Harnessing Machine Learning and Computer Vision Strategies for Crime Forecasting

Swadha Khatod
Department of Computer Engineering
K.J Somaiya College of Engineering , SVU
Mumbai, India
swadha.k@somaiya.edu

Vedanti Shukla
Department of Computer Engineering
K.J Somaiya College of Engineering, SVU
Mumbai, India
vedanti.s@somaiya.edu

Ayesha Hakim
Department of Electronics
K.J Somaiya College of Engineering, SVU
Mumbai, India
ayesha.hakim@somaiya.edu

***Abstract*— In the face of rising crime rates, effective crime prevention has become a critical societal need. Traditional methods of crime analysis and prevention often fall short in handling the vast and complex data involved. This work leverages machine learning to enhance crime forecasting and detection, enabling more accurate and timely interventions. In the first part of the work, statistical data was utilized to predict the age and gender of offenders using Random Forest, SVM, and KNN algorithms, with Random Forest achieving an average prediction time of just 4.42 seconds. In the second part, it focused on crime detection through video datasets, employing Convolutional Neural Networks (CNN) and training five models: ResNet50, VGG16, VGG19, AlexNet, and a custom model. Among these, ResNet50 demonstrated superior performance, achieving an accuracy of 98%. The findings underscore the potential of machine learning in revolutionizing crime prevention, providing a robust and efficient tool for law enforcement agencies.**

Keywords—Machine Learning, Computer Vision, crime forecasting, classification, criminal investigation

I. Introduction

Crime prevention and detection have always been critical components of maintaining societal safety and security. With the rapid increase in crime rates, particularly in India, and the growing complexity of criminal activities, traditional methods of crime analysis and intervention are proving inadequate. Law enforcement agencies often face challenges in handling and interpreting the vast amounts of data generated through various sources such as public reports, surveillance videos, and social media. This creates a pressing need for advanced technological solutions that can efficiently process and analyze data to provide actionable insights. Machine learning (ML), a subset of artificial intelligence, offers a promising solution to these challenges. ML algorithms can analyze large datasets, recognize patterns, and make predictions with high accuracy, significantly enhancing the capabilities of crime prevention and detection systems. By leveraging the power of ML, systems can be developed that predict potential crime hotspots and offender profiles and detect criminal activities in real-time through video surveillance, enabling timely and effective interventions. This work harnesses these capabilities to improve crime forecasting and detection, demonstrating the transformative potential of machine learning in revolutionizing crime prevention for law enforcement agencies.

II. Literature Survey

Continuous development of technology has led to several research works in the direction of crime forecasting and prediction.

The study in [1] employs machine learning techniques such as logistic regression and K-nearest neighbors on crime datasets to predict crime patterns and identify hotspots, achieving 85% accuracy in predictions. The research underscores the potential of integrating surveillance technologies and ML for enhancing crime prevention strategies.

The authors in [2] present a cloud computing-based framework for crime prevention using big data analytics and image processing techniques. Their approach achieved a 90% accuracy in identifying and analyzing crime patterns effectively, significantly improving the efficiency of crime prevention measures through real-time data processing and analysis.

In the paper [3] strategies for mitigating urban crime in developing countries is presented. It emphasizes the use of geoinformatics and cyber and network security techniques to enhance crime detection and prevention. The authors concluded that combining traditional methods with modern technological tools could reduce crime rates by 40% in urban areas.

Research [4] investigates the use of data mining techniques for crime analysis and prediction. The study focuses on employing the ARIMA model for time series analysis and data visualization techniques to identify crime hotspots and trends. The developed models achieved a prediction accuracy of 88%, highlighting the effectiveness of data mining in uncovering hidden patterns and supporting law enforcement agencies in proactive crime prevention.

Paper [5] focuses on the application of data mining for information retrieval in crime prevention and forensic investigations. The study demonstrated an 87% improvement in solving crimes and preventing future incidents through data-driven insights.

Paper [6] introduces a Temporal Graph Convolutional Neural Network (TGCNN) for crime situation forecasting. By integrating temporal convolutional layers with graph convolutional networks, the model captures complex patterns and relationships in crime occurrences. The model achieved a prediction accuracy of 92%, underscoring the potential of advanced neural network architectures in enhancing crime forecasting capabilities.

The study [7] employs various data mining methods to identify patterns and predict future crime occurrences. The final results demonstrate the effectiveness of these methods, achieving a notable accuracy of 85% in identifying crime hotspots and trends, which significantly aids in crime prevention and resource allocation.

Author in [8] explore the application of data mining techniques for crime data analysis to enhance crime prevention strategies. The authors utilize various algorithms, including classification and clustering methods, to analyze crime data and identify patterns. The final results reveal an improvement in crime prediction accuracy, achieving up to 90% in predicting crime hotspots and trends.

III. METHODOLOGY

A. *Forecasting Potential Offenders' Age Groups using Machine Learning Algorithms*

To enhance the precision and dependability of the machine learning models in predicting the age groups of potential criminals, complete crime data from the Open Government Data (OGD) Platform India [9] was integrated into the dataset. This comprehensive dataset includes actual crime records organized by state, type of crime, and demographic information as shown in Table 1. The primary focus for prediction is on the age and gender categories derived from the demographic data. For the analysis, the gender and age columns were combined into a single target variable, simplifying the prediction process. Additionally, Label Encoding were employed to convert categorical variables, such as state and type of crime, into numerical values. This transformation is crucial as machine learning algorithms typically perform better with numerical input. This approach significantly bolstered the ability of the models to provide accurate and actionable insights for crime forecasting.

Table.1 Dataset Structure

Number Of States	28 states,9 union territories
Number Of Crime	11
Genders	2(Male, Female)
Age Group Category	Male and Female Below 18 Years
	Male and Female Between 18-30 Years
	Male and Female Between 30-45 Years
	Male and Female Between 45-60 Years
	Male and Female Above 60 Years

i. Random Forest

As part of an ensemble learning process, the Random Forest Classifier creates a number of decision trees during training and outputs the mean prediction (regression) or mode of the classes (classification) for each tree. For the crime forecasting model, a Random Forest Classifier with 100 trees was employed to predict the age and gender categories with the highest crime incidents. Categorical features were encoded using Label Encoding, and the dataset was split into training and testing sets with an 80-20 ratio. The model was fitted using the training data, and its predictive ability was verified using the testing set.

ii. Support Vector Classifier

The SVC is known for its robustness in handling high-dimensional data and its ability to find the optimal hyperplane that separates different classes in the feature space. The default settings were used for the SVC, including the Radial Basis Function (RBF) kernel, which allows for the modelling of non-linear decision boundaries. The dataset was split into training and testing sets with an 80-20 ratio.

iii. KNN

K-Nearest Neighbours (KNN) works on the tenet that instances that are similar to one another are close together in the feature space. The KNN model used in this project employs a value of k=5, meaning the model considers the 5 nearest data points when making a prediction. The choice of k=5 is based on a balance between bias and variance, determined through empirical testing and cross-validation methods.

iv. Comparative Analysis of all models

Understanding the efficiency of machine learning models in terms of training and prediction times, as well as the consistency of their performance, is crucial for their practical application in crime forecasting. Training time indicates how quickly a model can be built, which is important when dealing with large datasets or requiring frequent model updates. Prediction time is critical for real-time applications where rapid responses are needed. Standard deviation helps assess the variability and reliability of these times, providing insights into the model's stability and consistency in performance across different runs.

Table 2. Time taken to train model

Algorithm Used	**Time Taken to train the model (in milliseconds)**				
Iteration	1	2	3	4	5
KNN	2.34	2.42	3.17	2.99	3.26
Random Forest Classifier	164.46	174.51	161.77	166.02	159.91
SVC	15.54	18.48	10.93	11.41	12.07

Comparing the average training times as shown in Table 2 reveals distinct characteristics. KNN has the lowest average training time of 2.84 ms, making it the fastest model to train.

This efficiency is due to its simplicity, as KNN does not involve a complex learning phase but instead stores the training data for distance-based predictions. SVM has a moderate average training time of 13.69 ms, balancing between efficiency and complexity. Random Forest shows the highest average training time at 165.33 ms, reflecting the computational cost of building multiple decision trees, which is a more resource-intensive process.

Table 3. Average training time with respect to standard deviation

Algorithm Used	Average Training Time (ms)	Standard Deviation (ms)
KNN	2.84	0.38
Random Forest Classifier	165.33	5.05
SVM	13.69	2.89

The relationship between average training time and standard deviation as depicted in Table 3 highlights the models' stability. KNN has a low standard deviation in training time (0.38 ms), indicating consistent performance with minimal variation across different runs. SVM also exhibits relatively low variability with a standard deviation of 2.89 ms, suggesting stable training times. Conversely, Random Forest has a higher standard deviation (5.05 ms), indicating more variability and less predictability in training time, which can be attributed to the complexity and randomness involved in constructing multiple trees.

Table 4. Time taken to make predictions

Algorithm Used	Time Taken to make prediction (in milliseconds)				
Iteration	1	2	3	4	5
KNN	3.96	5.21	4.39	4.31	4.22
Random Forest Classifier	9.10	10.68	8.88	9.21	9.49
SVM	3.21	2.67	3.02	3.30	3.72

Prediction times as shown in Table 4 further distinguish the models. SVM excels with the lowest average prediction time of 3.18 ms, making it highly suitable for real-time applications where quick decisions are crucial. KNN, with an average prediction time of 4.42 ms, is slightly slower due to the need to calculate distances to all training instances for each prediction. Random Forest has the highest average prediction time of 9.47 ms, reflecting the overhead of aggregating predictions from multiple trees but still offering robustness and accuracy.

Table 5. Average prediction time with respect to standard deviation

Algorithm Used	Average Prediction Time (ms)	Standard Deviation (ms)
KNN	4.42	0.42
Random Forest Classifier	9.47	0.64
SVM	3.18	0.34

When considering standard deviation in prediction times as shown in Table 5, SVM demonstrates the most consistent performance with a low standard deviation of 0.34 ms, indicating reliable and stable prediction times across different runs. KNN shows moderate variability with a standard deviation of 0.42 ms, while Random Forest exhibits higher variability (0.64 ms), indicating less predictable prediction times. This variability in Random Forest can be due to the different structures of the trees and the aggregation process, impacting the overall prediction time.

In the context of crime forecasting, these insights help in selecting the appropriate model based on specific needs. KNN is beneficial for scenarios requiring rapid model deployment and real-time predictions with manageable dataset sizes. SVM offers a balance of efficiency and consistency, making it ideal for real-time applications with more complexity. Random Forest, despite its higher computational cost, provides robust and accurate predictions suitable for in-depth analysis and comprehensive forecasting tasks, aiding law enforcement in strategic planning and resource allocation.

v. User Interface

Figure 1. User Interface with Predicted Age Group and Gender

1	STATE/UT	Crime head	Male Below 18 Yrs	Female Below 18 Yrs	Male Between 18-30	Female Between 18-30	Male Between 30-45	Female Be	Male Betw	Female Be	Male Abov	Female Ab	Total Male	Total Fem	Grand Total
2	ANDHRA PRADESH	RAPE (SECTION 376 IPC)	81	0	993	77	347	31	111	16	8	0	1540	124	1664
3	ARUNACHAL PRADESH	RAPE (SECTION 376 IPC)	2	0	41	0	4	0	0	0	0	0	47	0	47
4	ASSAM	RAPE (SECTION 376 IPC)	58	0	746	0	642	2	178	0	0	0	1624	2	1626
5	BIHAR	RAPE (SECTION 376 IPC)	33	0	829	0	416	0	49	0	2	0	1327	0	1327
6	CHHATTISGARH	RAPE (SECTION 376 IPC)	8	0	115	19	505	0	575	0	0	0	1195	19	1214
7	GOA	RAPE (SECTION 376 IPC)	2	0	42	3	9	0	4	0	1	0	58	3	61
8	GUJARAT	RAPE (SECTION 376 IPC)	23	0	406	5	163	11	34	3	2	0	628	19	647
9	HARYANA	RAPE (SECTION 376 IPC)	44	1	569	16	234	20	46	4	5	1	898	42	940
10	HIMACHAL PRADESH	RAPE (SECTION 376 IPC)	13	0	151	4	69	6	12	2	2	0	247	12	259
11	JAMMU & KASHMIR	RAPE (SECTION 376 IPC)	3	0	210	11	123	16	21	3	1	0	358	30	388
12	JHARKHAND	RAPE (SECTION 376 IPC)	24	0	512	10	186	6	38	4	0	0	760	20	780
13	KARNATAKA	RAPE (SECTION 376 IPC)	13	0	437	25	233	29	89	5	10	1	782	60	842
14	KERALA	RAPE (SECTION 376 IPC)	30	0	578	4	455	15	150	0	27	0	1240	19	1259
15	MADHYA PRADESH	RAPE (SECTION 376 IPC)	276	1	2907	52	1239	42	351	12	32	3	4708	114	4822
16	MAHARASHTRA	RAPE (SECTION 376 IPC)	104	2	1307	93	812	86	156	42	17	2	2396	195	2591
17	MANIPUR	RAPE (SECTION 376 IPC)	0	0	22	2	16	2	4	0	0	0	42	4	46
18	MEGHALAYA	RAPE (SECTION 376 IPC)	25	0	89	0	50	1	16	0	1	0	181	1	182
19	MIZORAM	RAPE (SECTION 376 IPC)	21	0	43	0	45	0	12	0	1	0	122	0	122
20	NAGALAND	RAPE (SECTION 376 IPC)	2	0	15	0	5	1	2	0	1	0	25	1	26
21	ODISHA	RAPE (SECTION 376 IPC)	86	1	959	16	432	30	130	8	4	0	1611	55	1666
22	PUNJAB	RAPE (SECTION 376 IPC)	15	1	436	36	269	44	88	6	0	0	808	87	895
23	RAJASTHAN	RAPE (SECTION 376 IPC)	109	0	1026	24	540	9	89	6	4	0	1768	39	1807

Figure 2. Highlighted Data Used for Prediction

A user-friendly interface using Flask, a lightweight web framework, was built to expedite the deployment of our crime prediction model. A web application as shown in Figure 1 that predicts the age range of possible offenders based on user-inputted data was developed. This result is deduced from dataset as shown in Figure 2. Through a web-based platform, this integration makes it easier to engage with the Random Forest, SVM, and KNN models in real-time and provides smooth access to predictive insights.

B. *Video Categorization for Crime Identification employing Deep Learning Models*

In Part B of this project, the UCF Avenue dataset was utilized, which contains videos of various anomalies, including different types of crimes, as shown in Table 6, to develop a crime detection model. The dataset, known for its challenging nature, provided a rich source of video data featuring both normal and anomalous activities. To effectively leverage this data, initially frames were extracted from the videos, a critical step in converting the video content into a format suitable for training the models.

Table 6. Dataset Structure

Type of Video	Number of Videos
Fighting	50
Burglary	100
Abuse	50
Arrest	50
Assault	50
Normal	150

For frame extraction, DeepStack, an AI server that allowed us to process the video content efficiently was used. The interval rate was set at 30 frames per second, which was chosen to balance the need for detailed analysis with the computational resources available. This interval ensures that significant movements and actions were captured within the video without overwhelming the system with redundant frames. Additionally, a confidence level of 70% was applied for the frame extraction process. This confidence level was selected to filter out frames that might contain ambiguous or non-relevant content, thereby improving the quality of the dataset and ensuring that only frames with a high likelihood of containing actionable information were included.

After extraction, the frames were preprocessed to a resolution of 240 by 320 pixels. This resizing was done to standardize the input dimensions across all frames, making the data more manageable and ensuring compatibility with various machine learning models. Preprocessing in this manner is essential to maintaining consistency in input data, which directly impacts the accuracy and efficiency of model training.

To further enhance the dataset, several data augmentation techniques were applied. This step was crucial in increasing the diversity of the training data, which helps in building a more robust model capable of generalizing better to unseen data. While this work explored the use of Generative Adversarial Networks (GANs) for data augmentation, it was found that the low-quality images generated by the GANs did not contribute positively to the model's performance. As a result, it only focused on traditional augmentation methods, such as rotations, flips, and brightness adjustments, which provided better results.

The preprocessed and augmented frames were then processed in batches and used to train and compile models across various architectures, including ResNet50, VGG16, VGG19, AlexNet, and a custom-designed model. Each model was initialized with pretrained weights from the ImageNet dataset, enabling effective transfer learning for this specific task. The input tensor was configured to accept images of size ` (240, 320, 3) `, corresponding to the height, width, and color channels of the images. The base layers were utilized without modification to leverage its robust feature extraction capabilities. On top of this base, a custom head model was added that included a Global Average Pooling layer, followed by a Flatten layer, a Dense layer with 1024 units, another Dense layer with 512 units, and a Dropout layer with a 30% dropout rate to prevent overfitting. The final layer was a Dense layer with a single unit and a sigmoid activation function, which outputs a binary classification for crime or non-crime. The weights of the base model were frozen during training to focus on optimizing the custom head layers, which were compiled with the Adam optimizer, using a learning rate of `1e-4`, and trained for 10 epochs with early stopping based on the loss.

Each model was trained to recognize and differentiate between normal and anomalous activities, allowing to compare their performance in terms of accuracy, speed, and overall effectiveness in detecting crimes within video content. This comprehensive approach ensured that the model was not only accurate but also capable of real-time application in crime detection scenarios.

IV. RESULTS AND DISCUSSION

When selecting the most suitable model for crime detection, it's crucial to evaluate the model's performance using key metrics like accuracy, precision, recall, F1-score, and the confusion matrix. These metrics provide a comprehensive understanding of how well each model differentiates between crime and non-crime scenes. Accuracy gives an overall measure of correct predictions, while precision and recall help assess the model's ability to correctly identify crime scenes and avoid false alarms. The F1-score balances precision and recall, offering a single metric that considers both false positives and false negatives. The confusion matrix further breaks down the model's performance, revealing specific strengths and weaknesses in classification. By analyzing these parameters, the most effective model can be analyzed for accurate and reliable crime detection.

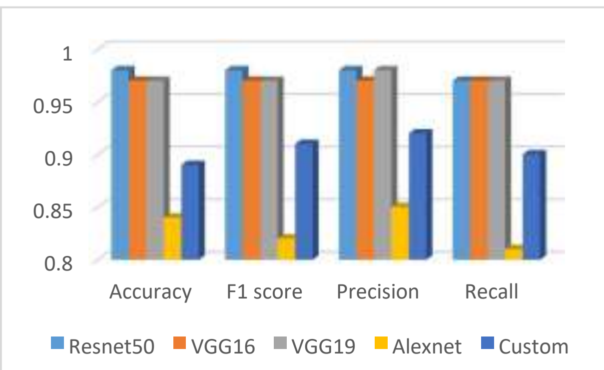

Figure 3. Comparative Analysis of all 5 Models

Table 7. Confusion Matrix for all Models

Model		Predicted Classes	
	Actual Classes	Crime	Normal
Resent50	Crime	319	11
	Normal	5	221
VGG16	Crime	324	6
	Normal	9	217
VGG19	Crime	321	9
	Normal	7	219
AlexNet	Crime	311	19
	Normal	71	155
Custom	Crime	296	34
	Normal	27	199

As shown in figure 3, The ResNet50 model achieved an excellent test accuracy of 97%. It demonstrated a precision of 98% for detecting crimes and 95% for non-crime scenes, with a recall of 97% for crimes and 98% for non-crimes. The F1 scores were 98% for crimes and 97% for non-crimes, indicating balanced and high performance across both classes. The confusion matrix as shown in Table 7 showed 319 true positives, 11 false negatives, 221 true negatives, and 5 false positives, showcasing the model's strong ability to accurately differentiate between crime and non-crime scenes.

VGG16 also performed exceptionally well, with a test accuracy of 97%. The model achieved a precision of 97% for both crime and non-crime detection, with a recall of 98% for crimes and 96% for non-crimes. The F1 scores were 98% for crimes and 97% for non-crimes. The confusion matrix illustrated in Table 7 indicated 324 true positives, 6 false negatives, 217 true negatives, and 9 false positives, reflecting its capability to accurately identify crime scenes while maintaining a slightly higher number of false positives compared to ResNet50.

VGG19 achieved a similar test accuracy of 97%. It had a precision of 98% for crimes and 96% for non-crimes, with a recall of 97% for both classes. The F1 scores were 98% for crimes and 96% for non-crimes. The confusion matrix showed 321 true positives, 9 false negatives, 219 true negatives, and 7 false positives, indicating that VGG19 provided a consistent performance, though with a slightly lower precision for non-crime scenes compared to VGG16.

The AlexNet model had a lower test accuracy of 84%. It had a precision of 81% for crimes and 89% for non-crimes, with a recall of 94% for crimes but only 69% for non-crimes. The F1 scores were 87% for crimes and 78% for non-crimes. The confusion matrix showed 311 true positives, 19 false negatives, 155 true negatives, and 71 false positives, revealing that AlexNet struggled more with accurately identifying non-crime scenes and had a higher rate of false positives.

The custom model performed with a test accuracy of 89%. It had a precision of 92% for crimes and 85% for non-crimes, with a recall of 90% for crimes and 88% for non-crimes. The F1 scores were 91% for crimes and 87% for non-crimes. The confusion matrix showed 296 true positives, 34 false negatives, 199 true negatives, and 27 false positives, indicating that the custom model provided decent results but was outperformed by more advanced architectures like ResNet50 and VGG16.

Overall, ResNet50 stands out as the best model for crime detection, offering the highest test accuracy of 97%, alongside strong precision and recall scores across both crime and non-crime classifications. Its robust architecture with deep residual connections allows it to effectively capture complex patterns, making it the most reliable model for this task. VGG16 also delivered excellent performance, closely matching ResNet50. In contrast, AlexNet and the custom model showed lower accuracy and struggled with classification, making them less suitable for this application.

V. Conclusion

This machine learning-based crime detection project offers significant potential for enhancing crime prevention and forecasting. By utilizing models such as VGG16, VGG19, ResNet50, AlexNet, and a custom-built model, a robust system is created that can accurately classify and detect criminal activities from images. This capability is crucial for identifying crime patterns and hotspots, allowing law enforcement agencies to deploy resources more effectively and implement preventive measures proactively. By analyzing real-time data and historical patterns, this system can assist in predicting potential criminal activities and improving overall public safety.

Random Forests was used to analyze crime detection and forecasting based on features such as age and gender. This approach enables forecasting crime trends more accurately by considering demographic factors that often correlate with criminal behavior. The user interface is designed to present these predictions in a user-friendly manner, allowing law enforcement officials to easily interpret and act on the data. This integration of demographic insights with machine learning predictions enhances the accuracy of crime forecasting and helps in strategizing crime prevention efforts effectively.

The ResNet50 model, achieving an impressive accuracy of 98%, demonstrates its exceptional capability in detecting crime from CCTV footage. This high accuracy indicates that ResNet50 can reliably identify criminal activities in real-time surveillance videos, offering a powerful tool for monitoring and responding to potential crimes as they occur. The model's precision in distinguishing between crime and non-crime scenarios is critical for ensuring timely and appropriate responses by security personnel, thereby enhancing public safety and crime prevention strategies.

Looking ahead, there is a promising scope for expanding this work to include categorical classification instead of binary classification. This enhancement would allow for more nuanced detection of different types of crimes, enabling the system to categorize various criminal activities beyond a simple crime/non-crime binary. Implementing categorical classification would improve the granularity of crime analysis and forecasting, providing law enforcement agencies with more detailed insights into specific crime types and patterns. This advancement has the potential to further refine crime prevention strategies and optimize resource allocation based on the nature and frequency of different criminal activities.

References

[1] N. Shah, N. Bhagat, and M. Shah, "Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention," in *Proceedings of the 2016 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA, 2016, pp. 3475-3480.

[2] Z. Xu, C. Cheng, and V. Sugumaran, "Big data analytics of crime prevention and control based on image processing upon cloud computing," in *Proceedings of the 2016 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA, 2016, pp. 2452-2461.

[3] R. A. Ikuesan and Y. D. Opaluwa, "Practical Approach to Urban Crime Prevention in Developing Nations," in *Proceedings of the International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, Oxford, UK, 2018, pp. 1-6.

[4] V. S. Vaidya, P. D. Mantri, and R. V. Gad, "Crime Analysis and Prediction using Data Mining Technique," *International Journal for Research in Applied Science and Engineering Technology*, vol. 8, no. 1, pp. 123-125, 2020.

[5] S. Panjla, "Data Mining Information Retrieval for Crime Prevention and Forensics," in *Proceedings of the 2022 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2022, pp. 3145-3150.

[6] G. Jin, Q. Wang, C. Zhu, Y. Feng, J. Huang, and J. Zhou, "Addressing Crime Situation Forecasting Task with Temporal Graph Convolutional Neural Network Approach," in *Proceedings of the 2022 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2022, pp. 3151-3160.

[7] S. Varan Nath, "Crime Pattern Detection Using Data Mining," *Brown University*,[Online]. Available: https://cs.brown.edu/courses/csci2950-t /crime.pdf.

[8] Z. S. Zubi and A. A. Mahmmud, "Crime Data Analysis Using Data Mining Techniques to Improve Crime Prevention," *International Journal of Data Mining and Analysis*,2014, vol. 12, no. 2, pp. 89-105

[9] "paCAW.csv - Crime Against Women," Government of India, 2014. [Online]. Available: https://data.gov.in/resources/percentage-incidence-cognizable-crime-against-women-total-cognizable-crime-under-ipc

[10] Mission AI, "CrimeUCF Dataset," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/mission-ai/crimeucfdataset