

CRIME PREDICTION USING MACHINE LEARNING AND DEEPMLEARNING

Dr P. Sreenivasulu¹,B Siva Sankar ², Ch.Yamini³

Professor¹ , Student² , Student³

,Electronics and Communications Engineering

Audisankara College of Engineering &Technology, Gudur, Tirupati, A.P, India.

Abstract

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. The field of machine learning is a subset of artificial intelligence that uses statistical models and algorithms to analyze and make predictions based on data.

1. Introduction

On the other hand, deep learning methods are 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 methodologies 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. Apart from

predicting 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. 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. Predictive policing refers to using data and analytic 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 hot spots 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.

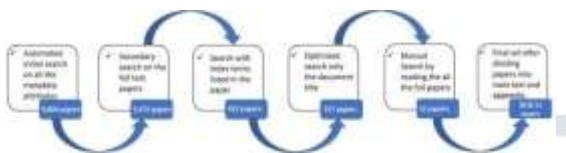
This information is used to create a predictive model that can be used to identify potential crime hotspots and predict future crime incidents. Another application of deep learning in crime prediction is computer vision and video analysis. This technology has been used to analyze video footage from surveillance 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 prediction. 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 data sets availability. We have highlighted distinct publicly available data sets 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.

1. RESEARCH METHODOLOGY

The primary research aims to find various efficient algorithms for predicting neighborhood crimes. In our previous work, we used statistical analysis to predict the crimes in New York 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 data sets 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 Table 7. In this survey, we have used a combination of an automated and manual search shown in Figure 2. In the initial stages, we focused on using the automatic digital

FIGURE 1. Research paper selection methodology. 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.

IEEE query: ((“Document Title”: “crime”) AND (“Document Title”: “predict” OR “Document Title”: “detect” OR “Document Title”: “recognize” OR “Document Title”: “machine learning” OR “Document Title”: “deep learning” OR “Document

Title”:“clustering” OR “Document Title”:“natural language processing”))

Science Direct Query: (“crime”) AND (“prediction” OR “detection” OR “recognition” OR “machine learning” OR “deep learning” OR “clustering” OR “natural language processing”)

ACM Query: “query”: { Full text:(“crime*”) AND (“predict*” OR “detect” OR “recognize” “machine learning” OR “deep learning” OR “clustering” OR “natural language processing”)) }

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 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 meta data 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

(<https://www.sciencedirect.com>)

Elsevier

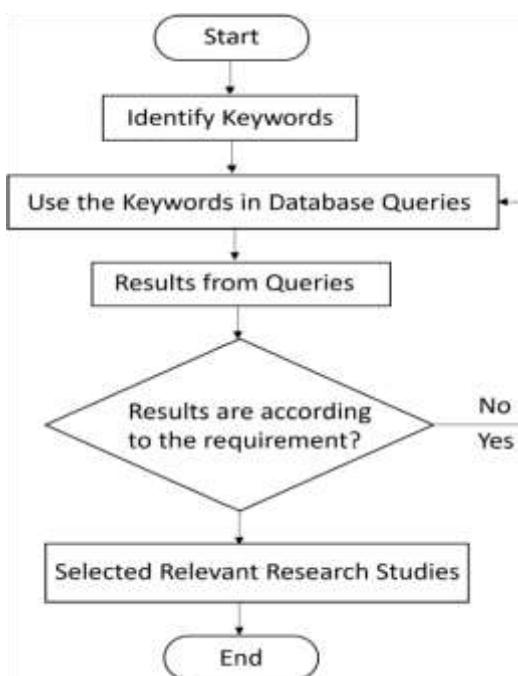


FIGURE 2. Steps involved for typical crime detection.

ACM (DL) Digital Library (<https://dl.acm.org>)

IEEE Explore Access Digital Library (<https://ieeexplore.ieee.org>)

After applying the automated filters as shown in Figure 1, 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, data sets 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 data sets they can use to apply the novel algorithms and get the results for their research.

3.LITERATURE ANALYSIS

3.1 PRE-ASSESSMENT LITERATURE ANALYSIS

An analysis of collected literature data from the distinct research databases is essential to receive information

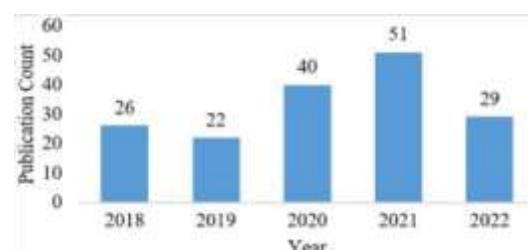


FIGURE 3. Research publication trends from 2018-2022.

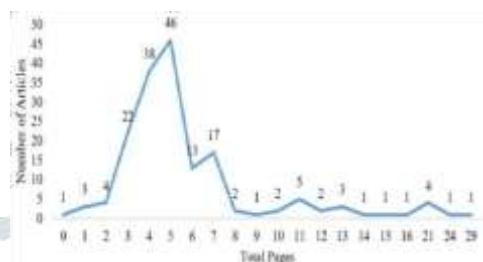


FIGURE 4. Distribution of article's page counts.

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, where 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. Figure 3 shows the research publication trends in neighborhood crime 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.

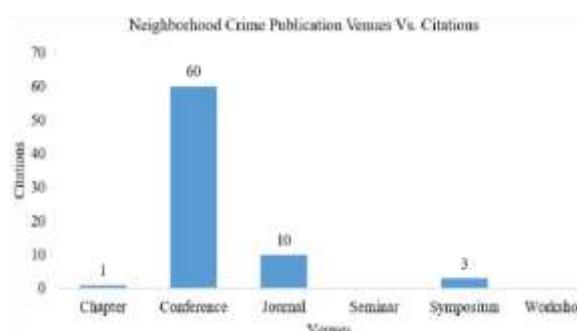


FIGURE 5. Distribution of article's citations.

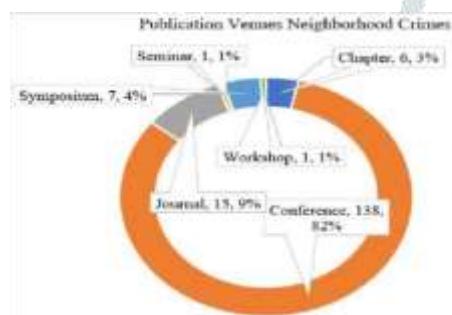


FIGURE 6. Distribution of research articles at various venues.

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 (0 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 venues. Wherein this figure depicts the venue name (e.g., conference, journal,), 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.

3.2. 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



selected articles.



FIGURE 8. Distribution of neighborhood crime-related selected article's technique types.

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.

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 of crime detection. Moreover, "prediction", "algorithms", and "techniques" are the words that indicate the aim of the existing studies. This word cloud is aligned with our selection criteria and objectives that further validate our final set of selected articles for this literature review study.

Figure 8 shows the Distribution of neighborhood crime-related selected article's technique types among Science Direct and IEEE databases, respectively. As shown in the Figure 8, among the neighborhood crime-related articles, machine learning (ML), the combination of machine learning and deep learning (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

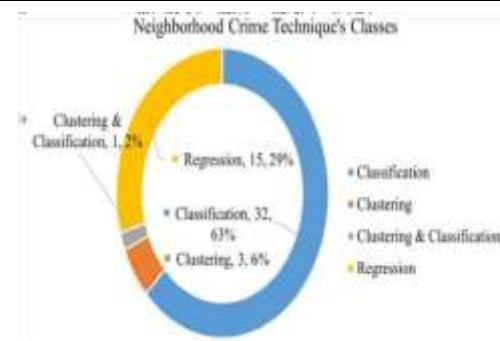


FIGURE 9. Distribution of neighborhood crime-related selected article's technique's classes.

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.

Figure 9 shows the distribution of neighborhood crime-related selected article's technique classes among Science Direct and IEEE databases. As shown in Figure 9, 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.

Figure 10 shows the Distribution of neighborhood Crime related selected article's Technique classes Versus technique Type among Science Direct and IEEE databases. Figure 10 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.

4.TOOLS

Hardware Tools:-

System : Intel i5 Processor.

Hard Disk : 500 GB.

Monitor : 15" LED.

Input Devices : Keyboard, Mouse.

Ram : 8 GB.

Software Tools:-

Operating system : Windows 11

Coding Language : Python 3

Web Framework : Flask.

Frontend : HTML, CSS, JavaScript.

6. 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.

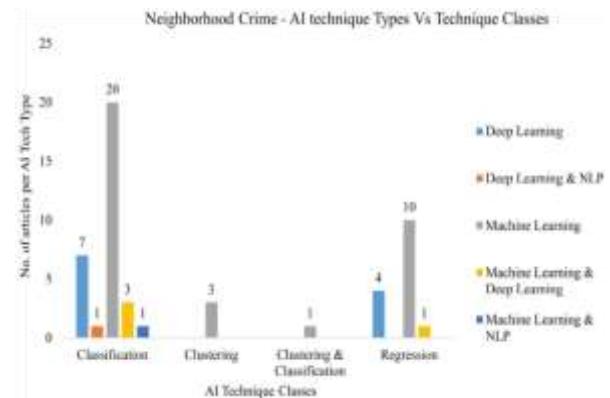


FIGURE 10. Distribution of neighborhood Crime related selected article Technique classes Versus technique Type.



FIGURE 11. Architecture flow of crime prediction.

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.

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

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 occurrence.

7.RESULTS

The results of crime detection using machine learning and deep learning can vary depending on factors such as the quality of data, the sophistication of algorithms, and the specific context of the crime. Generally, these techniques have shown promise in improving the accuracy of crime prediction and detection compared to traditional methods.

They can help law enforcement agencies in various ways, such as:

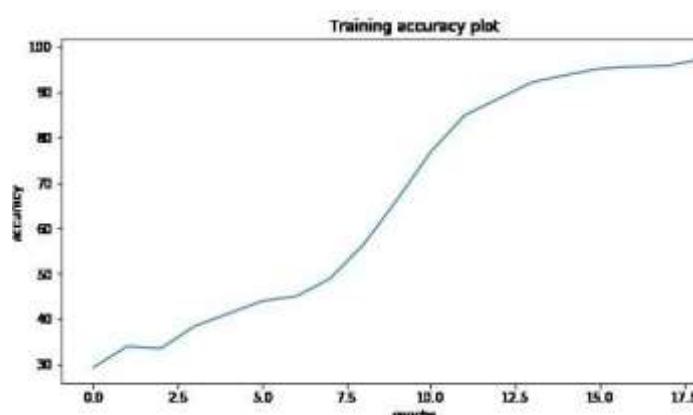
1. Predictive policing: Identifying high-risk areas and times for crime to allocate resources more effectively.
2. Suspect identification: Analyzing surveillance footage or forensic data to identify suspects.
3. Crime pattern recognition: Analyzing historical crime data to identify patterns and trends for proactive measures.
4. Anomaly detection: Identifying unusual behaviors or activities that could indicate criminal behavior.

privacy concerns, and the need for interpretability of results must be carefully addressed to ensure the fairness and effectiveness of crime detection systems.



8.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



limitations. Challenges such as bias in data,

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.

9.FUTURE WORK:-

The adoptions of machine 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 the current understanding of how these technologies can be effectively applied to the problem of crime prediction. In this section, we will discuss the potential

TABLE 6. Future Research Directions.

benefits of machine learning and deep learning algorithms for crime prediction and the future research.

One of the primary 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 smart watches, 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 needs 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 largescale systems.

Overall, 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.

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.

10. REFERENCES

N. Shah, N. Bhagat, and M. Shah, “Crime forecasting: A machine learning and computer vision approach to crime prediction and prevention,” Vis. Comput. Ind., Biomed., Art, vol. 4, no. 1, pp. 1–14, Apr. 2021.

S. A. Chun, V. A. Paturu, S. Yuan, R. Pathak, V. Atluri, and N. R. Adam, “Crime prediction model using deep neural networks,” in Proc. 20th Annu. Int. Conf. Digit. Government Res., Jun. 2019, pp. 512–514.

S. S. Kshatri, D. Singh, B. Narain, S. Bhatia, M. T. Quasim, and G. R. Sinha, "An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach," *IEEE Access*, vol. 9, pp. 67488–67500, 2021.

C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electron. Mark.*, vol. 31, no. 3, pp. 685–695, Apr. 2021.

W. Safat, S. Asghar, 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.

S. Kim, P. Joshi, P. S. Kalsi, and P. Taheri, "Crime analysis through machine learning," in Proc. IEEE 9th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON), Nov. 2018, pp. 415–420.

D. M. Raza and D. B. Victor, "Data mining and region prediction based on crime using random forest," in Proc. Int. Conf. Artif. Intell. Smart Syst. (ICAIS), Mar. 2021, pp. 980–987.

L. Elluri, V. Mandalapu, and N. Roy, "Developing machine learning based predictive models for smart policing," in Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP), Jun. 2019, pp. 198–204.

A. Meijer and M. Wessels, "Predictive policing: Review of benefits and drawbacks," *Int. J. Public Admin.*, vol. 42, no. 12, pp. 1031–1039, Sep. 2019.

S. Hossain, A. Abtahee, I. Kashem, M. M. Hoque, and I. H. Sarker, "Crime prediction using spatio-temporal data," in Computing Science, Communication and Security. Gujarat, India: Springer, 2020, pp. 277–289.