FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 8, April, 2026, pp 87 – 95
DOI: https://doi.org/10.33003/fjs-2026-1008-5296

Comparative Analysis of Supervised Machine Learning Models for Crime Classification on Chicago Data

***Yazid Salihu and Musab Hassan Bodinga**

Department of Information and Communication Technology, Federal University Birnin-Kebbi, Kebbi State, Nigeria.

*Corresponding authors' email: yazid.salihu@fubk.edu.ng

Received: 19 Jun. 2026
Accepted: 17 Jul. 2026
Published: 14 Aug. 2026

ABSTRACT
The increase in crime can be caused by several factors. As growth in population and density, criminal behavior, domestic crime growth, and so on. This causes difficulties for the law enforcement agencies to control and monitor on a regular crime basis since it requires forecasting and probabilities, significant progress has been made from traditional machine learning techniques (ML) to modern techniques neural networks (NNs), the advancement of these technologies have improved the operational efficiency of critical infrastructures but have also rendered these substantially more vulnerable to domestic crime. In this study, we explore how at least four machine learning techniques perform in prediction of crime under measured experimental conditions. A dataset comprising 3,000 crime records was used for the experiment, consisting of three balanced classes: narcotics, burglary, and robbery, with 1,000 samples allocated to each category. The results show that decision tree (DT) achieved accuracy of 90.33%, and random forest (RF) with an accuracy of 90.33% consistently outperformed logistic regression (LR), support Vector machine (SVM), the SVM model reached a higher 90.16%, while the accuracy obtained by LR is peak at 90.03%.

Keywords: Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Crime Prediction (CP), Cyber Security (CS)

INTRODUCTION

The advent of machine learning and deep learning algorithms has transformed various sectors, including the area of computer information systems, security and surveillance where its use from crime prevention, crime detection and crime prediction to target detection (Mandalapu et al., 2023) the advancement of these technologies have improved the operational efficiency of critical infrastructures but have also rendered these substantially more vulnerable to domestic crime (Opeyemi Babatunde et al., 2025). The cyber-attacks and other related crime exposure and financial concerns these attacks entail for an individual and organization can be demonstrated through arrange of example (Opeyemi Babatunde et al., 2025). Among the most recent breaches is that at Marriott that revealed personal details of approximately 5.2 million hotel guests and households (Opeyemi Babatunde et al., 2025). Also, the breach at Twitter allowed fraudulent tweets about Bitcoin generating more than 100,000 dollars Bitcoin worth, while the Solar winds hack managed to compromise multiple government systems along with many fortune 500 companies, globally (Tsiodra et al., 2023). The latter, resulted in an 8% fall in the share price of FireEye after it disclosed information about the attack and is expected to cost cyber insurers 90 million dollars for incident response and forensic services (Tsiodra et al., 2023).

With the increasing availability of crime historical data and using advancement of existing technologies, several researchers were provided with a unique opportunity to study and conduct research on crime detection and prediction using ML (machine learning) and DL (deep learning) methodologies. Based on the recent advances in the field of security and surveillance (Neil et al., 2013), (Chun et al., 2019), (Kshatri et al., 2021).

The field of ML is a subset of AI that uses statistical models and algorithms to analyze data and make predictions based on historical records (Bharadiya, 2023). On the other hand, DL techniques are a subset of ML that uses artificial neural networks (ANN) with multiple layers to model complex relationships or patterns between inputs and outputs (Sharma & Chaudhary, 2023). Both machine learning and deep learning methodologies have the potential to be applied to the problem of crime prediction in many ways (Safat et al., 2021). ML algorithms have been developed in crime prediction to analyze data and predict future crime patterns (Kim et al., 2018). For example, algorithms like, decision trees, random forests, K-Nearest Neighbor and support vector machines have been trained on crime data from different locations in a specific cities to predict crime patterns accurately (Raza & Victor, 2021). Apart from predicting crime fluently crime trends and patterns (Chapman et al., 2022). These ML abilities allow for deploying resources and tactics to combat crime effectively (Trinhammer et al., 2022). Additionally, ML algorithms can also be used to identify crime incidents and various environmental and demographic correlations between factors such as location, weather, and time of day (Elluri et al., 2019). The information can be used to develop algorithms that can predict criminal activities and provide prevention strategies to a given community's specific needs.

Predictive policing is also a significant application of ML algorithms for crime prediction (Meijer & Wessels, 2019). Predictive policing refers to using historical data and analytics tools to inform law enforcement efforts and reduce crime activities (Meijer & Wessels, 2019). DL algorithms, such as long short term memory, convolution and recurrent neural networks (LSTM, CNN and RNN), have also shown promise in crime prediction (Kapadiya & Patel, 2022). These technologies have been trained on crime data with either a temporal or spatial component to predict crime patterns accurately in specific cities (Kapadiya & Patel, 2022). This information is used to create a predictive model (ML or DL) that can be used to identify potential crime hotspots and predict future criminal incidents (Hossain et al., 2020).

Public and organizations have been encouraged to take advantage of any possible business opportunities by utilizing and adopting new-technologies (Saraiva et al., 2022), there is a huge misunderstanding of their criminal activities from the management perspective (Papathanasiou et al., 2025). Underestimation of security threats leads to an increase in their vulnerabilities and risks such as (fraud, human trafficking, narcotics drugs deals, burglary, robbery and so forth), which unfortunately can become actual challenges to them (Papathanasiou et al., 2025). Different authors aim at revealing the role of ML in addressing crime activities in present years (Saraiva et al., 2022), as found in the literature review, and to suggest opportunities for further research (Saraiva et al., 2022). This study would conduct a perspective that plays a key role of ML in crime prediction, which are current threats faced by public and businesses operators. Importantly, empirical research on crime predictions is needed.

ML and DL techniques are being widely adopted in crime activities (Saraiva et al., 2022), which are threats, behaviors, practices, awareness, and decision-making respectively (Saraiva et al., 2022). These techniques build prediction models using historical data and have applications across various domains (Gong et al., 2020). However, developing model that are accurate, efficient and interpretable remains challenging due to multifaceted influences (Saraiva et al., 2022). Various methods, including statistical models, prevention models, detection models, and prediction models, have been employed in this area (Saraiva et al., 2022).

This research explores current developments in ML and DL for crime prediction and discusses how these innovative technologies are being used to detect criminal activities, predict crime patterns from the available historical records, and prevent future crime. The primary goal of this research is to provide a comprehensive overview of recent developments in this field to provide the easiest way of handling the rise of crime activities and contribute to future research efforts.

Several models have been proposed, tested and evaluated by researchers (Mandalapu et al., 2023). However, despite progress, current literature lacks a clear evaluation of which architecture suitable per given prediction tasks (Alves et al., 2018; Araujo et al., 2019; Da Silva et al., 2020; Han et al., 2020; Ingilevich & Ivanov, 2018; Sathiyanarayanan et al., 2019) . Furthermore, many real-world applications operate under limited data (Algefes et al., 2022; Das et al., 2021; Nakib et al., 2018; Zhang et al., 2022) making it necessary to explore data-efficient training strategies that can enhance model performance without relying on large-scale datasets. These gaps make it difficult for practitioners and researchers to make informed decisions when selecting models for real-world applications. The aim of this research is to design, experiment, evaluate and compare the performance of four (4) machine learning algorithms (LR, SVM, DT and RF) to identify the best among them in terms of performance using small sample datasets. The primary aim of this research was achieved through the following objectives:

i. To implement innovative training methods designed to enhance model performance without requiring large datasets due to the lack of data.
ii. To experiment and evaluate the performance of these ML algorithms to improve generalization in limited data setups using appropriate metrics like accuracy, precision, recall, and F1-score for classification and regression tasks.
iii. To compare the performance of LR, SVM, DT and RFC based on the same sample datasets.

Review of Related Works

A novel method is proposed in the research study (Azhari & Utomo, 2018), the authors develop an application that can predict motorcycle robbery using ARIMAX – TFM with a single input with a technique to consider outside consequences. The accuracy of this technique is measured using metrics Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE), and the scores are 32.30 and 6.68.

The author in (Saraiva et al., 2022), provides a technique primarily based on auto-regressive fashions and spatial analysis to spontaneously locate excessive-hazard crime areas in city areas and reliably help in forecasting crime tendencies in each area, the experiment is use to detect crime using ARIMA , random forest, RepTree and ZeroR techniques. These experiments are performed on datasets gathered in New York City and Chicago.

Another study by (Ingilevich & Ivanov, 2018), compared multiple machine learning techniques to predict the criminal activities in various areas of a metropolis. This research explored three predictive models: logistic regression, linear regression and gradient boosting algorithms. The authors also select essential predictors by utilizing feature selection techniques. By applying these selection methods, there is an improved accuracy score, and it helped to avoid models from overfitting. After comparing the results of all three models, the authors found that the gradient boosting algorithm outperformed proven to be the best method to predict the crime rate in the urban area.

In another study (Da Silva et al., 2020), authors have looked at Brazil crimes as case, which have increased rapidly in recent times. Numerous predictive models or solutions use intelligent systems to identify when and where will a criminal offense will arise or happen next, which lets police send to those areas that are in danger for security measures. As part of the research, the authors investigated four machine-learning algorithms for identifying where a criminal offense will arise in Fortaleza, Brazil. Their results indicate that easy simpler algorithms are efficient in predicting crime. Also, they have seen that the decision tree and bagging regressor strategies obtained quality prediction outcomes than extra tree and random forest.

As mentioned above, numerous linear models are there to predict crime through the correlations between urban crime and metrics. However, due to non-Gaussian distributions and multi collinearity in urban attributes, we usually tend to make controversial conclusions on these attributes to predict crime. Ensemble-based machine learning algorithms can deal with such problems adequately. In the research work (Alves et al., 2018), authors applied RF (random forest regressor) to predict the crime and quantify the impact of urban attributes on homicides. Their approach has 88% accuracy in crime prediction, and the significance of city indicators is clustered and ranked equally.

In many research works (Araujo et al., 2019), authors have proven that machine learning models effectively predict crimes. In this research, the authors use visual analysis techniques to predict crimes using London crime dataset. The authors also evaluate the performance of this model using accuracy with a score of 87% still, could be more efficient in identifying which variables are significant in predicting crimes.

The research done by (Sathiyanarayanan et al., 2019), further examined the classification techniques KNN to predict different criminal incidents like analyzing the criminal reports, the authors uses crime endogenous data sources to

provide insights into existing public safety systems and platforms, the model achieved accuracy of 89%.
In the similar research done by (Elluri et al., 2019), examined the uses of classification techniques XGBOOST, RF and SVM to predict crime in neighborhoods using spatial sata analysis, the authors uses crime New York crime data to provide which achieved accuracy of 52%.
The research done in (Das et al., 2021) , authors have proven that machine learning models effectively predict crimes. The researchers applying particle swarm optimization-based classifier and a rule engine to classify crimes reports using USA and Indian crime dataset, the authors also evaluate the performance of this model using accuracy which achieved a score of 79% still, could be more efficient in identifying which variables are significant in predicting crimes.
In the study (Zhang et al., 2022), the authors have selected seventeen variables for prediction of crime, and the XGBoost algorithm is incorporated to train the model. A Shapley additive explanation (SHAP) and post hoc interpretable approach is used to parent the contribution of person variables. SHAP, a post hoc interpretable method, is used to determine the significance of individual variables. Among all these variables used in the research, the percentage of the non-neighborhood population and the populace aged from 25 to 44 contribute greater than different variables in predicting crime. The authors have also validated the SHAP values to demonstrate each variable's contribution to the crime prediction experimental findings; the model achieved an accuracy of 89% and AUC peak at 58%.
Similar study in (Algefes et al., 2022), the authors uses a combination of ML and DL models (CNN, KNN, SVM, DT, RF and NK) for focusing of crime on social media data, ANN are trained on crime-related text data to predict the likelihood of a crime occurring. These model uses Saudi Arabia tweets (2017 to 2021) to analyze the context and tone of the text data to classify patterns that may indicate criminal activity the model achieved an accuracy of 79%.

MATERIALS AND METHODS

The research followed an experimental approach**,** where these ML algorithms are implemented and evaluated based on their performance. These involve the following stages: Dataset Collection and Preparation**:** This stage involves cleaning, encoding, and scaling the data. Model Implementation and Evaluation**:** the ML algorithms (LR, SVM, DT and RF) are implemented each individually. Hyperparameters specific to these models are either tuned or left at default values to ensure optimal performance without overfitting. The models are evaluated based on their application in crime prediction. This is done using metrics such as Accuracy, F1-score, recall and precision. Additionally, a confusion matrix is used for a binary classification of performance.

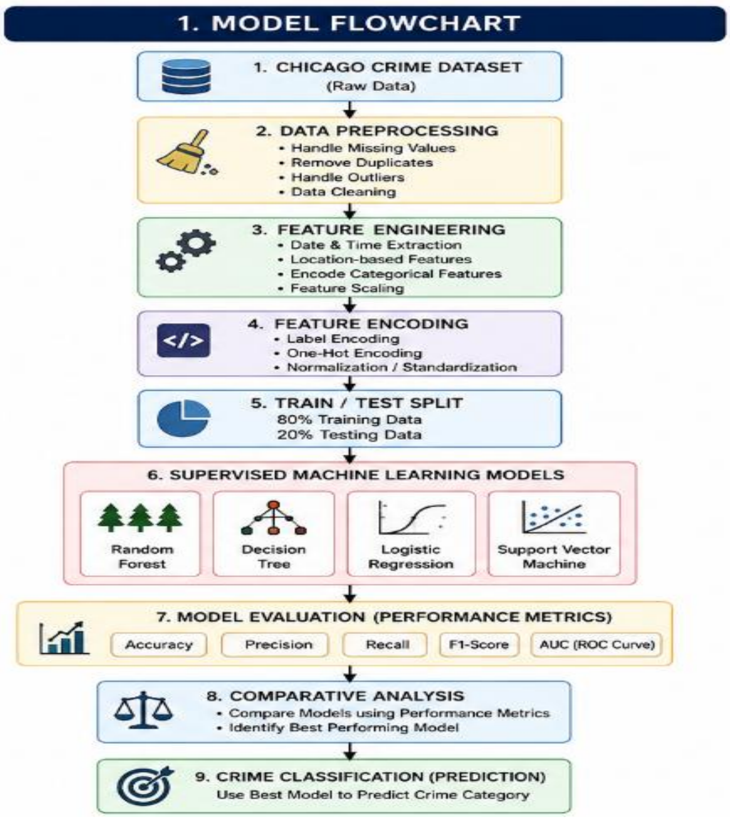

Figure 1: Model Flowchart

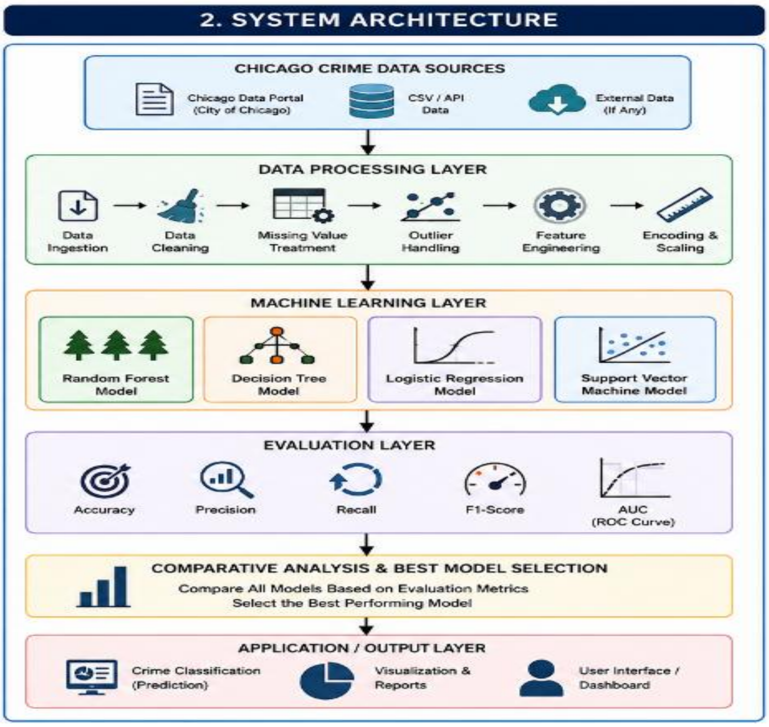

Figure 2: System Architecture

Materials and Tools
This section describes the key tools and technologies employed in implementing and evaluating the ML models for crime prediction. The choice of tools ensures efficient handling of data preprocessing, models training, and performance evaluation.

Table 1: Experimental Parameters

Parameter	Experimental Setup
Sample Size	3,000 sample records
CPU specification	CORE i7 with 8GB RAM, SSD 256, Windows 11 OS
Number of runs	multiple to account for variance
Bach size	32
Learning rate	LR: 1e-3 to le-4;
Loss function	categorical cross-entropy
Hyperparameter tuning method	Random
Train/validation/test splits	80%, 10%, 10%
Libraries	Numpy, pandas, scikit-learn, matplotlib and seaborn
Evaluation matrix	Accuracy, precision, recall, f1-score and AUC
Training Time	200 to 300 sec

Chicago Crime Dataset
The experiments used the publicly available Chicago Crime Dataset, which was accessed through Kaggle (https://www.kaggle.com/datasets/Chicago_crime_dataset.csv) the dataset consists of feature from thirty one (31) categories of samples crime, the three categories were extracted and used for the experiment consist of narcotics, burglary and robbery crime, 1,000 samples for each class was extracted and used totaling 3,000 sample records. This extracted dataset contains 9 features including records such as Case Number, Crime ID, Date, Primary Type, Description, Location Description, Arrest, Domestic and Year. Label encoding was applied to ensure that the dataset was in the correct format for the model and that features are standardized for models sensitive to feature scaling as shown in table 2 below.

Data preprocessing
The Chicago crime dataset underwent several preprocessing steps to improve data quality and ensure compatibility with the machine learning algorithms. Missing values, duplicate records, and irrelevant attributes were removed during the data-cleaning stage. Date and time attributes were transformed into meaningful features, such as year, month, day, and hour, to capture temporal crime patterns.
Categorical variables, including crime type, location description, district, ward, beat, arrest status, and domestic status, were transformed into numerical representations using label encoding and one-hot encoding techniques to ensure compatibility with the machine learning models. The unique crime identification number was treated as an identifier and excluded from the predictive process because it does not contribute to crime classification.
The independent variables (X) used in this study include case number, primary type, location description, date, description, arrest, domestic and year.
The dependent variable (y) is Crime category (crime classification label)
To ensure optimal model performance, feature scaling was applied to the numerical attributes using the Standard Scaler

technique. Standardization transformed the features to have a mean of zero and a standard deviation of one, which is particularly important for algorithms such as Logistic Regression and Support Vector Machine. Finally, a comparative analysis was performed to determine the most effective model for crime classification based on the evaluation metrics.

Table 2: Data Processing Result

Feature	Data Type	Encoding	Scaling
Case Number	String	Label encoded	Not Scaled
Primary Type	String	Label encoded	Standardized
Location Description	String	Label encoded	Standardized
Date	Datetime	Converted to numeric	Standardized
Description	String	Label encoded	Standardized
Arrest	Boolean	Label encoded	Standardized
Domestic	Boolean	Label encoded	Standardized
Year	Year	Converted to numeric	Scaled

Model development

Logistic Regression is one of the most used and simplest classification algorithms, assuming a linear relationship between the input features and the target variable. The model was also implemented using scikit-learn**.** It assumes that the target variable can be explained as a linear combination of the input features. The model was trained using the training set and then evaluated using the testing set. The model is easy to implement and interpret. It works well when the relationship between features and the target is approximately linear. The model may not perform well if the dataset contains non-linear relationships, high multicollinearity, or outliers.

Support Vector Machine is a powerful classification method that attempts to fit the data within a margin of tolerance, known as the epsilon-insensitive zone, while minimizing prediction errors. The SVM model is particularly useful for datasets with non-linear relationships between features. The model was implemented using the scikit-learn library. It was trained on the preprocessed Chicago crime dataset with optimized hyperparameters, including the kernel type (linear, radial basis function, etc.). SVM is effective for datasets with non-linear patterns and is resistant to overfitting, especially when the dataset has noise. SVM can be computationally expensive, particularly with large datasets, and requires careful tuning of hyperparameters to achieve optimal results.

Random Forest is an ensemble method that creates multiple decision trees and combines their outputs to improve prediction accuracy and generalization. The RF model was implemented using scikit-learn**,** with 100 decision trees (n_estimators) in the ensemble. It was trained on the Chicago crime dataset and evaluated on unseen test data. Random Forest is robust against overfitting, handles both linear and non-linear relationships well, and performs well with large datasets and high-dimensional data. Forest can be less interpretable than simpler models like Linear Regression and requires tuning several hyperparameters, such as the number of trees and the depth of the trees.

Decision Tree is a fundamental machine learning algorithm, is a supervised learning technique that splits data into branches based on decision rules derived from features, ultimately forming a tree-like structure that leads to predictions. The decision tree model was implemented using scikit-learn**,** with 10 decision trees (n_estimators) in the ensemble. It was trained on the Chicago crime dataset and evaluated on unseen test data.

Evaluation Metrics

Accuracy is a fundamental metric for evaluating the overall performance of ML and DL models. In this dataset, accuracy measures the proportion of correctly predicted instances out of the total instances from giving samples. A high accuracy score indicates that the model is effective in making correct predictions across both classes, thereby providing a reliable measure of its overall performance (Farhadpour et al., 2024).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

Recall, also referred to as sensitivity or the true positive rate, measures the ability of a detection model to correctly identify all actual dropout instances within the dataset. It is defined as the ratio of predicted dropout cases to the total number of actual dropout cases. A high recall score indicates that the model effectively detects most dropout instances, minimizing the occurrence of false negatives (Maxwell et al., 2021)

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

Precision is an essential evaluation metric used to measure the effectiveness of a detection model in accurately identifying true positive cases. It is defined as the ratio of correctly predicted positive instances to the total number of instances predicted as positive by the model. A high precision score indicates a low rate of false positives, meaning that non-dropout samples are rarely misclassified as dropout, thereby enhancing the reliability of the model (Farhadpour et al., 2024).

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

The F1-score is a composite evaluation metric that balances both precision and recall, providing a more comprehensive measure of a dropout detection model's performance. It is defined as the harmonic means of precision and recall, ensuring that both false positives and false negatives are considered. This metric is particularly useful when there is a need to achieve a balance between accurately identifying dropout instances and minimizing missed detections (Farhadpour et al., 2024).

$$F1-Score = \frac{2\times TP}{2\times TP+FP+FN} \quad (4)$$

RESULTS AND DISCUSSION

This section discusses the results obtained from the experiment conducted using SVM, LR, DT and RF models. The goal of the section is to implement these models with better performance and to evaluate them based on key metrics, such as accuracy, precision, recall, f1-score and ROC-AUC. These metrics provide insight into how well each model fits the data and how they differ in their prediction abilities. Visual representations are included in this section to aid in understanding the results.

Logistic Regression (LR) Model Results

The model achieved the following results on the test set: after model evaluation.

Table 3: Metrics Result for LR

Metric	Value
Accuracy	90.03%
Precision	92.14%
Recall	90.33%
F1-Score	90.20%
AUC	98.65%

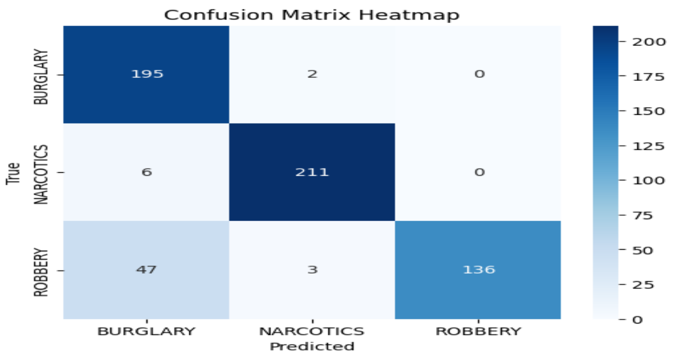

Figure 3: Confusion Matrix Classification Report for LR

Support Vector Machine (SVM) Model Results
The model achieved the following results on the test set: after model evaluation.

Table 4: Metrics Result for SVM

Metric	Value
Accuracy	90.16%
Precision	91.95%
Recall	90.16%
F1-Score	90.04%
AUC	98.70%

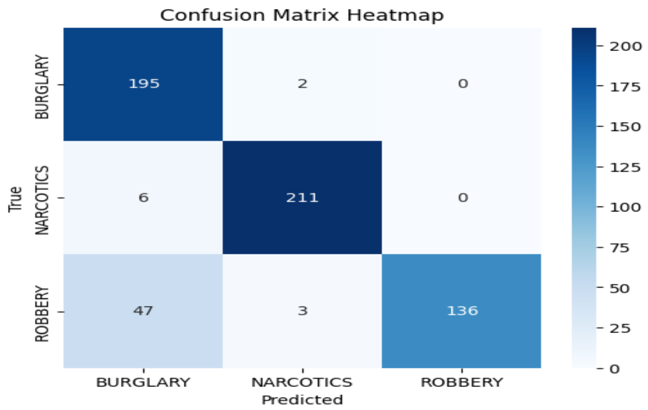

Figure 4: Confusion Matrix Classification Report for SVM

Decision Tree (DT) Model Results
The model achieved the following results on the test set: after model evaluation.

Table 5: Metrics Result for DT

Metric	Value
Accuracy	90.33%
Precision	92.14%
Recall	90.33%
F1-Score	90.20%
AUC	98.63%

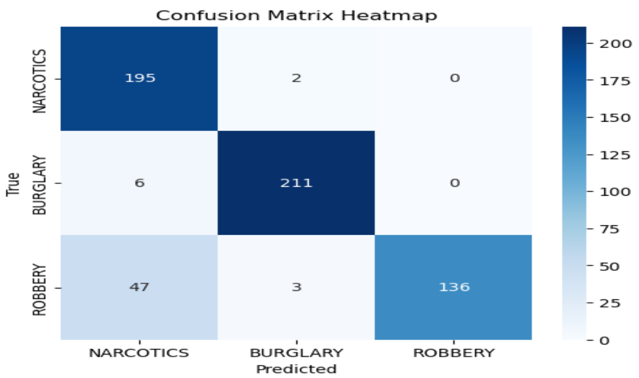

Figure 5: Confusion Matrix Classification Report for DT

Random Forest (RF) Model Results

The model achieved the following results on the test set: after model evaluation.

Table 6: Metrics Result for RF

Metric	Value
Accuracy	90.33%
Precision	92.14%
Recall	90.33%
F1-Score	90.20%
AUC	98.72%

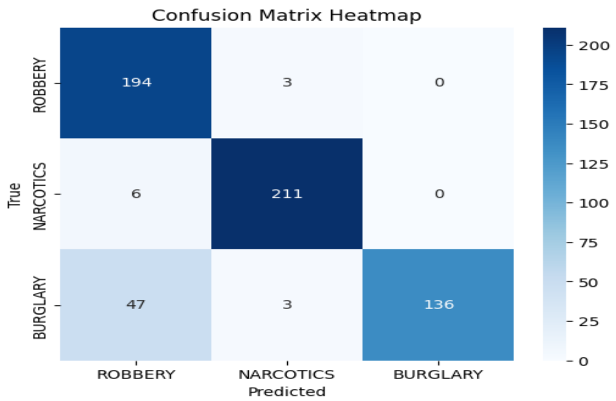

Figure 6: Confusion Matrix Classification Report for RF

DT and RF achieved higher accuracy in training and testing and maintained smoother accuracy. LR, on the other hand, achieved higher accuracy lower than DT and RF, possibly due to limitations in learning complex patterns while the SVM peak at 90.16% which has the lowest prediction accuracy; hence DT is considered the best Machine learning models for this research.

Models Comparison Results

Table 7: Metrics Result for ML Models

Metric	DT	SVM	RF	LR
Accuracy	90.33%	90.16%	90.33%	90.03%
Precision	92.14%	91.95%	92.14%	92.14%
Recall	90.33%	90.16%	90.33%	90.33%
F1-Score	90.20%	90.04%	90.20%	90.20%
AUC	98.63%	98.70%	98.72%	98.65%

Discussion

The results demonstrate that the proposed DT-based model significantly outperforms the benchmark system. The accuracy improvement from 79% to 90.33% indicates a robust enhancement in prediction capability. Furthermore, the inclusion of precision, recall, F1-score and AUC metrics confirm balanced performance in correctly identifying different crime categories without generating excessive false positives. In addition, the use of dynamic behavioral features allows the system to predict novel and polymorphic crime variants, overcoming the primary weakness of prediction capabilities. The ensemble nature of the decision tree further minimizes overfitting and enables faster decision-making through parallel processing.

Comparison Report between Benchmark System and Proposed System

The benchmark system, based on the work of (Algefes et al., 2022), employed a predict crime likelihood from social media. While this method achieved a detection accuracy of 79%, it exhibited notable limitations, including accuracy moderate; social media bias possible, a high rate of false positives and negatives, and lack of real-time crime prediction capabilities. The authors use a combination of ML and DL models (CNN, KNN, SVM, DT and RF) for focusing crime on social media data, ANN are trained on crime-related text data to predict the likelihood of a crime occurring. These model uses Saudi Arabia tweets (2017 to 2021) to analyze the context and tone of the text data to classify patterns that may indicate criminal activity. These limitations reduced the system's reliability and operational effectiveness, as summarized in Table 3.

Table 8: Limitations of the Benchmark System and DT-Based Solutions

Limitation	Impact	DT-Based Solution
Accuracy moderate; social media bias possible	Misses' new crime variants	Uses dynamic behavioral features
High false positives	Unreliable alerts	Ensemble learning reduces overfitting
Slow prediction speed	Delays response to prediction	Parallel decision trees for faster classification

CONCLUSION

This research presented the experimental evaluation and performance comparison of the ML models applied in this study for crime prediction analysis. The results demonstrated that tree-based ensemble methods (DT and RF) performed better overall compared to simpler models when applied to the dataset used. Specifically, the DT and RF models showed strong predictive capability, with DT achieving the most stable and reliable performance due to its ability to reduce overfitting through ensemble learning. RF also performed well and provided interpretability advantages, making it useful for understanding decision paths in the dataset. The findings indicate that for datasets with relatively structured and linear relationships, traditional ML models such as DT and RF are effective for classification tasks. However, the performance differences between models suggest that ensemble techniques offer improved robustness and generalization.

Overall, this study achieves its aim of evaluating and comparing ML models for crime-related data analysis, rather than relying on deep learning or sequence-based approaches. The results provide insight into model selection for similar structured datasets and highlight the effectiveness of classical supervised learning methods in this context. Limitations and future work, a few caveats are important. First, the dataset covers Chicago only; results may not generalize to other cities without retraining. Second, historical crime data can be biased (e.g. under-reporting in certain neighborhoods), which may limit model fairness. We did not explicitly address bias or ethical issues here, but these should be considered before deployment. Third, we used only basic features; richer data (precise location coordinates, victim information, weather, events) might improve accuracy further. Future research could integrate spatio-temporal models (e.g. deep-learning approaches) or test on real-time forecasting. Finally, a deeper hyperparameter search and cross-validation could refine model performance.

REFERENCES

Algefes, A., Aldossari, N., Masmoudi, F., & Kariri, E. (2022). A Text-mining approach for crime tweets in Saudi Arabia: From analysis to prediction. Proceedings - 2022 7th International Conference on Data Science and Machine Learning Applications, CDMA 2022, 109–114. https://doi.org/10.1109/CDMA54072.2022.00023

Alves, L. G. A., Ribeiro, H. V., & Rodrigues, F. A. (2018). Crime prediction through urban metrics and statistical learning. Physica A: Statistical Mechanics and Its Applications, 505, 435–443. https://doi.org/10.1016/j.physa.2018.03.084

Araujo, A., Cacho, N., Bezerra, L., Vieira, C., & Borges, J. (2019). Towards a Crime Hotspot Detection Framework for Patrol Planning. Proceedings - 20th International Conference on High Performance Computing and Communications, 16th International Conference on Smart City and 4th International Conference on Data Science and Systems, HPCC/SmartCity/DSS 2018, 1256–1263. https://doi.org/10.1109/HPCC/SmartCity/DSS.2018.00211

Azhari, & Utomo, P. E. P. (2018). Prediction the crime motorcycles of theft using ARIMAX-TFM with single input. Proceedings of the 3rd International Conference on Informatics and Computing, ICIC 2018, 1–7. https://doi.org/10.1109/IAC.2018.8780520

Bharadiya, J. (2023). Machine Learning in Cybersecurity: Techniques and Challenges. European Journal of Technology, 7(2), 1–14. https://doi.org/10.47672/ejt.1486

Chapman, A., Grylls, P., Ugwudike, P., Gammack, D., & Ayling, J. (2022). A Data-driven analysis of the interplay between Criminological theory and predictive policing algorithms. ACM International Conference Proceeding Series, 36–45. https://doi.org/10.1145/3531146.3533071

Chun, S. A., Pathak, R., Paturu, V. A., Atluri, V., Yuan, S., & Adam, N. R. (2019). Crime Prediction Model using Deep Neural Networks. ACM International Conference Proceeding Series, 512–514. https://doi.org/10.1145/3325112.3328221

Da Silva, A. R. C., De Paula Junior, I. C., Da Silva, T. L. C., De Macedo, J. A. F., & Silva, W. C. P. (2020). Prediction of crime location in a brazilian city using regression techniques. Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI, 2020-Novem, 331–336. https://doi.org/10.1109/ICTAI50040.2020.00059

Das, P., Das, A. K., Nayak, J., Pelusi, D., & Ding, W. (2021). Incremental classifier in crime prediction using bi-objective Particle Swarm Optimization. Information Sciences, 562, 279–303. https://doi.org/10.1016/j.ins.2021.02.002

Elluri, L., Mandalapu, V., & Roy, N. (2019). Developing machine learning based predictive models for smart policing. Proceedings - 2019 IEEE International Conference on Smart Computing, SMARTCOMP 2019, 198–204. https://doi.org/10.1109/SMARTCOMP.2019.00053

Gong, J., Zhang, H., & Du, W. (2020). Research on integrated learning fraud detection method based on combination classifier fusion (thbagging): A case study on the foundational medical insurance dataset. Electronics (Switzerland), 9(6). https://doi.org/10.3390/electronics9060894

Han, X., Hu, X., Wu, H., Shen, B., & Wu, J. (2020). Risk Prediction of Theft Crimes in Urban Communities: An Integrated Model of LSTM and ST-GCN. IEEE Access, 8, 217222–217230. https://doi.org/10.1109/ACCESS.2020.3041924

Hossain, S., Abtahee, A., Kashem, I., Hoque, M. M., & Sarker, I. H. (2020). Crime prediction using spatio-temporal data. In Communications in Computer and Information Science: 1235 CCIS. Springer Singapore. https://doi.org/10.1007/978-981-15-6648-6_22

Ingilevich, V., & Ivanov, S. (2018). Crime rate prediction in the urban environment using social factors. Procedia Computer Science, 136, 472–478. https://doi.org/10.1016/j.procs.2018.08.261

Kapadiya, K., & Patel, U. (2022). Blockchain and AI-Empowered Healthcare Insurance Fraud Detection : An Analysis, Architecture , and Future Prospects. IEEE Access, 10(August), 79606–79627. https://doi.org/10.1109/ACCESS.2022.3194569

Kim, S., Joshi, P., Kalsi, P. S., & Taheri, P. (2018). Crime Analysis Through Machine Learning. 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2018, (January), 415–420. https://doi.org/10.1109/IEMCON.2018.8614828

Kshatri, S. S., Singh, D., Narain, B., Bhatia, S., Quasim, M. T., & Sinha, G. R. (2021). An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach. IEEE Access, 9, 67488–67500. https://doi.org/10.1109/ACCESS.2021.3075140

Mandalapu, V., Elluri, L., Vyas, P., & Roy, N. (2023). Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions. IEEE Access, 11(April), 60153–60170. https://doi.org/10.1109/ACCESS.2023.3286344

Meijer, A., & Wessels, M. (2019). Predictive Policing: Review of Benefits and Drawbacks. International Journal of Public Administration, 42(12), 1031–1039. https://doi.org/10.1080/01900692.2019.1575664

Nakib, M., Khan, R. T., Hasan, M. S., & Uddin, J. (2018). Crime Scene Prediction by Detecting Threatening Objects Using Convolutional Neural Network. International Conference on Computer, Communication, Chemical, Material and Electronic Engineering, IC4ME2 2018, 1–4. https://doi.org/10.1109/IC4ME2.2018.8465583

Neil, S., Nandish, B., & Manan, S. (2013). Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. Visual Computing for Industry, Biomedicine, and Art, 7(9), 1–14.

Opeyemi Babatunde, G., Damilola Mustapha, S., Ike, C. C., Alabi, A. A., & Babatunde, G. O. (2025). A holistic cyber risk assessment model to identify and mitigate threats in us and canadian enterprises. Article in International Journal of Multidisciplinary Research and Growth Evaluation, 773–787. www.allmultidisciplinaryjournal.com

Papathanasiou, A., Liontos, G., Katsouras, A., Liagkou, V., & Glavas, E. (2025). Cybersecurity Guide for SMEs: Protecting Small and Medium-Sized Enterprises in the Digital Era. Journal of Information Security, 16(01), 1–43. https://doi.org/10.4236/jis.2025.161001

Raza, D. M., & Victor, D. B. (2021). Crime Using Random Forest. Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS-2021), 7, 980–987.

Safat, W., Asghar, S., & Gillani, S. A. (2021). Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques. IEEE Access, 9, 70080–70094. https://doi.org/10.1109/ACCESS.2021.3078117

Saraiva, M., Matijošaitienė, I., Mishra, S., & Amante, A. (2022). Crime Prediction and Monitoring in Porto, Portugal, Using Machine Learning, Spatial and Text Analytics. ISPRS International Journal of Geo-Information, 11(7). https://doi.org/10.3390/ijgi11070400

Sathiyanarayanan, M., Junejo, A. K., & Fadahunsi, O. (2019). Visual Analysis of Predictive Policing to Improve Crime Investigation. Proceedings of the 4th International Conference on Contemporary Computing and Informatics, IC3I 2019, 197–203. https://doi.org/10.1109/IC3I46837.2019.9055515

Sharma, S., & Chaudhary, P. (2023). Machine learning and deep learning. Quantum Computing and Artificial Intelligence: Training Machine and Deep Learning Algorithms on Quantum Computers, 71–84. https://doi.org/10.1515/9783110791402-004

Trinhammer, M. L., Merrild, A. C. H., Lotz, J. F., & Makransky, G. (2022). Predicting crime during or after psychiatric care: Evaluating machine learning for risk assessment using the Danish patient registries. Journal of Psychiatric Research, 152(June), 194–200. https://doi.org/10.1016/j.jpsychires.2022.06.009

Tsiodra, M., Panda, S., Chronopoulos, M., & Panaousis, E. (2023). Cyber Risk Assessment and Optimization: A Small Business Case Study. IEEE Access, 11(April), 44467–44481. https://doi.org/10.1109/ACCESS.2023.3272670

Zhang, X., Liu, L., Lan, M., Song, G., Xiao, L., & Chen, J. (2022). Interpretable machine learning models for crime prediction. Computers, Environment and Urban Systems, 94(November 2021), 101789. https://doi.org/10.1016/j.compenvurbsys.2022.101789

©2026 This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license viewed via https://creativecommons.org/licenses/by/4.0/ which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited appropriately.