The Indonesian Journal of Computer Science
www.ijcs.net
Volume 14, Issue 3, June 2025
https://doi.org/10.33022/ijcs.v14i3.4881

# Systematic Literature Review on Crime Prediction using Machine Learning Techniques

**Omobayo Ayokunle Esan[1], Bester Chimbo[2]**
oesan@wsu.ac.za[1], chimbb@unisa.ac.za[2]
[1]Depatment of Computer Science, University of South Africa
[2]Department of Information Technology, University of South Africa

**Article Information**

Received : 15 May 2025
Revised : 25 May 2025
Accepted : 9 Jun 2025

**Keywords**

Crime, prediction, machine learning, systematic literature review, preferred reporting item for systematic reviews

**Abstract**

Abstract contains problem statement, approaches/problem solving method, objectives and resulTo lower the crime rate in the community, many governments around the world have made preventive security measures their top priority. Thus, a major and extensively studied field is the use of machine learning in crime prediction. To investigate crime prediction using machine learning approaches, this study carried out a systematic literature review. The review assesses performance evaluation criteria, forecast methods, present issues, and potential future directions. From 2018 to 2024, a total of 100 research papers covering machine learning techniques for crime prediction were reviewed. The supervised learning approach is the most often used crime prediction technology, according to the review. The evaluation and performance criteria, the tools used to construct the models, and the difficulties they face in predicting crime were also covered. Machine learning approaches for crime prediction are an interesting area of research, and academics have used a number of machine learning models. Future studies should explore integrating different data sources, such as social media activity, economic indicators, and urban mobility data, to enrich the predictive models. Integration of two or more machine learning models can improve the predictive accuracy.

## A. Introduction

The swift development of technology, particularly in the areas of data science and machine learning, has created new opportunities to tackle intricate societal issues, such as law enforcement and crime prevention [1]. Using algorithms and analytical methods to forecast possible criminal activity, suspects, and crime hotspot regions is one such promising use, known as crime link prediction [2]. This strategy seeks to help law enforcement and security personnel identify crime hotspots, discover patterns in criminal activity, uncover the covert connections among offenders, and make well-informed judgements to prevent crime.

As crime data becomes more accessible and modern technologies advance, researchers have an excellent opportunity to study and analyze machine learning-based crime prediction [3]. Data is analyzed and predicted by machine learning, a subfield of artificial intelligence, using statistical models and algorithms [4]. Machine learning approaches have a wide range of potential applications in tackling the problem of crime prediction [5]. Crime prediction has employed machine learning algorithms to evaluate crime data and predict future crime trends [6]. To precisely forecast crime patterns, for instance, algorithms such as Random Forest (RF), Support Vector Machine (SVM), and Decision Trees (DT) have been trained on crime data from specific cities [6].

Applications of machine learning models include predictive police systems, where programs like PredPol use past crime data to forecast the locations and times of incidents. By concentrating on particular regions, police enforcement organizations can more efficiently distribute patrols, possibly averting crimes before they occur. These models frequently use algorithms such as decision trees and logistic regression to examine temporal and spatial trends of crime.

Also, in surveillance and threat detection to detect suspicious activities. For example, Convolutional Neural Networks (CNNs) for analyze live feeds to identify unusual movements or behaviors, alerting security personnel in real-time [7]. These systems are employed in public spaces like airports and train stations to enhance situational awareness. Furthermore, in crime link analysis for identification of hidden relationships between criminals, events, and locations. Graph-based models and clustering techniques can uncover networks of organized crime, fraud rings, or human trafficking operations, providing valuable insights to law enforcement agencies [8].

Another application areas of machine learning are smart city safety where machine learning is integrated with Internet of Things (IoT) devices to enhance urban safety [9]. For instance, sensors and ML algorithms monitor traffic flows, detect accidents, and analyze noise levels to identify disturbances that may indicate criminal activities. These systems provide city administrators with actionable intelligence to respond proactively. These applications demonstrate the potential of machine learning to enhance public safety and optimize resource allocation.

In addition to predicting crime patterns, these algorithms can offer insightful information about patterns and trends in crime, helping security guards allocate resources and strategies to effectively combat criminal behavior [10]. Identification of crime hotspots and future crime prediction can be achieved by applying machine learning to analyze crime data from a particular geographic area, such as a city or neighborhood [11]. Law enforcement activities can therefore be made more effective by utilizing this information to focus policing resources where they are most required [12]. These algorithms are designed to anticipate crime patterns in specific areas by analyzing crime data that has either a temporal or spatial component.

Several cities and countries have successfully implemented machine learning techniques to predict crime patterns and enhance law enforcement strategies. Notable examples include: Chicago, USA, Porto, Portugal [13], Chicago, United State of America (USA) [14], Brazil, Los Angeles, USA [15], etc. These implementations demonstrate the potential of machine learning in enhancing public safety through proactive crime prediction and resource management.

Before machine learning can be used to forecast crime, a number of challenges must be resolved. Finding reliable crime statistics is one of the difficulties [16]. It is challenging to get crime data, and what is available may need to be accurate or comprehensive [17]. Concerns of privacy and ethics are also linked to the collection of crime data. To effectively use machine learning for crime prediction, these issues need to be resolved. The interpretability of machine learning models is another difficulty. These models can be difficult to interpret and comprehend, which limits their value in the decision-making process [18].

This study conducted a systematic review of crime prediction to explore the existing body of research and methodologies applied in crime prediction domain. It seeks to identify trends, challenges, and comparing different predictive techniques, and evaluating their effectiveness. This study makes the following research contributions based on the aforementioned background:

- Comprehensive understanding of techniques: Identifies and categorizes various machine learning techniques (e.g., random forest, Naïve Bayes, support vector machines) applied in crime prediction.
- Evaluation of data sources: Analyzes commonly used datasets (e.g., public crime records, social media data).
- Identification of gaps and challenges: Explores challenges such as data privacy, bias in datasets, real-time applicability, and the interpretability of predictions.

The following is the structure of the remaining portions of this paper: Section II outlines the methodology, while Section III discusses the findings and conclusions. Please find the closing remarks in Section IV.

## B. Research Method

The study gathered relevant literature in order to investigate using the SLR methodology. Guidelines from the Preferred Reporting Items for Systematic Reviews (PRISMA) were used for the SLR. It is stated that PRISMA is "a minimal set of items for reporting in systematic reviews and meta-analyses that are based on evidence" [19]. This evaluation consists of the following four processes: inclusion, eligibility, screening, and identification. Figure 1 shows the PRISMA diagram for this study's literature review.

**Figure 1.** The chosen literature search and systematic review procedures are shown in the PRISMA flowchart

## Included Stage

The investigation of 100 papers used in this SLR and the distribution of studies by sources libraries is shown in Table 1.

**Table 1.** Distribution of studies by source libraries

| Source | Query Request | *Title and Abstract included* | Full Articles Reviewed | Full Articles Included |
|---|---|---|---|---|
| ACM | 100 | 55 | 22 | 13 |
| IEEE | 215 | 68 | 35 | 33 |
| WoS | 405 | 82 | 52 | 38 |
| SD | 265 | 89 | 38 | 16 |
| Total | n = 985 | n = 294 | n = 147 | n = 100 |

Table 1 presents a breakdown of the evaluated and incorporated papers for each review stage, organized based on the digital library they are affiliated with. To the following stages were taken into considerations: 1) the planning stage, 2) the review questions, 3) the search strategy, and 4) the selection criteria.

**Identification Stage**

At this stage, 300 studies were collected from the IEEE, Springer, ACM Digital Library, and Scopus databases. Duplicate articles were thrown out by the libraries. During this review, 80 research in total were disqualified. One hundred papers were chosen for full-text examination as a consequence of the screening procedure.

**Screening and Eligibility Stage**

The 220 publications' titles, abstracts, and keywords were examined as part of the screening procedure. To eliminate research published before 2018, those unrelated to machine learning-based crime prediction, and those not published or written in English, exclusion criteria were applied. 50 studies in all were eliminated at this point. An effort to retrieve 805 research from the four digital libraries using this query yielded 34 results, and 100 retrieved publications were evaluated for eligibility.

**Step Involved in the Systematic Literature Review**

The steps that are involved in conducting the systematic literature review are discussed in following sections.

**Planning Stage**

Planning is the first step in the systematic literature review (SLR), and it is an important step since it serves as the foundation for the entire process. Fig. 2 shows how many publications were published overall on crime prediction between 2018 and 2024.

Machine learning-based crime prediction has rapidly expanded, as shown in Figure 2, with more publications being published between 2018 and 2024. There were 154 papers published in 2019 about the application of machine learning algorithms for crime prediction, compared to 120 articles in 2018. In 2024, there were 530 relevant articles published, compared to 264, 456, 495, and 500 publications in 2020, 2021, 2020, and 2023.

**Figure 2.** Number of crime prediction using machine learning of paper publications from 2018 to 2024

**Reviews Research Question Stage**

Investigating machine learning-based crime prediction research is the goal of the current SLR. The SLR defines the study's scope in this phase by referring to the systematic literature review's objective, which is to investigate and assess machine learning-based crime prediction, after the planning phase has decided on the timeframe and databases. Five primary research questions are addressed to produce the research contributions.

**Research Question 1:** What are the predominant techniques used for crime prediction?

**Research Question 2:** What are the performance evaluation metrics associated with current crime prediction models?

**Research Question 3:** What is the predominant crime predictive implementation tools?

**Research Question 4:** What is the predominant crime predictive datasets?

**Research Question 5:** What is the crime predictive machine learning challenges?

**Search Stage**

The primary studies that address the research questions are listed in this section. Data sources and search keywords were the two stages of the search technique.

### The Search Terms Phase

These keywords include "crime prediction," "predictive crime technique," "machine learning crime detection," "crime data mining," "crime link prediction," and "crime forecasting models." The search terms phase involved identifying the key terms from the chosen research questions. To include the most recent developments in crime prediction methods, the search is restricted to papers released between 2018 and 2024 and the following steps were used to specify the formulation of the search terms:

- The study objectives and goals were used to determine the key search terms.
- The key phrases were to be replaced with new ones.
- To exclude unnecessary search results, boolean operators ORS and ANDS were used. Among the search terms used were "crime prediction" and "crime analysis" as well as "machine learning OR 'data mining' OR 'pattern recognition' OR 'nearest neighbour' OR 'regression tree' OR 'decision tree' OR 'classification tree'" both of which were linked to machine learning and crime prediction methods.

### Data Source Phase

Relevant publications are found and chosen for the SLR using databases that are supplied during the data source phase. Major academic databases, including the Association of Computing Machinery (ACM) Digital Library, Web of Science (WoS), Google Scholar, Scopus, Institute of Electrical and Electronics Engineers (IEEE) Xplore, and PubMed (for interdisciplinary studies), are used to conduct a systematic search for pertinent studies. Because these databases produced a substantial number of excellent research articles pertinent to our investigation, they were chosen.

### Selection Stage

After the documents are gathered, they are accessible in two stages to minimize the number of results that are appropriate for answering the research objectives. To gain access, the following two criteria were used: inclusion criteria and exclusion criteria.

### Inclusion Criteria

This involves findings and publications published in English and the inclusion criteria also include:

**Study Quality**: Peer-reviewed publications or articles from reputable journals and conferences, studies with clear methodologies and well-documented data sources and research using valid evaluation metrics (e.g., accuracy, precision, recall, F1-score, etc) for assessing machine learning models.

**Geographical Scope**: The geographical scope typically focuses on urban settings globally, with a preference for studies in high-crime-density areas or cities experiencing diverse types of crimes.

**Crime Type**: Also, the SLR focuses on the studies that address specific urban crime types such as theft, burglary, assault, drug-related offenses, and violent crimes.

**Machine Learning Focus**: The studies that applied machine learning such as supervised, unsupervised, or reinforcement learning methods to crime prediction are included in the SLR.

**Exclusion Criteria**

The studies that do not focus on crime prediction as a primary objective, non-peer-reviewed articles, commentaries, or opinion pieces, and duplicate studies or those with insufficient methodological detail.

**Dataset Used**

As shown in Table *2*, a variety of datasets have been employed in studies on crime prediction and detection. The Chicago Crime Dataset is one example, which contains data on crimes that have been reported in the Chicago area. Using this dataset, models have been created that predict the likelihood of specific crime types occurring in different areas of the city.

Research on crime prediction also uses the London Crime Dataset, which contains data on crimes that have been reported in London. A model has been created using this dataset to predict the likelihood of crimes in a certain area and how they relate to the socioeconomic traits of locals based on their geolocation.

**Table 2.** Dataset used in studie reviewed

| Dataset | Type of Data | Time Frame | Country |
|---|---|---|---|
| [20] | Crime | 2019 | San Francisco, United State of America. |
| [21] | Crime | 2021 | Los Angeles, United State of America. |
| [22] | Geospatial-Crime | 2022 | United Stated of America. |
| [23] | Temporal Crime | 2021 | Philadelphia. |
| [24] | Crime | 2018 | United Kingdom. |
| [25] | Crime | 2020 | India. |
| [26] | Crime | 2020 | India-Madhya Pradesh. |
| [27] | Crime | 2018 | India. |
| [6] | Crime | 2019 | Canada Vancouver. |
| [28] | Crime | 2021 | United Kingdom, London. |
| [15] | Crime | 2018 | United State, (New York). |
| [29] | Crime | 2018 | United State America. |
| [14] | Crime | 2018 | United State, Chicago. |
| [30] | Crime | 2018 | Brazil. |
| [20] | Crime | 2019 | San Francisco, United State of America. |

Furthermore, there are datasets that concentrate on video surveillance footage and different kinds of unusual activity for applications related to actual crimes. Examples of these datasets are the Los Angeles Crime Dataset, the New York City Crime Dataset, and the Philadelphia Crime Dataset. These datasets are frequently used in studies on crime detection and prediction and have been used to create models that predict the likelihood of particular types of crimes occurring in various locations, including data on crimes reported in the cities where they are located. All things considered, academics can utilize these statistics to build crime prediction models that could help law enforcement better prevent and deal with criminal activity.

## C. Result and Disucssion

This section presents the results and discussions of the Systematic Literature Review (SLR) conducted in this article, as shown in the following sections.

### Crime Prediction Techniques

This section summarizes the results of RQ1, which focusses on the most used Machine Learning (ML) methods for crime prediction. Machine Learning (ML) is an application that, without explicit programming, enables a system to automatically learn and get better from prior experiences [31, 32]. A precise pattern or information cannot always be identified after looking at the data [33, 34].

There are three types of machine learning: supervised, unsupervised, and semi-supervised [35]. An algorithm that uses labelled data to identify patterns and relationships to generate predictions or judgements is known as supervised learning [36].

Numerous researches have looked into different facets of supervised learning, such as applications and algorithms. k-Nearest Neighbours (k-NN), Naïve Bayes (NB), Random Forest (RF), Linear Regression (LR), Logistic Regression, SVM, Artificial Neural Network (ANN), Least Absolute Shrinkage and Selection Operation (LASSO), etc. are some of the basic techniques in supervised learning. In a variety of application fields, including marketing, finance, crime, health care, and natural language processing, it has been effectively used for both classification and regression tasks [37].

In the field of Artificial Intelligence (AI) known as "unsupervised machine learning," algorithms recognize patterns and structures in unlabeled data without the need for direct supervision [38]. The following are some examples of unsupervised learning techniques: k-means clustering, hierarchical clustering, Gaussian Mixture Models (GMM), Principal Component Analysis (PCA), Density-Based Spatial Clustering Application with Noise (DBSCAN), autoencoder, and t-Distributed etc.

Utilizing both labelled and unlabeled data for model training is the goal of semi-supervised learning approaches [39]. The objective of a semi-supervised learning model is to generate a better prediction result than what would be obtained from the model utilizing only the labelled data [40].

Graph-based techniques (e.g., label propagation, label spreading), generative models (e.g., semi-unsupervised Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), self-training, co-training, ensemble approaches, and hybrid methods are some examples. Due to its effectiveness in situations where obtaining labelled data is costly or difficult, semi-supervised learning has attracted a lot of attention. The frequency of machine learning approaches employed for crime prediction by the research papers analyzed in this study is depicted in Figure 3.

**Figure 3.** Different machine learning techniques for prediction of crime

The semi-supervised learning techniques aim to leverage both labelled and unlabeled data for model training [39]. The goal of a semi-supervised learning model is to provide a better outcome for prediction than that produced using the labelled data alone from the model [40]. Examples include, Graph-based technique (such as label propagation, label spreading), Generative models (semi-unsupervised Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs)), self-training, co-training, ensemble methods, and hybrid methods. Semi-supervised learning has gained significant attention due to its effectiveness in scenarios, where acquiring labelled data is expensive or impractical. Figure 3 shows the frequently of machine learning techniques used for crime prediction by research papers reviewed in this study.

Figure 3 shows that supervised learning is the most often used method for crime prediction, accounting for 36% of studies. Furthermore, 32% of the gathered research publications combined supervised and unsupervised methods, since some of them used multiple machine learning algorithms. Only 15% used unsupervised learning. Only 5% of respondents, which is surprising, employed the semi-supervised approach, indicating that it is not frequently used in the field of crime prediction. Lastly, 12% of the studies did not identify the methodology they used. Research publications have used a variety of methods to predict crime. Random Forest and Naïve Bayes (NB) are the most commonly used algorithms in crime prediction. Random Forest (RF), Naïve Bayes (NB), and decision tree algorithms were used in 25 studies.

## Crime Prediction Performance and Evaluation Metrics

In this section, the results of the assessment of research question (RQ2) are discussed in relation to the models' estimation accuracy. A list of the most widely used performance and assessment measures in crime prediction is compiled from 100 distinct studies.

A total of 12 research papers were found to have not addressed the assessment metrics. Additionally, the performance indicators are arranged in a table. Take the mean error category, which comprises Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Error Rates, and Root Mean Squared Error (RMSE). In addition, various measures like recall, detection rate, False Positive Rate (FPR), Predictive Efficiency Index (PEI), Recapture Rate Index (RRI), Directional Accuracy (DA), and True Positive Rate (TPR) also referred to as sensitivity were examined, as seen in Table 3.

**Table 3.** Frequency of different evaluation metrics used for machine learning crime prediction in studie reviewed

| Metric type | Specific metric | Frequency | Percentage (%) |
|---|---|---|---|
| Mean | MAPE<br>MAE<br>MSE<br>RMSE<br>MRE | 59 | 21% |
| Accuracy | EE<br>Directional Accuracy (DA)<br>PAI<br>PAI | 45 | 16% |
| True Positive Rate (TPR) | Sensitivity<br>Detection Rate<br>Recall<br>Hit Rate<br>Correctly Classified | 38 | 14% |
| False Positive Rate (FPR) | False Alarm Rate<br>Correct exclusion | 9 | 3% |
| Confusion Matrix | N/A | 5 | 1% |
| True Negative | Specificity False<br>Hit | 5 | 1% |
| False Negative Rate (FNR) | Misclassified<br>Incorrect exclusion | 4 | 1% |
| Prediction Index | Predictive Efficiency Index (PEI)<br>Recapture Rate Index (RRI) | 4 | 1% |
| Kappa Statistics | N/A | - | 1% |
| Memory Usage | N/A | - | 1% |
| Jensen Shannon Divergence (JS) | N/A | | 1% |
| No applied performance metrics | N/A | 19 | 21% |

Table 3 shows that the mean error is the most commonly used performance metric, appearing in 59 studies. Accuracy is used in 45 papers, while the true positive rate is used in 38 papers.

## Crime Prediction Implementation Tools

In this section, the findings of the research question (RQ3) regarding the most commonly used crime prediction implementation tools are presented. From the study, 18 different tools were used as presented in Figure 4, and the most applied tool is Weka was the most commonly used tools in research papers with frequency of 27%. Followed by Python programming language and R-language, as can be seen in Figure 4.

**Figure 4.** Tools used in the selected research papers for crime prediction

**Crime Prediction Commonly Used Datasets**

In this section, the findings of the research question (RQ4) regarding the commonly used dataset for crime prediction in machine learning are presented. There are different publicly available crime datasets. The common datasets used in crime prediction on machine learning techniques is shown in Figure 5.

**Figure 5.** Frequency of data set commonly used in crime prediction by countries

From Figure 5, one can see that Chicago crime dataset which is an open dataset is the most widely used with 19 research papers, followed by India crime dataset which was used in 12 research papers, then California and US crime datasets found in 12 research papers respectively.

## Challenges of Crime Prediction Using Machine Learning Techniques

This section presents the findings of RQ5, which concerns the challenges of crime prediction techniques using Machine Learning. There are different challenges recorded in literature that affect accurate prediction of crime with machine learning, these include:

**Data quality and availability:** Crime data is often incomplete, inaccurate, or inconsistent. Missing records, underreporting, or misclassification of crimes can distort predictions and hinder machine learning models' performance. For instance, crimes in remote or underprivileged areas may be underreported, leading to biased predictions.

**Temporal dynamics of crime:** Crime patterns evolve due to changes in socioeconomic conditions, laws, or policies. Static models may fail to adapt to these changes. Example: a model trained on pre-pandemic crime data might not predict post-pandemic trends accurately.

**Lack of ground truth for unreported crimes:** Many crimes go unreported, creating a gap between actual crime rates and the data available for modeling. Example: crimes like domestic violence and cybercrimes are notoriously underreported.

**Integration with socioeconomic and psychological factors**: Crime prediction requires integrating complex variables like socioeconomic status, cultural influences, and psychological factors, which are challenging to quantify. Example: incorporating mental health indicators into crime models is difficult due to privacy concerns and data availability.

**Real-time prediction**: Developing systems that predict crimes in real time requires significant computational resources and advanced algorithmic capabilities, which remain challenging. Example: a real-time burglary alert system needs instantaneous processing and low-latency responses.

## Discussion

This section discusses the various machine learning models that have proven to be effective for crime prediction. These models include support vector machines, logistic regression, and random forests which have been utilized to analyze crime data and identify patterns that can be used to predict criminal activity. Table 4 shows the advantages and limitations of some of these machine learning techniques.

**Table 4.** Advantages and limitation of some of the machine learning crime prediction models reviewed

| Ref. | Model | Advantages | Limitations |
|---|---|---|---|
| [41] | Random Forest (RF) | • Can process datasets with a large number of features, which is common in crime data.<br>• Due to ensemble learning, RF tends to generalize well.<br>• Can use class weights or sampling techniques internally to handle imbalance.<br>• Suitable for capturing complex interactions between features in crime patterns. | • Training can be slow with large datasets.<br>• Difficult to derive specific insights from the model.<br>• May create overly complex trees if data has noisy variables. |
| [42] | Support Vector Machines (SVM) | • Handles large feature sets well.<br>• Particularly when using the kernel trick.<br>• Kernels allow modeling of complex relationships in crime patterns. | • Slow for large datasets due to quadratic complexity.<br>• Choosing the right kernel and parameters (C, gamma) is critical.<br>• Struggles with classes that have significantly fewer samples; class weighting can mitigate this. |
| [43] | Linear Regression (LR) | • Provides a clear understanding of relationships between variables.<br>• Fast to train and deploy.<br>• Suitable for predicting numerical trends in crime rates. | • Performs poorly when relationships between variables are non-linear.<br>• Outliers can disproportionately influence predictions.<br>• Assumes uniform variance and struggles when certain crime patterns are underrepresented. |
| [44] | k-Nearest Neighbors (k-NN) | • Does not require a training phase; directly classifies based on proximity.<br>• Does not assume any underlying data distribution.<br>• Effective for localized crime trends. | • Distance calculation for all points can be slow for large datasets.<br>• Outliers can significantly impact predictions.<br>• Predominantly classifies into the majority class. |
| [45] | Gaussian Mixture Model (GMM) | • Can model multi-modal distributions, which are common in crime data.<br>• Effective for clustering crime patterns without labeled data.<br>• Provides the likelihood of data points belonging to different clusters. | • Poor initialization can lead to suboptimal results.<br>• May not perform well if the data does not follow Gaussian patterns. |

## D. Conclusion

Advances in technology have impacted crime prediction, creating difficult problems for law enforcement and security personnel. Recently, researchers have become more interested in utilizing machine learning to predict crime, with a focus on detecting criminal patterns and trends in crime occurrences.

This study examines over 900 publications to investigate the different machine learning methods used to predict crime. This study used a variety of machine learning algorithms and publicly available datasets to extract 100 articles that were carefully chosen. Decision trees, support vector machines, neural networks, and ensemble approaches are just a few of the machine learning techniques used for crime prediction, according to the report. These techniques have shown differing levels of efficacy and accuracy depending on the particular goals of the prediction task and the intricacy of the data.

Performance evaluation metrics, such as accuracy, precision, recall, and f1-score have been instrumental in assessing model efficiency. However, the inconsistent use of these metrics across studies has made it challenging to directly compare results. Moreover, studies indicate that the choice of metric significantly impacts the interpretation of a model's performance, especially in imbalanced datasets. The datasets commonly used in crime prediction research are diverse, ranging from publicly available data such as the city-specific crime reports to real-life datasets. While publicly available datasets provide accessible resources for initial research, their generalizability is often limited due to geographical and demographic differences.

Despite advancements, the field faces numerous challenges. Key issues include data imbalance, where certain crime types dominate the datasets, leading to biased predictions; data privacy and ethical concerns, which hinder data sharing and utilization; and real-world applicability, where models struggle to account for dynamic human behavior and external socio-political factors.
To address these challenges, future research should focus on developing standard benchmarks for evaluation, improving data collection methodologies, and incorporating interdisciplinary approaches that integrate sociology, psychology, and criminology. Moreover, enhancing collaboration between academia, industry, and law enforcement agencies can foster the development of more robust and practical crime prediction systems.

In conclusion, while significant strides have been made in crime prediction, there is still a need for more standardized, ethical, and contextually aware methodologies to ensure reliable and actionable insights that can aid in crime prevention and law enforcement.

## E. References

[1] L. A. Garcia-Segura, "The role of artificial intelligence in preventing corporate crime," *Journal of Economic Criminology,* vol. 5, no. 100091, 2024.

[2] F. Dakalbab, M. A. Talib, O. A. Waraga, A. B. Nassif, S. Abbas, and Q. Nasir, "Artificial intelligence & crime prediction: A systematic literature review," *Social Sciences & Humanities Open,* vol. 6, no. 1, p. 100342, 2022.

[3] P. D. B. Raut, S. Gawade, Y. Gadade, R. Garad, and A. Mathapati, "Criminal Combat:- Crime Analysis And Prediction Using Machine Learning,"

*International Research Journal of Modernization in Engineering Technology and Science,* vol. 06, no. 2, pp. 1039-1042, 2024.
[4] N. B. M. S. N. Shah, "Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention," *Vis. Comput. Ind. Biomed. Art* vol. 4, no. 9, 2021, doi: https://doi.org/10.1186/s42492-021-00075-z.
[5] X. Zhang, L. Liu, M. Lan, G. Song, L. Xiao, and J. Chen, "Interpretable machine learning models for crime prediction," *Computers, Environment and Urban Systems,* vol. 94, no. 101789, 2022.
[6] V. Mandalapu, L. Elluri, P. Vyas, and N. Roy, "Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions," *IEEE Acess,* vol. 11, pp. 60153-60170, 2023.
[7] O. A. Esan and I. O. Osunmakinde, "A computer vision model for detecting suspicious behaviour from multiple cameras in crime hotspots using convolutional neural networks," *International Conference on Practical Applications of Agents and Multi-Agent Systems,* pp. 197-209, 2022.
[8] D. Bright, R. Brewer, and C. Morselli, "Using social network analysis to study crime: Navigating the challenges of criminal justice records," *Social Networks,* vol. 66, pp. 50-64, 2021.
[9] S. M. A. Amin Ullah, Jianqiang Li, Lubna Nadeem, Tariq Mahmood, Amjad Rehman & Tanzila Saba "Smart cities: the role of Internet of Things and machine learning in realizing a data-centric smart environment," *Complex Intell. Syst.* , vol. 10, pp. 1607–1637, 2024, doi: https://doi.org/10.1007/s40747-023-01175-4.
[10] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN COMPUT. SCI.* , vol. 2, no. 160, 2021, doi: https://doi.org/10.1007/s42979-021-00592.
[11] J. He and H. Zheng, "Prediction of crime rate in urban neighborhoods based on machine learning," *Engineering Applications of Artificial Intelligence,* vol. 106, no. 104460, 2021.
[12] S. Sathish and S. Varadhan, "Location-Based Crime Prediction Analysis Using Machine Learning," *International Journal of Research Publication and Reviews,* vol. 5, no. 4, pp. 5288-5292, 2024.
[13] I. M. Miguel Saraiva , Saloni Mishra and Ana Amante "Crime Prediction and Monitoring in Porto, Portugal, Using Machine Learning, Spatial and Text Analytics," *ISPRS Int. J. Geo-Inf.* , vol. 11, no. 400, 2022, doi: https://doi.org/10.3390/ijgi11070400.
[14] C. Catlett, E. Cesario, D. Talia, and A. Vinci, "A data-driven approach for spatiotemporal crime predictions in smart cities," *In 2018 IEEE International Conference on Smart Computing (SMARTCOMP),* pp. 17-24, 2018.
[15] U. M. Butt *et al.*, "Spatio-temporal crime predictions by leveraging artificial intelligence for citizens security in smart cities," *IEEE Access,* vol. 9, pp.:47516–47529, 2021.
[16] E. Barbierato and A.T.Gatti, "The Challenges of Machine Learning: A Critical Review," *Electronics* vol. 13, no. 416, 2024, doi: https://doi.org/10.3390/electronics1302041.
[17] R. R. Bharath, H. K. Sulthan, R. M. Mingaz, and S. N. Kumaravengatesh, "A Machine Learning Approach To Crime Analysis And Forecasting For Prediction

And Prevention," *African Journal of Biological Science,* vol. 6, no. Si1, pp. 1301-1313, 2024.
[18] D. L. J. A. A. Tursunalieva, R. Dunne, J. Li, L. Riera, Y. Zhao, "Making Sense of Machine Learning: A Review of Interpretation Techniques and Their Applications," *Appl. Sci.* , vol. 14, no. 496, 2024, doi: https://doi.org/10.3390/app14020496.
[19] J. E. M. Matthew J. Page, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, Lesley A. Stewart, James Thomas, Andrea C. Tricco, Vivian A. Welch, Penny Whiting & David Moher "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *Syst Rev* vol. 10, no. 89, 2021, doi: https://doi.org/10.1186/s13643-021-01626-4.
[20] A. Sangani, C. Sampat, and V. Pinjarkar, "Crime prediction and analysis," *In 2nd International Conference on Advances in Science & Technology (ICAST),* 2019.
[21] X. Zhou, X. Wang, G. Brown, C. Wang, and P. Chin, "Mixed spatio-temporal neural networks on real-time prediction of crimes," *In 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA),* pp. 1749–1754, 2021.
[22] B. Zhou, L. Chen, S. Zhao, S. Li, Z. Zheng, and G. Pan, "Unsupervised domain adaptation for crime risk prediction across cities," *IEEE Transactions on Computational Social Systems,* 2022.
[23] J. He and H. Zheng, "Prediction of crime rate in urban neighborhoods based on machine learning.," *Engineering Applications of Artificial Intelligence,* vol. 106, no. 104460, 2021.
[24] H. K. R. ToppiReddy, B. Saini, and G. Mahajan, "Crime prediction&monitoring framework based on spatial analysis," *Procedia computer science,* vol. 132, pp. 696–705, 2018.
[25] P. Tamilarasi and R. U. Rani, "Diagnosis of crime rate against women using k-fold cross validation through machine learning," *In 2020 fourth international conference on computing methodologies and communication (ICCMC),* pp. 1034–1038, 2020.
[26] A. Kumar, A. Verma, G. Shinde, Y. Sukhdeve, and N. Lal, "Crime prediction using knearest neighboring algorithm," *In 2020 International conference on emerging trends in information technology and engineering (IC-ETITE),* pp. 1-4, 2020.
[27] S. Agarwal, L. Yadav, and M. K. Thakur, "Crime prediction based on statistical models," *In 2018 Eleventh International Conference on Contemporary Computing (IC3),* pp. 1-3, 2018.
[28] M. Sathiyanarayanan, A. K. Junejo, and O. Fadahunsi, "Visual analysis of predictive policing to improve crime investigation," *In 2019 International Conference on contemporary Computing and Informatics (IC3I),* pp. 197–203, 2019.
[29] A. M. Shermila, A. B. Bellarmine, and N. Santiago, "Crime data analysis and prediction of perpetrator identity using machine learning approach," *In 2018*

*2nd international conference on trends in electronics and informatics (ICOEI),* pp. 107-114, 2018.
[30] A. Araújo, N. Cacho, L. Bezerra, C. Vieira, and J. Borges, "Towards a crime hotspot detection framework for patrol planning," *In 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS),* pp. 1256–1263, 2018.
[31] K. Ahir, K. Govani, R. Gajera, and M. Shah, "Application on virtual reality for enhanced education learning, military training and sports," *Augment Hum,* vol. 5, no. 1, 2020, doi: https://doi.org/10.1007/s41133-019-0025-2.
[32] T. Talaviya, D. Shah, N. Patel, H.Yagnik, and M.Shah, "Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides," *Artif Intell Agric 4:58–73,* vol. 4, pp. 58-73, 2020, doi: https://doi.org/10.1016/j.aiia.2020.04.002.
[33] M. Pathan, N. Patel, H. Yagnik, and M. Shah, "Artificial cognition for applications in smart agriculture: a comprehensive review," *Artif Intell Agric,* vol. 4, pp. 81-95, 2020, doi: https://doi.org/10.1016/j.aiia.2020.06.001.
[34] R. Pandya, S. Nadiadwala, R. Shah, and M. Shah, "Buildout of methodology for meticulous diagnosis of K-complex in EEG for aiding the detection of Alzheimer's by artificial intelligence," *Augment Hum Res,* vol. 5, no. 1, 2020, doi: https://doi.org/10.1007/s41133-019-0021-6.
[35] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science,* vol. 2, no. 160, pp. 1-16, 2021.
[36] R. Pugliese, S. Regondi, and R. Marini, "Machine learning-based approach: global trends, research directions, and regulatory standpoints," *Data Science and Management,* vol. 4, pp. 19-29, 2021.
[37] M. Usman, R. Azam, M. A. Jaffar, I. Mehmood, H. S. M. Bilal, and H. Anwar, "A Comprehensive Review on Applications of Random Forest in Healthcare: Opportunities and Challenges," *Journal of Healthcare Engineering,* 2023, doi: 10.1155/2023/123456.
[38] J. Wang and F. Biljecki, "Unsupervised Machine Learning in Urban Studies: A Systematic Review of Applications," vol. 129, 2022, doi: https://doi.org/10.1016/j.cities.2022.103925.
[39] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science,* vol. 2, no. 60, pp. 1-21, 2021.
[40] H. J. Kim and S. Patel, "Enhancing Predictive Accuracy: Exploring the Role of Semi-Supervised Learning Models," *Journal of Machine Learning Research,* vol. 24, no. 3, pp. 210-225, 2023.
[41] H. V. R. L. Alves, & F. Rodrigues, "Crime prediction through urban metrics and statistical learning," *Physica A: Statistical Mechanics and Its Applications,* vol. 505, 2018.
[42] K. K. R. Alhajj, "Effectiveness of Support VEctor Machine for crime hot-spots prediction," *Applied Artificial Intelligence,* vol. 22, no. 5, pp. 433-458, 2008, doi: https://doi.org/10.1080/08839510802028405.
[43] A. I. A. C. G. Lozano, "Tax Crime Prediction with Machine Learning: A Case Study in the Municipality of São Paulo," *In ICEIS,* vol. 1, pp. 452-459, 2020.

[44] N. C. A. Araujo, 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),* pp. 1256–1263, 2018.
[45] M. V. Paulina Flores, Pablo Fuentes, Francisco Jaramillo, David Acuna, Aramis Perez, Marcos Orchard, "Modeling and Prediction of Criminal Activity Based on Spatio-Temporal Probabilistic Risk Functions," *Annual Conference of the Prognostics and Health Management Society,* pp. 1-14, 2015.