

Review

# Crime Prediction Using Machine Learning: A Comparative Analysis

Abdulrahman Alsubayhin, Muhammad Ramzan and Bander Alzahrani

Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

## Article history

Received: 07-10-2022

Revised: 19-07-2023

Accepted: 17-07-2023

## Corresponding Author:

Abdulrahman Alsubayhin  
Faculty of Computing and  
Information Technology, King  
Abdulaziz University, Jeddah,  
Saudi Arabia  
Email:  
aabdulrahmanalsubayhin@stu.kau.edu.sa

**Abstract:** In the past few years, Machine Learning (ML) methods have acquired extensive attention from researchers and even criminology practitioners for their potential in crime prediction and crime forecast. To date, many research studies have already been performed on crime prediction using machine learning techniques. Some common methods that have been used include KNNs, Decision trees, Naïve Bayes, and random forest, among others. In this research study, a comparative analysis of 51 research studies has been performed. The results indicate that the supervised learning approach is the most commonly used by researchers. In addition, random forest was the most commonly used method. In various studies, almost all proposed models and methods produced accurate results and outcomes with the exception of one study. In the future, it is critical to evaluate these ML based algorithms in a real-world situation and identify which situations or factors affect their accuracy. At the same time, it is also necessary to identify and evaluate which techniques can be used to improve the accuracy of each machine learning method and which algorithm or method might prove to be the best choice for police departments to use for predicting crime.

**Keywords:** Machine Learning, Crime Prediction, Crime Forecast, Crime Prediction Using Machine Learning, KNNs, Decision Trees, Naïve Bayes, Random Forest, ML Based Algorithms

## Introduction

Generally, crimes are rather common social issues, influencing a country's reputation, economic growth, and quality of life. They are perhaps a prime factor in influencing several critical decisions in a person's life, such as avoiding dangerous areas, visiting at the right time, and moving to a new place (ToppiReddy *et al.*, 2018). Crimes define and affect the impact and reputation of a community while placing a rather large financial burden on a country due to the need for courts and additional police forces (Saraiva *et al.*, 2022). With an increase in crimes, there is an increased need to reduce them systematically. In recent times, there has been a record increase in crime rates throughout the world. It is possible to reduce these figures by analyzing and predicting crime occurrences. In such a situation, preventive measures can be taken quickly (ToppiReddy *et al.*, 2018).

Crime forecasting in real-time is capable of helping save lives and prevent crimes, gradually decreasing the

crime rate (Wang *et al.*, 2019). With a comprehensive crime data analysis and modern techniques, crimes can be predicted and support can be deployed without delay. Machine learning has acquired significant attention in the past few years due to its potential and implications (Wang *et al.*, 2020). It has begun to be tested even in forecasting and predicting crime rates and previous studies have attested to its potential. In this study, a comparative analysis is carried out of previous studies on crime prediction through machine learning to identify the current techniques and schemes being used for crime predictions.

## Machine Learning

For most people, ML or machine learning is still a rather unexplored and mysterious field that appears difficult and complicated to explain. However, it has become very important in almost every field in just a few years. Regardless, it is quite a multidisciplinary field, dealing mainly with mathematics and programming, specifically density functions and

probability. Furthermore, it has multiple applications in various fields, including crime detection (Prathap, 2023). In order to work, machine learning generally applies three prime components including loss, data, and model. Meanwhile, different ML methods implement and use “trial and error” as a working principle. These methods consistently refine and validate a model based on a loss, which is incurred by its predictions about a specific occurrence or phenomenon as demonstrated by Fig. 1 (Ippolito and Lozano, 2020).

ML algorithms often learn data as they acquire experience through the adjustment of their parameters based on data patterns, characteristics, and features to obtain and acquire an optimal performance measure (Ippolito and Lozano 2020). In machine learning, a usual task is associated with classification. Algorithms that specialize in it need to specify and determine which category some data belongs to. Classification can be unsupervised or supervised. Meanwhile, for classification, performance measures can be based on several values, including true negatives, false negatives, false positives, and true positives (Ippolito and Lozano, 2020).

### Neural Networks

Fundamentally, neural networks are different models represented generally by interconnected nodes. Usually, these networks solve different types of classification problems. The neural network receives data features and values before calculating weights for features (Ippolito and Lozano, 2020). The underlying goal here is to reduce the error between the actual classification and forecasted classification. In addition, these weights are primarily adjusted to random values. Thus, in the sequence, these weights are adjusted by an iterative process, which helps minimize errors (Wu *et al.*, 2022).

### Naïve Bayes

Naïve Bayes (NB) is used for classification as a probabilistic algorithm and assumes independence among different features. It is based mainly on different conditional probabilities and its use for classification allows for probability calculation of classification in a class of data (Kshatri *et al.*, 2021).

### Decision Trees

Decision Trees are recognized as models based on structure for classification and are represented or shown as hierarchies that create trees in which nodes represent data features. Possible values are represented by arcs that come from a specific node (William *et al.*, 2023). Moving down to the hierarchy’s lowest level represents leaves, which indicates possible

classifications of different data elements. In a decision tree, the starting node corresponds to the feature of the data that partitions different elements of the data into homogeneous groups.

### Ensemble Learning

Ensemble Learning tends to combine the outcomes of different algorithms to yield better performance than the ones offered by the individual application of its different algorithms. Simple fusion methods are the most commonly used techniques for the combination of different classifiers. Furthermore, these methods tend to combine outcomes of different classifiers and if output is offered by probabilities, the last or final classification is measured by identifying the median, average, or maximum probability of all the given classifiers (Ippolito and Lozano, 2020).

### Random Forests

Random Forests (RF) are primarily based on the development of different decision trees and these trees are combined in such a manner that a random forest represents an ensemble of different decision trees. For each tree, a training set of various or multiple data elements is randomly drawn (Alves *et al.*, 2018). These sets are used for growing each tree without pruning. For the minimization of errors, it is essential for trees to be the least similar.

### Logistic Regression

Logistic regression is fundamentally an algorithm of classification and tends to classify data elements in at least two classes in its basic form. A logistic function is used by it, which is based on an NL or natural logarithm. Weights are fit by logistic regression to different data features, which minimizes the error between the true and predicted class (Ippolito and Lozano, 2020).

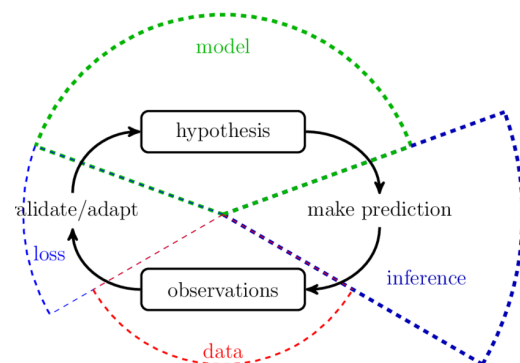


Fig. 1: General schematic of a machine-learning model

## Crime Prediction

The literature on criminology studies the association between different features and crime to develop approaches and methods for crime prediction and forecasting. Most works tend to focus on the identification of hotspots, areas where a high probability of crime exists (Gupta *et al.*, 2022). Some common methods include kernel density estimation, thematic mapping, and spatial and temporal analysis of crime. Meanwhile, in the case of techniques based on ML, ensemble learning, logistic regression, random forests, and deep learning are most often used (Azeez and Aravindhar, 2015). In some studies, these techniques are combined to acquire better accuracy and deliver more precise outcomes.

Some governments use ML-based frameworks for predicting crime. For example, police in Venice use ML for the prediction of crimes such as shootings and robberies. The IRS or Internal Revenue Service of the US uses ML to detect and identify pre-refund fraud and identity theft in the tax system (Aziz *et al.*, 2022). Italian Police also use ML to predict crime rates in different cities to prevent them and ensure that minimal damage is caused. Crime prediction is carried out to ensure the security and safety of citizens and several studies have indicated the use of multivariate time series, Artificial Neural Networks (ANNs), fuzzy theory, and SVM for crime predictions (Ippolito and Lozano, 2020).

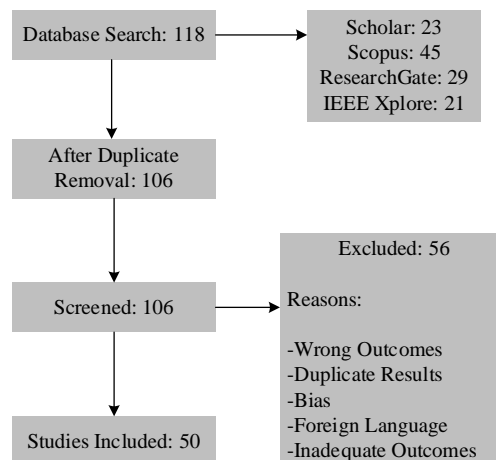
We present a comparative analysis of multiple existing studies on machine learning techniques for the prediction of crime to identify which methods tend to yield optimal accuracy. In the following section, a framework for the research and identification of previous studies has been outlined, presented, and discussed.

## Materials and Methods

The current study is based on a qualitative research design and performs a comparative analysis of credible research studies on the use of machine learning for the prediction of crime through a detailed review of studies. Secondary methods in the form of a literature review have been used to acquire sufficient and relevant information about the topic to perform the study.

### Search Strategy

In order to evaluate and review existing studies, a detailed search strategy was created and used. Several databases were used for searching and identifying relevant studies, such as Google Scholar, Scopus, ResearchGate, and IEEE Xplore. In these databases, different terms such as “Machine Learning and Crime Prediction,” “Crime Prediction using Machine Learning Methods,” and “Crime Forecasting and Machine Learning” were used to acquire a large number of studies.



**Fig. 2:** Schematic of the search strategy and inclusion/exclusion of studies

### Inclusion Criteria

To be included in the study, it was essential for the research study to be published in English, be credible and peer-reviewed, relevant, and indicate adequate outcomes. Studies needed to focus on machine learning-based methods for predicting and forecasting crimes. In addition, it was also necessary for research studies to be performed in the last few years (Fig. 2).

### Exclusion Criteria

All those research studies that did not meet the criterion identified above were excluded. For example, all research studies that were not in English or did not study machine learning methods and frameworks for predicting crime were excluded.

## Results

In this research, a comparative analysis of 50 research studies on crime prediction with the use of machine learning algorithms has been considered (Table 1). Out of all the studies, 6 research studies considered and used SVM for the prediction of crime. Meanwhile, 12 research studies used the KNN or K-Nearest algorithm to predict and forecast crime using large datasets. 13 research studies relied on Random Forest to analyze data and predict crime using it. 8 research studies relied on and considered using decision trees to predict and forecast crime using historical data. 10 research studies considered and used Naïve Bayes to predict and forecast crime in multiple locations.

The supervised learning method has been considered in almost all research studies. Few to no studies were encountered during this comparison that actually used an unsupervised learning method for the prediction of crime through ML algorithms. Some important themes have been identified upon the analysis of research studies and their comparison.

**Table 1:** Comparative analysis

No.	Methods	Results	Reference
[1]	A literature review is carried out to identify and evaluate various crime prediction techniques and methods	Artificial intelligence and multiple ML-based methods such as SVM and random forests are used	Grover <i>et al.</i> (2006)
[2]	A dynamic prediction model is proposed basis of the Bayesian learning method	The proposed model is capable of predicting the area where the next crime might happen or occur	Liao <i>et al.</i> (2010)
[3]	The study uses two different classification algorithms including Decision Tree and Naïve Bayesian for the prediction of crime categories for various US-based states	The results of the study indicate that the decision Tree algorithms tend to outperform Naïve Bayesian algorithms with an accuracy of 83% in the prediction of crime categories for various states of the US	Iqbal <i>et al.</i> (2013)
[4]	A literature review is performed and a comparative analysis of various data mining techniques is performed	Fuzzy time series and multivariate time series Clustering identified to perform better than others	Dubey and Chaturvedi (2014)
[5]	Real crime data sets are used from the 1990 US census and Back Propagation and Naïve Bayesian algorithms are compared for the crime category	The calculation of Naïve Bayes outperformed calculation of BP and acquired an accuracy of 90% for the first group and around 94.08% for the second group	Babakura <i>et al.</i> (2014)
[6]	WEKA as a data mining software is used for performing a comparative study. Decision Stump, Additive Regression and Linear Regression algorithms are the three have been implemented using a finite set of features the Crime Dataset offered by the University of California-Irvine repository	Upon comparison, it was identified that the linear regression algorithm performed effectively against selected or identified algorithms. It was the most accurate in the analysis and prediction of crime data	McClendon and Meghanathan (2015)
[7]	The authors rely on a graph database model that is based on deep Learning the component of the graph loader loads graphs on the graph data store for the analysis of graphs	The hybrid prediction algorithm is capable of predicting crime with an accuracy of 93% but it has a limitation associated with the management of a large amount of data	Azeez and Aravindhar (2015)
[8]	Three area-specific and two global models have been used for the prediction of crime	Area-specific models are capable of exhibiting higher accuracy and precision as compared to global models	Al Boni and Gerber (2016)
[9]	Crime forecasting is performed for PPB or Portland Police Bureau and multiple ML algorithms are used for the data including neural networks, gradient boosting machines, RF and SVM or support vector machine	For two different datasets, different results were achieved. For the first dataset, gradient-boosting machines and random forests were more accurate. On the second database, neural networks and SVM were better	Nguyen <i>et al.</i> (2017)
[10]	A comprehensive review is performed to determine different ML-based techniques for predicting crimes	Multivariate time series, SVM, ANN, and random forests are the prime methods that have been commonly used for carrying out studies	Shamsuddin <i>et al.</i> (2017)
[11]	A feature-level data fusion method is proposed with environmental context on the basis of a DNN or deep neural network. The dataset itself involves data, which is acquired from different online databases in Chicago. The DNN is configured with four layers including the joint feature representation, environmental, temporal, and spatial layers	Experimental performance outcomes indicate that the trained DNN is quite precise and accurate in the prediction of crime occurrence than other models of prediction. An accuracy of around 84.25% is achieved and a precision of around 74.35% is obtained	Kang and Kang (2017)
[12]	A survey of unsupervised and supervised learning techniques that have been used for crime identification	Each technique has its own specific drawbacks and benefits to be more Supervised learning methods are identified precise than others	David and Suruliandi (2017)
[13]	A survey of techniques and methods for the prediction of crime is performed	Various techniques under the supervised ML approach are often considered and used for predicting various types of crimes such as traffic violations and crimes against people	Prabakaran and Mitra (2018)
[14]	Machine learning algorithms and visualization techniques such as Google Maps and GoogleVis are used. KNNs were used for the extraction of knowledge from datasets	The proposed framework is capable of showing probabilities of crimes such as robbery and drugs among others in different areas	ToppiReddy <i>et al.</i> (2018)
[15]	A random forest regressor is used for the prediction of crime and quantification of the influence of different urban indicators on homicides,	The authors identify that the proposed model delivers with 97% accuracy the model also indicates that homicides are associated closely with variables like illiteracy and unemployment	Alves <i>et al.</i> (2018)
[16]	Crime data from Vancouver is acquired and analyzed using two different approaches to data processing including decision the boosted decision tree and KNN	For the crime data of 15 years, KNN and boosted trees are considered and used. The accuracy of crime prediction for the used algorithms is identified to be 39-44%	Kim <i>et al.</i> (2018)
[17]	The crime dataset is acquired from Caggle and KNN and Multi-linear regression is used in ML-based techniques	The model is capable of producing an accuracy of 0.97 for relationships, 0.96 for perpetrator sex and 0.60 for perpetrator age	Shermila <i>et al.</i> (2018)
[18]	Deep neural networks are used for making predictions using Portland and Chicago crime data. The crime counts are segregated into ten bins. Feed Forward Network, Convolutional Network, Recurrent Network and Recurrent Convolutional Networks are used	Using deep neural networks, The authors were able to acquire accuracy levels of around 65.3 and 75.6% for Portland and Chicago respectively. The results indicate the efficiency and effectiveness of neural networks for crime prediction	Stec and Klabjan (2018)
[19]	Human mobility data is used for crafting different features for crime prediction. Boosting and average ensemble techniques are used for studying crime data	Spatio-temporal and spatial features enhance the baseline models that rely on POI data and census data. The proposed models achieve $R_2$ absolute metrics of 89% on a temporal test set and 65% geographical set	Kadar and Pletikosa (2018)
[20]	The proposed system is implemented using VGGNet-19, a pre-trained deep-learning model. To further improve it, RCNN and Fast RCNN were used	The proposed model is able to detect crimes by identifying knives and guns in hands in less computation time than the model of GoogleNet inception V3	Navalgund and Priyadharshini (2018)
[21]	A survey of different ML-based techniques for the prediction of crime data is performed and carried out	KNN, Naïve Bayes, and KNNs are used primarily for the analysis and forecast of crimes using input data	Almaw and Kadam (2018)
[22]	Two deep neural network structures are considered including the adapted form and the second one excludes convolution. The method is implemented using Kera on top of the Theano	The methods used by the authors offer crime forecasting for every grid cell at an hourly temporal scale and the predictions themselves are accurate	Wang <i>et al.</i> (2019)

**Table 1:** Continue

	application. Tests were run on crime data for the last six months of 2015		
[23]	Deep Neural Networks are used for making predictions based on the history of arrest bookings. Data modeling based on Dynamic-Window is considered and used as well	The authors determine that DNNs can definitely help predicts criminal behavior and the data pooling method performs better than all others for multiple class crime predictions	Chun <i>et al.</i> (2019)
[24]	A systematic literature review is performed to identify the use of data mining for predicting crimes. 42 studies were included in the process	13 studies that were found to use supervised learning techniques and 11 were identified to use the applied clustering approach for the prediction of crime	Adesola <i>et al.</i> (2019)
[25]	The feature matrix is created on the basis of SNA metrics including walk-based indices or Katz-based indices and identifying common neighbor. DNN and SNA neural networks are used for the model	The TDRL-CNA link prediction model is capable and predicting links effectively and the proposed model exhibits a higher accuracy than the baseline DRL-CAN model	Lim <i>et al.</i> (2019)
[26]	CNN is used for the extraction of semantic and spatial features. LSTM is used for the extraction of temporal features and deep learning architectures including ResNet-50, Inception V3 and VGG16 are used	The ResNet50+LSTM-based model was identified to achieve and exhibits better accuracy than other models in predicting crimes	Rajapakshe <i>et al.</i> (2019)
[27]	Data mining methods are used for the analysis of data and the prediction of crimes based on it. Some important classifiers used include recall and accuracy, the number of correct classifications, and precision	Decision trees tend to exhibit better accuracy than others ML methods and data mining methods are capable of contributing to better predictions and forecasts	Llaha (2020)
[28]	A dataset is extracted from CLEAR and both Naïve Bayes and Decision Trees are used	The comparison of the two algorithms indicates that Decision trees performed and exhibited better outcomes than the Naïve Bayes algorithm	Aldossari <i>et al.</i> (2020)
[29]	A supervised learning technique is used for the prediction of crimes and the proposed system performs a prediction accuracy through the analysis of data. The system has two main algorithms including the k-nearest algorithm and decision trees. Meanwhile, the RF algorithm and Adaboost are used for increasing predictive accuracy	The proposed framework acquired a low predictive of 31.71% for unequally distributed classes frequencies After remodeling the data, random forests gained an accuracy of 99.16% as compared to the accuracy 68.03% of over methods	Hossain <i>et al.</i> (2020)
[30]	A dataset is acquired from the San Francisco Police Department for the analysis of data and crime prediction using ML algorithms like multilayer perceptron, and Gaussian Naïve Bayes, K-NN and CART. Cross-validation and validation were performed for testing results	Upon the comparison of results delivered by different algorithms, it was identified that the cart algorithm offered higher accuracy	Yadav <i>et al.</i> (2020)
[31]	A crime model is based on the SEMMA model and the manipulation of data, K-means are used for clustering before An artificial neural network is created with a perception structure	created ANN model provided positive results terms of forecasting as it was capable of predicting of crime statistics and occurrences based on the input data	Forradellas <i>et al.</i> (2020)
[32]	A systematic literature review is performed by consulting Several scientific databases including ACM, Scopus, and Springer, Science Direct and IEEE	Deep learning and time series analysis techniques are capable of offering positive results in predicting crime rates and crime prediction states are made more accurate by including temporal and spatial information in different crime datasets	Butt <i>et al.</i> (2020)
[33]	PRISMA guidelines are followed for reporting the systematic literature review and 32 papers are analyzed of techniques that were identified and selected from 786 papers	hotspots method is the most predominant type for prediction and forecasting. Mostly used methods were traditional ML methods and different approaches based on kernel density estimation. Common evaluation metrics include the Prediction Accuracy Index, prediction accuracy, and the F-1 Score. The train-test Split was the most common validation approach	Kounadi <i>et al.</i> (2020)
[34]	A spatiotemporal crime prediction method is proposed based on ML techniques and 2D Hotspot Analysis. A clustering algorithm has been used together with its	on A comparison of the model with conventional methods indicates a better performance when it relies on and uses hotspot analysis	Hajela <i>et al.</i> (2020)
[35]	Multiple ML algorithms are applied using KNIME, which enabled authors to compare the performance of various ML algorithms	The authors verified that random forests delivered the highest scores in the metrics of specificity, accuracy and precision for the fiscal data of 2018	Ippolito and Lozano (2020)
[36]	The authors used neural networks, random trees, and Bayesian networks and data mining for the analysis and prediction of crime rules from the acquired data. Data from 2015 to 2018 was used for the YD county	Through the comparison of results offered by algorithms, the authors determine that random trees delivered better outcomes and results than that of Bayesian Networks and neural networks	Wu <i>et al.</i> (2022)
[37]	An SBCPM or assemble-stacking-based crime prediction method that is based on SVM algorithms for the identification of predictions of crime using MATLAB. Additionally, the SVM algorithm is used and applied for the achievement of domain-specific configurations as compared to other models including the Random Forest, SMO Naïve Byes Bagging and J48 model	The proposed method acquires a classification accuracy of 99.5% and is identified to generate a more predictive effect than the baselines considered by others. It is also identified by the authors that the stacking ensemble model's accuracy is higher than that of any other individual classifier	Kshatri <i>et al.</i> (2021)
[38]	The study used multiple ML algorithms for predicting crime using crime data. A number of algorithms were tested including time series analysis, eXtreme gradient boosting, random forests, multilayer perceptron, decision trees, KNNs, Naïve Bayes, SVM, logistic regression and ARIMA model, and it is followed by KNN	The analysis of exploratory data predicts over 35 types crimes and indicated a yearly decline in the crime and rate in CHICAGO. The highest rate of efficiency is achieved by XGBoost for prediction accuracy	Safat <i>et al.</i> (2021)
[39]	The authors evaluated traditional deep learning techniques and compared results against SFTT as the proposed approach. In the SFTT framework, the data's spatial dimension is explored before the temporal structures' detection.	For data cell and classification, the SFTT approach tends to outperform traditional ML techniques and methods. Meanwhile, the TFTS approach tends to come in second place	Stalidis <i>et al.</i> (2021)
[40]	Long-term crime forecasts for robberies generated in Dallas at 200x200 grid cells that enable spatially changing	Random Forests outperformed kernel density estimation and Risk Terrain Models in terms of predicting	Wheeler and Steenbeck (2021)

**Table 1:** Continue

	associations of different demographic factors and crime and forecasting future crimes. Factors contributing to crime generators. After that, an interpretable model is used for facilitating the inner workings of the model	predictions that vary over space and are non-linear	
[41]	A comparative study of existing ML techniques is performed	The addition of a computer vision approach and machine learning are both capable of predicting crime rates and illustrating affected areas	Shah <i>et al.</i> (2021)
[42]	Crime data on 27 categories was acquired and obtained from Kaggle and diverse data analytics steps were applied. CRISP-DM methodology is used for the work. Linear regression technique is used for prediction	Linear regression is capable of predicting and forecasting crime rates to a significant extent	Obagbuwa and Abidoye (2021)
[43]	Transition probability, KNN, and K-clustering are used as a foundation of the proposed model and results are compared and With other ML-based methods,	The model proposed several features, which fed into Naïve Bayes, SVM, KNN, and random forests. Naïve Bayes was identified to produce the most successful results	Palanivinayagam <i>et al.</i> (2021)
[44]	Multiple ML-based algorithms are used for the prediction of crime in terms of severity and whether it would lead to is achieved by the neural network arrest. NN, AdaBoost, KNN, and random forests were developed. The models were tested on CLEAR by the Chicago police department	Among all the models, the highest accuracy of 90.77%	Tamir <i>et al.</i> (2021)
[45]	Naïve Bayes is used for the analysis of crime Data acquired from Kaggle	Naïve Bayes exhibits an accuracy of 93.07 with a precision of around 92.53%	Kanimozhi <i>et al.</i> (2021)
[46]	For the prediction of crimes, multiple ML algorithms are evaluated including impact learning, and K-Nearest neighbors, neural networks and decision trees	Of all the algorithms tested and evaluated, an optimal an accuracy of 81% is acquired with the use of algorithms based on decision trees. It is also determined that public Security is aided by using decision trees as algorithms	Adhikari <i>et al.</i> (2021)
[47]	There are multiple ML algorithms that have been considered and tested and Python has been used more accurately together with the algorithms	Clustering and decision trees are identified than other techniques that were tested.	Gupta <i>et al.</i> (2022)
[48]	120 research papers will be reviewed in 2021 and 2008 and the systematic review evaluates models comprehensively. 34 crime categories are studied by the researchers and 63 different ML techniques are studied	The supervised learning approach is the most the considered and used approaches and is capable of delivering precise outcomes	Dakalbab <i>et al.</i> (2022)
[49]	Data on human mobility and crime incidents is used for The evaluation of different neural networks and their ability to improve crime prediction	NbConv or neighbor convolution architectures that tend to model input features' spatiotemporal models produce and the best prediction accuracy. The use of mobility flow features in the absence of historical crime data are capable of improving the F1 scores for the model	Wu <i>et al.</i> (2022)
[50]	A systematic literature review was performed that focused primarily on 68 journal articles	A supervised machine-learning approach is considered and used by most authors	Jenga <i>et al.</i> (2023)

## Discussion

Generally, the adoption of machine learning algorithms for the prediction and detection of crime has demonstrated substantial promise in addressing the increasingly complex issue. Through the use of advanced algorithms and vast datasets, these technologies can enhance the effectiveness and accuracy of crime prediction models. On the basis of the results, there are a number of valuable insights that have been acquired. First, Random Forests have been considered and used the most by researchers and have been identified to offer quite accurate results. Second, all the studies, except for one study, showed the effectiveness and accuracy of ML algorithms and methods in predicting and forecasting crimes. Third, it is essential to actually perform thorough experiments and tests using the models and approaches discussed by the authors. Many authors have mentioned the need to test their models in the real world before using them. In addition, it might be challenging to use these models in a real situation due to the large amount of data that would be required to be considered and analyzed. Fourth, different models proposed by the authors tend to offer and produce varying results and findings in different situations using different types of datasets.

In addition, the authors have identified and faced several challenges while performing their experiments and training

their models. The very first challenge is associated with selecting the datasets. For instance, it is quite challenging for the researchers to select the right size of the dataset and then process it. For ML-based models and algorithms to work, clean and pre-processed data is required. When these models run on and use raw data, they do not produce promising results and their accuracy varies among different methods and approaches. The second challenge is associated with identifying and determining which factors influence and affect the accuracy and outcomes of algorithms.

Multiple factors influence and affect the accuracy of ML-based algorithms and models and these factors and features vary for each model and method. The third challenge is associated with using these algorithms and models in the real world. For instance, all the researchers have used the identified models and approaches in a simulated environment, not the real world. Due to this, there is no knowing how the studied models might perform in a real situation where raw data is required to be processed and analyzed for making predictions and forecasts. The fourth challenge is related to resource constraints, as testing and evaluating different models after training them requires a lot of time and excessive costs.

Even though machine learning algorithms are identified to support and even help with crime prediction, there is still a need for more implementable and interpretable models,

that can offer detailed and clear explanations of the predictions they make. It is particularly relevant and important for crime prediction because incorrect predictions can definitely lead to serious challenges and consequences for communities and individuals. Therefore, the fifth challenge is concerned with the lack of interpretable machine learning algorithms. Lastly, there is a need to go beyond just model-based prediction and explanation frameworks. It is also necessary to develop models that focus on identifying and highlighting the cause-and-effect relationship between different factors and crime patterns specific to each geographical area.

The current research study is close to the studies performed by Kounadi *et al.* (2020) in which the focus is on systematically reviewing studies on spatial crime forecasting and Mandalapu *et al.* (2023) in which the emphasis is on performing a systematic review on crime prediction using deep learning and machine learning (Kounadi *et al.*, 2020; Mandalapu *et al.*, 2023). Kounadi *et al.* (2020) determine that the hotspots method is the most predominant type of method for spatial forecasting. It was followed by traditional ML methods and kernel density estimation techniques. On the other hand, Mandalapu *et al.* (2023) highlight that there are still many challenges related to the use of ML and DL for crime prediction. Several challenges are highlighted and future directions are identified but there is a lack of focus on which techniques are used most commonly for crime prediction. In this sense, the current research study is unique in its findings.

## Conclusion

Overall, crimes have undoubtedly increased significantly in the past few years around the world. It has urged researchers and experts to study patterns and trends related to crime. At the same time, it has also allowed researchers to perform studies on the prediction of crime with the use of machine learning methods and techniques. In the past few years, a large number of studies have been performed to evaluate and identify the potential of different ML-based techniques and approaches, including Naïve Bayes, KNNs, ANN, Random Forest, and even Decision Trees. In this research, a comparative analysis of 51 existing research studies on the use of machine learning methods for crime prediction has been carried out. Some critical trends and findings have been obtained. At the same time, some challenges in performing these studies have been determined.

In the future, it is essential to evaluate the effectiveness of ML methods for crime prediction using larger datasets in real-world scenarios and then evaluate their accuracy and usefulness. In addition, it is important to identify different situations and scenarios

in which the accuracy of different ML methods is affected. Based on these insights, the most effective method should be proposed.

## Acknowledgment

The authors would like to thank King Abdulaziz University, which has supported and sponsored their studies and works.

## Funding Information

The authors received funding from King Abdulaziz University for this research work.

## Author's Contributions

**Abdulrahman Alsubayhin:** Gathered papers and data from many sources. Wrote a summary of each paper represented in a method used and, the result of the paper. Designed the outline of the paper tables and figures. Conducted literature search and performed comparative analysis.

**Muhammad Ramzan:** Provided valuable notes and comments to enhance the paper's quality. Reviewed the paper in general, sections and subsections correlation, verified method and strategy used.

**Bander Alzahrani:** Designed paper and work framework. Made a comprehensive revision to the collected papers and verified the used methods and techniques.

## Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

## References

- Adesola, F., Azeta, A., Oni, A., & Odun-Ayo, I. (2019). Systematic literature review of crime prediction and data mining. *Review of Computer Engineering Studies*, 6, 56-63.  
<https://doi.org/10.18280/rces.060302>
- Adhikari, M., Devkota, H. R., & Cesuroglu, T. (2021). Barriers to and facilitators of diabetes self-management practices in Rupandehi, Nepal- Multiple stakeholders' perspective. *BMC Public Health*, 21(1), 1269.  
<https://doi.org/10.1186/s12889-021-11308-4>
- Al Boni, M., & Gerber, M. S. (2016). Area-specific crime prediction models. 2016 15<sup>th</sup> *IEEE International Conference on Machine Learning and Applications (ICMLA)*, 671-76.  
<https://doi.org/10.1109/ICMLA.2016.0118>

- Aldossari, B. S., Alqahtani, F. M., Alshahrani, N. S., Alhammam, M. M., Alzamanan, R. M., Aslam, N., & Irfanullah, (2020). A comparative study of decision tree and naive bayes machine learning model for crime category prediction in Chicago. *Proceedings of 2020 the 6<sup>th</sup> International Conference on Computing and Data Engineering*, 34-38.  
<https://doi.org/10.1145/3379247.3379279>
- Almaw, A., & Kadam, K. (2018). Survey paper on crime prediction using ensemble approach. *International Journal of Pure and Applied Mathematics*, 118(8), 133-139. <https://www.acadpubl.eu/jsi/2018-118-7-9/articles/8/18.pdf>
- Alves, L., Ribeiro, H. V., & Rodrigues, F. (2018). Crime prediction through urban metrics and statistical learning. *Physica A: Statistical Mechanics and Its Applications*, 505.  
<https://doi.org/10.1016/j.physa.2018.03.084>
- Azeez, J., & Aravindhar, D. J. (2015, August). Hybrid approach to crime prediction using deep learning. *In 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1701-1710). IEEE.  
<https://doi.org/10.1109/ICACCI.2015.7275858>
- Aziz, R. M., Hussain, A., Sharma, P., & Kumar, P. (2022). Machine learning-based soft computing regression analysis approach for crime data prediction. *Karbala International Journal of Modern Science*, 8(1), 1-19.  
<https://www.iasj.net/iasj/download/8347851c22612f92>
- Babakura, A., Sulaiman, M. N. & Yusuf, M. A. (2014). Improved Method of Classification Algorithms for Crime Prediction. *2014 International Symposium on Biometrics and Security Technologies (ISBAST)*, 250-55.  
<https://doi.org/10.1109/ISBAST.2014.7013130>
- Butt, U. M., Letchmunan, S., Hassan, F. H., Ali, M., Baqir, A. & Sherazi, H. H. R. (2020). Spatio-Temporal Crime HotSpot Detection and Prediction: A Systematic Literature Review. *IEEE Access*, 8, 166553-74.  
<https://doi.org/10.1109/ACCESS.2020.3022808>
- Chun, S., Paturu, V., Yuan, S., Pathak, R., Atluri, V., & Adam, N. (2019). Crime prediction model using deep neural networks.  
<https://doi.org/10.1016/j.ssaho.2022.100342>
- Dakalbab, F., Talib, M. A., Waraga, O. A., Nassif, A. B., Abbas, S., & Nasir, Q. (2022). Artificial intelligence & crime prediction: A systematic literature review. *Social Sciences & Humanities Open*, 6(1), 100342.  
<https://doi.org/10.1016/j.ssaho.2022.100342>
- David, H. B., & Suruliandi, A. (2017). Survey on crime analysis and prediction using data mining techniques. *ICTACT Journal on Soft Computing*, 7, 1459-66.  
<https://doi.org/10.21917/ijsc.2017.0202>
- Dubey, N., & Chaturvedi, S. K. (2014). A survey paper on crime prediction technique using data mining. *Int. J. Eng. Res. Appl*, 4(3), 396-400.
- Forradellas, R., Nández Alonso, S., Jorge-Vázquez, J., & Rodriguez, M. (2020). Applied machine learning in social sciences: Neural networks and crime prediction. *Social Sciences*, 10.  
<https://doi.org/10.3390/socsci10010004>
- Grover, V., Adderley, R., & Bramer, M. (2006, December). Review of current crime prediction techniques. *In International Conference on Innovative Techniques and Applications of Artificial Intelligence* (pp. 233-237). London: Springer London.  
[https://doi.org/10.1007/978-1-84628-666-7\\_19](https://doi.org/10.1007/978-1-84628-666-7_19)
- Gupta, V. K., Shukla, S. K., & Rawat, R. S. (2022). Crime tracking system and people's safety in India using machine learning approaches. *International Journal of Modern Research*, 2(1), 1-7.
- Hajela, G., Chawla, M. & Rasool, A. (2020). A Clustering Based Hotspot Identification Approach For Crime Prediction. *Procedia Computer Science*, 167, 1462-70.  
<https://doi.org/10.1016/j.procs.2020.03.357>
- Hossain, S., Abtahee, A., Kashem, I., Hoque, M. M. & Sarker, I. H. (2020). Crime Prediction Using Spatio-Temporal Data. In N. Chaubey, S. Parikh, & K. Amin (Eds.), 1235, (pp.277-89).  
[https://doi.org/10.1007/978-981-15-6648-6\\_22](https://doi.org/10.1007/978-981-15-6648-6_22)
- Ippolito, A., & Lozano, A. C. G. (2020). Tax Crime Prediction with Machine Learning: A Case Study in the Municipality of São Paulo. *In ICEIS* (1) (pp. 452-459).
- Iqbal, R., Murad, M., Mustapha, A., Panahy, P. H. S., & Khanahmadliravi, N. (2013). An experimental study of classification algorithms for crime prediction. *Indian Journal of Science and Technology*, 6, 4219-25.  
<https://doi.org/10.17485/ijst/2013/v6i3.6>
- Jenga, K., Catal, C., & Kar, G. (2023). Machine learning in crime prediction. *Journal of Ambient Intelligence and Humanized Computing*, 14, 1-27.  
<https://doi.org/10.1007/s12652-023-04530-y>
- Kadar, C., & Pletikosa, I. (2018). Mining large-scale human mobility data for long-term crime prediction. *EPJ Data Science*, 7(1), 1-27.  
<https://doi.org/10.1140/epjds/s13688-018-0150-z>
- Kang, H. W., & Kang, H. B. (2017). Prediction of crime occurrence from multi-modal data using deep learning. *Plos One*, 12(4), e0176244.  
<https://doi.org/10.1371/journal.pone.0176244>
- Kanimozhi, N., Keerthana, N.V., Pavithra, G.S., Ranjitha, G., & Yuvarani, S. (2021). CRIME type and occurrence prediction using machine learning algorithm. *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 266-73.  
<https://doi.org/10.1109/ICAIS50930.2021.9395953>



- Kim, S., Joshi, P., Kalsi, P.S., Taheri, P., & Quasim, M.T. (2018). Crime analysis through machine learning. 2018 *IEEE 9<sup>th</sup> Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 415-20.  
<https://doi.org/10.1109/IEMCON.2018.8614828>
- Kounadi, O., Ristea, A., Araujo, A. & Leitner, M. (2020). A systematic review on spatial crime forecasting. *Crime Science*, 9(1), 7.  
<https://doi.org/10.1186/s40163-020-00116-7>
- 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, 67488-500.  
<https://doi.org/10.1109/ACCESS.2021.3075140>
- Liao, R., Wang, X., Li, L. & Qin, Z. (2010). A novel serial crime prediction model based on Bayesian learning theory. 2010 *International Conference on Machine Learning and Cybernetics*, 1757-62.  
<https://doi.org/10.1109/ICMLC.2010.5580971>
- Lim, M., Abdullah, A., Jhanjhi, N. & Khan, K., (2019). Link prediction in time-evolving criminal network with deep reinforcement learning technique. *IEEE Access*, pp. 1-1.  
<https://doi.org/10.1109/ACCESS.2019.2958873>
- Llaha, O. (2020, September). Crime analysis and prediction using machine learning. In *2020 43<sup>rd</sup> International Convention on Information, Communication and Electronic Technology (MIPRO)* (pp. 496-501). IEEE.  
<https://doi.org/10.23919/MIPRO48935.2020.9245120>
- 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*.  
<https://doi.org/10.1109/ACCESS.2023.3286344>
- McClendon, L., & Meghanathan, N. (2015). Using machine learning algorithms to analyze crime data. *Machine Learning and Applications: An International Journal*, 2, 1-12.  
<https://doi.org/10.5121/mlaj.2015.2101>
- Navalgund, U. V., & Priyadharshini, K. (2018). Crime intention detection system using deep learning. 2018 *International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET)*, 1-6.  
<https://doi.org/10.1109/ICCSDET.2018.8821168>
- Nguyen, T., Hatua, A., & Sung, A. (2017). Building a learning machine classifier with inadequate data for crime prediction. *Journal of Advances in Information Technology*, 141-47.  
<https://doi.org/10.12720/jait.8.2.141-147>
- Obagbuwa, I. C., & Abidoeye, A. P. (2021). South Africa crime visualization, trends analysis, and prediction using machine learning linear regression technique. *Applied Computational Intelligence and Soft Computing*, 2021, e5537902.  
<https://doi.org/10.1155/2021/5537902>
- Palanivinayagam, A., Gopal, S. S., Bhattacharya, S., Anumbe, N., Ibeke, E. & Biamba, C. (2021). An Optimized Machine Learning and Big Data Approach to Crime Detection. *Wireless Communications and Mobile Computing*, 2021, e5291528. <https://doi.org/10.1155/2021/5291528>
- Prabakaran, S. & Mitra, S. (2018). Survey of analysis of crime detection techniques using data mining and machine learning. *Journal of Physics: Conference Series*, 1000,012046.  
<https://doi.org/10.1088/1742-6596/1000/1/012046>
- Prathap, B. R. (2023). Geo-spatial crime density attribution using optimized machine learning algorithms. *International Journal of Information Technology*, 15(2), 1167-1178.  
<https://doi.org/10.1007/s41870-023-01160-7>
- Rajapakshe, C., Balasooriya, S., Dayarathna, H., Ranaweera, N., Walgampaya, N. & Pemadasa, N. (2019). Using CNNs RNNs and Machine Learning Algorithms for Real-Time Crime Prediction. 2019 *International Conference on Advancements in Computing (ICAC)*, 310-16.  
<https://doi.org/10.1109/ICAC49085.2019.9103425>
- 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), 400.  
<https://doi.org/10.3390/ijgi11070400>
- Shah, N., Bhagat, N., & Shah, M. (2021). Crime forecasting: A machine learning and computer vision approach to crime prediction and prevention. *Visual Computing for Industry, Biomedicine, and Art*, 4(1), 9.  
<https://doi.org/10.1186/s42492-021-00075-z>
- Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017). An overview on crime prediction methods. 2017 *6<sup>th</sup> ICT International Student Project Conference (ICT-ISPC)*, 1-5.  
<https://doi.org/10.1109/ICT-ISPC.2017.8075335>

- Shermila, A. M., Bellarmine, A. B., & Santiago, N. (2018, May). Crime data analysis and prediction of perpetrator identity using machine learning approach. *In 2018 2<sup>nd</sup> International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 107-114). IEEE. <https://doi.org/10.1109/ICOEI.2018.8553904>
- Stalidis, P., Semertzidis, T., & Daras, P. (2021). Examining deep learning architectures for crime classification and prediction. *Forecasting*, 3(4), 741-62. <https://doi.org/10.3390/forecast3040046>
- Stec, A., & Klabjan, D. (2018). Forecasting crime with deep learning. *arXiv preprint arXiv:1806.01486*. <https://doi.org/10.48550/arXiv.1806.01486>
- Tamir, A., Watson, E., Willett, B., Hasan, Q., & Yuan, J. S. (2021). Crime Prediction and Forecasting using Machine Learning Algorithms. *International Journal of Computer Science and Information Technologies*, 12(2), 26-33.
- ToppiReddy, H. Saini, B. & Mahajan, G. (2018). Crime prediction & monitoring framework based on spatial analysis. *Procedia Computer Science*, 132, 696-705. <https://doi.org/10.1016/j.procs.2018.05.075>
- Wang, B., Yin, P., Bertozzi, A. L., Brantingham, P. J., Osher, S. J. & Xin, J. (2019). Deep learning for real-time crime forecasting and its ternarization. *Chinese Annals of Mathematics, Series B*, 40(6),949-66. <https://doi.org/10.1007/s11401-019-0168-y>
- Wang, S., Cao, J., & Philip, S. Y. (2020). Deep learning for spatio-temporal data mining: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 34(8), 3681-3700. <https://doi.org/10.1109/TKDE.2020.3025580>
- Wheeler, A. & Steenbeek, W. (2021). Mapping the risk terrain for crime using machine learning. *Journal of Quantitative Criminology*, 37. <https://doi.org/10.1007/s10940-020-09457-7>
- William, P., Shrivastava, A., Shunmuga Karpagam, N., Mohanaprakash, T. A., Tongkachok, K., & Kumar, K. (2023). Crime analysis using computer vision approach with machine learning. *In Mobile Radio Communications and 5G Networks: Proceedings of Third MRCN 2022* (pp. 297-315). [https://doi.org/10.1007/978-981-19-7982-8\\_25](https://doi.org/10.1007/978-981-19-7982-8_25)
- Wu, J., Abrar, S. M., Awasthi, N., Frias-Martinez, E. & Frias-Martinez, V. (2022). Enhancing short-term crime prediction with human mobility flows and deep learning architectures. *EPJ Data Science*, 11(1), 1-20. <https://doi.org/10.1140/epjds/s13688-022-00366-2>
- Yadav, N., Kumar, A., Bhatnagar, R., & Verma, V. K. (2020). City crime mapping using machine learning techniques. *In The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019)* 4, (pp. 656-668). Springer International Publishing. [https://doi.org/10.1007/978-3-030-14118-9\\_65](https://doi.org/10.1007/978-3-030-14118-9_65)