International Journal of Scientific Research in Science, Engineering and Technology

Print ISSN - 2395-1990
Online ISSN : 2394-4099

OPEN ACCESS

Available Online at : www.ijsrset.com
doi : https://doi.org/10.32628/IJSRSET241134

Crime Prediction Using Machine Learning and Deep Learning

P. Karthik*[1] , P. Jayanth[2], K. Tharun Nayak[3] , K. Anil Kumar[4]

*[1]UG Student, Department of Computer Science and Engineering (Internet of Things), Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India

[4]Assistant Professor, Department of Computer Science and Engineering (Internet of Things), Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India

ARTICLEINFO

Article History:

Accepted: 01 May 2024
Published: 05 May 2024

Publication Issue :
Volume 11, Issue 3
May-June-2024
Page Number :
08-15

ABSTRACT

The utilization of machine learning and deep learning methods for crime prediction has become a focal point for researchers, aiming to decipher the complex patterns and occurrences of crime. This review scrutinizes an extensive collection of over 150 scholarly articles to delve into the assortment of machine learning and deep learning techniques employed in forecasting criminal behaviour. It grants access to the datasets leveraged by researchers for crime forecasting and delves into the key methodologies utilized in these predictive algorithms. The study sheds light on the various trends and elements associated with criminal behaviour and underscores the existing deficiencies and prospective avenues for advancing crime prediction precision. This thorough examination of the current research on crime forecasting through machine learning and deep learning serves as an essential resource for scholars in the domain. A more profound comprehension of these predictive methods will empower law enforcement to devise more effective prevention and response strategies against crime.

Keywords : Machine Learning , Deep Learning , Crime Prediction , Research Review , Algorithm Application, Dataset Analysis , Trend Identification , Criminal Activity Factors , Predictive Accuracy , Future Directions, Law Enforcement Strategies

I. INTRODUCTION

Crime prediction integrates data science and machine learning to anticipate crimes, aiding proactive policing. It uses historical data and algorithms to identify crime patterns and inform resource allocation. Advances in machine learning have enhanced crime prediction accuracy, supporting safer communities and smarter law enforcement. Understanding the factors behind crime helps address the challenges of urbanization and crime prevention.

II. EXISTING SYSTEM

The crime prediction system uses the Random Forest Classifier to forecast and categorize crimes effectively. It stands out for its accuracy, handling complex data, and combining multiple decision trees for reliable predictions. The dataset contains 15 features, aiding in comprehensive analysis. Tailored for law enforcement and policymakers, the system supports strategic decision-making and enhances public safety, demonstrating high accuracy without notable drawbacks.

Existing System Disadvantages :

1. Computational Intensity: The system's need for substantial computational power can be a hurdle in real-time applications or when resources are limited.
2. Lack of Interpretability: The complexity of the Random Forest algorithm can make it difficult to understand and explain the model's decisions, affecting transparency.

III. PROPOSED SYSTEM

The enhanced crime prediction system utilizes Python and integrates Decision Tree and Bagging Classifiers to improve accuracy in Portland's crime analysis from 2015 to 2023. It maintains high accuracy, processes a vast dataset with over half a million data points, and classifies 20 different crime types. Features like address, case number, and offense details provide rich input for the models.

This system is designed for adaptability and scalability, with user-friendly interfaces for stakeholders. It includes mechanisms for continuous model updates and hyperparameter tuning to adapt to changing crime patterns. The goal is to refine crime prediction techniques, optimize accuracy, and overcome previous limitations, making it a robust tool for law enforcement and policymakers.

Proposed System Advantages:

1. High Accuracy: The system's use of Decision Tree and Bagging Classifiers significantly boosts the accuracy of crime predictions.
2.Robustness and Interpretability: The Bagging Classifier enhances robustness against overfitting, while the Decision Tree Classifier maintains transparency in the predictive process.

IV. SYSTEM ARCHITECTURE

Our research outlines machine learning-based system architecture designed to differentiate between fake and real accounts. It starts with a dataset that undergoes pre-processing and feature selection to refine the data. This prepared data is then input into classifiers like Random Forest and Decision Tree, which predict the authenticity of an account. The results are labelled as either 'Fake' or 'Real' accounts, followed by a performance analysis and graphical representation of the outcomes.

In essence, the architecture captures the entire process from raw data to actionable insights. It emphasizes the importance of data preparation and the use of ensemble learning methods to ensure accurate predictions. The final step involves analysing the model's performance, likely to assess its effectiveness and accuracy in classifying accounts. This systematic approach is crucial for maintaining the integrity of digital platforms by identifying and filtering out fraudulent accounts.

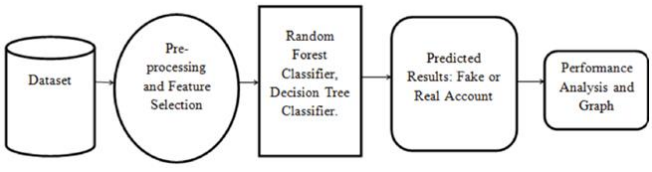

Figure 1 : System architecture

V. LITERATURE SURVEY

This paper explores the use of machine learning (ML) and computer vision to predict and prevent crime.

With crime rates rising, traditional methods fall short, prompting the need for innovative solutions. Our research demonstrates how ML and computer vision can transform law enforcement, offering faster, more accurate crime detection and prevention. We present case studies that validate the effectiveness of these technologies, suggesting a significant shift towards more advanced crime-fighting strategies.

This project evaluates the use of neural networks in machine learning to predict criminal behaviour from arrest records, addressing imbalanced data with data augmentation and a weighted loss function. Focusing on deep fully connected neural networks, suitable for complex feature relationships and limited domain knowledge, the study underscores the potential of machine learning in crime classification and suggests further exploration of these techniques for enhanced predictive accuracy.

Ensemble learning is a collective approach that combines the outputs of various classifiers to generate new predictions. Studies indicate that these ensemble models tend to be more dependable than single classifiers. Despite the availability of numerous ensemble techniques, selecting the right one for a specific dataset remains challenging. In the context of crime prediction in India, this complexity is heightened by the dynamic nature of criminal activities. Our research introduces a method called assemble-stacking based crime prediction method (SBCPM), utilizing SVM algorithms to enhance crime prediction accuracy using MATLAB. This method is compared with other models like J48, SMO, Naïve Bayes bagging, and Random Forest, showing that ensemble models can sometimes surpass others, especially in terms of correlation and error metrics. Achieving a 99.5% accuracy rate in tests, our method outperforms previous studies and aligns well with criminological theories, proving effective for predicting crimes and suggesting that ensemble models may offer superior accuracy over individual classifiers.

Crimes pose a significant challenge to justice and require effective control measures. This study leverages various machine learning algorithms, including logistic regression, SVM, Naïve Bayes, KNN, decision tree, MLP, random forest, XGBoost, and time series analysis with LSTM and ARIMA models, to enhance crime prediction and forecasting. The LSTM model showed promising results in terms of RMSE and MAE. Data analysis indicates a potential decrease in Chicago's crime rate and a minor rise in Los Angeles's, with February having fewer incidents. Predictions suggest a moderate future increase in Chicago's crime rate with a possible subsequent decline, while Los Angeles is expected to see a significant decrease as per the ARIMA model. These insights help in early crime detection, identifying high-risk areas, and informing police strategies for improved safety.

Crime significantly disrupts societal norms and hampers the social and economic stability of nations, a situation that Bangladesh also faces. This study concentrates on forecasting crime-prone areas, which can be instrumental in anticipating criminal incidents across various cities within specific timeframes. With insights into upcoming trends, law enforcement and detectives can implement preventive measures to decrease crime rates. The Random Forest algorithm is employed in this research to identify districts with varying levels of criminal activity.

VI.EXISTING ALGORITHM

Random Forest Classifier : The Random Forest Classifier is a sophisticated machine learning algorithm that belongs to the ensemble learning family. It constructs multiple decision trees during the training phase and outputs the class that is the mode of the classes for classification tasks, or the mean prediction for regression tasks. This method is known for its high accuracy, ability to handle large datasets with higher dimensionality, and robustness to overfitting. Random Forest works by creating a 'forest' of decision trees

from randomly selected subsets of the training set and then averaging the predictions to improve the overall result .

VII. PROPOSED ALGORITHM

The Decision Tree Classifier is a type of supervised learning algorithm that is used for classification tasks. It works by creating a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. It's like a flowchart where each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label . On the other hand, the Bagging Classifier, also known as Bootstrap Aggregating, is an ensemble meta-estimator that fits multiple base classifiers (like Decision Trees) on random subsets of the original dataset. It then aggregates their individual predictions to form a final prediction, which helps to improve the stability and accuracy of machine learning algorithms by reducing variance and avoiding overfitting .

VIII. RESULTS

User

Input : Enter Login name and Password

Output : If valid user name and password then directly open the home page otherwise show error message and redirect to the registration page.

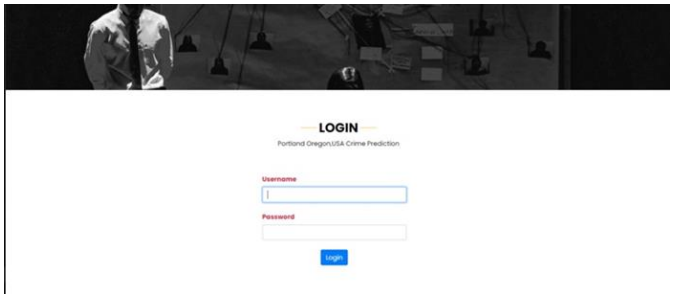

Figure 1 : Login Page

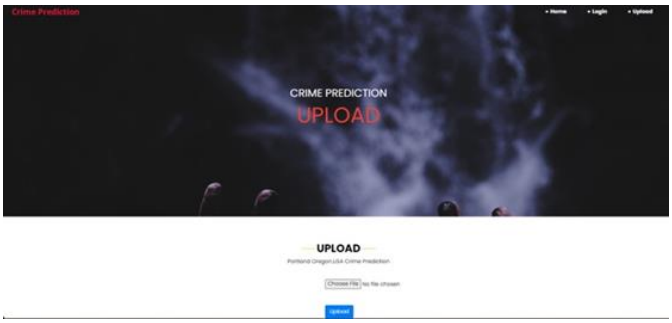

Figure 2: Dataset Upload Page

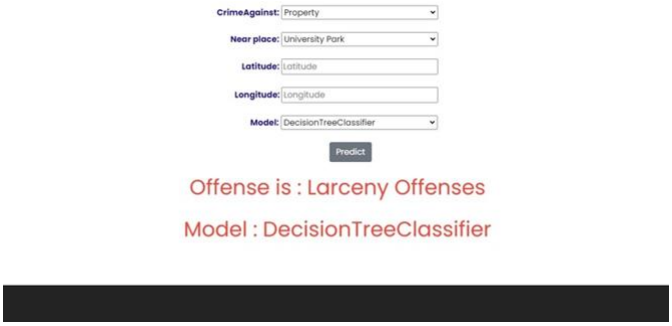

Figure 3: Model Running

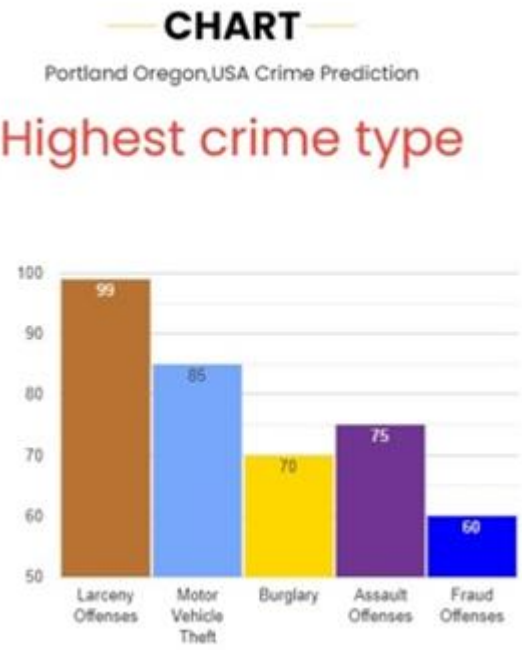

Figure 4: Classification of Crimes based on Dataset

IX. CONCLUSION

To sum up, the "Crime Prediction Using Machine Learning" initiative offers a sophisticated method for forecasting and sorting a range of criminal offenses in Portland, Oregon, USA. Utilizing Python's capabilities and machine learning techniques like the Decision

Tree and Bagging Classifiers, the project boasts impressive precision, with a 98% accuracy rate for the training dataset and 95% for the test dataset.

The project analyses crime trends from 2015 to 2023 using a dataset containing over half a million records, enabling a detailed examination of crime evolution. It categorizes 20 different types of crimes, ranging from 'Larceny Offenses' to 'Gambling Offenses,' offering a detailed perspective on the area's crime scene.

The Decision Tree Classifier enhances the system's transparency, promoting confidence among users, while the Bagging Classifier increases the system's strength, preventing overfitting and boosting prediction consistency.

With a comprehensive set of features like location, case number, type of crime, neighbourhood, occurrence date and time, offense category, offense type, and geographical coordinates, the models are well-equipped to deliver well-informed forecasts by considering various elements.

This system not only overcomes the shortcomings of previous models but also brings new features such as the ability to adapt to changing crime trends and scalability for various application contexts. The addition of user-friendly interfaces and the possibility of incorporating more data sources improve the system's usability and the precision of insights for law enforcement, urban planners, and policy makers.

Ultimately, the "Crime Prediction Using Machine Learning" project emerges as a potent and efficient instrument for improving public safety, optimizing the distribution of resources, and supporting active decision-making in city settings. The project's achievement in maintaining high accuracy and its comprehensive crime prediction strategy highlight its significance in bolstering security and shaping well-informed public policies.

X. REFERENCES

[1]. 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.

[2]. 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.

[3]. 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.

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

[5]. 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.

[6]. 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.

[7]. 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.

[8]. 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.

[9]. 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.

[10]. 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.

[11]. M. Saraiva, I. Matijosaitiene, S. Mishra, and A. Amante, "Crime prediction and monitoring in Porto, Portugal, using machine learning, spatial and text analytics," ISPRS Int. J. Geo-Inf., vol. 11, no. 7, p. 400, Jul. 2022.

[12]. O. Kounadi, A. Ristea, A. Araujo, and M. Leitner, "A systematic review on spatial crime forecasting," Crime Sci., vol. 9, pp. 1-22, Dec. 2020.

[13]. L. J. Morrisey, "Bibliometric and bibliographic analysis in an era of electronic scholarly communication," in Scholarly Communication in Science and Engineering Research in Higher Education. Evanston, IL, USA: Routledge, 2013, pp. 149-160.

[14]. M. Hofmann and A. Chisholm, Text Mining and Visualization: Case Studies Using Open-Source Tools, vol. 40. Boca Raton, FL, USA: CRC Press, 2016.

[15]. P. Tamilarasi and R. U. Rani, "Diagnosis of crime rate against women using K-fold cross validation through machine learning," in Proc. 4th Int. Conf. Comput. Methodologies Commun. (ICCMC), Mar. 2020, pp. 1034-1038.

[16]. A. Kumar, A. Verma, G. Shinde, Y. Sukhdeve, and N. Lal, "Crime prediction using K-nearest neighboring algorithm," in Proc. Int. Conf. Emerg. Trends Inf. Technol. Eng., Feb. 2020, pp. 1-4.

[17]. S. Agarwal, L. Yadav, and M. K. Thakur, "Crime prediction based on statistical models," in Proc. 11th Int. Conf. Contemp. Comput. (IC), Aug. 2018, pp. 1-3.

[18]. S. R. Bandekar and C. Vijayalakshmi, "Design and analysis of machine learning algorithms for the reduction of crime rates in India," Proc. Comput. Sci., vol. 172, pp. 122-127, Jan. 2020.

[19]. A. Gahalot, S. Dhiman, and L. Chouhan, "Crime prediction and analysis," in Proc. 2nd Int. Conf. Data, Eng. Appl. (IDEA), Feb. 2020, pp. 1-6.

[20]. B. Sivanagaleela and S. Rajesh, "Crime analysis and prediction using fuzzy C-means algorithm," in Proc. 3rd Int. Conf. Trends Electron. Informat. (ICOEI), Apr. 2019, pp. 595-599.

[21]. A. M. Shermila, A. B. Bellarmine, and N. Santiago, "Crime data analysis and prediction of perpetrator identity using machine learning approach," in Proc. 2nd Int. Conf. Trends Electron. Informat. (ICOEI), May 2018, pp. 107-114.

[22]. C. Catlett, E. Cesario, D. Talia, and A. Vinci, "A data-driven approach for spatio-temporal crime predictions in smart cities," in Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP), Jun. 2018, pp. 17-24.

[23]. C. Catlett, E. Cesario, D. Talia, and A. Vinci, "Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments," Pervas. Mobile Comput., vol. 53, pp. 62-74, Feb. 2019.

[24]. F. Yi, Z. Yu, F. Zhuang, X. Zhang, and H. Xiong, "An integrated model for crime prediction using temporal and spatial factors," in Proc. IEEE Int. Conf. Data Mining (ICDM), Nov. 2018, pp. 1386-1391.

[25]. S. K. Dash, I. Safro, and R. S. Srinivasamurthy, "Spatio-temporal prediction of crimes using network analytic approach," in Proc. IEEE Int. Conf. Big Data (Big Data), Dec. 2018, pp. 1912-1917.

[26]. X. Han, X. Hu, H.Wu, B. Shen, and J.Wu, "Risk prediction of theft crimes in urban communities: An integrated model of LSTM and ST-GCN," IEEE Access, vol. 8, pp. 217222-217230, 2020.

[27]. Z. Li, C. Huang, L. Xia, Y. Xu, and J. Pei, "Spatial-temporal hypergraph self-supervised learning for crime prediction," in Proc. IEEE 38th Int.

Conf. Data Eng. (ICDE), May 2022, pp. 2984-2996.

[28]. N. Tasnim, I. T. Imam, and M. M. A. Hashem, "A novel multi-module approach to predict crime based on multivariate spatio-temporal data using attention and sequential fusion model," IEEE Access, vol. 10, pp. 48009-48030, 2022.

[29]. B. Zhou, L. Chen, S. Zhao, S. Li, Z. Zheng, and G. Pan, "Unsupervised domain adaptation for crime risk prediction across cities," IEEE Trans. Computat. Social Syst., early access, Sep. 29, 2022, doi: 10.1109/TCSS.2022.3207987.

[30]. U. M. Butt, S. Letchmunan, F. H. Hassan, M. Ali, A. Baqir, T. W. Koh, and H. H. R. Sherazi, "Spatio-temporal crime predictions by leveraging artificial intelligence for citizens security in smart cities," IEEE Access, vol. 9, pp. 47516-47529, 2021.

[31]. S. Yao, M. Wei, L. Yan, C. Wang, X. Dong, F. Liu, and Y. Xiong, "Prediction of crime hotspots based on spatial factors of random forest," in Proc. 15th Int. Conf. Comput. Sci. Educ. (ICCSE), Aug. 2020, pp. 811-815.

[32]. M. Sathiyanarayanan, A. K. Junejo, and O. Fadahunsi, "Visual analysis of predictive policing to improve crime investigation," in Proc. Int. Conf. Contemp. Comput. Informat. (ICI), Dec. 2019, pp. 197-203.

[33]. A. Araujo, N. Cacho, L. Bezerra, C. Vieira, and J. Borges, "Towards a crime hotspot detection framework for patrol planning," in Proc. IEEE 20th Int. Conf. High Perform. Comput. Commun., IEEE 16th Int. Conf. Smart City, IEEE 4th Int. Conf. Data Sci. Syst. (HPCC/SmartCity/DSS), Jun. 2018, pp. 1256-1263.

[34]. A. A. Almuhanna, M. M. Alrehili, S. H. Alsubhi, and L. Syed, "Prediction of crime in neighbourhoods of New York City using spatial data analysis," in Proc. 1st Int. Conf. Artif. Intell. Data Analytics (CAIDA), Apr. 2021, pp. 23-30.

[35]. A. Baqir, S. U. Rehman, S. Malik, F. U. Mustafa, and U. Ahmad, "Evaluating the performance of hierarchical clustering algorithms to detect spatio-temporal crime hot-spots," in Proc. 3rd Int. Conf. Comput., Math. Eng. Technol. (iCoMET), Jan. 2020, pp. 1-5.

[36]. A. Algefes, N. Aldossari, F. Masmoudi, and E. Kariri, "A text-mining approach for crime tweets in Saudi Arabia: From analysis to prediction," in Proc. 7th Int. Conf. Data Sci. Mach. Learn. Appl. (CDMA), Mar. 2022, pp. 109-114.

[37]. S. P. C. W. Sandagiri, B. T. G. S. Kumara, and B. Kuhaneswaran, "Detecting crime related Twitter posts using artificial neural networks based approach," in Proc. 20th Int. Conf. Adv. ICT Emerg. Regions (ICTer), Nov. 2020, pp. 5-10.

[38]. M. A. Permana, M. I. Thohir, T. Mantoro, and M. A. Ayu, "Crime rate detection based on text mining on social media using logistic regression algorithm," in Proc. IEEE 7th Int. Conf. Comput., Eng. Design (ICCED), Aug. 2021, pp. 1-6.

[39]. X. Zhou, X. Wang, G. Brown, C. Wang, and P. Chin, "Mixed spatiotemporal neural networks on real-time prediction of crimes," in Proc. 20th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA), Dec. 2021, pp. 1749-1754.

[40]. R. Shenoy, D. Yadav, H. Lakhotiya, and J. Sisodia, "An intelligent framework for crime prediction using behavioural tracking and motion analysis," in Proc. Int. Conf. Emerg. Smart Comput. Informat. (ESCI), Mar. 2022, pp. 1-6.

[41]. N. Aldossari, A. Algefes, F. Masmoudi, and E. Kariri, "Data science approach for crime analysis and prediction: Saudi Arabia use-case," in Proc. 5th Int. Conf. Women Data Sci. Prince Sultan Univ. (WiDS PSU), Mar. 2022, pp. 20-25.

[42]. Y. Ma, K. Nakamura, E. Lee, and S. S. Bhattacharyya, "EADTC: An approach to interpretable and accurate crime prediction," in

Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC), Oct. 2022, pp. 170-177.

[43]. M. Boukabous and M. Azizi, "Multimodal sentiment analysis using audio and text for crime detection," in Proc. 2nd Int. Conf. Innov. Res. Appl. Sci., Eng. Technol. (IRASET), Mar. 2022, pp. 1-5.

[44]. L. G. A. Alves, H. V. Ribeiro, and F. A. Rodrigues, "Crime prediction through urban metrics and statistical learning," Phys. A, Stat. Mech. Appl., vol. 505, pp. 435-443, Sep. 2018.

[45]. J. He and H. Zheng, "Prediction of crime rate in urban neighborhoods based on machine learning," Eng. Appl. Artif. Intell., vol. 106, Nov. 2021, Art. no. 104460.

[46]. H. K. R. ToppiReddy, B. Saini, and G. Mahajan, "Crime prediction & monitoring framework based on spatial analysis," Proc. Comput. Sci., vol. 132, pp. 696-705, Jan. 2018.

[47]. A. Wolf, T. R. Fanshawe, A. Sariaslan, R. Cornish, H. Larsson, and S. Fazel, "Prediction of violent crime on discharge from secure psychiatric hospitals: A clinical prediction rule (FoVOx)," Eur. Psychiatry, vol. 47, pp. 88-93, Jan. 2018.

[48]. K. B. Sahay, B. Balachander, B. Jagadeesh, G. A. Kumar, R. Kumar, and L. R. Parvathy, "A real time crime scene intelligent video surveillance systems in violence detection framework using deep learning techniques," Comput. Electr. Eng., vol. 103, Oct. 2022, Art. no. 108319.

[49]. P. E. P. Utomo, "Prediction the crime motorcycles of theft using ARIMAXTFM with single input," in Proc. 3rd Int. Conf. Informat. Comput. (ICIC), Oct. 2018, pp. 1-7.

[50]. V. Ingilevich and S. Ivanov, "Crime rate prediction in the urban environment using social factors," Proc. Comput. Sci., vol. 136, pp. 472-478, Jan. 2018.