International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026

Next-Generation Machine Learning for crime prediction and community Safety

Puthi Lokesh[1], Mallula Kiran[2], Auvula karthik[3], V. Rani[4]

Department of Computer Science and Engineering (AI/ML), NRI Institute of Technology, Agiripalli, Andhra Pradesh

E-mail: rani.v@nriit.edu.in

Abstract

In order to promote proactive public safety management, this project offers a next-generation machine learning framework for secure communication safety and crime prediction. To more accurately forecast possible crime hotspots, the suggested approach combines past crime data with current contextual elements including weather, public events, and population dynamics. The platform detects high-risk areas and times by utilizing sophisticated machine learning techniques and spatiotemporal modelling, allowing for early preventive measures. Apart from forecasting, the system integrates a safe communication system that enables prompt and dependable sharing information between the public and law enforcement. By offering practical insights for the best use of resources, an interactive visualization dashboard helps decision-making even more.

Keywords— Crime prediction, Machine learning, Spatiotemporal analysis, Crime hotspot detection, public safety, Predictive policing, Secure communication, Real-time data analytics, Smart city safety, Urban crime analysis, Risk assessment, Law enforcement decision support, Data-driven policing, Ethical AI, Community safety

I.INTRODUCTION

Rapid urbanisation, population increase, and changing social dynamics have made crime prevention an increasingly difficult task. The majority of traditional crime analysis systems are reactive, concentrating on documenting and looking into crimes after they happen. These methods restrict law enforcement organisations' capacity to foresee criminal activity and put preventive measures in place beforehand. There is a great need for intelligent systems that can find hidden patterns and trends due to the increasing availability of large-scale crime datasets and urban contextual information. Predictive insights that enable proactive police and better public safety planning across metropolitan contexts are made possible by machine learning's potent tools for processing massive amounts of organised and unstructured data.One interesting avenue in crime analytics is the combination of past crime data with current contextual elements. Although they are frequently disregarded in conventional models, variables including weather, public events, population density, and temporal patterns have a substantial impact on crime occurrences. Predictive systems can become more accurate and flexible in response to shifting urban environments by including these dynamic aspects. Understanding how crime changes across time and space requires the use of spatiotemporal machine learning techniques. By identifying possible criminal hotspots, predicting high-risk times, and allocating resources more efficiently, this data-driven strategy helps authorities make preventive decisions and improve situational awareness.

Effective crime prevention relies on fast and secure communication between law enforcement and citizens in addition to prediction accuracy. Current communication systems are frequently disjointed or susceptible to security threats, which causes replies to be delayed and trust to be diminished. In order to overcome these issues, this project suggests a unified architecture that incorporates secure communication, visualisation support, and crime hotspot prediction. The technology promotes informed decision-making and community interaction by providing actionable insights via encrypted channels and user-friendly dashboards. Promoting moral, open, and proactive crime prevention tactics that enhance public safety while reducing prejudice and boosting confidence in intelligent policing systems is the main goal.

II.LITERATURE SURVEY

From conventional statistical and rule-based techniques to sophisticated machine learning-driven methodologies, crime prediction research has undergone substantial change. In order to detect high-crime locations using historical records, early research concentrated on hotspot mapping and kernel density estimation [10], [11]. In order to increase prediction accuracy, researchers started adding socioeconomic and demographic characteristics when large-scale urban data became available [7], [12]. The significance of contextual factors such physical environments, population density, and mobility patterns in influencing crime dynamics has beenhighlighted in recent studies [2], [14]. These investigations show that crime is impacted by temporal and spatial regularities rather than being random. Nevertheless, the majority of these methods continue to rely heavily on static historical data, which limits their ability to adjust to quickly evolving urban settings and the demands of real-time crime prevention [1], [5].

To improve openness and confidence in predictive policing systems, attention-based and comprehensible models like AIST have also been suggested [4]. Even though these methods demonstrate better predictive performance, many studies ignore deployment-oriented issues including real-time integration, scalability, and operational usefulness for law enforcement agencies in favour of algorithmic correctness [8], [13].

The impact of predictive policing systems on society and ethics is another significant aspect that has been emphasised in recent studies. Concerns about data bias, fairness, and the possibility that algorithmic decision-making would exacerbate already-existing societal injustices have been voiced by academics [16], [17]. In particular, when crime prediction algorithms interact with citizen-generated data and communication platforms, privacy and safe data management have become crucial challenges [15], [18]. Few research incorporate explainable AI and fairness-aware learning into end-to-end crime prediction frameworks [5], [20]. Furthermore, there is a clear research need that drives the proposed project because there is little study on secure communication systems that guarantee prompt and private information exchange between citizens and authorities [1], [19].

Inference of the Literature Survey

According to the literature review, research on crime prediction has advanced from simple statistical and hotspot mapping methods to sophisticated machine learning and deep learning-based spatiotemporal models. Previous research shows that adding geographical, temporal, and contextual elements greatly enhances hotspot detection and forecast accuracy. Attention mechanisms, hybrid spatiotemporal frameworks, and graph-based networks have demonstrated great promise in capturing intricate crime patterns in metropolitan settings. The review does, however, also highlight some significant shortcomings in the state of the art, most notably the excessive dependence on historical data and the absence of real-time contextual integration. Furthermore, ethical issues including prejudice, openness, and privacy are frequently neglected. Most significantly, secure communication techniques are rarely combined with prediction models in prior studies.

III. PROPOSED METHODOLOGY

1. Data Collection and Integration:

The first step in the suggested technique is the methodical gathering of structured crime data from reputable, openly accessible sources. The main dataset is made up of historical crime records that have been gathered from official government repositories and include crime statistics broken down by year and location. Standard data analysis tools are used to ingest this data, which is then arranged into a structured manner appropriate for further study. In order to facilitate time-based and spatial analysis, temporal consistency and regional indexing are prioritised. The system's basic layer is made up of the gathered data, which makes it possible to identify regional crime distributions and long-term crime trends. This stage guarantees that the predictive framework is based on the real, thorough, and well-structured data required for precise forecasting of crime hotspots.

2. Data Preprocessing and Cleaning:

To improve data quality and dependability, preprocessing is carried out following data collection. During this stage, duplicate records are eliminated, missing or incomplete values are handled, and dataset discrepancies are fixed. While categorical data like regions and crime kinds are encoded into machine-readable formats, numerical properties like crime counts are normalised to minimise scale bias. Effective time-series analysis is made possible by the temporal characteristics' structure, which maintains chronological order. To increase learning efficiency, irrelevant characteristics and noise are eliminated. By ensuring that the dataset is clean, consistent, and machine learning-optimized, this preprocessing phase directly contributes to increased model accuracy and stability in crime prediction.

3. Feature Engineering and Spatiotemporal Representation:

Preprocessed data is transformed into meaningful representations that describe crime dynamics through the use of feature engineering. Features like crime frequency, regional intensity, and temporal variation are obtained by aggregating crime records over spatial units and temporal frames. These characteristics enable the system to simulate the time evolution of crime patterns as well as their spatial

distribution. The framework approximates the spatiotemporal linkages necessary for hotspot detection by collecting correlations between location and time. Learning algorithms can identify recurrent patterns and abnormalities in crime incidents thanks to this structured feature representation, which offers a solid analytical foundation for predicting possible crime-prone regions.

4. Model Training and Crime Hotspot Prediction:

Using the designed features, supervised machine learning models are trained during the predictive phase. Because gradient boosting-based algorithms can model intricate, non-linear correlations in crime data, they are used. To guarantee objective assessment, the dataset is split into training and testing subsets. The program creates risk scores for various locations and eras by learning past crime trends during training. To determine high-risk areas that are probably going to see higher crime rates, these ratings are combined. Instead of depending just on retrospective analysis, this stage transforms analytical findings into predictive outputs, allowing for the proactive identification of criminal hotspots.

IV. WORKFLOW

1. Data Loading and Initial Exploration:

To handle datasets, visualise them, and train models, the workflow starts with the import of crucial machine learning and data analysis libraries. After that, the crime dataset is loaded from a CSV file that contains crime statistics broken down by district over a number of years. To comprehend the dataset structure, including the number of rows and columns, data kinds, and general schema, preliminary investigation is carried out. To guarantee data integrity and completeness, duplicate records and missing values are checked. To see trends in the distribution of crimes, descriptive statistics are produced. This preliminary investigation builds trust in the quality of the data and gets it ready for more analytical and predictive processing.

2. Visualisation and Analysis of Exploratory Data:

To find trends and patterns in the crime data, exploratory data analysis is carried out after data quality has been verified. In order to discover temporal fluctuations in the overall number of crimes, visualisations are created to analyse crime trends throughout several years. Districts with the greatest crime rates are highlighted using bar charts, which provide information on the geographical concentration of criminal activity. Understanding past criminal activity and regional differences is made easier by these visual studies. In order to ensure that the predictive model is based on significant and observable data patterns, the insights gleaned from exploratory analysis feed feature selection and model building.

3. Feature Selection and Target Variable Creation:

To build the prediction model, pertinent features are chosen from the dataset during this phase. To represent both spatial and criminal-specific characteristics, important parameters like state, district, year, and major crime categories are extracted. Next, a threshold based on the higher quartile of total crime counts is used to establish a target variable that represents the level of crime risk. High-risk areas are those that surpass this threshold, whereas low-risk areas are those that do not. This binary classification method turns the crime prediction problem into a hotspot identification-focused supervised learning challenge.

4. Model Training, Data Encoding, and Splitting:

Label encoding is used to transform the dataset's category variables—like states and districts—into numerical form that is appropriate for machine learning techniques. To guarantee objective model evaluation, the dataset is subsequently split into training and testing subsets. Because of its resilience to overfitting, robustness, and capacity to manage intricate feature interactions, a Random Forest classifier is chosen. In order to distinguish between high-risk and low-risk areas based on past crime characteristics, the model is trained using the training data.

5. Model Assessment and Preparation for Deployment:

The model's prediction performance is assessed using the test dataset when training is finished. To gauge the efficacy of classification, evaluation measures including accuracy, precision, recall, and F1-score are computed. The outcomes show how well the model can identify criminal hotspots. The trained model is serialised and saved for usage in deployment contexts after validation. In order to facilitate proactive crime monitoring and decision-making, this last phase makes it possible to integrate the predictive model into applications like dashboards or alarm systems.

SYSTEM ARCHITECTURE:

The proposed project's system architecture is built as a modular, layered framework that facilitates decision support and end-to-end crime hotspot prediction. Historical crime records are imported from structured datasets and kept in a central repository for analysis at the data layer. Data consistency, integrity, and effective processing access are guaranteed by this layer.

International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026

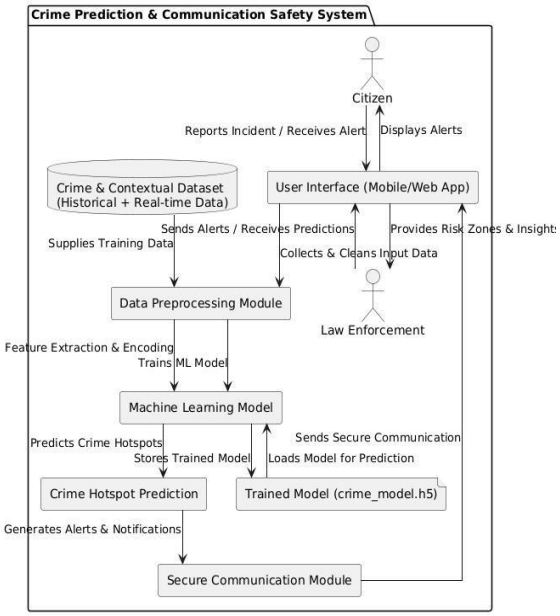


Fig 1 System Architecture

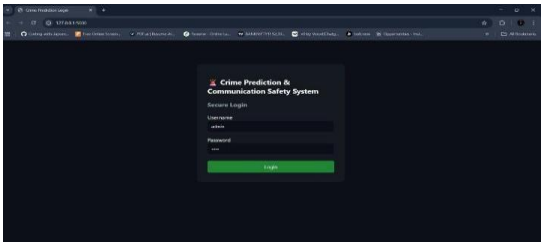
Preprocessing tasks including cleaning, normalisation, feature extraction, and categorical attribute encoding are carried out by the data processing layer above this. This layer prepares spatiotemporal representations of crime data through the use of exploratory analysis and aggregation processes. The machine learning layer, which houses the trained predictive models in charge of identifying crime trends and producing risk-level forecasts for various locations and time periods, receives the processed data after that.

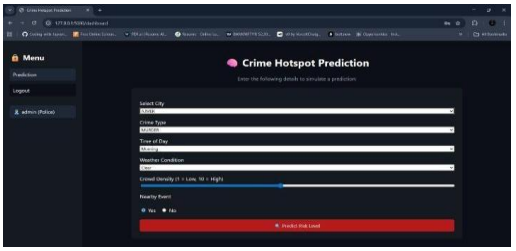
The highest part of the system architecture is the application and decision-support layer. This layer converts the machine learning models' prediction outputs into insights that may be put to use. Interactive dashboards that display crime patterns, high-risk areas, and temporal fluctuations in an understandable way are used to visualise risk scores and hotspot classifications. Authorities can be informed about possible crime-prone areas by integrating a threshold-based alert system, which enables preemptive action and efficient esource allocation. In order to preserve forecast accuracy over time, the design also allows for model persistence and recurring updates. In order to ensure that the framework can be expanded with more data sources, sophisticated models, or real-time inputs for improved public safety applications, the system architecture generally places an emphasis on scalability, modularity, and interpretability.

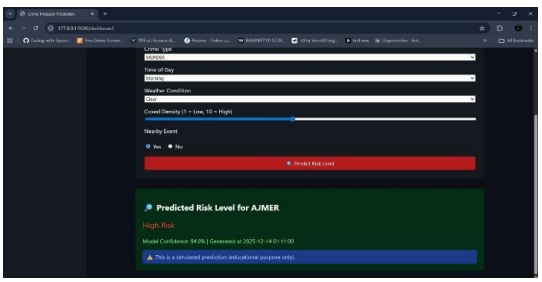
VI RESULTS AND DISCUSSION

The project's outcomes show how well the Random Forest-based crime hotspot prediction model can distinguish between high-risk and low-risk areas. The model obtained an overall accuracy of almost 98% using historical district- level crime data and certain crime-related variables, demonstrating a strong capacity to identify intricate patterns within the dataset. High recall and precision scores further attest to the model's ability to minimise misclassification while accurately identifying crime-prone locations. By combining predictions from several decision trees, the Random Forest algorithm's ensemble nature increases robustness and lessens overfitting. The results discussion emphasises how important threshold-based risk labelling and feature selection were in improving predictive performance.

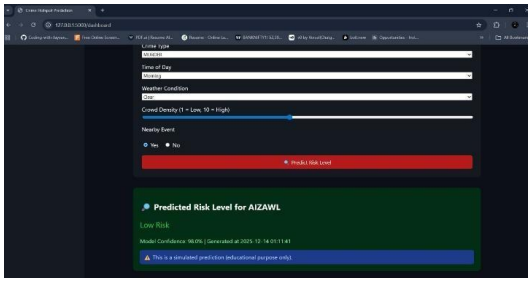
The project's debate emphasises the usefulness of using machine learning methods for public safety planning and crime hotspot prediction. Reliable classification of high-risk and low-risk areas results from the efficient management of intricate and non-linear correlations among crime-related indicators through the use of a Random Forest model. The high forecast accuracy suggests that preventative policing tactics can be supported by strong signals found in past crime trends. To prevent biassed or overconfident forecasts, the results further highlight the significance of rigorous feature selection and balanced risk thresholds. Although the current method performs well on organised historical data, its efficacy could differ depending on the location and time period. As a result, ongoing model assessment and upgrades are crucial.







International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026



VII CONCLUSION

To sum up, this study effectively illustrates how machine learning techniques can be applied for proactive crime hotspot prediction and decision assistance. The suggested technique successfully detects high-risk and low-risk areas by evaluating historical, district-level crime data and converting it through methodical preprocessing and feature engineering. The adoption of a Random Forest–based model enables robust learning of complicated criminal patterns, resulting in high predicted accuracy and reliable categorisation outputs. Law enforcement agencies can make well-informed decisions and allocate resources strategically when analytical insights are integrated with visualisation and alert mechanisms. The initiative also highlights the significance of data-driven and ethical methods in predictive policing.

REFERENCES

1. Ahmad, R., Nawaz, A., & Mustafa, G. (2024). CHART: Intelligent crime hotspot detection and real-time tracking using machine learning. Computers, Materials & Continua, 81(3), 4171–4194.
2. De Nadai, M., Xu, Y., Letouzé, E., González, M. C., & Lepri, B. (2020). Socio-economic, built environment, and mobility conditions associated with crime: A study of multiple cities. arXiv.
3. Zubair, T., Fatima, S. K., Ahmed, N., & Khan, A. (2025). Crime hotspot prediction using deep graph convolutional networks. arXiv.
4. Rayhan, Y., & Hashem, T. (2020). AIST: An interpretable attention-based deep learning model for crime prediction. arXiv.
5. Gupta, S., & Sayer, S. (2024). Machine learning for public good: Predicting urban crime patterns to enhance community safety. arXiv.
6. Wang, Y., Chen, Z., & Li, X. (2025). Research on a crime spatiotemporal prediction method integrating Informer and spatiotemporal graph convolutional network. Information, 9(7),
7. Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., & Pentland, A. (2014). Once upon a crime: Towards crime prediction from demographics and mobile data. Proceedings of the 16th ACM International Conference on Multimodal Interaction, 427–434.
8. Gerber, M. S. (2014). Predicting crime using Twitter and kernel density estimation. Decision Support Systems, 61, 115–125.
9. Wang, T., Rudin, C., Wagner, D., & Sevieri, R. (2017). Learning to detect patterns of crime. Machine Learning, 106(9–10), 1473–1495.
10. Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. Security Journal, 21(1–2), 4–28.
11. Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. Journal of the American Statistical Association, 106(493), 100–108.
12. Brantingham, P. J., & Brantingham, P. L. (1995). Criminality of place: Crime generators and crime attractors. European Journal on Criminal Policy and Research, 3(3), 5–26.
13. Eck, J. E., Chainey, S., Cameron, J. G., Leitner, M., & Wilson, R. E. (2005). Mapping crime: Understanding hotspots. National Institute of Justice.
14. Weisburd, D., Groff, E. R., & Yang, S. M. (2012). The criminology of place: Street segments and our understanding of the crime problem. Oxford University Press.
15. Kshetri, N. (2014). Big data's impact on privacy, security and consumer welfare. Telecommunications Policy, 38(11), 1134–1145.
16. Lum, K., & Isaac, W. (2016). To predict and serve? Significance, 13(5), 14–19.
17. Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. Advances in Neural Information Processing Systems, 29, 3315–3323.
18. Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. SAGE Publications.
19. National Crime Records Bureau. (2013). Crime in India 2012: Statistics. Ministry of Home Affairs, Government of India.
20. Rajanand. (2022). Crime in India dataset (2001–2012) [Data set].Kaggle.