

# **Project on CrimeVision: Advanced Crime Classification with Deep Learning**

## **ABSTRACT**

Predicting crime using machine learning and deep learning techniques has gained considerable attention from researchers in recent years, focusing on identifying patterns and trends in crime occurrences. This review paper examines over 150 articles to explore the various machine learning and deep learning algorithms applied to predict crime. The study provides access to the datasets used for crime prediction by researchers and analyzes prominent approaches applied in machine learning and deep learning algorithms to predict crime, offering insights into different trends and factors related to criminal activities. Additionally, the paper highlights potential gaps and future directions that can enhance the accuracy of crime prediction. Finally, the comprehensive overview of research discussed in this paper on crime prediction using machine learning and deep learning approaches serves as a valuable reference for researchers in this field. By gaining a deeper understanding of crime prediction techniques, law enforcement agencies can develop strategies to prevent and respond to criminal activities more effectively.

## **INTRODUCTION:**

Crime prediction is a complex problem requiring advanced analytical tools to effectively address the gaps in existing detection mechanisms. With the increasing availability of crime data and through the advancement of existing technology, researchers were provided with a unique opportunity to study and research crime detection using machine learning and deep learning methodologies. Based on the recent advances in this field, this article will explore current trends in machine learning and deep learning for crime prediction and discuss how these cutting-edge technologies are being used to detect criminal activities, predict crime patterns, and prevent crime. Our primary goal is to provide a comprehensive overview of recent advancements in this field and contribute to future research efforts.

Machine learning is a subset of artificial intelligence that uses statistical models and algorithms to analyze and make predictions based on data. On the other hand, deep learning is a subset of machine learning that uses artificial neural networks with multiple layers to model complex relationships between inputs and outputs. Both machine learning and deep learning technologies have the potential to be applied to the problem of crime prediction in many ways.

Machine learning algorithms have been utilized in crime prediction to analyze crime data and predict future crime patterns. For example, algorithms like decision trees, random forests, and support vector machines have been trained

on crime data from specific cities to predict crime patterns accurately [5]. Apart from predicting crime patterns, these algorithms can provide valuable insights into crime trends and patterns. These capabilities allow for deploying resources and tactics to combat crime effectively. Additionally, machine learning algorithms can also be used to identify correlations between crime incidents and various environmental and demographic factors such as location, weather, and time of day [6]. This information can be used to develop crime prediction and prevention strategies suitable to a given community's specific needs.

Predictive policing is also a significant application of machine learning for crime prediction [7]. Predictive policing refers to using data and analytics to inform law enforcement efforts and reduce crime. Machine learning algorithms can be used to analyze crime data from a specific geographic area, such as a city or neighborhood, to identify crime hotspots and predict future crime incidents. This information can then be used to direct policing resources to areas where they are most needed, increasing the effectiveness of law enforcement efforts.

Deep learning algorithms, such as convolution and recurrent neural networks, have also shown promise in crime prediction. These algorithms have been trained on crime data with either a spatial or temporal component to accurately predict crime patterns in specific cities. For example, deep learning algorithms have been used to analyze crime data, including the time, location, and type of crime incidents. This information is used to create a predictive model

that can be used to identify potential crime hotspots and predict future crime incidents. Another application of deep learning in crime prediction is computer vision and video analysis. This technology has been used to analyze video footage from surveillance cameras to detect and classify criminal activities, such as vandalism, theft, and assault. The advanced deep learning models are also integrated with drones and other aerial technologies to provide new opportunities to monitor and respond to criminal activities. These algorithms have also been used to analyze crime data from multiple sources, including crime reports, social media, and police records, providing a more comprehensive view of criminal activities. By automating this process, deep learning algorithms have the potential to enhance the ability to identify and respond to crime in real-time, providing a crucial tool in the fight against criminal activity.

Despite the promise of machine learning and deep learning for crime prediction, several challenges must be addressed. One of the biggest challenges is the availability of high-quality crime data. Crime data can be difficult to obtain, and the available data may need to be completed or reliable. Additionally, collecting and using crime data is associated with privacy and ethical concerns. These challenges must be addressed to fully realize the potential of machine learning and deep learning for crime prediction. Another challenge is the interpretability of machine learning and deep learning models. These models can be challenging to understand and interpret, limiting their usefulness in decision-making. To effectively apply these models to the problem of crime prediction, it is vital to develop interpretable models that can provide clear explanations of their predictions.

Moreover, the recent advancements in machine learning and deep learning for crime prediction show great promise in addressing this complex problem. However, significant challenges remain, and much work is still needed to realize these technologies' potential fully. This research article provides a comprehensive overview of recent trends in this field and offers insights into the potential applications of machine learning and deep learning for crime predic-

tion. By highlighting the potential of these technologies and the challenges that must be addressed, this research article contributes to the broader research community. It advances our understanding of the role of machine learning and deep learning in crime prediction. Hence, the key contributions of this work are as follows:- first, this paper provides the amalgamation of existing studies that utilized state-of-the-art machine learning and deep learning-based approaches in the realm of detecting neighborhood crime. Thereby extending the fathomable literature knowledge base. Second, this paper eliminates the limitation of the scarcity of potential datasets availability. We have highlighted distinct publicly available datasets related to neighborhood crime prediction that existing studies have utilized. Thereby archiving the data resources for future scholars. Third, this work drafted future research directions to eliminate the existing research gaps in neighborhood crimes. Thereby reasonably providing future research objectives/questions to the research community to pursue further.

## Research Methodology

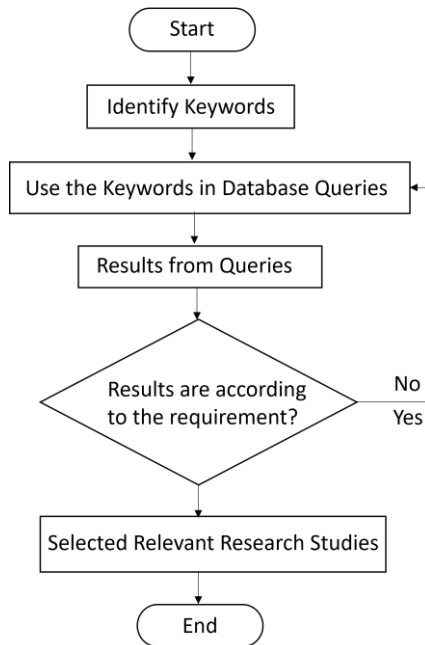
The primary research aims to find various efficient algorithms for predicting neighborhood crimes. In our previous work [6], we used statistical analysis to predict the crimes in Newyork city. Our paper got good attention from the researchers, so we wanted to look for the efficient machine learning and deep learning approaches used in this area. We have followed a systematic approach to select the papers for this review. As part of this research, we have considered the papers from multiple databases related to predicting crime.

For this review, we have considered all the primarily used terms in the papers focused on predicting crimes. To include all the possible alternative words of each term, we have used “\*” as a wild character for IEEE and ACM databases so that it contains zero or more characters after the string. The main target of this review is to check for all the existing research works to predict crime. In addition, we want to help the research community by identifying the different datasets used to apply the algorithms. Irrelevant studies are removed by applying multiple filters to our search queries. We also selected 30 papers to be part of the main text based on relevance and novelty, and 20 more papers are added in the appendix section. In this survey, we have used a combination of an automated and manual search shown in. In the initial stages, we focused on using the automatic digital search. In the final step, we manually read the entire paper to select a set of documents in machine learning and deep learning areas.

Firstly, we have identified the key terms to create the queries. We then used those keywords to construct the various research database-related queries based on respective syntax. Below are the queries used to explore IEEE, Science Direct, and ACM databases.

## Data Collection

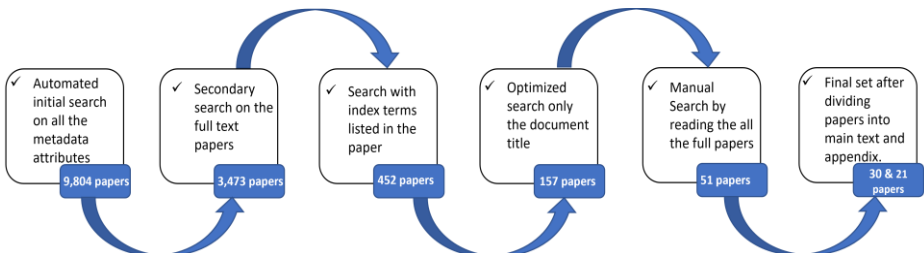
We focused on looking into individual research libraries rather than searching in google scholar. Because google scholar will have data from all these



**Fig. 1** Steps involved for typical crime detection

databases, there could be duplicates. Below are the database library homepage links where the research works were extracted using the keywords mentioned in search queries. Initially, we searched using all the metadata attributes available on each database. Next, we applied the filters only on the full-text papers. As we noticed that the number of documents is still high, we applied the filter on the index terms used in the article as the results will be more relevant. We have more than 450 papers from all the databases at this stage. Finally, in the last step, we applied the filter on the document title, where the total number of papers was 157.

- Science-Direct Elsevier (<https://www.sciencedirect.com>)
- ACM (DL) Digital Library (<https://dl.acm.org>)
- IEEE Xplore Access Digital Library (<https://ieeexplore.ieee.org>)



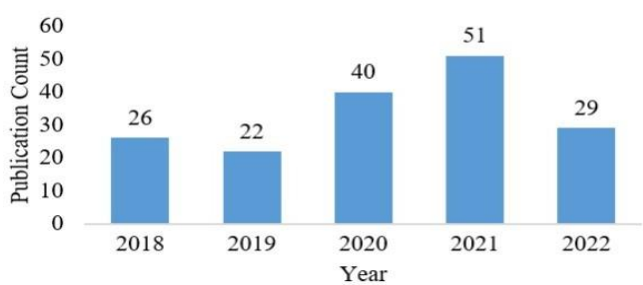
**Fig. 2** Research paper selection methodology

After applying the automated filters as shown in, all the authors manually divided the work and read the papers to select the final set. For the selection, we mainly looked into the essential elements like the focus or objective of the article, datasets that authors have used, algorithms applied, and the accuracy rates. The focus of this survey is not only to help the community know the various algorithms applied but also to let them know about the datasets they can use to apply the novel algorithms and get the results for their research.

## 2 Literature Analysis

### 2.1 Pre-assessment Literature Analysis

An analysis of collected literature data from the distinct research databases is essential [11] to receive information regarding the growth of an adopted research domain, scope across the research community, and popularity among the existing researchers. Thus, we have performed a detailed analysis of the collected literature data. We did pre-analysis and post-analysis, wherein pre-analysis comprises the exploration of initially collected literature ( i.e., the research papers that were collected immediately after performing our search query), and post-analysis comprises the investigation of those study's data that were finally selected after applying selection criteria. Although we have exhumed the two renowned research databases, IEEE and Science Direct, lately, it has been observed that Science direct discontinued the search result extraction. Hence following analytical charts are based on IEEE databases findings shows the research publication trends in neighborhood crime



**Fig. 3** Research publication trends from 2018-2022

from 2018 to 2022 (i.e., the last five years). It showed upward trends from 2018 to 2021 and downward trends during 2022. The years 2020 and 2021 are the apexes of COVID-19, which could be why existing researchers have utilized that time to explore more neighborhood crime research.

Figure 4 indicates the distribution of article page counts for the neighborhood crime research. The majority of articles (i.e., 46) comprise five pages. This may be because many researchers have published their work in conferences and symposiums rather than journals. Figure 4 depicts that very few articles comprise more than ten pages. Since conferences, seminars, workshops, and symposiums allowed the presentation of abstracts and short papers, this Figure 4 also depicts the 1,2,3 page-long articles.

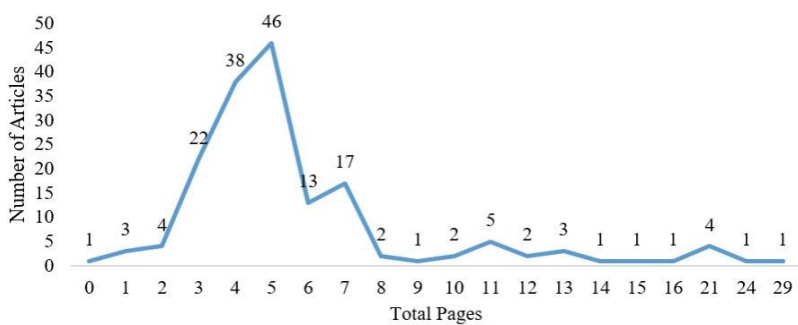


Fig. 4 Distribution of article’s page counts

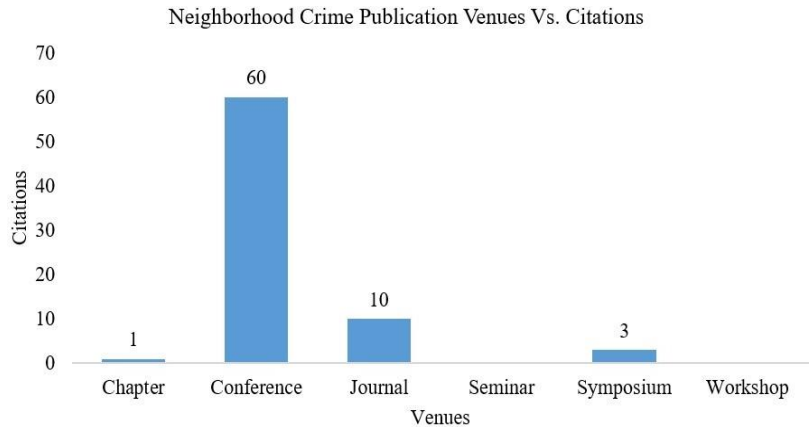
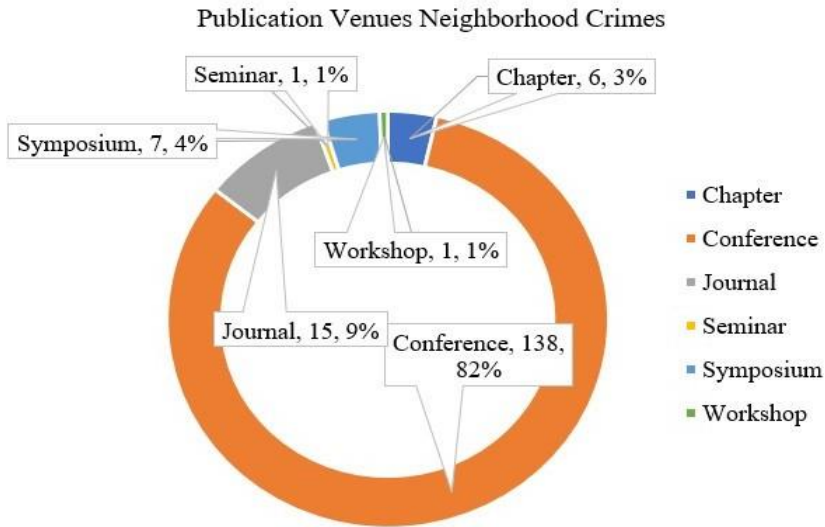


Fig. 5 Distribution of article’s citations

To assess the popularity of the neighborhood crime research articles, we have also performed the citation analysis shown in Figure 5. Wherein the conference papers have gained more citations (60) than the other journals (10 citations), chapter (1 citation), seminar (0 citations), symposium (3 citations), and workshop (o citation) articles in the neighborhood crime area, this may be because this area is less popular among the researchers. Figure 6, shows the distribution of neighborhood crime-related articles published in various



**Fig. 6** Distribution of research articles at various venues

venues. Wherein this figure depicts the venue name (e.g., conference, journal, etcetera), number of papers (i.e., in numeric value-1,7..), and percentage of published articles (i.e., 1%, 82%,...). Here, the majority of papers have been published in the conferences than the other venues like journals, workshops, symposiums, and seminars. In the neighborhood crime area, 82% of the articles were published in conferences, followed by 9% in the journal, 4% in the symposium, 3% in book chapters, 1% in seminars, and 1% in workshops.

## 2.2 Post-assessment Literature Analysis

To enhance the understanding of the neighborhood crime domain in combination with our above-cited selection criteria, we have also performed the post-assessment literature analysis because this is the crux of our literature survey to fulfill our identified objectives. We have created a word cloud for Neighborhood crime-related papers to fathom further the selected papers' underlying key concepts or themes. A word cloud often called a tag cloud, is a graphic depiction of the terms that appear the most frequently in a given text. Each word's magnitude in the word cloud reflects how frequently it appears in the text. Word clouds are frequently employed in literature reviews to swiftly pinpoint the key themes or topics within a sizable body of material. Additionally, they can be used to compare various texts and find trends and patterns in the data [12]

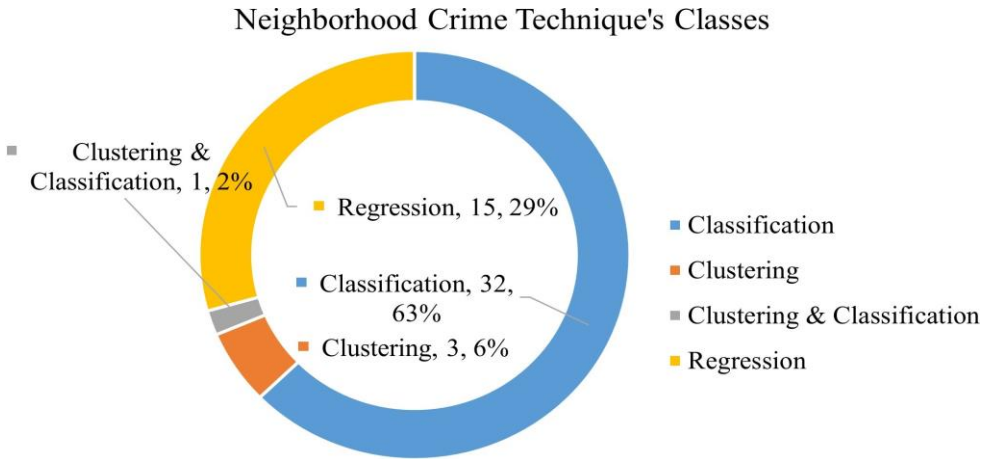
Figure 7 shows the word cloud for the neighborhood crime-related selected articles wherein many crime-related key terms have emerged as trends in the existing studies. "crime", "criminal", "policing", "enforcement", and "security" are the words that indicate the emphasis of researchers on the sub-areas





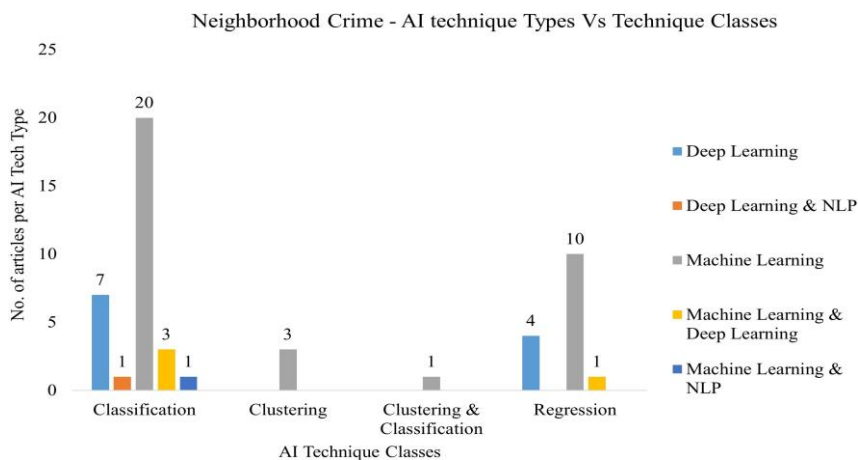
(DL), and the combination of machine learning and natural language processing (NLP), the combination of DL and NLP, and DL are the majorly used AI technique type. In the neighborhood crime domain – 67% of articles have used ML, 21% have used DL, 8% have used the combination of ML and DL, 2% have used the combination of DL and NLP, and 2% have used the combination of ML and NLP. It has been observed that machine learning techniques are popular in the neighborhood crime area.

shows the distribution of neighborhood crime-related selected article's technique classes among Science Direct and IEEE databases. As shown in Figure ??, among the neighborhood crime-related articles, the classification task is the prime focus utilizing various ML and DL-related techniques. In the neighborhood crime domain – 63% of articles have focused on the classification task, 29% have focused on the regression tasks, 6% have focused on the clustering task, and 2% of the studies have utilized the combination of classification and clustering. It has been observed that the Classification task is the prime focus of the studies on neighborhood crime.



**Fig. 9** distribution of neighborhood crime-related selected article's technique's classes

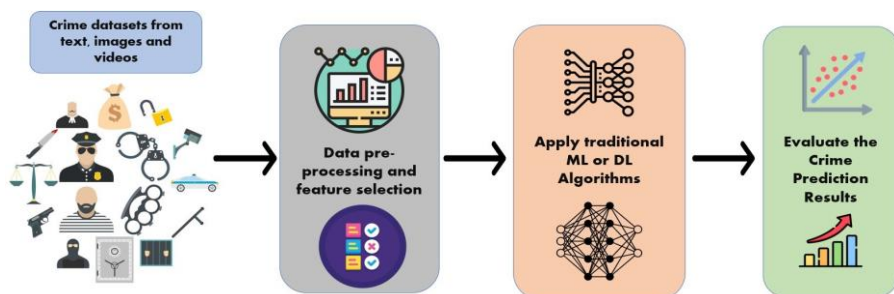
shows the Distribution of neighborhood Crime related selected article's Technique classes Versus technique Type among Science Direct and IEEE databases. answers the question- what AI Technique is used for which technique classes (classification, clustering, and regression)? Herein all five techniques (ML, DL, DL+NLP, ML+NLP, and ML+DL) have been used for the classification task whereas, for clustering and the combination of clustering and classification, the ML is solely used. In neighborhood crime articles NLP is also used for classification and regression tasks in addition to ML and DL. For the classification and regression tasks ML and DL both have been used.



**Fig. 10** Distribution of neighborhood Crime related selected article's Technique classes Versus technique Type

### 3 Crime Prediction Process & Datasets

Crime prediction using machine and deep learning involves several major steps as shown in Figure 11. The first step is data collection, which involves gathering relevant data such as crime statistics, demographics, and weather patterns. The next step is data preprocessing, which includes cleaning and transforming the data into a usable format. After data preprocessing, the data is split into training and testing sets for model development and evaluation. The next step is feature engineering, which involves selecting relevant features from the data that can be used to train the model. Once the features are selected, various machine and deep learning algorithms can be applied to the data for training and prediction purposes. Finally, the trained models are evaluated using various performance metrics to assess their accuracy and effectiveness in predicting crime. The results can be used to support decision-making in law enforcement and crime prevention efforts.



**Fig. 11** Architecture flow of crime prediction

**Table 1:** Crime detection papers with data science algorithms

Dataset Links	Type of Data	Country - Locality
<a href="https://data.gov.in/resource/incidence-crime-committed-against-women-india-during-2001-2012">https://data.gov.in/resource/incidence-crime-committed-against-women-india-during-2001-2012</a>	Crime	India
<a href="https://indore.mppolice.gov.in/">https://indore.mppolice.gov.in/</a>	Crime	India - Madhya Pradesh
<a href="https://ncrb.gov.in/en/node/3721">https://ncrb.gov.in/en/node/3721</a>	Crime	India
<a href="https://data.gov.in/catalogs">https://data.gov.in/catalogs</a> [	Crime	India
<a href="https://www.kaggle.com/datasets/murderaccountability/homicide-reports">https://www.kaggle.com/datasets/murderaccountability/homicide-reports</a>	Crime	Unites States
<a href="https://geodash.vpd.ca/opendata/">https://geodash.vpd.ca/opendata/</a>	Crime	Canada - Vancouver
<a href="https://plenar.io/explore/">https://plenar.io/explore/</a>	Crime - Spatio Temporal	Unites States - Chicago
<a href="https://data.cityofchicago.org/">https://data.cityofchicago.org/</a>	Crime	Unites States - Chicago
<a href="https://datasf.org/opendata/">https://datasf.org/opendata/</a>	Crime	United States - San Francisco
<a href="https://data.london.gov.uk/dataset/">https://data.london.gov.uk/dataset/</a>	Crime	United Kingdom - London
<a href="https://br-city.survey.okfn.org/place/rn.html">https://br-city.survey.okfn.org/place/rn.html</a>	Crime	Brazil - Natal
<a href="https://opendata.cityofnewyork.us/data/">https://opendata.cityofnewyork.us/data/</a>	Crime	United States - Newyork
<a href="https://www.twitter.com">www.twitter.com</a>	Social Media	Global - Twitter
<a href="https://www.kaggle.com/competitions/sf-crime/data">https://www.kaggle.com/competitions/sf-crime/data</a> [	Crime	United States - San Francisco
<a href="https://data.lacity.org/">https://data.lacity.org/</a>	Crime	United States - Los Angeles
<a href="https://dasci.es/transferencia/open-data/24705/">https://dasci.es/transferencia/open-data/24705/</a>	Crime- Weapons	Global
<a href="https://www.gdeltproject.org">https://www.gdeltproject.org</a>	Crime	Global
<a href="https://www.foursquare.com">www.foursquare.com</a>	Geospatial	Unites States
<a href="https://www.nyc.gov/taxi">https://www.nyc.gov/taxi</a> [Taxi	United States -	Newyork

<a href="https://www.wunderground.com/weather/api">https://www.wunderground.com/weather/api</a> <a href="https://www.icpsr.umich.edu/web/NACJD/studies/3355/datadocumentation">https://www.icpsr.umich.edu/web/NACJD/studies/3355/datadocumentation</a>	Weather	Global
<a href="https://roc-ng.github.io/XD-Violence/">https://roc-ng.github.io/XD-Violence/</a>	Crime - Video	United States Global
<a href="https://www.facebook.com">www.facebook.com</a>	Social Media	Global - Facebook
<a href="https://datasus.saude.gov.br/">https://datasus.saude.gov.br/</a>	Health	Brazil
<a href="https://www.phillypolice.com/crime-maps-stats/">https://www.phillypolice.com/crime-maps-stats/</a>	Crime - Spatio Temporal	Unites States - Philadelphia
<a href="https://mapstyle.withgoogle.com/">https://mapstyle.withgoogle.com/</a>	Geospatial	Global
<a href="https://www.mapbox.com/">https://www.mapbox.com/</a>	Geospatial	Global
<a href="https://data.police.uk/data/">https://data.police.uk/data/</a>	Crime	United Kingdom
<a href="https://polisen.se/en/services-and-permits/police-record-extracts/">https://polisen.se/en/services-and-permits/police-record-extracts/</a>	Crime	Sweden
<a href="https://www.openu.ac.il/home/hassner/data/violentflows/">https://www.openu.ac.il/home/hassner/data/violentflows/</a>	Crime - Video	Global

As shown in Table 1, there have been many datasets used in crime detection and prediction research articles. One example is the Chicago Crime Dataset, which contains data on crimes reported in the Chicago area. This dataset has been used to create models that predict the likelihood of specific types of crimes occurring in different areas of the city. Another dataset used in crime prediction research is the London Crime Dataset, which contains data on crimes reported in London city. This dataset has been used to create models that predict the likelihood of crimes occurring in specific areas and their relationship to the socio-economic factors of people based on their geo-locations in the area.

Other datasets commonly used in crime detection and prediction research include the Los Angeles Crime Dataset, the New York City (NYC) Crime Dataset, and the Philadelphia Crime Dataset. These datasets contain information on crimes reported in their respective cities and have been used to create models that predict the likelihood of specific types of crimes occurring in different areas. In addition to these, there are also global datasets that focus on CCTV video footage, types of aggression, and weapons for real-time crime predictions.

Overall, these datasets provide valuable information for researchers to build crime prediction models that could help law enforcement agencies prevent and respond to criminal activities more effectively. The location and access to datasets used by research articles surveyed in this paper are listed in Table 1.

## **4 Crime Prediction using Machine Learning techniques**

Traditional machine learning models have proven to be effective for crime prediction. Various types of models such as decision trees, support vector machines, logistic regression, and random forests have been utilized to analyze crime data and identify patterns that can be used to predict criminal activity. Unlike deep learning, which relies on large amounts of data and complex neural networks, traditional machine learning models require fewer data points and are easier to interpret. For example, a logistic regression model can be used to predict the likelihood of a certain type of crime occurring based on factors such as time of day, location, and demographics of the area. A decision tree model can be used to identify the most important factors that contribute to the occurrence of a particular crime. Random Forest (RF) models can be used to analyze a wide range of features and make predictions about crime patterns. In addition to these techniques, traditional machine learning models can also be used for anomaly detection and outlier analysis in crime data. By identifying unusual patterns or outliers in the data, law enforcement agencies can detect potential criminal activity and take action to prevent it. In the below sections 5.0.1 and 5.0.2, we discuss the latest research on using machine learning model-based regression and classification for crime prediction.

### **5.0.1 Machine Learning based Regression Methods for Crime Prediction**

Several crime detection scenarios are predicted using regression techniques as shown in 2. Researchers mainly focused on prevalent crimes like motorcycle robbery, losing property, and crimes in urban areas. Numerous factors may drive the boom in motorcycle robberies. For example, population growth and density, commuting conduct, bike usage, etc. These situations are problems for the police to govern and screen regularly since it requires forecasting and probabilities of robbery in a precise term. A novel method is proposed in the research study the authors created an application to predict motorcycle robbery with a technique to consider outside consequences using ARIMAX – TFM with a single input. The accuracy of ARIMAX is measured using Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE), and the scores are 32.30 and 6.68.

Rapid urbanization is a compelling challenge connected to city management and services. Cities with higher crime rates are difficult to manage public safety. To reduce crimes, new technologies are relieving police departments to access vast amounts of crime data to identify underlying trends and patterns. These technologies have doubtlessly grown the efficient deployment of police assets within a given region and ultimately guide greater powerful crime prevention. Researchers have worked on predictive models to use these datasets and predict crimes. Study provides a technique primarily based on spatial analysis and auto-regressive fashions to automatically locate excessive-hazard crime areas

in city areas and reliably forecast crime tendencies in each area. Experiments are performed on real-world datasets gathered in New York City and Chicago.

Another study compared multiple techniques to predict the crimes in various areas of a metropolis. This research explored three predictive models: linear regression, logistic regression, and gradient boosting. The authors utilized feature selection techniques to select essential predictors. By applying feature selection methods, there is an improvement in accuracy scored, and it helped to avoid model overfitting. After comparing the results of all four models, the authors found that the gradient boosting technique outperformed, proven to be the best method to predict the crime rate in the urban area.

In another study authors have looked at the crimes in Brazil, which have increased rapidly in recent times. Numerous predictive solutions use intelligent systems to identify when will a criminal offense will arise, which lets police send to those areas that are in danger. As part of their research, the authors looked into four machine-learning approaches for identifying where a criminal offense will arise in Fortaleza, Brazil. Their results indicate that easy algorithms are efficient in predicting crime. Also, they have seen that the Decision Tree and Bagging Regressor strategies obtained quality prediction outcomes.

As mentioned above, numerous linear models are there to predict crime through the correlations between urban metrics and crime. However, due to multicollinearity and nonGaussian distributions in urban attributes, we usually tend to make controversial conclusions on these attributes to predict crime. Ensemble-based machine learning algorithms can deal with such problems adequately. In the research work authors applied random forest regressor to predict the crime and quantify the impact of urban attributes on homicides. Their approach has 97% accuracy in crime prediction, and the significance of city indicators is clustered and ranked equally. Their research identifies the rank of urban indicators based on their significance in predicting crime. As per their results, unemployment and illiteracy are the essential variables for depicting homicides in Brazilian towns.

**Table 2:** Crime detection using machine learning regression techniques

Methodology	ML Algorithms	Dataset	Performance
ARIMAX method is used to predict the time series data from theft cases that affected the number of motorcycles.	ARIMAX algorithm	City of Yogyakarta	RMSE - 6.68

Crime data is used to detect and map crime-dense region (CDR) in NYC and Chicago, then ARIMA is applied on these crime hotspots.	ARIMA, Random Forest (RF), RepTree and ZeroR	Chicago Crimes - 2001 to present, New York City Crimes; 2006 to 2016	2016 RMSE for Area - 143.73, CDR 1 - 57.8, CDR2 - 29.85 & CDR 3 - 16.19
Clustering is used to understand distribution across city and then cluster are used to predict the number of crimes per location.	Linear regression (LR), LoR and Gradient boosting.	Saint-Petersburg Russia Crime (2014 - 2017)	R-Square - 0.9
Four machine learning methods are applied to predict the location of where a crime, more specifically, where a crime of robbery will occur.	RF, Bagging, DT, Extra tree	City of Fortaleza data	RMSE - 0.00231
Random forest regressor is used to predict crime and quantify the influence of urban. Results showed that unemployment and literacy are the two major indicators of crimes.	Random Forest Regressor	Department of Informatics of the Brazilian Public Health System — DATASUS.	Upto 97% Accuracy, Adjusted R square 80% on average.

### 5.0.2 Machine Learning based Classification Methods for Crime Prediction

Traditional regression techniques can successfully check the variables' significance but, they must be more reliable for crime prediction. In many research works mentioned in section, authors have proven that machine learning models effectively predict crimes. Still, they could be more efficient in identifying which variables are significant in predicting crimes. We further examined the classification techniques to predict different criminal incidents like analyzing the criminal reports as shown in Table. Studying those reviews for crime prediction enables regulatory authorities to deal with crime prevention strategies. However, collecting these reviews personally and determining their crime types is challenging. In one study authors have created a novel approach, an incremental classifier that learns the new data and dynamically predicts the results. In this research they have utilized



the Bi-objective Particle Swarm Optimization technique to develop an efficient incremental classifier for dynamically classifying and predicting crime reports. Crime reports from various countries have been collected from online newspapers to measure the performance of their classifier. Also, they evaluated the results manually with unprejudiced police witness narrative crime reports. They tested their approach on four datasets to measure their model’s statistical significance.

**Table 3:** Crime detection using machine learning classification techniques

Methodology	ML Algorithms	Dataset	Performance
OVR-XGBoost: One Vs Rest OVO-XGBoost: One Vs One. The primary difference is based on how the dataset is organized for classification.	XGBoost based algorithms	H City China: (2019)	Aggregate Accuracy - 85%
Applying particle swarm optimization-based classifier and a rules engine to classify crime reports.	Multiple classification algorithms	USA, India & UAE crime articles (2007 - 2017)	Aggregate Accuracy - 79%
17 spatiotemporal variables were used to train and test the XGBoost algorithm, and then SHAP is used to explain the model predictions.	XGboost model	ZG City, in Southeast China. (2017 to 2020 ).	Accuracy - 89% & AUC - 0.586
Random under and over sampling techniques are used to evaluate multiple algorithms.	Logistic Regression (LoR), RF, XGBoost and LightGBM	Danish national psychiatric patient register, (474 crime).	F1 Score - 76%
Graph theory based. Decision trees and heuristics are used for feature selection and an ensemble classifier is applied.	Multiple classification algorithms	USA, UAE & India Crime Articles (2008 - 2016)	Aggregate F1 Score - 88%

Identify if the crime is spread uniformly according to population density or whether specific socio-economic attributes account for increased or decreased crime.	Visual Analysis	London crime dataset	Aggregate Accuracy - 97%
Follows an algorithm-as-a-service architecture (AaaS), providing insights into existing public safety systems and platforms.	K Nearest Neighbor (KNN)	Crime endogenous data sources	Aggregate Accuracy - 90%
Prediction of crime in neighborhoods of New York city using spatial data analysis.	XGBOOST,	New York crime data	Aggregate Accuracy - 52%
EADT approach is used for Interpretable and Accurate Crime Prediction. Traditional data science steps are used to cluster crimes and then predict using classification modeling methods	RF and SVM Decision Tree (DT) K-Means	Pennsylvania state prisons National Crime Records Bureau (India)	Aggregate Accuracy - 77.6% Aggregate Accuracy - 78%

Another research focuses on predicting crime using the XGBoost algorithm. Based on the records of theft instances in H city, they developed an optimized decomposition and fusion method based on XGBoost and applied multi-class classification models like OVR-XGBoost and OVO-XGBoost. As the theft datasets have different classes, they have utilized the SMOTENN algorithm to process and make the dataset balanced. Their results show that OVR-XGBoost and OVO-XGBoost models' prediction accuracy is better than the baseline XGBoost models. In the study [the authors have selected 17 variables for crime prediction, and the XGBoost algorithm is adopted to train the prediction model. A post hoc interpretable approach, Shapley additive explanation (SHAP), is used to present the contribution of person variables. SHAP, a post hoc interpretable method, is used to determine the significance of individual variables. Among all 17 variables used in this research, the percentage of the non-neighborhood population and the populace aged 25–44 contribute greater than different variables in predicting crime. The higher the ambient population of elderly 25–44 in the vicinity, the more public crimes. The authors have also validated the SHAP values to demonstrate each variable's contribution to the crime prediction across the experimental findings.

These outcomes of the neighborhood techniques can assist the police in identifying the most important factors. At the same time, the global model identifies the essential features of the entire region.

Another research [54] focuses on predicting crime during or after psychiatric care. As modern threat-evaluation equipment is time-consuming to administer and offers constrained accuracy, this research looked to expand a predictive model designed to discover psychiatric patients liable to commit the crime. The authors utilized the longitudinal data of the affected Danish person registries, recognizing the 45,720 adult patients who had connected with the psychiatric system in 2014, of which 474 committed crimes leading to a forensic psychiatric treatment direction after discharge. Authors have used four machine learning models (Random Forest, Logistic Regression, XGBoost, and LightGBM) over various sociodemographic, judicial, and psychiatric variables. Their model identified 47% of future forensic psychiatric patients, making correct predictions in 57% of samples. This research demonstrates how a clinically useful preliminary risk assessment is achieved using machine learning classification techniques. Their research helps to flag possible forensic psychiatric patients while in contact with the general psychiatric system, which allows early intervention initiatives to be activated.

Another research work presents a graph-based ensemble classification approach for predicting crime reports better than traditional classifiers. Crime reports are graphically modeled to locate the maximal independent subset of features, and then they use decision tree classifiers on this set. Extensive experiments are performed to compare the overall performance of the proposed approach on numerous crime data sets. The developed ensemble classification model demonstrated better performance. Apart from predicting crime, researchers also focused on interpreting crime-related predictions. This will lead to a better understanding of what impacts crime detection.

## **4.1 Crime Prediction using Deep Learning techniques**

Deep learning has become a popular method for crime prediction in recent years. The studies included in the reference research articles use a range of deep learning algorithms, such as convolution neural networks (CNN), deep neural networks, and sentiment analysis, to analyze various types of data, including text, images, audio, and social media. These algorithms are capable of detecting patterns and anomalies in the data that can indicate criminal activity. One of the key strengths of deep learning is its ability to handle large and complex datasets, making it well-suited to the task of crime prediction. For example, image analysis algorithms can detect threatening objects in crime scenes and predict the likelihood of a crime occurring. Text mining techniques can be used to analyze crime-related tweets and make predictions about crime patterns. In addition, deep learning algorithms can detect anomalies in crime data in smart cities, which could indicate the presence of criminal activity. Researchers used these techniques to tackle both regression and classification problems in crime prediction as detailed in the below sections.

### 5.1.1 Deep Learning based Regression Methods for Crime Prediction

Deep learning algorithms in regression analysis are used as a tool for crime prediction to identify the factors most strongly associated with crime and use these relationships to make predictions about future crime patterns. The research articles in this area highlight the strengths of regression in modeling the relationship between multiple variables, including crime data, weather data, demographic data, social media data, and location data. A common theme among these research articles shown in Table 4 are the use of regression combined with deep learning techniques, such as convolution neural networks, recurrent neural networks, attention mechanisms, and sequential fusion models, to improve the accuracy of crime prediction.

In research focusing on theft crime prediction, the authors use regression to model the relationship between theft crime data, demographic data, and weather data. The regression model adopts two deep learning models, a Long Short-Term Memory (LSTM) network and a Spatio-Temporal Graph Convolutional Network (ST-GCN), to predict the likelihood of theft crimes in urban communities. The regression model can incorporate external information, such as weather data, which can influence crime patterns. The LSTM and ST-GCN models capture the temporal and spatial dependencies in the data, respectively. In another article the authors use regression to model the relationship between crime data, weather data, and social media data. The regression model is part of a more comprehensive, multi-module approach that uses attention mechanisms and sequential fusion models to predict the likelihood of crimes. This framework consists of four sub-modules, where the initial two modules adopt St-BiLSTM and ATTN-LSTM to process temporal and spatial features. Finally, two fusion models are used to abstract the data and make crime predictions on Chicago and San Francisco crime datasets.

In another research focused on using spatiotemporal data the authors use convolutional neural networks to develop a regression model on publicly available crime data in Los Angeles. The regression model is part of a more extensive, mixed spatiotemporal neural network that is designed to make real-time predictions about the likelihood of crimes. The authors claim that using regression in combination with the diverse spatiotemporal neural network results in improved accuracy and real-time performance. Another research that applies crime risk prediction across different cities uses regression to model the relationship between crime data and demographic data from other cities. The regression model is part of an unsupervised domain adaptation technique designed to predict the likelihood of crimes in new cities. The authors claim that using regression in combination with the unsupervised domain adaptation technique results in improved accuracy in crime prediction. A recent research article applied machine learning and deep learning methods to crime data from Xiaogan, a medium-sized city in China, to predict crime hourly. The models use weather, holiday, time slot ID, and Day of week information to extract

**Table 4** Crime detection using deep learning regression techniques

Methodology	DL Algorithm	Dataset	Perfor- mance
A Graph Convolution network in combination with ST-ResNet is used to perform spatiotemporal analysis and then LSTM is used to detect crimes in each community. GBDT is used to combine outputs of GCN and LSTM.	Long Short Term Memory (LSTM) in combination with Spatio Temporal Graph Convolution Network (ST-GCN) and Gradient Boosting Tree	Chicago Crime Data (2001 - 2020)	R-Squared - 0.84
Developed Attention-LSTM to process categorical-temporal data and Stacked Bidirectional LSTM model to process spatial information. The two were fused using feature and decision-level fusion. LA Crime data is separated into block-wise information based on the hour of the day, area, and city. These blocks are used to train CNN model.	ATTN-LSTM, St-Bi-LSTM, Fusion Models	San Francisco and Chicago Crime Data (2004 to 2017)	Aggre- gated R-Squared > 0.90
A model framework based on three phases i.e., intercity similar-grid matching, auxiliary features construction, and crime risk prediction using a dense CNN-based unsupervised domain adaptation.	Mixed Spatio Temporal Neural Network based on CNN	Los Angeles Crime Data	RMSE - 0.22
	CNN based on unsupervised domain adaptation model (UDAC)	New York City, Chicago, and Los Angeles Crime, weather and taxi Data (2015)	RMSE - 0.62
Weather, holiday, time slot ID, and Day of the week for spatial dependency and GRU-based features for Temporal dependencies to predict a number of incidents in different locations.	LoR, Support Vector Regressor, RF, GRU, Deep Crime and T-GCN	Xiaogan, data (2016). New York City data (2014). Additionally POI, Urban anomaly and Weather data are used.	Accuracy Xiaogan Dataset - 54%, NYC Dataset - 61%

spatial dependency (distance graph, poi similarity, and crime similarity). Temporal dependencies captured using GRU are used to predict the number of incidents in different locations.

These research articles highlight the versatility of regression as a tool that can be integrated with other techniques to enhance the performance of crime prediction models. Another commonality among the papers is the use of regression to model the relationship between crime data and other variables, such as weather and demographic data, to incorporate external information that may influence crime patterns. This allows for the creation of more comprehensive and accurate models of crime patterns. In summary, the five research papers demonstrate the strengths of regression as a tool for crime prediction, including its ability to model the relationship between multiple variables, its versatility in being integrated with other techniques, and its ability to incorporate external information that may influence crime patterns.

### **5.1.2 Deep Learning based Classification Methods for Crime Prediction**

Deep learning algorithms are trained on large amounts of data to classify instances into various categories. This makes them ideal for solving classification problems in crime detection. Deep learning models can accurately organize criminal activity and detect criminal intent by analyzing vast amounts of data, including images, audio, text, and social media data. For example, image-based data can provide detailed information about crime scenes, including the presence of weapons and other objects that may indicate criminal intent. Similarly, audio-based data can provide valuable insights into the tone and context of a conversation, helping to identify potential illegal activities. Another advantage of deep learning for classification problems in crime detection is the ability to identify hidden patterns in the data that traditional methods may miss. For example, deep neural networks can be trained to analyze crime-related tweets, uncovering patterns that indicate a potentially criminal act. The results of deep learning models in crime detection have been awe-inspiring. The papers reviewed under deep learning classification are listed in Table 5

In crime-related classification, two main types of deep learning algorithms are used: Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). CNN's are commonly used in image-based classification tasks, including crime scene prediction. In the research article focusing on crime scene data CNNs are trained to detect threatening objects in crime scenes, such as weapons. This allows the model to comprehensively analyze the crime scene, including the presence of things that may indicate criminal intent. On the other hand, RNN's are commonly used to study temporal patterns in data. In a research article focusing on crime prediction based on behavioral tracking the authors use a combination of deep learning algorithms, including CNN and RNN, to analyze behavioral tracking data and motion analysis data. The study shows that this approach can effectively predict criminal activities, such as theft and robbery, by analyzing patterns in the behavior and movements

**Table 5** Crime detection using deep learning classification techniques

Methodology	DL Algorithm	Dataset	Performance
R-CNN algorithm is adopted to detect multiple objects relevant to crime in real time videos and images	VGGNET19 based on Fast RCNN & RCNN	Youtube and Google Videos & Images	Train Accuracy - 100%
CNN is used to extract features from CCTV images to predict crime scene objects. Classify object or person in	CNN	2000 Images of weapons and blood scenes	Accuracy - 90.2%
video feed and track abnormal activity using Deep CNN (DCNN) and Recurrent Neural Network (RNN).	DCNN and RNN	1 Million video streams [Karpathy]	Mean Squared Error (MSE) - $1.38e - 10$
Extract tweet data and process text to identify text keywords that are relevant to weapons used for crime or criminal activity.	CNN, KNN, NB, Support Vector Machine (SVM), DT and RF	Saudi Arabia tweets (2017 - 2021).	Accuracy - 79%
Live CCTV data is used to detect faces and weapons using CNN- GRU model. [	RNN, Gradient Recurrent Unit (GRU) and LSTM.	CAVIAR Dataset	Accuracy - 95.97%
Preprocess crime data to extract features and train both traditional and Deep Learning models to predict crime region and type.	Traditional ML classifiers and Artificial Neural Network (ANN)	Saudi Arabia (2018 - 2020)	Accuracy for Crime Type - 99% & Crime Region - 81%
Audio data is used to extract Mel-Frequency Cepstrum Coefficients from sound waves and a CNN-RNN classifier is deployed & text data, features are extracted for BERT. Both models are combined by fusion model.	CNN & BERT	XD-Violence Audio & Video Data	Accuracy - 85.63%
Twitter data is processed by keyword filtering and labeled. The features vectors are generated to be inputted into models for training. Video	SVM and ANN	100000 Tweets (January 2020)	Accuracy - 90.3%
data is used to extract spatio temporal features and gestures to train DRNN for classification (Hostility & Violence).	Deep Reinforcement Neural Network (DRNN)	Crowd Violence, UCSD and Violent Flow Data.	Aggregate F1 score - 78%
Deep Learning and Machine learning models are applied to predict different types of crime and their relation to weather in Newyork city.	SVM, LR, DT, RF, CNN, RNN, LSTM and GRU	2018 NYPD and Weather data	Accuracy - 99%

of individuals in a given area. In the research articles [] focusing on social media data, Artificial Neural Networks are trained on crime-related text data to predict the likelihood of a crime occurring. These models analyze the context and tone of the text data to classify patterns that may indicate criminal activity.

In addition to the articles mentioned above, several other research studies have also used deep learning techniques for crime prediction and classification. These studies demonstrate the versatility of deep learning algorithms in crime-related classification tasks, as they can be applied to a wide range of data types, including images, text, audio, and social media data. For example, research focusing on crime anomaly detection [uses deep learning algorithms, including Autoencoders and CNN, to analyze crime patterns in smart cities. The study shows that this approach can effectively detect unusual crime patterns, which may indicate the presence of criminal activity. In another study focusing on audio and text data the authors use a multimodal deep learning model based on CNN and BERT that considers both audio and text data to classify crime-related events. This model is advantageous when audio data, such as 911 calls, is available and can provide a complete picture of the crime event.

These research articles including multiple other studies] highlight that deep learning algorithms, including CNN and RNN, have successfully applied to various data types for crime prediction and classification. These studies demonstrate the versatility of deep learning algorithms in this field and provide valuable insights into the factors contributing to criminal activity. By leveraging the strengths of these models, law enforcement agencies can gain a more comprehensive understanding of criminal activity and take proactive measures to prevent crime from occurring.

## **5 Discussion and future work**

The application of machine learning and deep learning algorithms to predict or detect crime has shown great promise in addressing this complex problem. By utilizing vast datasets and advanced algorithms, these technologies can potentially improve the accuracy and effectiveness of crime prediction models. However, despite the advances in this field, there are still significant gaps in our understanding of how these technologies can be effectively applied to the problem of crime prediction. In this section, we will discuss the potential benefits of machine learning and deep learning algorithms for crime prediction and the existing research gaps that need to be addressed in future research.

One of the major advantages of machine learning and deep learning algorithms for crime prediction is the ability to analyze large datasets and identify patterns in criminal activity or behavior. The ability of these algorithms to process vast amounts of data, including social media and other online sources, can provide valuable insights into criminal activities that are yet to be committed. Furthermore, deep learning algorithms like CNN and RNNs have been used to



analyze video footage from security cameras. This capability provides a more accurate and efficient means of detecting criminal activities. Another major benefit of machine learning and deep learning for crime prediction is the ability to develop real-time prediction models. These models can be used to analyze crime data in real-time and to predict future crime incidents. This supports law enforcement agencies to act quickly if a criminal activity is being committed. Additionally, integrating decentralized machine learning algorithms with wearable technology, such as body cameras and smartwatches, provides new opportunities to collect and analyze data related to criminal activities.

Even though machine learning and deep learning algorithms support effective crime prediction, there are still some significant challenges that need to be addressed. One of the major challenges in this area is the need for interpretable models that can provide clear explanations of their predictions. This is particularly important in the context of crime prediction, as incorrect predictions might lead to serious consequences for individuals and communities. Apart from the existing model-based explanation methods, it is also important to incorporate causal based explanations that focus on cause and effect relationship between crime patterns and relevant feature variables. Another challenge that needs to be addressed is the need for more accurate and reliable data. In order to effectively apply machine learning and deep learning for crime prediction, it is important to have access to high-quality and up-to-date crime data. As this study showed that many earlier researchers took advantage of data related to demographics, whether outside of crime-relevant datasets, there is a need to develop algorithms that can accurately handle data from multiple sources and integrate it into a single predictive model.

Another significant area of focus should be to research more on the ethical implications of using machine learning and deep learning for crime prediction. As these technologies are used to predict individuals and communities, it is important to ensure that they do not perpetuate existing biases or lead to discrimination. Furthermore, there is a need for more research on the privacy implications of using these technologies for crime prediction, this included but not limited to the potential risks of data breaches and the misuse of personal information. Another significant gap in the existing research is the need for more research studies on the effectiveness of machine learning and deep learning for crime prediction in the real world. While these technologies have shown great promise in this area, there is a need for more rigorous evaluations of their accuracy and effectiveness in real-world scenarios. Additionally, there is a need for more research on the scalability of these technologies and the challenges associated with their implementation in large-scale systems.

In conclusion, machine learning and deep learning methodologies have the potential to transform the field of crime prediction by providing more accurate and effective methods for predicting criminal activities. However, in order to fully realize the potential of these technologies, it is important to address the existing research gaps and challenges, including the need for interpretable models, accurate and reliable data, ethical considerations, and more rigorous

evaluations of their accuracy and effectiveness. By addressing these gaps, we can advance our understanding of the role of machine learning and deep learning algorithms in crime prediction and contribute to the development of more effective and efficient policing strategies.

**Table 6** Future Research Directions

Concept Areas and Research Questions
Feature-Oriented
1. How do visual features (e.g., images/videos) affect (i.e., positively or negatively) the crime identification?
2. How do audio features (e.g., voiceover, call recordings) affect (i.e., positively or negatively) the crime identification?
3. How do Textual features (e.g., investigating notes, complaint statements, crime interviews) affect (i.e., positively or negatively) the crime identification?
4. What are the visual features to be considered to detect neighborhood crime?
5. What are the Audio features to be considered to detect neighborhood crime?
Technique-Oriented
1. Are there any reinforcement learning techniques available to detect neighborhood crime?
2. Are there any transfer learning techniques available to detect neighborhood crime?
3. Are there any generative adversarial neural network techniques available to detect neighborhood crime?

As a future research goal and agenda, we have illustrated a range of prospective research directions in the area of neighborhood crime based on the importance and also current lack of focus in the areas. From Table 6, future researchers may want to address concerns like - "Are there any reinforcement learning techniques available to detect the neighborhood crime?", and What are the visual features to be considered to detect the neighborhood crime? The relevant datasets are available for such identified future research questions; such data may be utilized to accomplish the goal of early recognition of neighborhood crime.

The presented literature base and futuristic research goals offer a number of elements to direct future studies and thus theoretically support the effort to identify neighborhood crimes. Our systematic review offers a thorough grasp of the characteristics and methods used by earlier research to recognize and detect crimes. In addition, we have outlined 8 more research questions that fall into the categories of technique-oriented questions and feature-oriented questions in Table 6. Practically speaking, this systematic review serves as a guide for various researchers, practitioners, first responders, and crime analysts, to take into account the studied features and techniques to effectively understand and detect the crimes that, in part, foster the effort for early crime detection.

## 6 Conclusion

The complexity of crimes has increased along with technological development, creating difficult problems for law enforcement. Researchers' interest in utilizing machine learning and deep learning to predict crime has increased recently, with an emphasis on finding patterns and trends in crime occurrences. In order to analyze the various machine learning and deep learning algorithms used in predicting crime, this paper looks at more than 150 articles. We have significantly studied the selected 51 articles to extract the essence of utilized various ML and DL techniques along with the publicly available datasets. The use of machine learning and deep learning algorithms to anticipate or identify criminal activity has shown significant promise in resolving the crime detection problem. These advances may help to increase the precision and efficacy of crime prediction models by leveraging large datasets and sophisticated algorithms. Although there is a lack of literary wisdom on how these technologies can be used to solve the problem of crime prediction, despite the advancements in this sector. Thus our findings help to understand the implications of various ML and DL techniques. Also, our mentioned datasets and future directions will help the existing research community to pursue their research in the area of crime prediction.

GITHUB LINK:

<https://github.com/naanmudhalvan-SI/PBL-NT-GP-14610-1683893661>

