Tech Science Press

check for updates

Crime Prediction Methods Based on Machine Learning: A Survey

Junxiang Yin*

School of Social Science, University of Manchester, M13 9PL Oxford Road, Manchester, UK
*Corresponding Author: Junxiang Yin. Email: junxiang.yin@postgrad.manchester.ac.uk

Abstract: The objective of crime prediction, one of the most important technologies in social computing, is to extract useful information from many existing criminal records to predict the next process-related crime. It can aid the police in obtaining criminal information and warn the public to be vigilant in certain areas. With the rapid growth of big data, the Internet of Things, and other technologies, as well as the increasing use of artificial intelligence in forecasting models, crime prediction models based on deep learning techniques are accelerating. Therefore, it is necessary to classify the existing crime prediction algorithms and compare in depth the attributes and conditions that play an essential role in the analysis of crime prediction algorithms. Existing crime prediction methods can be roughly divided into two categories: those based on conventional machine learning and those based on contemporary deep learning. This survey analyses the fundamental theories and procedures. The most frequently used data sets are then enumerated, and the fundamental procedures of various algorithms are also analyzed in this paper. In light of the insufficient scale of existing data in this field, the ambiguity of data types used to predict crimes, and the absence of public data sets that have a significant impact on the research of algorithm models, this survey proposes the construction of a machine learning-based big data research model to address these issues. Future researchers who will enter this field are provided with a guide to the direction of future research development.

Keywords: Crime prediction; machine learning; artificial intelligence; deep learning

1 Introduction

As a special social phenomenon, the essence of crime is the interaction between people under certain conditions. It is the behaviour mode that the criminal entity causes harm or damage to other objects. The occurrence of these behaviors has a specific basis. However, due to the influence of technological means and the absence of collected crime-related data, it is relatively difficult to find the essential cause of crimes to a certain extent. Relevant studies indicate that from 2014 to 2019, the crime rate increased from 2.3% to 7.8%, some of which were jewelry theft and murder [1]. With the development of big data technology, the traditional analysis based on causality has been transformed

into the correlation-based analysis. It is assumed that crimes can be effectively predicted in terms of historical data and the possibility that these events can be predicted in advance, which helps to prepare and reduces the losses caused by relevant events to the society. Nevertheless, it is a great challenge to predict the occurrence of criminal events effectively. The occurrence factors of the crime events are related to many conditions, whereas the causal relationship is uneasy to judge. With the further development of information technology, especially artificial intelligence technology, the data-driven way transforms causality into correlation for processing. Machine-based learning can improve the accuracy of prediction, and the availability of the prediction model to some extent. The problem of crime prediction can be divided into traditional machine learning-based methods and modern artificial intelligence-based deep learning methods. According to relevant data, the most significant number of cases is theft, accounting for more than half of all cases of property abuse, followed by robbery. The factors of different cases vary greatly. For example, for the cases of property abuse, on the one hand, it is due to the rapid development of economic level, and on the other hand, citizens' safety awareness needs the improvement. Moreover, due to the rapid social development, unbalanced economic development and income inequality in various places, some unskilled people need to obtain primary channels of life through extreme means, thus adopting the desperate measures. Besides, many of the cases are recurring.

As an important part of computational sociology, how to fully integrate computational computation into these social data using existing data and completely understand the rules behind the data using computational techniques becomes more important. This survey attempts to understand some phenomena of crime in society utilizing models and analyze some complex, dynamic and evolving data in the criminal process to make predictions and how to design human-computer interaction in the constructed system to form powerful intelligence. Machine learning uses a large amount of existing data, and extracts data patterns by building algorithmic models to extract regular and valuable information. This study of machine learning-based crime prediction models can further activate the massive data resources in the public security system. Through the in-depth mining of data content, the value of these data can be released and high-quality data products can be effectively utilized. Synchronously, with the improvement of data volume and computing power, traditional machine learning methods have been gradually replaced by deep learning based on neural networks with better fitting ability. However, the amount of data involved in different crime prediction is different. The K-Nearest Neighbor (KNN), Support Vector Machines (SVM), decision tree, random forest, Naive Bayes, and so on are relatively more effective for a small volume of the crime data sets.

Fig. 1 displays the basic structure of crime prediction. The Part (a) is a scheme based on traditional machine learning. Firstly, the data set is divided into a training set and a test set. The basic requirement of the algorithm is to extract the essential information that can represent the crime event data, that is, to realize the effective representation of the crime data. As the types of crime data are different and diverse, it is necessary to select relevant attribute data, and then extract all the features of the data at one time to improve the effective expression ability of the data. Afterwards the crime is predicted through a variety of machine learning algorithms, such as linear regression model, the decision tree and support vector machine, and so on. Finally, the accuracy of crime prediction is output. For the purpose of making the model have good generalization, it is generally required that the accuracy of the training set and the accuracy of the test set must be balanced. Otherwise, there will be a state of over-fitting. In the testing process, for the test data of crime prediction, the model parameters obtained from the training data are directly applied to predict the crime data. However, the difference between the traditional machine learning-based [2,3] approach and the modern deep learning-based [4–6] approach in the prediction process is that the deep learning-based approach does not require

the selection of relevant attributes and feature extraction. The basic procedure of the algorithms on the basis of the deep learning is shown inf Part (b), where is not feature extraction part, so a pre-processing part should be added to clean the input data. Additionally, it is a black-box end-to-end input approach, so there is no need to select relevant attributes and extract features, only data need to be input into the neural network. The general parts of prediction techniques are Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Graph Neural Network (GNN), Long-Short Term Memory neural network (LSTM) and so on. With the powerful nonlinear high-order fitting ability of artificial neural network, the neural network can automatically realize the ability of data expression through neurons, thus achieving better prediction.
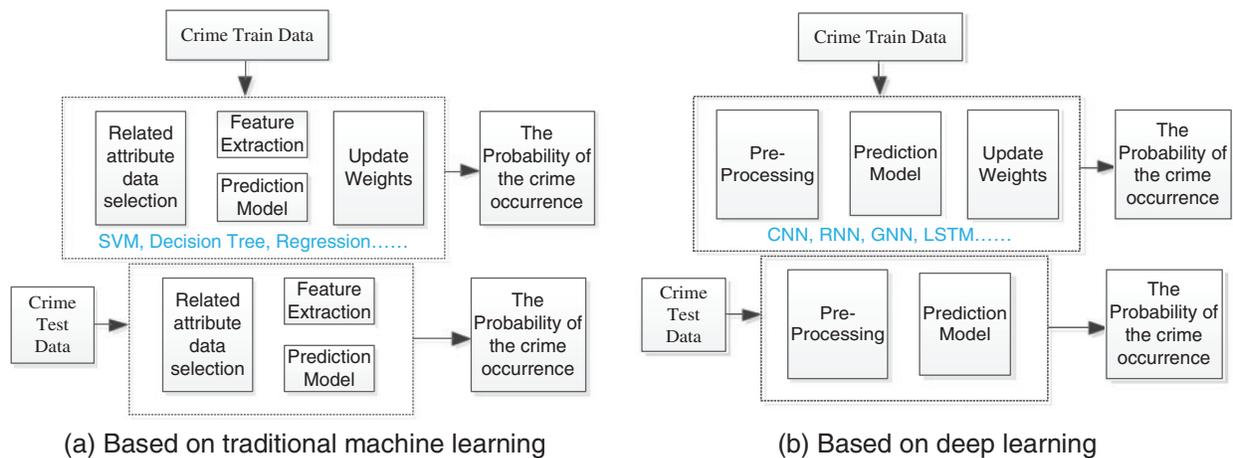

Figure 1: The basic procedure of the crime prediction methods: (a) is the scheme based on the traditional machine learning algorithm; (b) is the scheme based on deep learning algorithm

Based on aforementioned analysis, this paper conducts in-depth research on various existing crime prediction algorithm models, and studies the methods used from traditional machine learning-based and modern artificial neural network methods. The analysis of existing algorithms are presented in Section 2. Then then data set, basic techniques, and algorithm performance are given. Consequently, through these analyses, some problems existing in current crime prediction research are presented, such as the lack of public data set, too single data model used in the process of crime prediction, etc. Several possible solutions to this problem are proposed. At the same time, the crime prediction big data system is described, and each component is planned in detail to guide future researchers in this field.

2 The Analysis of the Existing Prediction Algorithm

The crime prediction based on the machine learning can not only help the police department know the possibility of crime occurrence, but also can help the people to protect themselves to avoid to be hurt. There are many papers related to this subject. In this part, we analyze these algorithms and classify them into different class.

2.1 The Analysis of Data Set

Data set is the most important thing to train the crime prediction scheme. However, the open source of crime data is scarce which limit the development of the crime prediction. Table 1 shows the

data set used in the surveyed algorithms. The volume of data set is will affect the performance of the crime prediction algorithm. When the volume is small, the traditional machine learning techniques get better performance. On the other hand, the methods based on the deep learning outperform than the traditional machine leaning on the big volume of the crime data set. The most two used dataset are from US and India for their data open system. Most of the papers test their algorithm based on the US dataset, especially on the New York City, San Francisco, and Chicago. The analysis of the crime prediction from India was always coarse granularity. The other dataset are from the UK, Canada, Portugal, Argentina, Swiss, Brazil, China, and so on. With the increment of open online competition source, some algorithms are test their performance on the Kaggle.

Table 1: Some information of the collected data set

Coutry	City	Data set description	Author and the published year
USA	New York City, San Francisco, Philadelphia, Chicago, Seattle, Minneapolis, DC metro, Portland, Dallas	Community data, National Criminal Record Database, National Archive of Criminal Justice Data, UCI community and Crime data set, Fraud and civil action	Buczak 2010, Somayeh2013,Iqbal 2013, Babakura 2014, McClendon 2015, Nguyen 2017, Kang 2017, Duan 2017, Mehmet 2017, Feng 2018, Kadar 2018, Huang 2018, Shermila 2018, ToppiReddy 2018, Nitta2019, Pradhan2019, Yi 2019, Wang2019, Umair 2020, Hajela 2020, Hossain 2020, Rayhan 2020,Wang 2020, Wei 2020, He2021, Matereke 2021, Safat 2021, Stalidis2021,Tamir2021, Wheeler 2021, Zhang 2020, Deepak 2021
India	Mumbai, Tamil Nadu, Andhra Pradesh, Telangana	Delhi police first information report (FIR) report, National Crime Record Bureau	Sivaranjani 2016, Aziz 2022,Bandekar 2020, Ghankutkar 2019, Deshmukh 2020, Kumar 2020, Tamilarasi 2020, Das 2019
Canada	Vancouver	Officially reported, BCAA, street information, Canadian Census	Bappee 2018, Kim 2018, Brantingham 2019, Bappee 2021
UK	London	Offence recorded by the police in England and Wales by office and police force area	Bogomolov 2014
Portugal	Porto	Official police data	Saraiva 2022
Argentina	Buenos Aires	Open crime data	Reier 2020

(Continued)

Table 1: Continued

Coutry	City	Data set description	Author and the published year
Swiss	Aargau	Burglary incidents in the Swiss canton	Kadar 2019
China	YD city, Gansu	Statistics released by the YD county	Li 2020, Liao 2010. Wu 2018
kaggele	kaggle	Kaggle open source crime data	Krishnendu 2020, Kanimozhi2020
Brazil	São Paulo	Taxes audits of the municipality	Ippolito 2020
Sri Lanka	Sri Lanka	Suspecious humans in video surveillance	Rajapakshe 2019

From the Table 1, we can get that the crime dataset in most of the existing papers are collected from the US open data source. The crime variables relationship of some big cities in US are deeply analyzed. Especially, with the development of the deep learning, the deep learning methods are on the basis of the big data. The volume of the data and the richness of the content are the core of the end-to-end algorithms.

2.2 The Analysis of Basic Algorithm

The basic approaches to crime event prediction based on the collected data can be divided into two parts. One is based on traditional machine learning and the other one is based on deep learning with neural networks as the basic framework structure. In the following, an in-depth analysis of the advantages and disadvantages of the existing methods will be developed from these two parts.

2.2.1 The Models Based on Traditional Machine Learning Methods

As early as the last century, Professor Brown [7] proposed the ReCAP plan, namely the regional crime analysis plan. Through data mining technology, various crime data can be effectively integrated, and characteristics and rules of different crimes can be found in the data to realize automatic prediction. Early methods used relatively little data and insufficient arithmetic power for crime prediction. Many algorithms used traditional machine learning methods, which are classified here as some clustering and classification techniques, Hot Spot, Bayesian, etc.

(1) Method Based on K-means, K-Nearest Neighbor, Support Vector Machine

Clustering is an unsupervised learning algorithm that divides similar data by designing certain rules. Since the crime data does not know which data attributes correspond to the crime, clustering methods are used to achieve the aggregation of different data. There are many clustering methods, and the more used algorithms are fuzzy clustering and k-means.

The method proposed by Brantingham et al. in 2009 [8] was based on fuzzy clustering and topological aggregation algorithms, and the officially reported crime data from 2001–2004, British Columbia Assessment Agency (BCAA) data from 2005, detailed street information from GIS Innovations in 2006, and Canadian Census data from 2001 were used to test the performance of the method. The algorithms were subjected to micro-macro-micro analysis to achieve validation of the

effectiveness of existing methods. The data used by Buczak et al. in 2010 [9] were the community-based data from the US Census, FBI Uniform Crime Report, US Law Enforcement Management and Administrative Statistics Survey. The National Criminal Record Database and National Archive of Criminal Justice Data were also used as the base material for the study. The fuzzy correlation rule was used to mine the data for crime patterns. Finally, the attribute that is very relevant to crime is the mean number people per household, race, age, salary, some type of income, poverty level, education, career, marriage, and so on. The method proposed by Somayeh et al. in 2013 [10] utilizes the better performing supervised classification learning algorithm. The experimental data were selected by the Communities and Crime Unnormalized dataset as the base object. And the datasets contain US Census 1990 socio-economic data, law enforcement data from the 1990 Law Enforcement Management and Administrative Statistics Survey. The crime case contains 2215 examples with a total of 147 attributes. The experimental results were compared to Area Under Curve, Bayesian, K-NN, Neural Network, decision tree, Support Vector Machine, and the effectiveness of the different methods is verified.

In addition, for some crime incidents, such as theft, homicide and drugs, Agarwal et al. 2013 [11] conducted an in-depth analysis of crime data from both quantitative and qualitative perspectives. The algorithm is based on the K-means algorithm in machine learning. It uses the basic idea of clustering to analyze crime record data from New South Wales, Australia. Homicide, attempted murder, child vandalism and careless driving resulting in death were analyzed to validate the method. The data set used by Gera et al. in 2014 [12] is the Delhi police's first information report (FIR) Report. The authors in this paper analyzed the main types of crime in the city and classified the crime type into heinous crime, non-heinous crime, and particular and local law violations. The area attributes in the paper can be divided into slums, residential areas, commercial areas, and zones-high security zone areas for very important people. Finally, these data are fed into the clustering algorithm, and the crime prediction system was developed. The method proposed by Sivaranjani et al. in 2016 [13] presents a methodology for Tamil Nadu data from the National Crime Records Bureau of India, using a cluster analysis approach to analyze crime information for 6 cities, and the time scale of the data is about 15 years. It is from 2000 to 2014, and it has 1760 instances and 9 attributes selected to describe the crime data. The basic structures include K-means, agglomerative clustering and density-based spatial clustering of Noise. The accuracy, recall and F-measure are used to test the performance of the method. The final clustering results are also visualized through integration with Google Maps. The method proposed by Mehmet et al. in 2017 [14] utilized the NIBRS dataset of 2013 crime records from the US, and the association rule was used to construct the relationships from the extracted features from the dataset to predict crime type. The data attributes include state, population, incident data hour, location, sex of victim, race, age. The effectiveness of the algorithm is measured by the confidence indicator.

Logistic linear regression models, as a type of clustering algorithm, are also commonly used in crime prediction. And the linear regression models are one of the simplest methods for making crime predictions and can be used to achieve data-based predictions by assigning weights to different attribute values (Eq. (1)).

$$Log[p/(1-p)] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_i x_i \tag{1}$$

Here, x_i is data attributes for crime prediction, such as the variables which have high relationship with the crime. For example, for the type of the theft, the time and the weather is the related properties. The i was the attribute weights in these data, and it was also known as the partial regression coefficient. In crime prediction procedure, the various attributes play different role in different crime prediction.

In theft, the time variable has higher relationship than the weather, and the weight of the time is bigger than the weight of the weather. *P* is the probability value of predicting the possibility of crime. In the prediction process, the existing crime data is input into the model for training, and the final regression coefficient value is obtained after several optimization. Finally, the model is verified in the test set to determine the final weight value. This model is quite simple, but because its linear characteristics may be inconsistent with real crime events, the generation mechanism of many crime events is complex and generally non-linear, so there may be a problem of low accuracy in the prediction of crime by logistic regression model.

McClendon et al. [15] use the Mississippi crime dataset as the training dataset. And the same finite set of feature extracted from the Communities and crime dataset are input to the linear regression, additive regression, decision stump algorithm, then the linear regression got the best performance among the selected models.

The method proposed by Ma et al. [16] used four models including logistic regression, classification and regression tree, Chi-square automatic interactive detection and multi-layer perceptron neural network to predict crime data and compared the accuracy of prediction, and sorted out the risk of prisoners and the current situation of recidivism assessment at home and abroad. Based on the data of the 2004 U.S. Bureau of Justice Statistics (BJS) Survey of Incarcerated Persons (SISFCF), the authors compared the effect and performance with random forest were compared. Through the evaluation indexes of sensitivity, specific efficiency, accuracy and SUC, the experimental results showed that, the accuracy of the method based on classification, the method based on regression tree model, and the method based on the chi-square automatic interaction detection model are similar, and the accuracy of logistic regression model got a slightly higher accuracy. The method based on multi-layer perceptron got the highest accuracy. So, the method provided a experimental support to the method based on artificial intelligence. And the method based on neural network structure can improve the performance of the algorithm to a certain extent and achieve better prediction of crime events when the volume of the used data arrives at a certain number.

In addition to clustering methods, classification methods are also used in the crime prediction process. The method proposed by Das et al. in 2019 [17] conducts an overall analysis on the crime trend of Indian states and federal territories through classification algorithm, and the results showed that the classification algorithm also achieves good results in the trend prediction process. For the analysis of violent crime, the fuzzy C-means method was used for clustering analysis by Yamini et al. in 2019 [18]. The proposed clustering methods include partitioning, hierarchical clustering, fuzzy clustering, density-based clustering and model-based clustering. The method used a dataset of US arrests in 1973, which included assault, murder, rape and other types, murder, rape and other types. The paper analyzed the significant relationships between relevant attributes and crime. The method proposed by Zhang et al. in 2020 [19] also use fuzzy association rules to mine the data form Chicago between 2012 and 2017, and the data form NSW between 2008 and 2012. Firstly, the pre-processing is done on the data by using the month as the base unit. Then a fuzzy function is constructed. Finally, predictive analysis is performed with fuzzy association rules to obtain the important attributes related to crime. The method proposed by Thomas in 2003 compared and tested prediction of violence risk with many methods. And the comparison results showed that decision trees performed better for crime prediction. There are several models of decision trees, and the algorithm uses mainly CART and CHAID, where CART is mainly used for classification, and the Gini Index minimum criterion is used to extract the appropriate features from existing crime data. CHAID is used to obtain more fine granularity results based on the interpretability of the interrelationships between the variables and to prevent the results from being influenced by incorrect parameters.

Most of the traditional prediction models are linear models, but the relationship between crime data and prediction results is relatively complex. This non-Gaussian distribution and multicollinearity affect the prediction performance of many methods. To solve this problem, Alves et al. [20] used the statistical learning method to the causality from the collecting data is so impossible that it is hard to classify the type of the crime. The relationship finding between the urban metrics and the crime becomes more important. The random forest regressor is selected to predict the crime, and the homicides in the urban is quantified. The precision can get 97% from this method. The results also found that the important urban indicator to predict the crime, and they are unemployment and the illiteracy.

Different regression methods: simple linear regression, multiple linear regression, decision tree regression, support vector regression, and random forest regression are selected to build a regression model to predict the crime in Aziz's paper [21]. The crime types are murder, rape, kidnapping and abduction, riots. The crime data collected from the official website of NCRB, and the data is covered from 2001 to 2012, and is classified by region-wise data and state-wise data. The performance criteria, such as R squared value and mean absolute percentage error, is used to evaluate the different advantages of different methods. The experimental results showed that the random forest regression got better performance from R squared value ans the error in region-wise data, and for the state-wise data, the best performance model is also the random forest regression. Different regression models based on regression algorithm are built [22], such as random forest regression, decision tree regression, multiple linear regression, simple linear regression, and support vector regression. The MySQL workbench and R programming are used to do the pre-processing of the collected crime data. The collected data covers from 2014 to 2020, and it is the data with state-wise, region-wise and year-wise. The evaluation criteria is also the R squared value and a MAPE value. And from the experimental results, the random forest regression got the better performance in the selected models. Feng et al. [23] used the series of the data in their proposed method. The time and the location of the prediction is given by the system. In order to balance the different class of crime data, multiple classes are merged into larger classes. There are many features are extracted, such as the dates, category, descript, DayofWeek, PdDistrict, resolution, address, the longitude of the location of a crime, the dome and the arrest situation. And the features are selected to make the accuracy be improved. Different machine learning method are performed in this paper, and they are the decision tree, K-NN and Naive Bayesian. The method also proved that the tree classification method is the better method in the selected methods.

The Geographical feature extracted from the selected crime data show the very important character in prediction of the crime, Hajela et al. [24] used the data with the variables, such as time, weather, location, census parameters like annual income, literature rate of the area. Some mapping techniques are used to make the hotspot identified in the crime place. The zero dimentional data, one dimentional data, and two dimentional data of the hot spot are analyzed. And the clustering and the grid-based approach are used to show the better performance of the proposed model. In Hossain's paper [25], it showed that the 50 percent of the total offence is commited by 10 percent offenders. And the four traditional machine learning methods including decision tree, and k-nearest, random forest, and adaboost are used to detect the crime pattern and give the classification type. The criminal activity dataset of San Francisco of 12 years from 2003 to 2015 is used to train the model. And the data set contains 878049 rows in a csv file. And the number of crime type is 39. The attributes of the crime data set are theft of property, other offences, non-criminal, physical assault, illegal drug, theft of vehicle, damage property, warrants, illegal entry, suspicious activity, kidnappings, robbery, fraud, forged, secondary codes. And the granularity of the crimes are in different months, days, hours, and police districts. Random forest gets the best performance in the selected models on the collected data.

Kim et al. [26] also used the K-NN and decision tree to predict the crime on the Vancouver Crime data set. And the dataset covered the 15 years from 2003. The geographical analysis was also done in this scheme. The hotspot was visualized on the map, and the incident points were clustered in 30-day time. The experimental results showed the boosted decision tree was super than the K-NN. Kumar et al. [27] classified the crime data are as Robbery, gambling, accident, violence, murder, kidnapping. Then the k-nearest neighboring algorithm was used to predict the crime. The root mean square error, mean absolute error and K-NN score were performed to test the effectiveness of the proposed algorithm. Tamilarasi et al. [28] survey the crime of woman offence in India, and gave the Indian state wise crime details. Independent features are standardized by the feature scaling technique. 6 different machine learning algorithms are compared on the test data, and they are decision tree, K-NN, linear regression CART, Naïve Bayes, and support vector machine. The K-NN outperformed than other machine learning algorithms.

Different with other data set, Forradellas et al. [29] analyzed the crime data of Buenos Aires, and a system named SEMMA was proposed. SEMMA means the methods processing, and they are sample, explore, modify, model and asses. The data covers the period from 2016 to 2019 which occurred in Buenos Aires. Some crime type was analyzed. There were homicides, theft, injuries and robberies. The key part of SEMMA is K-means, which was used to get the feature, and the neural network was used to predict the crime.

(2) Methods Based on Probability

The essence of prediction is to calculate the probability of a particular type of crime occurring based on existing data by mining the data for patterns and using inference algorithms to implement statistics on the characteristics embedded in the existing data. Bayesian networks are a supervised learning method that calculates probabilities using frequency substitution by constructing sample pairs of existing data and crime types and completes the basic inference process by using Bayes' law.

In order to further address the fact that many of the data obtained in the algorithm are only correlated, without any theoretical analysis of the determinants of crime, Keppens et al. in 2003 [30] introduces an inferential approach aimed at addressing the distinction between homicide, suicide, accidental and natural death. The method proposed by Thomas et al. in 2003 [31] focuses on violence crime and analyses long-term prediction based on more stable characteristics, while some short-term prediction generally uses some dynamic factors. A method of prediction based on a classification tree is proposed. A statistical analysis of the features and conclusions involved in the algorithm provides a good idea for further research. The method proposed by Agarwal in 2013 uses K-means clustering to extract crime patterns and predict crime based on the distribution of spatial data. The paper uses offence recorded by the police in England and Wales by office and police force area from 1990–2011, analyses homicide, attempted murder, child destruction, causing death by careless driving, and verifies the method's effectiveness. Unlike other methods, the method proposed by Sharma in 2014 [32] constructed an objective function to detect possible crimes using a decision tree approach and tries to minimize this function, and it was named by Zero-Crime. The method by Das in 2019 applies various classification techniques to analyze crime trends across Indian states and union territories in terms of accuracy, precision, recall and F-measure indicators for 28 states (Andhra Pradesh including Telangana) and 7 union territories in India. Federal Territories, crimes including kidnapping or abduction, murder, rape and dowry deaths were collected from 2001 to 2014, with different data types, including correlations such as for rape crimes, whose dataset contains information on the victim and the type of rape that occurred, for murder, the type of rape, and for murder cases, the age of the perpetrator and victim. During the model's training, KNN, Random Forest, Naïve Bayes, AdaBoost

and classification trees were used to implement the data prediction for the years 2001–2012, and the relevant parameters of the model were obtained. The construction of a crime prediction model was completed. The model prediction for kidnapping cases, where "purpose of kidnapping" was chosen as the category label, was constructed in a similar way to the algorithm theory, including the rape model. The method proposed by Rishabh et al. in 2020 [33] analyses common crime types in India, including cruelty by husbands and relatives, dowry, immoral trafficking, kidnapping, and molestation. The K-means method and the rapid miner tool were used to analyze the relationship between the variable attributes. The algorithm also analyzes the pattern of crime data formation through visualization. And the histograms is used to show the relationship between crime and related attributes and finds the likelihood of crime occurring at different times. The method proposed by Aldossari et al. in 2020 [34] used Chicago Police Department's CLEAR as the processed data, selected nine features from the dataset, and verified that the decision tree outperformed the Bayesian network. The method proposed by Krishnendu et al. in 2020 [35] analyses the region and age in which more crimes occur. And the Kaggle dataset on India crime data was used, and the dataset spans from 2001 to 2010, including 1053 values. The data was analyzed using K-means, and the results showed that the most female criminals and the most male criminals are from different places. The method proposed by Kanimozhi et al. in 2021 [36] also used the Kaggle dataset to investigate the relevance of attributes. This method can effectively overcome the problem of existing algorithms with treating different attributes in a isolation form, and it can locate the type of crime and hot areas from time and location of the case as attributes. The experimental results compared the methods in terms of accuracy, precision, recall and F1 score to verify the algorithm's effectiveness.

Ramasubbareddy et al. [37] analyzed that the apriori algorithm is used to analyze and forecast the chance of some crime. The decision tree is used to search the crime pattern, and the naïve Bayesian classifier is used to predict the crime in particular geographical location at a particular point of time. However, the dataset used in this paper is created hypothetically, and the representation performance of the data is questionable. The HTML and CSS along with PHP is used to construct the crime prediction system. The system can be used to predict the dangerous place and help people to raise the awareness.

In order to identify the different effected factors, find the high occurrence relations of crimes. Bandekar et al. [38] used Bayesian, Levenberg and Scaled algorithm to train and test the collected crime data which are form Public domain data, national crime records bureau. From the experimental results, the scaled algorithm got the best performance, and the statistical analysis was performed from the correlation, ANOVA and the graphs. In Babakura's method [39], the data is collected from US census in 1990, and from the US LEMAS survey from 1990, and the FBI UCR from the 1995. The type of the crimes is classified by two models, and they Naïve Bayesian and Back Propagation (BP). The evaluation criteria are accuracy, precision and recall value. And the Naïve Bayesian got the better performance.

The Naive Bayesian and Decision Tree are also implemented by Iqbal et al. [40] on the data included the socio-economic data from 1990 US census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR. The attributes are state, population, income, education level, employment situation and the crime category. They also fond that the decision tree is super than the Naive Bayesian algorithm.

In order to discovery the hidden relationship of the collected crime data, Niyonzima et al. [41] analyzed many machine learning methods, such as multilayered perceptron, Naive bayes classifiers, and support vector machine. The dataset is collected by the UCI, and contains 128 attributes and 1994 instances. The experimental showed the advantage of the decision tree. Liao et al. [42] put the factors

related to the geographical distribution into two parts. One part is the characteristics of the victims, such as the gender, age, jobs, race and etc. And the other part is public regions, such as school, subway, bus stop, hospital, square, park etc. The modus operandi and the criminal psychology were classified as the unrelated factors to the geographical distribution. Bayesian learning theory and geographic information probability distribution were combined to make the prediction of the crime.

(3) Methods Based on Hot Spot Analysis

The traditional approaches are primarily statistical, describing crime data separately, through basic statistical theoretical models. It focuses mainly on the attributes of the data itself, and lacks further exploration of the correlations between the data. However, in fact, in addition to the information in the spatial dimension, the information in the temporal dimension also plays an important role in predicting crime to some extent. The existing literature suggests that some crimes have Spatio-temporal hotspots, such as robbery, mugging, or theft have characteristics based on temporal aggregation.

The hotspot map is one of the most important methods for crime analysis. The method proposed by Chainey et al. in 2008 [43] is used thematic mapping of geographic area, point mapping, grid thematic mapping, kernel density estimation, and spatial ellipses for the area in Central/North London and Camden and Islington within the Metropolitan police force area. Chaney's method uses attributes such as location, size, shape, and orientation of clusters of crime incidents. Four types of crime were studied: street crime, burglary, theft of vehicles, and theft from vehicles. By comparing the experimental results, it has not only been verified that hotspot plays an important role in crime prediction, but also that the different mapping methods are used in different situations. The hotspot method proposed by Carter et al. in 2020 [44] investigates the criminal investigation data of Coral Gables, Florida, from 2004 to 2016 through the Emerging Hot Spot Analysis (EHSA) method. He focused on Residential and vehicular Burglary, and utilized the various sociodemographic variables in crime forecasting. These variables include commercial areas, Renter percentages, median household income, and multifamily households. The method of geographically weighted regression (GWR) in the clustering method was selected to predict. Finally, the prediction method was verfied by social disorganization theory and routine activity theory. This verification shows that using geospatial analysis to target specific locations has better advantages in solving property crimes. The method proposed by Cowen et al. in 2019 [45] utilizes techniques such as OLS regression models, harmonic analysis of diurnal patterns, and geospatial statistical technique. The sample data is from the 2007–2015 period of 782 Census Blocks in Miami-Dade County, Florida. Crime data contained 298,111 incidents. 97 different land-use types were included in the land-use data. Variables include walkability index, distance from public transportation (PT), distance from bike lanes (BLs), Intersection Density (ID), and Amenities (AM). Focusing on neighbourhood crime rates of larceny and aggravated assault, Cowen utilized conventional and geospatial analyses, routine activity and social disorganization theories of neighbourhood crime to analyze the algorithm. This analysis also verifies the validity of the method. The method proposed by Umair et al. in 2021 [46] introduces Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) to analyze the hot-spots (Hot-spots) segment. Meanwhile, the proposed algorithm is based on Seasonal Auto-Regressive Integrated Moving Average (SARIMA) to effectively predict spatio-temporal correlation data of crime, based on several New York boroughs, Bronx, Brooklyn, Manhattan, Queens and Staten Island. Also, the data from 2008 to 2017 were studied. Butt's method used MAE (Mean absolute error), MAPE, RMSE, ME as test indicators to verify the algorithm's effectiveness.

The spatial information from the collected data is extracted by geocoding in Bappee's method [47] to get the location feature of the crime. On the other hands, the creation of hotspots is the

another selected feature to help improve the performance of the prediction performance. The model named hierarchical density-based spatial clustering of application with Noise is used to find the shorted distance between the points, then the hotpoints is chosen. And the crime type is classified. Bogomolov et al. [48] get the aggregated and anonymized human behavior data and this data is collected from mobile network activity. The proposed combined the mobile network and the basic demographic information, which is different with the traditional method that used the historical knowledge or offenders' profiling. The dataset is the open data institute and MIT during the campus party Europe 2013 at the O2 Arena in London in 2013. and they are the geo-localized open data and the anonymized and aggregated human behavioral data. There are 5 modes used in this method, they are the decision trees, logistic regression, support vector machine, neural networks, and the different implementations of ensembles of tree classifiers with different parameters. From the criteria named Accuracy, F1 and AUC, the best model is random forest.

In order to solve the poor performance of the traditional method that only uses the single and the distribution of crime, He et al. [49] use the generative adversarial network to improve the performance and get the visualization of the crime distribution map. 2 million crime information collected from Philadephia are used to train the model, and the data covers from 2006 to 2018. The crime character of the low population density area has the strong sparseness, which make the data in the crime prediction system imbalance. In order to resolve this problem, Kadar et al. 2019 [50] use three features: spatial features, temporal features and crime features. Then the random forest, adaboost and logistic regression are used to train the data collected from the Swiss canton of Aargau from 2014 to 2017. The experimental results showed the advantage of the proposed on the low population.

For the selection of input attributes for machine learning, many reasons can lead to unexpected situations, such as weather, holidays and some unexpected events. For example, in weather, some people consider the following causes, such as atmospheric temperature, relative humidity, wind speed, precipitation, hours of sunshine, atmospheric pressure, cloud cover, fog visibility, etc., are all related. For example, temperature, relative humidity and sunshine hours are significantly correlated with the occurrence of criminal activities, and the frequency of violent assault crimes increases linearly with temperature. There is a more significant positive correlation between relative humidity and property crime (Cohn), and Lab demonstrates that the number of rape crimes is inversely proportional to sunshine hours. For example, most robberies occur in the summer, as the property violation can let the criminal subject know the basic property of the victim, which is easily exposed in the summer due to the relatively little clothing.

(4) Methods Based on Mixed Algorithms

In Nguyen's method [51], the support vector machine, neural network, random forest, and gradient boosting machines are compared, the demographic information of the area, educational background, economical and ethnic background are collected as the variables of the training data. And the crime data is from the jurisdiction of Portland police bureau. And the time of crime data covers 2012 to 2016. Experimental showed that the different algorithms have different performance on the different dataset. Reddy et al. [52] used the machine learning and the visualization tools to predict the crime rate and the crime area. The crime data of the UK police department are used to train the proposed method. The data included 5 attributes, and they are the crime type, location, data, latitude, and longitude. The data is collected from 2015 to 2017. K-nearest neighbor, Naïve Bayes was used to predict the crime. And the conclusion showed the effectiveness of the proposed method. For the randomness of the crime, the importance of the selected features, such as the time, data, area, or the geologically relevant feature is analyzed in Nitta's paper [53]. Different foresting models based on

machine learning like naïve Bayes and SVM are compared. The results showed that the LASSO-based feature is the best feature, and naïve Bayes classifier is the best classifier among the compared the selected classifiers.

The San Francisco crime dataset is used to analysis the internal relationship of the selected data in Pradhan's paper [54]. the dataset includes the incident number, day of the week that crime occurred, police district the crime occurred, resolution, street address, the longitudinal, the identifier, and the category of crime. The similar categories are combined. The Naïve Bayes, Decision Tree, Random Forest, K-NN and multinomial logistic regression are compared on the selected dataset. And the proposed can alleviate the imbalance dataset and improve the performance of the multi-class classification. Safat et al. [55] analyzed the crime rate, type and hotspot. The traditional machines learning methods, such as support vector machine (SVM), autoregressive integrated moving average (ARIMA) model, k-nearest neighbors (KNN), the logistic regression, Naïve Bayes, decision tree, eXtreme Gradient Boosting (XGBoost), multilayer perceptron (MLP), and random forest are analyzed in this proposed, and the deep learning methods which is used to analyze the time series analysis named long-short term memory (LSTM). The data set included the Chicago and the Los Angeles. The experimental results showed that the different algorithm got different performance on the different both dataset. Saraiva et al. [56] combined the data mining and the machine learning method into the proposed method. The crime dataset was collected from the Porto, Portugal, and covered from 2016 to 2018. The spatial patterns and the relevant hotspots are got from based on the random forest and decision tree. And the tweets related to insecurity was used to analyze in order to get the interpretation of the pattern. And the experimental results showed the effectiveness of the proposed methods.

In order to increase the efficiency of crime investigation system, Shemila et al. [57] proposed a novel crime prediction system based on the traditional methods, such as the multilinear regression, K-neighbors classifier and neural networks. And the system had two parts: crime analysis and prediction of perpetrator identity. The dataset is collected from the San Francisco, and covered from 1981 to 2014. The selected crime type is homicide. Tamir et al. [58] also analyzed the crime prediction performance based on the neural network on Chicago police open dataset. The city districts and the crime trend of year are deep analyzed. And the Folium is used to visualized the trend. Many feature variables are analyzed, such as the data, year, location information, primary type, and so on. The random forest, K-NN, and the AdaBoost were also performed, and the experimental showed the effectiveness of the Neural Network. The dataset of Zaidi's paper [59] combined the numerical and categorical features. Two traditional methods named random forest and support vector machine are used to compare the performance, and the random forest is verified with better performance than the support vector machine. Lin et al. 2018 [60] considered the importance of geographic information, and proposed grid-based crime prediction on the crime dataset collected from the Taoyuan city. The performance of method based on the proposed deep learning were compared with the traditional methods, such as the KNN, random decision tree, support vector machine. The criteria of F1 were used to verified the effect of the proposed method.

From the analysis, the basic pseudocode of the algorithm based on the conditional machine learning is as follows: The Input is training dataset: $D = \{(\mathbf{x}^{(n)}, y^{(n)})\}_{n=1}^{N}$, and the test dataset: V, the parameters of the model: θ

Step 1: Select the related data variables, $x_i \in \mathbf{x}^{(n)}$

Step 2: Get the representation (basic feature) of the selected crime data:

$$\mathbf{f}_i = \text{FeatureExtraction}(\mathbf{x}_i) \tag{2}$$

Step 3: For n = 1:N do

Compute the Loss from the models:

$$Loss(\theta, \boldsymbol{x}^{(n)}, y^{(n)}) = LossFuction(\text{model}(\boldsymbol{x}^{(n)}), y^{(n)}))$$ (3)

Update the parameters of the models:

$$\theta^* \leftarrow \theta - \alpha \frac{\partial Loss(\theta, \boldsymbol{x}^{(n)}, y^{(n)})}{\partial \theta}$$ (4)

Output: The models with parameters: θ^*

2.2.2 The Models Based on Deep Learning Methods

With the increase of the crime data volume and the more richness of the crime content, the deep learning methods are brought into the prediction of the crime data. The basic scheme of these methods are convolution part, pooling part, activation part, full-convolution parts, and so on. The deep learning methods can be classified into the basic convolution network, the network for the series, and the some special networks, such as, 3D neural network.

(1) Method Based on Convolution Network

A novel crime prediction model is proposed by Duan et al. [61] named spatio-temporal convolutional neural networks (STCN). This model can predict the possibility of the next day's crime risk. And the train data is the felony and 311 datasets in NewYork city from 2010 to 2015. The F1 and the AUC are used to assess the performance of the model, and the results show that the proposed deep learning method is effective. The model also make the visualization come true and help the citizens realize the relationship of the crime and the collected data. Given the sparse spatial and temporal distribution of the data used for crime prediction, many deep learning models do not work well. This paper uses non-emergency service request data as training data. Fine-grained crime prediction is achieved by using Deep inception-residual networks (DIRNet). The experimental data in New York City burglary-related crime data from 2010 to 2015 and the method's effectiveness is shown by comparing it with SVM, RandomForest, STCN and ST-ResNet using the positive F1 metric. The method proposed by Wang et al. in 2018 [62] proposes a network construction method based on spatio-temporal give influence to achieve crime prediction using node strength. The authors effectively combine the features and performance of degree, average degree and aggregation coefficient in complex networks, and use them as parameters of complex networks for performance analysis to achieve the extraction of key nodes in crime prediction and combine temporal and spatial distributions. Developed based on graph theory, complex networks are complex systems consisting of individuals and relationships with each other, formed based on dynamics. In general, points correspond to individuals in the physical world, and lines correspond to interrelationships between individuals. The combination of points and lines describes the phenomena in the existing physical world.

The history of the arrest bookings are used to predict the crime in chun's proposed method [63]. The full connected neural network is used to classify the crime. The collected crime dataset covers from 1997 to 2017. And there are 16841 unique people with 63133 records of arrest record, 42 unique crime types with 3 level of seriousness are recorded. And the data augmentation and some defined loss function are proposed to make the dataset meet the requirements. The proposed method showed that the data pooling method of looking at all the possible historical years per person is the best in the comparison methods. Kang et al. in [64] proposed a method base on feature fusion scheme bsed

on deep neural network. The crime data also presented multi model character. And the data used in this paper has various online databases of crime statistics, demographic and meteorological data, and images. And the prediction of crime occurrence highly nonlinear relationships, redundancies, and dependencies between multiple datasets. And the content of the data also has many types, such as crime occurrence reports, demographic, housing, economic, education, weather, and image data. 274,064 of 31 crime types cases in 2014, 2014 American Community Survey (ACS) data from American FactFinder, weather and image data by using the Weather Underground API and the Google Street View Image API, Weather data were captured from the daily weather history of Chicago. Lin et al. [65] used broken-windows theory and spatial analysis to predict the drug crime. The traditional machine learning methods, such as the random forest, Naïve Bayes, were compared with the proposed deep learning method. Experimental showed the effectiveness of the deep learning algorithm. Matereke et al. [66] compared three deep learning methods on the performance of the crime prediction, and they were deep multi view spatio-temporal network, spatio-temporal residual network, and the spatio-temporal dynamic network. The evaluation criteria were Mean absolute error, and the root mean square error. The experimental results showed that the best effectiveness of the three methods is spatio-temporal dynamic network. Wei et al. [67] use the data collected in New York city from 2014 to 2016 to train the proposed scheme. And the intricate spatial temporal-categorical correlations are analyzed in this method. Some evaluation matrix such as the root mean square error, mean absolute error are used to test the performance of the proposed method and the existing method, the results showed the effectiveness of the proposed method.

(2) Methods Based on Series Neural Network

The method proposed by Muthamizharasan et al. in 2022 [68] uses data from the NCRB dataset, which includes four aspects of crime, namely rape, murder, theft and offences against property. The use of CNN-LSTM as a model to predict crime, due to the good sequential predictive power of LSTM, The algorithm achieves better prediction accuracy compared to the original spatio-temporal statistical model-based algorithm. The method proposed by Deepak et al. in 2021 [69] introduces that the system's architecture consists of three different components: data preparation, classification and ontology construction. The paper uses the Bi-LSTM for classification problems as a fundamental part of the algorithmic framework, using data collected from Google News and Twitter as training data. In the initial labelling process, the Fuzzy c-means algorithm and the Term Frequency-Inverse Document Frequency Vectors method were utilised for labelling and feature extraction, and four different datasets were used (UCI community and Crime dataset, Fraud and civil action, CAIL, Crime in India). In the above datasets, the UCI community and Crime dataset was extracted from socio-economic data, specifically from 1990 US LEMAS, 1990 US census, and FBI data, all of which is real-time data. Fraud and civil action includes crime cases relevant to civil action, fraud, etc. There are 40256 cases in the CAIL dataset from Supreme People's Court. The CAIL dataset is from Supreme People's Court, and the Crime in India Dataset is a publicly available dataset from the machine learning competition Kaggle. The method proposed by Wang in 2018 proposes a network construction method based on spatio-temporal give influence to achieve crime prediction using node strength. The authors effectively combine the features and performance of degree, average degree and aggregation coefficient in complex networks, and use them as parameters of complex networks for performance analysis to achieve the extraction of key nodes in crime prediction and combine temporal and spatial distributions. Developed based on graph theory, complex networks are complex systems consisting of individuals and relationships with each other, formed based on dynamics. In general, points correspond to individuals in the physical world, and lines correspond to interrelationships between individuals. The combination of points and lines describes the phenomena in the existing physical

world. Wang et al. [70] also used the deep learning scheme named long short term memory (LSTM) to train on the crime dataset was collected in Atalanta from 2009 to 2016. The dependence in time lag and spatial distribution of criminal events are got by this proposed method. The time series make the length of the days is 50, and the spatial cell size is 0.05 degree. From the correlation coefficient value, the proposed method was effect and useful to the law enforcement agencies. Wawrzyniak et al. [71] also use the long short-term memory (LSTM) recurrent neural networks (RNN) and convolutional neural networks (CNN) to get the spatio-temporal distribution of the hot-spots.

(3) Method Based on Some Other Neural Network

Transfer learning [72] is a powerful learning method to use the related learning strategy by the different data to predict or forecast the collected data. The single and multi-domain representations are used to evaluate the performance of the classification. The data set is from three different cities. And it is collected form the regional police department, covers most of the dissemination areas. The Toronto and Vancouver cities is the source domains, and the Halifax is the target domain of the cross-domain transfer learning.

Irrelevant factors in the collected urban crime data may affect the performance of the crime prediction. In order to resolve the effect of the inhomogeneous noises—local outliers and irregular waves, Hu et al. [73] proposed a paradigm named Duronet, that is the Dual-robust enhanced spatial-temporal learning network. The proposed method is a encoder-decoder architecture can reduce the noise effect. Two types of noise are considered, and they are the local outliers and the irregular waves. Optimized spatial-temporal representations with different time slots are used to get the better representation pattern of the crime data. Self-attention module are also brought to reduce the influence of the irregular waves by the different weights. The data is the real-word crime records in Chicago from 2016 to 2017, and in New York from 2015 to 2016. The crime time, type, location and the arrest status are contained in the crime data. Compared with the traditional methods, such as the Support Vector Regression, Linear Regression and Random Field Regression, and some deep learning methods, such as LSTM, NN-CCRF, and some other series method, like ARIMA, the proposed method showed the better result than the other methods.

Inter-dependencies between crimes and other ubiquitous data [74]. Capture the relevant occurrence of different categories. The temporal, spatial, and the categorical signals are jointed embedded into the crime vectors. And the attentive hierarchical recurrent network are used to get the dynamic pattern. The data source is collected from New York city. From the visualization of the Geographic distribution of the occurrence with different categories, and compared with the other methods, such as auto-regression integrated moving average, multilayer perceptron, tensor decomposition, wide and deep learning, and gated recurrent unit, the proposed DeepCrime system got better performance on four types of crime, and they are burglary, robbery, felony assault, and grand larceny. The spatial dependencies and the temporal recurrence are used in Sun et al. [75], a prediction methods based on the Gated Recurrent Network with Diffusion Convolution modules is proposed, and is dubbed CrimeForecaster. For the generalization performance of the end-to-end scheme, the proposed method outperformed than other algorithms on the trained dataset. For the clearness of the basic theory of graph model, Wang et al. [76] analyzed the crime prediction based on the graph theory. The spatial temporal feature and the graph generation are got by Multi-Graph Convolutional Network (DT-MGCN) model to predict the crime rate. And the model included two graph model, and they were graph convolution network (GCN) and the encoder-decoder temporal convolutional network (EDTCN).

From the analysis of the existing algorithms based on deep learning, we can get that the algorithms applies relevant techniques such as natural language processing and video processing to crime prediction to build sequence-based analysis methods while being able to introduce attention mechanisms to find the most likely factors associated with crime attributes. And the basic procedure can be summarized as follows: the input: training dataset: $D = \{(x^{(n)}, y^{(n)})\}_{n=1}^{N}$, test dataset: V, the parameters of the model: $\boldsymbol{\theta}$.

Step1: Select the related data variables, $x_i \in \boldsymbol{x}^{(n)}$

Step2: Select the network model, such as LSTM, RNN, and so on:

Step3: Some network parts can be selected, such as Attention Network;

Step4: Train the network:

Step5: For n = 1:N do

Compute the Loss from the models:

$$Loss(\theta, \boldsymbol{x}^{(n)}, y^{(n)}) = LossFuction(\text{model}(\boldsymbol{x}^{(n)}), y^{(n)})) \tag{5}$$

Update the parameters of the models:

$$\theta^* \leftarrow \theta - \alpha \frac{\partial Loss(\theta, \boldsymbol{x}^{(n)}, y^{(n)})}{\partial \theta} \tag{6}$$

Output: The models with parameters

2.2.3 The Models Based on Different Media Type

The type of the crime data is multi-modal. The most modal of the existing algorithms is text. And the volume of the test data is relative small. So the traditional machine learning algorithms are used in these data, and they are outperformed than the deep learning methods. The article news includes much information of crime, then the natural language processing techniques are brought into the algorithms. In order to resolve the error of traditional method that people analyze the crime situation from the news and maybe consume vulnerable time, Ghankutkar et al. [77] proposed a crime prediction system base web to get a better estimate of future crimes. The online news articles with real-time character are used to be analyzed and then be classified by three classifiers. And the results are visualized on the system and it helps the people to know the crime situation of the selected predicted location. Li et al. [78] analyzed the text from the law department. The dataset was a real-world dataset contained 41481 judgment documents of the theft cases which can got from the China Judgements Online. The judgment-specific case features were extracted by the proposed algorithm. The Linear regression and the long short term memory were combined to make the prison term prediction and the natural language processing techniques, such as the word embedding, were also used to improve the performance. Umair et al. [79] transformed the spatio-temporal analysis into natural language processing scheme. The dataset used in this methods was the data crawled from the newspaper, such as the Dawn News, Dunya News, Ary News, The News, Daily Parkistan, Panistan Press Foundation, The Nation, and Journalism Pakistan by the Data Miner Tools. The Random forest and the k-nearest neighbor are used to train the crime dataset. Boukabous et al. [80] bring the Bert into the crime prediction which is usually in the natural language processing. In order to resolve the label problem by the lexicon based Bert of the crime prediction, this paper combined the deep learning and the lexicon-based. Lexicon-base method is used to label the Twitter dataset. Bert is used to train the dataset. The

criteria of evaluation is accuracy and F1-score. And the experimental result showed the effectiveness of the proposed method.

The other modal of crime is video collected from the video surveillance system. The video Chackravarthy et al. [81] predicted the crime by the video. Firstly, the video is separated into frames. Then the hybrid deep learning method is used to detect the person or object. The recurrent neural network is designed to extract temporal activities of a person. And the temporal behavior from the video streaming is combined with the feature of the former hybrid deep learning feature, then the model is built to predict the crime. Some anomalous behaviors, such as the car theft, the assault are detected by this method, and the experimental results show the effectiveness of the proposed method. Some anomalous behaviors, such as the car theft, the assault are detected by this method, and the experimental results show the effectiveness of the proposed method. 3D convolution neural network is the most popular model in multi-dimention data analysis, such as the behavior decision in video. Fan [82] bring the 3D in the crime prediction algorithm. And the source data is the open-source historical video. And the data is the behaviour, and 100 of a total of 600 groups are used to train the model. Because of the integrity protection of the feature in video by 3D CNN, the presentation power of the feature can improve the prediction performance of the crime. In order to resolve the effect of the noise in the video surveillance, and improve the prediction rate of the crime. Kumar et al. [83] proposed the lion-based deep belief neural paradigm (LbDBNP) based on the data of activities and handling tools. Three types of crime dataset were used to train the algorithm, and they were UCSDped1, UCSDped2, and avenue crime. The robustness ware outperformed than the other algorithms. Different with other algorithms, Crime Intension Detection System is developed by Navalgund et al. in [84], and they focus on the crime video and images. The VGGNet-19 is used to train the captured video or image from the circuit television cameras. And the SMS sending module is used to detail time information of the crime people. The GoogleNet Inception V3 is used to pre-trained. The FastRCNN and RCNN are also used to draw the bounding box to the detected object, such as the gun, knife, people, and so on. Shah et al. [85] analyzed the existing methods which are used in the crime prediction, and proposed the new scheme based on the machine learning and the computer vision. Rajapakshe et al. [86] combined the CNNs and RNNs to detect the abnormal behaviour from the video surveillance system. And the abnormal behavior can be categorized as the violent action, robberies, fast movements, vandalism, breaking and entering. The training data is the internet sources from the UCF crime dataset. The VGG16, GoogleLeNet (Inception), ResNet50, encode temporal feature based LSTM are used to compared the performance of each scheme. The traditional methods are also used to get the prediction of the crime, and the experiment showed the higher performance of the deep learning method.

Social media is an important carrier supporter to declare the peoples' idea, so Abbass et al. [87] analyzed the twitter tweets. Three steps are existing in this methods, and they are data pre-processing, data classification and crime prediction. Some machine learning methods, such as Multinomial Naïve Bayes (MNB), K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) are used to classify the crime type. And the best value of n is identified by N-Gram language model. The experiment is conducted by three criteria: precision, recall, and F-measure, and the results showed that SVM is the better method to predict social crime.

Ippolito et al. [88] used the data from the taxes audits of the municipality of São Paulo. Then some machine learning methods are used to train the data, and the results is that the random forest performs better than other methods, such as the neural networks, decision trees, naive bayes and ensemble learning.

The crowd sourcing information that collected from the social networks was used in Deshmukh's paper [89], and the reports from the 240 official first information report department. The author developed a app to predict the crime to help the people search the possible crime in Mumbai. The google maps sdk is used to get the GIS information, The random forest algorithm is used to train the collected data. The robbery, theft and the murder are the predicted crime types in this system. The experimental results showed the it was effect to the crime data.

The mobility data used in Kadar's method [90] also used the NYC open platform dataset, and the data covers from 2014 to 2015, includes 174,682 incidents across five boroughs. And five felony types are included, and they are grand larceny, robbery, felony assault, and grand larceny of motor vehicle. Otherwise, the data of Foursquare venues and checkins, subway rides, and taxi rides data are used to train the model. The tree-based model makes the crime prediction system more transparent.

3 Further Research Plans

3.1 Construction of Common Data Sets

Datasets play an important role in improving the performance of algorithms, and the release of publicly available shared datasets can facilitate comparisons between different algorithms, leading to improvements in individual methods. The relatively small number of publicly available datasets for crime prediction and the significant shortfall in the size of the datasets are important reasons why AI-based crime prediction methods are seriously lagging behind other fields. Therefore, constructing datasets of specific data sizes with richer data types and more complete annotations will play an essential role in enhancing the field's rapid development.

3.2 Study on the Interpretability of the Model

The interpretability of the model is essential through the study of the algorithm's mechanism, developing or finding the data attributes that are relevant to crime prediction. Many methods are only used in the laboratory, and the fundamental reason for this is the weak interpretability of the models. Constructing models with a more robust theoretical basis enables an adequate explanation of the algorithm. In the future, visualisation can be used to present each step of the algorithm visually and highlight the algorithm's specific process, thus making the model more usable in practice.

Interpretability is the most import part in the method base on machine learning, Zhang et al. [91] analyzed the contribution of the variables in the crime prediction based on machine learning. The routine activity theory and crime theory is used to analyze the relationship of the variable. The train model is XGBoost. The contribution of individual variables is discerned by the Shapley additive explanation. For example, the paper found that the theft has high relationship with the age, and the ambient population aged 25–44 have more tendency of theft. Rayhan et al. [92] concerned the interpretability and the accuracy of the crime prediction system. The interpretability can help the related people to realize the important thing of the crime data, and can help them to prevent the possible crime step. And an interpretable attention-based deep learning model named AIST is propose. The spatio-temporal correlation are analyzed by the past crime occurrence data. The more possible feature of the related features are extracted. And the experimental results showed the advantages of the proposed method. In order to improve the interpretability of the back box performance of the random forest, Wheeler et al. [93] analyze the robberies in Dallas by the proposed method. Some other methods were also used in the testing procedure, such as the risk terrain models and kernel density estimation. And the results showed the advantage of the proposed method on the micro place.

3.3 Multimodality

Single-attribute predictions may suffer from a lack of attribute richness. Traditional forecasting processes only have spatial data, such as location and other relevant factors, and some add air conditions to the forecasting process to improve accuracy to a certain extent. However, for deep learning approaches, including more effective and additional attribute data allows for a richer factorisation of the forecasting process. As there is no direct theoretical evidence that a particular factor or factors are associated with crime evidence, predictions can only be made in a form similar to a package of words (BoW). This approach requires that the prediction process simply contains factors that can influence crime prediction, as to which influencing factors again; due to the current lack of clarity in technology or data, causal factors are not put into the prediction process. For the time being, only relevant factors are considered.

Kang 2017 in proposed a method base on feature fusion scheme based on deep neural network. The crime data also presented multi model character. And the data used in this paper has various online databases of crime statistics, demographic and meteorological data, and images. And the prediction of crime occurrence highly nonlinear relationships, redundancies, and dependencies between multiple datasets. And the content of the data also has many types, such as crime occurrence reports, demographic, housing, economic, education, weather, and image data. 274,064 of 31 crime types cases in 2014, 2014 American Community Survey (ACS) data from American FactFinder, weather and image data by using the Weather Underground API and the Google Street View Image API, Weather data were captured from the daily weather history of Chicago.

In addition, for factors other than spatial ones, such as temporal evolution and other related types of data, there may be a certain degree of relationship with the ups and downs of the psychological state of occurrence of crime.

3.4 Multi-Granularity

Gorr et al. [94] states that in to improve prediction accuracy, the number of crimes per unit time needs to be above 30, so the model can only predict up to the month and cannot be trained for shorter periods. In order to improve the scale even further and make the granularity of crime prediction even more refined, and to be able to allow the scale of crime to be based on weeks, Liu et al. [95]. investigated this problem. They introduced a density shift model for spatio-temporal prediction based on the assumption that crime hotspots are characterized by invariance over short time scales. Where microscopic transfer probabilities need to be calculated, the method is further reduced in the prediction scale to achieve week-level prediction, with good consistency in spatial hotspot distribution and actual results. However, the crime system has certain dynamic characteristics and non-linear characteristics. As the method uses a linear descriptive means for predicting spatio-temporal transfer trends, it limits the method's predictive effect due to the lack of non-linear capabilities. How to make the algorithm can achieve better prediction effect on the short-term and long-term and other scales, making the model can be adaptive with the short-term linear ability and long-term non-linear ability, so as to improve the generalization ability of the model, which is worth in-depth research areas.

In addition, besides some traditional data, can we add video and other related data? With the development of safe cities, smart cities and other information infrastructure, related crime high-frequency areas have also increased security monitoring equipment, making many areas present a programme to prevent no dead ends. Meanwhile focuses on some behavioural videos of particular behaviours, thus building a multimodal information framework structure. To a certain extent, this can improve the accuracy of crime prediction even further. For example, the findings of the study, Critical

issues in policing: contemporary readings (BWT), suggest that disruptions in the social environment, such as broken windows, rubbish and crashed cars, are responsible for an increase in criminal activity, i.e., crime occurs in areas where visual perception is impaired.

3.5 Mixed Algorithms Based on the Traditional Algorithm and the Deep Learning

From the analysis, we can get the conclusion that the better performance is related the complex structure of the neural network. For the big volume of the crime data, the content of the data set become rich, and the representation of the data become strong. The robustness of the algorithm is better than the traditional machine learning for the end-to-end character. However, the understanding of the proposed structure of the method become confused black-box character. So, the mixed algorithm which combined the traditional techniques and the deep learning techniques becomes more and more popular.

In order to compare the performance of the crime prediction system, 5 examined datasets of the US are included in Stalidis et al. [96], and they are the Philadelphia, Seattle, Minneapolis, DC Metro, and San Francisco. The training data is the time series of crime types per location. 10 traditional methods and 3 different deep learning configurations are performed on the train data. The results showed the advantage of the methods based on deep learning. In order to resolve the sparseness of the crime data and the weakness of the crime signal of interest, Wang et al. [97] used spatial temporal residual network to predict the crime on the Los Angeles data. And the proposed was a real-time crime forecasting system. Some existing traditional machine learning algorithms are used to compare with the proposed method. The results showed the effectiveness of the proposed method. Wu et al. [98] used Bayesian network, random tree and neural network to predict the crime. And the dataset is collected in YD country from 2012 to 2015 and was divided into the crime of smuggling, selling, transporting, manufacturing and possessing of drugs. Illegal business, rape, fraud and gang fighting. With the dataset Chicago and NewYork dataset, Xia et al. [99] proposed a crime prediction system based on the Graph Neural Network. The hypergraph learning, temporal and spatial relation learning are used to visualize the learned weights. Many models are used to compare the performance, such as the ARIMA, SVM, ST-ResNet, DCRNN, STGCN, STtrans, DeepCrime, STDN, UrbanFM, ST-MetaNet, GMAN. The relationship of the different variable-wise are discovered by the proposed system, so the system got best performance than the compared methods. Yi et al. [100] combined the traditional method named conditional random field and the neural network named long short-term memory (LSTM) to improve the performance of the crime prediction. The traditional method have the strict process and the neural network have better performance based on the end to end structure. So the method can not only improve the performance, but also can improve the interpretability. 1,072,208 crime records in Chicago from 2013 to 2015, and 1,417,083 in New York City from 2015 to 2016 are used as the data set of the proposed crime prediction system. The results showed that the system is more effective and convenient to get the correlation of variables and the prediction system is large-scale and fine-grained.

3.6 Building a Multi-Knowledge Dimensional Crime Prediction and Analysis System

Finally, work towards building a data-driven crime prediction system, as shown in Fig. 2, which that should have multiple layers of crime data collection, cleaning, collation, storage, visual display and content understanding. In addition, different machine learning methods are constructed, specifically traditional machine learning methods and modern artificial intelligence methods, and crime is defined and classified according to specialisms.

(1) Crime data collection system: This system is a distributed, unit-oriented data collection system that can be shared and reused effectively. Various data collection tools, including established public data collection methods or customized methods, are used to collect various data. The data collected includes structured data, semi-structured data and even some unstructured crime-related data, capturing as many and comprehensive types of data as possible. A rich and comprehensive range of data types is essential for the predictive power of crime prediction models. Hence, a crime data collection system that is diverse in scale, comprehensive, realistic and accurate is an important prerequisite for building crime prediction models.

(2) Data cleansing component: The traditional non-data-driven approach to crime prediction requires to some extent finding attributes that are directly related to crime factors, the so-called orthogonality of the data. As the correlation between data attributes is limited by storage, it is desirable to exhibit orthogonality, i.e., to describe and express the essential aspects of the crime data using the minimum amount of data. Data that exhibits orthogonality is very effective at the request of traditional SVM-based, sliding average systems, Linear Discriminant Analysis, Principal Component Analysis and other methods. The redundancy in the description of these crime loud processes can make the crime prediction much less accurate. However, as technology develops further, especially artificial intelligence does not know which neurons are most relevant to the predictive accuracy of criminal activity due to its inability to take into account the correlation of things or the causal relationships between things, so this blind box approach can only replace the causal relationships between things by constructing correlations between data. Therefore, when using AI for crime prediction, a certain degree of redundant information enables relevant attribute factors to complement each other, improving crime prediction accuracy at the scale of massive amounts of data. However, it is not true that more redundancy is better. To a certain extent, noise pollution in the data can also make crime prediction less accurate. Therefore data cleaning is also required in crime prediction systems to ensure that the data stored is as strongly correlated with the prediction of crime as possible.

(3) Distributed crime database system: The large amount of data collected and the diversity of crime types, as well as the need to build different levels of viewing rights, require the construction of an effective database system with fast queries and easy access that is adaptive and extensible. The database allows for the normalization of crime data while ensuring data consistency. It provides the basis for further data processing.

(4) The Prediction Method Selection Library: No Free Lunch Theorem (NFL) states that if a method works on one problem, there is likely to be a possibility of under-performance in other domains. In the existing physical world, uniform, universally applicable machine learning methods with optimal performance on all problems do not yet exist. So each has its range of use based on different methods. And as each method has different types of parameters to be adjusted, it is even more important that a machine learning-based method for predicting crime should compare multiple methods for the same crime type and determine the final method based on the merits of the different results. Crime should compare multiple methods for the same crime type, thus determining the final method based on the strengths and weaknesses of the different results. In the prediction method selection library, multiple methods can be listed here. Then a comparative analysis of the different results can be carried out to hover over the relatively superior method. There are many new machine learning algorithms, such as the some supervised learning, and active learning, the new crime prediction can use these new algorithms to improve the performance.

(5) The visual analysis and results feedback part of the system: Although the fundamental mechanisms of many crimes are unclear, the relevant data attributes cannot yet be accurately

discovered. However, to provide interpretability of the model, i.e., to present the maximum possible correlation of the whole crime process, it needs to visualize the data, and here relevant issues such as attention mechanisms in deep learning can be utilized. The attention mechanism uses the construction of gating, which allows the strong correlates of data attributes to be visible and distinguishes them from other crime attribute factors so that in future when similar problems are encountered, the focus will be on the strong correlates of data attributes. In addition, the adjustment of this crime prediction model also requires more visualization of the results and the final real-world situation for effective adjustment, thus ensuring the robustness and generalization performance of the algorithm and model.

(6) Standardized development interfaces: crime forecasting involves different parts, while the hierarchical structure within the public security system is more responsible, so its maintenance requires a professional organization, as well as the standardization of the secondary development of these platforms to facilitate interfacing and utilization by third parties. Standardized interfaces will enable the sharing of research organizations and structures between the different levels of crime forecasting and maximize the use of crime data resources.
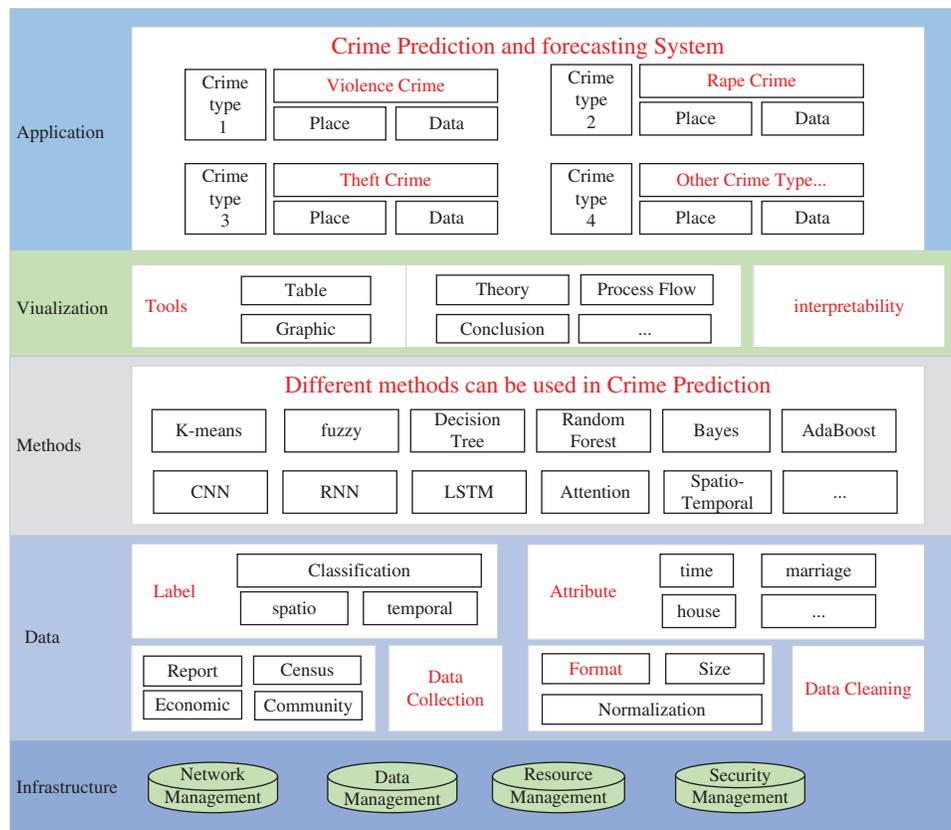

Figure 2: The architecture of the crime prediction system

As these systems have highly sensitive data, the security of the data is also essential. In the use process, criminal data needs to be encrypted or desensitized. In addition to data security, the system's security is also more important; the software's own security technology does not only protect it but also needs to be protected by various network security technologies and database security technologies.

Safeguarding data and system security, thus increasing the usability and effectiveness of this crime prediction system, accelerates the field's rapid development.

4 Conclusion

Predicting criminal activity is an integral component of social computing technology. In this paper, the existing traditional machine learning algorithms and data-driven deep learning methods are thoroughly examined, the related research algorithms are systematically analyzed, and the fundamental algorithm flow is coded. The most frequently used data sets are collected from the United States and India for their open-source crime data, and many algorithms have been tested on these data sets. The initial volume of crime data is relatively small, so effective methods with superior prediction performance are based on conventional machine learning techniques, such as the decision tree, support vector machine, random forest, AdaBoost, and Naive Bayes. Different data sets have different contents and types of crimes, and the performance of traditional machine learning methods depends on different extracted features; therefore, there is no normalized method with the best performance. With the development of big data and computing power, the volume of the crime data set expands, and its content becomes ever more detailed. Numerous structures based on deep learning are implemented in crime prediction. The performance of algorithms based on convolution neural networks and fully-connected convolution networks is superior to conventional machine learning. This paper also describes the commonly used algorithm measurement metrics, illustrates some of the problems in this field, and proposes a system for crime prediction based on massive data analysis.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] V. Srinidhi, P. Saranya and M. Ashok, "An affirmative learning techniques to analyse the crime scene in jewel theft murder," *International Research Journal of Multidisciplinary Technovation*, vol. 2, no. 5, pp. 311–315, 2020.

[2] R. Wazirali and R. Ahmad, "Machine learning approaches to detect DoS and their effect on WSNs lifetime," *Computers, Materials & Continua*, vol. 70, no. 3, pp. 4922–4946, 2022.

[3] U. Selvi and S. Pushpa, "Machine learning privacy aware anonymization using mapreduce based neural network," *Intelligent Automation & Soft Computing*, vol. 31, no. 2, pp. 1185–1196, 2022.

[4] M. Humayun and A. Alsayat, "Prediction model for coronavirus pandemic using deep learning," *Computer Systems Science and Engineering*, vol. 40, no. 3, pp. 947–961, 2022.

[5] X. Shao, "Accurate multi-site daily-ahead multi-step pm2.5 concentrations forecasting using space-shared cnn-lstm," *Computers, Materials & Continua*, vol. 70, no. 3, pp. 5143–5160, 2022.

[6] A. M. Almars, "Attention-based bi-lstm model for arabic depression classification," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 3091–3106, 2022.

[7] D. E. Brown, "The regional crime analysis program (ReCAP): A framework for mining data to catch criminals," in *Proc ICSMC'98*, San Diego, CA, USA, pp. 2848–2853, 1998.

[8]   P. L. Brantingham, P. J. Brantingham, M. Vajihollahi and K. Wuschke, "Crime analysis at multiple scales of aggregation: A topological approach," in *Putting Crime in its Place*, New York, NY, USA: Springer, pp. 87–107, 2009.

[9]   A. L. Buczak and C. M. Gifford, "Fuzzy association rule mining for community crime pattern discovery," in *ACM SIGKDD Workshop on Intelligence and Security Informatics*, New York, NY, USA, pp. 1–10, 2010.

[10]   S. Somayeh, M. Aida, S. Fatimah and A. J. Marzanah, "A study on classification learning algorithms to predict crime status," *International Journal of Digital Content Technology and its Applications*, vol. 7, no. 9, pp. 361–369, 2013.

[11]   J. Agarwal, R. Nagpal and R. Sehgal, "Crime analysis using k-means clustering," *International Journal of Computer Applications*, vol. 83, no. 4, pp. 1–4, 2013.

[12]   P. Gera and R. Vohra, "City crime profiling using cluster analysis," *International Journal of Computer Science and Information Technologies*,vol. 5, no. 4, pp. 5145–5148, 2014.

[13]   S. Sivaranjani, S. Sivakumari and M. Aasha, "Crime prediction and forecasting in Tamilnadu using clustering approaches," in *2016 Int. Conf. on Emerging Technological Trends (ICETT)*, Kollam, India, pp. 1–6, 2016.

[14]   S. Mehmet, K. Hacer and M. Ali Akcayol, "Crime analysis based on association rules using apriori algorithm," *International Journal of Information and Electronics Engineering*, vol. 7, no. 3, pp. 99–102, 2017.

[15]   L. McClendon and N. Meghanathan, "Using machine learning algorithms to analyze crime data," *Machine Learning and Applications: An International Journal (MLAIJ)*, vol. 2, no. 1, pp. 1–12, 2015.

[16]   G. Ma, Z. Wang and S. Ma, "Analysis of the effectiveness of machine learning model in predicting the risk of inmates," *Journal of Hebei University (Natural Science Edition)*, vol. 37, no. 4, pp. 426–433, 2017.

[17]   P. Das and A. K. Das, "Application of classification techniques for prediction and analysis of crime in India," *Computational Intelligence in Data Mining*, vol. 711, pp. 191–201, 2019.

[18]   C. Yamini and C. Premasundari, "A violent crime analysis using fuzzy c-means clustering approach," *ICTACT Journal on Soft Computing*, vol. 9, no. 3, pp. 1939–1944, 2019.

[19]   Z. Zhang, J. Huang, J. Hao, J. Gong and H. Chen, "Extracting relations of crime rates through fuzzy association rules mining," *Applied Intelligence*, vol. 50, pp. 448–467, 2020.

[20]   L. G. A. Alves, H. V. Ribeiro and F. A. Rodrigues, "Crime prediction through urban metrics and statistical learning," *Physica A: Statistical Mechanics and its Applications*, vol. 505, pp. 435–443, 2018.

[21]   R. M. Aziz, A. Hussain, P. Sharma and P. Kumar, "Machine learning-based soft computing regression analysis approach for crime data prediction," *Karbala International Journal of Modern Science*, vol. 8, no. 1, pp. 1–19, 2022.

[22]   R. M. Aziz, P. Sharma and A. Hussain, "Machine learning algorithms for crime prediction under Indian penal code," *Annals of Data Science*, pp. 1–32, 2022. https://doi.org/10.1007/s40745-022-00424-6.

[23]   M. Feng, J. Zheng, Y. Han, J. Ren and Q. Liu, "Big data analytics and mining for crime data analysis, visualization and prediction," in *Int. Conf. on Brain Inspired Cognitive Systems*, Xi'an, China, pp. 605–614, 2018.

[24]   G. Hajela, M. Chawla and A. Rasool, "A clustering based hotspot identification approach for crime prediction," *Procedia Computer Science*, vol. 167, pp. 1462–1470, 2020.

[25]   S. Hossain, A. Abtahee, I. Kashem, M. M. Hoque and I. H. Sarker, "Crime prediction using spatio-temporal data," in *Int. Conf. on Computing Science, Communication and Security*, Gujarat, India, pp. 277–289, 2020.

[26]   S. H. Kim, P. Joshi, P. S. Kalsi and P. Taheri, "Crime analysis through machine learning," in *IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conf. (IEMCON)*, Vancouver, BC, Canada, pp. 414–420, 2018.

[27]   A. Kumar, A. Verma, G. Shinde, Y. Sukhdeve and N. Lal, "Crime prediction using k-nearest neighboring algorithm," in *2020 Int. Conf. on Emerging Trends in Information Technology and Engineering (IC-ETITE)*, Vellore, India, pp. 1–4, 2021.

[28]   P. Tamilarasi and R. U. Rani, "Diagnosis of crime rate against women using k-fold cross validation through machine learning," in *2020 Fourth Int. Conf. on Computing Methodologies and Communication (ICCMC)*, Erode, India, pp. 1034–1038, 2020.

[29]   R. F. R. Forradellas, S. L. N. Alonso, J. Jorge-Vazquez and M. L. Rodriguez, "Applied machine learning in social sciences: Neural networks and crime prediction," *Social Sciences*, vol. 10, no. 1, pp. 1–20, 2020.

[30]   J. Keppens and J. Zeleznikow, "A model based reasoning approach for generating plausible crime scenarios from evidence," in *Proc the 9th Int. Conf. on Artificial Intelligence and Law*, Edinburgh, Scotland, UK, pp. 51–59, 2003.

[31]   S. Thomas and M. Leese, "A green-fingered approach can improve the clinical utility of violence risk assessment tools," *Criminal Behaviour & Mental Health*, vol. 13, no. 3, pp. 153–158, 2003.

[32]   M. Sharma, "Z-CRIME: A data mining tool for the detection of suspicious criminal activities based on decision tree," in *2014 Int. Conf. on Data Mining and Intelligent Computing (ICDMIC)*, Delhi, India, pp. 1–6, 2014.

[33]   S. Rishabh, R. Rishabh, K. Vidhi and C. Prathamesh, "K-means clustering analysis of crimes on Indian women," *Journal of Cyber Security and Information Management (JCIM)*, vol. 4, no. 1, pp. 5–25, 2020.

[34]   S. W. Aldossari, F. Alqahtani, N. S. Alshahrani, M. M. Alhammam, R. M. Alzamanan *et al.,* "A comparative study of decision tree and naive Bayes machine learning model for crime category prediction in Chicago," in *Proc the 6th Int. Conf. on Computing and Data Engineering*, Sanya, China, pp. 34–38, 2020.

[35]   S. G. Krishnendu, P. P. Lakshmi and L. Nitha, "Crime analysis and prediction using optimized k-means algorithm," in *Proc 2020 Fourth Int. Conf. on Computing Methodologies and Communication (ICCMC)*, Erode, India, pp. 915–918, 2020.

[36]   N. Kanimozhi, N. V. Keerthana, G. S. Pavithra, G. Ranjitha and S. Yuvarani, "Crime type and occurrence prediction using machine learning algorithm," in *Proc 2021 Int. Conf. on Artificial Intelligence and Smart Systems (ICAIS)*, Coimbatore, India, pp. 266–273, 2021.

[37]   S. Ramasubbareddy, T. A. S. Srinivas, K. Govinda and S. S. Manivannan, "Crime prediction system," *Innovations in Computer Science and Engineering*, vol. 103, no. 3, pp. 127–134, 2020.

[38]   S. R. Bandekar, and C. Vijayalakshmi, "Design and analysis of machine learning algorithms for the reduction of crime rates in India," *Procedia Computer Science*, vol. 172, no. 3, pp. 122–127, 2020.

[39]   A. Babakura, M. N. Sulaiman and M. A. Yusuf, "Improved method of classification algorithms for crime prediction," in *2014 Int. Symp. on Biometrics and Security Technologies (ISBAST)*, Kuala Lumpur, Malaysia, pp. 250–255, 2014.

[40]   R. Iqbal, M. A. A. Murad, A. Mustapha and R. H. S. Panahy, "An experimental study of classification algorithms for crime prediction," *Indian Journal of Science and Technology*, vol. 6, no. 3, pp. 4219–4225, 2013.

[41]   I. Niyonzima, E. Ahishakiye, D. Taremwa and E. Opiyo, "Crime prediction using decision tree (J48) classification algorithm," *International Journal of Computer and Information Technique*, vol. 6, no. 3, pp. 188–195, 2017.

[42]   R. Liao, X. Wang, L. Li and Z. Qin, "A novel serial crime prediction model based on Bayesian learning theory," in *2010 Int. Conf. on Machine Learning and Cybernetics*, Qingdao, China, vol. 4, pp. 1757–1762, 2010.

[43]   S. Chainey, L. Tompson and S. Uhlig, "The utility of hotspot mapping for predicting spatial patterns of crime," *Security Journal*, vol. 21, no. 1, pp. 4–28, 2008.

[44]   J. Carter, E. R. Louderback, D. Vildosola and S. S. Roy, "Crime in an affluent city: Spatial patterns of property crime in coral gables, florida," *European Journal on Criminal Policy and Research*, vol. 26, no. 4, pp. 547–570, 2020.

[45]   C. Cowen, E. R. Louderback and S. S. Roy, "The role of land use and walkability in predicting crime patterns: A spatiotemporal analysis of miami-dade county neighborhoods, 2007–2015," *Security Journal*, vol. 32, no. 3, pp. 264–286, 2019.

[46] B. Umair, S. Letchmunanet, F. H. Hassan, M. Ali, A. Baqir *et al.,* "Spatio-temporal crime predictions by leveraging artificial intelligence for citizens security in smart cities," *IEEE Access*, vol. 9, pp. 47516–47529, 2021.

[47] F. K. Bappee, A. S. Júnior and S. Matwin, "Predicting crime using spatial features," in *Canadian Conf. on Artificial Intelligence*, Toronto, Canada, pp. 367–373, 2018.

[48] A. Bogomolov, B. Lepri, J. Staiano, N., Oliver, F. Pianesi *et al.,* "Once upon a crime: Towards crime prediction from demographics and mobile data," in *Proc. of the 16th Int. Conf. on Multimodal Interaction*, Istanbul, Turkey, pp. 427–434, 2014.

[49] J. He and H. Zheng, "Prediction of crime rate in urban neighborhoods based on machine learning," *Engineering Applications of Artificial Intelligence*, vol. 106, no. c, pp. 104460, 2021.

[50] C. Kadar, R. Maculan and S. Feuerriegel, "Public decision support for low population density areas: An imbalance-aware hyper-ensemble for spatio-temporal crime prediction," *Decision Support Systems*, vol. 119, no. 26, pp. 107–117, 2019.

[51] T. T. Nguyen, A. Hatua and A. H. Sung, "Building a learning machine classifier with inadequate data for crime prediction," *Journal of Advances in Information Technology*, vol. 8, no. 2, pp. 141–147, 2017.

[52] H. K. R. ToppiReddy, B. Saini and G. Mahajan, "Crime prediction and monitoring framework based on spatial analysis," *Procedia Computer Science*, vol. 132, pp. 696–705, 2018.

[53] G. R. Nitta, B. Y. Rao, T. Sravani, N. Ramakrishiah and M. Balaanand, "LASSO-based feature selection and naïve Bayes classifier for crime prediction and its type," *Service Oriented Computing and Applications*, vol. 13, no. 3, pp. 187–197, 2019.

[54] I. Pradhan, K. Potika, M. Eirinaki and P. Potikas, "Exploratory data analysis and crime prediction for smart cities," in *Proc. of the 23rd Int. Database Applications and Engineering Symp.*, Athens, Greece, pp. 1–9, 2019.

[55] W. Safat, A. Sohail, and S. A. Gillani, "Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques," *IEEE Access*, vol. 9, pp. 70080–70094, 2021.

[56] M. Saraiva, I. Matijošaitienė, S. Mishra and A. Amante, "Crime prediction and monitoring in porto, Portugal, using machine learning, spatial and text analytics," *ISPRS International Journal of Geo-Information*, vol. 11, no. 7, pp. 400, 2021.

[57] A. M. Shermila, A. B. Bellarmine and N. Santiago, "Crime data analysis and prediction of perpetrator identity using machine learning approach," in *2018 2nd Int. Conf. on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, pp. 107–114, 2018.

[58] A. Tamir, E. Watson, B. Willett, Q. Hasan and J. S. Yuan, "Crime prediction and forecasting using machine learning algorithms," *International Journal of Computer Science and Information Technologies*, vol. 12, no. 2, pp. 26–33, 2021.

[59] N. A. S. Zaidi, A. Mustapha, S. A. Mostafa and M. N. Razali, "A classification approach for crime prediction," in *Int. Conf. on Applied Computing to Support Industry: Innovation and Technology*, Ramadi, Iraq, pp. 68–78, 2019.

[60] Y. Lin, M. Yen and L. Yu, "Grid-based crime prediction using geographical features," *ISPRS International Journal of Geo-Information*, vol. 7, no. 8, pp. 298, 2018.

[61] L. Duan, T. Hu, E. Cheng, J. Zhu and C. Gao, "Deep convolutional neural networks for spatio-temporal crime prediction," in *Proc. of the Int. Conf. on Information and Knowledge Engineering (IKE). The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing*, Las Vegas, NV, USA, pp. 61–67, 2017.

[62] Z. Wang, X. Liu, J. Lu, W. Wu, H. Zhang, "Construction and spatial-temporal analysis of crime network: A case study on burglary," *Geomatics and Information Science of Wuhan University*, vol. 43, no. 5, pp. 759–765, 2018.

[63] S. A. Chun, V. A. Paturu, S. Yuan, R. Pathak, V. Atluri *et al.,* "Crime prediction model using deep neural networks," in *Proc. of the 20th Annual Int. Conf. on Digital Government Research*, Dubai, United Arab Emirates, pp. 512–514, 2019.

[64] H. Kang and H. Kang, "Prediction of crime occurrence from multi-modal data using deep learning," *PLoS One*, vol. 12, no. 4, pp. e0176244, 2017.

[65] Y. Lin, T. Chen and L. Yu, "Using machine learning to assist crime prevention," in *2017 6th IIAI Int. Congress on Advanced Applied Informatics (IIAI-AAI)*, Hamamatsu, Japan, pp. 1029–1030, 2017.

[66] T. Matereke, C. N. Nyirenda and M. Ghaziasgar, "A comparative evaluation of spatio temporal deep learning techniques for crime prediction," in *2021 IEEE AFRICON*, Arusha, United Republic of Tanzania, pp. 1–6, 2021.

[67] Y. Wei, W. Liang, Y. Wang and J. Cao, "CrimeSTC: A deep spatial-temporal-categorical network for citywide crime prediction," in *2020 the 3rd Int. Conf. on Computational Intelligence and Intelligent Systems*, Tokyo, Japan, pp. 75–79, 2020.

[68] M. Muthamizharasan and R. Ponnusamy, "Forecasting crime event rate with a CNN-LSTM model," in *Innovative Data Communication Technologies and Application*, Singapore, pp. 461–470, 2022.

[69] G. Deepak, S. Rooban and A. Santhanavijayan, "A knowledge centric hybridized approach for crime classification incorporating deep bi-LSTM neural network," *Multimedia Tools and Applications*, vol. 80, no. 18, pp. 28061–28085, 2021.

[70] B. Wang, P. Yin, A. L. Bertozzi, P. J. Brantingham, S. J. Osher *et al.,* "Deep learning for real-time crime forecasting and its ternarization," *Chinese Annals of Mathematics, Series B*, vol. 40, no. 6, pp. 949–966, 2019.

[71] Z. M. Wawrzyniak, S. Jankowski, E. Szczechla and Z. Szymanski, "Data-driven models in machine learning for crime prediction," in *2018 26th Int. Conf. on Systems Engineering (ICSEng)*, Sydney, NSW, Australia, 2018.

[72] F. K. Bappee, A. Soares, L. M. Petry and S. Matwin, "Examining the impact of cross-domain learning on crime prediction," *Journal of Big Data*, vol. 8, no. 1, pp. 1–27, 2021.

[73] K. Hu, L. Li, J. Liu and D. Sun, "Duronet: A dual-robust enhanced spatial-temporal learning network for urban crime prediction," *ACM Transactions on Internet Technology (TOIT)*, vol. 21, no. 1, pp. 1–24, 2021.

[74] C. Huang, J. Zhang, Y. Zheng and N. V. Chawla, "Deepcrime: Attentive hierarchical recurrent networks for crime prediction," in *Proc. of the 27th ACM Int. Conf. on Information and Knowledge Management*, Turin, Italy, pp. 1423–1432, 2018.

[75] J. Sun, M. Yue, Z. Lin, X. Yang and L. Nocera *et al.,* "Crimeforecaster: Crime prediction by exploiting the geographical neighborhoods' spatiotemporal dependencies," in *Joint European Conf. on Machine Learning and Knowledge Discovery in Databases*, Ghent, Belgium, pp. 52–67, 2020.

[76] Y. Wang, L. Ge, S. Li and F. Chang, "Deep temporal multi-graph convolutional network for crime prediction," in *Int. Conf. on Conceptual Modeling*, Vienna, Austria, pp. 525–538, 2020.

[77] G. Ghankutkar, N. Sarkar, P. Gajbhiye, S. Yadav, D. Kalbande *et al.,* "Modelling machine learning for analysing crime news," in *2019 Int. Conf. on Advances in Computing, Communication and Control (ICAC3)*, Mumbai, India, pp. 1–5, 2019.

[78] S. Li, H. Zhang, L. Ye, S. Su, X. Guo *et al.,* "Prison term prediction on criminal case description with deep learning," *Computers, Materials and Continua*, vol. 62, no. 3, pp. 1217–1231, 2020.

[79] A. Umair, M. S. Sarfraz, M. Ahmad, U. Habib, M. H. Ullah *et al.,* "Spatiotemporal analysis of web news archives for crime prediction," *Applied Sciences*, vol. 10, no. 22, pp. 210–235, 2020.

[80] M. Boukabous and M. Azizi, "Crime prediction using a hybrid sentiment analysis approach based on the bidirectional encoder representations from transformers," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 25, no. 2, pp. 1131–1139, 2022.

[81] S. Chackravarthy, S. Schmitt and L. Yang, "Intelligent crime anomaly detection in smart cities using deep learning," in *2018 IEEE 4th Int. Conf. on Collaboration and Internet Computing (CIC)*, Philadelphia, PA, USA, pp. 399–404, 2018.

[82] Y. Fan, "Criminal psychology trend prediction based on deep learning algorithm and three-dimensional convolutional neural network," *Journal of Psychology in Africa*, vol. 31, no. 3, pp. 292–297, 2021.

[83]   K. K. Kumar and H. V. Reddy, "Crime activities prediction system in video surveillance by an optimized deep learning framework," *Concurrency and Computation: Practice and Experience*, vol. 34, no.11, pp. cpe.6852, 2022.

[84]   U. V. Navalgund and K. Priyadharshini, "Crime intention detection system using deep learning," in *2018 Int. Conf. on Circuits and Systems in Digital Enterprise Technology (ICCSDET)*, Kottayam, India, pp. 1–6, 2018.

[85]   N. Shah, N. Bhagat and M. Shah, "Crime forecasting: A machine learning and computer vision approach to crime prediction and prevention," *Visual Computing for Industry, Biomedicine, and Art*, vol. 4, no.1, pp. 1–14, 2021.

[86]   C. Rajapakshe, S. Balasooriya, H. Dayarathna, N. Ranaweera, N. Walgampaya *et al.,* "Using cnns rnns and machine learning algorithms for real-time crime prediction," in *2019 Int. Conf. on Advancements in Computing (ICAC)*, Malabe, Sri Lanka, pp. 310–316, 2019.

[87]   Z. Abbass, Z. Ali, M. Ali, B. Akbar and A. Saleem, "A framework to predict social crime through twitter tweets by using machine learning," in *2020 IEEE 14th Int. Conf. on Semantic Computing (ICSC)*, San Diego, CA, USA, pp. 363–368, 2020.

[88]   A. Ippolito and A. C. G. Lozano, "Tax crime prediction with machine learning: A case study in the municipality of São paulo," *ICEIS*, vol. 1, pp. 452–459, 2020.

[89]   A. Deshmukh, S. Banka, S. B. Dcruz, S. Shaikh and A. K. Tripathy, "Safety app: Crime prediction using GIS," in *2020 3rd Int. Conf. on Communication System, Computing and IT Applications (CSCITA)*, Mumbai, India, pp. 120–124, 2020.

[90]   C. Kadar and I. Pletikosa, "Mining large-scale human mobility data for long-term crime prediction," *EPJ Data Science*, vol. 7, no. 1, pp. 1–27, 2018.

[91]   X. Zhang, L. Liu, M. Lan and G. Song, "Interpret able machine learning models for crime prediction," *Computers, Environment and Urban Systems*, vol. 94, no. 2, pp. 101789, 2022.

[92]   Y. Rayhan and T. Hashem, "AIST: An interpretable attention-based deep learning model for crime prediction," arXiv preprint arXiv:2012.08713, 2020.

[93]   A. P. Wheeler and W. Steenbeek, "Mapping the risk terrain for crime using machine learning," *Journal of Quantitative Criminology*, vol. 37, no. 2, pp. 445–480, 2021.

[94]   W. Gorr and A. Olligschlaeger and Y. Thompson, "Short-term time series forecasting of crime," *The International Journal of Forecasting*, vol. 19, no. 4, pp. 579–594, 2003.

[95]   H. Liu and D. E. Brown, "Criminal incident prediction using a point-pattern-based density model," *International Journal of Forecasting*, vol. 19, no. 4, pp. 603–622, 2003.

[96]   P. Stalidis, T. Semertzidis and P. Daras, "Examining deep learning architectures for crime classification and prediction," *Forecasting*, vol. 3, no. 4, pp. 741–762, 2021.

[97]   S. Wang and K. Yuan, "Spatiotemporal analysis and prediction of crime events in atlanta using deep learning," in *2019 IEEE 4th Int. Conf. on Image, Vision and Computing (ICIVC)*, Xiamen, China, pp. 346–350, 2019.

[98]   S. Wu, C. Wang, H. Cao and X. Jia, "Crime prediction using data mining and machine learning," in *Int. Conf. on Computer Engineering and Networks*, Shanghai, China, pp. 360–375, 2018.

[99]   L. Xia, C. Huang, Y. Xu, P. Dai, L. Bo *et al.,* "Spatial-temporal sequential hypergraph network for crime prediction with dynamic multiplex relation learning," in *IJCAI*, Montreal, Canada, pp. 1631–1637, 2021.

[100]   F. Yi, Z. Yu, F. Zhuang and B. Guo, "Neural network based continuous conditional random field for fine-grained crime prediction," in *IJCAI*, Macao, China, pp. 4157–4163, 2019.