TerrorXG provides accurate global predictions of terrorist attacks and outperforms other machine learning methods. In regions in which discrepancies between the predicted and actual attacks were identified, an analysis of the multisource driving factors adequately explained terrorist incidents. The results indicate that compared with natural factors, social factors have substantially stronger effects on terrorist attacks, reflecting the inherent relationship between terrorist activities and government policies. The proposed GCI has the strongest influence on the prediction of terrorist attacks, with a SHAP value that is 42.4% higher than that of the second-ranking population factor, highlighting the effectiveness and practical relevance of the proposed indicator. Additionally, regions characterised by high socioeconomic inequality and densely populated urban areas are more vulnerable to terrorism. In contrast to the findings of previous studies, our findings suggest that ethnicity exerts only limited influence on the risk of terrorism.

This study offers a theoretical foundation for assessing terrorist risk and provides practical guidance for policy-makers in terms of the efficient allocation of counterterrorism resources. The findings suggest that to mitigate terrorism risk, reducing the intensity of national conflicts, addressing socioeconomic inequalities, and strengthening governance in densely populated areas should be prioritised. Moreover, several limitations remain, including the lack of explicit consideration of political systems, potential bias from the use of single-source conflict data, and insufficient coverage of emerging forms of terrorism, such as cyberterrorism. In future research, more diverse datasets should be incorporated, conflict typologies should be refined, and the role of political and institutional factors in shaping terrorism dynamics should be further explored.

Disclosure statement

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.

Funding

This work was supported by the National Natural Science Foundation of China [42471491], the National Key Research and Development Programme of China [2023YFB3906803], the Key Laboratory of Earth Surface System and Human-Earth Relations, Ministry of Natural Resources of China [LBXT2023YB03], and the 'CUG Scholar' Scientific Research Funds at China University of Geosciences (Wuhan) [2022034]. This work was partly supported by a project grant from the Co-creation Centre for Disaster Resilience, IRIDeS, Tohoku University (ID: 2-QR001).

Author contributions

CRediT: **Xiang Zhang:** Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing; **Xin Qiu:** Formal analysis, Writing – original draft; **Chenglong Yu:** Conceptualization, Data curation, Methodology, Validation; **Ziheng Chen:** Investigation, Methodology; **Geyuan Zhu:** Methodology, Writing – review & editing; **Liangyang Dai:** Investigation, Resources, Writing – review & editing; **Qingfeng Guan:** Funding acquisition, Project administration, Supervision; **Yao Yao:** Project administration, Writing – review & editing.

Notes on contributors

Xiang Zhang is a graduate student at China University of Geosciences (Wuhan), China and an intern student at LocationMind Institute, LocationMind Inc., Japan. His research interests include GeoAI and human mobility.

Xin Qiu is a graduate student at China University of Geosciences (Wuhan), China. Her research interests include GeoAI and geospatial big data mining.

Chenglong Yu is a graduate student at China University of Geosciences (Wuhan), China and an intern student at LocationMind Institute, LocationMind Inc., Japan. His research interests include GeoAI and Large Language Model.

Ziheng Chen is a graduate student at Institute of Seismology, China Earthquake Administration, Wuhan. His research interests include GeoAI and geospatial big data mining.

Geyuan Zhu is a graduate student at China University of Geosciences (Wuhan), China and an intern student at LocationMind Institute, LocationMind Inc., Japan. His research interests are intelligent agriculture, and large language model.

Liangyang Dai is a phd student at Wuhan University. His research interests are geospatial big data mining and health geography.

Qingfeng Guan is a Professor at China University of Geosciences (Wuhan). His research interests include high-performance spatial intelligence computation and urban computing.

Yao Yao is a Professor at China University of Geosciences (Wuhan), Hitotsubashi University, and Reitaku University. He is also a Senior Scientist at LocationMind Inc. He has previously served as a Project Researcher at the University of Tokyo. His research interests include spatiotemporal big data mining, social geographic computing, and urban geographic information systems.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Ackerman, G., and H. Peterson. 2020. "Terrorism and COVID-19: Actual and Potential Impacts." *Perspectives on Terrorism* 14 (3): 59–73. https://www.jstor.org/stable/26918300.

Alonso, R. 2023. "Why Did so Few Become Terrorists: A Comparative Study of Northern Ireland and the Basque Country." *Terrorism and Political Violence* 35 (2): 264–283. https://doi.org/10.1080/09546553.2021.1905631.

Al-Sabbagh, T. A., B. Mohamadi, M. O. Abu Ghazala, and A. Younes. 2026. "Land Use as a Target for Terrorism: A Geospatial and Temporal Framework of Terrorist Preferences Analysis in Egypt." *Land Use Policy* 160: 107822. https://doi.org/10.1016/j.landusepol.2025.107822.

Asongu, S. A., S. Le Roux, and P. Singh. 2021. "Fighting Terrorism in Africa: Complementarity between Inclusive Development, Military Expenditure and Political Stability." *Journal of Policy Modeling* 43 (5): 897–922. https://doi.org/10.1016/j.jpolmod.2020.01.014.

Atoum, M. S., A. A. Alarood, E. A. Alsolmi, A. Obeidat, and M. Alazab. 2025. "Predictive Analysis of Global Terrorist Attacks Using Lexical Patterns Across Multiple Datasets." *Expert Systems* 42 (1): e13808. https://doi.org/10.1111/exsy.13808.

Broeders, D., F. Cristiano, and D. Weggemans. 2023. "Too Close for Comfort: Cyber Terrorism and Information Security Across National Policies and International Diplomacy." *Studies in Conflict & Terrorism* 46 (12): 2426–2453. https://doi.org/10.1080/1057610X.2021.1928887.

Cai, Z., Z. Huang, Y. Guo, M. Gong, Y. Wang, and Y. Chang. 2024. "Spatial-Temporal Evolution Analysis of Terrorist Attacks in Pakistan from 2001 to 2020." *World Regional Studies* 33 (5): 60. https://doi.org/10.3969/j.issn.1004-9479.2024.05.20222175.

Chen, T., and C. Guestrin. 2016. XGBoost: A Scalable Tree Boosting System In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, Association for Computing Machinery, New York, NY, USA. 785–794. https://doi.org/10.1145/2939672.2939785.

Chen, W., G. Zhai, Z. Shen, and Y. Shi. 2020. "Analysis of Temporal and Spatial Characteristics of Terrorist Attacks in the Middle East from 1970 to 2015." *World Regional Studies* 29 (4): 659. https://doi.org/10.3969/j.issn.1004-9479.2020.04.2019039.

China Academy of Information and Communications Technology. 2020. Study on the Governance of Telecommunication Network Fraud under the New Circumstances. https://www.caict.ac.cn/kxyj/qwfb/ztbg/.

Choi, S.-W. 2022. "Leader Nationalism, Ethnic Identity, and Terrorist Violence." *British Journal of Political Science* 52 (3): 1151–1167. https://doi.org/10.1017/S0007123421000144.

Dmello, J. R., A. Perliger, and M. Sweeney. 2022. "The Violence of Political Empowerment: Electoral Success and the Facilitation of Terrorism in the Republic of India." *Terrorism and Political Violence* 34 (7): 1281–1304. https://doi.org/10.1080/09546553.2020.1761342.

Galla, D. and J. Burke. 2018. Predicting Social Unrest Using Gdelt In *Machine Learning and Data Mining in Pattern Recognition*P Perner, ed. 103–116. Springer International Publishing. https://doi.org/10.1007/978-3-319-96133-0_8

Ge, Q., M. Hao, F. Ding, D. Jiang, J. Scheffran, D. Helman, and T. Ide. 2022. "Modelling Armed Conflict Risk under Climate Change with Machine Learning and Time-Series Data." *Nature Communications* 13 (1): 2839. https://doi.org/10.1038/s41467-022-30356-x.

Haghani, M., E. Kuligowski, A. Rajabifard, and P. Lentini. 2022. "Fifty Years of Scholarly Research on Terrorism: Intellectual Progression, Structural Composition, Trends and Knowledge Gaps of the Field." *International Journal of Disaster Risk Reduction* 68: 102714. https://doi.org/10.1016/j.ijdrr.2021.102714.

Hastie, T., J. Friedman, and R. Tibshirani. 2001. The Elements of Statistical Learning. Springer. https://doi.org/10.1007/978-0-387-21606-5

He, R., Y. Lu, C. Jiang, Y. Deng, X. Li, and D. Shi. 2023. "Progress in Research and Practice of Spatial-Temporal Crime Prediction over the Past Decade." *Journal of Geo-Information Science* 25 (4): 866–882. https://doi.org/10.12082/dqxxkx.2023.220808.

Huamaní, E. L., A. M. Alicia, and A. Roman-Gonzalez. 2020. "Machine Learning Techniques to Visualize and Predict Terrorist Attacks Worldwide Using the Global Terrorism Database." *International Journal of Advanced Computer Science and Applications (IJACSA)* 11(4): 562–570. https://doi.org/10.14569/IJACSA.2020.0110474.

Institute for Economics & Peace. 2024. Global Terrorism Index 2024: Measuring the Impact of Terrorism, Sydney, February 2024. http://visionofhumanity.org/resources.

Karlsrud, J. 2017. "Towards UN Counter-Terrorism Operations?" *Third World Quarterly* 38 (6): 1215–1231. https://doi.org/10.1080/01436597.2016.1268907.

Kiyani, G., J. Pickering, M. Raza, and C. Webb. 2025. "Civil-Military Relations and Domestic Terrorism." *Journal of Conflict Resolution* 69 (1): 127–151. https://doi.org/10.1177/00220027231224022.

Koutidis, G., K. Loumponias, T. Tsikrika, S. Vrochidis, and I. Kompatsiaris. 2024. "Towards AI-Driven Prediction of Terrorism Risk Based on the Analysis of Localized Web News." *IEEE Access* 12: 167288–167299. https://doi.org/10.1109/ACCESS.2024.3485913.

Krieger, T., and D. Meierrieks. 2019. "Income Inequality, Redistribution and Domestic Terrorism." *World Development* 116: 125–136. https://doi.org/10.1016/j.worlddev.2018.12.008.

LaFree, G., and L. Dugan. 2007. "Introducing the Global Terrorism Database." *Terrorism and Political Violence* 19 (2): 181–204. https://doi.org/10.1080/09546550701246817.

Leetaru, K., and P. A. Schrodt. 2013. "GDELT: Global Data on Events, Location, and Tone, 1979-2012." *ISA Annual Convention* 2 (4): 1–49. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.686.6605.

Li, Z. 2022. "Extracting Spatial Effects from Machine Learning Model Using Local Interpretation Method: An Example of SHAP and Xgboost." *Computers, Environment and Urban Systems* 96: 101845. https://doi.org/10.1016/j.compenvurbsys.2022.101845.

Liu, L., W. Liu, W. Liao, H. Yu, C. Jiang, R. Lin, J. Ji, and Z. Zhang. 2018. "Comparison of Random Forest Algorithm and Space-Time Kernel Density Mapping for Crime Hotspot Prediction." *Progress in Geography* 37 (6): 761–771. https://doi.org/10.18306/dlkxjz.2018.06.003.

Lundberg, S. M., and S.-I. Lee. 2017. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems* 30: 4765–4774. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

Luo, L., C. Qi, H. Wang, and L. Wang. 2020. "Prediction of Terrorist Attack Events Time Based on LSTM Model." *Journal of Systems Engineering Journal* 35 (2): 163–172. https://doi.org/10.13383/j.cnki.jse.2020.02.003.

Ma, M., G. Zhao, B. He, Q. Li, H. Dong, S. Wang , and Z. Wang. 2021. "XGBoost-based Method for Flash Flood Risk Assessment." *Journal of Hydrology* 598: 126382. https://doi.org/10.1016/j.jhydrol.2021.126382.

Matereke, T., C. N. Nyirenda, and M. Ghaziasgar. 2021. "A Comparative Evaluation of Spatio Temporal Deep Learning Techniques for Crime Prediction." *2021 IEEE AFRICON*, 1–6. https://doi.org/10.1109/AFRICON51333.2021.9570858.

McClanahan, B., and T. Linnemann. 2018. "Darkness on the Edge of Town: Visual Criminology and the "Black Sites" of the Rural." *Deviant Behavior* 39: 512–524. https://doi.org/10.1080/01639625.2017.1407111.

Meng, C., and N. Ghafoori. 2024. "The Economic Impact of Terrorism in South Asia." *Socio-Economic Planning Sciences* 96: 102084. https://doi.org/10.1016/j.seps.2024.102084.

Mo, H., X. Meng, J. Li, and S. Zhao. 2017. "Terrorist Event Prediction Based on Revealing Data." *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*. 239–244. https://doi.org/10.1109/ICBDA.2017.8078815.

Murphy, M., E. Sharpe, and K. Huang. 2024. "The Promise of Machine Learning in Violent Conflict Forecasting." *Data & Policy* 6: e35. https://doi.org/10.1017/dap.2024.27.

Nabin, M. H., M. T. H. Chowdhury, S. Bhattacharya, and P. M. Sgro. 2022. "Terrorism and Uneven Economic Development." *Economic Modelling* 113: 105870. https://doi.org/10.1016/j.econmod.2022.105870.

Omelicheva, M. Y., and L. Markowitz. 2019. "Does Drug Trafficking Impact Terrorism? Afghan Opioids and Terrorist Violence in Central Asia." *Studies in Conflict & Terrorism* 42 (12): 1021–1043. https://doi.org/10.1080/1057610X.2018.1434039.

Ozili, P. K. 2023. The Acceptable R-square in Empirical Modelling for Social Science Research. Social research methodology and publishing results: A guide to non-native English speakers. 134–143. IGI Global. https://doi.org/10.2139/ssrn.4128165.

Python, A., J. Brandsch, and A. Tskhay. 2017. "Provoking Local Ethnic Violence – a Global Study on Ethnic Polarization and Terrorist Targeting." *Political Geography* 58: 77–89. https://doi.org/10.1016/j.polgeo.2017.02.001.

Qi, L., K. Qin, P. Luo, B. Yao, and Z. Zhu. 2021. "Quantitative Expression of Conflict Intensity and Conflict Event Detection Based on GDELT News Data." *Journal of Geo-Information Science* 23 (11): 1956–1970. https://doi.org/10.12082/dqxxkx.2021.210172.

Qin, K., P. Luo, and B. Yao. 2019. "Networked Mining of GDELT and International Relations Analysis." *Journal of Geo-Information Science* 21 (1): 14–24. https://doi.org/10.12082/dqxxkx.2019.180674.

Rubinstein, R. Y. and D. P. Kroese. 2016. Cross-Entropy Method In *Simulation and the Monte Carlo Method*. 257–306. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118631980.ch8

Sagi, O., and L. Rokach. 2021. "Approximating XGBoost with an Interpretable Decision Tree." *Information Sciences* 572: 522–542. https://doi.org/10.1016/j.ins.2021.05.055.

Schuurman, B. 2020. "Research on Terrorism, 2007–2016: A Review of Data, Methods, and Authorship." *Terrorism and Political Violence* 32 (5): 1011–1026. https://doi.org/10.1080/09546553.2018.1439023.

Shen, S., C. Song, C. Cheng, J. Gao, and S. Ye. 2020. "GDELT: Big Event Data for Sensing Global Social Dynamics." *World Regional Studies* 29 (1): 71. https://doi.org/10.3969/j.issn.1004-9479.2020.01.2019800.

Shwartz-Ziv, R., and A. Armon. 2022. "Tabular Data: Deep Learning Is Not All You Need." *Information Fusion* 81: 84–90. https://doi.org/10.1016/j.inffus.2021.11.011.

Stottlemyre, S. 2014. The Effect of Country-Level Income on Domestic Terrorism: A Worldwide Analysis of the Difference between Lone-Wolf and Group Affiliated Domestic Terrorism. Georgetown University. https://repository.digital.georgetown.edu/handle/10822/709882.

Tollefsen, A. F., K. Bahgat, J. Nordkvelle, and H. Buhaug. 2016. "PRIO-GRID V. 2.0 Codebook." *Journal of Peace Research* 49(2): 363–374. https://doi.org/10.1177/0022343311431287.

Van de Poel, E., A. R. Hosseinpoor, N. Speybroeck, T. Van Ourti, and J. Vega. 2008. "Socioeconomic Inequality in Malnutrition in Developing Countries." *Bulletin of the World Health Organization* 86(4): 282–291. https://doi.org/10.2471/BLT.07.044800.

Van den Broeck, G., L. Anton, S. Maximilian, and S. Dan. 2022. "On the Tractability of SHAP Explanations." *Journal of Artificial Intelligence Research* 74: 851–886. https://doi.org/10.1613/jair.1.13283.

Van Der Does, R., J. Kantorowicz, S. Kuipers, and M. Liem. 2021. "Does Terrorism Dominate Citizens' Hearts or Minds? the Relationship between Fear of Terrorism and Trust in Government." *Terrorism and Political Violence* 33 (6): 1276–1294. https://doi.org/10.1080/09546553.2019.1608951.

Voukelatou, V., I. Miliou, F. Giannotti, and L. Pappalardo. 2022. "Understanding Peace through the World News." *EPJ Data Science* 11 (1): 2. https://doi.org/10.1140/epjds/s13688-022-00315-z.

Wan, G., and X. Cao. 2016. "The Historical Evolution and Reflection of Geospatial Information Grid." *Acta Geodaetica et Cartographica Sinica* 45 (S1): 15–22. https://doi.org/10.11947/j.AGCS.2016.F002.

Wang, W., Y. Lin, Z. Xu, X. Zeng, and C. Wang. 2021. "Research on Damage Assessment and Spatial-Temporal Evolution of Globle Terrorist Attacks from 2000 to 2018." *Human Geography* 36 (5): 96–100. https://doi.org/10.13959/j.issn.1003-2398.2021.05.012.

Wang, D., S. Thunéll, U. Lindberg, L. Jiang, J. Trygg, and M. Tysklind. 2022. "Towards Better Process Management in Wastewater Treatment Plants: Process Analytics Based on SHAP Values for Tree-Based Machine Learning Methods." *Journal of Environmental Management* 301: 113941. https://doi.org/10.1016/j.jenvman.2021.113941.

Weidmann, N. B., J. K. Rød , and L.-E. Cederman. 2010. "Representing Ethnic Groups in Space: A New Dataset." *Journal of Peace Research* 47 (4): 491–499. https://doi.org/10.1177/0022343310368352.

Yang, M., Y. Wang, Y. Chen, Y. Zhou, and Q. Jiang. 2019. "Impact of Maternal HIV Infection on Pregnancy Outcomes in Southwestern China—A Hospital Registry Based Study." *Epidemiology and Infection* 147: e124. https://doi.org/10.1017/S0950268818003345.

Zhang, X., L. Sun, and L. Wang. 2015. "Statistical Data Gridding in China'S First Geographic Condition Survey." *Remote Sensing Information* 30 (3): 27–32. https://doi.org/10.3969/j.issn.1000-3177.2015.03.006.

Zhang, X., Y. Yao, C. Yu, Z. Hu, G. Zhu, M. Shibasaki, L. Dai, Y. Guo, Q. Guan, and R. Shibasaki. 2026. "MGIM: A Masked Modeling Framework for Land Parcel-Level Geo-Inference." *International Journal of Geographical Information Science* 0(0): 1–27. https://doi.org/10.1080/13658816.2026.2630403.

Zhao, Z., K. Liu, and M. Wang. 2024. "A Novel Spatial-Aware Deep Learning Approach for Exploring the Environmental Context of Terrorist Attacks and Armed Conflicts." *International Journal of Disaster Risk Reduction* 114: 104921. https://doi.org/10.1016/j.ijdrr.2024.104921.