IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 26 September 2023, accepted 16 November 2023, date of publication 24 November 2023, date of current version 8 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3336811

RESEARCH ARTICLE

AI-Driven Counter-Terrorism: Enhancing Global Security Through Advanced Predictive Analytics

FAHAD ALI KHAN[1], GANG LI[1,2], ANAM NAWAZ KHAN[3], QAZI WAQAS KHAN[3], MYRIAM HADJOUNI[4], AND HELA ELMANNAI[5]

[1]College of Urban and Environmental Sciences, Northwest University, Xi'an 710127, People's Republic of China (PRC)
[2]Shaanxi Key Laboratory of Earth Surface System and Environmental Carrying Capacity, Northwest University, Xi'an 710127, People's Republic of China (PRC)
[3]Department of Computer Engineering, Big-Data Research Center, Jeju National University, Jeju 63243, Republic of Korea
[4]Department of Computer Sciences, College of Computer and Information Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[5]Department of Information Technology, College of Computer and Information Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

Corresponding authors: Hela Elmannai (hselmannai@pnu.edu.sa) and Gang Li (lig@nwu.edu.cn)

The authors would like to thank Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R193), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

ABSTRACT Recent terrorist attacks have emerged as a formidable menace to global peace and security, giving rise to an acute humanitarian and economic crisis characterized by the loss of numerous lives and the incurring of substantial financial damages. In response to this pressing concern, the scholarly community has introduced a range of AI-driven predictive analytics methodologies as prospective instruments in the fight against terrorism. Machine Learning (ML) techniques, commonly employed in counter-terrorism efforts, face the formidable challenge of accurately forecasting terrorism activities due to the escalating complexity and voluminous nature of the underlying data. In this context, we present a solution in the form of an optimal weighted voting ensemble classifier, specially tailored for the classification of weapon types in terrorist attacks. Leveraging the computational capabilities of the Particle Swarm Optimization (PSO) algorithm, we determine the optimal weight assignments for the base learners, namely Random Forest and Xtreme Gradient Boosted Machines. Additionally, we developed several machine learning models to predict the likelihood of casualties in the aftermath of a terrorist incident. Rigorous validation of these models is carried out utilizing the publicly available Global Terrorism Database (GTD), necessitating meticulous data preprocessing to rectify anomalies and address class imbalances, a task effectively accomplished through the application of the Synthetic Minority Over-sampling Technique (SMOTE). To underscore the effectiveness of our proposed model, we conduct an extensive comparative analysis, bench-marking it against state-of-the-art machine learning models. The comprehensive experimental results unequivocally demonstrate the superiority of our models, as they consistently achieve the lowest error rates, thereby highlighting their enhanced generalizability and performance. Our detailed experiments show that the PSO-based classifier achieves 95% accuracy in weapon classification, while in case of casualty prediction, The XGB model showcases the most consistent and reliable performance across all datasets from Pakistan, India, and Afghanistan, with the lowest Mean Squared Error. The findings derived from our study furnish invaluable insights to empower counter-terrorism agencies, facilitating data-driven decision-making and the proactive implementation of measures to mitigate and counteract the scourge of terrorist acts.

INDEX TERMS Predictive analytics, machine learning, spatial patterns, artificial intelligence.

I. INTRODUCTION

Terrorism has deeply impacted the global landscape, notably over the past two decades, leading to a setback in economic development and societal advancement. As defined by the Global Terrorism Index (GTI), terrorism is delineated as the application of physical force or the threat thereof by non-state entities to enforce compliance and instill apprehension within specific groups, thereby advancing the entity's political, economic, religious, or social agenda [1]. Terrorism is

The associate editor coordinating the review of this manuscript and approving it for publication was Sawyer Duane Campbell.

characterized by the intentional employment of violence and intimidation, mainly targeting civilian populations, with the intent to fulfill certain political, religious, or ideological aims. Actions such as bombings, shootings, kidnappings, and cyber-attacks fall within its scope, typically directed at critical sites or demographics to maximize psychological impact. The term further encompasses the adaptive quality of terrorist methods, which persistently transform to align with advancements in technology and shifts in the global socio-political landscape, thereby posing a persistent and intricate challenge.

The multifaceted repercussions of terrorism, as posited by [2], encompass the wastage of both human and material resources. However, the gravest consequence of terrorist attacks lies in the loss of innocent lives. According to a study conducted by [3], terrorism is responsible for the demise of 29,376 individuals in the year 2015.

Along with stock market volatility, financial loss, and increased security threats, terrorism also has secondary implications like a decrease in trade flows, tourism, and FDI. Many facets of terrorism, such as its causes, consequences, and underlying causes, have already been explored and analyzed in the existing literature [4]. Terrorist acts are found to have a negative correlation with both progress and education levels. Mineral wealth and authoritarian governments both played beneficial roles in the events that took place. Two large areas of Eurasia provided the data for their analysis [5]. Using data on terrorist attacks and fatalities from 96 countries, [6] ran a series of multiple regression studies and found no correlation between terrorism and a host of economic and social variables. The results disprove a widely held belief that those factors contribute to an increase in terrorist activity. Nonetheless, more research is needed into the topic because the number of terrorist attacks in developing nations is far higher than in developed nations. In an effort to better understand the factors that contribute to terrorist attacks and to counteract them, the United States established the Global Terrorism Database (GTD), an open-access global database on terrorism [7]. According to the Global Terrorism Database, 3,800 terrorist attacks have been reported on average every year between 1970 and 2017. There has been a dramatic increase in the annual number of terrorist incidents from 2005 to 2015, but it has been progressively declining ever since.

Unfortunately, because of the complexities of terrorism, it is challenging to establish an efficient counter-terrorism strategy that can guarantee the safety of its victims. As preventing and resolving terrorist attacks is incredibly challenging, therefore development of an efficient method for predicting the number of casualties and classifying the weapon types have become a de facto necessity.

The global patterns and results of terrorism have been the subject of analysis and prediction by a number of researchers. ML classifiers and regressors are extremely important [8] and have made significant contributions to the development of highly accurate predictive systems for counter-terrorism. Researchers have used ML methods to examine several dimensions of terrorism [9]. An ensemble machine learning model was developed by Olusola et al. [10] to foresee potentially terrorist-prone continents, combining support vector machine (SVM) and K-nearest neighbor (KNN). Widespread use of Deep learning architectures (DL) for predictive modeling in diverse domains. DL is a subset of ML that utilizes hidden layers to reveal previously obscured spatial and temporal relationships in large data-sets [11].

A major factor in the wide adoption of NN and, more specifically, DNN is huge amounts of labeled data being readily available [12]. In addition to terrorist risk prediction, various safety and security sectors use data-driven [13] and machine learning [14] techniques. The aforementioned studies provide a wealth of technical details for the present work. The predictive abilities of these machine learning models have been found to much exceed those of traditional statistical models.

The review of literature suggests that the current body of work is inadequate because of its failure to take into account important factors such as weapon type. Moreover, the estimates of casualties only consider the likelihood of fatalities or injuries, not the actual number of casualties, which is more crucial for risk management. Lastly, the developed methods lack sufficient accuracy for their wide adaption and time critical real-word applications. Nonetheless, an efficient predictive analytics framework is required. To resolve these issues, we developed a machine learning based framework comprising of an optimal weighted voting ensemble classification model with PSO based weight assignment and TabNet based regression model for classifying weapon types and predicting casualties respectively. The key contributions of the proposed work are as follows:

- Our research endeavors to create a range of Deep Learning and traditional Machine Learning models aimed at accurately forecasting forthcoming terrorist attack incidents, with a particular emphasis on discerning the weaponry involved in such assaults.
- To ensure the accurate classification of weaponry utilized by terrorists, we have devised an innovative weighted voting ensemble classifier, strategically leveraging the Particle Swarm Optimization (PSO) algorithm to assign optimal weights to the base learners for improved model performance.
- To bolster preemptive and efficacious counter-terrorism efforts, our study demonstrates the utility of applying SMOTE to address class imbalance within the GTD dataset and effectively resolved it for improved performance of prediction and classification models.
- The developed system contributes significantly to the field through a comprehensive examination of the GTD dataset, encompassing the identification of trends and patterns pertaining to the location, timing, and methodology of terrorist attacks.
- The developed system carries substantial implications for law enforcement agencies and the academic

community, offering valuable insights that can inform the development of more efficient counter-terrorism strategies through advance predictive analytics.

The rest of the paper is organized as follows Section II presents state of the art that contextualize our findings and highlight current development in this field. Section III describes the proposed predictive analytic framework, as well as detailes an analysis and explanation of the machine learning algorithms utilized in predictive modeling. Section IV presents a detailed data analysis of the GTD dataset. In the section V, we present experimental results and discuss the findings of the proposed predictive analytic for counter-terrorism. lastly Section VI conclude the findings of developed framework and provide future insights.

II. RELATED WORK

The research on terrorism research employs two distinct types of quantitative methods, namely statistical techniques, and machine learning approaches. In order to provide a comprehensive examination, we presented a review of literature that explores the application of statistical and machine learning techniques in the analysis of terrorist attacks and resulting casualties.

Before the extensive integration of Artificial Intelligence (AI) and ML methodologies, statistical approaches are the prevailing method employed for the analysis of data pertaining to terrorist attacks. Multiple prominent studies, such as the research conducted by [15] and [16], have presented statistical estimation methodologies that seek to predict the probability of significant terrorist acts on a wide scale by utilizing historical data. Borooah et al. did a comprehensive analysis on terrorism in India from 1998 to 2004. Their research included statistical analysis to examine the influence of different types of assaults and distinct terrorist groups on the number of casualties. In addition, the authors of [17] did a comparative investigation on the casualties and other repercussions arising from different types of explosions. It is crucial to acknowledge that the statistical approaches under consideration largely prioritize the tasks of data exploration, organization, and description, while placing a very limited emphasis on predicted outcomes.

ML is commonly utilized as an expansion of statistical learning in practical scenarios, offering algorithmic and technical assistance for the resolution of real-life predicaments. The utilization of ML has attracted substantial attention in the realm of emergency management and decision-making due to its proficient capacity to collect and distribute disaster-related data in real-time [18]. The process of data mining frequently encompasses tasks related to data classification, which aims to establish the correlation between features and their respective labels [19]. In recent times, there has been a widespread adoption of ML methodologies to tackle classification problems that are specifically associated with terrorist attacks [20]. For instance [21] proposed a machine learning based solution to process and evaluate information related to acts of terrorism. The objective of the proposed work is to predict the forthcoming acts of terrorism using Logistic Regression (LR), Decision Trees (DT), Gaussian Bayesian Network (GBN), AdaBoost (AB), and Random Forests (RF).

In the work conducted [22], an innovative hybrid classifier that harnesses the power of Big Data for the prediction of terrorist attacks is introduced. The methodology outlined in their study encompasses multiple phases, including data acquisition, pre-processing, and the creation of a hybrid classification model. To enhance the predictive capabilities of this hybrid classifier, Genetic Algorithm (GA) is employed to optimize the classifier weights. The study's outcomes indicate that the hybrid classifier outperformed standalone classifiers in terms of predictive performance.

The objective of the research undertaken by [23] is to develop a hybrid machine learning framework that could be utilized to forecast the participation of organizations in acts of terrorism. In [24] the authors conducted a study wherein they utilized multiple ML models, such as Artificial Neural Networks (ANN), Naïve Bayes (NB), Support Vector Machines (SVM), Random Forest, and Decision Trees. The objective of their research was to forecast the type of a terrorist attack as well as its spatial distribution.

A distinct study by [25] aimed at developing a comprehensive framework to predict the actions of terrorist organizations. In this work the researchers employed a methodology that incorporated a range of techniques, such as social network analysis, wavelet transform, and pattern recognition. This methodology facilitated the acquisition of profound understandings regarding the behavior of these collectives and the ability to formulate accurate forecasts pertaining to their attack strategies. The experimental findings provided clear evidence that SVM exhibited superior performance compared to other benchmark methods in all evaluation measures.

Numerous prior research endeavors have utilized deep learning methodologies in the examination of terrorism. Previous research in the fields of ML and Deep Learning (DL) has faced challenges in accurately predicting terrorist activities, particularly in the context of bi-label and multi-label classification tasks. This issue has been highlighted by [26] and discussed further by [27] in their respective studies. Convolutional Neural Networks (CNNs) have exhibited exceptional performance in tasks such as image classification, while other DL approaches have also demonstrated proficiency in multi-label classification. However, these methods face difficulties when dealing with complex and nonlinear data-sets. Furthermore, the high accuracy rates exhibited by these models are frequently constrained to smaller and less complex data-sets, which is incongruous with the real-world attributes of terrorism datasets. The existence of these differences highlights the urgent requirement to improve the precision of ML techniques, specifically when used for bi-label and multi-label classification problems as in case of terrorist datasets.

In addition, the existing literature lacks precision in predicting the exact number of casualties in terrorist attacks. Previous studies primarily focus on determining whether casualties occur or not, rather than providing detailed casualty numbers, which are crucial for effective risk management. Moreover, imbalance in the datasets adversely affect the accuracy of casualty predictions, as the majority of samples often represent incidents with no casualties. Nevertheless, the task of predicting the casualty count in terrorist attacks is challenging, entailing a multitude of factors, including inherent attributes and human-related elements such as the quantity of assailants and their psychological state.

Therefore, it is essential to develop more efficient machine learning methods to accurately predict casualties. Nonetheless, it is worth noting that there exists a significant gap in the literature with regards to the accurate identification of future terrorist activities, as well as the projection of critical variables such as the likelihood of causalities and utilization of specific weapon types. The existence of this gap in research highlights the significance of creating models and techniques utilizing deep learning and weighted ensemble techniques in order to accurately predict terrorist weapon types as well as the expected casualties. To this aim our proposed solution tackles the limitations observed in the existing literature, aiming to improve the accuracy and effectiveness of predicting terrorist activities and casualties. By leveraging this approach, we aim to provide valuable early warning and decision support for the management of terrorist attacks.

III. PROPOSED METHODOLOGY

In this paper we developed a predictive modeling framework based on advance ML algorithms for the prediction of number of causalities as well as an optimal weighted voting ensemble model for predicting the weapon type. The ensemble model comprises of RF, XGB and DT bagging classifier. While PSO algorithm is used to allocate optimal weights to the base classifiers so that the classification accuracy can be improved and total error can be minimized. Figure 1 shows the operational overview of the proposed methodology for weapons classification and causalities prediction.

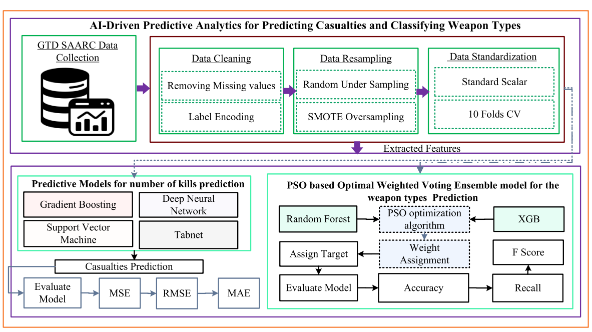


FIGURE 1. Proposed architecture for AI-driven predictive analytic framework for counter-terrorism.

Figure 1 illustrates that the GTD SAARC data-set is passed as input to the pre-processing module. The pre-processing module removes the missing values and performs label encoding, re-sampling, and data standardization. The next sections provide a comprehensive examination of the intricate aspects of data preparation methodologies. After completing the initial step of dataset preparation, the pre-processed data is fed into the ML models in order for training. The trained model is then used for prediction and classification tasks. Afterwards, a comprehensive review is conducted on the results of each model using performance assessment indicators.

A. PREDICTIVE MODELING

This study focuses on the development of appropriate machine learning models for the classification of weapons and the prediction of number of casualties. The subsequent subsections elucidate the intricacies of the machine learning models that have been constructed for the purpose of predictive modelling and analytics.

1) GRADIENT BOOSTING

GB is a machine learning technique that leverages the concept of ensemble learning to construct a robust model by aggregating multiple weak learners. This technique has the potential to be employed in both classification and regression tasks. The process of GB involves initially computing the mean of the target variable, followed by the computation of the residual error. Subsequently, construct a DT and utilise all trees to make predictions for the target variable. The calculation of residuals is repeated, and this iterative procedure continues until the termination condition is satisfied. The n-estimators parameter is utilised to specify the quantity of trees in the algorithm. The learning rate parameter governs the rate at which the algorithm converges. The max features parameter is employed to determine the amount of splits, while the max depth parameter is utilised to establish the depth of the tree.

2) SUPPORT VECTOR MACHINE

Support Vector Regression (SVR) is a regression model that applies the same ideas as SVM for classification tasks. SVR is a computational algorithm that addresses both linear and non-linear problems. The parameters of SVM are denoted as C and *Gamma*. The parameter C plays a crucial role in determining the balance between achieving accurate classification of training points and the creation of smooth decision boundaries. The parameter *Gamma* is responsible for determining the extent to which a single training example can exert its influence. The model is trained using the RBF kernel with $C = 0.8$ and $gamma = 0.01$.

3) DEEP NEURAL NETWORK

Deep Neural Networks (DNNs) represent a sophisticated variant of artificial neural network structures characterized by their intricate architecture, featuring multiple hidden layers positioned between the input and output layers. Within the DNN model, data flows from initiation at the input

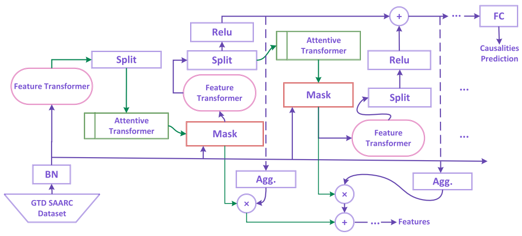


FIGURE 2. **Proposed Tabnet architecture for AI-driven predictive analytics for predicting casualties.**

layer to culmination in the production of the network's ultimate output at the output layer. This study employed a fully connected neural network design, often referred to as a feed-forward neural network, in which each neuron within a given layer establishes connections with every other neuron in the subsequent layer. The core objectives of our research revolved around harnessing the capabilities of the machine learning models for the tasks of classifying weapon types and predicting casualties. To optimize the network's performance, we leveraged the Adam optimizer, a highly efficient tool for fine-tuning the network's weight parameters during the training process. Specifically, we applied the categorical cross-entropy loss function for weapon type classification, while the mean squared error loss function served as our choice for casualty prediction. In the hidden layers of the network, we implemented the Rectified Linear Unit (ReLU) activation function, a widely embraced choice in the realm of deep learning. In the context of weapon type classification, we strategically deployed the Softmax activation function in the output layer to enhance the precision of weapon classification, thereby enabling the assignment of probabilities to distinct weapon classifications. In contrast, for casualty prediction, we judiciously adopted the linear activation function at the output layer, facilitating the generation of continuous numerical predictions.

4) TABNET MODEL

The TabNet model has the ability to map spatial and temporal correlations in terrorism data. Figure 2 shows the architecture of TabNet model. The architecture of TabNet, an encoding system designed for GTD SAARC data, is illustrated in Figure 2. This approach effectively handles categorical data by utilizing trainable embeddings and capitalizes on the inherent value of raw numerical features. Rather than employing a uniform feature normalization method, TabNet utilizes batch normalization (BN). Each decision step in TabNet operates on the same D-dimensional features. TabNet follows a sequential multi-processing encoding procedure with N steps. At each step, the processed information from the previous (i-1)th stage is used to determine the most relevant features. The final step involves aggregating the processed feature representation to make informed decisions. The concept of top-down attention, successfully applied in processing visual and text input [28] and reinforcement learning [29], serves as inspiration for its application in the sequential format. By efficiently identifying a concise subset of relevant information within high-dimensional input, this method aims to improve connectivity and enable accurate predictions.

B. OPTIMAL WEIGHTED VOTING ENSEMBLE MODEL

Ensemble Learning is a machine learning technique that integrates the predictive capabilities of numerous classifiers to enhance overall performance. Ensemble Voting is a strategic approach that involves aggregating the outcomes of numerous classifiers through a voting mechanism. The Ensemble Weighted Voting approach involves assigning weights to each classifier based on their performance. In the context of ensemble voting, it is observed that each classifier makes an equal contribution. The utilization of a weighted voting approach is employed to allocate weights that are contingent upon the performance of the classifier. This study employs Random Forest, XGB, and Bagging Classifier as the base learners. The rationale behind utilizing RF, XGB is attributed to their superior performance when applied to Tabular data. The present work utilized the PSO technique in order to optimize the weights of the basic classifier. PSO is a meta-heuristic optimization technique that draws inspiration from the collective behavior observed in bird flocking.

The PSO algorithm involves working with a population of particles within a designated search space, aiming to identify the optimal solutions within the provided search space. Every individual element within a given search space is indicative of a potential solution and possesses both a velocity and a position. The velocity of a particle is indicative of the amplitude of its movement, while the position of the particle reflects the potential solutions being explored. The algorithm begins by initializing the population, ensuring that the initial solutions adhere to the specified constraints of the search space. The initial boundaries in this investigation were established as [0.1, 0.99], and the number of particles is set to 25. The maximum number of iterations is defined as 30. The primary goal of PSO is to determine the optimal weights for each individual classifier in an ensemble model, with the aim of maximizing overall performance. The objective function is specified in equation 1.

$$Fitnessvalue = \frac{1}{n} = \sum_{j=1}^{n} \text{Accuracy}^{j} \tag{1}$$

where n represents the total number of classifiers and accurately represents each candidate solution's accuracy performance. To find better solutions, the goal of this function is to maximized the prediction accuracy.

C. DATASET

The primary dataset utilized in this study is drawn from the GTD [30], which is a publicly accessible repository dedicated to cataloging and managing information related to terrorist incidents worldwide. This comprehensive database is diligently maintained by the National Consortium for the Study

Algorithm 1 Operational Overview of the Developed System

Require: GTD SAARC Dataset $data = (x_1, x_2, x_3, \ldots, x_n)$
Ensure: Weapon type and Number of Causalities

Initialization:
$inputdata \leftarrow$ Read_Data()
$inputdata \leftarrow$ Remove_Missing_Values($inputdata$)
$inputdata \leftarrow$ Label_Encoding($inputdata$)
$X, Y \leftarrow$ FeatureSplit($inputdata$)
$X, Y \leftarrow$ Data_Resampling(X, Y)
$Scaled\_Data \leftarrow$ Data_Standardization(X)
$testset, Trainset \leftarrow$ TrainTestSplit($Scaled\_Data, Y$)
$Models \leftarrow$ ModelParametersInitialization()
for each $model$ in $Models$ **do**
    $TrainedModel \leftarrow$ ModelTraining($Trainset$)
    $\hat{Y} \leftarrow$ ModelPrediction($testset$)
    Evaluate:
    $Accuracy \leftarrow \frac{TP+TN}{TP+FN+FP+TN}$
    $Precision \leftarrow \frac{TP}{TP+FP}$
    $Recall \leftarrow \frac{TP}{TP+FN}$
    $F\text{-}Score \leftarrow \frac{2 \times Precision \times Recall}{Precision + Recall}$
    $MSE \leftarrow \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2$
    $MAE \leftarrow \sum_{i=1}^{m}|y_i - \hat{y}_i|$
    $RMSE \leftarrow \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}$
end for

of Terrorism and Responses to Terrorism (START) [31], a distinguished research and educational institution situated at the University of Maryland in the United States. For the purposes of this research, we have chosen a temporal range spanning from January 1998 to December 2017, covering a period of two decades. The reason behind this specific timeframe is rooted in the significant upsurge in terrorism, particularly in Pakistan, following the 9/11 attacks in the United States. These attacks led to U.S. military intervention in Afghanistan, with profound repercussions for the entire region, particularly India, Pakistan, and Afghanistan.

A wide range of geographical areas and incidents are covered by the GTD, providing scholars with a comprehensive perspective on the dynamics of global terrorism. Specific information about individual instances of terrorism, including characteristics such as geographical coordinates, date and time and method of occurrence, number of casualties, and specified targets, is included in each dataset entry. Furthermore, extensive biographies of the individuals involved in these occurrences, encompassing their personal identities, organizational associations, driving forces, and ideological foundations, are provided by the information. Additionally, the armaments and techniques utilized in these attacks are systematically documented, thereby furnishing significant background information for understanding the operational approaches employed by terrorist organizations. To undertake a comprehensive analysis of the societal impacts of terrorism, a method of categorization that separates victims into discrete groups, including those who have lost their lives, sustained injuries, and been held as hostages, is employed by the GTD. Data related to coordinated attacks and other significant incidents is also encompassed by the dataset, therefore offering unique perspectives on the extent and complexity of terrorist actions. It is crucial to acknowledge that the data is obtained from a diverse range of recognized and authoritative sources, including news articles, government publications, and scientific research, thus ensuring its authenticity and dependability.

1) STUDY AREA

Pakistan, being a significant partner of the United States in the worldwide counter-terrorism efforts, underwent substantial consequences throughout this duration. Hence, the selected temporal scope provides a crucial perspective for analyzing the progression and ramifications of terrorism in the regions of Afghanistan, Pakistan, and India. Pakistan is geographically located in the northwestern part of South Asia, and it shares its borders with Afghanistan and Iran to the west, India to the east, and China to the northeast. The geographical location of Pakistan confers considerable geostrategic significance to the country, with an estimated land size of around 803,940 square kilometers. Pakistan's administrative framework comprises four provinces, namely Punjab, Sindh, Khyber Pakhtunkhwa (KPK), and Balochistan. Furthermore, the nation contains three administrative territories, specifically Gilgit-Baltistan, Azad Jammu and Kashmir (AJK), and the Federally Administered Tribal Areas (FATA). Significantly, the aforementioned administrative regions underwent amalgamation with the neighboring province of KPK in 2018, as a component of a substantial endeavor to restructure the administrative framework.

The primary subject of investigation in this scholarly study pertains to the region of South Asia, with particular emphasis on the countries of Pakistan, India, and Afghanistan. These three nations exemplify a spectrum of geographical and geopolitical heterogeneity. The strategic significance of this region is evident due to its inclusion of four provinces and numerous territories. India, situated to the east of Pakistan, is a vast nation including numerous states and union territories, each of which has a significant cultural heritage. Afghanistan, situated to the west of Pakistan and north of India, possesses a multifaceted historical narrative characterized by periods of conflict and geopolitical importance, alongside challenging geographical features such as the imposing Hindu Kush mountain range. The rationale for choosing these three countries is based on their significant exposure to the characteristics being examined, rendering them very pertinent to the objectives of this research endeavor.

D. DATA PREPROCESSING

In addition to the aforementioned data pre-processing procedures designed to ensure the dataset's suitability for machine learning model training, several crucial data preparation operations are conducted to achieve optimal model performance.

E. FEATURE ENCODING

Feature Encoding plays a pivotal role in the data preparation process, primarily concentrating on the conversion of string-based categorical variables into numerical representations. The primary aim is to make these variables compatible with various machine learning techniques that exclusively operate with numeric data. To achieve this, Python's label encoding method was utilized. It facilitated the transformation of attributes like country, region, attack type, and the target variable (weapon type) into numerical values. This encoding process effectively incorporated categorical features into the machine learning models, thereby improving their interpretability and usability.

1) DATA RESAMPLING

Addressing the issue of class imbalance emerged as a pivotal consideration within the realm of data preparation, particularly in scenarios where one class exhibited a significant disproportion relative to others. To effectively mitigate this concern, a meticulous deployment of data resampling techniques is executed. The two principal resampling methodologies, namely under-sampling and over-sampling, are extensively applied. Under-sampling involves the removal of data instances from the majority class, while over-sampling entails the purposeful addition of data instances to the minority class, constituting the core of our approach. In this specific study, the random under-sampling method is employed as a strategic means to rectify class imbalance, thereby bolstering the statistical robustness of our analyses. Additionally, we employed the Synthetic Minority Over-sampling Technique (SMOTE) to rectify the asymmetry observed in the distribution of class labels for different types of weapons. SMOTE adeptly generated synthetic data instances, strategically placed within the feature space to smooth out inconsistencies. This method not only addressed the disparity in class representation but also safeguarded the minority group against potential biases that might favor the majority. Consequently, the machine learning models utilized in our research were fortified in terms of their robustness and fairness.

2) DATA STANDARDIZATION

The final critical aspect of data preparation pertained to Data Standardization. Within the GTD SAARC dataset, the features exhibited variations in scale, which had the potential to impede the performance of certain machine learning algorithms. To mitigate this concern, the standard scalar standardization method was systematically applied. This process rescaled feature values, normalizing their ranges to prevent features with larger scales from exerting disproportionate influence over the modeling process. Data standardization played a pivotal role in ensuring that all input features contributed equitably to the learning process, ultimately resulting in more consistent and precise model predictions.

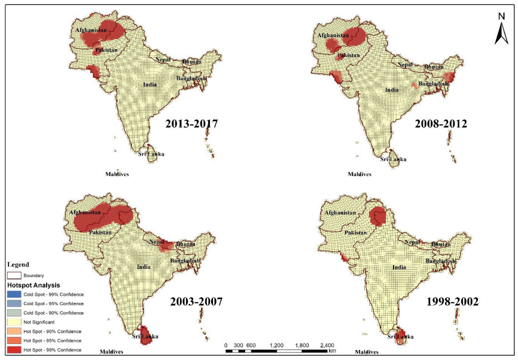


FIGURE 3. Distribution of terrorist attacks over time.

3) FEATURE SELECTION

The masking operation is carried out multiplicatively, according to $M[i] \cdot f$. In order to create the masks, an attentive transformer, as shown in Fig. 2, takes into account the processed features from the previous step, denoted as $a[i-1]$: $M[i] = \text{sparsemax}(P[i-1] \cdot h_i(a[i-1]))$ [32] introduced the sparse-max normalization method maps the Euclidean projection onto the probabilistic simplex to increase sparsity. This method has proven effective and fits in with the goal of selecting sparse features for enhanced interpretability. The goal of selecting sparse features for easier interpretation is consistent with this strategy and has demonstrated superior performance to others. It should be noted that the sum of each row in $M[i]$ should be 1, $\sum_{j=1}^{D} M[i]b[j] = 1$. As depicted in Figure 4, the trainable function h_i consists of a fully connected (FC) layer followed by batch normalization (BN). The prior scale term $P[i$ quantifies the usage of each feature in earlier stages and is computed as follows: $P[i] = Q_i \sum_{j=1}^{D}(\gamma - M[j])$ *gamma* is a relaxation parameter that, when set to 1, mandates the incorporation of a feature.

IV. ANALYSIS OF TERRORIST ATTACK DATA

This section presents a comprehensive data analysis of the terrorist data, with the goal of gaining significant insights into the how and when of the events that occurred.

A. ANALYSIS OF TERRORIST ATTACKS IN MOST AFFECTED COUNTRIES

Fig. 3 shows the number of affected countries by attacks. In this dataset, we provide an exploratory data analysis of the terror attacks from 1970-2017, finding the most affected countries, the most notorious groups, their motives, etc. From 2003-2007 the total number of terrorist attack hotspots rose significantly in Afghanistan mainly due to heavy resistance from Taliban and serious unrest in the country after the United States and its coalition forces declared war on Afghanistan in late 2001 following 9/11 attacks on the American soil that killed nearly 3,000 people. The highest number of hot-spots can be seen in Afghanistan and Pakistan, followed by Sri Lanka and Nepal (Figure 1). Militant groups fighting against the U.S. in Afghanistan also carried out terrorist attacks in Pakistan for being an ally of Washington in Afghanistan war. A surge in the terror attacks can also be observed in

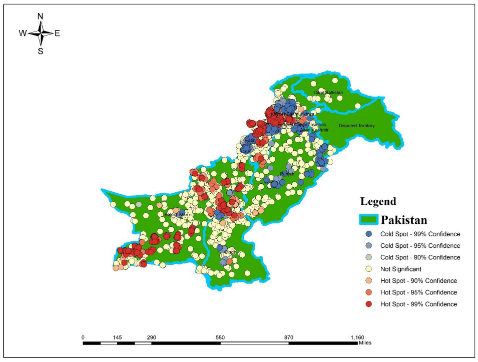


FIGURE 4. High risk areas of terrorist attacks using hotspots analysis.

the disputed Kashmir territory between India and Pakistan. On the other hand, almost the entire Sri Lanka is filled with terrorist attack hotspots amid full scale war between Sri Lankan government and LTTE during the period. The hotspots also expanded in Nepal as a result of a civil war between the communist Party of Nepal (Maoist) and the Nepal government.

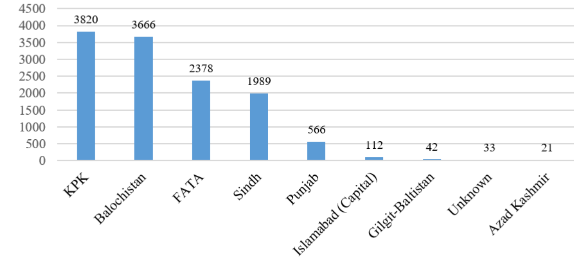
B. ANALYSIS OF HIGH AND LOW CONCENTRATION AREAS OF TERRORIST ATTACKS

The identifications of hotspots of terrorist attacks is highly useful since it highlights regions more susceptible to terrorism. Figure 4 represents the hot spot map of terrorist attacks in Pakistan. The dark red color represents areas with highest number of terrorist incidents, while dark blue exhibits regions with low incidents count. Terrorist attacks are more concentrated in the KPK and Federally Administered Tribal Area (FATA), which lie in north-western part of the country along the Pakistan-Afghanistan border. FATA was a semi-autonomous tribal region until it merged with the adjacent province KPK in 2018. However, we will discuss the two regions separately, as they are merged one year after the study duration selected for this research.

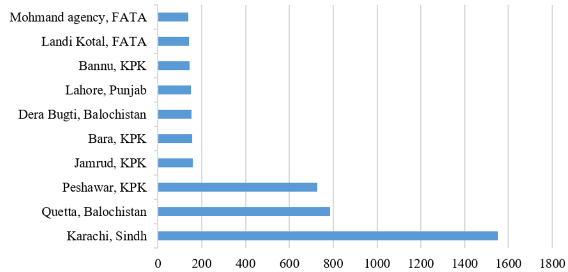
Post 9/11, Pakistan assisted U.S. in fighting militant groups in Afghanistan and as a result those groups later turned against Pakistan and carried out deadly attacks in the areas adjacent to Pak-Afghan border where they use to have active presence, and the main reason behind the high concentration of attacks in FATA and KPK. The second highest number of hotspots can be seen in Balochistan, predominantly along its southwestern coast while some are scattered in the north of the province. Comparatively, the highest the number of terrorist incident cold spots can be observed in KPK and FATA, followed by Punjab, while very few can be spotted in Balochistan and Sindh.

C. RANKING REGIONS AND CITIES BASED ON THE TERRORIST ATTACKS COUNT

To further determine the severity of attacks in different regions across Pakistan, frequency distributions analysis is performed. Figure 5 demonstrates KPK (Figure5a) is the



(a) Region-wise terrorist attacks count in Pakistan (1998-2017)



(b) Top ten cities with highest number of terrorist attacks (1998-2017)

FIGURE 5. Region-wise terrorist attacks analysis.

most affected province in terms of total number of terrorist attacks happened between 1998-2017, followed very closely by Balochistan.

Third and fourth highest number of attacks happened in FATA and Sindh, respectively as shown in Figure 5a. Despite being the most populous province of the country, Punjab stands at fifth in the list, while Gilgit Baltistan and Azad Kashmir are the least affected regions. Existing literature suggests that the probability of conflict spillover escalates with a bigger refugee inflow from adjacent war-torn countries. After United States invaded Afghanistan in 2001 following 9/11 attacks on the American soil, millions of Afghan citizens took refuge in Pakistan and most of them were given shelter in KPK. Moving forward, Figure 5b. exhibits the ten most affected cities in terms of terrorist attacks count. The highest number of terrorist incidents occurred in Karachi, which is the most populous city of Pakistan. Densely populated regions appear more vulnerable and typically prone to terrorist activities than less populated regions. Additionally, lethal incidents are expected to occur within or close to big cities as they have an effect on a larger audience. Quetta received the second highest terrorist attacks, while Peshawar is the third hardest hit city by terrorism incidents. The rest of the cities in the list have comparatively experienced considerably fewer terrorist attacks. Among the top-ten most affected cities, four are situated in KPK province, two belong to Balochistan and FATA respectively, while both Sindh and Punjab have just one city each in the list.

D. TOP MILITANT GROUPS BASED ON TOTAL ATTACKS CARRIED OUT

While referring to Pakistan Security Report (2008-2015) [33] reported that TTP and its affiliates are responsible

TABLE 1. Frequency distribution of terrorist groups based on total incidents (1998-2017).

Terrorist Group Name	Q	P
Unknown	9497	77.857026
Tehrik-i-Taliban Pakistan (TTP)	1335	10.944417
Baloch Republican Army (BRA)	313	2.5659944
Baloch Liberation Front (BLF)	185	1.5166421
Baloch Liberation Army (BLA)	181	1.4838498
Lashkar-e-Jhangvi	134	1.0985407
Lashkar-e-Islam (Pakistan)	124	1.0165601
Khorasan Chapter of the Islamic State	96	0.7870143
United Baloch Army (UBA)	89	0.7296278
Sindhu Desh Liberation Army (SDLA)	57	0.4672897
Taliban	54	0.4426955
Taliban (Pakistan)	38	0.3115265
Lashkar-e-Balochistan	32	0.2623381
Militants	32	0.2623381
Al-Qaida	31	0.25414
Total	12198	100

for 88% of the terrorist attacks in the country. While only 12% belonged to political and separatist terrorism. The group with the second highest number of terrorist attacks in the list is BRA, an armed separatist group based in Balochistan, followed by BLF and BLA based in Afghanistan. Majority of militant organizations listed in the Table 1 are sectarian groups that are involved in the sectarian violence, which is usually inspired by difference between various sects of one religion within a nation.

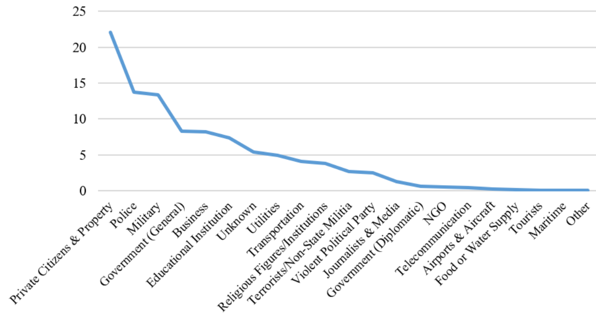
E. ANALYSIS OF TARGET TYPES

Private citizens and property is the most vulnerable target type accounting for about 22% of the overall incidents, and the apparent reason is that they are soft targets for terrorists and easy to hurt. Police and military are third and fourth most affected types receiving about 13% of the total attacks. Government, businesses, educational institutions also among the leading target types accounting for about 8% of total attacks (Figure 6a and 6b). Maritime, tourists, food supply, airport and air-crafts fall under the least affected target types.

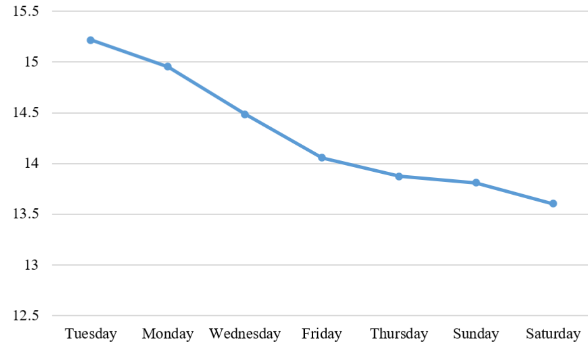
Total terrorist attacks occurred in Pakistan between 1998-2017 are almost evenly distributed among weekdays i.e. between 13.5 percent to 15.5 percent. The highest number of attacks are carried out on Tuesday (15.2 percent), followed by Monday (14.9 percent), Wednesday (14.4 percent) and Friday (14.05 percent), while the least number of attacks occurred on Saturday (Figure 6b).

F. ANALYSIS OF HIGH AND LOW CONCENTRATION AREA OF CASUALTIES AND TERRORIST GROUPS

Figures 7 to 12 shows the number of terrorist activities per year. Analyzing the number of terrorist group activities each year can be an important aspect of understanding the overall trends and patterns of terrorism and identifying potential changes in the threat landscape. By examining the frequency of terrorist incidents over time, we can see that



(a) Percentage distribution of target types (1998-2017)



(b) Terrorist attacks pattern on weekdays (1998-2017)

FIGURE 6. Analyzing patterns in terrorist attacks.

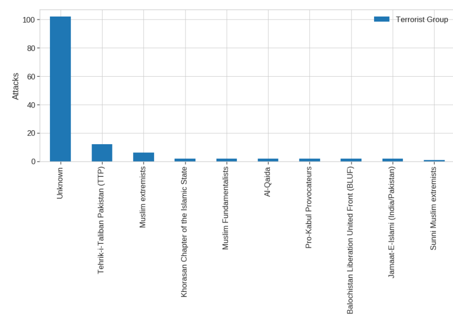


FIGURE 7. Analysis of terrorist activities in Islamabad capital territory.

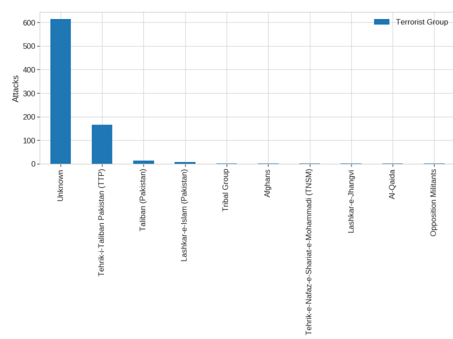


FIGURE 8. Analysis of terrorist activities in KPK.

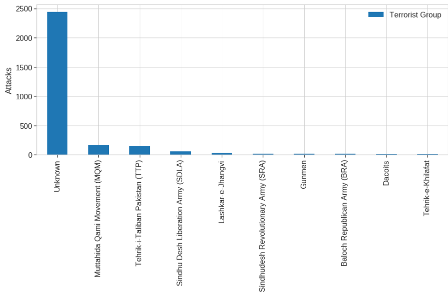


FIGURE 9. Analysis of terrorist activities in Sindh.

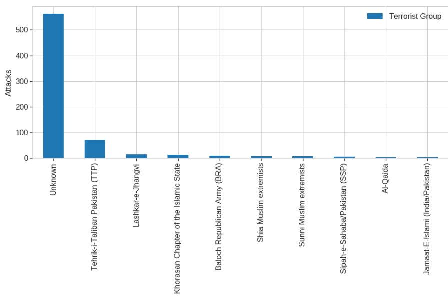


FIGURE 10. Analysis of terrorist activities in Punjab.

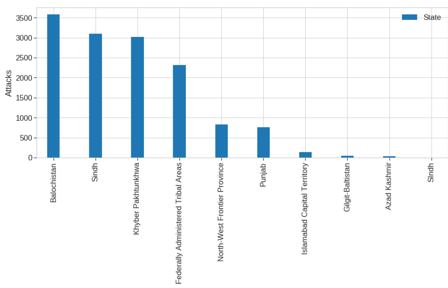


FIGURE 11. Analysis of terrorist activities in most affected states in Pakistan.

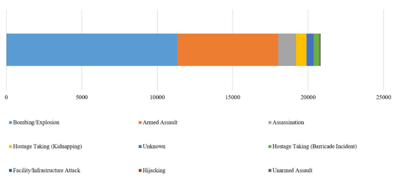


FIGURE 12. Frequency distribution of casualties according to attack type (1998-2017).

TTP and unknown terrorist group remains on the top of the list of attacking groups, while Baluchistan Sindh and KPK are the top affected provinces. The figure also represents the total casualties against each attack type. The highest number of casualties, more than 11,000, occurred because of bombing/explosion. Further, the attacks involving explosions represent significant mass killings seen in terrorist activities worldwide.

The highest number of casualty hotspots with confidence level of more than 95 % can be seen in KPK and FATA, mainly along the Pakistan-Afghanistan border (Figure 13). Militant organizations managed to roam freely across the Pak-Afghan border by exploiting the transit trade deals between the two countries. The selected routes designated for trade between Pakistan and Afghanistan pass through major cities of KPK, FATA and Balochistan province, where terrorist groups, such as Tehrik-i-Taliban Pakistan (TTP), use to have active presence, which is why areas along the Pak-Afghan borders ended up with highest terrorist attack and casualty count. Punjab has the second highest casualty hotspots that are scattered throughout the province, while Sindh and Gilgit-Baltistan have the lowest casualty count with no hotspots. Comparatively, casualty cold spots are more concentrated in Balochistan, followed by Sindh, while some cold spots in the south of Punjab.

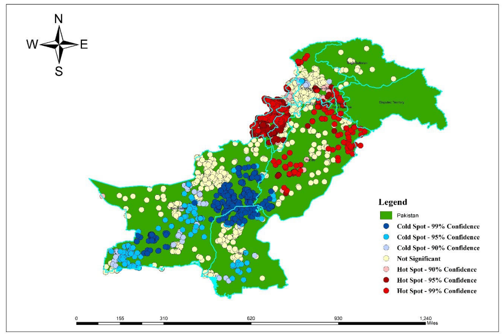
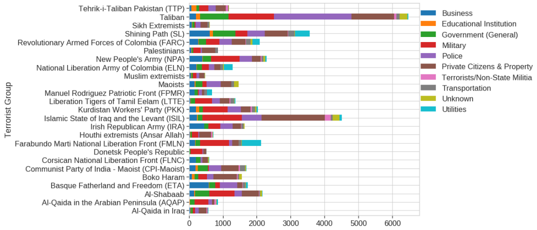


FIGURE 13. Hotspot distribution map of casualties (1998-2017).

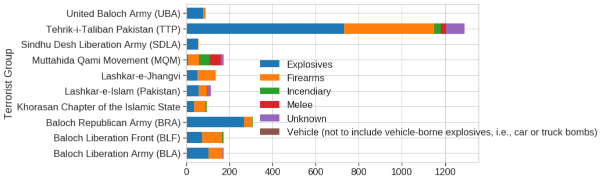
In Figure 14 and 15 we presents a detailed analysis of weapon types attack types and favourite targets. The figure represents the frequency distribution of various weapons used by terrorist groups. As evident from the figure, Explosives is the most widely used weapon type selected by terrorist groups in the region, followed by Firearms. However, the third highest percentage of weapon types remain Unknown. Meanwhile, Incendiary and Melee accounted for the fourth and fifth most widely used weapon types, while other categories were insignificant. Figure 14a presents south-Asia Favorite targets of top 25 most active terrorist groups. The 25 most active terrorist groups in South Asia and their top targets have been categorized. It can be analyzed from the figure that private citizens and properties received the highest percentage of attacks in the region. Moreover, police, military, and government are also among the top five favorite targets of these groups. Meanwhile, business, education, and utilities received relatively lower number of attacks.

Figure 14b shows the frequency distribution of different weapons used in Pakistan. Explosives were the most widely used weapon type, followed by firearms. Meanwhile, Melee, Incendiary, and unknown categories were insignificant. If we look at the groups, the Tehreek-e-Taliban-e-Pakistan carried out the highest number of explosives attacks in the country, followed by the Baloch Republican Army.

Figure 15a shows that "Bombing/explosion" is the most famous attack type used by terrorists in Pakistan. Bombings

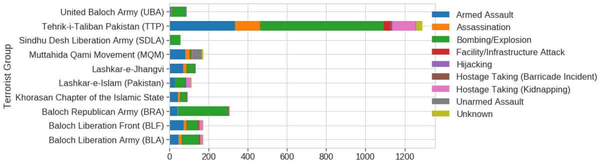


(a) South-Asia Favorite targets of top 25 most active terrorist groups

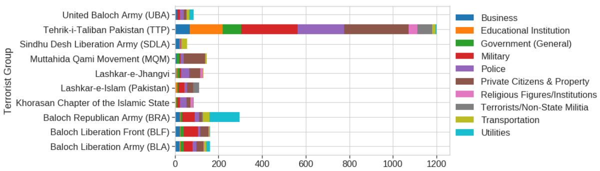


(b) Weapons used by various terrorist groups in Pakistan

FIGURE 14. South-Asia favorite targets and weapons used in Pakistan.



(a) Attack type by various terrorist groups in Pakistan



(b) Favorite targets of terrorist groups in Pakistan

FIGURE 15. Attack type and favorite targets in Pakistan.

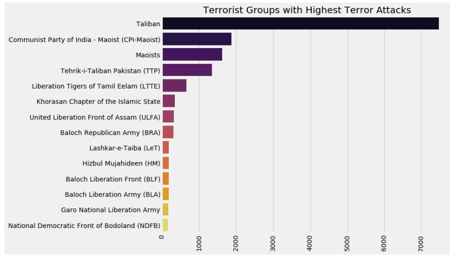


FIGURE 16. Most notorious groups with highest terrorist attacks in the study region.

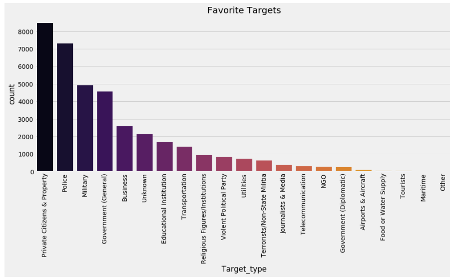


FIGURE 17. South-Asia favorite targets count.

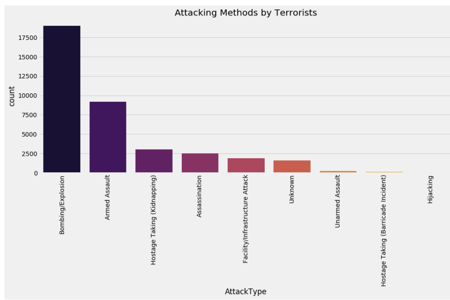


FIGURE 18. Analysis of terrorist attack methods.

have been widely used by terrorists around the world due to their devastating effects as they are highly lethal with high accuracy. Meanwhile, Armed assault is the second most widely used attack type, followed by Assassination. The remaining attack types were relatively insignificant. As evident in Figure 15b, private citizens and property remained the most vulnerable target type, since they are soft targets for terrorists and easy to hurt. Moreover, military and police were also among the top three targets of terrorist groups in the country. In comparison, government, businesses and educational institutions were among the less targeted categories.

Figure 16 shows the most Notorious Groups with highest terrorist attacks in the study. The figure also depicts the south-Asia favorite targets and attack method (Figure 17). From the analysis Taliban are the most notorious group while private citizen and property is the favorite target. While bombing and explosion remains the favorite attack method (Figure 18). These insight help analysts to identify periods of increased or decreased activity and assess whether certain counter-terrorism efforts or other external factors may have had an impact. They can also use this information to predict future trends and adjust strategies accordingly. Additionally, tracking the number of terrorist activities each year can be useful for comparing the threat levels in different regions and countries and identifying areas that may require more attention or resources. It can also help to prioritize counter-terrorism efforts and allocate resources more effectively.

V. EXPERIMENTS AND RESULTS

The findings from our correlation analysis on the GTD dataset are presented in Figure 19. This graphical representation includes a correlation matrix that effectively communicates the correlation coefficients, which serve as crucial indicators of the degree of linear relationship between each pair of variables in the data-set. These correlation coefficients span a range from -1 to $+1$, and their interpretation is pivotal in understanding the data. Specifically, a positive correlation coefficient suggests a strong tendency for two variables to move in sync; when one variable increases or decreases, the other follows suit. In simpler terms, when we observe a

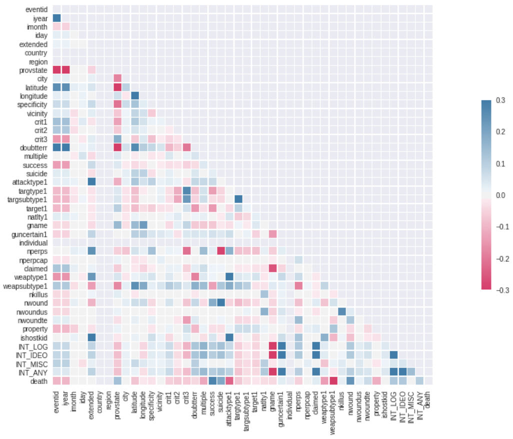


FIGURE 19. **Correlation analysis of the given data-set.**

positive correlation between variables X and Y, it means that high values of X consistently align with high values of Y, and low values of X consistently correspond to low values of Y. Contrarily, negative correlation is evident when a high value of X is associated with low values of Y and vice versa. This understanding of correlation is essential for gaining insights into the relationships within the data-set.

A. PERFORMANCE EVALUATION MEASURES

1) EVALUATION METRICS FOR WEAPON TYPE CLASSIFICATION

F score, accuracy, recall, and precision metrics are used to evaluate the performance of each method. The accuracy metrics demonstrate that how efficiently our model classifies the data points's among all classes. Precision metrics are used to find how correctly our model is classifying the positive class. Recall that metrics are used to find the performance percentage of actual positive from all the predictions. F-score is a Harmonic Mean between the Recall and Precision.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{F-Measure} = \frac{2 \cdot (Precision \cdot Recall)}{Precision + Recall} \tag{5}$$

2) EVALUATION METRICS FOR CASUALTIES PREDICTION

Mean Square, Root Mean Square, and Mean Absolute Error are used to evaluate the prediction model's performance for causalities prediction. The MSE is calculated to measure how the predicted value is nearest to the actual value. While RMSE measures the average magnitude of a difference between a data point's actual and predicted value. However, MAE also measures the average magnitude of an actual and predicted value without considering the direction of the error.

TABLE 2. **Experimental results without re-sampling.**

Method	Accuracy	Precision	Recall	F score
GB	0.8347	0.834	0.835	0.8255
DNN	0.9049	0.9146	0.8949	0.905
Tabnet	0.907	0.909	0.9000	0.902
WVC	0.9134	0.9125	0.913	0.910

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{6}$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{7}$$

$$sMSE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i| \tag{8}$$

B. WEAPON TYPE CLASSIFICATION RESULTS

Table 2 presents the empirical results obtained from an investigation into machine-learning models tasked with the classification of weapon types, all without the implementation of data re-sampling techniques. The exclusion of data re-sampling is pivotal, as it simulates real-world scenarios where class imbalances often exist and allows for an assessment of model performance under these conditions. Data re-sampling is a common preprocessing method in machine learning used to mitigate imbalances within datasets. The experimental findings reveal compelling insights into the comparative predictive performance of the models under scrutiny. A noteworthy discovery is the exceptional performance of PSO based weighted voting classifier. This particular model demonstrates superior predictive capabilities when juxtaposed with the other machine learning models considered in this study. The PSO-based classifier leverages optimization principles inspired by swarm behavior, exhibiting robustness in tackling the complexities associated with weapon type classification tasks. Conversely, the results indicate that GB, a well-established ensemble learning technique, displays lower predictive accuracy compared to its counterparts. This observation prompts a deeper inquiry into the factors contributing to the reduced performance of the GB model, with potential implications for its optimization in the context of weapon type classification. These findings contribute substantively to the field of machine learning, particularly concerning its applications within the realm of security and defense. The emphasis on real-world conditions and the commendable performance of the PSO-based classifier underscore its viability for practical applications, while the under-performance of the GB model calls for further investigation and potential enhancements to bolster its efficacy in similar classification tasks.

The results of our investigation into machine-learning models entrusted with the classification of weapon types

TABLE 3. Experimental results with under and over resampling.

Method	Accuracy	Precision	Recall	F score
GB	0.877	0.860	0.8767	0.8771
DNN	0.8827	0.8976	0.8703	0.884
Tabnet	0.821	0.810	0.820	0.825
WVC	0.9181	0.919	0.9181	0.9184

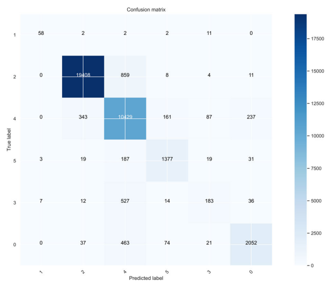


FIGURE 20. Results without resampling WVC.

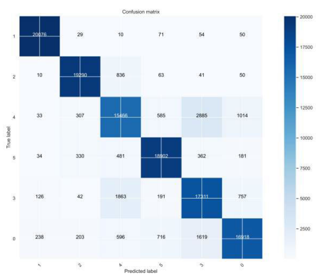


FIGURE 21. Results with resampling GB.

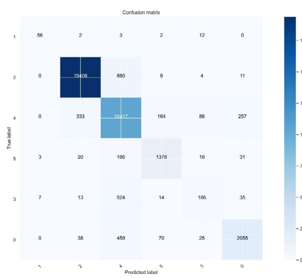


FIGURE 22. Results with resampling WVC.

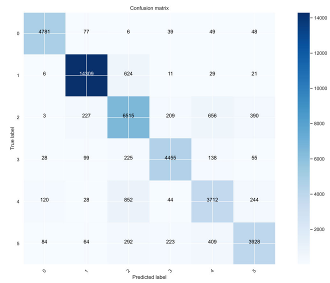


FIGURE 23. Results with under and over resampling GB.

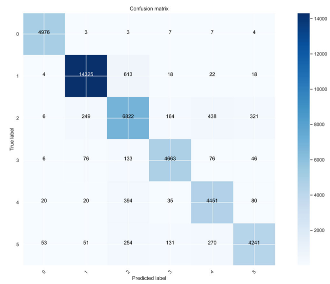


FIGURE 24. Results with under and over resampling WVC.

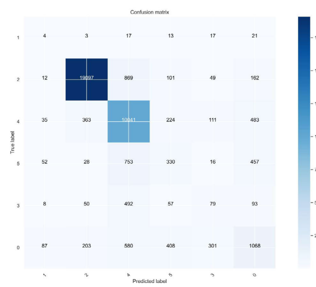


FIGURE 25. Results without resampling GB.

are summarized in detail in Table 3. Notably, these outcomes were calculated using both under-sampling and over-sampling techniques. Incorporating these sampling strategies is essential for a comprehensive evaluation of model performance, as they address the challenges presented by imbalanced class distributions that are frequently encountered in real-world scenarios.

The quantitative analysis of the machine learning models under these sampling techniques reveals a noteworthy trend: the performance of the models improves significantly when under-sampling and over-sampling techniques are employed as shown in Figure 20 to 25. The numbers 0 to 5 denotes the weapon types namely Chemicals, Explosives, Firearms, Incendiary, Melee and Unknown. This observation demonstrates the effectiveness of these our developed weighted voting classifier in mitigating the negative effects of class imbalances in the data-set. The improved model performance has far-reaching impacts for the practical applicability of machine learning in the context of weapon type classification, where impartial and reliable predictions are of the utmost importance. The significance of addressing class imbalances through sampling techniques is highlighted by these findings, which contribute considerably to the advancement of machine learning applications in security and defense. The improved performance of the models employing these strategies demonstrates their applicability in real-world situations, where class imbalances are frequently prevalent.

In Table 4, we present the experimental outcomes resulting from our investigation of machine-learning models customized for the classification of weapon categories using the SMOTE. This method employs the generation of synthetic samples to address class imbalance, a prevalent issue in real-world data-sets. The results clearly demonstrate that SMOTE oversampling significantly improves model performance,

TABLE 4. Experimental results with over resampling.

Method	Accuracy	Precision	Recall	F score
GB	0.8868	0.88779	0.8870	0.8867
DNN	0.9055	0.9160	0.8964	0.9061
Tabnet	0.8353	0.8390	0.8352	0.8355
WVC	0.95098	0.95101	0.9510	0.9508

TABLE 5. Causalities prediction for Pakistan.

Method	MSE	RMSE	MAE
SVR	39.37	6.27	1.62
GB	36.85	6.07	2.17
DNN	33.64	5.80	2.17
TabNet	40.94	6.52	2.065

surpassing both the absence of sampling and traditional over- and under-sampling techniques.

Remarkably, our findings demonstrate the remarkable accuracy of the PSO-based weighted voting classifier, which achieves a 95% accuracy rate. This performance considerably exceeds that of alternative prediction models, making the PSO-based method the best performer in this context. These results not only demonstrate the effectiveness of SMOTE oversampling, but also the potential of the PSO-based weighted voting classifier for weapon classification tasks. These findings have important implications for the field of machine learning, particularly in the security and defense domain, where accurate and objective predictions are of the utmost importance. The demonstrated benefits of SMOTE oversampling support its value as a valuable instrument for addressing class imbalance issues, while the exceptional performance of the PSO-based weighted voting classifier indicates its potential for use in real-world scenarios.

C. EXPERIMENTAL RESULTS OF CAUSALITIES PREDICTION

In Table 5 we present the results of our experimentation with machine learning models created to predict casualties in Pakistan. The experimental outcomes provide essential insights into the predictive capabilities of these models. Compared to other machine learning models, the Deep Neural Network (DNN) model has significantly lower prediction errors, as measured by Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). These results demonstrate the accuracy with which the DNN model predicts casualties in Pakistan. Lower prediction errors indicate a better fit between the model's predictions and the actual casualty data, indicating the model's superior performance in this predictive task. In contrast to the other machine learning models evaluated, the Tabnet model exhibits a relatively larger prediction error. This indicates that the Tabnet model may not perform as well in this particular prediction task as the DNN model or the other models being considered. The greater prediction error may be indicative of limitations or sub-optimal predictive patterns within the architecture of the Tabnet model.

TABLE 6. Causalities prediction for India.

Method	MSE	RMSE	MAE
SVR	7.29	2.70	0.96
GB	6.89	2.62	1.22
DNN	32.54	5.70	1.52
TabNet	52.73	7.46	1.56

These findings have substantial implications for the implementation of machine learning in the field of casualty prediction, particularly in Pakistan. The DNN model's superior performance makes it a promising instrument for accurately forecasting casualties, which could aid disaster preparedness and response efforts. In contrast, the higher prediction error observed in the Tabnet model necessitates additional research and optimization efforts if this model is to be effectively deployed in similar predictive scenarios.

The Table 6 summarizes the results of our investigation into machine learning models designed to predict casualties in India. These experimental findings shed light on the predictive capabilities of the models under consideration. In comparison to alternative machine learning models, the GB model demonstrates significantly lower prediction errors in terms of MSE, RMSE, and MAE.

The GB model's superior performance demonstrates its ability to generate accurate predictions of casualties in the Indian context. The reduced prediction errors indicate a strong correlation between the model's predictions and the actual casualty data, highlighting the model's potential utility for predictive tasks within the field of casualty forecasting.In contrast, the Tabnet model exhibits a relatively higher prediction error than the other machine learning models considered in this research. This increased prediction error suggests that the Tabnet model may be less suited for predicting casualties in India. If the Tabnet model is to be utilized effectively in similar predictive scenarios, the observed performance disparity may require additional analysis and possibly model-specific optimizations. These results have significant implications for the implementation of machine learning techniques in the field of casualty prediction, particularly in India. The GB model's remarkable accuracy makes it a promising instrument for precise casualty forecasting, which could prove invaluable for disaster preparedness and response efforts. In the meantime, the elevated prediction error observed in the Tabnet model highlights the significance of model selection and optimization for achieving optimal predictive performance in casualty prediction tasks.

The prediction errors of machine learning models are described in Table. Several machine learning models have been used to attempt to estimate casualties in Afghanistan, and their respective prediction errors are displayed in Table 7. MSE, RMSE, and MAE are used to quantify these prediction errors and provide a thorough evaluation of model performance in this setting. One model that stands out from the rest is the GB model, which regularly has less prediction errors than the other models. This finding

TABLE 7. Causalities prediction for Afghanistan.

Method	MSE	RMSE	MAE
SVR	42.15	6.49	2.63
GB	38.47	6.20	3.04
DNN	85.25	9.23	3.19
TabNet	86.58	9.55	3.33

demonstrates how well the GB model can estimate casualties in Afghanistan. The GB model's superior performance in fatality forecasting is indicated by the fact that its prediction errors are smaller than those of competing models. We found that the Tabnet model is more prone to making incorrect predictions than the other machine learning models we tested. The increased prediction inaccuracy raises concerns about the Tabnet model's suitability for use in the challenging setting of fatality prediction in Afghanistan.

Insights into the effectiveness of machine learning algorithms for casualty prediction have been provided, and these findings are especially relevant when considered in the context of Afghanistan. The GB model has shown remarkable performance, making it a valuable tool for accurate casualty forecasts that might be used in disaster management and response. However, the Tabnet model's larger prediction error highlights the necessity of individualized model enhancements to optimize its performance in the context of fatality prediction jobs in Afghanistan.

D. DISCUSSION

The experimental results from the machine learning models in weapon classification and casualty prediction highlight several key strengths crucial for counter-terrorism and security applications. Firstly, the exceptional performance of the PSO-based classifier in weapon classification demonstrates its robust pattern recognition capabilities, making it highly effective for identifying complex and nuanced patterns in security-related data. The success of under and over-sampling techniques, particularly with the PSO model, illustrates the models' adaptability in handling imbalanced datasets, a common challenge in real-world scenarios. Furthermore, the notable accuracy of the DNN model in casualty prediction, especially in certain regions, indicates its strong potential in providing precise and reliable forecasts, essential for planning and response in security operations. These models' capabilities in processing large and varied datasets efficiently and accurately highlight their significant role in enhancing data-driven decision-making processes in counter-terrorism efforts. Collectively, these strengths underline the transformative impact that advanced machine learning techniques can have in bolstering national and global security measures.

VI. CONCLUSION

In conclusion, this study addresses the pressing global issue of terrorism by harnessing the power of AI-driven predictive analytics. As the escalating threat of terrorist attacks has created a severe humanitarian and economic crisis. To combat this menace, we developed advanced machine learning techniques for forecasting terrorism activities, focusing on weapon type classification and casualty prediction. Our investigation, rigorously evaluated these machine learning models. Notably, the GB model demonstrated exceptional predictive capabilities, offering reliable and accurate casualty estimates for Afghanistan. This model's performance underscores its potential for applications in disaster management and response, where accuracy is paramount. Furthermore, we developed an optimal weighted voting ensemble classifier tailored for weapon type classification. This WVC, optimized using the PSO algorithm, enhances the accuracy of classifying weapon types in terrorist attacks. In summary, this research contributes significantly to counter-terrorism efforts by providing vital insights and model recommendations. The GB model, with its exceptional predictive power, stands out as a promising tool for addressing the complexities of terrorism-related challenges, while the WVC introduces an effective approach to weapon type classification. These findings collectively enhance global security and peace efforts.

REFERENCES

[1] S. P. Das, *Economics of Terrorism and Counter-Terrorism Measures*. New York, NY, USA: Springer Books, 2022.
[2] M. Tahir, U. Burki, and T. Azid, "Terrorism and environmental sustainability: Empirical evidence from the MENA region," *Resour., Environ. Sustainability*, vol. 8, Jun. 2022, Art. no. 100056.
[3] R. Wided and A. A. Alfalih, "Extremism immunity through artificial intelligence networks: Extremism awareness and social intelligence," *Int. J. Data Netw. Sci.*, vol. 7, no. 1, pp. 341–356, 2023.
[4] O. A. Odeniyi, M. E. Adeosun, and T. P. Ogundunmade, "Prediction of terrorist activities in Nigeria using machine learning models," *Innovations*, vol. 71, pp. 87–96, Dec. 2022.
[5] M. Chaudhary and D. Bansal, "Open source intelligence extraction for terrorism-related information: A review," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 12, no. 5, p. e1473, 2022.
[6] J. A. Piazza, "Rooted in poverty?: Terrorism, poor economic development, and social Cleavages1," *Terrorism Political Violence*, vol. 18, no. 1, pp. 159–177, Mar. 2006.
[7] P. Agarwal, M. Sharma, and S. Chandra, "Comparison of machine learning approaches in the prediction of terrorist attacks," in *Proc. 12th Int. Conf. Contemp. Comput. (IC)*, Aug. 2019, pp. 1–7.
[8] S. A. Robbins, "Machine learning and counter-terrorism: Ethics, efficacy, and meaningful human control," Delft Univ. Technol., Delft, The Netherlands, Tech. Rep., 2021, doi: 10.4233/uuid:ad561ffb-3b28-47b3-b645-448771eddaff.
[9] J. K. Saini and D. Bansal, "Computational techniques to counter terrorism: A systematic survey," *Multimedia Tools Appl.*, pp. 1–26, 2023.
[10] O. A. Olabanjo, B. S. Aribisala, M. Mazzara, and A. S. Wusu, "An ensemble machine learning model for the prediction of danger zones: Towards a global counter-terrorism," *Soft Comput. Lett.*, vol. 3, Dec. 2021, Art. no. 100020.
[11] J. A, Vikas, J. Samuel, Bharanishwar, Srinivasan, G. Balan, and J. Premkumar, "Enhancing public safety through license plate recognition for counter terrorism through deep learning technique," in *Proc. 4th Int. Conf. Signal Process. Commun. (ICSPC)*, Mar. 2023, pp. 96–100.
[12] D. Jiang, J. Wu, F. Ding, T. Ide, J. Scheffran, D. Helman, S. Zhang, Y. Qian, J. Fu, S. Chen, X. Xie, T. Ma, M. Hao, and Q. Ge, "An integrated deep-learning and multi-level framework for understanding the behavior of terrorist groups," *Heliyon*, vol. 9, no. 8, Aug. 2023, Art. no. e18895.
[13] M. Alam and M. Tahernezhadi, "Machine learning and statistical analysis techniques on terrorism," in *Fuzzy Systems and Data Mining VI*. Amsterdam, The Netherlands: IOS Press, 2020, pp. 210–222.
[14] U. S. Onyekachi, T. Norbert, and E. D. Uzoamaka, "Data mining approach to counterterrorism," *Comput., Inf. Syst., Develop. Informat. Allied Res. J.*, vol. 11, no. 2, p. 15, 2020.

[15] J. Gill, "Discussion of 'estimating the historical and future probabilities of large terrorist events' by Aaron clauset and ryan woodard," *Ann. Appl. Statist.*, vol. 7, no. 4, pp. 1881–1887, 2013.

[16] A. Clauset and R. Woodard, "Estimating the historical and future probabilities of large terrorist events," *Ann. Appl. Statist.*, vol. 7, no. 4, pp. 1838–1865, Dec. 2013.

[17] J. Arnold, P. Halpern, M.-C. Tsai, and H. Smithline, "Mass casualty terrorist bombings: Comparison of outcomes by bombing type," *Prehospital Disaster Med.*, vol. 17, no. S2, pp. S47–S48, Dec. 2002.

[18] L. Zheng, C. Shen, L. Tang, C. Zeng, T. Li, S. Luis, and S.-C. Chen, "Data mining meets the needs of disaster information management," *IEEE Trans. Hum.-Mach. Syst.*, vol. 43, no. 5, pp. 451–464, Sep. 2013.

[19] L. Jiang, G. Kong, and C. Li, "Wrapper framework for test-cost-sensitive feature selection," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 3, pp. 1747–1756, Mar. 2021.

[20] R. Bridgelall, "Applying unsupervised machine learning to counterterrorism," *J. Comput. Social Sci.*, vol. 5, no. 2, pp. 1099–1128, Nov. 2022.

[21] Y. Gao, X. Wang, Q. Chen, Y. Guo, Q. Yang, K. Yang, and T. Fang, "Suspects prediction towards terrorist attacks based on machine learning," in *Proc. 5th Int. Conf. Big Data Inf. Anal. (BigDIA)*, Jul. 2019, pp. 126–131.

[22] X. Meng, L. Nie, and J. Song, "Big data-based prediction of terrorist attacks," *Comput. Electr. Eng.*, vol. 77, pp. 120–127, Jul. 2019.

[23] M. M. H. Khorshid, T. H. M. Abou-El-Enien, and G. M. A. Soliman, "Hybrid classification algorithms for terrorism prediction in middle east and north Africa," *Int. J. Emerg. Trends Technol. Comput. Sci.*, vol. 4, no. 3, pp. 23–29, 2015.

[24] M. I. Uddin, N. Zada, F. Aziz, Y. Saeed, A. Zeb, S. A. A. Shah, M. A. Al-Khasawneh, and M. Mahmoud, "Prediction of future terrorist activities using deep neural networks," *Complexity*, vol. 2020, pp. 1–16, Apr. 2020.

[25] Z. Li, D. Sun, B. Li, Z. Li, and A. Li, "Terrorist group behavior prediction by wavelet transform-based pattern recognition," *Discrete Dyn. Nature Soc.*, vol. 2018, pp. 1–16, 2018.

[26] F. Saidi and Z. Trabelsi, "A hybrid deep learning-based framework for future terrorist activities modeling and prediction," *Egyptian Informat. J.*, vol. 23, no. 3, pp. 437–446, Sep. 2022.

[27] R. Zhao, X. Xie, X. Zhang, M. Jin, and M. Hao, "Spatial distribution assessment of terrorist attack types based on I-MLKNN model," *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 8, p. 547, Aug. 2021.

[28] D. A. Hudson and C. D. Manning, "Compositional attention networks for machine reasoning," 2018, *arXiv:1803.03067*.

[29] A. Mott, D. Zoran, M. Chrzanowski, D. Wierstra, and D. J. Rezende, "Towards interpretable reinforcement learning using attention augmented agents," in *Proc. Adv. Neural Inf. Process. Syst.*, 32, 2019, pp. 1–10.

[30] B. Herre, "The global terrorism database: How do researchers measure terrorism?" Our World Data, 2023. [Online]. Available: https://ourworldindata.org/the-global-terrorism-database-how-do-researchers-measure-terrorism

[31] R. Q. Al-Fayez, M. Al-Tawil, B. Abu-Salih, and Z. Eyadat, "GTDOnto: An ontology for organizing and modeling knowledge about global terrorism," *Big Data Cognit. Comput.*, vol. 7, no. 1, p. 24, Jan. 2023.

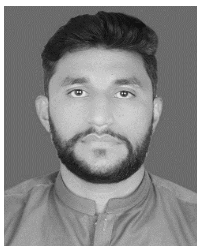
[32] A. Martins and R. Astudillo, "From softmax to sparsemax: A sparse model of attention and multi-label classification," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1614–1623.

[33] M. K. Khan and L. Wei, "When friends turned into enemies: The role of the national state vs. Tehrik-i-Taliban Pakistan (TTP) in the war against terrorism in Pakistan," *Korean J. Defense Anal.*, vol. 28, no. 4, pp. 597–626, 2016.

FAHAD ALI KHAN is currently a Doctoral Researcher with the College of Urban and Environmental Sciences, Northwest University, Xi'an, China. As a committed Ph.D. candidate, he delves into the intersecting realms of human geography, big data, and crime geography. His academic endeavors are characterized by a profound commitment to elucidating the spatial dynamics of criminal behavior and its interplay with societal constructs. His work stands out for its insightful examination of the geographic underpinnings of crime and its implications for urban development and policy planning.

GANG LI received the B.S. and Ph.D. degrees in physical geography from Lanzhou University in 2002 and 2008, respectively. He is a Professor of Human Geography of College of Urban and Environmental Sciences and Director of Joint Lab of Spatio-Temporal Big Data of Northwest University, P.R. China. He was a Visiting Scholar of the University of North Carolina at Chapel Hill, USA, from December 2009 to June 2010, and the University of Heidelberg, Germany, from March 2016 to February 2017, a Postdoc at the University of Graz, Austria, September 2010 to June 2011. His current research interests include human-environment interaction and space security, disaster geography and crime geography, spatio-temporal big data minning, digital humanities and computational social sciences.

ANAM NAWAZ KHAN received the B.S. and M.S. degrees in computer science from COMSATS University Islamabad, Islamabad, Pakistan, in 2016 and 2019, respectively. She joined various reputable academic institutions and later decided to expand her knowledge by continuing her academic career. In 2020, she moved to the Republic of Korea and started working as a Ph.D. Research Fellow with the Department of Computer Engineering, Jeju National University, Jeju, South Korea. She is currently working on integrating the AI-enabled IoT platforms and edge computing to provide autonomous intelligent services for smart environments. Her research interests include the development and optimization of artificial intelligence and machine learning-based applications in smart spaces.



QAZI WAQAS KHAN received the Bachelor of Science (B.Sc.) degree in statistics and computer science from the University of the Punjab, Lahore, Pakistan, in 2017, and the M.S. degree in computer science and the M.C.S. degree from COMSATS University Islamabad, Attock Campus, Punjab, Pakistan, in 2022 and 2019, respectively. He is currently pursing the Ph.D. degree in computer engineering with Jeju National University, South Korea. He was awarded Merit Scholarship in M.S. (CS). He was also awarded PEEF Scholarship in the B.Sc. and M.C.S. degrees. His research interests include applied machine learning, federated learning, data mining, natural language processing, computer vision, and the Internet of Things.

MYRIAM HADJOUNI received the Ph.D. degree in computer science from Paris XI (actual new name Paris Saclay) University, France, in 2012, the degree (Hons.) from Manouba University, Tunisia, and the M.Sc. degree (Hons.) from the Higher Institute of Management of Tunis, University of Tunis, Tunisia, in 2005. Currently, she is an Assistant Professor with the Computer Sciences Department, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Her research interests include information retrieval, data science, data analytic, big data, and artificial intelligence.

HELA ELMANNAI received the Ph.D. degree in information technology from SUPCOM, Tunisia. She is currently an Associate Professor with the Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Saudi Arabia. Her research interests include artificial intelligence, networking, blockchain, and engineering applications.