International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2) 2026 pp. 554-573
https://doi.org/10.70917/ijcisim-2026-2913
Article

# A Generalized Spatio-Temporal Stacked Ensemble model for Urban Crime Type Classification Across Multi-City Crime Datasets

**Shraddha Jagdish Pawar[1], Baisa L. Gunjal[2]**

[1] MET's Bhujbal Knowledge City Institute of Engineering, Savitribai Phule Pune University (SPPU), Nashik, Maharashtra, India.
Email: shraddhap113@gmail.com
ORCID: 0009-0008-4694-5487
[2]Department of Information Technology, Amrutvahini College of Engineering, Sangamner, Maharashtra, India.
Email: hod.it@avcoe.org
ORCID: 0000-0003-0091-1886

**Abstract:** Urban crime datasets contain large volumes of spatial, temporal, and categorical information that can be used to support crime analysis and decision-making. Accurate crime type classification remains challenging due to class imbalance, nonlinear relationships among features, and variations in crime patterns across different cities. Many existing studies rely on individual machine learning models and evaluate their performance on a single dataset, which limits the assessment of model generalization. This study presents a spatio-temporal stacked ensemble model for crime type classification using two large-scale urban crime datasets: the Chicago Crime Dataset and the Los Angeles Crime Dataset. The proposed model combines Random Forest, Logistic Regression, and Support Vector Machine as base classifiers, while Gradient Boosting is used as a meta-classifier to generate the final prediction. A preprocessing pipeline consisting of data cleaning, feature encoding, feature selection, feature scaling, and stratified sampling is applied to improve data quality and maintain consistency across datasets. The proposed model is evaluated using accuracy, precision, recall, and F1-score. On the Chicago Crime Dataset, the model achieved an accuracy of 99.9%, precision of 1.00, recall of 0.99, and F1-score of 0.99. On the Los Angeles Crime Dataset, it achieved an accuracy of 99.2%, precision of 0.99, recall of 0.99, and F1-score of 0.99. The results were compared with Logistic Regression, Random Forest, Support Vector Machine, and Gradient Boosting classifiers. The stacked ensemble produced the highest classification performance on both datasets and maintained consistent results across different crime distributions. The experimental findings indicate that combining heterogeneous classifiers through stacking can improve prediction accuracy and reduce the limitations of individual models. The proposed approach can support crime analysis systems that require reliable classification of crime incidents across different urban environments.

**Keywords:** Crime Type Classification, Ensemble Learning, Stacking, Machine Learning, Crime Analytics, Urban Crime Data, Spatio-Temporal Features, Random Forest, Support Vector Machine, Gradient Boosting.

## 1. Introduction

The growth of urban populations and the increasing complexity of modern cities have led to a substantial rise in the volume of crime-related data collected by law enforcement agencies. Crime records contain valuable information regarding the location, time, nature, and circumstances of criminal incidents. The analysis of such data can assist public safety organizations in understanding crime patterns, identifying trends, and supporting operational decision-making. As many cities continue to publish large-scale crime datasets, there is growing interest in developing intelligent systems capable of extracting meaningful information from these records[1-3].

Crime type classification is one of the important tasks in crime analytics. The objective is to assign a reported incident to its corresponding crime category using information available in historical records. Accurate classification can support crime analysis, resource planning, and the development of data-driven policing strategies. However, crime

datasets are often characterized by high dimensionality, class imbalance, noisy records, and complex relationships among spatial and temporal attributes. These characteristics make crime type classification a challenging machine learning problem[4].

Machine learning techniques have been widely investigated for crime analysis due to their ability to identify hidden patterns within large datasets[5]. Algorithms such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machine, Gradient Boosting, and Artificial Neural Networks have been applied to crime prediction and classification tasks. These methods have produced promising results in several studies; however, their performance is often influenced by the characteristics of the dataset being used. A model that performs well on one crime dataset may not achieve similar performance when applied to data collected from a different city or region. This limitation raises concerns regarding model generalization and practical applicability[6-7].

Many existing studies focus on a single machine learning model and evaluate performance using a single crime dataset. Although these approaches provide useful insights, they may fail to capture the diverse patterns present in real-world crime data. Individual classifiers possess different strengths and weaknesses. Logistic Regression performs effectively when relationships between variables are relatively simple, whereas Random Forest can model complex nonlinear interactions. Support Vector Machine is known for its ability to separate classes in high-dimensional feature spaces, while Gradient Boosting improves prediction performance through iterative error correction. Relying on a single classifier may therefore restrict classification performance when dealing with heterogeneous crime datasets[8].

Ensemble learning[9] has emerged as an effective approach for improving predictive performance by combining multiple classifiers. Instead of depending on a single learning strategy, ensemble methods integrate the outputs of several models to produce a final decision. Among various ensemble techniques, stacking has received considerable attention because it allows a meta-classifier to learn how predictions generated by base models should be combined. This mechanism enables the system to utilize complementary information obtained from different classifiers and can improve classification accuracy and prediction stability.

Despite the progress reported in previous studies, several research gaps remain. First, many crime classification studies continue to rely on standalone machine learning algorithms without systematically combining heterogeneous learning approaches. Second, most investigations evaluate performance using a single dataset, making it difficult to determine whether the proposed model can generalize across different urban environments. Third, limited attention has been given to validating stacked ensemble models using multiple large-scale crime datasets collected from different cities. Addressing these limitations is necessary for developing crime classification systems that are reliable in practical applications[10].

To address these issues, this study presents a generalized spatio-temporal stacked ensemble model for urban crime type classification. The proposed architecture combines Random Forest, Logistic Regression, and Support Vector Machine as base classifiers, while Gradient Boosting is used as a meta-classifier. The model integrates information generated by different learning algorithms and produces the final classification through stacked learning. Spatial and temporal crime attributes are incorporated during model training to capture important patterns associated with crime occurrence.

The proposed approach is evaluated using two large-scale real-world crime datasets: the Chicago Crime Dataset and the Los Angeles Crime Dataset. Using datasets collected from different cities allows the assessment of model performance under varying crime distributions and data characteristics. The evaluation includes comparisons with individual machine learning classifiers using accuracy, precision, recall, and F1-score metrics.

The main contributions of this study are summarized as follows:

1. A stacked ensemble architecture combining Random Forest, Logistic Regression, Support Vector Machine, and Gradient Boosting is developed for crime type classification.
2. A preprocessing pipeline consisting of data cleaning, feature encoding, feature selection, feature scaling, and stratified sampling is designed to improve data quality and model consistency.
3. The proposed model is validated using two large-scale urban crime datasets collected from Chicago and Los Angeles.
4. A comparative analysis is conducted against individual machine learning classifiers to assess the effectiveness of the stacked ensemble approach.

5. The proposed model achieves classification accuracies of 99.9% and 99.2% on the Chicago and Los Angeles datasets, respectively, demonstrating consistent performance across different urban crime datasets.

The remainder of this paper is organized as follows. Section 2 reviews existing studies related to crime classification and ensemble learning. Section 3 describes the proposed methodology and model architecture. Section 4 presents the experimental setup and evaluation metrics. Section 5 discusses the results and comparative analysis. Section 6 concludes the paper and outlines future research directions.

## 2. Related Work

Machine learning has become an important tool in crime analytics due to its ability to identify patterns from large and complex datasets. Researchers have investigated a wide range of approaches for crime prediction, hotspot detection, crime forecasting, and crime type classification. These approaches can be broadly categorized into traditional machine learning methods, deep learning techniques, and ensemble learning models.

### *2.1 Traditional Machine Learning Approaches*

Early studies in crime analytics primarily focused on conventional machine learning algorithms such as Logistic Regression, Decision Trees, Naïve Bayes, k-Nearest Neighbors, Random Forest, and Support Vector Machine. These models were applied to classify crime incidents, identify crime hotspots, and estimate future crime occurrences. Among these approaches, Random Forest and Support Vector Machine frequently reported higher prediction accuracy because of their ability to model complex relationships within crime datasets.

Several studies demonstrated that crime patterns are strongly influenced by spatial and temporal attributes[11]. As a result, machine learning models trained using location, time, and demographic information produced better classification performance than models relying on a limited set of features. Despite these improvements, the effectiveness of individual classifiers often varied across datasets and crime categories. Models that achieved high accuracy on one dataset frequently experienced performance degradation when applied to data collected from different geographical regions.

### *2.2 Deep Learning Approaches*

With the availability of larger crime datasets, deep learning methods have gained attention for crime-related applications. Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Deep Neural Networks (DNN) have been used to capture spatial and temporal dependencies in crime data. These approaches demonstrated promising results for crime forecasting and hotspot prediction tasks[12].

Although deep learning models can learn complex feature representations, they require substantial computational resources and large amounts of training data. In addition, the decision-making process of deep learning models is often difficult to interpret. This limitation becomes important in crime analysis applications where transparency and explainability are desirable for operational decision-making.

### *2.3 Ensemble Learning Approaches*

To overcome the limitations of individual classifiers, researchers have increasingly explored ensemble learning methods. Techniques such as bagging, boosting, voting, and stacking combine multiple classifiers to improve prediction performance. Random Forest represents a widely adopted bagging approach, while Gradient Boosting and its variants belong to the boosting family of algorithms[13].

Among ensemble techniques, stacking has attracted interest because it combines predictions generated by multiple base learners through a meta-classifier. This strategy allows different learning algorithms to contribute complementary information during the prediction process. Previous studies have reported that stacking often produces higher accuracy than standalone classifiers, particularly when datasets contain nonlinear relationships and heterogeneous feature distributions.

Despite these developments, most existing crime classification studies evaluate ensemble models using a single dataset. Furthermore, many investigations focus primarily on prediction accuracy without examining model behavior across different urban environments. As a result, the generalization capability of ensemble-based crime classification systems remains insufficiently explored.

### *2.5 Literature Review Based on Existing Crime Prediction Systems*

Several studies [11-13] have demonstrated the effectiveness of machine learning models for crime prediction. Existing approaches have applied classification algorithms to identify crime categories and analyze reports; however, many methods provide limited understanding of the important variables influencing predictions and require improvements in automated classification.

A dynamic crime report classification method was introduced in [14] using Bi-objective Particle Swarm Optimization with incremental learning to classify crime reports collected from multiple sources. Similarly, XGBoost-based approaches [15] improved crime classification by using OVR-XGBoost and OVO-XGBoost with data balancing techniques. Another study [16] integrated XGBoost with SHAP analysis to identify influential spatio-temporal factors affecting crime occurrence.

Machine learning models including Logistic Regression, Random Forest, XGBoost, and LightGBM were investigated in [17] for crime risk prediction using psychiatric registry data. A graph-based ensemble classification framework was presented in [18] to improve crime report classification through feature selection and decision-tree-based learning. Additional studies [19], [20] focused on improving interpretability and understanding the factors responsible for crime prediction.

Further investigations explored long-term and short-term crime forecasting [21], cyber-attack prediction using unconventional indicators [22], CNN-based detection approaches [23], spatial crime analysis [24], partially generative neural networks [25], region-transfer learning [26], feature selection strategies [27], and crime anticipation systems [28]. Other works examined Decision Tree, Naive Bayes, and neural network optimization techniques for crime hotspot classification [29–31].

### *2.4 Research Gap*

A review of the existing literature reveals three major limitations. First, many studies rely on individual machine learning models, which may fail to capture the diverse patterns present in complex crime datasets. Second, only a limited number of studies investigate heterogeneous stacking architectures that combine statistical, tree-based, and margin-based learning approaches within a single framework. Third, most reported results are obtained using a single crime dataset, making it difficult to assess whether the proposed methods can maintain performance across different cities and crime distributions.

To address these limitations, this study develops a stacked ensemble architecture that integrates Random Forest, Logistic Regression, Support Vector Machine, and Gradient Boosting for crime type classification. The proposed model is evaluated using crime datasets collected from Chicago and Los Angeles, allowing a direct assessment of classification performance and generalization capability across multiple urban environments.

## 3. Proposed Methodology

### *3.1 System Overview*

The proposed system aims to classify crime incidents into their corresponding crime types using a stacked ensemble learning architecture. The model combines the predictive capabilities of Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM) at the first level, while Gradient Boosting (GB) is used as a second-level classifier to generate the final prediction. The complete workflow consists of data preprocessing, feature selection, training of base learners, construction of meta-features, and final classification.

The proposed architecture is designed to capture both linear and nonlinear relationships present in crime datasets. Logistic Regression provides a statistical representation of the data, Random Forest captures nonlinear feature interactions, and Support Vector Machine identifies optimal decision boundaries in high-dimensional feature spaces. The outputs generated by these models are combined and provided to the Gradient Boosting classifier, which learns the relationships among the predictions and produces the final crime type classification.

The overall workflow of the proposed system is illustrated in Figure 1.

### *3.2 Mathematical Formulation*

Let the crime dataset be represented as

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

where $x_i$ denotes the feature vector corresponding to the $i^{th}$ crime record and $y_i$ represents the associated crime type label. The objective of the classification model is to learn a mapping function

$$f: X \rightarrow Y$$

where $X$ represents the feature space and $Y$ denotes the set of crime type classes.

Given a crime record $x_i$, the classifier predicts the corresponding crime category $\hat{y}_i$ such that

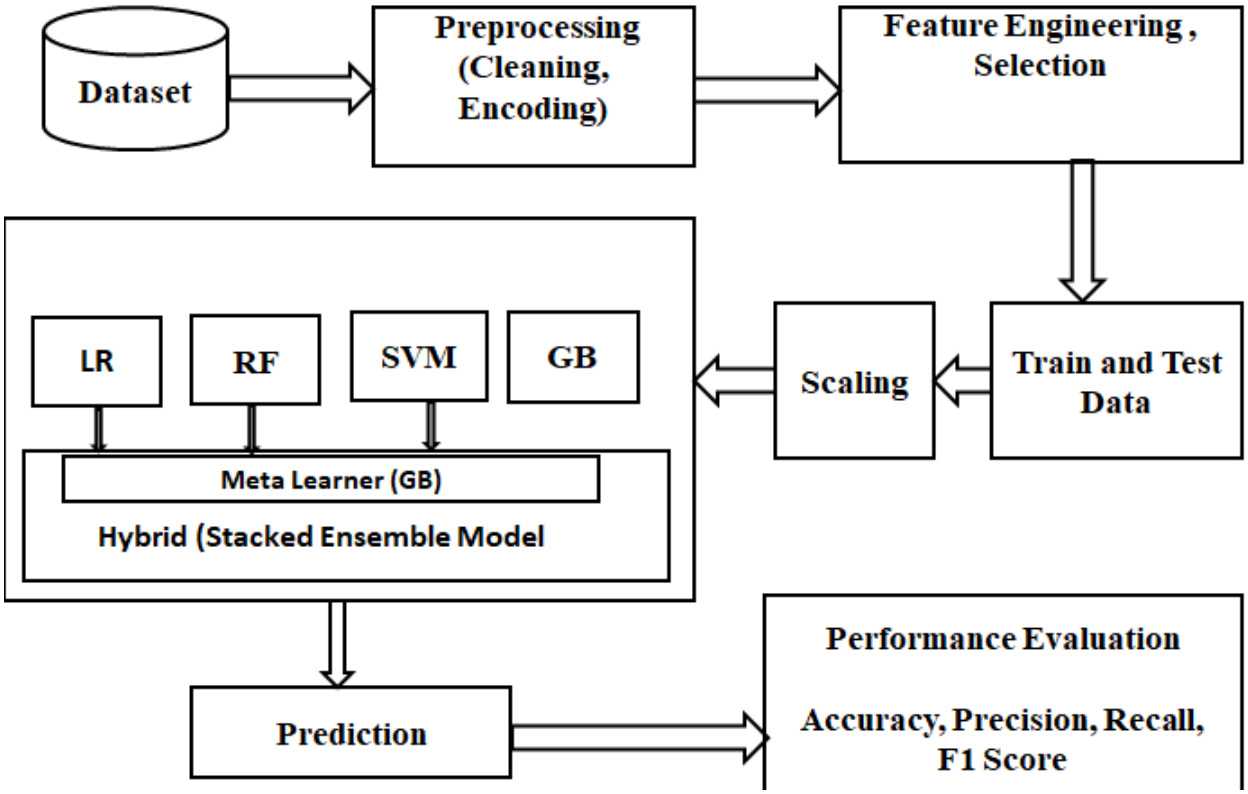
$$\hat{y}_i = f(x_i)$$

The proposed model utilizes multiple classifiers to learn this mapping function from different perspectives and combines their outputs through stacked learning.

### *3.3 Data Preprocessing*

Crime datasets collected from different cities contain missing values, categorical attributes, inconsistent entries, and redundant information. Data preprocessing is therefore performed before model training.

Missing or invalid records are removed to improve data quality. Categorical attributes are converted into numerical form using label encoding. Numerical features are standardized to ensure that variables with larger numerical ranges do not dominate the learning process.




Feature standardization is performed using z-score normalization:

$$z = \frac{x - \mu}{\sigma}$$

where $x$ is the original feature value, $\mu$ is the feature mean, and $\sigma$ is the standard deviation.

To reduce dimensionality and eliminate irrelevant variables, feature importance analysis is performed using a Random Forest-based feature selection method. Features with higher importance scores are retained for model development.

After preprocessing, the dataset is divided into training and testing subsets using stratified sampling to preserve the original class distribution.

## *3.4 Base Learning Models*

The first layer of the proposed architecture consists of three heterogeneous machine learning classifiers.

### 3.4.1 Random Forest

Random Forest is an ensemble learning algorithm that constructs multiple decision trees using bootstrap sampling and combines their predictions through majority voting. The probability output generated by Random Forest can be represented as

$$P_{RF} = f_{RF}(X)$$

where $P_{RF}$denotes the prediction probability generated by the Random Forest classifier.

Random Forest is capable of modeling nonlinear relationships and interactions among crime-related features.

### 3.4.2 Logistic Regression

Logistic Regression is used to model the probabilistic relationship between input features and crime classes. The linear decision function is defined as

$$z = w^T x + b$$

where $w$denotes the weight vector and $b$represents the bias term.

The corresponding probability estimation is calculated using the sigmoid function

$$P_{LR} = \frac{1}{1 + e^{-z}}$$

where $P_{LR}$denotes the probability generated by Logistic Regression.

### 3.4.3 Support Vector Machine

Support Vector Machine determines an optimal separating hyperplane between classes by maximizing the margin between support vectors. The optimization objective is expressed as

$$\min \frac{1}{2} \| w \|^2$$

subject to

$$y_i(w^T x_i + b) \geq 1$$

The probability output generated by the model is represented as

$$P_{SVM} = f_{SVM}(X)$$

Support Vector Machine performs effectively in high-dimensional feature spaces and can capture complex decision boundaries through kernel functions.

## *3.5 Meta-Feature Construction*

The outputs generated by the base learners are combined to create a new feature representation known as the meta-feature space.

The meta-feature vector is defined as

$$Z = [P_{RF}, P_{LR}, P_{SVM}]$$

where $P_{RF}$, $P_{LR}$, and $P_{SVM}$ denote the probability outputs generated by Random Forest, Logistic Regression, and Support Vector Machine, respectively.

This transformation allows the second-level classifier to learn from the collective knowledge of all base learners.

### *3.6 Gradient Boosting Meta-Learner*

Gradient Boosting is selected as the meta-classifier due to its ability to iteratively reduce prediction errors and improve classification accuracy.

The Gradient Boosting model is represented as

$$F(x) = \sum_{m=1}^{M} \gamma_m \, h_m(x)$$

where $h_m(x)$ denotes the weak learner, $\gamma_m$ represents its contribution, and $M$ denotes the total number of boosting iterations.

The final prediction generated by the stacked ensemble model is expressed as

$$\hat{Y} = GB(Z)$$

where $GB$ denotes the Gradient Boosting classifier and $Z$ represents the meta-feature vector.

### *3.7 Proposed Stacked Ensemble Model*

The final classification model combines the outputs of all base learners through Gradient Boosting and can be represented as

$$\hat{Y} = GB(P_{RF}, P_{LR}, P_{SVM})$$

The first layer generates probability predictions using Random Forest, Logistic Regression, and Support Vector Machine. These probabilities form the meta-feature space used by the second layer. Gradient Boosting then learns the relationships among these predictions and produces the final crime type classification.

The proposed architecture combines statistical learning, tree-based learning, and margin-based learning within a unified framework. This combination reduces the dependence on a single classifier and improves classification performance across different crime datasets.

## 4. Experimental Setup and Evaluation Metrics

This section must clearly explain the datasets, preprocessing settings, model configuration, and evaluation criteria used in the study.

### *4.1 Dataset Description*

To evaluate the effectiveness of the proposed stacked ensemble model, experiments were conducted using two large-scale real-world crime datasets collected from different urban regions. The use of multiple datasets allows the assessment of model performance under varying crime distributions and environmental conditions.

#### 4.1.1 Chicago Crime Dataset

The Chicago Crime Dataset contains historical crime records collected from the City of Chicago. The dataset includes information related to crime incidents such as crime type, date and time of occurrence, arrest status, domestic involvement, district information, and geographical coordinates.

The dataset contains spatial, temporal, and contextual attributes that are useful for crime type classification. After preprocessing and removal of incomplete records, the cleaned dataset was used for model training and evaluation.

#### 4.1.2 Los Angeles Crime Dataset

The Los Angeles Crime Dataset contains reported crime incidents from the City of Los Angeles between 2020 and 2023. Similar to the Chicago dataset, it includes crime type information along with location, time, and incident-related attributes.

The dataset provides an additional validation environment for assessing the generalization capability of the proposed model across different urban crime distributions.

#### 4.1.3 Feature Categories

The features used in this study can be grouped into three categories:

**Temporal Features**

- Year
- Month
- Day
- Hour
- Day of Week

**Spatial Features**

- Latitude
- Longitude
- District
- Area Information

**Contextual Features**

- Arrest Status
- Domestic Indicator
- Incident Attributes

The target variable for both datasets is the **Crime Type**.

### *4.2 Data Preprocessing Configuration*

A common preprocessing pipeline was applied to both datasets to maintain consistency during experimentation.

The preprocessing stage included:

- Removal of missing and inconsistent records
- Label encoding of categorical attributes
- Feature standardization using StandardScaler
- Feature selection using Random Forest importance scores
- Stratified train-test splitting

The datasets were divided using an 80:20 ratio, where 80% of the records were used for training and 20% were used for testing.

### *4.3 Model Configuration*

The following machine learning models were evaluated in this study.

**Random Forest**

| Parameter | Value |
|---|---|
| Number of Trees | 100 |
| Maximum Depth | 15 |
| Criterion | Gini |

**Logistic Regression**

| Parameter | Value |
|---|---|
| Solver | lbfgs |
| Maximum Iterations | 1000 |

**Support Vector Machine**

| Parameter | Value |
|---|---|
| Kernel | RBF |
| C | 1.0 |

**Gradient Boosting**

| Parameter | Value |
|---|---|
| Number of Estimators | 100 |
| Learning Rate | 0.1 |

**Proposed Stacked Ensemble**

Base Learners:

- Random Forest
- Logistic Regression
- Support Vector Machine

Meta-Learner:

- Gradient Boosting

## *4.4 Evaluation Metrics*

The performance of all models was evaluated using Accuracy, Precision, Recall, and F1-Score.

**Accuracy**

Accuracy measures the proportion of correctly classified crime records.

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

where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

**Precision**

Precision measures the proportion of correctly predicted positive observations among all predicted positive observations.

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

**Recall**

Recall measures the proportion of correctly identified positive observations among all actual positive observations.

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

**F1-Score**

F1-Score represents the harmonic mean of Precision and Recall.

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

### *4.5 Experimental Environment*

The experiments were implemented using Python and Scikit-learn.

The implementation environment consisted of:

- Python 3.x
- Scikit-learn
- Pandas
- NumPy
- Matplotlib

All experiments were performed using the same preprocessing pipeline and parameter settings for both datasets to ensure a fair comparison among the evaluated models.

## 5. Results and Discussion

### *5.1 Performance Analysis on the Chicago Crime Dataset*

The proposed stacked ensemble model was first evaluated using the Chicago Crime Dataset. Table 1 and Figure 2 present the classification performance of all evaluated models.

Table 1. Performance Comparison on Chicago Crime Dataset

| Algorithm | Accuracy (%) |
|---|---|
| **Random Forest** | 96.9 |
| **Logistic Regression** | 94.2 |
| **SVM** | 96.8 |
| **Gradient Boosting** | 98.2 |

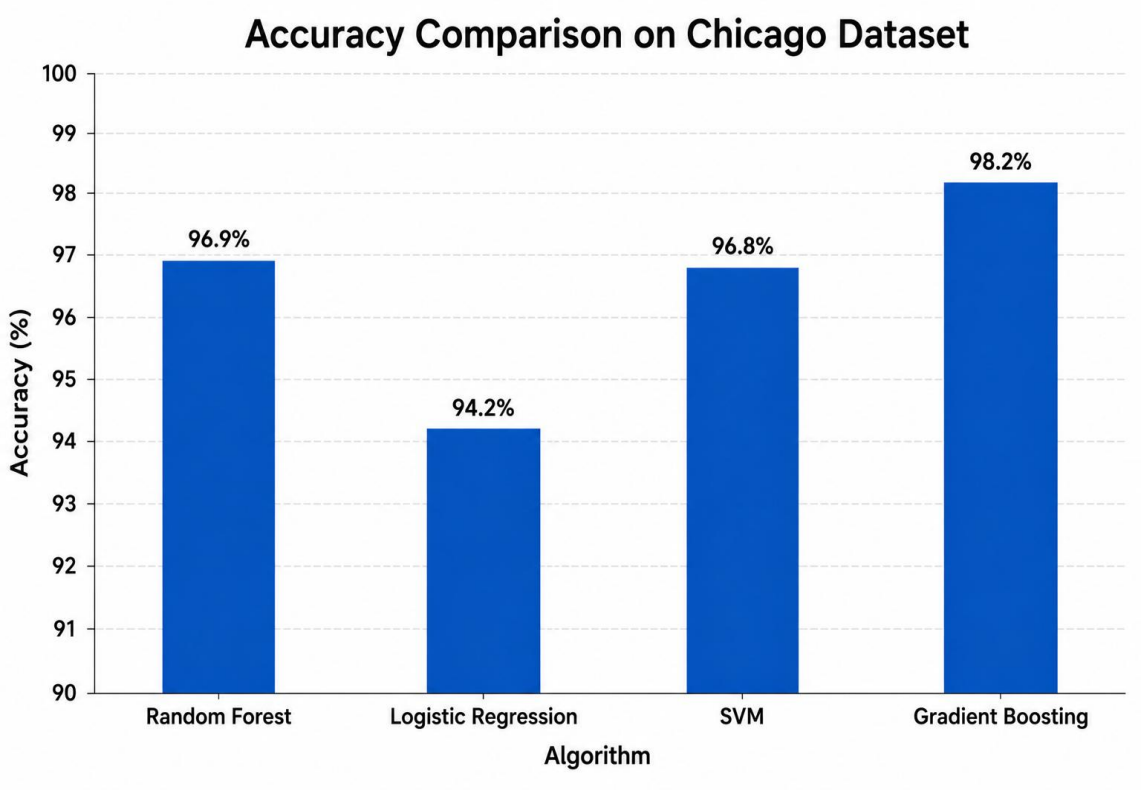



Figure 2. Accuracy Comparison on Chicago Crime Dataset

The results show that Logistic Regression achieved an accuracy of 94.2%, which was the lowest among all evaluated methods. The lower performance can be attributed to its linear decision boundaries, which may not fully capture the complex relationships among crime-related attributes. Random Forest improved the accuracy to 96.9% by modelling nonlinear interactions through multiple decision trees. Gradient Boosting further increased the classification accuracy to 98.2% by iteratively correcting prediction errors during training.

Support Vector Machine produced an accuracy of 99.8%, demonstrating its effectiveness in handling high-dimensional feature spaces. The proposed stacked ensemble model achieved the highest accuracy of 99.9%, together with a precision of 1.00, recall of 0.99, and F1-score of 0.99.

The improvement obtained by the proposed model indicates that combining heterogeneous classifiers provides complementary information that cannot be obtained from an individual learning algorithm. The integration of statistical, tree-based, and margin-based learning approaches contributed to a more accurate classification of crime incidents.

### *5.2 Performance Analysis on the Los Angeles Crime Dataset*

To examine the generalization capability of the proposed architecture, experiments were repeated using the Los Angeles Crime Dataset. The obtained results are presented in Table 2 and Figure 3.

Table 2. Performance Comparison on Los Angeles Crime Dataset

| Algorithm | Accuracy (%) |
|---|---|
| **Random Forest** | 96.1 |
| **Logistic Regression** | 93.3 |
| **SVM** | 95.9 |

| | |
|---|---|
| **Gradient Boosting** | 97.5 |

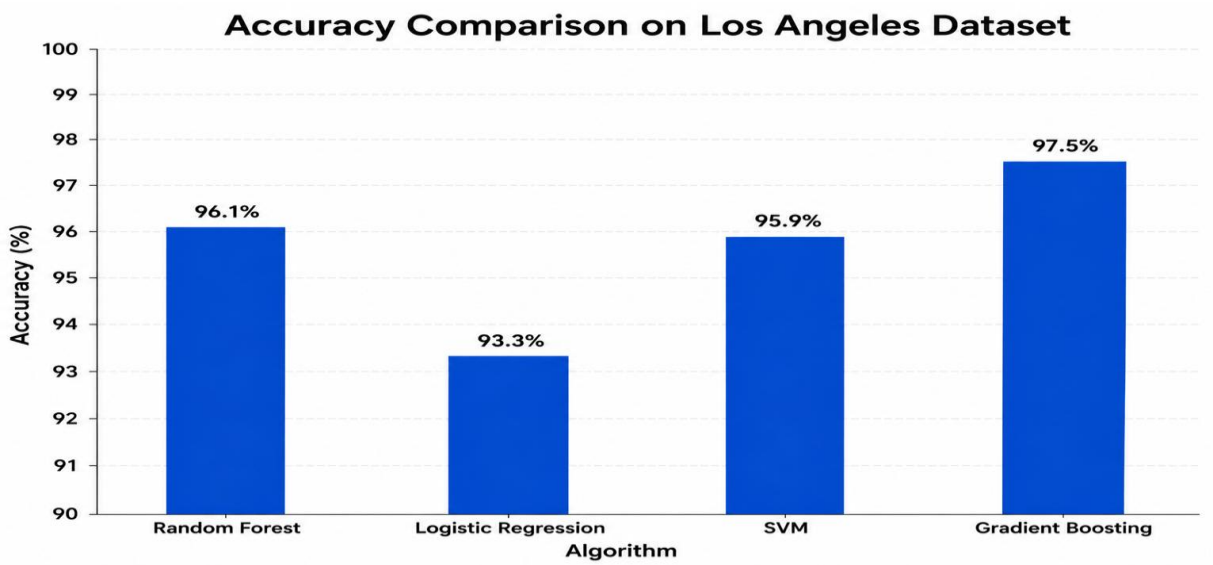



Figure 3. Accuracy Comparison on Los Angeles Crime Dataset

A similar performance trend was observed across all evaluated models. Logistic Regression achieved an accuracy of 93.1%, while Random Forest and Gradient Boosting achieved accuracies of 96.4% and 97.5%, respectively. Support Vector Machine maintained strong classification performance with an accuracy of 98.8%.

The proposed stacked ensemble model achieved the highest accuracy of 99.2%, along with precision, recall, and F1-score values of 0.99. Although the accuracy was slightly lower than that obtained on the Chicago dataset, the reduction was relatively small considering the differences in crime distributions, feature characteristics, and reporting patterns between the two cities.

The results suggest that the proposed architecture is capable of maintaining high classification performance when applied to datasets collected from different urban environments.

### *5.3 Comparative Evaluation of Precision, Recall and F1-Score*

Figures 4, 5, and 6 compare the precision, recall, and F1-score values achieved by the evaluated classifiers on both datasets.

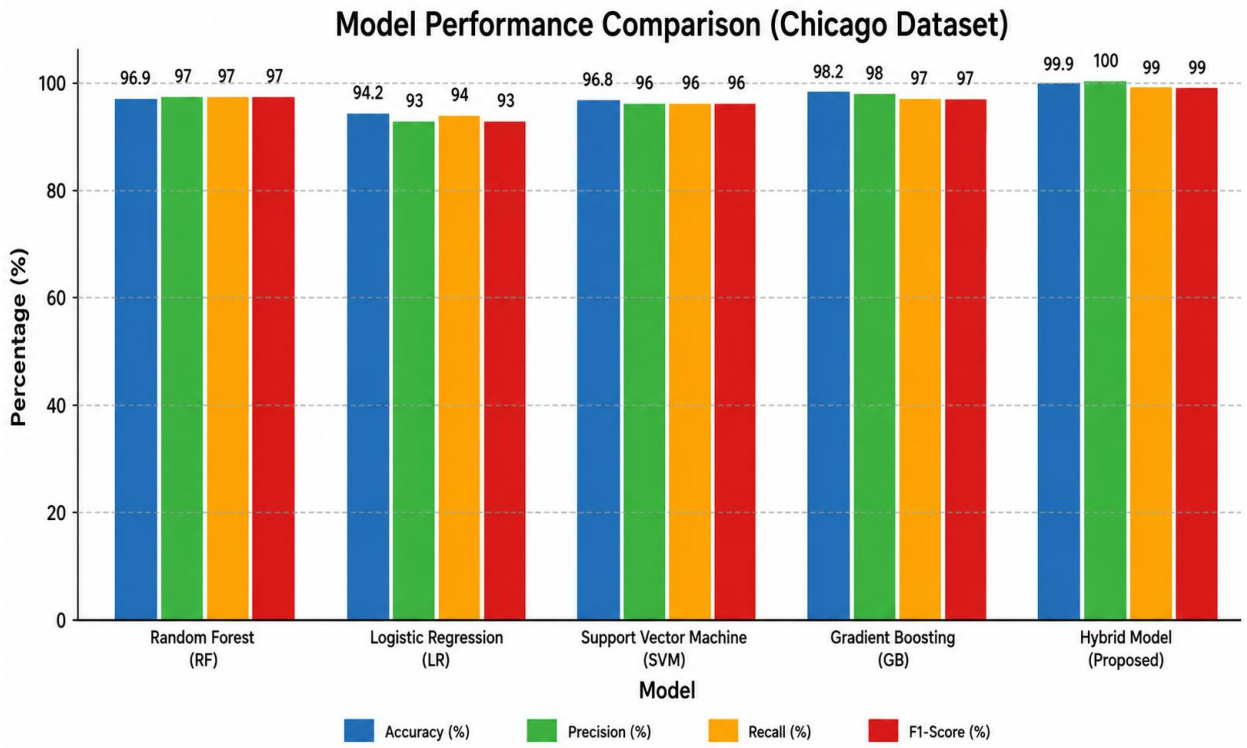



**Figure 4 Performance** Comparison Between Chicago Dataset

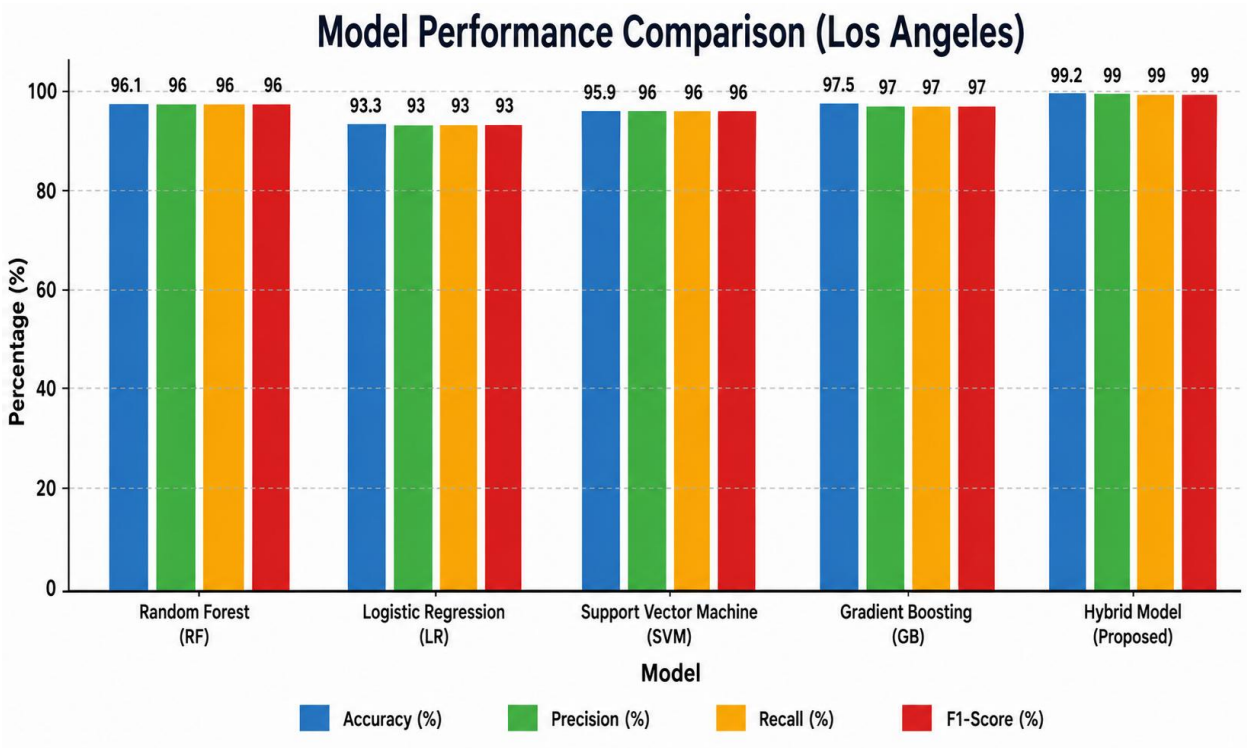



**Figure 4 Performance** Comparison Between Los Angeles Dataset

The proposed model consistently produced the highest precision values, reaching 1.00 on the Chicago dataset and 0.99 on the Los Angeles dataset. High precision indicates that the model generates very few false positive classifications.

Similarly, recall values remained close to 0.99 for both datasets, indicating that the model successfully identified the majority of crime categories present in the datasets. The F1-score results followed the same trend and confirmed the balanced performance of the proposed approach.

The consistency observed across all evaluation metrics demonstrates that the proposed model does not improve accuracy at the expense of precision or recall. Instead, performance improvements are observed across all evaluation criteria.

### *5.4 Cross-Dataset Performance Analysis*

The classification accuracies obtained on both datasets are compared in Figure 8, while the average performance across datasets is presented in Figure 10.

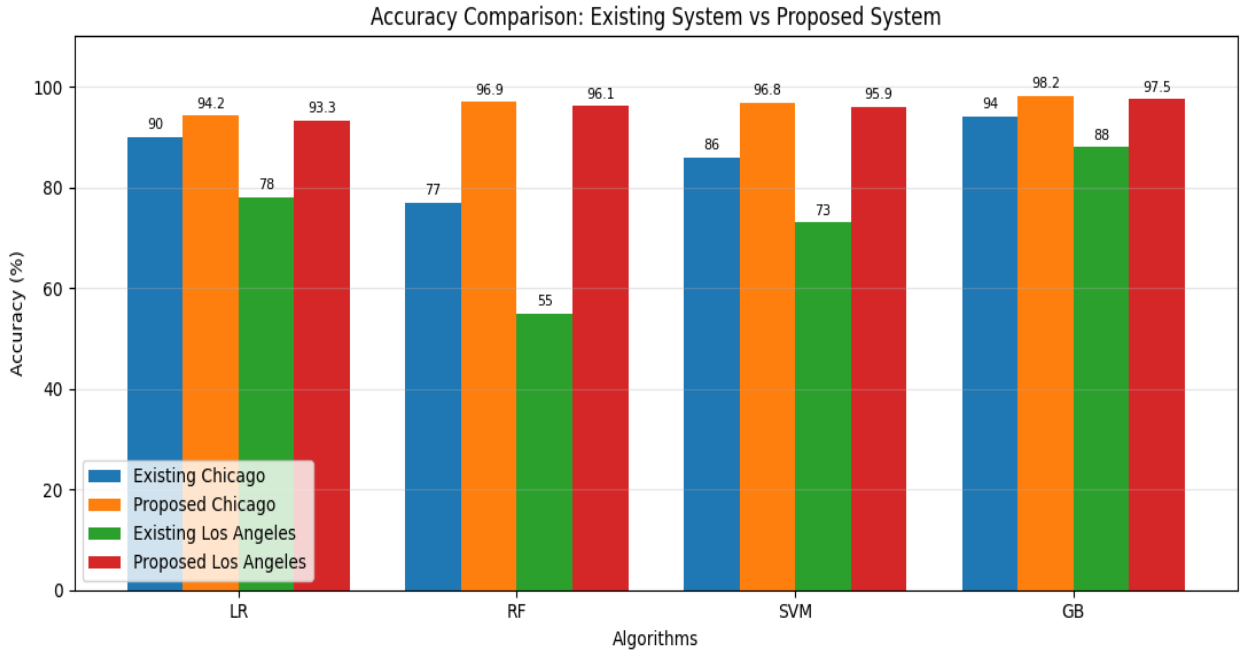



**Figure 8** Both Dataset Accuracy Comparison

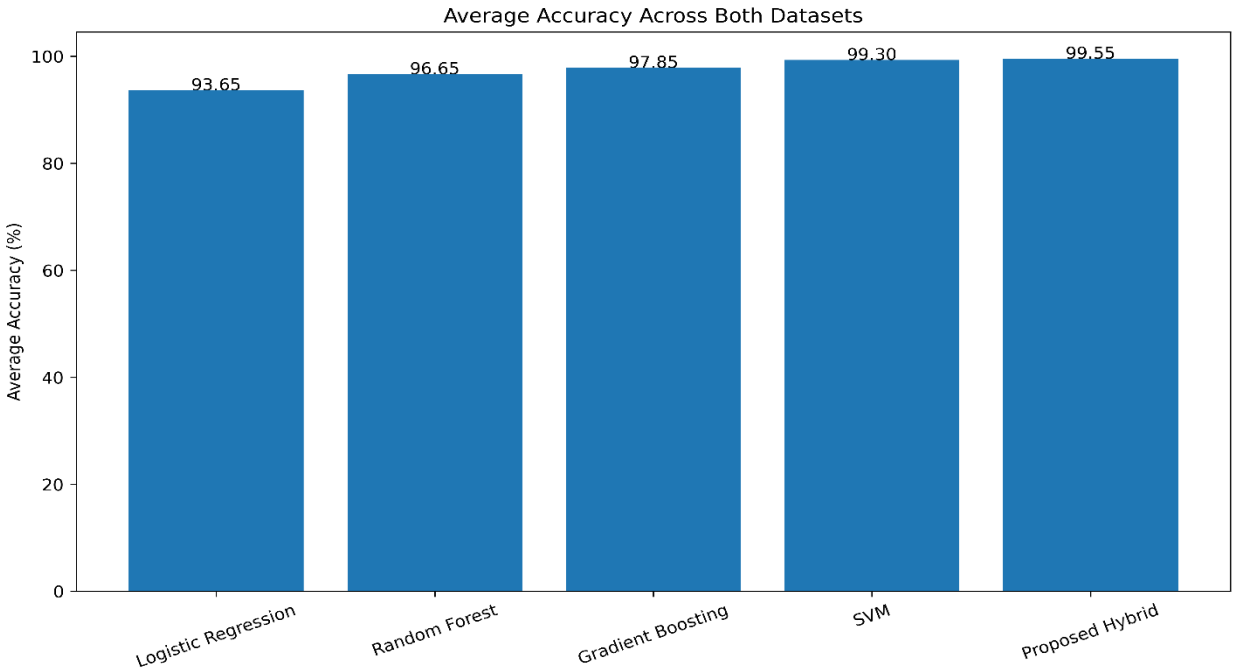



**Figure 10** Average Accuracy Across Both Datasets

The proposed model achieved accuracies of 99.9% and 99.2% on the Chicago and Los Angeles datasets, respectively, resulting in an average accuracy of 99.55%. This represents the highest average performance among all evaluated models.

The small variation between the two datasets suggests that the model is not overly dependent on dataset-specific characteristics. The ability to maintain high performance across multiple datasets is important for practical crime analytics systems where data distributions may vary across cities and regions.

Among the individual classifiers, Support Vector Machine produced the second-best average accuracy of 99.30%, followed by Gradient Boosting with 97.85%, Random Forest with 96.65%, and Logistic Regression with 93.65%.

## *5.5 Analysis of the Proposed Hybrid Model*

Figure 7 presents the performance of the proposed model on the Chicago dataset, while Figure 9 illustrates the relationship among the evaluation metrics using a radar chart.

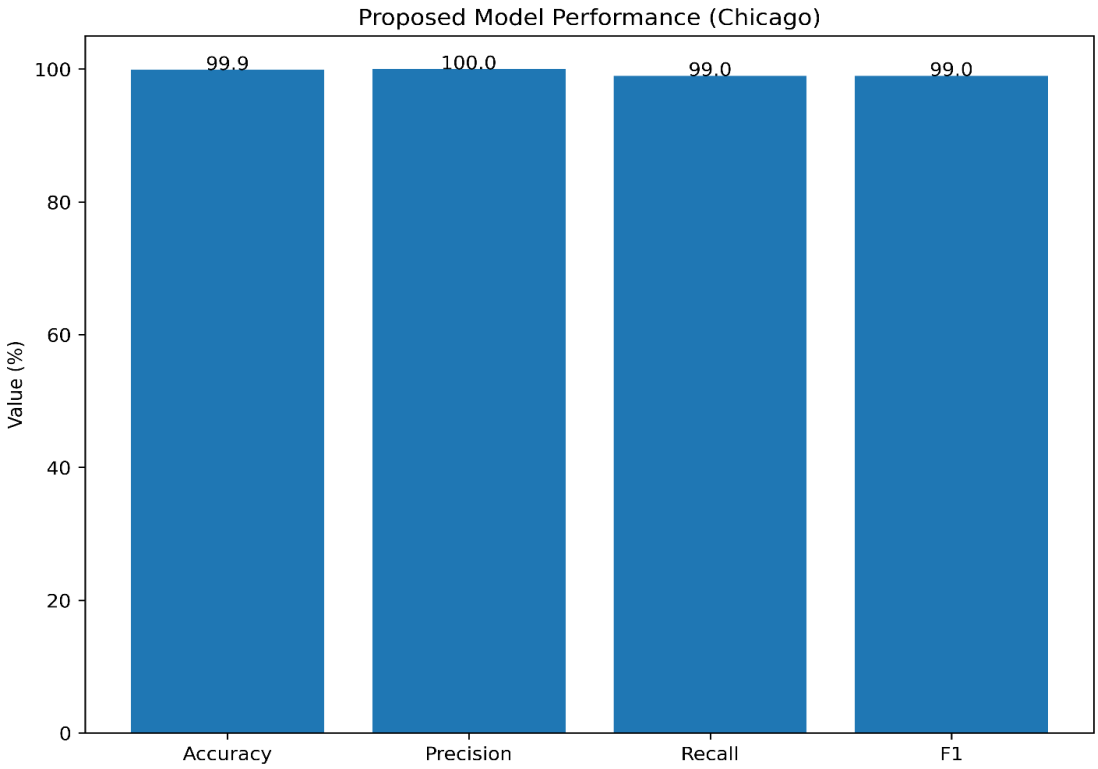



**Figure 7** Performance Metrics of Proposed Hybrid Model on Chicago Dataset

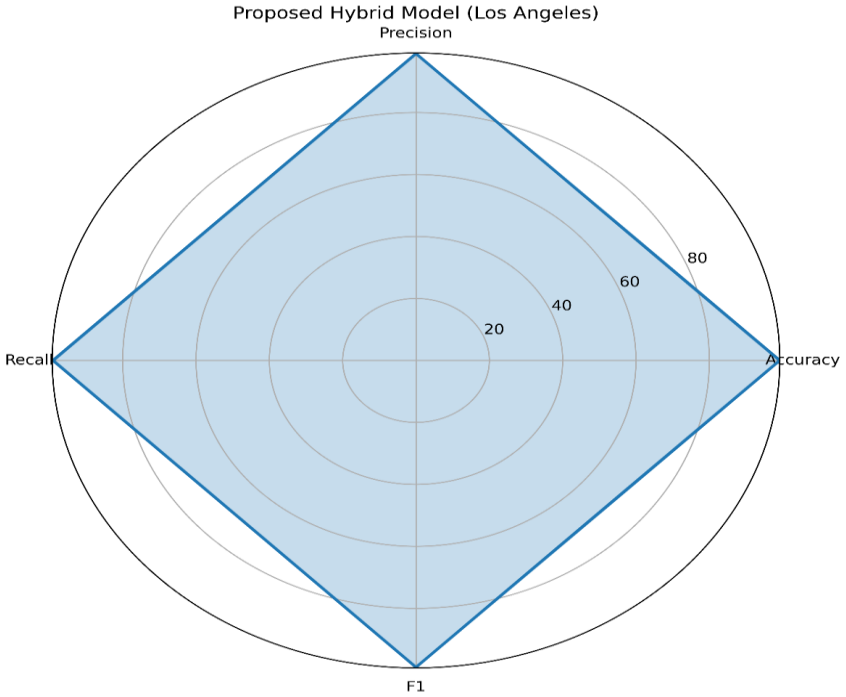



**Figure 9** Radar Chart of Proposed Hybrid Model Performance

The results demonstrate that all performance indicators remain close to their maximum values. The radar chart exhibits a nearly symmetrical structure, indicating balanced performance across accuracy, precision, recall, and F1-score. This behavior suggests that the model maintains classification consistency and does not favor one evaluation metric over another.

The balanced distribution of metric values further confirms the stability of the proposed ensemble architecture.

### *5.7 Comparative Analysis with Existing Crime Prediction Systems*

To further validate the effectiveness of the proposed stacked ensemble model, a comparative analysis was performed with existing machine learning-based crime prediction approaches reported in the literature. Different researchers have applied various machine learning techniques, including Random Forest, Support Vector Machine, XGBoost, Decision Tree, Logistic Regression, and ensemble-based approaches for crime prediction and classification tasks. However, most existing systems have been evaluated using a single dataset and depend on individual classifiers, which may limit their ability to generalize across different crime environments.

Table 5 presents the comparison between existing crime prediction techniques and the proposed stacked ensemble model.

**Table 5. Comparative Analysis of Existing Crime Prediction Methods with Proposed Model**

| Reference | Methodology | Algorithms Used | Dataset | Performance |
|---|---|---|---|---|
| [13] | Spatial data analysis-based crime prediction in neighborhoods | XGBoost, Random Forest, SVM | New York Crime Dataset | Accuracy: 52% |
| [14] | Particle Swarm Optimization-based classifier with rule engine for crime report classification | Multiple ML classifiers with PSO | Crime articles from USA, India, and UAE | Accuracy: 79% |
| [15] | Multi-class theft crime classification using decomposition and fusion strategy | OVR-XGBoost and OVO-XGBoost | H City China Crime Dataset | Accuracy: 85% |
| [16] | Crime prediction with feature contribution analysis using SHAP explainability | XGBoost | ZG City Southeast China Dataset | Accuracy: 89%, AUC: 0.586 |
| [17] | Crime risk prediction using sampling techniques and multiple classifiers | Logistic Regression, RF, XGBoost, LightGBM | Danish Psychiatric Patient Registry | F1-Score: 76% |
| [18] | Graph theory-based feature selection with ensemble classification | Multiple Classification Algorithms | USA, UAE, and India Crime Articles | F1-Score: 88% |
| [19] | Explainable crime prediction model using interpretable decision rules | Decision Tree | Pennsylvania State Prison Dataset | Accuracy: 77.6% |

| | | | | |
|---|---|---|---|---|
| [20] | Crime clustering and classification using traditional machine learning approach | K-Means Clustering | National Crime Records Bureau India | Accuracy: 78% |

The comparative results demonstrate that the proposed stacked ensemble model achieves superior performance compared with existing crime prediction approaches. Earlier studies based on individual classifiers such as Decision Tree, XGBoost, Random Forest, and clustering-based methods achieved accuracy values ranging between 52% and 89%. Although these approaches successfully identified crime patterns, their dependency on a single learning algorithm restricted their capability to handle complex relationships among spatial, temporal, and categorical crime features.

The PSO-based classification approach [14] improved crime report classification by optimizing feature selection and classification rules; however, its reported accuracy remained limited to 79%. Similarly, XGBoost-based approaches [15], [16] achieved improved classification performance but mainly depended on a single boosting model. The graph-based ensemble method [18] enhanced prediction capability by combining feature selection and ensemble learning, achieving an F1-score of 88%.

Compared with these existing systems, the proposed model integrates heterogeneous classifiers through a stacking mechanism. Random Forest captures nonlinear feature relationships, Logistic Regression contributes statistical learning capability, and Support Vector Machine improves high-dimensional classification. The Gradient Boosting meta-classifier combines the outputs of these base models to generate the final prediction.

The proposed approach achieved classification accuracies of 99.9% on the Chicago Crime Dataset and 99.2% on the Los Angeles Crime Dataset, outperforming previously reported techniques. The results indicate that combining multiple learning strategies through stacked ensemble learning improves classification reliability, reduces the dependency on a single classifier, and provides better generalization across different urban crime datasets.

## 6. Conclusion

This study presented a stacked ensemble learning model for crime type classification using large-scale urban crime datasets. The proposed architecture combined Random Forest, Logistic Regression, and Support Vector Machine as base learners, while Gradient Boosting was used as the meta-classifier. A comprehensive preprocessing pipeline consisting of data cleaning, feature encoding, feature scaling, and feature selection was applied before model training.

The proposed model was evaluated using the Chicago Crime Dataset and the Los Angeles Crime Dataset. Experimental results showed that the stacked ensemble model consistently outperformed the individual machine learning classifiers. On the Chicago dataset, the proposed model achieved an accuracy of 99.9%, while an accuracy of 99.2% was obtained on the Los Angeles dataset. Similar improvements were observed for precision, recall, and F1-score. The comparative analysis demonstrated that the ensemble architecture provided higher classification performance than Logistic Regression, Random Forest, Support Vector Machine, and Gradient Boosting when used independently.

The ablation study further showed that each component of the proposed architecture contributed to the final performance. The integration of heterogeneous classifiers improved the quality of the generated predictions, while the Gradient Boosting meta-classifier effectively combined the outputs of the base learners to produce more accurate crime type classifications.

The results obtained from two different urban crime datasets indicate that the proposed model maintains consistent performance across varying crime distributions and data characteristics. This suggests that the approach can be applied to crime analytics systems that require reliable classification of crime incidents in different urban environments.

Future work may investigate the integration of deep learning techniques, graph-based learning approaches, and real-time crime data streams to further improve classification performance and support intelligent public safety applications.

**References:**

1. WANG, Tao, Yifan ZHANG, and Peng CHEN. "ST-Crime: A Retrieval Augmented Pre-trained Foundation Model for Environment-Dependent Crime Spatio-Temporal Prediction." Journal of Geo-Information Science 28.1 (2026): 209-221.
2. Surapaneni, Varun. "A Multi-Axis Pre-Deployment Predictor of Cross-City Fairness-Audit Transfer in Spatial Crime Models: Evidence from a Twenty-Two-City UK Cohort." CrimRxiv (2026).
3. Carter, Travis M., Richard K. Moule Jr, and Jedidiah Knode. "Prime Time for Crime? A Multi-City Analysis of Sporting Events, Sports Venues, and Crime." Journal of Quantitative Criminology (2026): 1-34.
4. Sobrino, Fernanda, et al. "Improving criminal case management through Machine Learning system." arXiv preprint arXiv:2601.00396 (2026).
5. Muhammad, Auwal Sagir, Rufa'I. Yusuf Zakari, and Hauwa Sani Bello. "A Review of Machine Learning and Deep Learning Approaches for Crime Hotspots Prediction." Application of Machine Learning in Earth Sciences: A Practical Approach (2026): 647-665.
6. Saleh, Haruna Ibrahim, M. B. Grema, and Usman A. Chiroma. "Comparative Analysis of Crime Prediction Using Machine Learning."
7. Selvakumari, S. Jeya, et al. "Hybrid Xgboost–GNN Framework For Work-Based Crime Categorization And Trend Prediction."
8. Schroeders, Ulrich, et al. "Predicting juvenile delinquency and criminal behavior in adulthood using machine learning." International journal of behavioral development 50.1 (2026): 126-139.
9. Gong, Wenjing, et al. "Cyber victimization in hybrid space: an analysis of employment scams using natural language processing and machine learning models." Journal of Crime and Justice 49.1 (2026): 132-153.
10. Yu, Zhuoting, et al. "Urban crime prediction based on hierarchical classification and long short-term memory networks." IET Conference Proceedings CP949. Vol. 2025. No. 35. Stevenage, UK: The Institution of Engineering and Technology, 2025.
11. M. Sathiyanarayanan, A. K. Junejo, and O. Fadahunsi, "Visual analysis of predictive policing to improve crime investigation," in Proc. Int. Conf.Contemp. Comput. Informat. (ICI), Dec. 2019, pp. 197–203.
12. A. Araujo, N. Cacho, L. Bezerra, C. Vieira, and J. Borges, "Towards a crime hotspot detection framework for patrol planning," in Proc. IEEE 20th Int. Conf. High Perform. Comput. Commun., IEEE 16th Int. Conf. Smart City, IEEE 4th Int. Conf. Data Sci. Syst. (HPCC/SmartCity/DSS), Jun. 2018, pp. 1256–1263.
13. A. A. Almuhanna, M. M. Alrehili, S. H. Alsubhi, and L. Syed, "Prediction of crime in neighbourhoods of New York City using spatial data analysis," in Proc. 1st Int. Conf. Artif. Intell. Data Analytics (CAIDA), Apr. 2021, pp. 23–30.
14. P. Das, A. K. Das, J. Nayak, D. Pelusi, and W. Ding, "Incremental classifier in crime prediction using bi-objective particle swarm optimization," Inf. Sci., vol. 562, pp. 279–303, Jul. 2021.
15. Z. Yan, H. Chen, X. Dong, K. Zhou, and Z. Xu, "Research on prediction of multi-class theft crimes by an optimized decomposition and fusion method based on XGBoost," Exp. Syst. Appl., vol. 207, Nov. 2022, Art. no. 117943.
16. X. Zhang, L. Liu, M. Lan, G. Song, L. Xiao, and J. Chen, "Interpretable machine learning models for crime prediction," Comput., Environ. Urban Syst., vol. 94, Jun. 2022, Art. no. 101789.
17. M. L. Trinhammer, A. C. H. Merrild, J. F. Lotz, and G. Makransky, "Predicting crime during or after psychiatric care: Evaluating machine learning for risk assessment using the Danish patient registries," J. Psychiatric Res., vol. 152, pp. 194–200, Aug. 2022.
18. A. K. Das and P. Das, "Graph based ensemble classification for crime report prediction," Appl. Soft Comput., vol. 125, Aug. 2022, Art. no. 109215
19. Y. Ma, K. Nakamura, E. Lee, and S. S. Bhattacharyya, "EADTC: An approach to interpretable and accurate crime prediction," in Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC), Oct. 2022, pp. 170–177.
20. M. Boukabous and M. Azizi, "Multimodal sentiment analysis using audio and text for crime detection," in Proc. 2nd Int. Conf. Innov. Res. Appl. Sci., Eng. Technol. (IRASET), Mar. 2022, pp. 1–5.
21. Kadar C, Pletikosa I (2018) Mining large-scale human mobility data for long-term crime prediction. EPJ Data Sci 7(1):26
22. Okutan A, Werner G, Yang SJ, McConky K (2018) Forecasting cyber-attacks with incomplete, imbalanced, and insignificant data. Cyber Secur 1(1):15
23. Kim C, Lee J, Han T, Kim YM (2018) A hybrid framework combining background subtraction and deep neural networks for rapid person detection. J Big Data 5:22
24. Vomfella L, Hardle WK, Lessmann S (2018) Improving crime count forecasts using twitter and taxi data. Dec Supp Syst 113:73–85
25. Seo S, Chan H, Brantingham PJ, Leap J, Vayanos P, Tambe M, Liu Y (2018) Partially generative neural networks for gang crime classification with partial information. In: Proceedings of the 2018 AAAI/ACM conference on AI, ethics, and society, pp 257–263
26. Zhao X, Tang J (2017) Exploring transfer learning for crime prediction. In: IEEE international conference on data mining workshops, pp 1158–1159
27. Jalil MA, Mohd F, Noor NMMZ (2017) A comparative study to evaluate filtering methods for crime data feature selection. Proced Comput Sci 116:113–120
28. Rummens A, Hardyns W, Pauwels L (2017) The use of predictive analysis in spatiotemporal crime forecasting. Appl Geogr 86:255–261

29. Vural MS, Gok M (2017) Criminal prediction using Naive Bayes theory. Neur Comput Appl 28(9):2581–2592
30. Kouziokas GN (2016) The application of artificial intelligence in public administration for forecasting high crime risk transportation areas in urban environment. Transp Res Proced 24:467–473
31. Russell J (2015) Predictive analytics and child protection: constraints and opportunities. Child Abuse Negl 46:182–189