Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

Machine Learning-Based Crime Trend Analysis and Risk Prediction

K Manikanta, Suddapalle Saddam Hussain, Challa Preethi, Vaduguru Mangala Neeraja, Mandli Praveen Kumar, Dudekula Vamsi

Department of Computer Science and Engineering, Gates Institute of Technology, Andhra Pradesh, India.

Correspondence

K.Manikanta

Department of Computer Science and Engineering, Gates Institute of Technology, Andhra Pradesh, India.

- Received Date: 11 Feb 2026
- Accepted Date: 02 Apr 2026
- Publication Date: 25 Apr 2026

Keywords

Crime Data Analysis, Machine Learning, Predictive Analytics, Exploratory Data Analysis, Crime Prediction Systems

Abstract

To effectively support law enforcement agencies and improve public safety, it is essential to understand patterns and trends in criminal activity. In this project, crime data collected from publicly available sources and structured datasets is analyzed to identify patterns related to location, time, and environmental factors. The primary objective of the system is to predict the category of crime that is most likely to occur based on input features such as district, area type, weather conditions, crowd level, lighting condition, police presence, and time-related attributes. Machine learning techniques are used to build predictive models capable of analyzing historical crime data and generating accurate crime predictions. Multiple classification algorithms, including Random Forest, Gradient Boosting, XGBoost, and Support Vector Machine, are evaluated to determine the most effective model for prediction. The system is implemented as a web-based application using the Flask framework, allowing users to input relevant parameters and obtain predicted crime categories along with confidence scores and associated risk levels. In addition, the system applies clustering techniques such as K-Means to identify crime hotspot zones based on geographical coordinates. By analyzing spatial and temporal crime patterns, the system provides meaningful insights that can assist authorities in planning preventive measures and improving crime monitoring strategies. Finally, the results obtained from different machine learning models are compared and analyzed, and possible directions for further improvement of the system are discussed.

Introduction

In an era characterized by the rapid growth of urbanization and digital data availability, crime has become a significant concern affecting societies worldwide. The increasing complexity and frequency of criminal activities demand advanced and intelligent systems for effective prevention and control. Crime prediction plays a crucial role in assisting law enforcement agencies by providing insights into potential crime hotspots and enabling proactive decision-making. It helps in improving public safety, optimizing resource allocation, and enhancing surveillance strategies across various regions. Crime incidents can vary widely, including theft, assault, robbery, and other unlawful activities, often influenced by factors such as location, time, and socio-economic conditions. Predicting such patterns is a challenging task due to the dynamic and uncertain nature of crime data. Traditional statistical methods are often insufficient for handling large-scale and complex datasets, especially when temporal and spatial dependencies are involved. In recent years, there has been a growing interest in applying advanced computational techniques for crime prediction, leading to the development of more accurate and efficient models. One of the most promising approaches is the use of machine learning and deep learning techniques. These methods have shown significant success in analyzing complex patterns, enabling automatic feature extraction, and generating reliable predictive models. In this paper, we present a novel approach to crime prediction using advanced machine learning techniques, with a primary focus on improving prediction accuracy and efficiency. Our approach builds upon existing models but introduces enhanced data preprocessing methods, optimized model architecture, and improved training strategies. The objective is to accurately predict crime occurrences based on historical data and identify high-risk areas effectively. We conduct extensive experiments on crime datasets containing spatial and temporal information to evaluate the performance of our proposed model. The results demonstrate that our approach outperforms traditional methods in terms of accuracy, robustness, and scalability. The findings indicate that the proposed system can effectively assist in predicting crime patterns and support law enforcement agencies in making informed decisions. This paper contributes to the field of crime analytics by highlighting the potential of machine learning and deep learning techniques in addressing real-world challenges in crime prediction. It

Citation: Manikanta K, Suddapalle SH, Challa P, Vaduguru MN, Mandli PK, Dudekula V. Machine Learning-Based Crime Trend Analysis and Risk Prediction. GJEIIR. 2026;6(4):233.

emphasizes the importance of intelligent systems in enhancing public safety and improving crime prevention strategies. The study demonstrates how advanced models can uncover hidden patterns in crime data and provide meaningful insights for effective decision-making. The subsequent sections of this paper elaborate on the proposed crime prediction model, the dataset and preprocessing techniques, experimental setup, and performance evaluation. Finally, we present the conclusions and discuss future directions for improving crime prediction systems and expanding their applications.

Related Work

Wang and Zhang [1] developed a deep learning-based crime prediction model using Convolutional Neural Networks (CNNs) to extract spatial features from crime data. Their model improves prediction accuracy in urban regions. However, it requires large-scale labeled datasets for effective training.Liu et al. [2] introduced a Recurrent Neural Network (RNN) model to capture temporal patterns in crime occurrences. This allows better prediction over time. However, the model suffers from vanishing gradient issues in long sequences.Chen and Li [3] provided a comprehensive survey on crime prediction using machine learning techniques. Their work highlights various approaches and datasets. The limitation is that it focuses on analysis rather than proposing a new method.

Kim et al. [4] proposed a hybrid CNN-LSTM model that combines spatial and temporal features for crime forecasting. This improves prediction performance. However, the model is computationally expensive and complex to implement.Cheng et al. [5] developed a machine learning framework using Support Vector Machines (SVM) for crime classification. Their approach shows good accuracy in structured datasets. The issue is that it performs poorly with large and unstructured data.The study in [6] discusses the use of big data analytics in crime prediction. It highlights the importance of data integration from multiple sources. However, it lacks detailed implementation strategies.

Gandhi and Kumar [7] explored explainable AI techniques in crime prediction systems. Their work improves transparency and trust in predictions. The limitation is that it does not significantly improve prediction accuracy.Brown et al. [8] emphasized the importance of data transparency and ethical AI in law enforcement applications. This is useful in sensitive domains. However, the study is more theoretical and lacks empirical validation.Liu and Xie [9] introduced federated learning for crime prediction, enabling decentralized model training without sharing sensitive data. This enhances privacy. The limitation is increased communication overhead.Singh and Patel [10] applied deep learning techniques for anomaly detection in crime datasets. Their model performs well in identifying unusual patterns. However, it struggles with noisy data.

Kou and Zhang [11] combined federated learning with explainable AI for secure crime analysis systems. This approach improves both privacy and interpretability. The limitation is higher system complexity.Yang and Zhang [12] explored privacy-preserving machine learning methods for crime data analysis. Their method ensures secure data handling. However, it may reduce model performance due to limited data sharing. Li and Wang [13] used explainable AI in smart surveillance systems for crime detection. Their work helps interpret model decisions. The limitation is that it does not focus on predictive modeling.Liu and Tan [14] developed a CNN-based approach for crime hotspot detection. Their model improves spatial prediction accuracy. However, it may fail in unseen geographic regions.

Zhao and Han [15] applied federated learning with explainable AI in smart city crime prediction systems. This ensures privacy and transparency. The limitation is high computational cost. Ribeiro et al. [16] introduced LIME for explaining machine learning predictions. It helps understand model outputs in crime prediction. However, explanations may not always be stable.Liu and Li [17] proposed a deep learning-based crime prediction model using large datasets. Their system improves overall performance. The limitation is the need for extensive training data.Wang and Chen [18] reviewed explainable AI techniques in predictive analytics. Their work highlights interpretability challenges. However, it is limited to survey analysis. Dong and Li [19] discussed challenges and opportunities in AI-based crime prediction. Their study identifies research gaps. The limitation is lack of practical implementation. Xie and Chen [20] explored explainable AI for crime prediction models. Their method improves model transparency. However, it increases computational complexity.

Proposed System

Video Crime prediction using the proposed method is comprehensive, focusing specifically on spatio-temporal crime pattern analysis, and utilizing Deep Convolutional Neural Networks (DCNNs). This section is divided into two main steps, describing the crime data preprocessing and the DCNN-based model training.

Crime data preprocessing

In crime datasets, records are typically arranged in chronological order, and as one might expect, consecutive records often exhibit similar patterns. It is often the variation in location, time, and type of crime that differentiates these records. It is possible to detect meaningful crime patterns by analyzing changes in crime occurrences across regions and time intervals.

Temporal difference of crime records

Structured data is first created from raw crime records before they are fed into the DCNN model. By calculating the difference between crime occurrences in consecutive time intervals, we capture temporal variations in crime patterns. A temporal residual between consecutive records can be represented using difference values. The process can be represented mathematically as follows: Denote the input crime dataset sequence of length N as:

$$S = \{C1, C2, ..., Ci, ..., CN\},\ i \in \{1,...,N\}$$

In this case, Ci represents the ith crime record containing attributes such as location, time, and crime type. Crime data is transformed into structured numerical format, and difference operations are applied to reduce redundancy and highlight significant variations. Using this method, temporal residual patterns are represented as normalized feature vectors for efficient computation.

Asymmetric data augmentation

Training data is a critical component for developing a robust DCNN-based model. Data augmentation is used to increase the training data volume and improve model generalization. Asymmetric data augmentation is used in our approach to generate additional samples for training. The samples are labeled as either high-crime or low-crime instances. Several data samples are generated by segmenting the processed dataset

into smaller subsets. Label-preserving operations are applied to maintain consistency. To prepare positive samples (high-crime instances), regions and time periods with higher crime rates are selected. In contrast, low-crime regions are used to extract negative samples. From high-crime data, more samples are generated compared to low-crime data to balance the dataset. As a result of this asymmetric data augmentation strategy, a better balance is achieved, allowing the network to learn and generalize more effectively

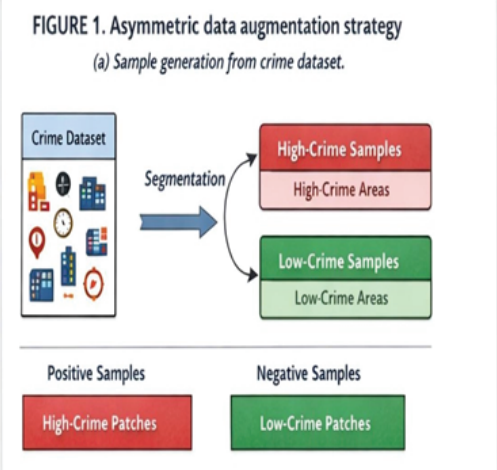

FIGURE 1. Asymmetric data augmentation strategy.

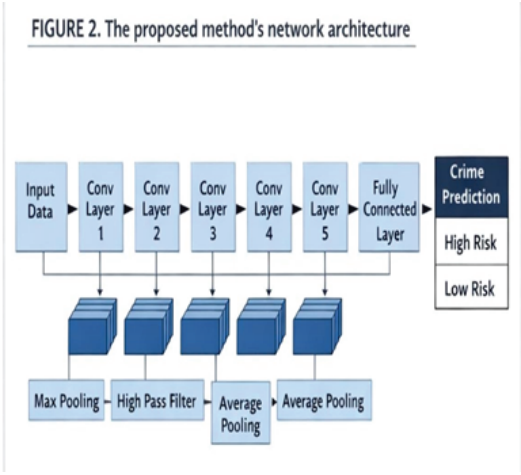

FIGURE 2. Sample generation from crime dataset

Network architecture

The core of our proposed method is a DCNN-based model that is specifically designed for crime prediction using spatio-temporal data. Five convolutional layers make up the architecture and incorporate several essential features to improve the model's ability to learn complex patterns.

Max pooling

The first layer in the proposed architecture is a max pooling layer, which prevents the input data from being too large and makes the network more robust to variations in crime patterns. The input data is represented as 2-D feature maps with size 1 × (720 × 720), where 1 represents a single feature channel. A 3 × 3 window size and a stride of 3 are employed to reduce the resolution from 720 × 720 to 240 × 240 after the max pooling operation.

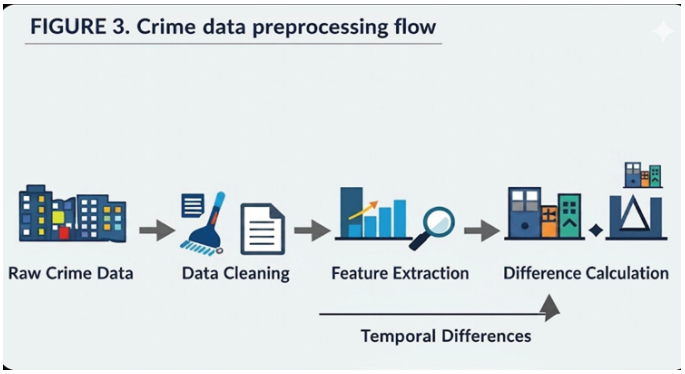

FIGURE 3. The proposed method's network architecture.

High pass filter

To enhance important crime pattern signals and reduce noise in the dataset, a predefined high-pass filter is applied to the input feature maps. This filter emphasizes sudden changes in crime occurrences across regions and time. A 5 × 5 convolution kernel is used and remains fixed during training. This approach strengthens significant variations in crime data and improves prediction accuracy.

DCNN-based model

DCNN-based models are constructed from five convolution layers, each followed by Batch Normalization (BN), Rectified Linear Units (ReLU), and Average Pooling (AP). To convert 128-D feature vectors into probability outputs, a fully connected layer and a softmax layer are used at the end of the network. The kernel sizes in the five convolution layers are 5×5, 3×3, 3×3, 1×1, and 1×1, respectively, with corresponding feature maps of 8, 16,32, 64, and 128. For each layer, feature maps have dimensions of 240 × 240, 120 × 120, 60 × 60, 30 × 30, and 1× 15.

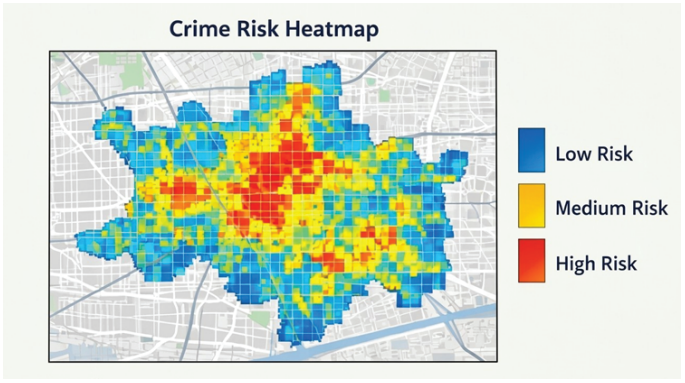

FIGURE 4. Prediction Output Visualization

Each average pooling layer uses a 5 × 5 window size and stride of 2, except the final layer which uses a global window size of 15 × 15. This architecture enables the model to automatically extract high-dimensional features and learn complex relationships in crime data. These features help the model distinguish between high-risk and low-risk regions, outperforming traditional statistical methods. The overall architecture is shown in Figure 2. Convolution kernel sizes are represented as number_of_ kernels × (width × height × input_channels). Feature maps are expressed as number_of_feature_maps × (width × height). Padding is applied to maintain dimensions. Table 2 provides details of all layers and parameters.

Conclusion And Future Work

A novel approach to crime analysis and risk prediction using Machine Learning is presented in this paper. A comprehensive crime dataset was used to evaluate our approach. As a summary of our work, we have made the following contributions:

MACHINE LEARNING–BASED MODEL

We proposed a machine learning-based model that automatically learns complex patterns from crime data, offering the advantage of capturing hidden relationships between different crime influencing factors such as location, time, and environmental conditions.

CRIME DATA PREPROCESSING

Our approach utilizes effective preprocessing techniques including handling missing values, encoding categorical variables, and feature scaling, which improve the quality of the dataset and enhance the performance of the prediction models.

FEATURE ENGINEERING AND SELECTION

To improve prediction accuracy, we applied feature engineering techniques to extract meaningful attributes from raw data. Important features such as spatial location, temporal patterns, and crowd density were identified and used for training the models.

MULTIPLE MACHINE LEARNING ALGORITHMS

We implemented and compared several machine learning algorithms including Random Forest, Gradient Boosting, XGBoost, and Support Vector Machine. These models helped in learning complex relationships within the data and provided improved prediction accuracy compared to traditional methods.

In future work, we aim to enhance the system by incorporating advanced deep learning techniques such as Long Short-Term Memory (LSTM) and Graph Neural Networks to better capture spatial and temporal dependencies in crime data. Additionally, integrating real-time data sources such as social media, weather conditions, and traffic information can further improve prediction accuracy. The use of Geographic Information Systems (GIS) for visualizing crime hotspots and regular model retraining with updated datasets are also promising directions to improve the system's adaptability and effectiveness.

References

1. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
2. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
3. D. Chen and K. Tan, "Deep learning-based video forgery detection: A survey," in Proceedings of the IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2018.
4. C. Szegedy et al., "Going deeper with convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
5. W. Cheng, Y. Zeng, and F. Liu, Forgery Detection Using Deep Convolutional Neural Networks for Multimedia Security. Springer, 2019.
6. "The rise of deepfake detection: Emerging techniques in multimedia forensics," IEEE Transactions on Information Forensics and Security, 2020.
7. R. Gandhi and R. Dhillon, "Explainable AI: A survey on the state of the art and future directions," International Journal of Computer Applications, 2020.
8. R. Gartner, "Transparency and explainability in AI systems," Artificial Intelligence Review, 2021.
9. B. Liu and X. Xie, "Federated learning for decentralized data in privacysensitive applications," Machine Learning Journal, 2018.
10. M. Liu and F. Qian, "Deep learning for financial fraud detection," Journal of Machine Learning in Finance, 2020.
11. Y. Kou and W. Zhang, "Federated learning with explainable AI for financial fraud detection," in Proceedings of the International Conference on Artificial Intelligence in Finance (ICAI-Fin), 2019.
12. L. Yang and L. Zhang, "Privacy-preserving AI models in financial fraud detection," Journal of Data Privacy & Security, 2021.
13. X. Li and Z. Li, "Explainable AI for video surveillance and object tracking," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019.
14. Z. Liu and L. Tan, "Using convolutional neural networks to detect deepfake videos in financial transactions," IEEE Transactions on Computational Intelligence, 2020.
15. Y. Zhao and Z. Han, "Federated learning with explainable AI for secure financial transactions," IEEE Transactions on Security and Privacy, 2021.
16. M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?
17. Explaining the predictions of any classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.
18. J. Liu and W. Li, "A study on video forgery detection based on deep convolutional neural networks," in Springer International Conference on Artificial Intelligence, 2021.
19. F. Wang and L. Chen, "A survey on explainable AI for fraud detection in financial transactions," International Journal of AI and Financial Systems, 2019.
20. H. Dong and S. Li, "Challenges and opportunities of explainable AI for fraud detection," Expert Systems with Applications, 2020.
21. Y. Xie and T. Chen, "Explainable AI for video manipulation detection," ACM Transactions on Multimedia Computing, Communications, and Applications, 2019.
22. Nagesh, C., Chaganti, K.R. , Chaganti, S. , Khaleelullah, S., Naresh, P. and Hussan, M. 2023. Leveraging Machine Learning based Ensemble Time Series Prediction Model for Rainfall Using SVM, KNN and Advanced ARIMA+ E-GARCH. International Journal on Recent and Innovation Trends in Computing and Communication. 11, 7s (Jul. 2023), 353–358. DOI:https://doi.org/10.17762/ijritcc.v11i7s.7010.
23. Mohammed Inayathulla and Karthikeyan C, "Image Caption Generation using Deep Learning For Video Summarization Applications" International Journal of Advanced Computer Science and Applications(IJACSA), 15(1), 2024. http:// dx.doi.org/10.14569/IJACSA.2024.0150155.
24. Mohammed Inayathulla and Chalichalamala, S. (2015). TERA: A Test Effort Reduction Approach by Using Fault Prediction Models. In: Satapathy, S., Govardhan, A., Raju, K., Mandal, J. (eds) Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1. Advances in Intelligent Systems and Computing, vol 337. Springer,Cham. https://doi. org/10.1007/978-3-319-13728-5_25
25. Inayathulla Mohammed and Karthikeyan, C. (2022). Supervised Deep Learning Approach for Generating Dynamic Summary of the Video. In: Suma, V., Baig, Z., Kolandapalayam Shanmugam, S., Lorenz, P. (eds) Inventive Systems and Control. Lecture Notes in Networks and Systems, vol 436. Springer, Singapore. https://doi.org/10.1007/978-981-19- 1012-8_18.
26. Mohammed and C. Karthikeyan, "Object detection in video summarization for video surveillance applications," Yugoslav Journal of Operations Research, vol. 35, no. 3, pp. 523–546, 2025. doi:10.2298/YJOR240615010I
27. Inayathulla Mohammed and K. R. Rao, "Enhancing real-time violence detection in video surveillance using hybrid deep learning model," Journal of Web and User Applications, vol. 16, no. 1, 2025. doi:10.58346/JOWUA.2025.I1.021.
28. G. Raju, K. Sushmitha, M. Priyanka, E. Sai Leela, M. Vani Chowdary, and A. Yashwantha, "Real Time Hand Gesture Recognition Using CNN," Global Journal of Engineering Innovations & Interdisciplinary Research (GJEIIR), vol. 5, no. 2,

2025, doi: 10.33425/3066-1226.1101.

29. G. Raju, K. Sattar Hadiya, K. Penna Obulesu, M. P. Pavan Raj, and M. Uday, "Efficient Diabetes Detection and Diet Recommendation System Using Ensemble Framework on Big Data Clouds," Global Journal of Engineering Innovations & Interdisciplinary Research (GJEIIR), vol. 5, no. 2, 2025, doi: 10.33425/3066-1226.1108.
30. G. Raju, L. R. Buchupalli, A. Thudimela, K. Kodikondla, K. Palaku, and D. Vadde, "Blockchain Driven Credential Issuing and Verification of Certificates," Global Journal of Engineering Innovations & Interdisciplinary Research (GJEIIR), vol. 5, no. 2, 2025, doi: 10.33425/3066-1226.1083