International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0177
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1180 - 1184

A Spatiotemporal and NLP-Based Framework for Crime Prediction and Prevention

Alisha Jabeen[1], Ansif Azeez[2], Abhinav S[3], Jenas Renjan Parakkal[4], Adhwaith K G[5], Anandha Krishnan[6]
[1]Assistant Professor, Department of Computer Science, Yenepoya Deemed to be University, India.
[2,3,5]MSc Data science with Big Data Analytics, YIASCM, Bangalore, Karnataka, India.
[4]MCA AI and ML with in Data Science, YIASCM, Bangalore, Karnataka, India.
[6]MSc Cybersecurity and Ethical Hacking with Cyber forensics, YIASCM, Bangalore, Karnataka, India.
Emails: *b.ramakrishna.blr@yenepoya.edu.in[1], 46666@yenepoya.edu.in[2], 46656@yenepoya.edu.in[3], 47321@yenepoya.edu.in[4], 47397@yenepoya.edu.in[5], 46667@yenepoya.edu.in[6]*

Abstract
The way police prevent crime keeps on changing. Earlier it was about reacting to things after a crime has occurred and it relied on the past. This led to slower responses and poor resource utilization. But now they use evidence to solve and stop crimes before they happen. This study uses techniques like Machine Learning, Geographic Information Systems, and Predictive Modeling that will help us to find high-crime areas and patterns that occur frequently by using resources effectively. Spatiotemporal Analysis and Natural Language Processing (NLP) are the main two ways to examine it. This will help with finding out trends and frequency of crime occurrence. This research accords a foundation that combines Machine Learning with Spatiotemporal and NLP-driven analysis to strengthen crime forecasting. Moreover, NLP techniques can also draw out useful pieces of information from unstructured text data like police reports, witness statements, and posts on social media. This research accords a foundation that combines Machine Learning with Spatiotemporal and NLP-driven analysis to strengthen crime forecasting. Overall, the study concludes that Ethical Incorporation of data science can significantly reinforce public safety while maintaining ethical values and social norms.

Keywords: *Crime Prevention, Data Science in Law Enforcement, Machine Learning, Geographic Information Systems, Behavioral patterns, Predictive Model, Spatiotemporal Analysis, Natural Language Processing, Practical Feasibility, Ethical Incorporation.*

1. Introduction
A society that works well urges for good communication and public safety. As cities outscape quickly and digital data grows abundantly, it has become difficult for police to stop crime. The traditional way of policing mostly uses crime records from the past, manual analysis, and responses that happen after crimes happen. These methods can help us understand how crime has happened in the past, but they don't always accurately tell us how crimes will happen in future or help us intervene fastly. Because of this, there is a growing need for smart ways to handle systems that help the police make decisions before they are committed or occurred. New technologies in data science, machine learning, and computers have made it easier to predict and stop crime to an extent. Researchers and police use predictive analytics to search for trends in large datasets, discover dimensions that are probably suspicious, and predict crimes that might occur in the future. Spatiotemporal Analysis and Natural Language Processing (NLP) are the main two ways to examine it. Spatiotemporal analysis distinguishes how crime is scattered over different places and times. This will help us find high-crime areas and patterns that occur frequently. This aids the police in figuring out circumstances where crimes are probable

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0177
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1180 - 1184

to occur. Moreover, NLP techniques can also draw out useful pieces of information from unstructured text data like police reports, witness statements, and posts on social media. This research accords a foundation that combines Machine Learning with Spatiotemporal and NLP-driven analysis to strengthen crime forecasting [1]

1.1. Types of Crime Data Sources

The different types of Crime Data Sources based on where they are collected from are:

- **Historical Crime Records:** The data regarding old or past criminal activities. This will also include when, where and how the crime took place and the seriousness of it.
- **Geographical Data:** Location based records will be included in this type. That is, spatial data such as city maps, neighbourhoods etc for finding out high-crime areas.
- **Socioeconomic Data:** Details about the data are collected in the form of socioeconomic and infrastructural characteristics. It is recorded using population rate, income distribution and other factors.
- **Textual Crime Reports:** Written or typed data like FIR, news stories, social media posts, reports from police etc will be a part of this type [2].

2. Methodology

The steps that are involved in this study are:

- Data Collection
- Data Preprocessing
- Training and Testing the model
- Spatiotemporal Analysis

So the data we collected will be cleaned and important features are found out from it so that we can train & test the model using them. The algorithm uses historical crime data, social factors and other variables to prepare a predictive outcome. It predicts a range of analytical perspectives: The Spatiotemporal spread of crime combined with Natural Language Processing (NLP) which is an important tool to extract operational intelligence out of unstructured text formats. Also, at the technical level, the research yearns ethical considerations of algorithmic decision making such as data bias risks, model transparency and possible dangers to individual privacy. It also strengthens accountability mechanisms and human-machine collaboration to ensure that predictive systems are responsible.

2.1. Data Collection

The most important source of data is old crime records, which are provided with characteristics such as type of crime, geographical coordinates, time of crime, and description of the crime. Other details about the data are collected in the form of socioeconomic and infrastructural characteristics. All this information is gathered together for the next step [3].

2.2. Data Preprocessing

Preprocessing the data includes cleaning and converting the raw data into a state that is consistent and reliable. Missing values in the data are taken care with the correct imputation methods, and the categorical values are encoded to allow the processing of the data using machine learning algorithms. The spatial coordinates are mapped using Geographic Information System (GIS) tools. For textual data, the preprocessing steps include:

- Tokenization
- Removing stop words
- Lemmatization
- Feature extraction using TF-IDF or word embeddings [4]

2.3. Machine Learning Model

The identified spatial, temporal, and textual features are embodied into machine learning models such as:

- Random Forest
- Support Vector Machine (SVM)
- Gradient Boosting

2.4. Spatiotemporal Analysis

Spatiotemporal analysis is performed to identify patterns of crime that happened in different regions of space and intervals of time. In this regard, Kernel Density Estimation (KDE) and hotspot mapping are utilized to identify regions of high density of crime. In addition, temporal analysis of crime frequency is executed, including daily, weekly, and seasonal patterns [5].

3. System Architecture

The framework for predicting crime uses Spatiotemporal Analytics and Natural Language Processing (NLP). The framework will be composed

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0177
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1180 - 1184

of five main layers [6]:

3.1. Data Acquisition Layer

In this layer, heterogeneous data sets are obtained from different sources. Data sets involve:

- Historical crime data sets
- Geographic data sets (latitude and longitude)
- Socioeconomic data sets
- Police reports/textual narratives about crime

3.2. Data Preprocessing Layer

In this layer, data preprocessing is performed to 2ensure consistency and quality in the gathered data sets. Data preprocessing includes:

- Data cleaning
- Handling missing values in data sets
- Normalization for numerical data sets
- Tokenization for textual data sets
- Stop-word removal for textual data sets
- Feature extraction for textual data sets using TF-IDF [7]

3.3. Feature Engineering Layer

In this layer, spatial, temporal, and textual features will be arranged from the preprocessed data sets.

Spatial Features:

- Geographic coordinates
- Crime density in the vicinity

Temporal Features:

- Hour of the day
- Day of the week
- Seasons [8]

Textual Features:

- Keywords using NLP techniques
- Sentiment and topic modeling from reports

3.4. Machine Learning Layer

In this layer, machine learning algorithms are utilized for predicting crime probability by understanding and observing the engineered features from the data sets.

- Random Forest Classifier
- Support Vector Machine (SVM)
- Gradient Boosting Models [9]

3.5. Visualization and Decision Support Layer

In this final layer, outputs will be generated for the law enforcement authorities by using the results acquired from the machine learning algorithms. The outputs include:

- Crime hotspot maps
- Crime prediction dashboards [10]
- Resource allocation recommendations for proactive policing strategies Shown in Figure 1- 3.

4. Figures

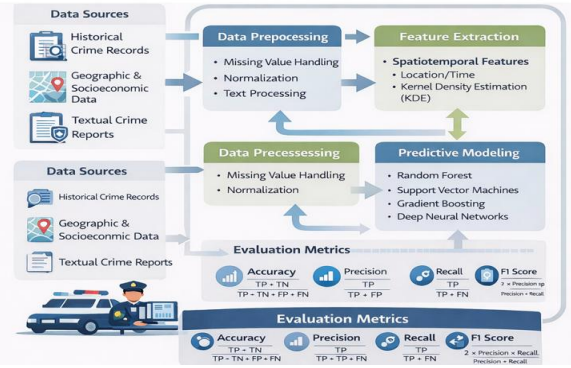

Figure 1 Architecture of the Proposed Spatiotemporal and NLP-based Crime Prediction Framework Showing Data Acquisition, Preprocessing, Feature Extraction, Machine Learning Modeling, and Decision Supports for Proactive Law Enforcement

Figure 2 Graphical Analysis of Crime Prediction Using Spatiotemporal Patterns and NLP-Derived Textual Insights, Highlighting Crime

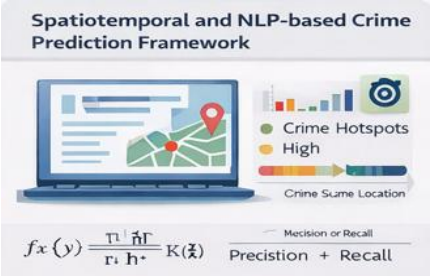

Figure 3 Visualization of Spatial Crime Hotspots Identified Using Kernel Density Estimation (KDE) Based on Historical Crime Location Data

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0177
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1180 - 1184

5. Future Work

Future research can be aimed at the following:

- Integration with real-time IoT and surveillance information
- Use of deep learning algorithms such as LSTM and CNN [11]
- Development of adaptive crime prediction systems using reinforcement learning
- Use of social media analytics to improve early crime detection

These features can be a great contribution in improvising the accuracy and real-time monitoring of crimes.

6. Results and Discussion

6.1. Results

The results from the experiments indicate that the suggested framework increases the accuracy of crime prediction on a large scale when compared to onset models that depend only on old crime data statistics. The combining of NLP feature data from text reports adds extra contextual data that enlarges the prediction ability of the models [12-15].

6.2. Discussions

The results from doing spatial analysis show that crime is clustered in concentrated population areas, and from temporal analysis, it is clear that crime frequency increases in the evenings and at night. The machine learning models were carried out well in detecting potential crime hotspots.

Conclusion

The proposed system serves as a thorough solution for identifying patterns of crime and predicting the occurrence of potential criminal activities using Machine Learning Algorithms, Spatial analysis, and Natural Language Processing (NLP). It strongly supports the benefits of hybrid analytical approaches in enhancing situational awareness for Problem-Oriented Policing. Further studies may assist in this process by merging real-time data streams with deep learning algorithms for more exact predictions.

Acknowledgements

We would like to thank our faculty members and institution for giving us this opportunity and for their continuous support and guidance during the preparation of this research work. We would also like to thank our mentors and colleagues who provided valuable suggestions and encouragement throughout the study. Their support helped us successfully complete this project on Spatiotemporal and NLP-based Framework for Crime Prediction and Prevention.

References

[1]. Chainey, S., & Ratcliffe, J. (2005). GIS and Crime Mapping. Wiley.
[2]. Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. Journal of the American Statistical Association, 106(493), 100–108.
[3]. Wang, T., Rudin, C., Wagner, D., & Sevieri, R. (2013). Learning to detect patterns of crime. Machine Learning and Knowledge Discovery in Databases (ECML PKDD).
[4]. Weisburd, D. (2015). The law of crime concentration at places. Criminology, 53(2), 133–157.
[5]. Brantingham, P. L., & Brantingham, P. J. (1995). Criminality of place: Crime generators and crime attractors. European Journal on Criminal Policy and Research, 3(3), 5–26.
[6]. Neill, D. B. (2009). Expectation-based scan statistics for monitoring spatial time series data. International Journal of Forecasting, 25(3), 498–517.
[7]. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
[8]. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.
[9]. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.
[10]. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357.
[11]. Aggarwal, C. C. (2015). Data Mining: The Textbook. Springer.
[12]. Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0177
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1180 - 1184

Cambridge University Press.
[13]. Mitchell, T. M. (1997). Machine Learning. McGraw-Hill.
[14]. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.
[15]. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.