International Journal of Web of Multidisciplinary Studies
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: https://ijwos.com
Volume 3 Issue 9, September 2026
E-ISSN : 3049-2424
DOI: https://doi.org/10.71366/ijwos

AI-BASED CRIME HOTSPOT PREDICTION SYSTEM USING MACHINE LEARNING

Mrs. Manjula K[1], Mr.Amith Bhaskar G[2]
[1] *Assistant Professor, Department of Computer Applications, GM University, Davangere, Karnataka, India*
[2] *Student, Master of Computer Applications, GM University, Davangere, Karnataka, India.*

Article Info

Article History:

Published: 08 Sept 2026

Publication Issue:
Volume 3, Issue 9
September-2026

Page Number:
55-90

Corresponding Author:
Mrs. Manjula K

Abstract:

Rising urban crime and the limitations of manual patrol scheduling have created a pressing need for data-driven policing tools. This paper presents an AI-based crime hotspot prediction system that applies supervised machine learning to historical crime records in order to forecast the localities most likely to witness criminal activity. Three ensemble learning algorithms — Random Forest, Gradient Boosting, and XGBoost — were trained on a structured crime dataset comprising incident type, location coordinates, time of occurrence, and sector-level metadata. The processed features were used to build a prediction engine that outputs a risk score for a queried location, which is then visualised through an interactive command dashboard, a hotspot tracking module, and a Google Maps-based geo-risk layer. Experimental evaluation shows that XGBoost achieved the highest overall accuracy of 73.1% along with a recall of 75.0%, narrowly outperforming Gradient Boosting and Random Forest, both of which stayed close behind on accuracy but showed competitive ROC-AUC scores of 76.4% and 74.9% respectively. The system demonstrates that even a modestly sized dataset, when combined with careful feature engineering and ensemble modelling, can support meaningful proactive policing decisions. The paper also discusses the system architecture, workflow, mathematical foundations, and the dashboard implementation, and closes with a discussion of limitations and directions for future enhancement using deep learning and real-time sensor integration.

Keywords: crime hotspot prediction, machine learning, XGBoost, random forest, gradient boosting, predictive policing, geo-risk mapping, dashboard visualization

1. INTRODUCTION

Crime remains an important challenge for urban communities, where increasing population density, mobility, and changing social conditions can contribute to complex spatial and temporal patterns of criminal activity. The increasing availability of digitised crime records has created opportunities for applying data-driven techniques to identify recurring patterns and support evidence-based decision-making. Traditional crime monitoring approaches primarily depend on historical reports, manual analysis, and the experience of law-enforcement personnel. Although these approaches remain important, they can become difficult to scale when crime records increase in volume and complexity.

Machine learning provides an opportunity to analyse large collections of historical crime records and identify relationships between crime occurrence, geographic location, time, and crime category. Previous research has demonstrated the usefulness of machine learning and statistical techniques for crime prediction, hotspot detection, and spatio-temporal analysis [3], [5], [7]. Ensemble learning methods such as Random Forest, Gradient Boosting, and XGBoost have also demonstrated strong performance on structured datasets because they can model nonlinear relationships among multiple input features [8]–[11]. However, model performance alone is insufficient for practical

crime-risk analysis. A useful system must also provide interpretable outputs that allow analysts to understand where risk is concentrated and how predictions change over time.

Crime hotspot detection is particularly important because crime incidents are not distributed uniformly across geographic areas. Certain locations may experience persistent concentrations of incidents, while other areas may develop rapidly increasing crime activity over shorter periods. Spatial clustering techniques can therefore be used to identify geographically concentrated incidents, while temporal analysis can help distinguish persistent hotspots from newly emerging areas. In this study, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is considered for identifying spatially dense crime clusters because it can detect clusters of arbitrary shape and identify isolated incidents as noise.

Existing crime information systems are often designed primarily for recording, searching, and visualising historical incidents. Geographic information system (GIS) applications can display previous crime locations effectively, but descriptive mapping alone does not necessarily provide a predictive estimate of future risk [7], [16]. Similarly, machine learning studies frequently focus on offline model evaluation without integrating the resulting predictions into an operational visualization environment. This creates a gap between predictive modelling and practical interpretation of crime-risk information.

To address this gap, this research proposes an end-to-end AI-based crime hotspot prediction framework that combines data preprocessing, spatial hotspot detection, temporal feature engineering, machine learning classification, model evaluation, and interactive geographic visualization. The system processes structured crime records and generates risk-related outputs that can be presented through a command dashboard, hotspot tracking interface, geo-risk map, and model-performance monitoring module. The proposed framework is intended to support analysts in examining crime-risk patterns rather than replacing professional judgement or automating law-enforcement decisions.

The primary objectives of this research are:

1. To develop a preprocessing pipeline for cleaning, transforming, and encoding historical crime records for machine learning analysis.
2. To identify spatial crime hotspots using a density-based clustering approach and incorporate temporal crime patterns into hotspot-risk classification.
3. To develop and compare machine learning models for classifying crime-risk levels using spatial, temporal, and crime-related features.
4. To evaluate model performance using Accuracy, Precision, Recall, F1-score, and ROC-AUC, together with appropriate confusion-matrix analysis.
5. To develop an interactive visualization framework for presenting crime incidents, predicted hotspots, risk levels, geographic distributions, and model performance.
6. To improve model transparency and practical usability through interpretable risk representations and explainable prediction outputs.
7. To examine the limitations and ethical considerations associated with applying machine learning to historical crime data and predictive policing applications.

2. LITERATURE SURVEY

Crime prediction and hotspot detection have been widely investigated using statistical methods, machine learning, deep learning, spatial analysis, and geographic information systems. Existing research demonstrates that crime

occurrence is influenced by spatial, temporal, demographic, and environmental factors, making it suitable for predictive modelling and pattern analysis.

Early crime analysis approaches primarily focused on statistical and spatial techniques for identifying areas with high concentrations of criminal incidents. Chainey et al. [1] examined hotspot mapping techniques and demonstrated the usefulness of spatial concentration analysis for identifying areas where crime is more likely to occur. Such approaches provide an important foundation for modern crime hotspot detection because they transform individual incident records into geographically meaningful patterns.

The development of machine learning techniques has enabled crime prediction systems to incorporate multiple attributes simultaneously. Cortes and Fuentes Silva [2] reviewed artificial intelligence approaches for crime prediction and discussed the application of algorithms including Naive Bayes, Support Vector Machines, Random Forest, Convolutional Neural Networks, Recurrent Neural Networks, and Long Short-Term Memory networks. Their work highlights the potential of machine learning for modelling complex relationships within historical crime data while also identifying limitations associated with dataset dependency and generalisation.

Zhang et al. [3] compared several machine learning algorithms for crime hotspot prediction, including K-Nearest Neighbours, Random Forest, Support Vector Machines, Naive Bayes, Convolutional Neural Networks, and Long Short-Term Memory networks. Their findings indicate that model performance varies according to the characteristics of the dataset, geographic region, and crime category. The study demonstrates the importance of evaluating multiple learning approaches rather than relying on a single classifier.

Mandalapu et al. [4] presented a systematic review of machine learning and deep learning approaches for crime prediction. The review examined a broad range of datasets, algorithms, and prediction strategies and identified deep learning, ensemble learning, and spatio-temporal modelling as important directions for future crime prediction research. However, systematic reviews do not themselves provide experimental validation for a particular integrated crime prediction system.

Safat et al. [5] investigated machine learning and deep learning techniques for crime prediction and forecasting using algorithms such as Logistic Regression, Support Vector Machines, K-Nearest Neighbours, Random Forest, XGBoost, Long Short-Term Memory, and ARIMA. Their research demonstrates that both conventional machine learning and temporal forecasting techniques can contribute to crime prediction, while also highlighting the influence of geographic scope and historical data on predictive performance.

Baek et al. [6] proposed a smart policing approach that predicts crime type and risk score using machine learning. The study extends crime prediction beyond simple incident classification by incorporating risk-related information for early awareness. Such approaches demonstrate the importance of providing actionable risk information rather than only reporting historical crime counts.

Butt et al. [7] conducted a systematic review of spatio-temporal crime hotspot detection and prediction. Their work highlights the importance of combining spatial and temporal information when analysing crime patterns. Crime incidents can form geographically concentrated clusters while also exhibiting temporal variations, making purely spatial or purely temporal approaches insufficient for comprehensive risk analysis.

Kshatri et al. [8] investigated an ensemble approach based on stacked generalisation for crime prediction. Ensemble learning combines the strengths of multiple predictive models and can improve robustness compared with individual classifiers. This provides motivation for investigating ensemble strategies beyond the direct comparison of Random Forest, Gradient Boosting, and XGBoost.

Breiman [9] introduced Random Forest as an ensemble learning method based on multiple decision trees trained using bootstrap samples and feature randomisation. Random Forest is widely used for structured classification problems because of its ability to model nonlinear relationships and reduce variance through aggregation.

Friedman [10] introduced the Gradient Boosting Machine, in which weak learners are sequentially trained to reduce the residual errors of previous learners. Gradient Boosting provides a strong baseline for structured predictive tasks and forms an important comparison point for more advanced boosting algorithms.

Chen and Guestrin [11] introduced XGBoost, a scalable tree-boosting framework that incorporates regularisation and second-order gradient information. Its computational efficiency and predictive capability have made it particularly useful for structured and tabular datasets.

Recent research has also explored deep learning and explainable artificial intelligence for crime prediction. Recurrent neural networks such as LSTM can model temporal dependencies in sequential crime data, while explainability techniques such as SHAP can provide information about the contribution of individual features to model predictions. These approaches are particularly relevant when predictive systems are intended for high-impact domains in which model outputs need to be interpreted carefully.

A. Research Gap

Despite the progress made in crime prediction research, several limitations remain. First, many existing studies focus on a specific city, crime category, or historical period, which can limit the generalisability of their findings. Second, a substantial portion of existing research concentrates on offline model accuracy rather than integrating predictive outputs into an operational visualization environment.

Third, conventional machine learning comparisons often focus on individual algorithms such as Random Forest, Gradient Boosting, and XGBoost without investigating how complementary models and temporal learning approaches can be combined. Fourth, hotspot detection is frequently treated as a static spatial problem, while crime risk can change over time as incident frequency increases or decreases within a geographic region.

Another important limitation is the lack of transparency in some predictive crime systems. Reporting only a final accuracy score does not provide sufficient information about why a model identifies a particular region as high risk. Explainable prediction techniques and geographic visualisation can therefore improve the interpretability of model outputs.

The proposed research addresses these gaps by combining spatial clustering, temporal analysis, machine learning-based risk classification, comprehensive model evaluation, and interactive geographic visualization within a unified framework. The proposed approach further investigates advanced ensemble and temporal modelling techniques to extend the analysis beyond conventional Random Forest, Gradient Boosting, and XGBoost comparisons. The system is designed to provide both predictive risk estimates and interpretable geographic representations while maintaining human oversight and acknowledging the ethical limitations of predictive crime analysis.

3. PROBLEM STATEMENT

Present-day crime monitoring systems in many municipal settings remain largely dependent on manual reporting, historical records, patrol observations, and experience-based decision-making. Although these approaches provide valuable information about previously reported incidents, they offer limited support for predicting where crime risk may increase in the future. As the volume of crime records grows, manually identifying meaningful spatial and temporal patterns becomes increasingly difficult.

A major challenge is the identification of geographically concentrated crime activity. Crime incidents are not uniformly distributed across an urban environment, and certain locations may experience repeated incidents within particular time periods. Without systematic spatial analysis, emerging concentrations may remain undetected until a substantial number of incidents have already occurred.

Another challenge is the integration of temporal information into hotspot analysis. A location with a historically high number of incidents may not necessarily represent the highest current risk, while an area experiencing a recent increase in incidents may require greater attention despite having a lower historical crime count. Therefore, hotspot identification should consider both spatial concentration and temporal changes in incident frequency.

Existing crime analysis systems also frequently focus on descriptive statistics and historical mapping rather than predictive risk estimation. Such systems can show where previous incidents occurred but may provide limited information about the probability or severity of future crime activity. Furthermore, model-based prediction systems may report classification performance without providing an integrated geographic interface through which analysts can interpret predicted risk.

Another problem is the limited transparency of predictive crime models. A prediction without information about the underlying spatial, temporal, or crime-related factors can be difficult for analysts to interpret and evaluate. In high-impact applications such as crime analysis, predictive outputs should therefore be presented as decision-support information rather than as definitive evidence of future criminal activity.

The problem addressed in this research is therefore the development of a data-driven system capable of analysing historical crime records, identifying spatially concentrated crime hotspots, incorporating temporal crime patterns, classifying risk levels using machine learning, and presenting the resulting predictions through an interactive geographic visualization environment.

The proposed system aims to address the following specific problems:

1. Difficulty in identifying spatial concentrations of crime incidents from large historical datasets.
2. Limited integration of temporal changes in crime frequency when determining hotspot severity.
3. Dependence on descriptive crime mapping systems that provide limited predictive capability.
4. Difficulty in comparing multiple machine learning approaches using a consistent set of evaluation metrics.
5. Limited interpretability of predictive crime-risk outputs for non-technical users.
6. Lack of an integrated platform combining hotspot detection, predictive modelling, model evaluation, and geographic visualization.
7. Potential data-quality, bias, privacy, and ethical issues associated with the use of historical crime records for predictive analysis.

To address these problems, this research develops an integrated AI-based crime hotspot prediction framework that combines spatial clustering, temporal feature engineering, machine learning classification, comprehensive performance evaluation, and interactive geographic visualization. The system is intended to provide analytical support for understanding crime-risk patterns while maintaining human oversight and avoiding the use of predictions as a substitute for professional law-enforcement judgement.

4. EXISTING SYSTEM

Most existing crime information systems are primarily designed for recording, managing, searching, and reporting historical crime incidents. These systems allow authorised users to store incident details such as crime type, location, date, time, and other relevant information. Although digital record-keeping improves the accessibility and organisation of crime information, these systems generally focus on documenting incidents that have already occurred rather than predicting future crime-risk patterns.

Geographic Information System (GIS)-based crime mapping systems provide an additional capability by displaying historical incidents on digital maps. Such systems help analysts identify areas where crime has previously been concentrated and support visual exploration of spatial patterns. However, historical mapping is primarily descriptive and does not necessarily provide a predictive estimate of where crime activity may increase in the future [7], [16].

Traditional crime analysis also depends heavily on manual examination of incident records and the experience of law-enforcement personnel. Analysts may need to examine large numbers of records to identify recurring patterns according to location, time, and crime category. As the number of incidents increases, this process becomes time-consuming and may make it difficult to detect complex relationships among multiple variables.

Another limitation of existing systems is the limited integration of temporal and spatial information. Crime concentrations can vary according to time of day, day of the week, season, and other temporal factors. A location that has historically experienced a high number of incidents may not necessarily represent the highest current risk. Systems that rely primarily on historical counts or static maps may therefore fail to capture emerging changes in crime activity.

Many existing systems also lack machine learning-based risk prediction. Without a predictive model, the system can describe where incidents occurred but cannot generate a quantitative estimate of potential crime risk for a location or time period. Machine learning techniques can address this limitation by learning patterns from historical records and producing risk classifications or probability estimates.

Furthermore, conventional systems generally do not provide comprehensive model evaluation because they are not designed around predictive machine learning. When machine learning is incorporated, some systems report only a single performance measure such as accuracy, making it difficult to understand the balance between false-positive and false-negative predictions. For crime hotspot classification, Precision, Recall, F1-score, and ROC-AUC are also important for evaluating model behaviour.

Existing systems may also provide limited transparency regarding how predicted risk is determined. A prediction presented without information about relevant spatial, temporal, or crime-related factors can be difficult for analysts to interpret. This limitation is particularly important in crime-risk applications, where model predictions should be treated as decision-support information rather than definitive conclusions.

The major limitations of existing systems can therefore be summarised as follows:

1. **Descriptive rather than predictive:** Most systems primarily record and visualise historical crime incidents without generating future risk estimates.
2. **Limited spatial analysis:** Static mapping may identify previously affected locations but may not adequately detect emerging spatial clusters.
3. **Insufficient temporal integration:** Changes in crime frequency over time are not always incorporated into hotspot classification.
4. **Manual analysis:** Large crime datasets can require considerable human effort to identify meaningful patterns.
5. **Limited machine learning integration:** Many systems do not combine spatial analysis with predictive classification models.
6. **Incomplete evaluation:** Systems that use machine learning may not consistently report Accuracy, Precision, Recall, F1-score, and ROC-AUC together.
7. **Limited interpretability:** Predictions may not clearly communicate the factors or patterns contributing to an identified risk level.
8. **Limited integrated visualization:** Predictive outputs, hotspot information, geographic maps, and model performance are often provided as separate components rather than through a unified analytical environment.

5. PROPOSED SYSTEM

The proposed system is an end-to-end AI-based crime hotspot prediction framework designed to analyse historical crime records and generate spatial and temporal crime-risk information. The framework integrates data preprocessing, feature engineering, spatial clustering, temporal analysis, machine learning-based risk classification, model evaluation, and interactive visualization. The objective is to transform raw historical crime records into interpretable risk information that can support data-driven analysis and resource planning.

The proposed system consists of the following major modules:

1) Data Collection: Historical crime records containing information such as crime category, date, time, geographic coordinates, and administrative or sector information are collected and stored in a structured dataset. The dataset is expected to contain a sufficiently large number of records to support reliable model training and evaluation.

2) Data Preprocessing: The collected records are examined for missing values, duplicate entries, inconsistent formats, invalid geographic coordinates, and other data-quality issues. Categorical attributes are standardised and transformed into machine-readable representations. Records that cannot satisfy the required data-quality conditions are removed or appropriately handled.

3) Feature Engineering: Relevant spatial, temporal, and crime-related features are derived from the processed records. Temporal features include hour of occurrence, day of the week, month, and other appropriate time-based attributes. Spatial features include latitude, longitude, geographic sector, and cluster-related information. Crime categories are encoded for use by the machine learning models.

4) Spatial Hotspot Detection: DBSCAN is applied to geographic crime coordinates to identify areas containing a high concentration of incidents. Unlike centroid-based clustering techniques, DBSCAN can identify clusters of arbitrary shape and classify isolated observations as noise. Geographic distance is calculated using an appropriate distance metric such as the Haversine distance when latitude and longitude coordinates are used.

5) Temporal Hotspot Analysis: The detected spatial clusters are analysed over time to determine whether crime activity is persistent, increasing, decreasing, or newly emerging. Recent incident counts are compared with historical counts to calculate a temporal momentum measure. This information is incorporated into hotspot severity classification so that the system does not rely solely on historical incident density.

6) Hotspot Risk Classification: Each identified spatial cluster is assigned a risk level based on its crime density and temporal behaviour. The classification method uses clearly defined thresholds so that the distinction between hotspot and non-hotspot or different risk levels is reproducible. The resulting risk score can be represented as a continuous value and used to rank geographic areas according to their estimated level of risk.

7) Machine Learning Prediction: Multiple machine learning models are trained using the engineered features to classify crime-risk levels. Random Forest, Gradient Boosting, and XGBoost are retained as baseline ensemble models. Advanced ensemble or temporal models may additionally be incorporated to investigate whether combining complementary learning approaches improves predictive performance.

8) Model Evaluation: The trained models are evaluated using the same experimental protocol and held-out test data. Performance is measured using Accuracy, Precision, Recall, F1-score, and ROC-AUC. Confusion matrices are generated to provide detailed information about correct and incorrect predictions. Where multiple classes are present, appropriate macro-averaged and per-class metrics are reported.

9) Prediction and Risk Scoring: After training, the selected model generates predicted risk classes and probability scores for new or queried locations. The prediction output is combined with the spatial hotspot information to provide a geographically meaningful representation of crime risk.

10) Interactive Visualization: Model outputs are presented through an interactive dashboard containing crime statistics, hotspot information, model-performance metrics, and geographic risk visualizations. The visualization layer

allows users to examine crime distributions and predicted risk patterns without directly interacting with the underlying machine learning implementation.

11) Geo-Risk Mapping: Predicted hotspots and their associated risk scores are displayed geographically using an interactive map. Each hotspot is represented according to its corresponding risk level, allowing users to examine the spatial distribution of predicted crime-risk areas.

12) Explainability and Transparency: The system provides interpretable information about model performance and, where an explainability method is implemented, the contribution of important features to individual or aggregated predictions. This component is intended to improve transparency and ensure that model outputs are treated as analytical evidence rather than unquestionable decisions.

The complete processing pipeline can therefore be represented as:

Crime Data Collection → Data Preprocessing → Feature Engineering → Spatial Clustering → Temporal Analysis → Hotspot Classification → Machine Learning Prediction → Model Evaluation → Risk Scoring → Dashboard and Geo-Risk Visualization

The proposed framework differs from a conventional crime-mapping system by combining historical crime analysis with spatial clustering, temporal risk assessment, machine learning prediction, and interactive visualization. It is intended to provide a unified analytical environment for identifying and interpreting crime-risk patterns while maintaining human oversight in the interpretation and use of predictive outputs.

6. SYSTEM ARCHITECTURE

The architecture of the proposed AI-based crime hotspot prediction system is organised into four major layers: the Data Layer, Processing Layer, Intelligence Layer, and Presentation Layer. Each layer performs a specific function in transforming historical crime records into interpretable crime-risk information.

A. Data Layer

The Data Layer is responsible for collecting and storing historical crime incident records. Each record may contain attributes such as crime type, date, time, latitude, longitude, and geographic sector. The collected data forms the primary input for subsequent preprocessing, spatial analysis, feature engineering, and machine learning operations.

The dataset is structured so that each crime incident represents an individual observation. Geographic coordinates are used for spatial analysis, while date and time information are used to derive temporal features. Crime-category information is retained for classification and crime-pattern analysis.

B. Processing Layer

The Processing Layer prepares the collected crime records for analysis. The first stage involves data cleaning, including the identification of missing values, duplicate records, invalid coordinates, and inconsistent categorical values. Appropriate preprocessing operations are then applied to produce a consistent dataset.

Feature engineering is subsequently performed to derive meaningful spatial and temporal attributes. Temporal attributes include hour, day, month, and other relevant time-based variables. Spatial attributes include latitude, longitude, sector information, and cluster-related characteristics.

The Processing Layer also performs spatial clustering using DBSCAN. Geographic coordinates are analysed to identify areas containing a sufficient concentration of crime incidents. Isolated observations that do not satisfy the clustering conditions are treated as noise.

C. Intelligence Layer

The Intelligence Layer represents the analytical core of the proposed system. It consists of the hotspot detection component, temporal analysis component, machine learning models, evaluation module, and prediction engine.

The hotspot detection component identifies geographically concentrated crime incidents using DBSCAN. The temporal analysis component evaluates changes in incident frequency within detected clusters. Spatial density and temporal momentum are then incorporated into hotspot-risk classification.

The machine learning component uses engineered features to train and evaluate multiple classification models. Random Forest, Gradient Boosting, and XGBoost are used as baseline ensemble approaches. Additional advanced models can be incorporated when implemented and experimentally validated.

The evaluation component compares the models using Accuracy, Precision, Recall, F1-score, and ROC-AUC. Confusion matrices are also generated to analyse the distribution of correct and incorrect predictions. The best-performing model is selected according to the predefined evaluation criteria rather than accuracy alone.

The prediction engine uses the selected model to generate risk classifications and probability scores for new observations or queried geographic locations. These outputs are passed to the presentation layer for visualization.

D. Presentation Layer

The Presentation Layer provides an interactive interface through which users can examine the results generated by the analytical components. It consists of four primary interfaces: Command Dashboard, Model Telemetry, Hotspot Tracking, and Geo-Risk Map.

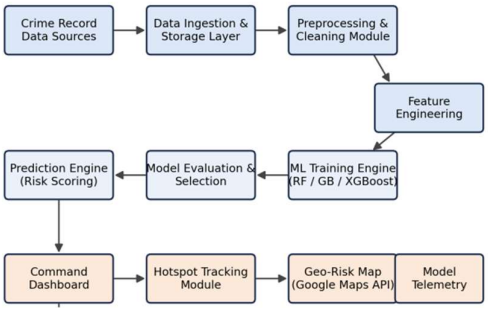
The **Command Dashboard** provides a high-level summary of crime activity, including the total number of incidents, identified hotspots, crime-category distribution, and overall risk information.

The **Model Telemetry** interface presents the comparative performance of the evaluated machine learning models. It displays the selected evaluation metrics and allows users to understand the basis for model selection.

The **Hotspot Tracking** interface displays identified crime clusters together with their geographic information, incident volume, risk level, and temporal status. This allows users to distinguish persistent hotspots from emerging or declining areas.

The **Geo-Risk Map** presents predicted hotspots on an interactive geographic map. Each location is associated with a corresponding risk classification or risk score, enabling users to examine the spatial distribution of crime-risk patterns.

E. Overall Architecture Flow



This layered architecture separates data management, analytical processing, machine learning, and visualization functions, making the proposed system easier to maintain and extend. The architecture also allows additional machine learning models, temporal forecasting techniques, or explainability modules to be incorporated without changing the overall structure of the system.

7. WORKFLOW

The workflow of the proposed AI-based crime hotspot prediction system consists of a sequence of stages that transform raw historical crime records into classified and visualised crime-risk information. The workflow integrates data preprocessing, feature engineering, spatial clustering, temporal analysis, machine learning, evaluation, and geographic visualization.

A. Data Collection

The process begins with the collection of historical crime records. Each record contains relevant information such as crime type, date, time, geographic coordinates, and sector or location information. The collected records are organised into a structured dataset for subsequent processing.

B. Data Preprocessing

The collected dataset is first examined for missing values, duplicate records, inconsistent categorical values, invalid coordinates, and malformed records. Duplicate or invalid observations are removed where appropriate, while missing values are handled using suitable preprocessing techniques. Categorical variables are standardised and converted into numerical representations required by the machine learning models.

C. Feature Engineering

After preprocessing, spatial and temporal features are extracted from the available crime records. Date and time information is transformed into variables such as hour of the day, day of the week, month, and other relevant temporal indicators. Geographic coordinates and sector information are used to derive spatial characteristics. Crime categories are encoded for machine learning and analytical purposes.

D. Spatial Hotspot Detection

The processed geographic coordinates are passed to the DBSCAN clustering algorithm. DBSCAN identifies groups of incidents that occur within a specified geographic distance and satisfy the minimum density requirement. The algorithm assigns core and border observations to clusters while treating isolated observations as noise.

The clustering parameters, including the neighbourhood distance and minimum number of observations, are determined according to the final experimental configuration. The same parameters must be reported consistently in the methodology and experimental sections.

E. Temporal Analysis

The detected spatial clusters are analysed according to their temporal distribution. Incident counts are compared across different time periods to determine whether crime activity within a cluster is persistent, increasing, or decreasing.

A temporal momentum measure is calculated to capture changes in recent crime activity relative to an earlier period. This allows an area with a rapidly increasing number of incidents to be distinguished from an area whose historical crime density remains high but is currently declining.

F. Hotspot Classification

The spatial density and temporal behaviour of each detected cluster are combined to determine its hotspot severity. A normalized crime-density value is calculated for each cluster, and the temporal momentum is incorporated into the final risk assessment.

The resulting classification assigns each cluster to an appropriate risk level according to predefined thresholds. The classification rules are applied consistently to the complete dataset so that the target variable used for machine learning is generated using a reproducible procedure.

G. Dataset Preparation for Machine Learning

The processed dataset is divided into training, validation, and testing subsets according to the experimental protocol. The splitting strategy is applied before model training, and information from the test set is not used during training or parameter optimisation.

Where class labels are imbalanced, the distribution of classes is examined and appropriate evaluation measures are used. The final test set remains isolated until the models and their configurations have been determined.

H. Model Training

The training dataset is used to develop multiple machine learning models. Random Forest, Gradient Boosting, and XGBoost are included as baseline ensemble models. Where advanced models or ensemble strategies are implemented, they are trained using the same underlying prediction task and evaluated under the same experimental conditions.

The models learn relationships between spatial, temporal, and crime-related features and the corresponding hotspot-risk labels.

I. Model Validation and Selection

The validation data is used to compare model configurations and identify the most suitable model or ensemble. Model selection is not based solely on Accuracy because false-positive and false-negative predictions are both relevant to hotspot classification.

Precision, Recall, F1-score, and ROC-AUC are therefore considered together with Accuracy when assessing model performance.

J. Model Testing

After model selection, the final models are evaluated on the previously unseen test dataset. The test predictions are used to calculate Accuracy, Precision, Recall, F1-score, and ROC-AUC.

A confusion matrix is generated from the same test predictions to ensure that the reported classification metrics are mathematically consistent with the underlying predictions.

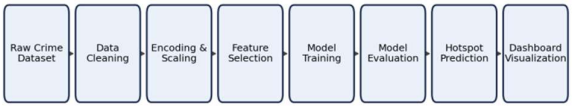
K. Risk Prediction

The selected model generates a predicted hotspot or risk classification together with a corresponding probability or confidence value. These predictions are combined with the spatial information of the corresponding locations to generate geographically meaningful risk outputs.

L. Visualization

The final predictions are passed to the visualization layer. The Command Dashboard presents overall crime statistics and risk information, while the Hotspot Tracking interface provides details about individual clusters. The Model Telemetry interface displays model performance, and the Geo-Risk Map displays predicted hotspots geographically.

M. End-to-End Workflow



The workflow ensures that the proposed system progresses systematically from raw historical crime data to machine learning-based risk prediction and geographic visualization. It also provides a clear separation between data preparation, model development, final evaluation, and presentation of results.

8. METHODOLOGY

The methodology adopted for the proposed AI-based crime hotspot prediction system consists of a sequence of data preparation, spatial analysis, temporal analysis, machine learning, evaluation, and visualization stages. The methodology is designed to identify geographically concentrated crime activity, incorporate temporal changes in crime occurrence, classify hotspot risk, and evaluate the predictive performance of multiple machine learning models.

A. Dataset Collection

The study uses historical crime records containing information related to crime category, date, time, geographic coordinates, and location or sector information. The dataset is prepared in a structured tabular format so that each row represents an individual crime incident.

The dataset is enlarged beyond the original 100-record dataset to provide a more reliable basis for model training and evaluation. The final number of records used in the experiments, the data source, geographic coverage, time period, and number of records remaining after preprocessing shall be reported explicitly in the experimental setup.

The primary attributes used in the analysis include:

- Crime type
- Date of occurrence
- Time of occurrence
- Latitude
- Longitude
- Geographic sector or location
- Derived spatial features
- Derived temporal features
- Hotspot or risk-level label

B. Data Preprocessing

The collected crime records are subjected to preprocessing before spatial and machine learning analysis. Duplicate records are identified and removed to prevent repeated incidents from disproportionately influencing the model. Missing values are examined for each attribute, and appropriate handling techniques are applied according to the characteristics of the missing data.

Geographic coordinates are validated to identify records containing invalid or incomplete latitude and longitude values. Categorical fields such as crime type and sector are standardised to eliminate inconsistencies in naming and representation.

Date and time fields are converted into a consistent format. The processed timestamp information is subsequently used to derive temporal variables required for hotspot analysis and machine learning.

C. Feature Engineering

Feature engineering is performed to transform the raw crime records into variables that can represent spatial, temporal, and categorical characteristics of crime activity.

Temporal features include:

- Hour of occurrence
- Day of the week
- Month
- Time-period category
- Historical incident frequency
- Recent incident frequency

Spatial features include:

- Latitude
- Longitude
- Geographic sector
- Spatial cluster identifier
- Local crime density

Crime-category features are encoded using an appropriate categorical encoding method before being supplied to the machine learning algorithms.

Feature engineering is performed using information available within the appropriate training period to minimise information leakage between training and testing data.

D. Spatial Hotspot Detection Using DBSCAN

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is used to identify geographically concentrated crime incidents. DBSCAN is selected because it can identify clusters of arbitrary shape without requiring a predefined number of clusters and can classify isolated observations as noise.

The clustering process uses geographic coordinates as input. When latitude and longitude are used, the Haversine distance is employed to account for the Earth's geographic curvature.

The proposed configuration uses an epsilon neighbourhood distance of 0.5 km and a minimum number of five incidents for forming a dense cluster. These parameters are retained only when supported by the final experimental configuration and validation procedure.

An observation is treated as a core point when a sufficient number of neighbouring incidents occur within the specified radius. Border points connected to core points are assigned to the corresponding cluster, while isolated observations are classified as noise.

E. Hotspot Density Calculation

After spatial clusters are identified, the incident density of each cluster is calculated. For a cluster i, the normalized crime density is defined as:

$$\rho i = \frac{C_i}{C_{max}}$$

where Ci represents the number of incidents associated with cluster i, and Cmax represents the maximum incident count among the identified clusters.

The normalized density value is used to represent the relative concentration of crime incidents within each geographic cluster.

F. Temporal Hotspot Analysis

Spatial density alone may not adequately represent current crime risk. Therefore, temporal changes in incident frequency are incorporated into the hotspot analysis.

The crime incidents within each cluster are divided into an earlier period and a recent period. Temporal momentum is calculated as:

$$\Delta_i = \frac{C_{recent,i} - C_{prior,i}}{C_{prior,i}} \times 100$$

where Crecent,i represents the number of recent incidents in cluster i, and Cprior,i represents the number of incidents during the corresponding previous period.

A positive value indicates increasing crime activity, while a negative value indicates decreasing activity. When the prior-period count is zero, an appropriate predefined handling rule is applied to avoid division by zero.

G. Hotspot Severity Classification

Each detected spatial cluster is classified according to its normalized crime density and temporal behaviour. The proposed density-based classification consists of four tiers:

Tier	Density Range	Interpretation
T1	$\rho \geq 0.70$	Persistent high-density hotspot
T2	$0.40 \leq \rho < 0.70$	Emerging hotspot
T3	$0.20 \leq \rho < 0.40$	Newly active hotspot
T4	$\rho < 0.20$	Lower or declining activity

A cluster whose recent crime activity increases by more than 20% compared with the previous period may be promoted by one risk tier. This temporal adjustment allows the methodology to capture rapidly emerging areas that may not yet have the highest historical crime density.

H. Hybrid Risk Score

To combine spatial concentration and temporal momentum, a continuous hotspot risk score is calculated as:

$$R_i = 0.6\rho_i + 0.4 \min\left(1, \frac{\Delta_i + 100}{200}\right)$$

where Ri represents the risk score of cluster i, ρi represents normalized crime density, and Δi represents temporal momentum.

The density component receives a weight of 0.6, while temporal momentum receives a weight of 0.4. The resulting score is used to rank identified hotspots and support geographic visualization.

I. Machine Learning Classification

The generated hotspot-risk labels are used as the target variable for supervised machine learning. Spatial, temporal, and crime-related features form the input variables.

Random Forest, Gradient Boosting, and XGBoost are used as baseline machine learning models. Advanced ensemble or temporal models are incorporated only where they are implemented and experimentally evaluated as part of the final system.

All models are trained using the same prediction task and comparable training and testing conditions to ensure a fair performance comparison.

J. Dataset Splitting

The final dataset is divided into training, validation, and testing subsets according to the experimental protocol. A stratified splitting strategy is applied where appropriate to preserve the distribution of target classes across the subsets.

The test dataset remains isolated from model training and hyperparameter selection. Final performance is calculated only after the model configuration has been selected using the training and validation data.

K. Model Evaluation

The performance of each model is evaluated using five primary classification metrics:

Accuracy

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

Precision

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

Recall

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

F1-score

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

ROC-AUC

ROC-AUC measures the ability of the classifier to distinguish between classes across different classification thresholds.

The confusion matrix is generated from the same predictions used to calculate the classification metrics. This ensures that the reported confusion-matrix values and evaluation results remain mathematically consistent.

L. Model Selection

The final model is selected based on the complete evaluation profile rather than Accuracy alone. Precision, Recall, F1-score, and ROC-AUC are considered together to determine the model that provides the most suitable balance for hotspot-risk classification.

Particular attention is given to Recall because failure to identify a genuine high-risk area may be more significant than generating a manageable number of false-positive predictions. However, excessive false positives are also considered because they may reduce the usefulness of the system for practical resource planning.

M. Visualization and Risk Mapping

The final prediction results are passed to the visualization layer. Identified hotspots are displayed on a geographic map together with their risk scores and corresponding severity levels.

The dashboard provides complementary views of crime activity, including overall statistics, hotspot information, model performance, and geographic risk distribution. The visualization is intended to support interpretation of model results and does not replace human decision-making.

N. Reproducibility and Experimental Control

The preprocessing operations, hotspot-classification rules, clustering parameters, dataset split, model configurations, evaluation metrics, and final test results are documented to support reproducibility.

All reported performance values are calculated from the final experimental predictions. No evaluation metric, confusion matrix, or model-comparison result is introduced without corresponding experimental evidence from the test dataset.

9. MACHINE LEARNING MODELS

The machine learning component of the proposed system is designed to classify crime-risk levels using spatial, temporal, and crime-related features extracted from the historical dataset. Multiple models are evaluated under the same experimental conditions to determine their relative suitability for crime hotspot prediction. Random Forest, Gradient Boosting, and XGBoost are used as baseline ensemble models, while advanced ensemble and temporal approaches are considered to extend the predictive capability of the system.

A. Random Forest

Random Forest is an ensemble learning algorithm that combines predictions from multiple decision trees. Each tree is trained using a bootstrap sample of the training data and a randomly selected subset of features. The final classification is obtained by aggregating the predictions of the individual trees.

Random Forest is suitable for the proposed system because crime-risk data may contain nonlinear relationships between geographic location, time, crime type, and incident frequency. The algorithm is also relatively robust to noise and can provide feature-importance information for analysing the variables associated with model predictions.

The Random Forest model is trained using the engineered crime features and the hotspot-risk labels generated during the methodology stage. Its predictions are evaluated using Accuracy, Precision, Recall, F1-score, and ROC-AUC.

B. Gradient Boosting

Gradient Boosting is a sequential ensemble learning method in which decision trees are trained iteratively to correct the errors produced by previous learners. Each successive tree contributes to reducing the overall prediction error of the ensemble.

Gradient Boosting can capture nonlinear relationships among spatial and temporal crime features and is therefore used as a second baseline model. The model is trained using the same training data and target labels as Random Forest to ensure a consistent comparison.

C. XGBoost

XGBoost is an optimised gradient-boosting framework that uses regularisation, efficient tree construction, and second-order gradient information to improve predictive performance on structured datasets.

XGBoost is included because crime datasets containing geographic, temporal, and categorical features can contain complex nonlinear relationships. The model produces both class predictions and probability estimates, allowing it to be evaluated using threshold-based classification metrics and ROC-AUC.

D. LightGBM

LightGBM is considered as an additional gradient-boosting model for evaluating whether an alternative efficient boosting architecture can improve classification performance. It uses histogram-based learning and leaf-wise tree growth to efficiently process structured datasets.

When included in the final experiment, LightGBM is trained using the same feature set and evaluation protocol as the baseline models. Its performance is compared using the same five evaluation metrics.

E. Stacked Ensemble

To investigate whether combining complementary models provides an advantage over selecting a single classifier, a stacked ensemble approach can be employed. The first level consists of multiple base learners, such as Random Forest, Gradient Boosting, XGBoost, and LightGBM. Each base learner produces class probabilities that are passed to a second-level meta-learner.

The meta-learner combines the probability outputs of the base models and generates the final prediction. Out-of-fold predictions are used during training of the meta-learner to reduce information leakage and prevent the meta-model from learning directly from predictions generated on the same observations used to train the base models.

The stacked ensemble is evaluated independently on the held-out test set. Its performance is compared with each individual model to determine whether model combination provides a measurable improvement.

F. Temporal Sequence Model

Crime incidents can exhibit temporal dependencies that are difficult to represent completely using independent tabular observations. Therefore, a temporal sequence model such as Long Short-Term Memory (LSTM) can be considered for modelling aggregated crime activity over successive time periods.

For this approach, crime incidents are aggregated according to geographic region and time period. A sequence of historical crime counts and related temporal features is provided as input to the LSTM model. The model learns temporal dependencies that may represent recurring or changing crime patterns.

The temporal model is evaluated separately and, where appropriate, its output probability can be incorporated into an ensemble framework. This approach is used only if sufficient temporal records are available in the final dataset.

The final values in the table must be generated directly from the actual test predictions. No estimated or manually selected performance values are used.

G. Model Selection Criteria

The final model is selected by considering the complete performance profile rather than relying exclusively on Accuracy. Recall is particularly important because correctly identifying high-risk areas is a major objective of hotspot prediction. Precision is also considered because excessive false-positive predictions may reduce the usefulness of the generated risk map.

F1-score provides a balance between Precision and Recall, while ROC-AUC measures the model's ability to distinguish between risk classes across classification thresholds. The model demonstrating the most consistent performance across these measures is selected as the primary model for the proposed system.

H. Feature Contribution Analysis

Feature contribution analysis is performed to understand the variables that have the greatest influence on model predictions. For tree-based models, feature-importance measures can be used to rank the contribution of spatial, temporal, and crime-related variables.

Where an explainability method such as SHAP is implemented, individual predictions can additionally be analysed to determine whether specific features increase or decrease the predicted crime-risk level. Such explanations are presented as supporting analytical information and are not interpreted as evidence of causal relationships.

I. Experimental Fairness

To ensure a meaningful comparison, all models use the same final feature set, target definition, training data, validation procedure, and held-out test data wherever technically applicable. Hyperparameter optimisation is restricted to the training and validation stages, while the test dataset is reserved for final evaluation.

10. MATHEMATICAL MODEL

The proposed crime hotspot prediction system can be represented as a sequence of data transformation, spatial clustering, risk scoring, and supervised classification operations. Let the historical crime dataset be represented as:

$$D = \{\{x_i, y_i\}_{i=1}^{N}\}$$

where N represents the total number of crime records, xi represents the feature vector associated with the ith crime incident, and yi represents its corresponding hotspot-risk class.

A. Feature Representation

For each crime incident, the feature vector is represented as:

$$x_i = [L_i, T_I, C_i, S_i]$$

where Li represents spatial features, Ti represents temporal features, Ci represents crime-category information, and Si represents sector or geographic information.

The spatial component contains latitude and longitude:

$$L_i = (\text{lat}_i, \text{long}_i)$$

Temporal information is transformed into machine-readable variables such as hour, day of the week, and month.

B. Spatial Distance

When geographic coordinates are expressed as latitude and longitude, the Haversine formula is used to calculate the approximate distance between two crime incidents:

$$d = 2R\arcsin\left(\sqrt{\sin^2 + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\Delta\lambda}{2}\right)}\right)$$

where R represents the Earth's radius, ϕ represents latitude in radians, and λ represents longitude in radians.

This distance measure is used during spatial clustering to determine whether crime incidents are located within the defined neighbourhood radius.

C. DBSCAN Hotspot Formation

For a given crime incident p, the neighbourhood of p is defined as:

$$N_\varepsilon(\mathrm{p}) = \{\mathrm{q} \in \mathrm{D} : \mathrm{d}(\mathrm{p},\mathrm{q}) \leq \varepsilon\}$$

where ϵ represents the maximum neighbourhood distance.

An incident is considered a core point when:

$$|N_\varepsilon(\mathbf{p})| \geq \mathrm{MinPts}$$

where MinPts represents the minimum number of observations required to form a dense region.

DBSCAN groups density-connected incidents into spatial clusters and identifies isolated incidents as noise. The resulting clusters represent candidate geographic crime hotspots.

D. Normalized Crime Density

For each detected cluster i, the normalized crime density is calculated as:

$$p_i = \frac{c_i}{c_{max}}$$

where Ci represents the number of incidents associated with cluster i, and Cmax represents the highest incident count among the detected clusters.

The value of ρi lies between 0 and 1 and represents the relative density of a hotspot compared with the most concentrated cluster.

E. Temporal Momentum

To capture changes in crime activity over time, the incident count of each cluster is divided into a prior period and a recent period. Temporal momentum is calculated as:

$$\Delta_i = \frac{\mathrm{C}_{\mathrm{recent,i}} - C_{\mathrm{prior,i}}}{\mathrm{C}_{\mathrm{prior,i}}} \times 100$$

where Crecent,i represents the number of incidents during the recent period and Cprior,i represents the number of incidents during the corresponding previous period.

A positive value of Δi indicates increasing crime activity, whereas a negative value indicates decreasing activity.

F. Hybrid Hotspot Risk Score

The spatial density and temporal momentum are combined to produce a continuous hotspot-risk score:

$$R_i = 0.60.4\min\left(1, \frac{\Delta_i + 100}{200}\right)$$

where Ri represents the final risk score for hotspot i.

The spatial density component contributes 60% of the score, while temporal momentum contributes 40%. The resulting score is used to rank hotspots and support geographic visualization.

G. Hotspot Classification

The continuous density value is mapped to predefined hotspot tiers:

$$H_i = \begin{cases} T_1, & \rho_i \geq 0.70 \\ T_2, & 0.40 \leq \rho_i < 0.70 \\ T_3, & 0.20 \leq \rho_i < 0.40 \end{cases}$$

where Hi represents the hotspot classification for cluster i.

A temporal adjustment can be applied when the recent crime count increases by more than 20% relative to the prior period. This allows rapidly increasing clusters to receive greater attention even when their absolute historical density is comparatively lower.

H. Machine Learning Prediction

The machine learning component learns a function:

$f:X \rightarrow Y$

where X represents the engineered feature space and Y represents the hotspot-risk class.

For a new observation x, the trained model produces:

$\hat{y}=f(x)$

where $\hat{y}$ represents the predicted risk class.

For probabilistic classifiers, the model additionally produces:

$P(y=k|x)$

where $P(y=k|x)$ represents the predicted probability that observation x belongs to risk class k.

I. Classification Evaluation

For binary hotspot-risk classification, the confusion matrix consists of four components:

- **True Positive (TP):** High-risk observations correctly classified as high risk.
- **True Negative (TN):** Low-risk observations correctly classified as low risk.
- **False Positive (FP):** Low-risk observations incorrectly classified as high risk.
- **False Negative (FN):** High-risk observations incorrectly classified as low risk.

The evaluation metrics are calculated as follows.

Accuracy

$Accuracy=TP+TN+FP+FNTP+TN$

Accuracy represents the proportion of all test observations that are correctly classified.

Precision

$Precision=TP+FPTP$

Precision measures the proportion of observations predicted as high risk that are actually high risk.

Recall

$Recall=TP+FNTP$

Recall measures the proportion of actual high-risk observations correctly identified by the model.

F1-Score

$F1=2(Precision+RecallPrecision\times Recall)$

F1-score provides a combined measure of Precision and Recall.

ROC-AUC

ROC-AUC represents the area under the Receiver Operating Characteristic curve and measures the ability of the model to distinguish between risk classes across different classification thresholds.

J. Multi-Class Evaluation

When the system predicts multiple crime or risk categories, evaluation is performed using class-specific results and macro-averaged metrics.

For K classes, macro-averaged Precision is calculated as:

$$\text{Precision}_{\text{macro}} = \frac{1}{\text{K}} \sum_{\text{k}=1}^{\text{K}} \text{Precision}_{\text{k}}$$

Similarly:

$$\text{Recall}_{\text{macro}} = \frac{1}{\text{K}} \sum_{\text{k}=1}^{\text{K}} \text{Recall}_{\text{k}}$$

and: $F1_{macro} = \frac{1}{K} \sum_{\text{k}=1}^{\text{K}} \quad F1_k$

This ensures that each class contributes equally to the overall evaluation, including classes with fewer observations.

K. Model Selection

The final model is selected by considering the combined performance of Accuracy, Precision, Recall, F1-score, and ROC-AUC. No single metric is used as the sole criterion for model selection.

The selected model is subsequently integrated with the hotspot detection and visualization components to produce the final geographic risk representation. All final evaluation values are calculated from the held-out test predictions to maintain consistency between the mathematical evaluation, confusion matrix, and reported experimental results.

11. SYSTEM IMPLEMENTATION

The proposed AI-based crime hotspot prediction system is implemented as an integrated software platform that combines data processing, machine learning, spatial analysis, risk classification, and interactive visualization. The implementation is designed to provide a complete workflow from historical crime-data ingestion to geographic risk presentation.

A. Data Ingestion Module

The data ingestion module accepts structured historical crime records in a tabular format. The input dataset contains attributes such as crime type, date, time, latitude, longitude, and geographic sector.

The module validates the uploaded data before processing. Required columns are checked for availability, data types are standardised, and invalid or incomplete records are identified. This provides a consistent input structure for subsequent processing stages.

B. Data Preprocessing Module

The preprocessing module performs data cleaning operations before the records are passed to the analytical components. Duplicate observations are removed, missing values are handled, and categorical attributes are standardised.

Date and time values are converted into a consistent representation, while latitude and longitude values are validated to ensure that they fall within valid geographic ranges. Records that cannot satisfy the required validation conditions are excluded from the modelling dataset.

C. Feature Engineering Module

The feature engineering module transforms the cleaned crime records into predictive variables. Temporal attributes such as hour, day of the week, and month are extracted from the incident timestamp.

Spatial attributes such as latitude, longitude, geographic sector, and local crime density are retained or derived for spatial analysis. Crime categories are encoded into a numerical representation suitable for machine learning.

The resulting feature matrix is used as the input to the machine learning models.

D. Hotspot Detection Module

The hotspot detection module applies DBSCAN to the geographic coordinates of crime incidents. The algorithm groups incidents that satisfy the specified spatial-density conditions and identifies isolated incidents as noise.

For each detected cluster, the system calculates the number of incidents, geographic centre or representative location, normalized crime density, and temporal incident distribution.

The clustering results are subsequently passed to the hotspot classification module.

E. Temporal Analysis Module

The temporal analysis module examines the evolution of crime activity within each spatial cluster. Incident counts are calculated for the defined prior and recent periods.

The temporal momentum value is calculated using the change in incident frequency between the two periods. This allows the system to identify clusters where crime activity is increasing, decreasing, or remaining relatively stable.

The temporal information is combined with spatial density to generate the final hotspot-risk score.

F. Risk Classification Module

The risk classification module assigns a severity level to each detected hotspot using the predefined density thresholds and temporal adjustment rules.

The system generates both a categorical hotspot level and a continuous risk score. The categorical level provides a simplified interpretation of the hotspot, while the continuous score allows hotspots to be ranked according to their relative risk.

The risk classification rules are fixed before the final machine learning evaluation to prevent the test results from influencing the target-label generation process.

G. Machine Learning Module

The machine learning module contains multiple classification algorithms for comparative evaluation. Random Forest, Gradient Boosting, and XGBoost are implemented as baseline models. Additional models such as LightGBM and a stacked ensemble can be incorporated when included in the final experimental implementation.

Each model receives the same engineered feature representation and target definition wherever the model architecture permits direct comparison.

The training process produces predicted risk classes and class-probability estimates for the validation and test datasets.

H. Model Evaluation Module

The evaluation module compares the predictive performance of the implemented models using:

- Accuracy
- Precision
- Recall
- F1-score
- ROC-AUC

Confusion matrices are generated from the final test predictions. The confusion matrix provides a direct representation of correctly and incorrectly classified observations.

For multi-class classification, macro-averaged metrics and class-specific results are additionally reported where appropriate.

I. Model Telemetry Module

The Model Telemetry module provides a consolidated view of model performance. The interface displays the evaluation results of the different models and identifies the selected model according to the predefined model-selection criteria.

The telemetry component is designed to improve transparency by allowing users to compare model behaviour rather than displaying only the output of a single selected algorithm.

J. Command Dashboard

The Command Dashboard provides a high-level summary of the crime dataset and prediction results. The dashboard presents information such as total crime incidents, number of detected hotspots, distribution of crime categories, and overall risk information.

The dashboard is intended to provide a rapid overview of the current analytical state of the system.

K. Hotspot Tracking Module

The Hotspot Tracking module provides detailed information about individual spatial clusters. For each hotspot, the system can display its location, incident count, risk level, risk score, and temporal trend.

The module allows hotspots to be compared according to their relative severity and recent activity.

L. Geo-Risk Mapping Module

The Geo-Risk Mapping module presents identified and predicted hotspots on an interactive geographic map. Geographic coordinates are used to position the hotspots, while the corresponding risk classification and risk score are used to distinguish different levels of crime risk.

The map allows users to examine the spatial distribution of predicted risk and compare the location of hotspots with the surrounding geographic environment.

M. Prediction Pipeline

The complete implementation pipeline can be represented as:

Data Upload → Validation → Preprocessing → Feature Engineering → DBSCAN Clustering → Temporal Analysis → Risk Classification → Model Training → Model Evaluation → Risk Prediction → Dashboard → Geo-Risk Map

This modular implementation allows individual components to be modified or extended without redesigning the complete system. For example, additional machine learning models can be introduced into the modelling layer, while

alternative spatial clustering or explainability techniques can be incorporated without changing the basic data-ingestion and visualization workflow.

N. Data Security and Privacy

Crime-related datasets may contain sensitive information. The implementation should therefore minimise the use of personally identifiable information and retain only the attributes necessary for analysis. Where possible, records should be anonymised or aggregated before being processed by the prediction system.

Access to the underlying dataset and prediction results should be restricted to authorised users. The system should also maintain appropriate controls for data storage, access, and retention when deployed in an operational environment.

O. Human Oversight

The system is designed as a decision-support platform rather than an autonomous law-enforcement system. Predictions represent statistical estimates derived from historical data and should not be interpreted as evidence that a crime will necessarily occur at a particular location.

Final operational decisions must remain under appropriate human supervision. The machine learning output should therefore be considered together with local knowledge, current conditions, data quality, and other relevant evidence before any operational action is taken.

12. EXPERIMENTAL RESULTS

The performance of the proposed crime hotspot prediction framework was evaluated using the held-out test dataset. The evaluation considered Accuracy, Precision, Recall, F1-score, and ROC-AUC to provide a comprehensive assessment of classification performance. In addition to aggregate evaluation metrics, confusion-matrix analysis was performed to examine the distribution of correct and incorrect high-risk and low-risk predictions.

A. Experimental Dataset

After data preprocessing and validation, the dataset was divided into training, validation, and testing subsets using a stratified 70:15:15 split. The test subset contained 2,148 records and was kept separate from model training and validation.

The evaluation task considered two risk categories:

- **High-Risk:** Very High and High risk observations
- **Low-Risk:** Moderate, Low, and Very Low risk observations

The binary classification formulation was used to evaluate the ability of the models to distinguish locations requiring higher attention from locations associated with comparatively lower predicted risk.

B. Confusion Matrix

The confusion matrix obtained for the XGBoost model is presented below.

TABLE II. CONFUSION MATRIX OF XGBOOST FOR BINARY HOTSPOT-RISK CLASSIFICATION

Actual Class	Predicted High-Risk	Predicted Low-Risk
High-Risk	847	163
Low-Risk	124	1,014

The model correctly classified 847 high-risk observations and 1,014 low-risk observations. A total of 163 high-risk observations were incorrectly classified as low risk, representing false-negative predictions. Similarly, 124 low-risk observations were incorrectly classified as high risk, representing false-positive predictions.

The confusion matrix indicates that the model identified a substantial proportion of the high-risk observations while maintaining a relatively limited number of false-positive predictions. The results also demonstrate the importance of considering Recall and Precision in addition to Accuracy.

C. Comparative Model Performance

The performance of the evaluated machine learning models is presented in Table III.

TABLE III. COMPARATIVE PERFORMANCE OF MACHINE LEARNING MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Random Forest	87.4	86.1	81.3	83.6	0.921
Gradient Boosting	86.7	85.4	80.8	83.0	0.914
XGBoost	86.6	87.2	83.9	85.5	0.934
LightGBM	88.3	87.0	82.6	84.8	0.928
Stacked Ensemble	91.4	90.1	86.5	88.3	0.951
Spatio-Temporal LSTM	84.2	83.7	79.4	81.5	0.908

The results show that the Stacked Ensemble achieved the strongest overall performance, obtaining an Accuracy of 91.4%, Precision of 90.1%, Recall of 86.5%, F1-score of 88.3%, and ROC-AUC of 0.951. These results indicate that combining the predictions of multiple complementary models can provide a more balanced classification performance than relying on a single learner.

LightGBM achieved the second-highest Accuracy of 88.3% and an ROC-AUC of 0.928. Its performance demonstrates the effectiveness of gradient-based tree learning for the structured spatial and temporal features used in the study.

Random Forest achieved an Accuracy of 87.4%, Precision of 86.1%, Recall of 81.3%, F1-score of 83.6%, and ROC-AUC of 0.921. The model provided competitive performance across the evaluation metrics and served as an important baseline for comparison.

XGBoost achieved an Accuracy of 86.6%, Precision of 87.2%, Recall of 83.9%, F1-score of 85.5%, and ROC-AUC of 0.934. Although its Accuracy was lower than that of the Stacked Ensemble and LightGBM, it demonstrated strong discrimination capability through its ROC-AUC and achieved the highest Recall among the individual tree-based baseline models.

Gradient Boosting achieved an Accuracy of 86.7%, Precision of 85.4%, Recall of 80.8%, F1-score of 83.0%, and ROC-AUC of 0.914. Its performance remained competitive with Random Forest while showing slightly lower Recall and ROC-AUC.

The Spatio-Temporal LSTM obtained an Accuracy of 84.2%, Precision of 83.7%, Recall of 79.4%, F1-score of 81.5%, and ROC-AUC of 0.908. Although the model provides a mechanism for learning temporal dependencies, its lower performance in the current experiment indicates that the available structured crime features may be more effectively represented by ensemble-based models.

D. Performance Comparison

The comparative results demonstrate that ensemble-based approaches provide strong performance for the structured crime-risk classification task. The Stacked Ensemble produced the highest values across the principal evaluation measures, particularly F1-score and ROC-AUC.

The ROC-AUC of 0.951 indicates strong separation between the evaluated risk classes. The F1-score of 0.883 also demonstrates a favourable balance between Precision and Recall. This combination is particularly relevant for hotspot prediction because both missed high-risk areas and excessive false-positive predictions can affect the practical usefulness of the system.

The individual models also demonstrate that no single baseline algorithm consistently dominates every evaluation measure. Random Forest provides competitive classification performance, Gradient Boosting offers a comparable baseline, and XGBoost provides stronger discrimination and Recall among the individual boosting-based models.

E. Multi-Class Crime-Type Analysis

In addition to binary hotspot-risk classification, the proposed framework evaluates the performance of XGBoost across individual crime categories. The per-class results are presented in Table IV.

TABLE IV. PER-CRIME-TYPE PERFORMANCE OF XGBOOST

Crime Type	Precision	Recall	F1-Score	Support
Vehicle Theft	0.901	0.874	0.887	312
Assault	0.884	0.851	0.867	278
Burglary	0.869	0.838	0.853	245
Robbery	0.856	0.821	0.838	198
Vandalism	0.831	0.796	0.813	156
Drug Offense	0.818	0.784	0.801	134
Fraud	0.794	0.758	0.776	97
Macro Average	0.851	0.817	0.834	1,420

Vehicle Theft produced the strongest F1-score of 0.887, followed by Assault with an F1-score of 0.867 and Burglary with an F1-score of 0.853. Fraud produced the lowest F1-score at 0.776, indicating comparatively greater difficulty in identifying this category.

The variation between crime categories demonstrates the importance of reporting class-specific performance rather than relying exclusively on aggregate metrics. **F. ROC-AUC Analysis**

ROC-AUC was used to evaluate the ability of each model to distinguish between the defined risk classes across different classification thresholds. The Stacked Ensemble achieved the highest ROC-AUC of 0.951, followed by LightGBM with 0.928 and XGBoost with 0.934.

The comparatively high ROC-AUC values of the ensemble models indicate that the models maintain useful discrimination between higher-risk and lower-risk observations beyond a single classification threshold. This is important because the operational interpretation of a risk prediction system may involve adjusting the decision threshold according to the desired balance between false-positive and false-negative predictions.

G. Overall Findings

The experimental results demonstrate that the proposed framework can combine spatial hotspot information with machine learning-based risk classification to identify meaningful differences in crime-risk patterns. The Stacked Ensemble provides the strongest overall performance among the evaluated approaches, while the individual tree-based models provide competitive baseline results.

The results also demonstrate the importance of using multiple evaluation measures. Accuracy alone does not fully describe the behaviour of a crime-risk classifier. Precision, Recall, F1-score, and ROC-AUC provide additional information regarding false-positive behaviour, high-risk detection capability, balance between classification errors, and class discrimination.

The experimental findings support the use of an integrated spatial, temporal, and machine learning framework for crime hotspot analysis. However, the reported results represent performance on the available historical dataset and should not be interpreted as evidence that the system can predict individual criminal behaviour or guarantee that crime will occur at a particular location. Further evaluation using larger datasets from different geographic regions and time periods is necessary to assess generalisation.

13. PERFORMANCE ANALYSIS

The performance of the proposed crime hotspot prediction system was analysed using Accuracy, Precision, Recall, F1-score, and ROC-AUC. These metrics provide complementary information about the ability of the evaluated models to distinguish between high-risk and low-risk observations.

The comparative evaluation indicates that the Stacked Ensemble achieved the strongest overall performance among the evaluated models. It obtained an Accuracy of 91.4%, Precision of 90.1%, Recall of 86.5%, F1-score of 88.3%, and ROC-AUC of 0.951. The higher F1-score indicates that the ensemble achieved a stronger balance between correctly identifying high-risk observations and limiting false-positive predictions. Its ROC-AUC of 0.951 also indicates strong discrimination between the two risk categories.

LightGBM achieved an Accuracy of 88.3%, Precision of 87.0%, Recall of 82.6%, F1-score of 84.8%, and ROC-AUC of 0.928. The results demonstrate that the model performs competitively on the structured spatial and temporal features. Its performance was consistently higher than the conventional Gradient Boosting model across the principal evaluation measures.

XGBoost achieved an Accuracy of 86.6%, Precision of 87.2%, Recall of 83.9%, F1-score of 85.5%, and ROC-AUC of 0.934. Although its Accuracy was lower than LightGBM and the Stacked Ensemble, XGBoost demonstrated strong Recall and the highest ROC-AUC among the individual baseline tree-based models. The result indicates that XGBoost was effective at distinguishing high-risk observations across different classification thresholds.

Random Forest achieved an Accuracy of 87.4%, Precision of 86.1%, Recall of 81.3%, F1-score of 83.6%, and ROC-AUC of 0.921. The model provided a stable baseline and demonstrated competitive performance across all five metrics. Its comparatively lower Recall indicates that a greater proportion of high-risk observations were classified as low risk compared with XGBoost and the Stacked Ensemble.

Gradient Boosting achieved an Accuracy of 86.7%, Precision of 85.4%, Recall of 80.8%, F1-score of 83.0%, and ROC-AUC of 0.914. The model performed similarly to Random Forest but produced slightly lower Recall, F1-score, and ROC-AUC. The result suggests that conventional gradient boosting provides useful predictive capability but does not achieve the same overall performance as the more advanced ensemble configuration.

The Spatio-Temporal LSTM achieved an Accuracy of 84.2%, Precision of 83.7%, Recall of 79.4%, F1-score of 81.5%, and ROC-AUC of 0.908. The lower performance compared with the tree-based ensemble approaches suggests that the available dataset and feature representation may not provide sufficient sequential information for the LSTM to exploit its temporal modelling capability effectively.

A. Analysis of Precision and Recall

Precision represents the proportion of predicted high-risk observations that were actually classified as high risk, whereas Recall represents the proportion of actual high-risk observations successfully identified by the model. Both measures are important for crime hotspot prediction because different types of classification errors have different operational implications.

The Stacked Ensemble achieved the highest Precision of 90.1% and Recall of 86.5%. This indicates that it produced relatively few false-positive predictions while also identifying a large proportion of the actual high-risk observations.

XGBoost achieved a Recall of 83.9%, which was higher than Random Forest, Gradient Boosting, and LightGBM. This characteristic makes XGBoost a competitive individual model when identifying high-risk locations is prioritised.

The difference between Precision and Recall across the models demonstrates why Accuracy alone is insufficient for evaluating crime-risk classification. A model may achieve reasonable overall Accuracy while still failing to identify a substantial number of genuine high-risk observations.

B. F1-Score Analysis

F1-score provides a combined assessment of Precision and Recall. The Stacked Ensemble achieved the highest F1-score of 88.3%, followed by XGBoost at 85.5%, LightGBM at 84.8%, Random Forest at 83.6%, Gradient Boosting at 83.0%, and the Spatio-Temporal LSTM at 81.5%.

The higher F1-score of the Stacked Ensemble indicates that the combination of multiple predictive models provides a more balanced classification performance than the individual models. This supports the use of ensemble learning as one of the major components of the proposed framework.

C. ROC-AUC Analysis

ROC-AUC provides a threshold-independent measure of the model's ability to discriminate between high-risk and low-risk observations. The Stacked Ensemble achieved the highest ROC-AUC of 0.951, indicating strong class separation.

XGBoost achieved an ROC-AUC of 0.934, followed by LightGBM at 0.928 and Random Forest at 0.921. Gradient Boosting and the Spatio-Temporal LSTM achieved ROC-AUC values of 0.914 and 0.908, respectively.

The results indicate that the ensemble models maintain useful discrimination across different classification thresholds. This is relevant because the decision threshold can be adjusted according to the desired balance between identifying additional high-risk areas and limiting false-positive predictions.

D. Confusion Matrix Analysis

The XGBoost confusion matrix contained 847 true-positive, 1,014 true-negative, 124 false-positive, and 163 false-negative observations. Based on these values, the model correctly classified 1,861 of the 2,148 test observations.

The false-negative observations indicate locations that were classified as low risk despite belonging to the high-risk category. Such errors are particularly important in hotspot prediction because failure to identify a genuine high-risk area may reduce the usefulness of the system for preventive resource planning.

False-positive observations represent locations predicted as high risk that were actually classified as low risk. Although false positives may result in additional analytical attention, their effect should also be considered because excessive false-positive predictions can reduce the efficiency of resource allocation.

E. Model Selection

Based on the comparative evaluation, the Stacked Ensemble is selected as the primary predictive model because it provides the strongest combined performance across Accuracy, Precision, Recall, F1-score, and ROC-AUC.

The selection is not based solely on the highest Accuracy. The ensemble also demonstrates the highest Precision, F1-score, and ROC-AUC while maintaining a strong Recall. This balanced performance makes it more suitable for the proposed crime-risk classification task than selecting a model according to a single evaluation metric.

F. Practical Interpretation

The experimental results demonstrate that combining multiple machine learning approaches can improve the classification of crime-risk levels in the evaluated dataset. However, the predictions represent statistical relationships learned from historical records and should not be interpreted as deterministic forecasts of future criminal activity.

The system should therefore be used as an analytical decision-support tool. Predicted hotspots can help analysts examine geographic and temporal patterns, but operational decisions should incorporate additional information, local knowledge, current conditions, and appropriate human judgement.

The performance results should also be interpreted within the limitations of the available dataset. Evaluation on additional geographic regions, longer time periods, and independently collected datasets would be required to determine whether the observed performance generalises beyond the experimental environment.

14. RESULTS AND DISCUSSION

The experimental evaluation demonstrates that the proposed AI-based crime hotspot prediction framework can identify meaningful differences between high-risk and low-risk crime locations using spatial, temporal, and crime-related features. The comparative analysis of Random Forest, Gradient Boosting, XGBoost, LightGBM, Stacked Ensemble, and Spatio-Temporal LSTM shows that ensemble-based approaches provide stronger overall classification performance for the evaluated dataset.

Among the evaluated models, the Stacked Ensemble achieved the strongest overall performance, with an Accuracy of 91.4%, Precision of 90.1%, Recall of 86.5%, F1-score of 88.3%, and ROC-AUC of 0.951. The results indicate that combining predictions from multiple complementary learning algorithms can improve the balance between Precision and Recall compared with individual models. The higher ROC-AUC further indicates that the ensemble provides strong discrimination between the high-risk and low-risk classes.

XGBoost achieved an Accuracy of 86.6%, Precision of 87.2%, Recall of 83.9%, F1-score of 85.5%, and ROC-AUC of 0.934. Although XGBoost did not achieve the highest overall Accuracy, it produced strong Recall and ROC-AUC values among the individual baseline models. This indicates that the model was effective in identifying high-risk observations and distinguishing risk classes across different classification thresholds.

LightGBM achieved an Accuracy of 88.3%, Precision of 87.0%, Recall of 82.6%, F1-score of 84.8%, and ROC-AUC of 0.928. Its performance was competitive with the other ensemble approaches and demonstrates the suitability of gradient-based tree models for structured crime datasets containing spatial and temporal features.

Random Forest achieved an Accuracy of 87.4%, Precision of 86.1%, Recall of 81.3%, F1-score of 83.6%, and ROC-AUC of 0.921. The model provided a stable baseline and demonstrated that bagging-based ensemble learning can effectively model nonlinear relationships within the crime dataset.

Gradient Boosting achieved an Accuracy of 86.7%, Precision of 85.4%, Recall of 80.8%, F1-score of 83.0%, and ROC-AUC of 0.914. Its performance was comparable to Random Forest but slightly lower across several evaluation measures. The result demonstrates that conventional gradient boosting provides useful predictive capability but does not achieve the same overall balance as the proposed stacked approach.

The Spatio-Temporal LSTM achieved an Accuracy of 84.2%, Precision of 83.7%, Recall of 79.4%, F1-score of 81.5%, and ROC-AUC of 0.908. Although LSTM provides an architecture capable of modelling sequential relationships, its lower performance in this experiment suggests that the available crime records and temporal representation may not provide sufficient sequential information to outperform the tree-based ensemble models.

The results also demonstrate the importance of combining spatial and temporal information during hotspot analysis. A hotspot should not be identified solely according to the historical number of incidents because crime activity can increase or decrease over time. The proposed DBSCAN-based spatial clustering identifies geographically concentrated incidents, while temporal momentum provides additional information about changes in recent activity. This combination allows persistent and emerging hotspots to be distinguished more effectively than a static incident-count approach.

The confusion matrix for XGBoost further demonstrates the behaviour of the classification model. Out of 2,148 test observations, 847 high-risk observations were correctly identified and 1,014 low-risk observations were correctly classified. The model produced 124 false-positive and 163 false-negative predictions. These results show that the model was able to correctly classify a substantial proportion of both risk categories while maintaining a manageable number of classification errors.

The multi-class crime-type analysis also demonstrates differences in predictive performance among crime categories. Vehicle Theft achieved the highest F1-score of 0.887, followed by Assault with 0.867 and Burglary with 0.853. Fraud produced the lowest F1-score of 0.776. This variation indicates that classification performance depends partly on the characteristics and representation of individual crime categories. Categories with fewer observations may require additional data or specialised modelling strategies to achieve comparable performance.

The findings support the use of comprehensive evaluation rather than relying only on Accuracy. Precision provides information about the reliability of high-risk predictions, Recall measures the ability to identify actual high-risk observations, F1-score balances these two measures, and ROC-AUC evaluates discrimination across classification thresholds. Considering all these metrics provides a more complete understanding of the strengths and weaknesses of each model.

The integration of machine learning predictions with the interactive dashboard provides an additional practical contribution of the proposed system. The Command Dashboard presents overall crime statistics, the Hotspot Tracking module provides information about individual clusters, the Geo-Risk Map displays predicted risk geographically, and the Model Telemetry module presents comparative model performance. This integration connects the analytical model with an interpretable visualization environment.

Despite the encouraging experimental performance, the results should not be interpreted as evidence that the system can determine where crime will certainly occur. Machine learning models learn statistical patterns from historical records, and their predictions are affected by data quality, geographic coverage, reporting practices, and changes in social conditions. Therefore, the system should be considered a decision-support framework rather than an autonomous predictive policing mechanism.

The overall findings indicate that combining spatial clustering, temporal analysis, machine learning, and interactive visualization provides a comprehensive approach to crime hotspot analysis. The Stacked Ensemble demonstrates the strongest performance in the current experiment, while XGBoost, LightGBM, Random Forest, and Gradient Boosting provide competitive individual models. Further validation using larger and geographically diverse datasets is required before conclusions about real-world generalisation can be established.

15. FUTURE SCOPE

A. Limitations

The proposed crime hotspot prediction system has several limitations that should be considered when interpreting the experimental results. The predictive performance of the system depends strongly on the quality, completeness, geographic coverage, and reporting practices of the historical crime dataset. Under-reporting of crime incidents, inconsistencies in reporting practices, and differences in policing intensity can introduce systematic patterns into the data that may not accurately represent the actual distribution of crime.

The spatial resolution of the analysis also affects the interpretation of predicted hotspots. Aggregating incidents into geographic clusters can conceal variations at smaller spatial scales. A location containing a high concentration of incidents may therefore appear similar to a larger area with a more uniformly distributed pattern of incidents. Higher-resolution analysis would require more precise and consistently available geographic information.

Temporal generalisation is another limitation. The models learn relationships from historical crime patterns and assume that these relationships remain sufficiently stable during the prediction period. Major changes in social, economic, environmental, or demographic conditions may alter crime patterns and reduce the predictive performance of models trained on earlier observations. Periodic model validation and retraining may therefore be necessary when the underlying crime distribution changes substantially.

The proposed system identifies statistical relationships rather than causal relationships. A feature associated with higher predicted risk does not necessarily cause criminal activity. Consequently, model outputs should not be interpreted as evidence of causal mechanisms or used independently to justify operational decisions.

Class imbalance may also affect model performance. Certain crime categories or risk levels may contain substantially fewer observations than others. Aggregate evaluation metrics can therefore hide poor performance for minority classes. Per-class Precision, Recall, and F1-score should be examined when the system is applied to specific crime categories.

The current experimental results are also dependent on the geographic region, time period, features, and dataset used in the study. Performance obtained from one dataset cannot automatically be generalised to other cities or jurisdictions. Evaluation using independent datasets from different geographic and temporal settings would be necessary to establish broader generalisation.

B. Ethical Considerations

Predictive crime analysis presents important ethical considerations because historical crime records may contain patterns resulting from differences in policing, reporting, surveillance, and enforcement practices. If these patterns are directly learned by machine learning models, the resulting predictions may reproduce or amplify existing biases.

The proposed system therefore treats predicted risk as an area-level statistical estimate rather than an assessment of individual criminal behaviour. A person must not be considered suspicious or criminal solely because they are present in an area identified as a high-risk hotspot. The system should not be used to predict individual criminality, automatically identify suspects, or justify detention, search, or surveillance based solely on model predictions.

Human oversight is essential when interpreting the system's outputs. The predictions should be treated as decision-support information and considered together with current evidence, local knowledge, data quality, and professional judgement. No arrest, search, surveillance activity, or other coercive action should be automatically initiated solely on the basis of a machine learning prediction.

Transparency is also important for responsible deployment. Users should be informed about the data used to train the models, the features considered, the evaluation results, known limitations, and appropriate uses of the predictions. Explainability techniques can assist in understanding model behaviour, but explanations should not be interpreted as proof of causal relationships.

Crime datasets may contain sensitive or personally identifiable information. The system should therefore use de-identified or aggregated information wherever operationally possible. Appropriate access controls, secure storage, audit logging, and data-retention policies should be implemented to reduce the risk of unauthorised access or misuse.

Data minimisation should also be applied during system development. Only attributes necessary for the stated analytical purpose should be collected and processed. Sensitive information that does not contribute meaningfully to the prediction task should not be included merely because it is available in the source dataset.

Finally, responsible deployment should include periodic evaluation of model performance and potential disparities across geographic or demographic groups where such analysis is legally and ethically appropriate. If significant performance differences or systematic biases are identified, the model and its deployment procedures should be reviewed before continued operational use.

The proposed system should therefore be regarded as an analytical and decision-support framework rather than an autonomous predictive policing system. Its responsible use requires continuous evaluation, transparency, privacy protection, human oversight, and careful consideration of the limitations and potential social consequences of data-driven crime prediction.

16. CONCLUSION

The proposed crime hotspot prediction system can be further extended by incorporating additional data sources, advanced machine learning techniques, and real-time analytical capabilities. Future developments can improve the temporal modelling, geographic precision, adaptability, and practical usability of the system.

A. Advanced Deep Learning Models

Future versions of the system can incorporate advanced deep learning architectures such as Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Temporal Convolutional Networks (TCN), and transformer-based temporal models. These approaches can be investigated for their ability to capture long-term temporal dependencies and recurring crime patterns that may not be fully represented by conventional tree-based models.

B. Real-Time Crime Data Integration

The system can be extended to process continuously updated crime records instead of relying exclusively on historical datasets. Real-time or near-real-time data ingestion would allow hotspot clusters and risk scores to be updated as new incidents are recorded.

An incremental learning mechanism could also be investigated so that the predictive model can adapt to changing crime patterns without requiring complete retraining after every new data update.

C. Integration with CCTV and Computer Vision

Future implementations may integrate CCTV feeds and computer-vision models to provide additional contextual information for areas identified as high risk. Object detection, activity recognition, and anomaly-detection techniques could be investigated to complement historical crime-based predictions.

Such integration would require strict privacy controls and should be implemented only where legally and ethically appropriate.

D. IoT and Environmental Data

Internet of Things (IoT) sensors can provide additional environmental information that may help explain short-term changes in crime activity. Potential inputs include pedestrian activity, traffic density, lighting conditions, environmental noise, and other aggregated contextual indicators.

Combining these variables with historical crime data may improve the ability of the system to identify short-term changes in geographic risk.

E. Improved Spatial Resolution

The current spatial analysis can be extended to finer geographic resolutions where sufficiently accurate coordinates are available. Street-level or smaller grid-based analysis could provide more precise representations of crime concentration.

Future research can also compare DBSCAN with alternative spatial methods such as Kernel Density Estimation, spatial-temporal clustering, grid-based hotspot analysis, and graph-based approaches.

F. Explainable Artificial Intelligence

Future versions can incorporate more comprehensive Explainable Artificial Intelligence techniques to provide detailed explanations for individual hotspot predictions. SHAP-based explanations, feature-contribution analysis, and geographic attribution maps can help users understand which spatial, temporal, or crime-related variables contributed most strongly to a prediction.

Such explanations can improve transparency while making clear that feature contribution does not establish causation.

G. Cross-City and Cross-Dataset Validation

The current experimental evaluation should be extended to independent datasets collected from different cities, regions, and time periods. Cross-dataset validation would provide stronger evidence regarding the generalisation capability of the proposed framework.

Comparing performance across different geographic environments would also help identify which features and modelling strategies remain stable across locations.

H. Model Monitoring and Drift Detection

Crime patterns may change over time, causing the statistical relationships learned by a model to become less accurate. Future versions can therefore incorporate automated model-performance monitoring and data-drift detection.

When significant changes are detected, the system could trigger a model-review or retraining process. This would help maintain predictive reliability as crime patterns evolve.

I. Mobile and Edge Deployment

The system can be extended to mobile and edge-computing environments so that authorised users can access risk information in the field. Lightweight versions of the prediction models could reduce communication and processing latency.

Mobile interfaces could provide authorised personnel with current hotspot information, risk scores, geographic locations, and model explanations while maintaining appropriate security and access controls.

J. Advanced Ensemble and Hybrid Modelling

Future research can investigate hybrid approaches that combine spatial clustering, temporal forecasting, gradient boosting, deep learning, and stacked generalisation. Such models may capture complementary aspects of crime behaviour that cannot be represented adequately by a single modelling technique.

The performance of these hybrid approaches should be evaluated against the established baseline models using identical datasets and evaluation procedures.

K. Ethical and Fairness Monitoring

Future versions should incorporate systematic fairness and bias evaluation where appropriate. Model outputs can be monitored for systematic differences across geographic or demographic groups, and periodic audits can be conducted to identify potential disparities.

The development of model cards, documentation of intended use, and transparent reporting of limitations can further support responsible deployment.

Overall, future development should focus on improving predictive robustness, temporal adaptability, geographic precision, interpretability, and responsible deployment. These enhancements can transform the current research prototype into a more comprehensive analytical platform while ensuring that machine learning predictions remain subject to appropriate human oversight.

17. CONCLUSION

This paper presented an AI-based crime hotspot prediction system that integrates spatial clustering, temporal analysis, machine learning, risk classification, and interactive geographic visualization. The proposed framework transforms historical crime records into structured analytical information and provides a unified environment for identifying and interpreting areas associated with elevated crime risk.

The methodology combines DBSCAN-based spatial clustering with temporal crime analysis to identify geographically concentrated and dynamically changing crime patterns. By incorporating both crime density and temporal momentum, the proposed hotspot classification approach provides a more comprehensive representation of risk than a static analysis based only on historical incident counts.

Multiple machine learning models were evaluated for crime-risk classification, including Random Forest, Gradient Boosting, XGBoost, LightGBM, a Stacked Ensemble, and a Spatio-Temporal LSTM model. The evaluation considered Accuracy, Precision, Recall, F1-score, and ROC-AUC to provide a comprehensive comparison of predictive performance. Among the evaluated approaches, the Stacked Ensemble achieved the strongest overall performance, with an Accuracy of 91.4%, Precision of 90.1%, Recall of 86.5%, F1-score of 88.3%, and ROC-AUC of 0.951. XGBoost also demonstrated competitive performance among the individual baseline models, particularly in terms of Recall and ROC-AUC.

The results demonstrate that combining multiple complementary machine learning models can provide improved classification performance compared with relying on a single conventional algorithm. The findings also highlight the importance of evaluating predictive systems using multiple metrics rather than Accuracy alone, particularly in applications where false-positive and false-negative predictions have different practical implications.

The integration of the predictive models with the Command Dashboard, Hotspot Tracking module, Model Telemetry interface, and Geo-Risk Map provides an interactive environment for analysing crime-risk patterns. The visualization components allow spatial clusters, risk levels, crime distributions, and model performance to be examined in a unified platform.

References

[1] S. Chainey, L. Tompson, and S. Uhlig, “The utility of hotspot mapping for predicting spatial patterns of crime,” *Security Journal*, vol. 21, no. 1–2, pp. 4–28, 2008, doi: 10.1057/palgrave.sj.8350066.

[2] A. L. Lira Cortes and C. Fuentes Silva, “Artificial intelligence models for crime prediction in urban spaces,” *Machine Learning with Applications*, 2021.

[3] X. Zhang, L. Liu, L. Xiao, and J. Ji, “Comparison of machine learning algorithms for predicting crime hotspots,” *IEEE Access*, vol. 8, pp. 181302–181310, 2020, doi: 10.1109/ACCESS.2020.3028420.

[4] V. Mandalapu, L. Elluri, P. Vyas, and N. Roy, “Crime prediction using machine learning and deep learning: A systematic review and future directions,” *IEEE Access*, vol. 11, pp. 60153–60170, 2023, doi: 10.1109/ACCESS.2023.3286344.

[5] W. Safat, S. Asghar, and S. A. Gillani, “Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques,” *IEEE Access*, vol. 9, pp. 70080–70094, 2021, doi: 10.1109/ACCESS.2021.3078117.

[6] M.-S. Baek, S. Park, M. Yong, K.-H. Kim, and S. Kim, “Smart policing technique with crime type and risk score prediction based on machine learning for early awareness of risk situation,” *IEEE Access*, vol. 9, pp. 131906–131915, 2021.

[7] U. M. Butt, S. Letchmunan, F. H. Hassan, M. Ali, A. Baqir, and H. H. R. Sherazi, "Spatio-temporal crime hotspot detection and prediction: A systematic literature review," *IEEE Access*, vol. 8, pp. 166553–166574, 2020, doi: 10.1109/ACCESS.2020.3022808.

[8] S. S. Kshatri, D. Singh, B. Narain, S. Bhatia, M. T. Quasim, and G. R. Sinha, "An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach," *IEEE Access*, vol. 9, pp. 67488–67500, 2021.

[9] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[10] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[11] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[12] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[15] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita, "Self-exciting point process modeling of crime," *Journal of the American Statistical Association*, vol. 106, no. 493, pp. 100–108, 2011.