



Machine learning in crime prediction

Karabo Jenga¹ · Cagatay Catal² · Gorkem Kar¹

Received: 13 April 2022 / Accepted: 10 January 2023 / Published online: 2 February 2023
© The Author(s) 2023

Abstract

Predicting crimes before they occur can save lives and losses of property. With the help of machine learning, many researchers have studied predicting crimes extensively. In this paper, we evaluate state-of-the-art crime prediction techniques that are available in the last decade, discuss possible challenges, and provide a discussion about the future work that could be conducted in the field of crime prediction. Although many works aim to predict crimes, the datasets they used and methods that are applied are numerous. Using a Systematic Literature Review (SLR) methodology, we aim to collect and synthesize the required knowledge regarding machine learning-based crime prediction and help both law enforcement authorities and scientists to mitigate and prevent future crime occurrences. We focus primarily on 68 selected machine learning papers that predict crime. We formulate eight research questions and observe that the majority of the papers used a supervised machine learning approach, assuming that there is prior labeled data, and however in some cases, there is no labeled data in real-world scenarios. We have also discussed the main challenges found while conducting some of the studies by the researchers. We consider that this research paves the way for further research to help governments and countries fight crime and decrease this for better safety and security.

Keywords Machine learning · Artificial Intelligence · Feature selection · Crime prediction · Neural networks · Crime analysis

1 Introduction

A crime is a form of violence or illegal act done by a perpetrator against another person that can cause harm or property damage and is punishable by the law of the governing state of authority in which the crime was carried out. Law authorities apply crime-solving techniques to take preventive measures but in many cases, they cannot deliver effective results (Dakalbab et al. 2022). Over the years, crimes have continued to increase within countries. In another study, it is stated that the top three countries with the highest crimes

as of 2021 were South Africa, Venezuela, and Papua New Guinea (Materoke et al. 2021). As the population of a country increases, crime rates within that country increase (Walczak 2021), this also impacts the accurate prediction of crime (Pratibha et al. 2020). The safety and security of a country are a very important part of its growth and economic development, crime prediction will decrease economic loss and increase public safety in exchange (ToppiReddy et al. 2018; Rumi et al. 2018). Governments are responsible for ensuring the safety of their citizens to control and maintain crime incidents, threats, and data being collected to better the lives of their citizens (Butt et al. 2020).

Data mining and ML are both versatile fields that involve the use of computers and mathematics where the programming is completed for the system to perform certain tasks, these are both important parts of crime prevention and detection (Bandekar and Vijayalakshmi 2020). Data mining can be considered as the process where discovers of new patterns from large data sets involving methods from statistics and AI, but also database management.

Different techniques have been developed for crime prediction and in recent years, more researchers are publishing

✉ Cagatay Catal
ccatal@qu.edu.qa

Karabo Jenga
karabo.jenga@bahcesehir.edu.tr

Gorkem Kar
gorkem.kar@eng.bau.edu.tr

¹ Department of Computer Engineering, Bahcesehir University, Istanbul, Turkey

² Department of Computer Science and Engineering, Qatar University, Doha, Qatar

papers and finding interest in the topic, which can be seen in the upward trend of publications between 2021 and 2022 in Fig. 2. Crime is a growing pandemic that contributes to social and economic issues which in turn negatively affect the development of a country and its communities (Matereke et al. 2021). Both social and economic elements, a.k.a., features, contribute to a perpetrator committing a crime or a victim being the subject of a crime, these features can be used to predict the future occurrence of future crimes and help decrease crime rates across many communities (Hajela et al. 2020).

The availability of enormous volume of data being made available by certain governments has given motivation to researchers to further pursue research in the field of crime. Historical data has made it an interesting subject that sparked attention in research, many researchers have proposed several different models for predicting the future occurrence of crimes (Pratibha et al. 2020). In some areas law authorities have restrictions over their data and may not make this available to researchers in the area, causing further frustration and disappointment.

Governments and police have access to a large amount of data that could be used to help reduce the crime rate (Yuki et al. 2019). Crime pattern theory also suggests that offenders prefer not to venture into unknown territories or areas and that they would rather commit opportunistic and violent crimes by taking advantage of familiar areas they have previously encountered (Jalil et al. 2017). Crimes are also not distributed evenly and uniformly nor are they random in an area or city. A hotspot is defined as an area or region where crimes would commonly take place. Based on this knowledge, it is worth noting that mapping crime hotspots can help researchers understand the reasons behind frequent crimes in a specific area (Kadar and Pletikosa 2018; Kadar et al. 2019).

Machine learning (ML) is a subfield of Artificial Intelligence being used across many different fields today to predict the future occurrence of certain events as well as better decision making. ML can be understood as the study of computer algorithms that can automatically improve on their own through experience/learning and by the use of data. Deep Learning (DL) is a subset of machine learning that is inspired by how our brains function, this technique is an artificial neural network that includes many different layers and layer types (e.g., pooling layer, convolution layer, fully-connected layer, dropout layer) that attempt to replicate the behavior of our brains. There exist four types of learning types, which are supervised, semi-supervised, unsupervised, and reinforcement learning. AI comprises both computer and mathematics (i.e., statistics) aspects where the programming is performed for the system to perform a certain action, commonly associated with humans (He and Zheng 2021).

To develop a highly accurate crime prediction model, it is important to understand the nature of a crime (Elluri et al. 2019). The nature of a crime could include features relating to the crime such as the offender(s) age, gender, location, number of offenders, education status, income, the weapon used, victim(s) age, gender, location, economic status, education status, time, date, day of the week, year, month, to name a few.

In Fig. 1, we have identified the objectives for each of the articles in our paper, the articles have been categorized into 6 main objectives namely: Social media crime prediction, Novel crime prediction, Suspect or Offender prediction, Spatial-temporal crime hotspot prediction, Feature selection, and Crime patterns and mapping. We need to understand the objectives of the articles to draw accurate findings from our research, understand the direction, gaps, and challenges faced by other researchers in the study of crime prediction, and motivate the research taken.

Researchers have made numerous amounts of contributions to crime investigation and prediction. Unlike most industries; health care, transportation, agriculture, finance, retail, and customer services, crime prediction has a lack of comprehensive and systematic literature reviews, which can help to organize, and summarize existing literature, evidence, and challenges they encounter. This SLR study focuses on the prediction of future occurrences of crime using machine learning techniques between the years 2010 and 2022, to answer the identified research questions and provide further research on the gaps that have been recognized in the field. We seek to broaden the opportunity for further research on crime investigation and machine learning and our motivation for this work was not to only publish research on previous and current studies on crime prediction, but to:

- Demonstrate the need for the availability of crime data to researchers
- Help to decrease the crime rate in our communities by synthesizing the existing knowledge
- Identify the gaps in current findings
- Help to improve existing models

The purpose of this paper is to provide a better understanding of ML algorithms used in crime prediction and analysis. Contributions to crime prediction are ongoing. This SLR aims to cover literature between the Jan 2010 and August 2022 and is guided by the work (Catal 2012; Lighthart et al. 2021; Kitchenham and Charters 2007). These guidelines are used widely in literature for conducting SLR.

The rest of the paper is organized into the following sections: Sect. 2 presents the related work. Section 3 discusses the adopted research methodology followed

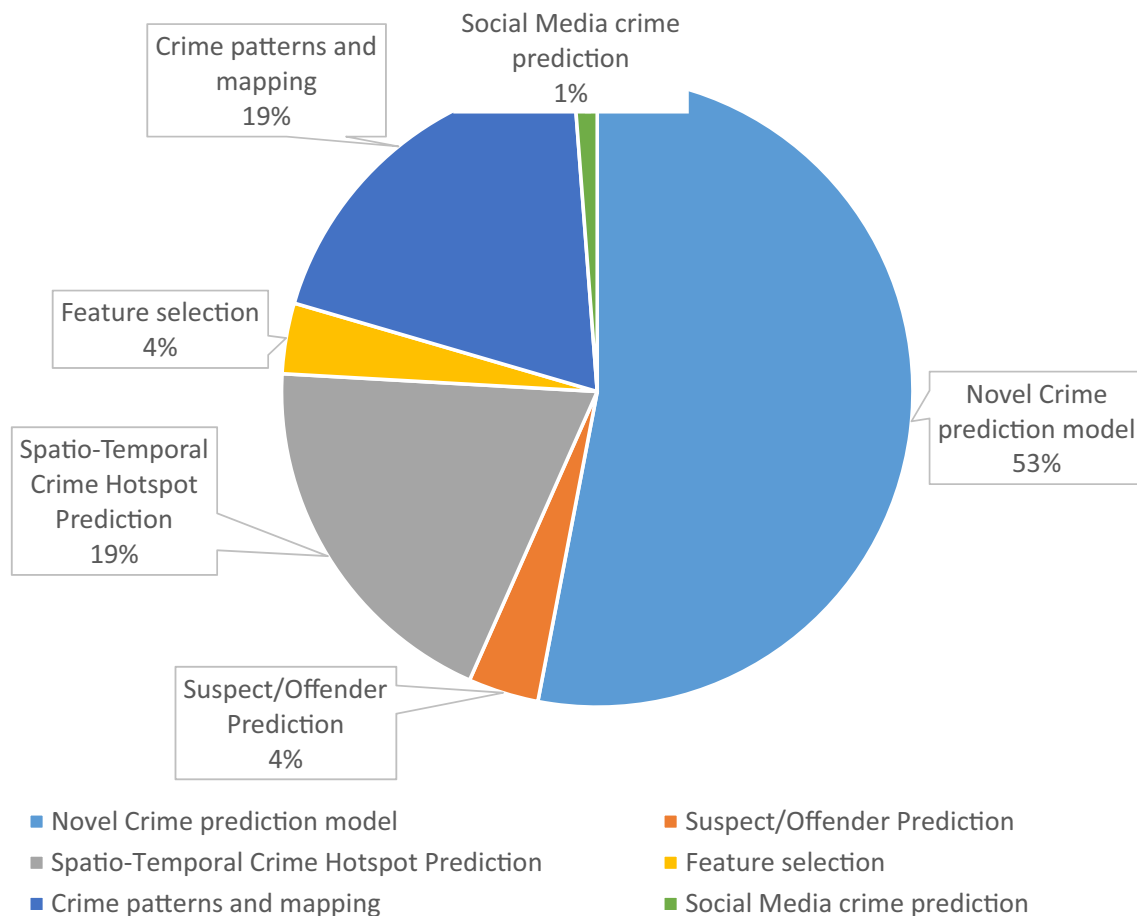


Fig. 1 Distribution of objectives

throughout the paper. Section 4 presents the results found in the SLR for the research questions we have identified. Section 5 discusses our findings and results in detail and finally, in Sect. 6 we present our conclusion and future work.

2 Related work

This study seeks to explore different machine learning algorithms and techniques used in crime prediction. Our findings are reported together with the challenges identified by the researchers to help gain a better understanding of the current state-of-the-art that is available to conduct further research and build better performing and more accurate models to fight crime.

A review (Falade et al. 2019) is conducted where they focused on crime prediction using data mining and further concluded that crime prediction using data mining is a hot research topic due to the impact of crime on the socio-economic development of a nation (Kounadi et al. 2020). They

stated that the data being used could be statistical reports of crimes in an area or region. They also note that crime prediction research will positively benefit society in assisting law enforcement agencies and governments understand the multiple layers that contribute to the cause of a crime. This research may also aid governments in making better decision-making approaches for the security and safety of their citizens and provide a more proactive approach towards the improvement of communities and decrease in crime.

In 2020, a review (Butt et al. 2020) is published that discusses the approaches of Spatio-temporal crime hotspot detection and prediction. The researchers state that the availability of large amounts of data that are being collected and made available to the public has made it more possible to conduct and pursue further research in the field of crime and crime investigation. The availability of historical data allows for the forecasting of future crimes, with it being a growing point of interest to develop significant machine learning models to assist in discovering different features that are related to crime prediction (Butt et al. 2020). It is also worth noting that the researchers proposed in their future

work to explore the enhancement of clustering approaches for crime hotspot detection (DBSCAN), enhancement of time series analysis for crime forecasting (ARIMA), exploration of transfer learning in crime prediction, and the use of LSTM with exponential smoothing for crime prediction. The researchers concluded by stating, that for a novel spatial–temporal model, datasets should be produced to improve the proposed methods and that regions or countries should have datasets available or should collect data to help scientists in the development of better crime prediction methods to decrease crime and grow the country.

There is another a paper on spatial crime forecasting, spatial referring to space and time (Kounadi et al. 2020). The researchers investigated crime both in time and space in this paper. They followed the “PRISMA” (Liberati et al. 2009) reporting process in their review to report their findings and gain knowledge and analyze the state-of-the-art techniques and methods in current studies of crime prediction. Some of the limitations outlined by the researchers in the study included the terminology, which is not consistent and could be due to the different academic backgrounds of the stakeholders (e.g., criminology, data analyst, computer science, geoscience, and public policy) and those significant details of some experiments are partly or not at all reported.

Our last related work-study (Ippolito and Lozano 2020) focuses on tax crime prediction using machine learning. It was noted as a significant paper as this crime is known as a white-collar crime, most papers in the study focused on violent crimes. The researchers seek to better decision-making in fiscal audit plans relating to service taxes in the municipality of São Paulo. In the study, the researchers applied Neural Networks, Naïve Bayes, Decision trees, Ensemble learning, Random forests, and Logistic regression. The researchers also conducted manual face-to-face data collection of fiscal audits and plans from 2016 to 2019. The results obtained showed that Random Forest provides the highest accuracy of 66.2%. Machine learning can better assist governments in making better decisions and plans around tax audits. Tax crime prediction allows governments to prepare for these crimes before they are committed. They conclude by stating that machine learning helps make predictions on crimes against law systems, and predictions can guide better decision-making and planning of fiscal audits more assertively.

Governments and law enforcement agencies are the gatekeepers to crime datasets (historical and present) and ensure their maintenance thereof (Falade et al. 2019). Accurate predictions of future crimes can have a positive impact on society and the economy (Butt et al. 2020). It is seen from Ippolito et al. that a crime detection system can alert users and make them cautious of the crime they are about to commit and the likelihood of them being caught, the implementation of such a system in society could leverage the same

impact on violent crime (Ippolito and Lozano 2020). Safer and more secure conditions lead to economic growth and sustainable development. The literature presents many different approaches to crime prediction using machine learning, compared to other review studies, to our knowledge this study includes recent papers and has different research questions applied. As such, the observations and suggestions are different than those of previous studies (Table 1).

3 Research methodology

Systematic literature reviews (SLRs) are and have been used in many different fields of academia from engineering to medicine to gather and summarize data on a certain research topic, we employed guidelines (Kitchenham et al. 2010) to follow a structured approach for our study. By using an SLR, we can also identify the challenges and possible solutions that can be employed. It defines the guidelines for the following three (3) main phases:

1. Planning the review: Gathering related data and research work related to our research topic (i.e., the use of machine learning in crime prediction), defining research questions and systematic search protocols that include the selection of keyword strings to be used for related papers, and how this criterion will be applied to the papers. The planning phase also includes the following phases:
 - a. Identification of the need for a review
 - b. Commissioning a review
 - c. Specifying the research question(s)
 - d. Developing a review protocol
 - e. Evaluating the review protocol
2. Conducting the review: Putting a classification schema in place, details on how the data and papers will be separated for analysis where features are grouped based on similar or common attributes. The papers are subject to inclusion and exclusion criteria. The papers that pass the criteria should meet the minimum quality assessment threshold, which is mostly selected as the mean value. Critical data from these papers are extracted and synthesized to present a general overview of the understanding of how machine learning can be used for crime prediction to identify possible gaps and opportunities in the selected field. The phases in the conducting phase include:
 - a. Identification of research
 - b. Selection of primary studies
 - c. Study quality assessment
 - d. Data extraction and monitoring

- e. Data synthesis
3. Reporting the review: The final step discusses and presents our findings, the research questions which were identified in step 1 are addressed and visually presented with graphs, figures, and tables β if needed and may include:
 - a. Specifying dissemination mechanisms
 - b. Formatting the main report
 - c. Evaluating the report

3.1 Research questions

The research questions that are investigated are presented as follows:

1. What are the objectives of the paper? To answer this question, we need to understand what the researcher aims to achieve. What is the desired output variable? (e.g., suspect detection, crime hotspot detection, etc.).
2. What data source types have been used to collect data?
3. What independent variables are (a.k.a., features) used in the publications?
4. Which ML algorithm(s) performed best in the study?
5. Which public datasets have been used?
6. Which evaluation metrics have been applied?
7. What are the top 5 ML categories?
8. What challenges have the authors identified?

3.2 Databases

We used five top reliable and trustworthy scientific databases to find relevant papers for our study. We did not select Google Scholar because it indexes non-peer-reviewed papers and sometimes non-reputable journals as well. For searching the databases, we applied search strings, which are defined in Sect. 3.4.

- ACM Digital Library (<https://dl.acm.org/>)
- IEEE Xplore (<http://ieeexplore.ieee.org>)
- Springer Link (<https://link.springer.com/>)
- Science Direct (<http://www.sciencedirect.com>)
- Scopus (<https://www.scopus.com/>)

As shown in Table 4, IEEE Xplore provides the largest number of papers before and after the exclusion criteria, ACM Digital library being the lowest. Forward and backward snowballing was also applied in this SLR study to find more relevant papers. The papers that were collected using these methods were included in the study selection criteria and those that passed were added to the list of selected publications (Fig. 2).

3.3 Search strategy

We applied search parameters to the databases to collect the relevant papers for the study. An advanced search was applied to the selected databases, which included the following terms: “crime”, “crime prediction”, and “machine learning”. To help broaden our search string parameters, we read the abstract, introduction, and conclusion sections of the recent papers to find synonyms that could be included in the search string. Some of the identified words included “Neural Networks”, “Artificial Intelligence”, “Data mining” and/or “Crime patterns”. The search string that was used within each database is presented in the following subsection.

3.4 Search strings

We had to adjust the search string for each database to gather a more detailed search on our related papers for the SLR.

These search strings shown in Table 2 were applied to the databases in the title, abstract, and keywords fields. In Science Direct, we searched for “Crime Prediction” AND “Machine Learning” with Year between 2010 – 2022, this would ensure that we retrieved all the latest papers for the last 10 years. IEEE Xplore returned the largest amount of papers that we could use for our study, it is also one of the largest online scientific research databases, we applied “Crime Prediction” to all metadata, and “Crime Prediction” AND “Machine Learning” were applied to the document title. Finally, Springer Link was used and the search string applied with the title “Crime Prediction” AND “Machine learning” and a year range of 2010 and before 2023. Figure 3 shows the yearly trend of the selected papers.

3.5 Selection criteria

We define the selection criteria in the first phase of the guidelines to decrease any possibility of bias in the selection of the publications for the study. It is also employed to ensure that we process relevant studies, in our case, we would define these studies as those that can assist us in answering the research questions. We establish a criterion before we begin with the study to avoid and reduce the chance of a biased criterion (Kitchenham et al. 2007). Papers should respond false to all the exclusion criteria and true to all the inclusion criteria.

The following criteria have been defined as the exclusion criteria:

1. Only abstract is available
2. The paper is not in English
3. The publication is a review/survey paper

Table 1 Related review studies

Title	Description	Year	Reference
Survey on crime analysis and prediction using Data mining techniques	To perform a survey on supervised and unsupervised ML techniques for criminal investigation	2017	David and Suruliandi (2017)
Survey of analysis of crime detection techniques using data mining and machine learning	A survey paper on different data mining techniques used in crime investigation	2018	Prabakaran and Mitra (2018)
A survey of research into artificial neural networks for crime prediction	The paper focuses only on the previous 8 years, to help identify the time and place of the next crime occurrence	2019	Abbas et al. (2019)
Systematic literature review of crime prediction and data mining	Crime prediction and data mining techniques to aid in the prediction of future occurrences of crime	2019	Falade et al. (2019)
Spatio-temporal Crime HotSpot detection and prediction: A systematic Literature Review	Aims to provide core, bases for research on spatio-temporal crime hotspot detection and prediction applications	2020	Butt et al. (2020)
A systematic review on spatial crime forecasting	The aim of the paper is to provide research and evaluate state-of-the-art spatial crime forecasting, more focused on the design and technical aspects	2020	Kounadi et al. (2020)
Tax Crime Prediction with Machine Learning: A Case Study in the Municipality of São Paulo	To use ML to aid in decision-making in fiscal audit plans when referring to service taxes in Sao Paulo	2020	Ippolito and Lozano (2020)
Predicting crime and other uses of neural networks in police decision making	Reviews and examines current research on the utilization of neural networks for forecasting crime and other police decision-making	2021	Steven Walczak (2021)
Survey on crime analysis and prediction using data mining and ML techniques	Explores opportunities in analyzing large data repositories, analyzing socioeconomic factors assigned with crime incidents, finding the outliers, categorizing patterns, and designing effective models to predict crimes using data mining and ML techniques	2021	Saravanan et al. (2021)
Comparative analysis of time series model and machine testing systems for crime forecasting	Uses time series and machine testing systems for crime estimation, predicting the number of crimes	2021	Jha et al. (2021)
A survey on crime analysis and prediction	Analyzes historical data using deep learning, statistical models, and algorithms, and identifies different approaches used to predict and forecast crime occurrences	2022	Thomas and Sobhana (2022)
Artificial intelligence and crime prediction: A systematic literature review	Evaluates models from different views; crime analysis type, crime studied types, prediction technique, performance metrics and evaluations, strength and weaknesses, limitation, and future works	2022	Dakalbab et al. (2022)

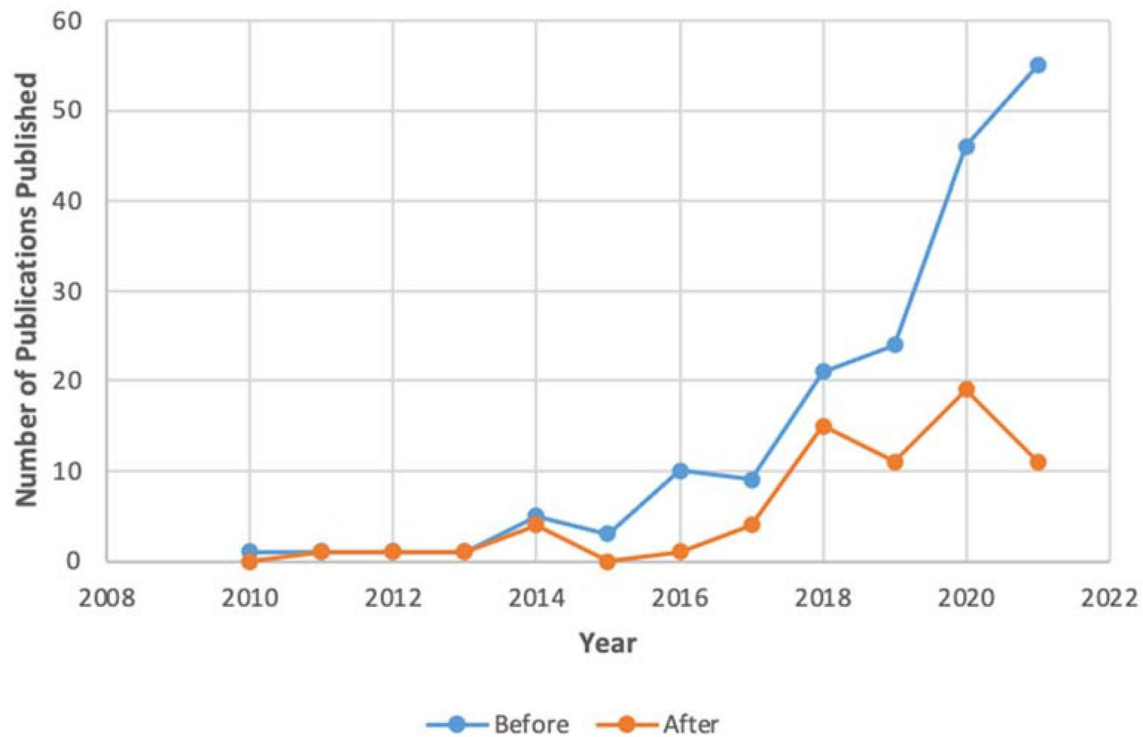


Fig. 2 Yearly trend of publications

4. The publication is duplication and already retrieved from another database
5. The paper does not explain in detail how Machine Learning was applied

3.6 Collecting and filtering

After collecting and finding various publications by applying the defined search strings to each database, we also used snowballing, the results obtained were then gathered in an excel spreadsheet to analyze and synthesize the data. All the papers that passed the selection criteria were subject to a quality assessment, which would enable us to narrow down the publications further, for high-quality papers to

be included as primary studies. It defines eight (8) quality assessment questions which we applied and then further scored a paper with 1 (yes/non-compliant), 0 (no/comply), and 0.5 (somewhat) against each question. If the score of the related paper is lower than 4.5 that paper was excluded, if that paper scored above 4.5, it was included. The quality assessment questions were defined as follows:

1. Are the aims of the study clearly defined?
2. Are the scope and experimental design of the study defined?
3. Is the technology assessed or used clearly defined?
4. Is the research process clearly defined?
5. Are all the study questions answered?

Table 2 Databases and the search strings used for the study

Database	Search string
ACM Digital Library	27 results for: [[Title: "crime"] OR [[Title: "crime prediction"] AND [Title: "machine learning"]]] AND [[Abstract: "crime prediction"] OR [Abstract: "machine learning"]] AND [Publication Date: (01/01/2010 TO 08/31/2022)]
IEEE Xplore	116 for: ("All Metadata": crime prediction) AND ("Document Title": machine learning) AND ("Document Title": crime OR ("Document Title": crime prediction)) Filters applied: 2010–2023
Springer Link	25 Result(s) for: 'Crime prediction' AND 'Machine learning' within 2010–2022
Science Direct	108 results for: Year: 2010–2022 Title: Crime OR Crime prediction AND Machine learning
Scopus	77 document results: TITLE ("CRIME" OR "CRIME PREDICTION" AND "MACHINE LEARNING") PUB-YEAR > 2010 PUBYEAR < 2023 KEY ("crime" AND "machine learning")

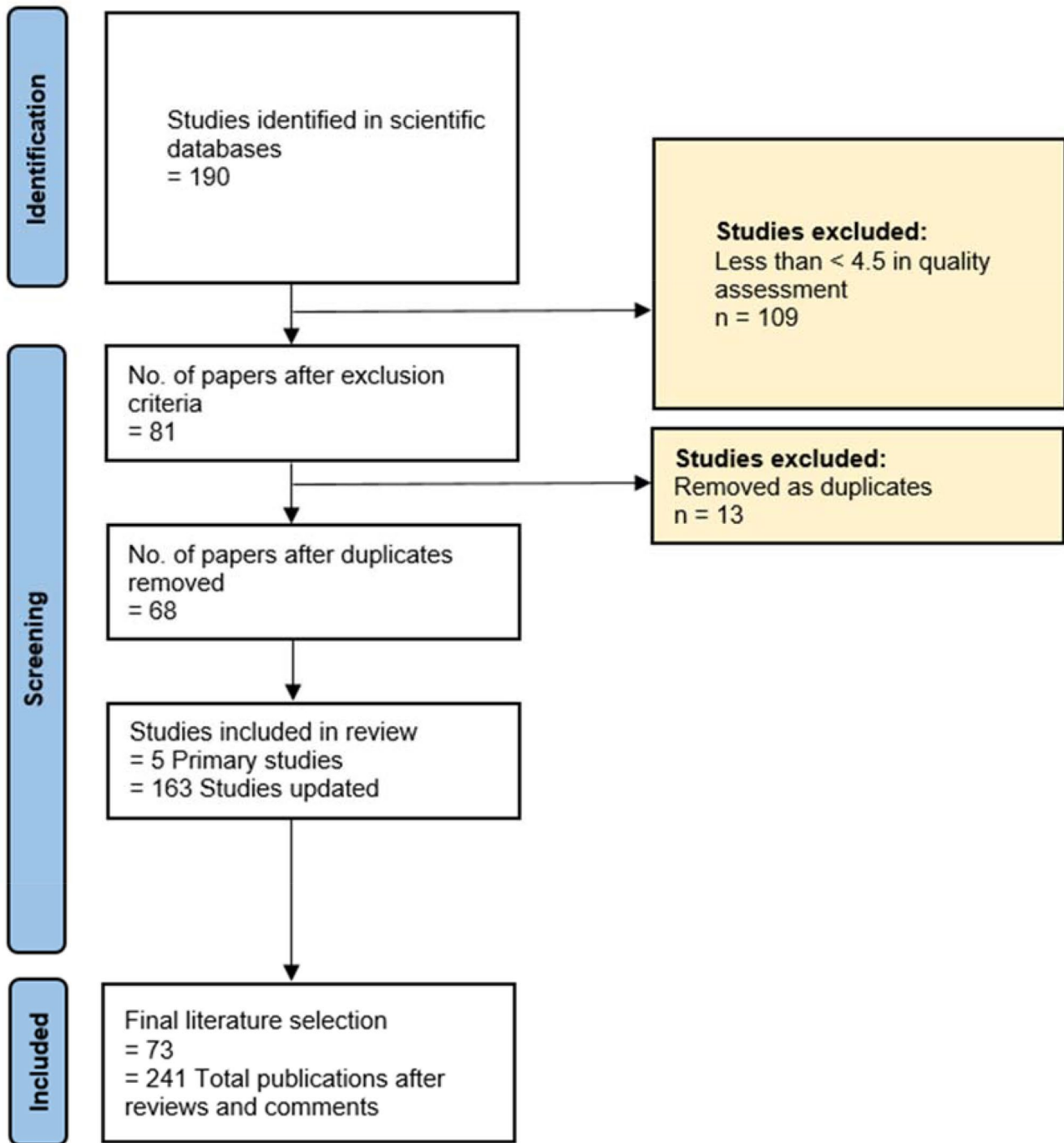


Fig. 3 Search methodology applied to identify relevant papers

6. Are the challenges, limitations, and negative findings clearly defined?
7. Are the main findings on the creditability, validity, and reliability stated?
8. Does the conclusion relate to the aim of the purpose of the study?

3.7 Data extraction, synthesis, and reporting

To answer the research questions with the most reliable data, all publications needed to pass the quality assessment to ensure that only reliable and trustworthy publications had been included in this next phase. All the relevant data were

collected from primary studies we read the articles in full to gather all the relevant data needed. We assigned each publication to a specific row and each answer was assigned to a different column in the extraction process. We found that with some of the research questions, some of the answers and features which returned were categorized into fewer groups to allow for better handling of the data. For research question 1 we identified the following categories: Novel crime prediction model, suspect/offender prediction, Spatio-temporal crime hotspot prediction, feature selection, crime patterns, and mapping, and social media crime prediction. In research question 2, we found that some answers were synonyms and we then combined them (e.g., meteorological and weather are both weather data features, social media, websites, and newspapers also are categorized into one column. In research question 3, we broke down the independent variables into crime parameters, data parameters, time parameters, offender parameters, victim parameters, and location parameters. The algorithms in research question 4 were categorized into decision-tree algorithms, regression algorithms, neural network algorithms, and others (e.g., Naive Bayes, and Support Vector Machines). Lastly, the final step in the SLR process is to report on our findings and answer all of the research questions.

4 Results

In this section, we present our findings using various visual aid graphs, and tables to illustrate all findings effectively. In total, 353 publications were found across all databases, after we applied the study selection criteria, and the quality assessment, 68 publications remained as relevant publications that we could then use throughout our study. Table 3 represents all the 68 primary studies found. Table 4 shows the number of publications retrieved after the exclusion criteria. The research questions that were identified are addressed in the following section in detail one after the other. In Fig. 3, we have applied the Prisma flow process, which is a technique used for reporting systematic reviews and meta-analyses (Kounadi et al. 2020).

4.1 Research Question 1—what are the objectives of the paper?

We first need to understand what the researchers are trying to achieve through the study. What is their desired output variable? What do they intend to predict? (e.g., suspect detection, developing a novel crime prediction model). We found that several different researchers would have different objectives and to manage this data efficiently, we would need to group

our findings into more narrow / detailed categories. We then identified 6 categories in which the publication could be grouped as the researchers found a common objective within their target variable. To find these categories, we need to read all the publications. The objectives have been categorized into novel crime category prediction model included all papers that sought to find a crime prediction model to detect the future occurrence of crimes, the researcher did not focus on a single or common crime type as stated by (Yu et al. 2011) crime analysis is done by using historical data to predict the time and place where a crime could take place. Suspect / offender prediction used technology to predict the future occurrence of an offender committing a crime using location data based on the fact that offenders or crime committers do not venture into crimes in new locations and thus would repeat crimes in common areas. Only three papers focused on offender prediction. Spatio-temporal crime hotspot is the study of crime using space and time, researchers focused on space (i.e., location parameters) and time (i.e., date parameters) to build models that could predict crime using space and time, the likely location of the potential crime (Zhang et al. 2020), hotspots and cold spots have an unbalance in data and that is because cold spots are much more prevalent than hotspots. Feature selection, knowing the right type of features and variables to use is a vital part of the machine learning process and could aid in the development of an accurate predictive model. The crime patterns and mapping category focused on understanding different crime types and how certain crimes occurred in various areas or locations to aid in the prediction of future crimes. Social media crime prediction uses Natural Language Processing (NLP) methods and textual data from social platforms, newspapers, etc. to aid in predicting crimes. Figure 3 shows the distribution of these objectives.

4.2 Research Question 2—what data source types have been used to collect data?

Due to the sensitivity of the data and the information held within the datasets, five authors did not report on the data source type that they used to collect the data resulting in five Not Applicable (NA) values. We further divided the rest of our findings into five categories: Actual crime records (datasets collected and maintained by governments and law enforcement agencies), traffic data (data from traffic segments or taxi flow data), location data (data from points of interest locations), visual data (data collected by using CCTV surveillance systems), text data (data from social media platforms, newspapers and other text data sources). All categories are presented in Fig. 4.

Table 3 Primary studies

Title	ML Technique(s) used	Best Performing algorithm(s)	References	Dataset		Dataset source
				Private	Public	
A Clustering Based Hotspot Identification Approach For Crime Prediction	Naive Bayes Decision Tree	Decision Tree	Hajela et al. (2020)	x		https://www.kaggle.com/datasets/kaggle/san-francisco-crime-classification
A Comparative Evaluation of Spatio Temporal Deep Learning Techniques for Crime Prediction	Deep learning: Spatio Temporal Residual Network (ST-ResNet) Deep Multi-View Spatio Temporal Network (DMVST-Net) Spatio-Temporal Dynamic Network (STD-Net)	STD-Net	Materke et al. (2021)	x		https://data.cityofchicago.org
A comparative study of decision tree and Naive Bayes machine learning model for crime category prediction in Chicago	Naive Bayes Decision Tree	Decision Tree	Aldossari et al. (2020)	x		https://data.cityofchicago.org
A Comparative Study to Evaluate Filtering Methods for Crime Data Feature Selection	Naive Bayes Support Vector Machine Logistic Regression Random Forest Random Tree Multi-class classifier Decision Stump	Random Forest	Jalil et al. (2017)	x		Communities and crime dataset using WEKA, a data mining tool (software)
A data-driven agent-based simulation to predict crime patterns in an urban environment	Decision Tree	Decision Tree	Rosés et al. (2021)	x		NYPD complaint data, Census tract and street segment data, LBSN data, Taxi trip data, weather data, land use information, population density, points of interest, calls for service, public transportation stops, tree census data, and school locations
A Neural-Network-Based Model of Charge Prediction via the Judicial Interpretation of Crimes	Neural Network CNN RNN	LSTM -Neural network	Li et al. (2020)	x		China Find Law Online
An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach	Support Vector Machine J48 Naive Bayes Random Forest	Ensemble-stacking-based crime prediction method – SVM algorithms	Kshatri et al. (2021)	x		India National Crime record bureau
Analysis and Classification of Crime Tweets	Naive Bayes Random Forest J48 ZeroR	Random Forest	Lal et al. (2020)	x		Twitter.com

Table 3 (continued)

Title	ML Technique(s) used	Best Performing algorithm(s)	References	Dataset		Dataset source
				Private	Public	
Applied machine learning in social sciences: Neural networks and crime prediction	Neural Network	Neural Network	Forradellas et al. (2020)	x	x	City of Buenos Aires
Behavior analysis and crime prediction using big data and machine learning	Naïve Bayes	Naïve Bayes	Jha et al. (2019)	x		https://data.gov.in/data https://ncrb.gov.in/
Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data	Neural Networks	LSTM—Neural network	Feng et al. (2019)	X		https://data.sfgov.org/ https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/fjzp-q8t2 https://www.opendataphilly.org/dataset/crime-incidents
Comparison of Machine Learning Algorithms for Predicting Crime Hotspots	Random Forest KNN Algorithm Support Vector Machine LSTM Naïve Bayes CNN	LSTM	Zhang et al. (2020)	x		Coastal city in the southeast of China
Crime analysis and prediction using data mining	Naïve Bayes	Naïve Bayes	Sathyadevan et al. (2014)	x		Data collected from various web-sites
Crime analysis and prediction using machine learning	Decision Tree Neural Network Naïve Bayes Support Vector Machine	Decision Tree	Llaha (2020)	x		Communities and crime dataset using WEKA
Crime Analysis and Prediction using Optimized K-Means Algorithm	K-means	K-means	Krishnendu et al. (2020)	x		Crimes in India between 2001 and 2011
Crime Analysis Through Machine Learning	KNN Decision Tree	Decision Tree	Kim et al. (2018)	x		Vancouver Police Department between 2003 and 2018
Crime Anatomization Using QGIS	K-means SVM Random Forest	K-means	Joshi et al. (2019)	x		Different NGOs and police stations were contacted in order to obtain adequate datasets in the City of Mumbai
Crime Data Analysis and Prediction of Perpetrator Identity Using Machine Learning Approach	Multilinear Regression KNN Neural Network	KNN	Shermila et al. (2018)	x		San Francisco Homicide dataset (1981–2014)

Table 3 (continued)

Title	ML Technique(s) used	Best Performing algorithm(s)	References	Dataset		Dataset source
				Private	Public	
Crime event prediction with dynamic features	Random Forest Neural network SVM Logistic Regression	Random Forest	Rumi et al. (2018)	x	x	Brisbane, the capital city of Queensland, Australia New York City, USA
Crime Forecasting Using Data Mining Techniques	SVM J48 KNN Neural Network Naïve Bayes	Neural Network	Yu et al. (2011)	x	x	Collected from various agencies
Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention	Decision Tree (J48) KNN Naïve Bayes	Naïve Bayes	Shah et al. (2021)	x	x	NA
Crime Hotspot Detection With Clustering Algorithm Using Data Mining	K-means	K-means	Aarthi et al. (2021)	x	x	Public survey between 2016 and 2018
Crime Prediction & Monitoring Framework Based on Spatial Analysis	KNN Naïve Bayes	KNN	ToppiReddy et al. (2018)	x	x	https://data.police.uk/data/
Crime Prediction and Analysis	K-means	K-means	Pratibha et al. (2020)	x	x	UCI
Crime prediction by data-driven Green's function method	Prospective Hotspot maps Expectation Maximization Data-driven Greens function	Data-driven greens function	Kajita and Kajita (2020)	x	x	https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2
Crime prediction through urban metrics and statistical learning	Random Forest	Random Forest	Alves et al. (2018)	x	x	Brazilian Public Health System
Crime Prediction Using Multiple-ANFIS Architecture and Spatiotemporal Data	Fuzzy Inference systems Adaptive Neuro-Fuzzy Inference systems	Adaptive Neuro-Fuzzy Inference system	Islam et al. (2018)	x	x	Dhaka Metropolitan Police (DMP) website
Crime prediction using patterns and context	NA	NA	Baloian et al. (2017)	x	x	Chilean Police
Crime Prediction Using Spatio-Temporal Data	Random Forest Decision Tree KNN Adaboost	Random Forest	Hossain et al. (2020)	x	x	https://data.sfgov.org/
Crime rate prediction in the urban environment using social factors	Linear regression Logistic regression Gradient boosting	Gradient boosting	Ingilevich and Ivanov (2018)	x	x	Ministry of Internal Affairs of the Russian Federation
Crime Rate Prediction Using Machine Learning and Data Mining	KNN Linear regression Naïve Bayes	KNN	Mahmud et al. (2021)	x	x	Bangladesh

Table 3 (continued)

Title	ML Technique(s) used	Best Performing algorithm(s)	References	Dataset		Dataset source
				Private	Public	
Crime risk analysis through big data algorithm with urban metrics	Statistical models Ordinary Least Square Linear regression Random Forest Artificial Neural Network	Least Absolute Shrinkage and Selection Operator Extra—Trees	Wang et al. (2020)	x		Chinese Ministry of Public Security
Crime Scene Prediction by Detecting Threatening Objects Using Convolutional Neural Network	CNN	CNN	Nakib et al. (2018)	x		Pictures of knives, guns (short gun, revolver and machine gun) and blood
Crime Spatiotemporal Prediction With Fused Objective Function in Time Delay Neural Network	Artificial Neural Network Time Delay Neural Network Nonlinear Autoregressive with Exogenous Input	Nonlinear Autoregressive with Exogenous Input	Ghazvini et al. (2020)	x		Malaysian crime spatial dataset Three benchmark time series
Crime Prediction Model using Deep Neural Networks	Deep Neural Network	Deep Neural Network	Chun et al. (2019)	x		Historical booking data between 1997 to 2017 from the Korean Government
Criminal shortlisting and crime forecasting based on modus operandi	K-means Hierarchical clustering with Average linking	Hierarchical with mean linking	Munasinghe et al. (2014)	x		Sri Lanka Police
DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction	Deep Neural Network	Deep Neural Network	Huang et al. (2018)	x		New York City Police Department
Design and Analysis of Machine Learning Algorithms for the reduction of crime rates in India	Statistical Analysis Bayesian Neural Network Levenberg Algorithm Scaled Algorithm	Statistical Analysis	Bandekar and Vijayalakshmi (2020)	x		India National Crime record bureau
Developing Machine Learning-Based Predictive Models for Smart Policing	Deep learning AutoMLP (Multilayer Perceptron) Decision Tree Logistic Regression Random Forest Neural Network SVM	LSTM	Elluri et al. (2019)	x		New York City Police Department
Dynamics of crime activities in the network of city community areas	KNN Decision Tree Naïve Bayes SVM	Decision Tree	Niu et al. (2019)	x		Citizen Law Enforcement Analysis and Reporting (CLEAR) system—City of Chicago – 2002 to 2017) Census Data – 2000, 2008, 2013, and 2017
Examining the impact of cross-domain learning on crime prediction	Gradient boosting TrAdaBoost TrResampling	Gradient boosting	Bappee et al. (2021)	x		http://data.torontopolice.on.ca/

Table 3 (continued)

Title	ML Technique(s) used	Best Performing algorithm(s)	References	Dataset		Dataset source
				Private	Public	
Extracting crime information from online newspaper articles	Conditional Random Field (CRF)	Conditional Random Field (CRF)	Arulanandam et al. (2014)	x		Otago Daily Times New Zealand Herald
The improved method of classification algorithms for crime prediction	Naïve Bayes Back Propagation	Naïve Baye	Babakura et al. (2014)	x		1990 US census Law enforcement data from 1990 US LEMAS survey 1995 FBI UCR
Improving crime count forecasts using Twitter and taxi data	Random Forest Gradient Boosting Feed-forward Artificial Neural Network	Artificial Neural Network	Vomfell et al. (2018)	x		Twitter.com Taxi Data in New York US Census Bureau
Learning to Detect Patterns of Crime	Series finder	Series finder	Wang et al. (2013)	x		Cambridge Police department – House break ins
Machine learning-based analytical approach for geographical analysis and prediction of Boston City crime using a geospatial dataset	Decision Tree Information gain Gini impurity Chi-square Random Forest	Random Forest with PCA (Principal Component Analysis)	Sharma et al. (2021)	x		Boston Kaggle crimes dataset
Mining large-scale human mobility data for long-term crime prediction	Random Forest Extra Trees Gradient Boosting	Extra Trees	Kadar et al. (2019)	x		New York City – Crime data, Census data, Foursquare venues data, Subway usage data, Taxi usage data,
Mixed spatial-temporal characteristics based on Crime Hot Spots Prediction	KNN	KNN	Zhang et al. (2016)	x		Crimes in Nanchang, China
Multi-UAV Allocation Framework for Predictive Crime Deterrence and Data Acquisition	Neural Network	Neural Network	Miyano et al. (2020)	x		Data from audio and visual sensors in residential and commercial areas
Optimization of criminal hotspots based on underlying crime controlling factors using geospatial discriminative pattern	Geospatial Discriminative Patterns	Geospatial Discriminative Patterns	Wang et al. (2012)	x		A north-eastern city in the United States
Predicting Crime Using Time and Location Data	Random Forest Decision Tree Extra Trees Bagging AdaBoost	Bagging	Yuki et al. (2019)	x		Chicago crime dataset
Prediction Analysis of Crime in India Using a Hybrid Clustering Approach	Naïve Bayes KNN	Naïve Bayes	Kiran and Kaishveen (2018)	x		NA

Table 3 (continued)

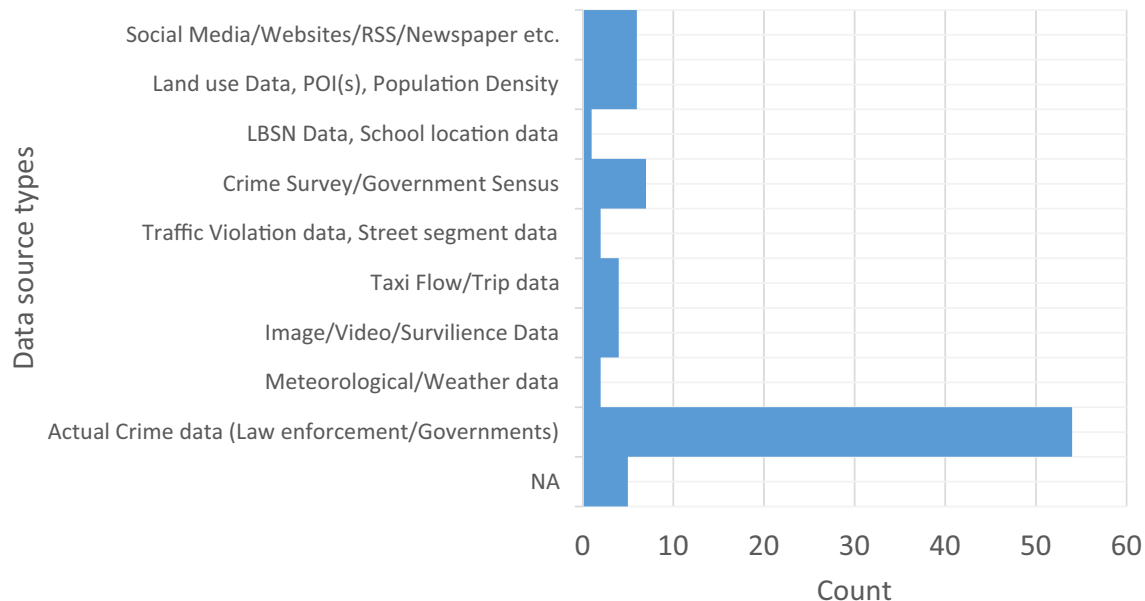
Title	ML Technique(s) used	Best Performing algorithm(s)	References	Dataset		Dataset source
				Private	Public	
Prediction of Crime Hotspots based on Spatial Factors of Random Forest	Random Forest	Random Forest	Yao et al. (2020)	x		San Francisco Dataset
Prediction of Crime Occurrence using Information Propagation Model and Gaussian Process	Information Propagation	Information Propagation	Morimoto et al. (2019)		x	Kobe crime dataset Boston crime dataset
Prediction of crime rate in urban neighbourhoods based on machine learning	GAN Neural Network	GAN Neural Network	He and Zheng (2021)	x		Philadelphia crime data from 2006 to 2018
Predictive Policing in the Context of Residential Burglary: An Empirical Illustration based on a Pilot Project in Baden-Württemberg, Germany	Near-repeat ratio	Near-repeat ratio	Gerstner (2018)		x	City of Karlsruhe
Public decision support for low population density areas: An imbalance-aware hyper-ensemble for Spatio-temporal crime prediction	Logistic Regression Bagging Boosting	Logistic Regression Bagging Boosting	Kadar et al. (2019)		x	Burglary incidents in the Swiss canton of Aargau between January 14, 2014 to January 12, 2017
Risk Prediction of Theft Crimes in Urban Communities: An Integrated Model of LSTM and ST-GCN	LSTM Spatial-Temporal Graph Convolutional Network	LSTM Spatial-Temporal Graph Convolutional Network	Han et al. (2020)	x		Chicago crime data
Smart Policing Technique with Crime Type and Risk Score Prediction Based on Machine Learning for Early Awareness of Risk Situation	DNN CNN SVM Naïve Bayes	DNN CNN	Baek et al. (2021)	x		Korean police security data from KICS
Spatio-Temporal Crime Predictions by Leveraging Artificial Intelligence for Citizen's Security in Smart Cities	Seasonal Auto-regressive Integrated Moving Average (SARIMA) Hierarchical Density-Based Spatial Clustering of Applications with Noise	Hierarchical Density-Based Spatial Clustering of Applications with Noise	Butt et al. (2021)	x		New York City between 2008 and 2017
Spatio-Temporal Prediction of Baltimore Crime Events Using CLSTM Neural Networks	CNN + LSTM	CNN + LSTM	Esquivel et al. (2020)	x		City of Baltimore
StreetNet: preference learning with a convolutional neural network on urban crime perception	CNN	CNN	Fu et al. (2018)	x		Street view image

Table 3 (continued)

Title	ML Technique(s) used	Best Performing algorithm(s)	References	Dataset		Dataset source
				Private	Public	
The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context	Logistic Regression Neural Network	Logistic regression + Neural Network	Rummens et al. (2017)	x		Crime anticipation system
Towards a Crime Hotspot Detection Framework for Patrol Planning	Random Forest Multi-Layer Perceptron KNN	KNN	Araujo et al. (2018)	x		Natal, Brazil – Crime data, Exogenous data sources
Unsupervised identification of crime problems from police free-text data	Latent Dirichlet Allocation	Latent Dirichlet Allocation	Birks et al. (2020)	x		Crime records in UK
Using CNN's RNNs and Machine Learning Algorithms for Real-time Crime Prediction	CNN RNN SVM Logistic Regression Decision Tree Random Forest Gradient Boosting	Gradient Boosting	Rajapakshe et al. (2019)	x		Sri Lanka, videos clips
Using Machine Learning to Assist Crime Prevention	Random Forest Naïve Bayes Deep Learning	Deep learning	Lin et al. (2017)	x		
Predicting Crime and Other Uses of Neural Networks	Neural Network	Neural Network	Walczak (2021)	x		City of Detroit— https://data.detroitmi.gov/datasets/rms-crime-incidents

Table 4 Number of publications retrieved before and after exclusion criteria per database

Databases	Publications before exclusion criteria	Publications after exclusion criteria
ACM Digital Library (https://dl.acm.org/)	27	6
IEEE Xplore (http://ieeexplore.ieee.org)	116	29
Springer Link (https://link.springer.com/)	25	8
Science Direct (http://www.sciencedirect.com)	108	13
Scopus (https://www.scopus.com/)	77	11
Total	353	68

**Fig. 4** Data source types used

4.3 Research Question 3—what independent variables are (a.k.a features) used in the publications?

A total of 44 independent variables were found, 5 of them were identified as common variables: Crime ID, date, crime type, latitude, and longitude. 23 of the publications did not clarify the independent variables they used to achieve their objectives. The remaining independent variables were then grouped into the following categories to better understand and easily separate them across the study: (1) Crime parameters that contain information regarding the crime that had occurred, (2) Date parameters that contain information regarding the date like the day of the week, week, year, month, etc., (3) Time parameters that contain information on the time the crime took place, (4) Offender parameters that contain information regarding the offender of the crime age, gender, address, income, etc., and (5) Victim parameters that represent information regarding the offender of the crime age, gender, address, income, etc. and (6) Location

parameters that present information regarding the location of the crime. Table 5 shows the distribution of these categories. Based on this table, the most-used category is the location and the second most used one is related to the crime parameters category. The % of papers column represents the % of how many times the variables appeared in the papers, the % will be above 100% due to the fact the researchers use more than one variable in some studies. E.g. Crime parameters appeared 75% of the time in the papers, however, had an overall appearance of 24% out of 100% (Fig. 5).

4.4 Research question 4—what ML algorithm(s) performed best in the study?

The study identified 14 different machine learning algorithms. Each publication stated the best algorithms they used and various evaluation metrics to conclude the best-performing algorithm. If a publication only applied one algorithm, that algorithm was then noted as the best-performing algorithm from the study. We grouped the best-performing algorithms

Table 5 Independent variables by category and their distribution

Category	Independent Variable	% Appearance of variable in papers	% out of 100
Crime parameters	Record ID, Crime category, description, police district, resolution	75	24
Date parameters	Date, year, month, day of the week	61	20
Time parameters	Time, the hour of the day	18	6
Location parameters	Address, longitude, latitude, location description, home location of offender, home location of the victim, state, weather conditions	96	31
Offender parameters	Offender age, offender sex, offender income, education status, no. of offenders, gang activity, family structure, weapon	39	13
Victim parameters	Victim age, victim sex, victim race, the victim count	20	6

Table 6 Best performing algorithms by category

Algorithm category	Algorithms	# of papers
Decision Tree-based	Random Forest, Decision Tree, Extra Trees, Gradient Boosting	16
Regression algorithms	Support Vector Regression	3
Artificial Neural Networks	Long Short Term Memory, Deep Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks	24
Others	K-means, K-nearest neighbor, Naïve Bayes	14

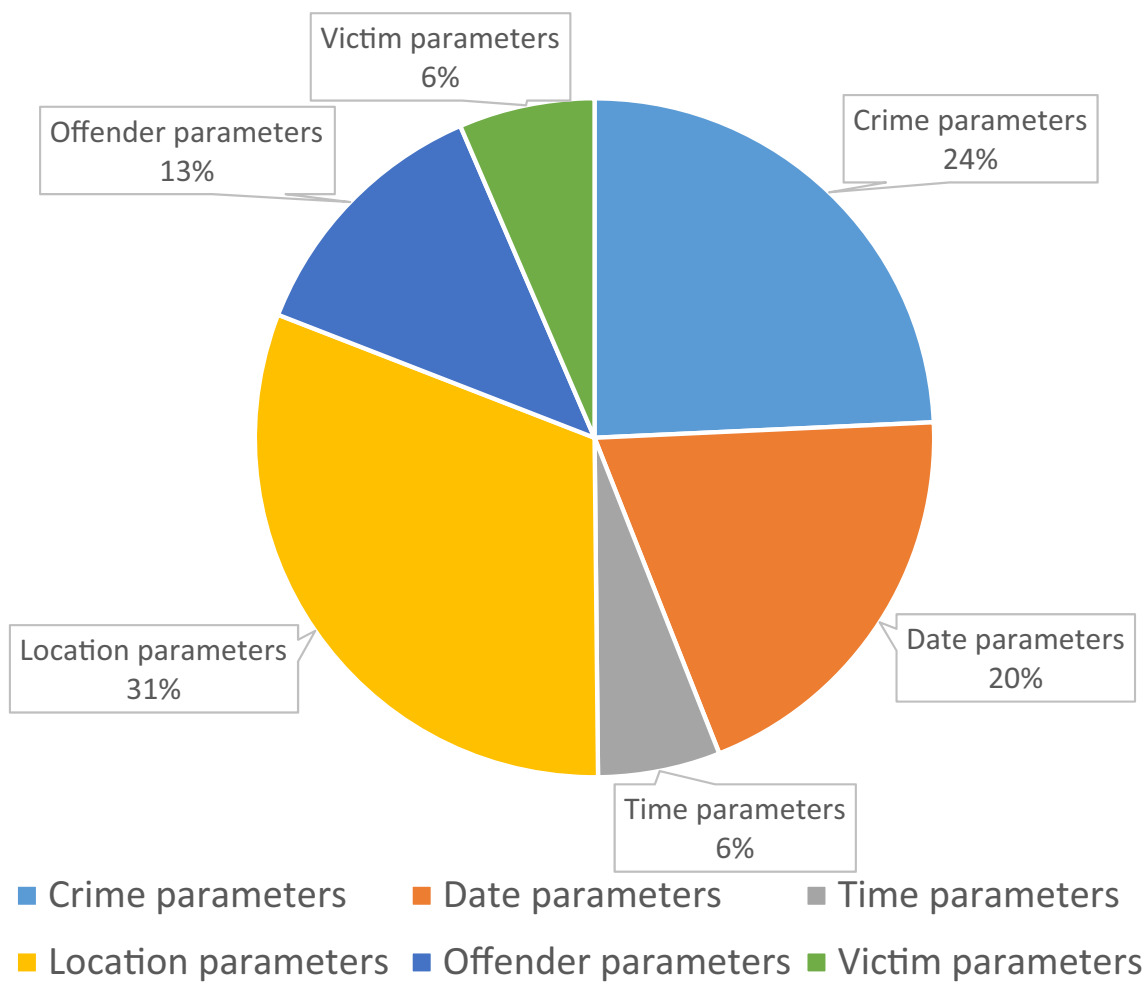


Fig. 5 Features by categories

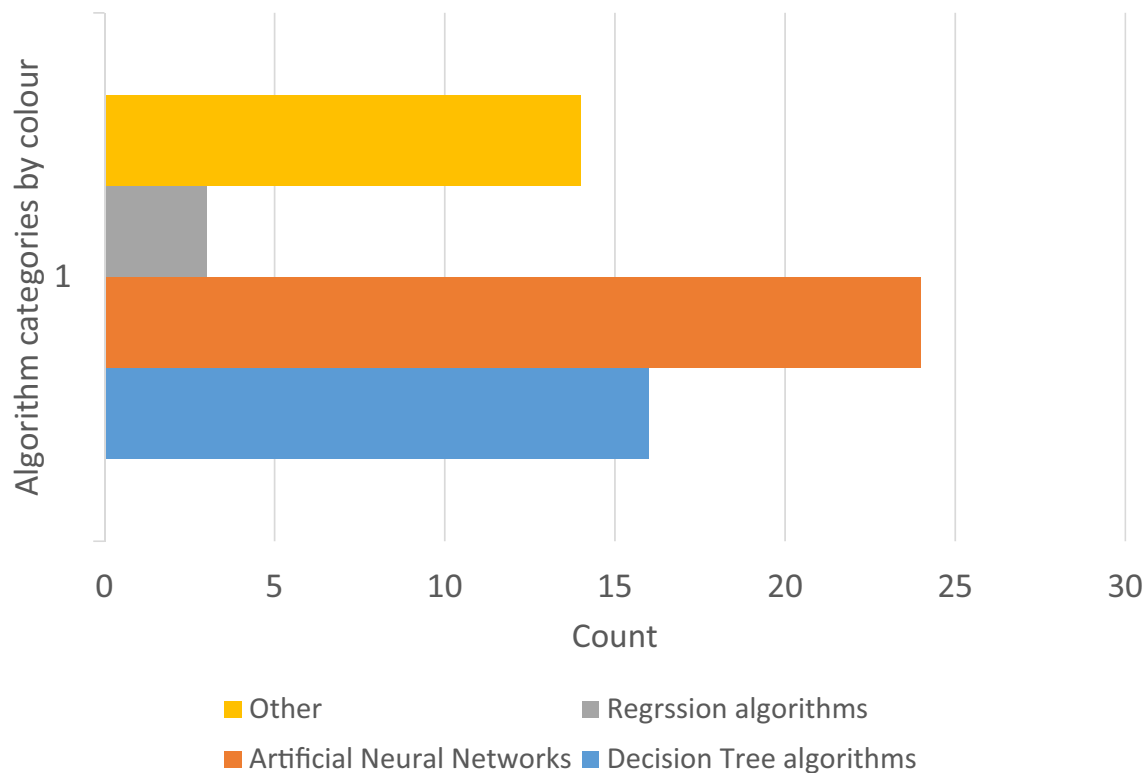


Fig. 6 Algorithms grouped by categories

into decision tree-based algorithms, artificial neural networks, regression algorithms, and others. However, the most commonly used algorithms were Artificial Neural Networks, Random Forest, and KNN algorithms. Table 6 shows the algorithms by category and the number of times they appeared in papers (Fig. 6).

4.5 Research Question 5—**which public datasets have been used?**

The majority of the datasets used throughout the study were collected from actual crime records datasets collected and maintained by governments or law enforcement agencies that made their data available to the general public via their online platform where web scrapping can be performed or third-party platforms that have the datasets available for the public such as kaggle.com or UCI. The majority of the datasets being used today show that governments and law enforcement are taking an interest in the collection and maintenance of these datasets. Figure 7 shows the distribution of public and open datasets that have been used it also indicates that most of the studies preferred public datasets, which are very useful for repeatability, verifiability, and even refutability of the experiments., and we have also grouped the commonly used datasets with links in Table 7.

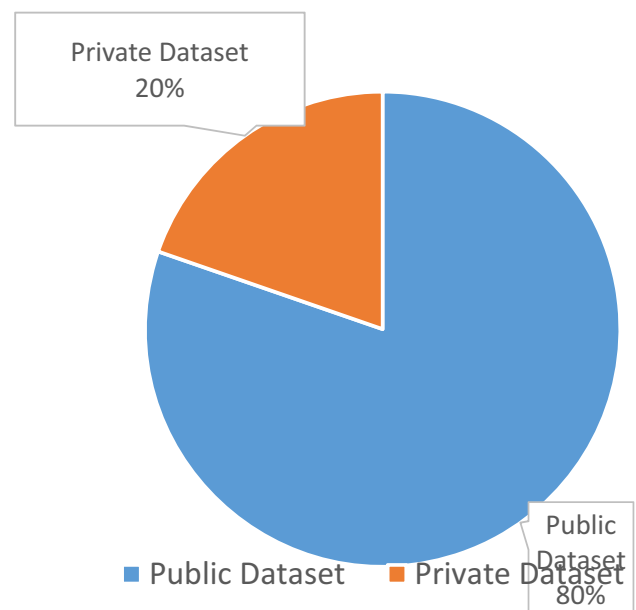


Fig. 7 Distribution of Public & Private datasets used

4.6 Research Question 6—**which evaluation metrics have been applied?**

14 different evaluation metrics were identified in the study that researchers applied to evaluate the model performance.

Table 7 Used datasets in the study and their distribution

Dataset location	Public	Private	Description	Distribution of papers
Kaggle (kaggle.com), UCI, Open crime data portals	x		San Francisco https://www.kaggle.com/c/sf-crime Boston https://www.kaggle.com/datasets/AnalyzeBoston/crimes-in-boston Chicago https://www.chicago.gov/city/en/dataset/crime.html NYPD https://compstat.nypdonline.org/ City of Detroit crime data https://detroitmi.gov/node/14916 Brazil Crime data https://www.kaggle.com/datasets/inquisitivecrow/crime-data-in-brazil	41
Official police department and law enforcement records and sensitive data	x		UK police website https://www.met.police.uk/sd/stats-and-data/met/crime-data-dashboard/ , Dhaka metropolitan https://dmp.gov.bd/crime-data/ Chilean police department Saint Petersburg police crime data Bangladesh police crime data Malaysian police crime data Sri Lanka police crime data Cambridge police Philadelphia police crime data Amsterdam police crime data China police crime data	25
Text data		X	Social media, newspaper articles	2

Some of the researchers used more than one evaluation metric, which resulted in a higher % value than 100%. In some of the papers, the researchers did not clearly state the evaluation method they had applied and in others, they did not state it at all, we found that 21 papers did not state their evaluation metrics in detail. In most cases, these papers were those where only one machine learning algorithm was applied in the study. The following evaluation metrics were used more than 5 times (Joshi 2016): Accuracy, Root Mean Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Error (MAE), Precision, Recall, Area Under the ROC curve (AUC), F1 score, Hit rate, and R2 score and the rest of the metrics were applied 4 times or less (Fig. 8).

4.7 Research Question 7—what are the top 5 ML categories?

In research question 4, we identified the machine learning algorithms used in the study. We deduced this information further to understand the type of machine learning

categories. As data is fed into the algorithms, the machines learn from this data and optimize their tasks to better their performance and intelligence over time. Several papers used multiple machine learning algorithms. The most applied was supervised machine learning, these tasks are those which contain labeled data for the algorithms to learn from. In semi-supervised learning, a mixture of both labeled and unlabeled data is fed into the algorithms, unsupervised algorithms learn from unlabeled data, and reinforcement algorithms are those, which are rewarded based on the desired behaviors/target outcome or punished for undesired outcomes. Reinforcement learning algorithms learn through a trial and error basis (Fig. 9).

4.8 Research Question 8—what are the challenges / limitations and possible solutions?

Researchers need to pay careful attention to the challenges they face when implementing machine learning algorithms or when seeking to develop a robust crime prediction model due to the sensitivity of the outcome variable and data

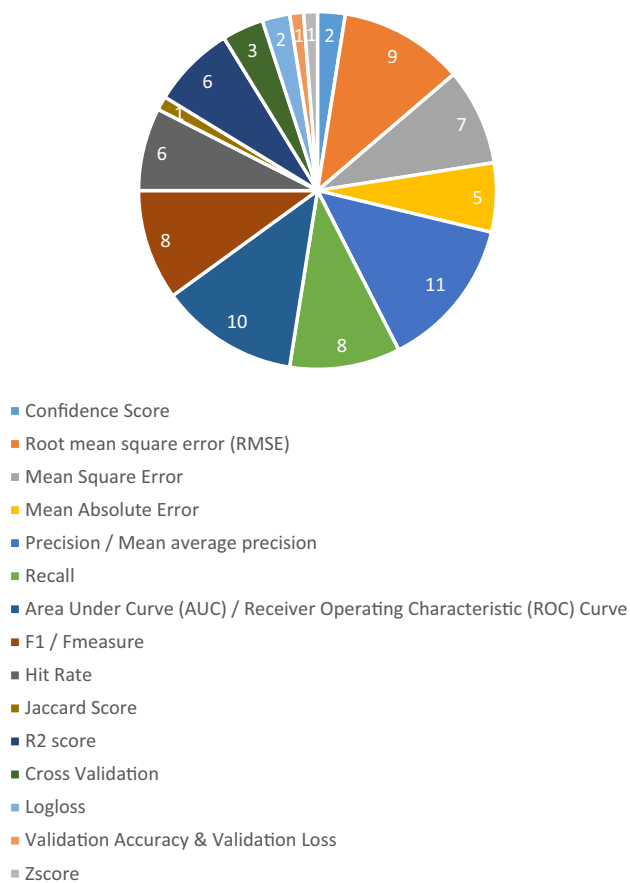


Fig. 8 Distribution of evaluation metrics applied

contained. We picked up one primary and common challenge when conducting crime prediction, which is the accuracy and trustworthiness of the available data. Just above 50% of the papers failed to provide or detail the challenges they had faced, we then indicated these as “Challenge NA”. The same applied when looking at the possible solutions, we found that more than 50% also lacked to identify further solutions and we deduced this as “Solution NA”. Several internal and external factors can contribute to the limitations of crime data and prediction. It is worth noting that one of the major challenges across the world is the reporting of crimes within communities. Underreporting is when people do not come forward with information on a crime or report a crime that has taken place to them due to fear of life and or any other threatening factors, and due to this, these statistics cannot be added or included to official crime statistics. In the work (Butt et al. 2020) it is stated that a survey by Malaysian and British police concluded that about 50% failed to report a crime. The challenges presented forward by the researchers have been narrowed down into 4 commonly faced challenges in machine learning: (1) Data collection, (2) data storage and security, (3) data pre-processing, and

(4) performance issues. We have not found any issue with visualization (Table 8).

5 Discussion

In this section, we present our discussion and feedback on each research question in Sect. 5.1, and we also discuss the potential threats to validity in Sect. 5.2.

5.1 General discussion

Firstly, we identified 350+ publications in total when conducting our general search into crime prediction publications. After we applied our exclusion criteria and quality assessment to our results, our total number of publications narrowed down to 68 primary studies. From our results, we have managed to identify different machine learning approaches, challenges, and possible solutions in current and past crime prediction trends. Crime prediction is a broad term that can be used by different fields of academia such as criminology and social development. We further discuss our responses to each research question in the following subsections.

5.2 Research Question 1—what are the objectives of the paper?

The results obtained from the data concluded that the objective of 68 of the publications was to produce a novel crime prediction model, it is also worth noting that the results would exceed the number of publications as some publications had more than one objective in their study (e.g., some included feature selection and a novel crime prediction model). We can conclude that an interesting point of interest would be to narrow down a crime prediction model to focus on common crimes in a specific area and develop a novel crime prediction model from this data.

5.3 Research Question 2—what data source types have been used to collect data?

To understand the approach and methodology of the researcher, the data source used was an important part of the study, also to learn the various types of data sources being used currently in crime prediction. We noted that there are several different data source types such as actual crime records, Twitter tweets, Facebook posts, census data, weather data, cellphone tower data, and point of interest data. As a result, we further narrowed down our results, most publications have made use of actual crime records in their

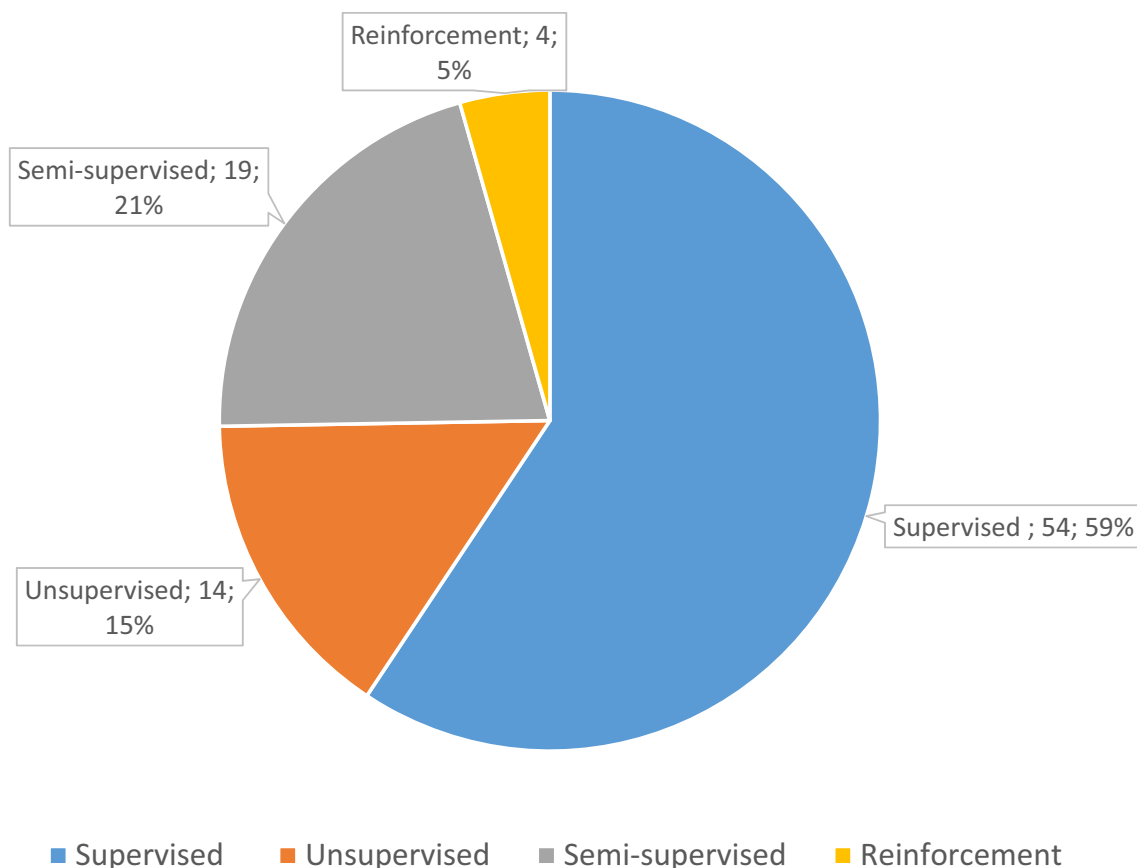


Fig. 9 Distribution of machine learning categories

studies. The use of actual crime data could improve crime prediction models, decrease crime rates within communities, and help researchers gain a better understanding of crime patterns (Yuki et al. 2019).

5.4 Research Question 3—what independent variables are (a.k.a features) used in the publications?

Crime is affected by both social and economic factors (Matereke et al. 2021); features from both the offenders and victims can contribute to the occurrence of a crime. Many of the researchers proposed that crime is a time and space problem and looked into pursuing prediction models about Spatio-temporal crime prediction. As an indication of this, researchers used a lot of location data, it is also suspected that criminals or offenders repeat a crime in places they are most familiar with (Bandeekar and Vijayalakshmi 2020).

5.5 Research Question 4—which ML algorithm(s) performed best in the study?

Artificial Neural Networks came up as the most used machine learning algorithm for crime prediction. More complex models can be built around neural networks with ensemble algorithms combined, adding boosting parameters can help improve performance and accuracy in the models for real-time predictions.

5.6 Research Question 5—Which public datasets have been used?

80% of the publications retrieved are from public datasets, which are openly accessible to the public, we have also shown this information in Table 7. Most of the data concluded that a lot of research around crime prediction is being done in India, and one paper relieved a study in Cape Town, South Africa (Matereke et al. 2021) where the Chicago data

Table 8 Challenges and Solutions proposed by researchers in primary studies

Category	Challenges	Proposed Solutions	References
Data Collection	Researchers can only predict crime records known to the police Noisy data, Noisy Twitter data	Victims should report more crimes Researchers achieved higher predictions when using the CPLIC model (4% improvement), proposed text data cleaning methods	Kshatri et al. (2021) Forraddellas et al. (2020); Kajita and Kajita (2020)
Data Storage and Security	Data used was not for actual police crime record Data imbalance Security and infrastructure costs of data	Use of actual crime records for actual or novel models Data augmentation Researchers have not provided a solution to the constantly changing environments	Yuki et al. (2019); Kiran and Kaishveen (2018) Saravanan et al. (2021) Nakib et al. (2018); Han et al. (2020)
Data pre-processing	The sparsity of data	Reliable and secure environments, large infrastructure There is no viable solution to this challenge, Dynamic information is very sparse, and a matrix factorization-based approach can be considered	Hajela et al. (2020) Huang et al. (2018); Wang et al. (2013)
Performance issues	Low accuracy, computational issues with KNN, low accuracy for Decision Tree and KNN algorithms, negative values by linear regression	Use of neural networks for future works, Use of manhattan distance instead of Euclidean distance, Random Forest approach with Adaboost for better accuracy, Improved results with Gradient boosting	Baek et al. (2021), Falade et al. (2019); Ippolito and Lozano (2020)

portal was used to obtain the data. Some publications noted that due to the sensitivity of the data and how it could be used to public advantage in the wrongful hands of others, they could not disclose who the data source owner was nor where the dataset was collected from.

5.7 Research Question 6—**which evaluation metrics have been applied?**

It was difficult to gather some information and data from some of the publications due to the researchers not giving full or detailed information on their study, some 20 + did not disclose the type of evaluation metrics they used within the study, and only detailed their results and findings during the study process, and accuracy percentage was given in these sections in replacing. Accuracy, area under the ROC curve and precision came as the top three evaluation metrics identified in the study. Equation 1 represents how to calculate the precision parameter and also, shows how to calculate the recall metric. Area under the ROC curve (AUC) value can be understood as the ability to differentiate between classes by a classifier, and accuracy refers to the % out of 100 for the prediction made.

$$\text{Precision} = \frac{\text{True Positive}}{\text{Actual Results}}$$

$$\text{or } \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (1)$$

$$\text{Recall} = \text{True Positive} \frac{\text{True Positive}}{\text{Predicted results}}$$

$$\text{or } \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (2)$$

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total}} \quad (3)$$

The precision is the ratio of True Positive (TP) over True Positive (TP) + False Positive (FP). The recall is the ratio of True Positive (TP) over True Positive (TP) + False Negative (FP) (Rumi et al. 2018).

5.8 Research Question 7—**what are the top 5 machine learning categories?**

In this study, the authors used supervised machine learning as the common machine learning category (Sharma et al. 2021). This means that more research can be performed for the other categories, particularly semi-supervised learning, and unsupervised learning because sometimes the number of labeled data points is quite limited or does not exist, therefore, we need models that can be used in these cases.

5.9 Research Question 8—what are the challenges/limitations and possible solutions?

We categorized our results into the following categories: Data collection, data pre-processing, data storage and security, and performance issues. It is worth noting that during our study we did not find any publication which had data visualization issues, this does not mean the challenge doesn't exist within the domain however, it is something that could be investigated in the future. Data collection in this domain remains a general issue with data owners, governments, or law enforcement agencies not making this data available to the public, not properly maintaining this data, or not keeping such data in reliable and safe storage locations. As we can see from Fig. 2, the rise in interest in this domain as a research topic is primarily due to the increased availability of data by governments in some parts of the world.

5.10 Available commercial tools

At present, ML is being used by law enforcement and other government agencies to predict crime. These known predictive policing software are Crime anticipation system, PreCobs, PredPol, and Hunchlab, these systems also allow for more efficient allocation of resources for law enforcement agencies and provide crime prevention strategies (Carvalho and Pedrosa 2021). It is also worth noting that these systems are fairly new and may come with limitations making the evaluation of its impact in crime rates more difficult.

5.11 Potential threats to validity

To ensure that quality research in software engineering is conducted, it is important to analyze threats to validity. Our current research has some limitations to other secondary studies (Dinter et al. 2021), and is discussed as follows. Regarding construct validity, we performed an automated search instead of manually reading titles in electronic journals, this means that we might have missed some relevant papers due to our automated search. Also, query phrasing is an important construct validity threat, and each electronic database has different options for executing the corresponding query. To minimize the potential risks, query design was discussed among authors before executing it in the databases. Another potential construct validity threat is related to the data extraction forms; we might have missed some useful data fields although we have updated these fields several times during our review process. Since we carefully specified the research questions, we reached our objectives adequately and we consider that there is no internal validity threat. Regarding the conclusion validity, we followed a well-defined SLR protocol, and the process was discussed among authors before performing this research. Conclusions

were derived from the collected data and personal/subjective opinions were not included.

5.12 Limitations

Even though some of these methods may present many benefits, there are also limitations. Some of these limitations may be noted as technical limitations to ML models, which may come with some of these benefits in crime prediction:

- ML techniques do not produce accurate results or predictions immediately, because they need to learn from previous data.
- The relationship between urban metrics and population size is not linear (Alves et al. 2018)
- Data availability and a limited amount of resources
- System performance issues (technical)
- Data storage (technical)
- Data sparsity (pre-processing of data)

5.13 Future research outlooks

We identified the following research directions to pave the way for further research:

1. The development of Explainable Artificial Intelligence (XAI) models / interpretable machine learning models
2. The use of new deep learning models such as transformers to improve the performance of models
3. The development of unsupervised learning-based crime prediction models
4. The design and implementation of a benchmarking tool that can evaluate the performance of different machine learning models on public datasets
5. The development of semi-supervised learning-based crime prediction models
6. The development of new publicly available datasets for researchers
7. The use of new features to build crime prediction models
8. The analysis of cross-country crime prediction models to analyze the commonalities between models
9. The development of open source tools for crime prediction

6 Conclusion and future work

This SLR provided an overview of crime prediction analysis and the various machine learning algorithms used in the field. Based on the primary selected studies, we have found that crime is affected by many internal and external factors. In this study, we systematically reviewed the critical aspects of crime prediction by following the guidelines of the work

(Kitchenham et al. 2007). Researchers have made numerous amounts of contributions to the study of crime investigation and prediction. Unlike most industries; health care, transportation, agriculture, finance, retail, and customer services crime prediction has a lack of comprehensive and systematic literature reviews that can help to organize, and summarize existing literature, evidence, and potential challenges they encounter. To our knowledge, this is the most up-to-date review on the use of machine learning in crime prediction between 2010 and 2022. The study showed that the researchers in the field are interested in crime prediction and there is a growing need for interest in the field. Secondly, we noted that most of the publications used crime data records from police stations/law enforcement agencies. These datasets contained a variety of variables that made up the data from crime type, crime case, victim, perpetrator, date, time, weather, and many more. This SLR concludes that crime prediction can help assist law enforcement, governments, and police department across the world to improve their communities and economy. A novel approach to crime prediction would be an optimal approach to solving crime prediction and for a safer economy.

In this SLR, the limitations are restricted to journal articles, and reviews between 2010 and 2022 related to ML and AI in crime prediction. A large number of irrelevant articles were omitted in the exclusion criteria phase of the search approach. This made it possible for us to only look at the papers that passed the criteria for the study. We believe that the quality of this study has been further improved and strengthened by the addition of more sources and articles.

Acknowledgements We thank Dr. Nosisi Jenga for taking the time to assist in the proofreading of this manuscript. We also thank our universities for providing the infrastructure and scientific database access rights.

Funding Open Access funding provided by the Qatar National Library.

Data availability statement The paper has no associated data to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aarhi S, Samyuktha M, Sahana M (2019) Crime hotspot detection with clustering algorithm using data mining. In: 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), IEEE, pp 401–405
- Abbas AJF (2019) A survey of research into artificial neural networks for crime prediction. *Перспективы науки*, 33
- Aldossari BS, Alqahtani F M, Alshahrani NS, Alhammam MM, Alzamanan RM, Aslam N (2020) A comparative study of decision tree and naive bayes machine learning model for crime category prediction in chicago. In: Proceedings of 2020 the 6th International Conference on Computing and Data Engineering, pp 34–38
- Alves LG, Ribeiro HV, Rodrigues FA (2018) Crime prediction through urban metrics and statistical learning. *Physica A* 505:435–443
- Araújo A, Cacho N, Bezerra L, Vieira C, Borges J (2018) 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; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), IEEE, pp 1256–1263
- Arulanandam R, Savarimuthu BTR, Purvis MA (2014) Extracting crime information from online newspaper articles. *Proc Sec Austral Web Conf* 155:31–38
- Babakura A, Sulaiman MN, Yusuf MA (2014) The improved method of classification algorithms for crime prediction. In: 2014 International Symposium on Biometrics and Security Technologies (ISBAST), IEEE, pp 250–255
- Baek MS, Park W, Park J, Jang KH, Lee YT (2021) Smart Policing Technique with Crime Type and Risk Score Prediction Based on Machine Learning for Early Awareness of Risk Situation. *IEEE Access* 9:131906–131915
- Baloian N, Bassaletti CE, Fernández M, Figueroa O, Fuentes P, Manasevich R, Vergara M (2017) Crime prediction using patterns and context. In: 2017 IEEE 21st international conference on computer supported cooperative work in design (CSCWD), IEEE, pp 2–9
- Bandekar SR, Vijayalakshmi C (2020) Design and analysis of machine learning algorithms for the reduction of crime rates in India. *Procedia Comput Sci* 172:122–127
- Bappee FK, Soares A, Petry LM, Matwin S (2021) Examining the impact of cross-domain learning on crime prediction. *J Big Data* 8(1):1–27
- Birks D, Coleman A, Jackson D (2020) Unsupervised identification of crime problems from police free-text data. *Crime Sci* 9(1):1–19
- Butt UM, Letchmunan S, Hassan FH, Ali M, Baqir A, Sherazi HHR (2020) Spatio-temporal crime HotSpot detection and prediction: a systematic literature review. *IEEE Access* 8:166553–166574
- Butt UM, Letchmunan S, Hassan FH, Ali M, Baqir A, Koh TW, Sherazi HHR (2021) Spatio-temporal crime predictions by leveraging artificial intelligence for citizens' security in smart cities. *IEEE Access* 9:47516–47529
- Carvalho A, Pedrosa I (2021) Data analysis in predicting crime: predictive policing. In: 2021 16th Iberian Conference on Information Systems and Technologies (CISTI), IEEE, pp 1–7
- Catal C (2012) On the application of genetic algorithms for test case prioritization: a systematic literature review. In: Proceedings of the 2nd international workshop on evidential assessment of software technologies, pp 9–14
- Chun SA, Avinash Paturu V, Yuan S, Pathak R, Atluri V, Adam NR (2019) Crime prediction model using deep neural networks. In:

- Proceedings of the 20th Annual International Conference on Digital Government Research, pp 512–514
- Dakalbab F, Abu Talib M, Elmutasim O, Bou Nassif A, Abbas S, Nasir Q (2022) Artificial intelligence & crime prediction: a systematic literature review. *Soc Sci Humanit* 6:100342
- David H, Suruliandi A (2017) Survey on crime analysis and prediction using data mining techniques. *ICTACT J Soft Comput* 7(3):1459
- Elluri L, Mandalapu V, Roy N (2019) Developing machine learning-based predictive models for smart policing. In: 2019 IEEE International Conference on Smart Computing (SMARTCOMP), IEEE, pp 198–204
- Esquivel N, Nicolis O, Peralta B, Mateu J (2020) Spatio-temporal prediction of Baltimore crime events using CLSTM neural networks. *IEEE Access* 8:209101–209112
- Falade A, Azeta A, Oni A, Odun-ayo I (2019) Systematic literature review of crime prediction and data mining. *Rev Comput Eng Stud* 6(3):56–63
- Feng M, Zheng J, Ren J, Hussain A, Li X, Xi Y, Liu Q (2019) Big data analytics and mining for effective visualization and trends forecasting of crime data. *IEEE Access* 7:106111–106123
- Forradellas RFR, Nández Alonso SL, Jorge-Vazquez J, Rodriguez ML (2020) Applied machine learning in social sciences: neural networks and crime prediction. *Social Sci* 10(1):4
- Fu K, Chen Z, Lu CT (2018) Streetnet: preference learning with a convolutional neural network on urban crime perception. In: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp 269–278
- Gerstner D (2018) Predictive policing in the context of residential burglary: an empirical illustration on the basis of a pilot project in Baden-Württemberg, Germany. *Eur J Secur Res* 3(2):115–138
- Ghazvini A, Abdullah SNHS, Hasan MK, Kasim DZAB (2020) Crime spatiotemporal prediction with fused objective function in time-delay neural network. *IEEE Access* 8:115167–115183
- Hajela G, Chawla M, Rasool A (2020) A clustering-based hotspot identification approach for crime prediction. *Procedia Comput Sci* 167:1462–1470
- Han X, Hu X, Wu H, Shen B, Wu J (2020) Risk prediction of theft crimes in urban communities: an integrated model of lstm and st-gcn. *IEEE Access* 8:217222–217230
- He J, Zheng H (2021) Prediction of crime rate in urban neighborhoods based on machine learning. *Eng Appl Artif Intell* 106:104460
- Hossain S, Abtahee A, Kashem I, Hoque MM, Sarker IH (2020) Crime prediction using spatio-temporal data. International conference on computing science communication and security. Springer, Singapore, pp 277–289
- Huang C, Zhang J, Zheng Y, Chawla NV (2018) DeepCrime: attentive hierarchical recurrent networks for crime prediction. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp 1423–1432
- Ingilevich V, Ivanov S (2018) Crime rate prediction in the urban environment using social factors. *Procedia Comput Sci* 136:472–478
- Ippolito A, Lozano ACG (2020) Tax crime prediction with machine learning: a case study in the municipality of São Paulo. *ICEIS* 1:452–459
- Islam M, Karim R, Roy K, Mahmood S, Hossain S, Rahman RM (2018) Crime prediction using multiple-anfis architecture and spatiotemporal data. In: 2018 International Conference on Intelligent Systems (IS), IEEE, pp 58–65
- Jalil MMA, Mohd F, Noor NMM (2017) A comparative study to evaluate filtering methods for crime data feature selection. *Procedia Comput Sci* 116:113–120
- Jha P, Jha R, Sharma A (2019) Behavior analysis and crime prediction using big data and machine learning. *Int J Recent Technol Eng* 8(1):1
- Jha S, Yang E, Almagrabi AO, Bashir AK, Joshi GP (2021) Comparative analysis of time series model and machine testing systems for crime forecasting. *Neural Comput Appl* 33(17):10621–10636
- Joshi R (2016) Accuracy, precision, recall & f1 score: interpretation of performance measures. Retrieved 1(2018):2016
- Joshi N, Shaikh S, Tripathy A, Sen S, Janson J, Karnik S, Varghese B (2019) Crime anatomization using qgis. In: 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), IEEE, pp 1–4
- Kadar C, Pletikosa I (2018) Mining large-scale human mobility data for long-term crime prediction. *EPJ Data Sci* 7(1):1–27
- Kadar C, Maculan R, Feuerriegel S (2019) Public decision support for low population density areas: An imbalance-aware hyperensemble for Spatio-temporal crime prediction. *Decis Support Syst* 119:107–117
- Kajita M, Kajita S (2020) Crime prediction by data-driven Green's function method. *Int J Forecast* 36(2):480–488
- Kim S, Joshi P, Kalsi PS, Taheri P (2018) Crime analysis through machine learning. In: 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), IEEE, pp 415–420
- Kiran J, Kaishveen K (2018) Prediction analysis of crime in India using a hybrid clustering approach. In: 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud)(I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics, and Cloud)(I-SMAC), 2018 2nd International Conference on IEEE, pp 520–523
- Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering. Technical report EBSE 2007-001, Keele University and Durham University Joint Report
- Kitchenham BA, Mendes E, Travassos GH (2007) Cross versus within-company cost estimation studies: a systematic review. *IEEE Trans Softw Eng* 33(5):316–329
- Kitchenham BA, Brereton P, Turner M, Niazi MK, Linkman S, Pretorius R, Budgen D (2010) Refining the systematic literature review process—two participant-observer case studies. *Empir Softw Eng* 15(6):618–653
- Kounadi O, Ristea A, Araujo A, Leitner M (2020) A systematic review on spatial crime forecasting. *Crime Sci* 9(1):1–22
- Krishnendu SG, Lakshmi PP, Nitha L (2020) Crime analysis and prediction using optimized k-means algorithm. In: 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), IEEE, pp 915–918
- Kshatri SS, Singh D, Narain B, Bhatia S, Quasim MT, Sinha GR (2021) An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: an ensemble approach. *IEEE Access* 9:67488–67500
- Lal S, Tiwari L, Ranjan R, Verma A, Sardana N, Mourya R (2020) Analysis and classification of crime tweets. *Procedia Comput Sci* 167:1911–1919
- Li X, Kang X, Wang C, Dong L, Yao H, Li S (2020) A neural-network-based model of charge prediction via the judicial interpretation of crimes. *IEEE Access* 8:101569–101579
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Moher D (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 62(10):e1–e34
- Lighthart A, Catal C, Tekinerdogan B (2021) Systematic reviews in sentiment analysis: a tertiary study. *Artif Intell Rev* 54(7):4997–5053
- Lin YL, Chen TY, Yu LC (2017) Using machine learning to assist crime prevention. In: 2017 6th IIAI international congress on advanced applied informatics (IIAI-AAI), IEEE, pp 1029–1030
- Llaha O (2020) Crime analysis and prediction using machine learning. In: 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), IEEE, pp 496–501

- Mahmud S, Nuha M, Sattar A (2021) Crime rate prediction using machine learning and data mining. *Soft computing techniques and applications*. Springer, Singapore, pp 59–69
- Materike T, Nyirenda C, Ghaziasgar M (2021) A comparative evaluation of spatio temporal deep learning techniques for crime prediction (No. 5648). *EasyChair*.
- Miyano K, Shinkuma R, Shiode N, Shiode S, Sato T, Oki E (2020) Multi-UAV allocation framework for predictive crime deterrence and data acquisition. *Internet of Things* 11:100205
- Morimoto S, Kawamukai H, Shin K (2019) Prediction of crime occurrence using information propagation model and gaussian process. In: 2019 14th Asia Joint Conference on Information Security (AsiaJCIS), IEEE, pp 80–87
- Munasinghe AKDMP, Udeshini S, Perera H, Weerasinghe R (2014) Criminal shortlisting and crime forecasting based on modus operandi. At 2014 14th International Conference on Advances in ICT for Emerging Regions (ICTer)
- Nakib M, Khan RT, Hasan MS, Uddin J (2018) Crime scene prediction by detecting threatening objects using a convolutional neural network. In: 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), IEEE, pp 1–4
- Niu X, Elsisy A, Derzsy N, Szymanski BK (2019) Dynamics of crime activities in the network of city community areas. *Appl Netw Sci* 4(1):1–17
- Prabakaran S, Mitra S (2018) Survey of analysis of crime detection techniques using data mining and machine learning. *J Phys* 1000(1):012046
- Pratibha AG, Uprant SD, Chouhan L (2020) Crime prediction and analysis. *Int Conf Data Eng Appl*. <https://doi.org/10.1109/IDEA49133.2020.9170731>
- Rajapakshe C, Balasooriya S, Dayarathna H, Ranaweera N, Walgampaya N, Pemadasa N (2019) Using CNN's RNNs and machine learning algorithms for real-time crime prediction. In: 2019 International Conference on Advancements in Computing (ICAC), IEEE, pp 310–316
- Rosés R, Kadar C, Malleon N (2021) A data-driven agent-based simulation to predict crime patterns in an urban environment. *Comput Environ Urban Syst* 89:101660
- Rumi SK, Deng K, Salim FD (2018) Crime event prediction with dynamic features. *EPJ Data Sci* 7(1):43
- Rummens A, Hardyns W, Pauwels L (2017) The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context. *Appl Geogr* 86:255–261
- Saravanan P, Selvaprabu J, Arun Raj L, Khan AA, Sathick KJ (2021) Survey on crime analysis and prediction using data mining and machine learning techniques. *Advances in smart grid technology*. Springer, Singapore, pp 435–448
- Sathyadevan S, Devan MS, Gangadharan SS (2014) Crime analysis and prediction using data mining. In: 2014 First International Conference on Networks & Soft Computing (ICNSC2014), IEEE, pp 406–412
- Shah N, Bhagat N, Shah M (2021) Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. *Vis Comput Indu Biomed Art* 4(1):1–14
- Sharma HK, Choudhury T, Kandwal A (2021) Machine learning-based analytical approach for geographical analysis and prediction of Boston City crime using geospatial dataset. *GeoJournal*. <https://doi.org/10.1007/s10708-021-10485-4>
- Shermila AM, Bellarmine AB, Santiago N (2018) Crime data analysis and prediction of perpetrator identity using machine learning approach. In: 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), IEEE, pp 107–114
- Thomas A, Sobhana NV (2022) A survey on crime analysis and prediction. *Maters Today* 58:310–315
- ToppiReddy HKR, Saini B, Mahajan G (2018) Crime prediction and monitoring framework based on spatial analysis. *Procedia Comput Sci* 132:696–705
- Van Dinter R, Tekinerdogan B, Catal C (2021) Automation of systematic literature reviews: a systematic literature review. *Inf Softw Technol* 136:106589
- Vomfell L, Härdle WK, Lessmann S (2018) Improving crime count forecasts using Twitter and taxi data. *Decis Support Syst* 113:73–85
- Walczak S (2021) Predicting Crime and Other Uses of Neural Networks in Police Decision Making. *Front Psychol* 12:587943
- Wang D, Ding W, Stepinski T, Salazar J, Lo H, Morabito M (2012) Optimization of criminal hotspots based on underlying crime controlling factors using geospatial discriminative patterns. International conference on industrial engineering and other applications of applied intelligent systems. Springer, Berlin, pp 553–562
- Wang T, Rudin C, Wagner D, Sevieri R (2013) Learning to detect patterns of crime. Joint European conference on machine learning and knowledge discovery in databases. Springer, Berlin, pp 515–530
- Wang J, Hu J, Shen S, Zhuang J, Ni S (2020) Crime risk analysis through big data algorithm with urban metrics. *Physica A* 545:123627
- Yao S, Wei M, Yan L, Wang C, Dong X, Liu F, Xiong Y (2020) Prediction of crime hotspots based on spatial factors of random forest. In: 2020 15th International Conference on Computer Science and Education (ICCSE), IEEE, pp 811–815
- Yu CH, Ward MW, Morabito M, Ding W (2011) Crime forecasting using data mining techniques. In: 2011 IEEE 11th International Conference on Data Mining Workshops, IEEE, pp 779–786
- Yuki JQ, Sakib MMQ, Zamal Z, Habibullah KM, Das AK (2019) Predicting crime using time and location data. In: Proceedings of the 2019 7th International Conference on Computer and Communications Management, pp 124–128
- Zhang X, Liu L, Xiao L, Ji J (2020) Comparison of machine learning algorithms for predicting crime hotspots. *IEEE Access* 8:181302–181310
- Zhang Q, Yuan P, Zhou Q, Yang Z (2016) Mixed spatial-temporal characteristics based on crime hot spots prediction. In: 2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD), IEEE, pp 97–101

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.