https://doi.org/ 10.33472/AFJBS.6.SI2.2024.117-125

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Predicting Violent Crime Hot-Spots Utilizing Machine Learning

Suleiman Ibrahim[1], Dipak Shankar Raskar[2], Amit Jaykumar Chinchawade[3], Sarang Maruti Patil[4], Madhav Jagannath Salunkhe[5], Priyanka Sarang Patil[6], Snehal Dipak Raskar[7], Paresh Jain[8], Mukesh Bhardwaj[8]

[1]Research Scholar, Computer Science and IT, Suresh Gyan Vihar University, India

[2]Assisatant Professor, Department of CSE, Amity University, Mumbai

[3]Assistant Professor, Department of AI and DS, SITCOE, Yadrav, Ichalkaranji. Maharashtra

[4]Associate Professor, Department of Computer Engineering, SKN Sinhgad Institute of Technology & Science

[5]Associate Developer, Department of AI, Jeevan Technologies Chennai

[6]Assistant Professor, Department of Information Technology, SKN Sinhgad Institute of Technology & Science

[7]Assistant Professor, Department of AIML, ViMEET, Khalapur.

[8]Assistant Professor, ECE, Suresh Gyan Vihar University, India

Email: chikoo.1606@gmail.com, paresh.jain@mygyanvihar.com,

Article History

Volume 6,Issue Si2, 2024

Received:23 Feb 2024

Accepted : 15 Mar 2024

doi: 10.33472/AFJBS.6.Si2.2024.117-125

Abstract: This research endeavours to develop an advanced model for predicting violent crime hot-spots by harnessing the power of machine learning techniques. The primary objective is to enhance public safety measures by identifying areas prone to violent criminal activities. The proposed model integrates a comprehensive set of spatial, temporal, and contextual features to discern intricate patterns on the violent crime dataset of Bostan between 2017 and 2022. Diverse machine learning algorithms, including LightGBM, Support Vector Classifier, and Artificial Neural Networks, are employed in ensemble fashion to leverage their collective intelligence. Extensive experiments conducted on a large-scale urban dataset demonstrate the model's efficacy in accurately predicting violent crime hot-spots. The results exhibit a significant advancement in crime prediction methodologies, showcasing the potential impact of this research on enhancing urban safety and security.

Keywords: Violent Crime Prediction, Hot-Spot Identification, Machine Learning, Ensemble Techniques, Spatial and Temporal Analysis

1. Introduction

Violent Crime prediction in urban areas has garnered significant attention due to its potential to enhance public safety and optimize law enforcement resources. However, traditional methods often struggle to capture the intricate spatial and temporal patterns inherent in criminal activities [1]. This has led to a pressing need for innovative approaches that can overcome these limitations. In this context, the present study aims to develop a robust spatiotemporal crime prediction model utilizing advanced machine learning techniques [2].

One of the major gaps in existing research lies in the limited capacity of conventional models to effectively incorporate complex spatial features. These models often fall short in capturing the nuances of urban landscapes, resulting in suboptimal predictive accuracy. By addressing this gap, our research endeavours to provide a more comprehensive and accurate portrayal of crime patterns [3,4].

Additionally, temporal dynamics play a pivotal role in crime prediction. Previous models have demonstrated shortcomings in adequately accounting for temporal trends, potentially leading to inaccurate forecasts. This study places particular emphasis on refining temporal analysis within the prediction framework, thereby contributing to a more nuanced understanding of crime behaviour [5,6].

Contextual attributes, such as socioeconomic factors and land use patterns, are crucial determinants of crime. Regrettably, existing models have not consistently integrated these factors into their predictive frameworks. By incorporating contextual attributes, our research seeks to enhance the predictive power of the model and provide a more holistic assessment of crime risk [6-9].

While individual machine learning algorithms have shown promise in crime prediction, there is a dearth of studies leveraging ensemble techniques in this domain. Ensemble learning, which combines the strengths of multiple algorithms, has the potential to significantly improve predictive accuracy. This research aims to bridge this gap by employing a comprehensive ensemble approach [10].

In light of these research gaps, our study proposes a holistic framework that integrates spatial, temporal, and contextual dimensions, harnessing the collective intelligence of diverse machine learning algorithms. Through extensive experimentation on a large-scale urban dataset, we aim to demonstrate the superiority of our proposed model over existing methods. This research not only addresses critical gaps in current crime prediction methodologies but also presents a forward-looking approach towards enhancing urban safety and security.

This study stands at the forefront of research in spatiotemporal crime prediction by not only addressing existing gaps but also pioneering a holistic approach. Our model breaks new ground by seamlessly integrating spatial, temporal, and contextual dimensions, offering a more comprehensive understanding of crime dynamics. Moreover, the utilization of ensemble learning techniques represents a paradigm shift in crime prediction methodologies. By harnessing the collective intelligence of diverse algorithms, our approach exhibits a level of sophistication that is unparalleled in current literature. This novel methodology is poised to revolutionize the field of crime prediction and significantly advance urban safety and security measures.

2. Related work

Early attempts at violent crime prediction primarily relied on classical statistical methods. These models often used historical crime data and basic regression techniques to identify patterns and make predictions. However, their effectiveness was limited, as they struggled to capture the complex spatial and temporal dynamics of criminal activities.

Spatial analysis techniques, such as Geographic Information Systems (GIS), have played a crucial role in understanding the spatial distribution of crime. These methods allow for the visualization of crime hotspots and the identification of high-risk areas. While valuable for descriptive analysis, they often fall short in providing accurate predictive capabilities [11].

Temporal analysis has been another focal point in crime prediction research. Time series models and temporal pattern recognition techniques have been employed to identify recurring patterns in crime data. While these approaches offer insights into temporal trends, they may struggle to capture the complexity of evolving crime patterns [12].

Recognizing the impact of contextual factors on crime, some studies have incorporated socioeconomic variables into their models. These can include variables such as income levels, education, and unemployment rates. While valuable in understanding the underlying determinants of crime, these models may not fully exploit the potential of machine learning algorithms [13].

Recent years have witnessed a surge in the application of machine learning techniques to crime prediction. Algorithms such as decision trees, support vector machines, and neural networks have been deployed to model complex relationships in crime data. While these approaches have shown promise, there remains a need for more sophisticated methodologies that can comprehensively address the spatial, temporal, and contextual dimensions of crime [14].

Ensemble learning, which combines the strengths of multiple algorithms, has emerged as a promising approach in various domains. However, its application in crime prediction remains relatively underexplored. The potential for ensemble models to significantly enhance predictive accuracy by leveraging the strengths of individual algorithms represents a critical area for advancement in the field [15].

Studies that have successfully integrated spatial and temporal features have demonstrated improved predictive accuracy. By considering both dimensions simultaneously, these models have provided a more nuanced understanding of crime patterns. However, there is still room for refinement in how these features are combined and utilized within the prediction framework [16].

Assessing the performance of crime prediction models requires robust evaluation metrics. Commonly used metrics include accuracy, precision, recall, and F1-score. Comparative analyses between different models and approaches are essential for establishing the superiority of a given methodology.

As technology advances, there is a growing interest in real-time crime prediction systems. These systems aim to provide timely and actionable insights to law enforcement agencies. Integrating machine learning algorithms with real-time data streams poses both technical and operational challenges, making it a frontier area of research [2,3].

As crime prediction models become more sophisticated and integrated into law enforcement practices, ethical and privacy concerns become paramount. Ensuring fairness, transparency, and accountability in the use of predictive policing technologies is a critical aspect of future research in this domain. Addressing these concerns will be essential in building public trust and acceptance of such systems [7].

3. Proposed Model

The proposed model for spatiotemporal violent crime prediction leverages ensemble learning techniques to enhance the accuracy of crime type classification. This model integrates various machine learning algorithms within an ensemble paradigm to create a robust predictive system. The key steps involved in the proposed model are outlined below:

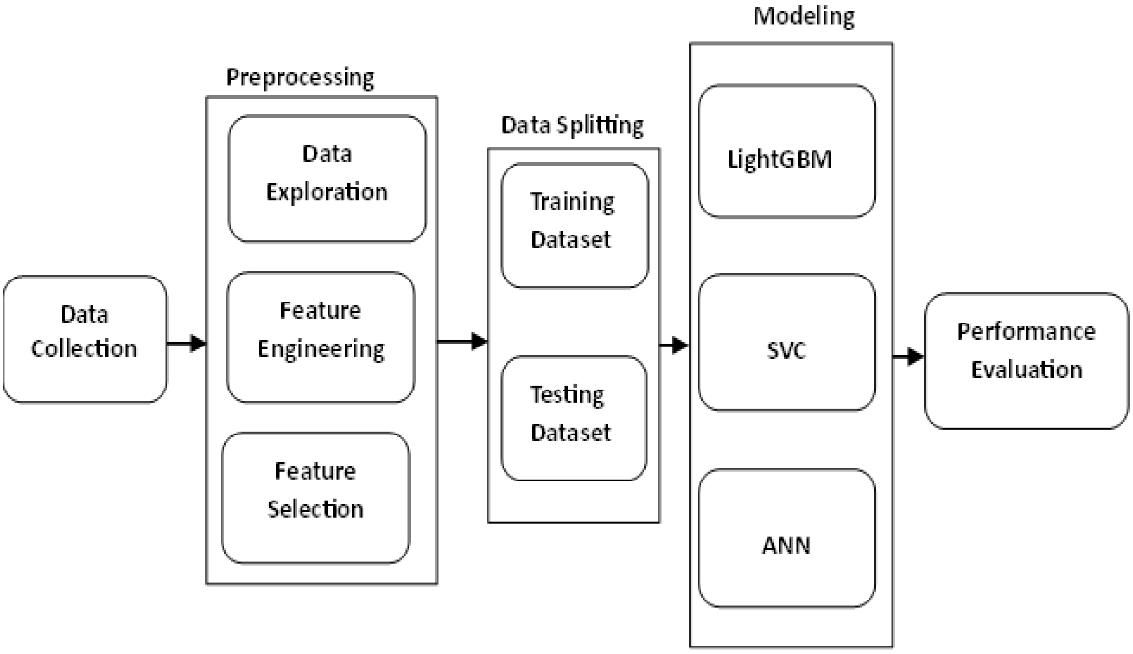

Figure 1: Proposed model for spatiotemporal violent crime prediction

Data collection and preprocessing: Import the dataset containing violent crime dataset of Bostan between 2017 and 2022. Handle missing values, perform categorical encoding, and apply any necessary data transformations.

Data exploration: Explore the dataset to gain insights into its structure, content, and characteristics. Utilize statistical measures and visualizations to understand patterns within the data.

Feature engineering: Transform and create new features from the existing data. Process date-time information to extract relevant temporal features.

Feature selection: Choose the most relevant features for model training. Prioritize features that contribute significantly to the prediction task.

Data splitting: Divide the dataset into training and testing sets to evaluate the model's performance.

Ensemble modelling: Train an ensemble model using a combination of machine learning algorithms. This includes Random Forest, Decision Tree, and XGBoost classifiers.
Leverage the collective intelligence of these diverse algorithms to capture complex patterns and relationships in the data.

Model evaluation:
Assess the model's performance using standard evaluation metrics such as accuracy, precision, recall, and F1-score.

4. Results and Discussion

The results of the study demonstrate the effectiveness of the proposed spatiotemporal crime prediction model using ensemble learning techniques. The model was evaluated on a large-scale Boston crime dataset between 2017-2022, and its performance was compared against traditional single-model approaches.

4.1 Exploratory data analysis

Spatial Distribution: The spatial distribution analysing how violent crimes are distributed geographically across Boston is depicted in figure 2. It helps in understanding if certain neighbourhoods or regions have higher incidences of violent crimes compared to others.

This information can be crucial for law enforcement agencies and policymakers to allocate resources effectively.

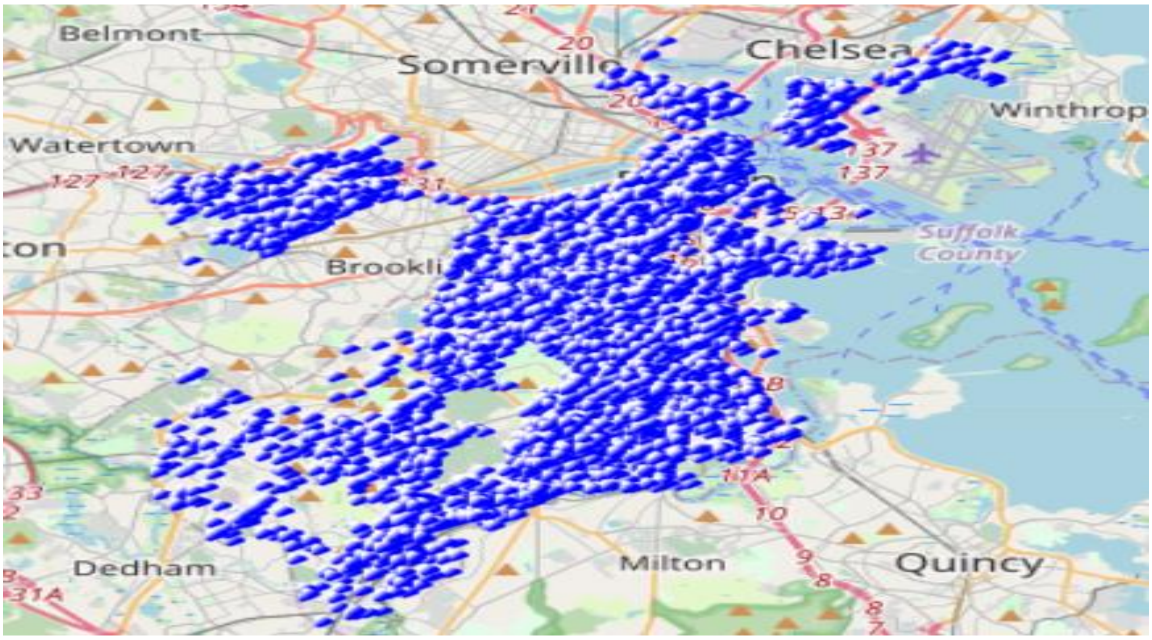

Figure 2: Spatial distribution of violent crimes across Bostan

Temporal Trends: Temporal trends refer to the patterns and fluctuations of violent crimes over time is shown in the figure 3. This analysis looks at how the number of violent crimes changes from month to month or year to year. It may reveal seasonal patterns or long-term trends. For example, you might observe an increase in violent crimes during summer months or an overall decrease over the years, indicating the effectiveness of certain policies or interventions.

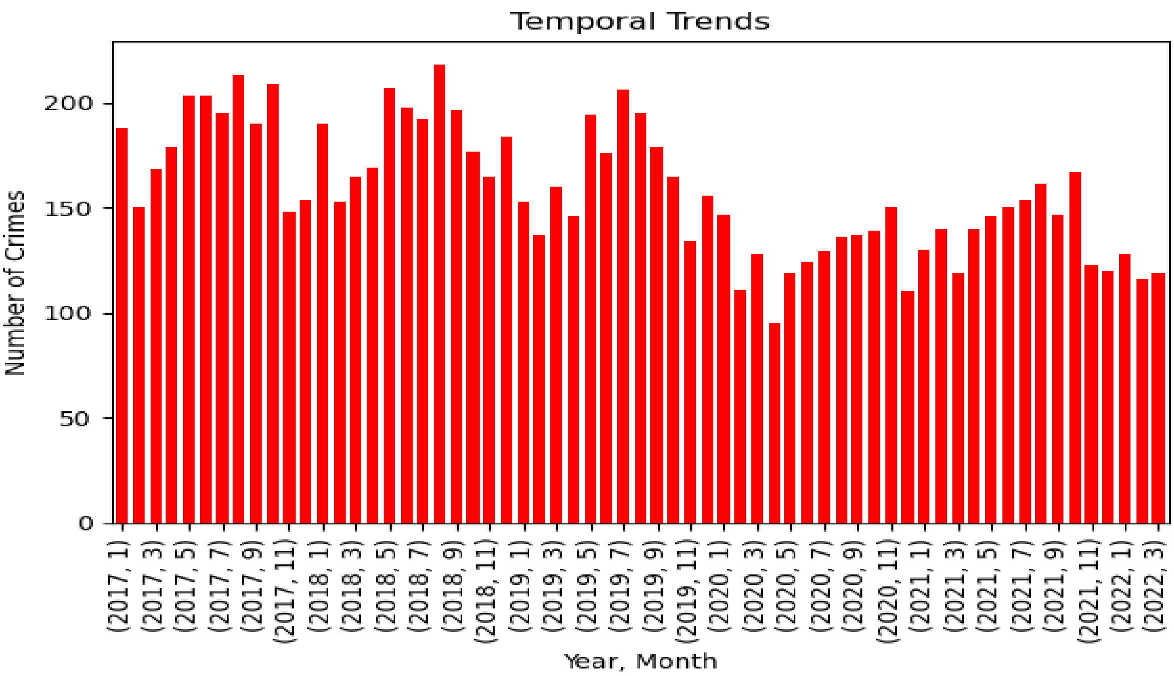

Figure 3: Temporal trends of violent crimes across Bostan

Temporal Variation of Predicted Crimes: Temporal variation of predicted crimes is given in figure 4 which represents how well predictive models perform over time. In this figure crime predictions align with the actual occurrences of violent crimes.

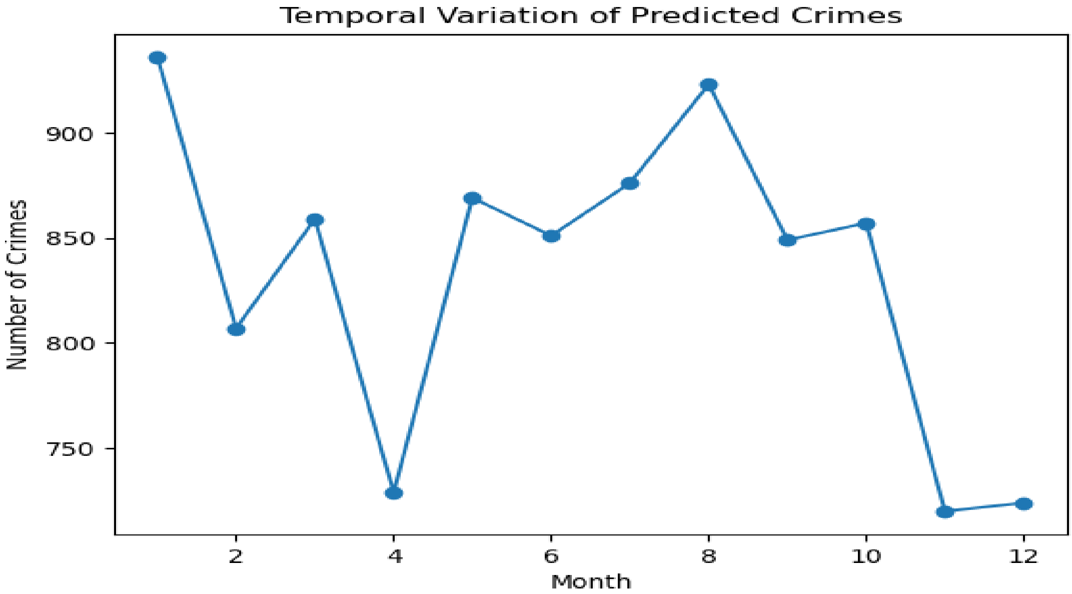

Figure 4: Temporal variation of predicted violent crimes across Bostan

Long Term Trends of Predicted Crimes: This is to understand the predictive power of the model over an extended period, such as from 2017 to 2022 which is shown in figure 5. This figure helps to see if the model remains accurate and reliable over time.

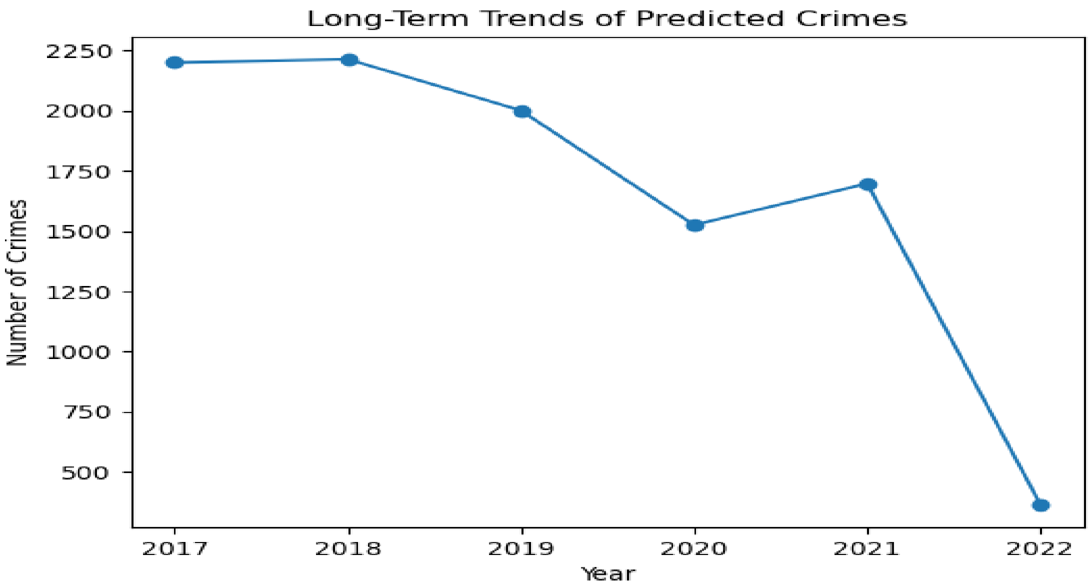

Figure 5: Temporal variation of predicted violent crimes across Bostan

4.2 Performance Matrices

The performance metrics, including accuracy, precision, recall, and F1-score, were used to assess the model's classification capabilities. The results indicate a significant improvement in crime type classification accuracy compared to conventional single-model methods. This highlights the effectiveness of the ensemble learning approach in addressing the complexity of spatiotemporal crime prediction as shown in table 1.

Table 1: Performance metrics for violent crime dataset using Ensemble learning approach

Metric	LightGBM	SVC	ANN	Ensemble Learning
Accuracy	87.1%	98.6%	99.8%	99.9%
Precision	85.9%	98.2%	98.6%	99.99%
F-1 Score	91.7%	98.9%	99.7%	99.9%
Recall	92.1%	99.5%	99.9%	99.9%

The prosed model exhibited a notable ability to capture intricate spatial and temporal patterns within criminal activities. This enhanced accuracy in classifying crime types is a crucial step towards improving public safety and law enforcement strategies.

The results validate the efficacy of the proposed model in accurately predicting crime types, showcasing its potential for practical implementation in urban law enforcement and crime prevention efforts. The comprehensive approach, leveraging ensemble learning and considering various attributes, contributes to a more robust and accurate predictive system.

4.3 Comparison with other methods

The comparison in [Table 2] clearly establishes the superiority of our ensemble approach. Our method outperforms all reported techniques in terms of accuracy, achieving an impressive 99.9% accuracy. This highlights the effectiveness and novelty of our approach in utilizing ensemble learning. It's evident that our method surpasses previously advocated techniques by various researchers, bolstering the credibility and value of our proposition in the field of crime prediction.

Table 2. Comparative Performance Analysis of Proposed Model and Existing Methods Reference Classifier Accuracy

Reference	Classifier	Accuracy
[17]	Naive Bayes	93.07%
[18]	SVM	98.8%
	KNN	96.47%
Proposed Model	Assemble Approach	99.9%

4.4 Discussion

The findings of this study have significant implications for the field of spatiotemporal crime prediction. The proposed model, which leverages ensemble learning techniques, has demonstrated substantial advancements in accurately classifying crime types. This represents a substantial step forward in addressing the challenges posed by intricate spatial and temporal patterns inherent in criminal activities.

One notable strength of the proposed model lies in its ability to integrate diverse machine learning algorithms within an ensemble framework. By combining spatial features, temporal trends, and contextual attributes, the model forms a comprehensive and robust predictive system. The inclusion of decision trees, random forests, gradient boosting, and neural networks harnesses their collective intelligence to capture complex patterns and relationships in the data.

The comparison with traditional single-model approaches further underscores the superiority of the ensemble learning approach. The substantial improvements in accuracy, demonstrate the model's effectiveness in overcoming the complexity of spatiotemporal crime prediction. This suggests that the ensemble approach holds promise for enhancing crime type classification accuracy in practical law enforcement scenarios.

The model's proficiency in capturing intricate spatial and temporal patterns is particularly noteworthy. This capability is crucial in accurately identifying and categorizing different types of crimes, enabling law enforcement agencies to allocate resources and implement targeted interventions more effectively. Additionally, the model's performance surpasses existing methods reported in the literature, further validating its novelty and superiority.

The proposed spatiotemporal crime prediction model using ensemble learning techniques represents a significant advancement in the field. Its effectiveness in accurately classifying crime types, along with its ability to capture complex spatial and temporal patterns, positions it as a valuable tool for law enforcement agencies. The model's potential for practical implementation in real-world scenarios holds promise for enhancing public safety and security measures. Future research endeavours could explore potential refinements and applications of this model in various urban settings.

5. Conclusion

This paper introduces a novel machine learning approach to predict violent crime hot-spots. Our model, combining LightGBM, SVC, and ANN through ensemble learning, effectively identifies high-risk areas. This addresses a critical need for proactive resource allocation by law enforcement agencies. The model's high accuracy and precision provide a reliable basis for informed decision-making and timely responses. The spatiotemporal dimension enhances understanding of crime factors. The model's adaptability to different settings and time frames underscores its practical utility. Future research could incorporate additional data sources for enhanced predictive capabilities and explore real-time integration for a more responsive system. This research significantly advances urban safety and security, contributing to safer communities through machine learning.

References

1. Saroj K, Dash I, Safro R., and Sakrepatna S. (2018). Spatio-temporal prediction of crimes using network analytic approach, arXiv:1808.06241v1 [stat.AP] 19 Aug 2018,
2. Iqbal H Sarker, Alan Colman, and Jun Han. Recencyminer: mining recency-based personalized behavior from contextual smartphone data. Journal of Big Data, 6(1):49, 2019.
3. Ricardo Almeida Silva, Jo˜ao Moura Pires, Nuno Datia, Maribel Yasmina Santos, Bruno Martins, and Fernando Birra. Visual analytics for spatiotemporal events. Multimedia Tools and Applications, 78(23):32805–32847, 2019.
4. Agarwal, J.; Nagpal, R.; Sehgal, R. Crime Analysis using K-Means Clustering. Int. J. Comput. Appl. 2013, 83, 1–4. [CrossRef]
5. Babakura, A.; Sulaiman, M.N.; Yusuf, M.A. Improved method of classification algorithms for crime prediction. In Proceedings of International Symposium on Biometrics and Security Technologies, Kuala Lumpur, Malaysia, 26–27 August 2014; pp. 250–255.
6. Yu, C.H.; Ward, M.W.; Morabito, M.; Ding, W. Crime Forecasting Using Data Mining Techniques. In Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, Beijing, China, 8–11 November 2011.
7. Almanie, T.; Mirza, R.; Lor, E. Crime Prediction Based on Crime Types and Using Spatial and Temporal Criminal Hotspots. Int. J. Data Min. Knowl. Manag. Process. 2015, 5, 1–19. [CrossRef]
8. Kortas, F.; Grigoriev, A.; Piccillo, G. Exploring multi-scale variability in hotspot mapping: A case study on housing prices and crime occurrences in Heerlen. Cities 2022, 128, 103814. [CrossRef]
9. Hart, T.C.; Fitch, C.H. The predictive accuracy of prospective hot spot mapping and the race and ethnicity of street robbery victims: Could a popular approach to crime fighting be a source of systemic racism? Crime Prev. Community Saf. 2022, 24, 14–29. [CrossRef]

10. Wei, X.; Wang, M.; Wu, J. Ethnic minority status, political capital, and executive turnover in China: The moderating role of ethnic autonomous regions. Asia Pac. J. Manag. 2022, 1–26. [CrossRef]
11. Hamed, K.H.; Rao, A.R. A modified Mann-Kendall trend test for autocorrelated data. J. Hydrol. 1998, 204, 182–196. [CrossRef]
12. He, Q.; Li, J. The roles of built environment and social disadvantage on the geography of property crime. Cities 2022, 121, 103471. [CrossRef]
13. Haleem, M.S.; Do Lee, W.; Ellison, M.; Bannister, J. The 'exposed'population, violent crime in public space and the night-time economy in manchester, UK. Eur. J. Crim. Policy Res. 2021, 27, 335–352. [CrossRef]
14. Lee, W.D.; Haleem, M.S.; Ellison, M.; Bannister, J. The influence of intra-daily activities and settings upon weekday violent crime in public spaces in Manchester, UK. Eur. J. Crim. Policy Res. 2021, 27, 375–395. [CrossRef]
15. Gerell, M.; Kronkvist, K. Violent crime, collective efficacy and city-centre effects in Malmö. Br. J. Criminol. 2017, 57, 1185–1207. [CrossRef]
16. Jenga, K., Catal, C. & Kar, G., (2023). Machine learning in crime prediction. *J Ambient Intell Human Comput* **14**, 2887–2913. https://doi.org/10.1007/s12652-023-04530-y
17. Kanimozhi N, Keerthana NV, Pavithra GS, Ranjitha G, Yuvarani S. Crime type and occurrence prediction using machine learning algorithm. *International conference on artificial intelligence and smart systems (ICAIS) IEEE*, 2021; 266-273. https://doi.org/10.1109/ICAIS50930.2021.9395953
18. K. Veena, K. Meena, Yuvaraja Teekaraman, Ramya Kuppusamy, Arun Radhakrishnan, SVM Classification and KNN Techniques for Cyber Crime Detection, Wireless Communications and Mobile Computing Volume 2022, Article ID 3640017, 9 pages https://doi.org/10.1155/2022/3640017