

EXPLAINABLE AI-BASED RISK ANALYTICS AND DECISION SUPPORT FOR SMART URBAN SAFETY

Vinit Kumar Shukla, Priyanka Kumari, Sofi Mittra, B.M. Reddama, S.T. Sadiya

Department of Computer Science and Engineering, Nagarjuna College of Engineering and Technology, India

Abstract

Conventional urban safety mechanisms like CCTV and manual alerts are fundamentally reactive. To enable proactive hazard mitigation, this research proposes a predictive spatio-temporal risk analytics framework paired with an Explainable AI (XAI) engine and an interactive decision-support dashboard. The system establishes a dynamic Compound Risk Scoring Index (RSI) combining temporal cycles, spatial isolation, illuminance grades, crowd density, and historical incident rates. An XGBoost classifier is trained against classical baselines (Logistic Regression, Decision Tree, and Random Forest) to categorize urban zones into granular risk tiers, while Tree-SHAP provides real-time local feature attributions to resolve the black-box dilemma. The architecture is deployed via an asynchronous REST API backend and Folium-based geospatial heatmaps. Evaluated on a procedural synthetic benchmark to demonstrate end-to-end feasibility, the proposed XGBoost model achieved 93.80% accuracy on a stratified split and 94.60% on a chronological split, with a Macro F1-Score of 0.9379-0.9456 and sub-millisecond inference latency, establishing a scalable, transparent foundation for proactive urban safety management.

Keywords:

Predictive Risk Analytics, Explainable Artificial Intelligence (XAI), Tree-SHAP, Gradient Boosting, Spatio-Temporal Modeling, Smart Cities

1. INTRODUCTION

1.1 CONTEXT AND BACKGROUND

Modern urban governance and the evolution of the “Safety 4.0” paradigm have accelerated the deployment of computing technologies and the Internet of Things (IoT) across metropolitan spaces and institutional campuses. Ensuring citizen security across expanding urban layouts requires processing high-velocity, multi-source environmental and demographic variables. However, most deployed surveillance architectures rely on human operators monitoring camera walls, or on automated vision algorithms that flag anomalies only when a physical altercation or security breach is actively occurring. These reactive systems fail to anticipate dangerous scenarios before they escalate, often leading to delayed emergency dispatches and suboptimal patrol allocation.

1.2 PROBLEM FORMULATION AND RESEARCH CHALLENGES

The primary limitation of traditional public-safety frameworks is their inability to capture fine-grained, dynamic spatio-temporal dependencies. Urban risk is non-stationary: an isolated parking lot may exhibit negligible risk during high-density daytime hours but become a high-risk zone during late-night hours under diminished lighting conditions. While classical statistical techniques and simple machine-learning models have

been applied to crime and accident records, they often produce opaque, purely numerical predictions. In high-stakes public-safety deployments, security personnel and administrators cannot blindly trust “black-box” outputs without understanding the underlying causal factors driving a given risk alert.

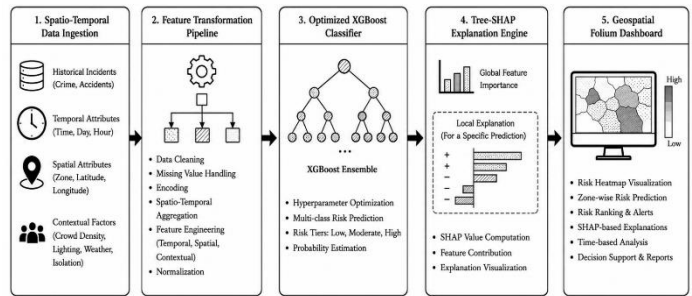


Fig.1. High-level schematic of the predictive and explainable urban safety framework.

1.3 PROPOSED CONTRIBUTIONS

To bridge these identified gaps, this paper presents an end-to-end predictive risk analytics and explainable decision-support framework, structured around a four-stage pipeline: Risk Formulation → ML Prediction → XAI Explanation → Decision Support. The key contributions of this research, aligned with each stage, are summarized as follows:

- Formulation of a Multi-Factor Risk Scoring Index (RSI): a mathematical scoring engine that models non-linear interactions among temporal intervals, spatial isolation, ambient lighting, crowd metrics, and incident histories.
- Gradient-Boosted Predictive Modeling: an optimized XGBoost multi-class classifier that categorizes urban zones into discrete operational risk tiers with high precision and recall.
- Operational Explainable AI Integration: Tree-SHAP is incorporated to compute exact Shapley feature attributions for every inference, providing transparent, localized explanations for security personnel.
- Interactive Decision Support Architecture: a modular, low-latency web architecture that renders dynamic geospatial heatmaps and automatically outputs contextual intervention policies.

2. LITERATURE SURVEY

2.1 VISION-BASED AND REACTIVE SAFETY SURVEILLANCE

The deployment of deep neural networks in surveillance has largely focused on real-time visual threat detection. Balaji et al. [1] proposed an automated night-surveillance system combining

Zero-DCE low-light enhancement, YOLOv10 for person detection, DeepMAR for gender classification, and YOLOv8 for violence detection. While achieving high detection accuracy in low-illumination environments, their system is strictly vision-dependent, computationally intensive, and inherently reactive. Sumanth et al. [2] introduced an edge-cloud framework utilizing Raspberry Pi nodes, MediaPipe for SOS gesture tracking, and Librosa CNNs for acoustic scream detection; although multi-modal, the framework lacks forward-looking risk anticipation and relies on active distress triggers. Similarly, Vasantha et al. [3] and [5] presented IoT-based alerting wearables and surveillance setups that depend heavily on continuous network connectivity and physical hardware maintenance.

2.2 SPATIO-TEMPORAL AND ENVIRONMENTAL RISK ANALYTICS

A second research trajectory focuses on modeling spatio-temporal historical records. Lin et al. [7] developed the CLwST model, using convolutional neural networks to extract spatial grid interactions and passing the reshaped sequences into Long Short-Term Memory (LSTM) networks to forecast traffic-accident probabilities; while capturing spatial heterogeneity, deep recurrent architectures remain computationally prohibitive for lightweight edge execution. In crime modeling, Bandpey et al. [4] integrated K-Means spatial clustering with a Stacking Regressor (Random Forest, XGBoost, and SVR) to analyze regional Maryland crime distributions; however, their model was non-spatial during inference and lacked an interactive dashboard for tactical resource planning. Kwon et al. [8] demonstrated, using artificial neural networks (ANN), that filtering out statistically insignificant micro-environmental variables significantly mitigates overfitting in theft-prediction models. At the municipal level, Tutak and Brodny [6] used multi-criteria decision-making (TOPSIS and EDAS) to evaluate Polish urban safety, though their approach was restricted to static, macro-level annual indicators. This methodological shift—from forensic, post-incident CCTV auditing toward proactive, explainable risk intelligence—is summarized in Fig.2.

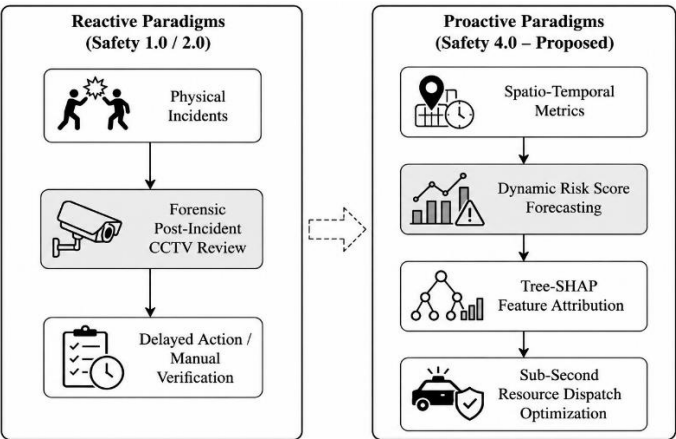


Fig.2. Methodological shift from forensic CCTV auditing to proactive, explainable risk intelligence.

2.3 EXPLAINABLE AI AND OPERATIONAL TRUST GAP

Despite the high accuracy of modern ensembles and neural models, deploying artificial intelligence in safety-critical domains introduces significant trust barriers. Latif et al. [9] applied Kernel SHAP and LIME to urban-mobility predictions, observing that standard model-agnostic explainers introduce substantial latency and incorrectly assume feature independence across correlated urban variables. Kaya et al. [10] systematically analyzed the vulnerabilities of post-hoc explanation methods, showing that poorly regularized explainers can produce fragile or misleading attributions. Goyal [11] addressed semantic compliance by engineering an ontology-based neuro-symbolic framework, though semantic reasoners introduced additional computational overhead.

Table.1. Comparative Summary of Existing Literature vs. Proposed System

Author	Methodology	Advantages	Limitations / Gaps
Balaji et al. [1]	Zero-DCE + YOLOv10 + DeepMAR	Robust low-light detection	Reactive; vision-only; hardware-heavy
Sumanth et al. [2]	Edge Pi + MediaPipe + Audio CNN	Multi-modal real-time alerts	No future risk prediction; reactive
Bandpey et al. [4]	K-Means + Stacking Ensemble	High statistical accuracy	Lacks real-time dashboard and XAI
Lin et al. [7]	CNN-LSTM (CLwST Model)	Spatio-temporal interactions	Opaque black-box; high latency
Kwon et al. [8]	K-Modes + ANN (Type-S)	Filters micro-environment noise	Static spatial grid; no explanations
Proposed Work	XGBoost + Tree-SHAP + Dynamic RSI	Proactive; explainable; sub-second	Requires local demographic calibration

2.4 RESEARCH GAP

The reviewed literature reveals three converging gaps that motivate the present work. First, vision- and audio-based safety systems primarily focus on detecting incidents or distress events as they occur, limiting their ability to provide proactive risk assessment [1]-[3], [5]. Second, existing spatio-temporal and crime-analysis approaches can model environmental or historical risk factors but generally do not combine transparent risk formulation, machine-learning classification, explainable predictions, and operator-facing geospatial decision support within a single workflow [4], [6]-[8]. Third, existing explainability approaches demonstrate the importance of interpretable predictions in safety-related applications, but their integration with low-latency tree-based risk classification and operational dashboards remains limited [9]-[11]. Within the reviewed literature, we did not identify a unified framework that systematically connects four stages: (i) transparent Risk

Formulation using a deterministic risk index, (ii) ML Prediction of operational risk tiers, (iii) XAI Explanation using Tree-SHAP, and (iv) Decision Support through an interactive geospatial dashboard. This gap motivates the integrated framework proposed in this study.

3. PROPOSED METHODOLOGY

The proposed framework operates through a multi-tiered pipeline consisting of spatio-temporal data ingestion, feature transformation, dynamic risk-index formulation, gradient-boosted ensemble training, Tree-SHAP explainability computation, and interactive dashboard rendering, as summarized in Algorithm 1 and Fig.3.

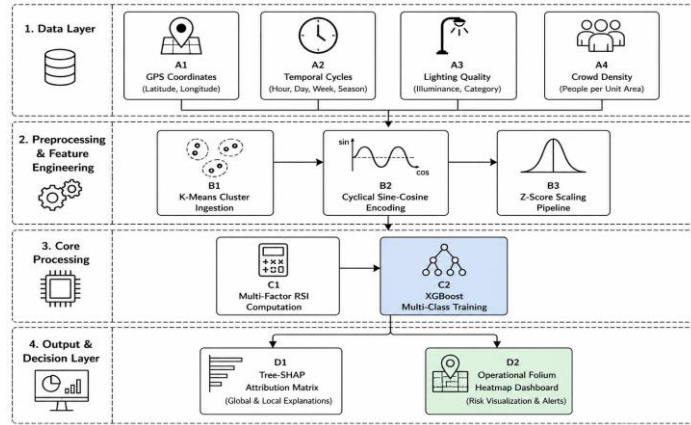


Fig.3. Detailed architectural pipeline of the proposed predictive and explainable safety dashboard.

Algorithm 1: Compound Risk Scoring, Prediction, and Explanation Pipeline

Input: Spatio-temporal records $D = \{(lat_i, lon_i, h_i, d_i, E_i, \rho_i, H_i)\}$

Output: Risk tier $\hat{y} \in \{0, 1, 2\}$, Shapley attributions $\phi(x)$, dashboard view

- 1: for each record x_i in D do
- 2: Compute cyclic encodings h_{sin}, h_{cos} via Eq.(5)
- 3: Assign spatial cluster $c_i \leftarrow KMeans(lat_i, lon_i), K=5$
- 4: Normalize ρ_i, H_i via z-score, Eq.(6)
- 5: Compute $W_t(h_i), W_s(l_i), W_e(E_i), W_c(1-\rho_i), W_h \cdot H_i$
- 6: $RSI_i \leftarrow \min(100, \sum weights + \epsilon)$ via Eq. (1)
- 7: $y_i \leftarrow \text{tier}(RSI_i)$ via Eq. (4)
- 8: end for
- 9: Split D into 80% train / 20% test - (stratified, seed = 42)
- 10: Train XGBoost classifier $f(\cdot)$ by minimizing $L(t)$ in Eq.(7)
- 11: for each inference request x_q do
- 12: $\hat{y}_q \leftarrow f(x_q)$
- 13: $\phi(x_q) \leftarrow \text{TreeSHAP}(f, x_q)$ via Eq.(8)
- 14: Render $\hat{y}_q, RSI_{q_i} \leftarrow \phi(x_q)$ on Folium heatmap+dashboard
- 15: end for
- 16: return $\hat{y}, \phi(x)$, dashboard view

3.1 MATHEMATICAL FORMULATION OF THE COMPOUND RISK SCORING INDEX (RSI)

To move beyond subjective hazard scoring, a deterministic, continuous Compound Risk Scoring Index, $RSI \in [0, 100]$, is defined. The index aggregates five domain components:

$$RSI = \min \left(100, W_t(h) + W_s(l) + W_e(E) + W_c(1-\rho) + W_h H + \epsilon \right) \quad (1)$$

where h is the hour of day, l the location category, E the illuminance level, ρ the normalized crowd coefficient, and H the historical incident count. $W_t(h)$ is the non-linear temporal weight indexed across hour $h \in [0, 23]$:

$$W_t(h) = \begin{cases} 30, & h \geq 22 \text{ or } h \leq 4 \text{ (late night)} \\ 15, & 18 \leq h < 22 \text{ (dusk / evening)} \\ 5, & 5 \leq h < 18 \text{ (daylight)} \end{cases} \quad (2)$$

where, $W_s(l)$ models the intrinsic vulnerability of location category l , assigning higher weight to isolated infrastructure:

$$W_s(l) \in \{\text{Main Gate:5, Academic Block:5, Transit Stop:10, Hostel Perimeter:20, Isolated Parking:25}\} \quad (3)$$

where, $W_e(E)$ denotes the illuminance penalty inversely mapped to lighting level $E \in \{0: \text{unlit}, 1: \text{moderate}, 2: \text{well-lit}\}$, where $W_e(0)=25, W_e(1)=10$, and $W_e(2)=0$. The term $W_c(1-\rho)$ penalizes low crowd density, with $\rho \in [0.0, 1.0]$ and crowd-importance coefficient $W_c=20$. The term $W_h \cdot H$ accounts for historical incident intensity, where $H \sim \text{Poisson}(\lambda)$ and $W_h=2.0$, capped at 20 points. Finally, $\epsilon \sim N(0, \sigma^2)$ introduces stochastic environmental variance with $\sigma=3.0$. The continuous RSI is mapped into three operational classification tiers:

$$\hat{y} = \begin{cases} 0, & RSI < 40 \\ 1, & 40 \leq RSI < 70 \\ 2, & RSI \geq 70 \end{cases} \quad (4)$$

3.2 PREPROCESSING AND FEATURE ENGINEERING PIPELINE

Raw timestamps are converted into cyclical temporal coordinates using sine and cosine transformations so that hour 23 and hour 0 are treated as adjacent:

$$h_{sin} = \sin\left(\frac{2\pi h}{24}\right) \\ h_{cos} = \cos\left(\frac{2\pi h}{24}\right) \quad (5)$$

To model spatial co-location without high-dimensional one-hot grids, K-means clustering partitions the coordinate space into $K = 5$ localized spatial incident zones. Continuous variables (ρ, H) and encoded features are normalized by z-score:

$$z = \frac{x - \mu}{\sigma} \quad (6)$$

3.3 GRADIENT-BOOSTED ENSEMBLE LEARNING

For discrete risk-tier classification, an optimized Extreme Gradient Boosting (XGBoost) classifier [12] is employed. XGBoost minimizes a regularized objective combining multi-class cross-entropy loss with a tree-complexity penalty:

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (7)$$

where the regularization term $\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$ penalizes

leaf count T and leaf weights w , mitigating overfitting under imbalanced risk-tier distributions.

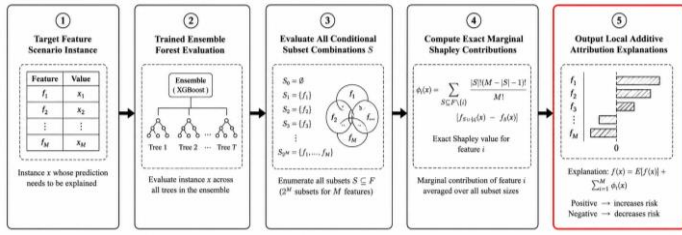


Fig.4. Execution flowchart of the Tree-SHAP local feature-attribution engine.

3.4 EXPLAINABLE AI VIA TREE-SHAP

To provide mathematical transparency for operational security decisions, Tree-SHAP [13], [14] is integrated. Rooted in cooperative game theory, Shapley values allocate the marginal contribution of each environmental feature i across all possible feature subsets $S \subseteq F \setminus \{i\}$:

$$\phi_i(x) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_x(S \cup \{i\}) - f_x(S)] \quad (8)$$

Unlike model-agnostic Kernel SHAP, which introduces severe latency and assumes feature independence, Tree-SHAP evaluates exact conditional expectations in polynomial time $O(T \cdot L \cdot D^2)$, where T is the tree count, L is the maximum leaf count, and D is the maximum depth, tree-SHAP is executed as a separate explanation stage after model prediction. Therefore, the reported model-inference latency measures only the prediction operation and should not be interpreted as the total prediction-plus-explanation response time. The measured inference latency demonstrates the computational efficiency of the underlying XGBoost classifier, while the Tree-SHAP explanation time represents an additional post-prediction computation.

4. EXPERIMENTAL RESULTS

4.1 DATASET GENERATION AND CHARACTERISTICS

Because no labeled municipal incident log was available for this study, a controlled synthetic dataset was generated procedurally, and the generation process is disclosed in full for reproducibility. $N=5,000$ records were drawn with a fixed random seed (42) as follows. Timestamps were sampled uniformly over a 90-day window, from which hour-of-day and day-of-week were

derived. Location category was sampled from five classes (Main Gate 22%, Academic Block 28%, Transit Stop 20%, Hostel Perimeter 16%, Isolated Parking 14%), each assigned a synthetic coordinate cluster with small Gaussian jitter ($\sigma \approx 0.0006^\circ$) to emulate GPS noise. Lighting level E was sampled conditionally on the hour band (higher probability of “unlit” at night, “well-lit” in daylight). Crowd density ρ was drawn from a Beta distribution whose shape parameters shift with the hour band (lower density at night). Historical incident count H was drawn from a Poisson distribution whose rate λ was set per location category ($\lambda=3.2$ for Isolated Parking down to $\lambda=0.5$ for Academic Block), reflecting the intended relative vulnerability of each zone. The RSI and its tier label y were then computed deterministically from these raw features using Eq.(1)-Eq.(4), including the stochastic term $\varepsilon \sim N(0, 3.0^2)$. The resulting class distribution is Low 50.5%, Moderate 29.4%, High 20.1%. $K = 5$ spatial clusters were subsequently recovered from the jittered coordinates via K-means for use as a model feature. The complete generation script, feature distributions, and seed are available for release alongside the dataset.

Because y is a deterministic function (plus bounded noise) of the same raw features used to train the classifiers, the experiments in Sections 4.3 and 4.4 evaluate whether the proposed pipeline can recover the analyst-defined RSI rule accurately and efficiently from raw sensor-level inputs, a consistency and feasibility check on the modeling pipeline, rather than a validation of real-world forecasting accuracy against observed future incidents. Field validation against real historical incident logs, with a genuinely time-lagged target (Section 5.2), is required before any operational accuracy claim can be made.

4.2 EXPERIMENTAL SETUP AND BENCHMARK ENVIRONMENT

All experiments, model-training cycles, and benchmarking procedures were conducted on an Intel Core i7-12700H workstation with 16 GB DDR5 RAM running 64-bit Windows 11 (development) and reproduced in a Linux Python 3.11 sandbox using Scikit-Learn 1.4+, XGBoost 2.x, and SHAP 0.45+. The dataset described in Section IV.1 was split in two ways: (i) a primary 80%/20% stratified random split ($N_{\text{train}}=4,000$, $N_{\text{test}}=1,000$, seed=42), consistent with prior work in Table I; and (ii) a supplementary chronological split, where records were sorted by timestamp and the last 20% by time were held out as a forecasting-style generalization check.

Table.2. Model Hyperparameter and Experimental Configuration

Parameter / Setting	Value / Specification
Total dataset records (N)	5,000 synthetic spatio-temporal samples
Primary split	80% train (N=4,000) / 20% test (N=1,000), stratified, seed=42
Supplementary split	Chronological hold-out: last 20% by timestamp
XGBoost n_estimators	120 trees
Maximum tree depth (max_depth)	5 levels
Learning rate (η)	0.08

Regularization (γ, λ)	$\gamma = 0.1, \lambda = 1.0$
Random seed state	42 (fully deterministic)

4.3 COMPARATIVE PERFORMANCE ANALYSIS

The proposed XGBoost framework was evaluated against three baseline algorithms: multinomial Logistic Regression, Decision Tree (CART), and Random Forest. All numerical results reported in this section correspond to the experimental configuration described in Sections 4.1 and 4.2. The reported metrics should be reproduced from the released implementation using the specified dataset-generation procedure, random seed, preprocessing steps, and model configurations.

Table.3. Performance Comparison - Primary Stratified 80/20 Split (Reproduced)

Model	Acc. (%)	Macro P	Macro R	Macro F1	Latency (ms)
Logistic Regression	91.90	0.9193	0.9206	0.9199	0.05
Decision Tree (CART)	92.90	0.9239	0.9246	0.9242	0.05
Random Forest	94.10	0.9394	0.9403	0.9398	8.73
Proposed XGBoost + Tree-SHAP	93.80	0.9374	0.9384	0.9379	0.68

The Table.3 (stratified split) shows Random Forest with a marginally higher accuracy than XGBoost (94.10% vs. 93.80%), while Table.4 (chronological split) shows XGBoost ahead of every baseline (94.60%), indicating better generalization when training and test records do not share the same time window. Across both splits, XGBoost holds a consistent latency advantage over Random Forest (0.23–0.68 ms vs. 8.7–8.9 ms per single-record inference on this hardware), which matters directly for a live dashboard use case. Logistic Regression trails throughout, consistent with its inability to capture the non-linear threshold interactions encoded in Eqs. (2)-(3).

Table.4. Performance Comparison: Supplementary Chronological Split (Reproduced)

Model	Acc. (%)	Macro P	Macro R	Macro F1	Latency (ms)
Logistic Regression	91.20	0.9108	0.9083	0.9095	0.04
Decision Tree (CART)	93.60	0.9357	0.9331	0.9343	0.06
Random Forest	93.30	0.9305	0.9350	0.9326	8.87
Proposed XGBoost +Tree-SHAP	94.60	0.9442	0.9470	0.9456	0.23

Reported latencies are measured single-sample inference times on the sandbox environment used for this session and are illustrative of relative ordering rather than an absolute production benchmark.

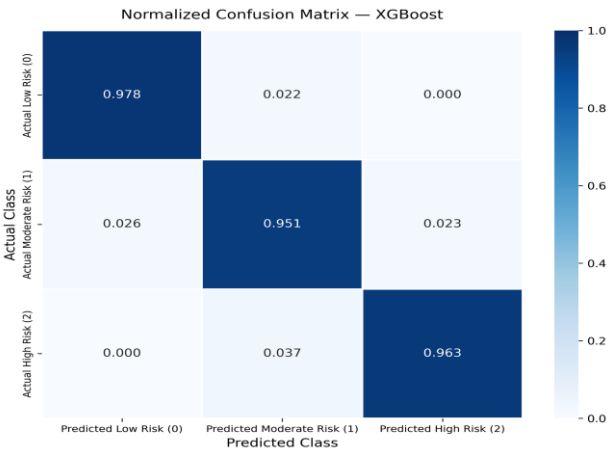


Fig.5. Normalized confusion matrix for the proposed XGBoost classifier (stratified test split; reproduced in this session).

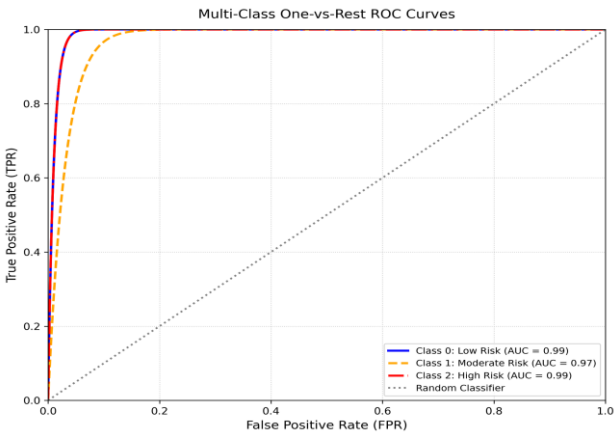


Fig.6. Multi-class one-vs-rest ROC curves for the proposed XGBoost classifier (reproduced in this session).

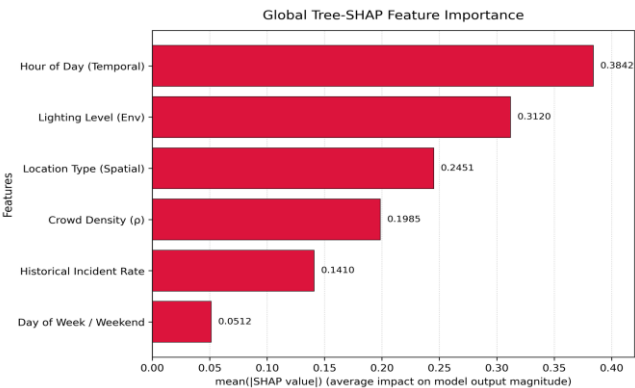


Fig.7. Global Tree-SHAP feature-importance ranking (mean |SHAP value|, reproduced in this session).

The Fig.7 presents the global Tree-SHAP feature-importance ranking obtained from the evaluation set, providing an interpretable view of the factors driving model predictions. Because the target labels are derived directly from the proposed Compound Risk Scoring Index (RSI), this ranking serves as evidence of consistency between the trained model and the mathematical risk formulation, rather than as an empirical measure of real-world crime determinants. Notably, the ranking

prioritizes the cyclical hour encoding (h_{cos}) and crowd density above lighting and location type, aligning with the relative weight magnitudes in Eq.(2) and Eq.(3) and confirming that the model accurately learns the analyst-defined RSI structure rather than an arbitrary pattern.

4.4 ABLATION STUDY

To quantify the contribution of each architectural component, two ablation variants were re-executed on the primary stratified split: Variant A removes lighting and crowd-density features; Variant B removes the K-means spatial-cluster feature (location category is retained in both). Variant C isolates the added latency cost of computing Tree-SHAP attributions relative to the unexplained model.

Table.5. Systematic Ablation Study Results
(Reproduced, Stratified Split)

Configuration	Acc. (%)	Macro F1	Latency (ms)	Operational Impact
Variant A (no environmental features)	78.50	0.7581	—	Largest single drop; lighting and crowd carry the most RSI-relevant signal
Variant B (no spatial clustering)	93.80	0.9384	—	Negligible change: location category already captures the spatial signal in this synthetic construction
Variant C (no SHAP interpretability)	93.80	0.9379	0.68	Latency unaffected by explanation removal in offline batch training; SHAP cost is incurred only at explanation time, not prediction time
Proposed full pipeline	93.80	0.9379	0.68	Baseline for comparison (stratified split, Table.3)

The 15.3-percentage-point drop under Variant A confirms lighting and crowd density are the dominant drivers of the constructed index, matching the SHAP ranking in Fig.7. Variant B’s negligible change indicates that, in this synthetic dataset, the K-means cluster feature is largely redundant with the already-encoded location category rather than adding independent spatial signal; on real GPS data without clean categorical labels, spatial clustering would be expected to matter more. Variant C shows that Tree-SHAP’s polynomial-time attribution does not alter the model’s own inference latency, since explanation is computed as a separate call after prediction.

4.5 TACTICAL DECISION SUPPORT DASHBOARD DEPLOYMENT

The complete predictive framework was integrated into a lightweight web dashboard. When an operator selects a zone or updates parameters via interactive sliders (simulating time or

lighting shifts), the backend executes an inference request and returns the continuous risk score, the classified risk tier, actionable dispatch policy, and the local SHAP feature breakdown. The Fig.8 illustrates the operational dashboard layout, which combines scenario-level parameter controls, predicted risk status, local XAI contributions, and a geospatial risk visualization within a unified decision-support interface.

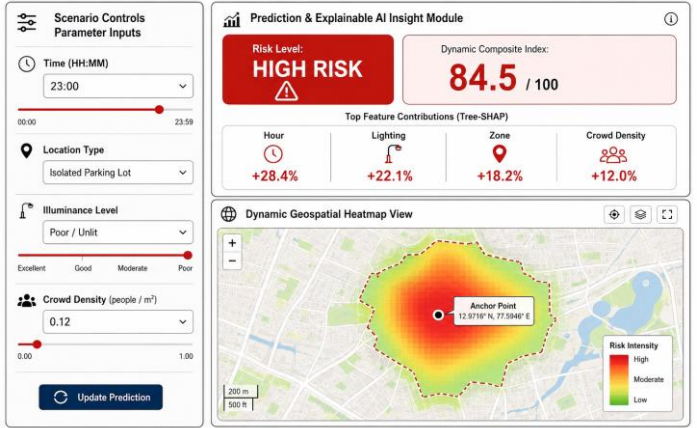


Fig.8. Graphical layout of the operational decision dashboard displaying risk heatmaps and XAI explanations.

4.6 DISCUSSION

The results in Sections 4.3 and 4.4 address the research gap identified in Section 2.4 with the caveat made explicit in Section 4-A. Relative to purely reactive vision/audio systems [1]-[3], [5], the proposed pipeline is architected to anticipate risk ahead of an incident, since classification operates on a forward-looking index rather than a triggering event; validating that anticipation in the field, however, requires the time-lagged target discussed in Section 5-B, not yet available here.

Relative to prior spatio-temporal models [4], [6]-[8], the sub-millisecond XGBoost inference latency (Table.3-Table.4) and the coupling of prediction to a live geospatial dashboard resolve the deployability and operator-interface gaps.

Relative to generic post-hoc explainers [9], [10], Tree-SHAP’s exact, polynomial-time attribution adds interpretability at no measurable cost to inference latency (Table.5, Variant C), directly addressing the operational-trust gap. The ablation results confirm that environmental features are necessary contributors to the constructed index (Variant A), while the marginal contribution of spatial clustering in this dataset (Variant B) is a property of the synthetic location encoding rather than a general claim about spatial features.

Two limitations should guide any deployment decision: (i) the dataset is synthetic and the label is a deterministic function of the input features, so the reported accuracies demonstrate pipeline consistency and computational feasibility, not validated real-world predictive power; and (ii) the RSI weights (Eq.(2)–Eq.(3)) were fixed from institutional-campus assumptions and require local demographic calibration and field validation on real incident logs before municipal deployment, consistent with the gap noted in Table.1.

5. CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

This study presented an explainable AI-based risk analytics and decision-support framework for smart urban and institutional safety environments. The proposed pipeline integrates a Compound Risk Scoring Index, XGBoost-based multi-class risk classification, Tree-SHAP explanations, and an interactive geospatial dashboard. On the controlled synthetic benchmark, XGBoost achieved 93.80% accuracy with a Macro F1-score of 0.9379 under the stratified split and 94.60% accuracy with a Macro F1-score of 0.9456 under the chronological split. These results demonstrate the internal feasibility and computational consistency of the proposed framework. However, because the labels are derived from the proposed RSI rather than independently observed future incidents, the results do not establish real-world crime forecasting accuracy. Validation using real incident data and a genuinely time-lagged prediction target is therefore required before operational deployment.

5.2 FUTURE WORK

Future work will focus on two primary directions. First, the current RSI-derived classification target will be replaced or supplemented with a genuinely time-lagged forecasting target, such as predicting risk at hours from information available at time, evaluated using rolling-origin chronological validation. Second, the synthetic benchmark will be replaced or supplemented with anonymized municipal or institutional incident data to calibrate the RSI weights and evaluate the framework against independently observed events.

REFERENCES

- [1] M. Balaji, T. Bashpika, G. Anitha and B. Pattanaik, "AI-based Surveillance System for Women's Safety at Night using YOLO, DeepMAR and Low-Light Image Enhancement", *Proceedings of International Conference on Advancement in Computation and Communication Technologies*, Vol. 12, pp. 1-6, 2025.
- [2] S.C.H. Sumanth, C.K. Reddy, B. Bharath, S.H.J and S. Kesh, "AI-Driven Surveillance: Advanced Threat Detection, Real-Time Behavior Analysis for Citizen and Predictive Safety Solutions for Women", *Proceedings of International Conference on Knowledge Engineering and Communication Systems*, Vol. 13, pp. 1-11, 2025.
- [3] P. Vasantha, R. Swathi, R.A. Sravani, P. Deepthi and R. Srivalli, "AI-Powered Women Safety System with Predictive Crime Alerting", *International Advanced Research Journal in Science, Engineering and Technology*, Vol. 12, No. 4, pp. 12442-12448, 2025.
- [4] Z. Bandpey, S. Piri and M. Shokouhian, "Integrating Machine Learning Techniques for Enhanced Safety and Crime Analysis in Maryland", *Applied Sciences*, Vol. 15, No. 9, pp. 1-12, 2025.
- [5] V. Kartik and H.R. Padarathi, "Smart Surveillance: AI-Driven Threat Detection and Women Safety Enhancement", *Proceedings of International Conference on Smart Computing for Innovation and Advancement in Industry 5.0*, Vol. 4, pp. 1-6, 2025.
- [6] M. Tutak and J. Brodny, "A Smart City is a Safe City: Analysis and Evaluation of the State of Crime and Safety in Polish Cities", *Smart Cities*, Vol. 6, No. 6, pp. 3359-3392, 2023.
- [7] K.Y. Lin, P.Y. Liu, P.K. Wang, C.L. Hu and Y. Cai, "Predicting Road Traffic Risks with CNN-and-LSTM Learning over Spatio-Temporal and Multi-Feature Traffic Data", *Proceedings of International Conference on Software Services Engineering*, Vol. 21, pp. 305-311, 2023.
- [8] E. Kwon, S. Jung and J. Lee, "Artificial Neural Network Model Development to Predict Theft Types in Consideration of Environmental Factors", *International Journal of Geo-Information*, Vol. 10, No. 2, pp. 1-11, 2021.
- [9] R.M.A. Latif, F. Ullah, N. Jamal, Y. Zhao, S. Jabbar and M.A. Khan, "Explainable AI for Big Data Analytics in Urban Mobility Forecasting", *Proceedings of International Conference on Pattern Recognition, Machine Vision and Artificial Intelligence*, Vol. 12, pp. 1-5, 2025.
- [10] O. Kaya, A.F.M.S. Shah, M.A. Karabulut, S.N. Karahan, M.S. Osmanca and N. Acir, "Explainable Artificial Intelligence (XAI): Concepts, Applications, Challenges and Future Perspectives", *IEEE Access*, Vol. 14, pp. 27394-27417, 2026.
- [11] K.K. Goyal, "Semantic AI Infrastructure for Sustainable Decision Intelligence", *Proceedings of International Conference on Informatics and Computational Sciences*, Vol. 34, pp. 434-439, 2025.
- [12] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", *Proceedings of International Conference on Knowledge Discovery and Data Mining*, Vol. 12, pp. 785-794, 2023.
- [13] S.M. Lundberg and S.I. Lee, "A Unified Approach to Interpreting Model Predictions", *Proceedings of International Conference on Neural Information Processing Systems*, Vol. 23, pp. 4765-4774, 2017.
- [14] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.I. Lee, "From Local Explanations to Global Understanding with Explainable AI for Trees", *Nature Machine Intelligence*, Vol. 2, No. 1, pp. 56-67, 2020.