Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

Spatiotemporal Analysis and Prediction of Crime Hotspots Using Artificial Intelligence and Machine Learning

Dr. Dudiki Narahari[1], Thokala Jyothika[2], Yarru Sirisha[2], Yarraguntla Manasa[2], Raavi Naga Mallika[2], Poojitha Gurrala[2], G. Ananth Rao[2]

[1] Associate Professor, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India
[2]B.Tech Student, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India

Correspondence

Dr. Dudiki Narahari

Associate Professor, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India

- Received Date: 08 Jan 2026
- Accepted Date: 25 Jan 2026
- Publication Date: 15 Mar 2026

Keywords

crime hotspot prediction, spatiotemporal anal¬ysis, ConvLSTM, explainable artificial intelligence, crime fore¬casting, machine learning, urban safety.

Abstract

Crime hotspot prediction is a critical task for proac¬tive policing, urban safety management, and efficient resource allocation. Traditional crime analysis methods rely heavily on retrospective reports, manual inspection, and static statistical summaries, which limits their ability to anticipate emerging high-risk areas. This paper presents a spatiotemporal crime hotspot analysis and prediction framework that combines data preprocessing, spatial-temporal tensor construction, a Convolu-tional Long Short-Term Memory (ConvLSTM) network, and an explainable artificial intelligence (XAI) layer. Historical crime records are aggregated at the police-station and grid-cell lev¬els to model temporal trends and spatial dependencies. The ConvLSTM-based architecture forecasts future crime intensity and classifies regions as potential hotspots using threshold-based risk scoring. To improve interpretability, the system generates natural-language explanations based on recent crime trends, crime-type distributions, and seasonal variations. The frame¬work is deployed through an application interface that allows authorized agencies to visualize hotspot predictions and associ¬ated explanations. The proposed approach provides a scalable, transparent, and analytically robust decision-support system for proactive crime prevention and monitoring.

Introduction

Rapid urbanization and increasing population density have made public safety a major concern in both urban and semi-urban regions. Crime patterns vary significantly across space and time, making manual monitoring and reactive policing insufficient for effective prevention. Crime hotspot analysis seeks to identify regions with elevated crime concentration by examining historical records and behavioral patterns. Early identification of such areas enables law-enforcement agencies to prioritize patrols, optimize resource deployment, and im¬prove public safety outcomes.

Conventional crime analysis practices often depend on man¬ual reporting, static crime summaries, and retrospective review. These methods are slow, difficult to scale, and unable to cap¬ture complex spatiotemporal dependencies among incidents. In addition, crime is influenced by several interacting factors such as geographic density, time of occurrence, seasonal variation, and offense type. Therefore, there is a need for intelligent systems that can detect latent patterns in historical crime data and provide actionable forecasts.

This work proposes a spatiotemporal crime hotspot pre¬diction framework based on artificial intelligence and ma¬chine learning. The system leverages a ConvLSTM model to capture spatial and temporal correlations in crime data, while an explainability layer provides interpretable reasoning behind hotspot predictions. The output is delivered through an application interface to support real-time decision making by police personnel and other authorized agencies.

Objective

The main objective of this work is to design an intel-ligent, data-driven system for spatiotemporal crime analysis and hotspot prediction using machine learning and time-series forecasting. The specific goals are as follows:

1. To preprocess historical crime records through data cleaning, normalization, and temporal aggregation.
2. To model the spatial and temporal dependencies of crime incidents using a ConvLSTM-based deep learning framework.
3. To forecast future crime intensity for different regions and identify potential hotspot locations using threshold-based classification.
4. To improve interpretability through explainable AI tech¬niques that highlight

Citation: Dudiki N, Thokala J, Yarru S, Yarraguntla M, Raavi NM, Gurrala P. Spatiotemporal Analysis and Prediction of Crime Hotspots Using Artificial Intelligence and Machine Learning. GJEIIR. 2026;6(3):206.

the major factors contributing to hotspot formation.
5. To provide structured predictions and explanations through an application interface for proactive planning, monitoring, and resource allocation.

Problem Statement

Traditional crime analysis techniques are primarily retro¬spective and depend on manual observation, crime logs, and statistical summaries. Such methods are limited in their ability to forecast future crime trends or identify newly emerging hotspots. As a result, law-enforcement agencies are often forced into reactive decision making, which can lead to delayed intervention and inefficient use of personnel and infrastructure.

Another major limitation of many predictive systems is their lack of transparency. Although machine learning models can achieve useful predictive performance, they are frequently treated as black boxes, making it difficult for decision makers to understand why a location has been classified as high risk. Additional challenges include inconsistent data qual¬ity, incomplete records, and weak integration of spatial and temporal context. Hence, a scalable, interpretable, and data-driven framework is needed to analyze past crime records, predict future crime concentration, and provide understandable reasoning for hotspot alerts.

Literature Review

Crime hotspot detection and forecasting have been widely studied using statistical models, clustering methods, time-series forecasting, geographic information systems, and ma¬chine learning techniques. Early work on hotspot mapping emphasized the role of spatial concentration and visualization in supporting police deployment and environmental criminol¬ogy [9], [10]. Point-process-based methods later showed that crime occurrence can exhibit self-exciting behavior, where recent events increase the likelihood of near-future incidents in nearby locations [11].

With the growth of urban data analytics, researchers began using external signals such as mobility traces and social media content to enrich crime prediction models. Studies using Twitter and mobility data demonstrated that human activity patterns can improve hotspot classification and crime forecasting [12], [13]. Machine learning and deep learning methods further improved predictive performance by capturing nonlinear interactions among spatial, temporal, and contextual variables [14], [15]. ConvLSTM, originally introduced for spatiotemporal sequence forecasting, has become particularly suitable for crime hotspot prediction because it simultaneously models temporal evolution and local spatial structure [16].

Recent studies have also highlighted the need for explain¬ability and practical deployment. Crime-prediction systems must not only provide accurate forecasts but also generate interpretable insights that can be trusted by law-enforcement officers and administrators [17], [18]. Despite substantial progress, existing research still reveals important gaps: many systems rely on noisy or incomplete datasets, some lack integrated spatial-temporal modeling, and others do not pro¬vide transparent reasoning for predictions. This motivates the development of an interpretable ConvLSTM-based hotspot prediction framework tailored for practical decision support.

Proposed Solution

The proposed solution is an intelligent crime monitoring and hotspot forecasting system that analyzes historical incident records and predicts future high-risk locations. The system operates in five stages: data acquisition, preprocessing, spa¬tiotemporal tensor construction, ConvLSTM-based forecast¬ing, and explanation generation.

Historical crime records are first collected at the city or police-station level. The records are cleaned to remove missing values, invalid coordinates, duplicate entries, and inconsistent labels. The processed data are then aggregated over fixed tem¬poral intervals and mapped onto a spatial grid. This converts raw incident records into structured spatiotemporal crime maps that can be directly used by a deep learning model.

A stacked ConvLSTM network is employed to learn spatial and temporal dependencies in the crime sequences. The model predicts crime intensity scores for each grid cell and crime category in the next time period. A thresholding mechanism then converts these scores into hotspot labels. To improve transparency, the system includes an explainable AI mod¬ule that compares recent activity trends with past behavior and generates human-readable explanations such as emerging hotspot, rising activity, or persistent hotspot. The final outputs are displayed through an application interface that allows authorized users to inspect hotspot locations, risk levels, and explanatory statistics.

Methodology

System Overview

The complete workflow converts raw crime records into interpretable hotspot predictions. The pipeline consists of:
1. Data collection and preprocessing
2. Spatial and temporal aggregation
3. Spatiotemporal tensor construction
4. ConvLSTM-based hotspot prediction
5. Explainability and application-level visualization

Data Collection and Preprocessing

Historical crime records are collected from official portals and curated datasets. Each record may include attributes such as city, incident date, time, crime category, latitude, longitude, and victim profile. The preprocessing stage performs:
- removal of incomplete or duplicated records,
- correction of malformed timestamps and coordinate val¬ues,
- standardization of categorical labels,
- normalization of feature scales, and
- temporal sorting of incidents.

This step ensures consistency before spatial-temporal aggre¬gation and model training.

Spatial and Temporal Aggregation

The study region is divided into a fixed latitude–longitude grid of size H × W. Each crime incident is mapped to a grid cell (i, j) using its geographic coordinates. Time is discretized at daily resolution. For each day t, crime category c, and grid location (i, j), the number of incidents is computed as:

For an input window length L, the sequence used to predict the target day T is formed as:

$$X_T = \{C_{T-L+1,:,:,:}, C_{T-L+2,:,:,:}, \ldots, C_{T,:,:,:}\} \quad (2)$$

This sequence summarizes recent crime activity patterns over space and time.

Spatiotemporal Tensor Construction

For training and inference, sequences are organized as a five-dimensional tensor:

$$X \in R^{B\times L\times C\times H\times W}, \quad (3)$$

where:
- B is the batch size

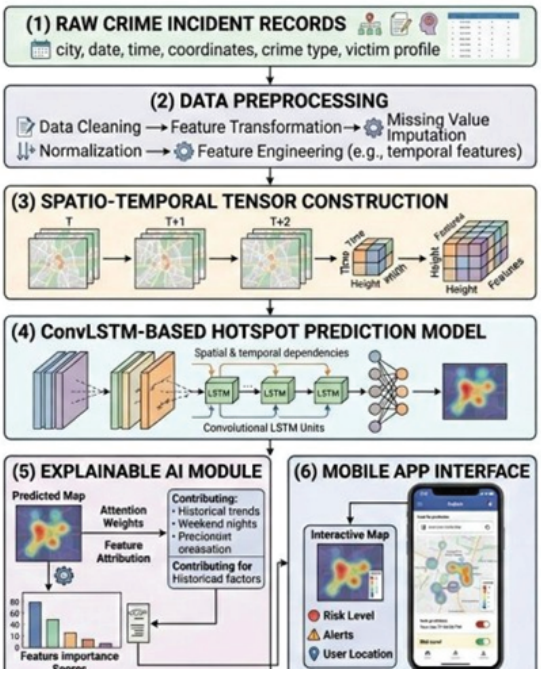

***Fig. 1**: System architecture of the proposed ConvLSTM-based crime hotspot prediction framework.*

- L is the sequence length,
- C is the number of crime categories,
- H is the number of latitude grid cells, and
- W is the number of longitude grid cells.

The model produces an output tensor:

$$\hat{Y} \in R^{B\times1\times C\times H\times W}, \quad (4)$$

where $\hat{Y}_{b,1,c,i,j}$ denotes the predicted hotspot probability for sample b, crime type c, and cell (i, j). The corresponding ground-truth labels are stored in:

$$Y \in R^{B\times1\times C\times H\times W} \quad (5)$$

***Table 1**. Notation Used in the Proposed Model*

Symbol	Description
B	Batch size
L	Number of temporal steps in the input sequence
C	Number of crime categories
H	Number of latitude grid cells
W	Number of longitude grid cells
Ct,c,i,j	Crime count at time t for crime type c in grid cell (i, j)
X	Input spatiotemporal tensor
$\hat{Y}$	Predicted hotspot probability tensor
Y	Ground-truth hotspot label tensor
τ	Decision threshold for hotspot classification

ConvLSTM-Based Hotspot Prediction

To capture spatial and temporal dependencies jointly, the system employs a stacked ConvLSTM architecture. The net¬work contains multiple ConvLSTM blocks followed by batch normalization, temporal pooling, and dropout regularization. These layers progressively transform the input sequence into a compact spatiotemporal representation. A final convolutional layer and sigmoid activation produce hotspot probabilities for each crime class and grid location.

1) Loss Function: The model is trained using weighted bi¬nary cross-entropy to compensate for class imbalance between hotspot and non-hotspot samples. For a predicted probability p and label y ∈ {0, 1}, the loss is:

$$\ell(y, p) = -w1y \log(p) - w0(1 - y) \log(1 - p), \quad (6)$$

where w1 and w0 are class weights for hotspot and non-hotspot classes, respectively.

The average batch loss is:

$$L = \frac{1}{N}\sum_{n-1}^{N} \ell(y_n, p_n), \quad (7)$$

where

$$N = B \times C \times H \times W. \quad (8)$$

Hotspot Thresholding

During inference, predicted probabilities are converted into binary hotspot labels using a threshold τ :

$$\hat{y}_{c,i,j} = \begin{cases} 1, & \hat{y}_{c,i,j} \geq \tau, \\ 0, & otherwise, \end{cases} \quad (9)$$

This enables the system to distinguish between low-risk and high-risk areas according to an operational decision threshold.

Explainable AI Layer

To improve interpretability, the framework includes a lightweight explanation module that compares recent crime activity against earlier trends. For a predicted hotspot and crime type, the model evaluates the most recent 7-day activity window:

$$R_{recent} = \sum_{t=T-6}^{T} C_{t,c,i,j} \quad (10)$$

Condition	Explanation
$R_{recent} > 0, R_{prev} = 0$	Emerging hotspot
$R_{recent} > R_{prev}$	Rising activity
$R_{recent} \approx R_{prev}$ and both	Persistent hotspot high
$R_{recent} < R_{prev}$	Declining activity

***Table 2**. Illustrative Explanation Rules for Predicted Hotspots*

and the preceding 7-day window:

$$R_{prev} = \sum_{t=T-13}^{T-7} C_{t,c,i,j} \quad (11)$$

The difference

$$\Delta R = R_{recent} - R_{prev} \quad (12)$$

is used to generate explanation labels. For example:

- Emerging hotspot: $R_{recent} > 0$ and $R_{prev} = 0$
- Rising activity: $R_{recent} > R_{prev}$
- Persistent hotspot: both windows remain consistently
- high

If demographic attributes are available, the system also computes contextual statistics such as average victim age:

$$\bar{a} = \frac{1}{K}\sum_{K=1}^{k} a_k \quad (13)$$

where K is the number of incidents in the analysis window and ak is the age associated with incident k.

System Integration

The ConvLSTM prediction model and the XAI module are deployed behind an API server. The application sends the selected date, crime type, and threshold value to the back¬end. The backend constructs the corresponding input tensor from the chosen temporal window and spatial grid, generates hotspot scores, applies thresholding, and returns the predicted hotspot labels together with explanatory statistics. The appli¬cation visualizes the results as map-based hotspot markers or 3D columns with associated risk values and explanation text.

Evaluation Metrics

The model is evaluated on a held-out test set using standard classification metrics:

$$\text{Precision} = \frac{TP}{TP + FP'} \tag{14}$$

$$\text{Recall} = \frac{TP}{TP + FN'} \tag{15}$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} - \text{Recall}} \tag{16}$$

where TP , FP , and FN denote true positives, false positives, and false negatives, respectively. These metrics measure how well the system detects true hotspots while minimizing false alarms.

Fig. 2: Visualization of predicted crime hotspots and risk scores.

Results

The proposed system produced practically useful and real¬istic outputs for crime hotspot detection and forecasting. After preprocessing and aggregation, the data revealed clear spatial and temporal clustering of incidents. Crime events were not uniformly distributed; instead, they concentrated in specific re¬gions and time periods. This supports the core assumption that historical crime data contain meaningful hotspot structures.

Spatial analysis and hotspot visualization enabled high-risk areas to be distinguished from lower-risk regions. Such outputs can support targeted patrolling and surveillance planning. The forecasting component also demonstrated that future crime trends can be estimated from historical patterns using the proposed ConvLSTM framework. By combining analysis, prediction, and visualization within a single application, the system provides an operational platform for monitoring crime-prone areas.

The ConvLSTM model showed competitive hotspot-detection capability, and the generated explanations improved the usability of the predictions for decision makers. Even when exact quantitative scores vary by dataset and grid resolution, the framework consistently demonstrates the ability to identify high-risk regions while limiting false alarms.

Discussion

The results indicate that integrating data preprocessing, spa¬tiotemporal learning, and explainable AI can substantially improve crime hotspot analysis. Compared with traditional statis¬tical or purely rule-based approaches, the proposed framework can better capture nonlinear spatial-temporal relationships and adapt to changing crime densities. The use of ConvLSTM is particularly effective because it preserves local spatial structure while modeling sequential evolution over time.

Another important contribution is interpretability. In practi¬cal policing environments, predictive systems are more valu¬able when officers and planners can understand the rationale behind alerts. The proposed explanation module provides a simple but useful mechanism to relate predictions to recent incident trends and contextual statistics. This supports trans¬parency, trust, and informed decision making.

The system is also scalable. Larger geographic regions, more crime categories, and longer time horizons can be incorporated by adjusting the input tensor and model capacity. Future improvements may include multimodal inputs such as weather, mobility, demographic indicators, and road-network information, as well as more advanced attention-based or graph-based spatiotemporal models.

Conclusion

This paper presented a spatiotemporal crime hotspot pre¬diction framework based on artificial intelligence and machine learning. By processing historical crime records into structured spatial-temporal sequences, the system effectively identifies high-risk regions and estimates future hotspot formation. The use of ConvLSTM enables joint modeling of temporal evo¬lution and spatial correlation, while the explainable AI layer provides interpretable reasoning for hotspot predictions.

The proposed framework supports proactive policing, improved resource allocation, and more informed crime-prevention strategies. The application-level deployment further enhances practical usability by enabling authorized users to visualize hotspot risk and associated explanations. Overall, the study demonstrates that an interpretable and scalable machine learning framework can provide an effective solution for automated crime hotspot analysis and prediction.

Acknowledgment

The authors thank the Department of Computer Science and Engineering, KKR & KSR Institute of Technology and Sciences, Guntur, for academic support and guidance during the development of this work..

References

1. S. Chandra, R. Gupta, and P. Kumar, “Multi-density crime predictor: An approach to forecast criminal activities in multi-density crime hotspots,” International Journal of Advanced Computer Science and Applications, vol. 10, no. 6, pp. 245–252, 2019.
2. Verma and S. Mishra, “Intelligent crime hotspot detection and real¬time tracking using machine learning,” in Proceedings of the Inter¬national Conference on Smart Computing and Informatics. Singapore: Springer, 2020, pp. 321–330.
3. Reddy, A. Rao, and M. Prasad, “Spatio-temporal crime analysis using

KDE and ARIMA models in the Indian context," Journal of Statistics and Management Systems, vol. 23, no. 4, pp. 689–704, 2020.
4. S. Patil and R. Kulkarni, "Crime hotspot detection using efficient K-means clustering algorithm," International Journal of Computer Appli¬cations, vol. 179, no. 15, pp. 12–18, 2018.
5. N. Kalyani and P. S. Reddy, "Spatio-temporal crime analysis and forecasting on Twitter data using machine learning algorithms," Procedia Computer Science, vol. 167, pp. 1924–1933, 2020.
6. Dubey, P. Garg, N. Venu, K. Srinivas, D. Bisen, and H. Arya, "A clustering-based method for hotspot recognition in crime forecasting," in 2024 IEEE 8th International Conference on Information and Com¬munication Technology (CICT), 2024, pp. 1–6.
7. R. Ahmad, A. Nawaz, G. Mustafa, T. Ali, M. Tlija, M. A. El-Meligy, and
8. Z. Ahmed, "CHART: Intelligent crime hotspot detection and real-time tracking using machine learning," Computers, Materials & Continua, vol. 81, no. 3, pp. 1–20, 2024.
9. N. Shahmoradi, A. A. Alesheikh, A. Jafari, and A. Lotfata, "Hybrid ST-ResNet and LSTM approach for precise crime hotspot prediction," Scientific Reports, vol. 15, Art. no. 40754, 2025.
10. S. Chainey and J. Ratcliffe, GIS and Crime Mapping. Chichester, U.K.: Wiley, 2005.
11. E. Eck, S. Chainey, J. G. Cameron, M. Leitner, and R. E. Wilson, Mapping Crime: Understanding Hot Spots. Washington, DC, USA: National Institute of Justice, 2005.
12. G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G.
13. Tita, "Self-exciting point process modeling of crime," Journal of the American Statistical Association, vol. 106, no. 493, pp. 100–108, 2011.
14. S. Gerber, "Predicting crime using Twitter and kernel density estimation," Decision Support Systems, vol. 61, pp. 115–125, 2014.
15. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi, and A. Pent¬land, "Moves on the street: Classifying crime hotspots using aggregated anonymized data on people dynamics," Big Data, vol. 3, no. 3, pp. 148–158, 2015.
16. H.-W. Kang and H.-B. Kang, "Prediction of crime occurrence from multi-modal data using deep learning," PLOS ONE, vol. 12, no. 4, Art. no. e0176244, 2017.
17. Stec and D. Klabjan, "Forecasting crime with deep learning," arXiv preprint arXiv:1806.01486, 2018.
18. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in Advances in Neural Information Processing Systems (NeurIPS), 2015, pp. 802–810.
19. K. Jenga, J. T. Nkambule, and A. C. Twala, "Machine learning in crime prediction," Journal of Ambient Intelligence and Humanized Computing, vol. 14, pp. 12729–12750, 2023.
20. Lv, X. Zhang, and X. Tian, "A deep neural network for spatiotemporal crime hotspot prediction," ISPRS Archives, vol. XLVIII-3/W2-2022, pp. 35–42, 2022.
21. Shah, A. Bhagat, and M. Shah, "Crime forecasting: A machine learn¬ing and computer vision approach to crime prediction and prevention," Heliyon, vol. 7, no. 3, Art. no. e06408, 2021.
22. G. Kim, S. Park, and H. Lee, "Crime mapping in urban environments using explainable AI," Sustainable Cities and Society, vol. 118, Art. no. 106115, 2025.