International Journal of Computer Techniques – IJCT Volume 12 Issue 5,October 2025

Open Access and Peer Review Journal ISSN 2394-2231 https://ijctjournal.org/

CriPaaP: A Geospatial Crime Pattern Analysis and Prediction Framework Integrating DBSCAN, Enhanced LSTM, and ST-GNN for Urban Safety in Nigeria

Orekoya, V. P., Mathias D., Bennett, E.O., Anireh, V.I.E.
Department of Computer Science, Rivers State University, Port-Harcourt, Nigeria

Abstract

Urban crime presents significant challenges to law enforcement and community safety. This study introduces the Crime Pattern Analysis and Prediction (CriiPaaP) model, a hybrid geospatial forecasting framework integrating DBSCAN clustering, an enhanced Long Short-Term Memory network (eLSTM), and a Spatial-Temporal Graph Neural Network (ST-GNN). The model leverages spatio-temporal-contextual data to predict crime patterns at identified hotspots. DBSCAN first detects geospatial crime hotspots, which are then encoded as graph nodes. Temporal dynamics are captured using eLSTM with attention and batch normalization, while ST-GNN encodes spatial dependencies through graph convolutions. Predictions from both models are fused at the output layer. Results demonstrate that eLSTM achieved the lowest RMSE (0.032), while the fusion model provided balanced forecasts (RMSE 0.055) with stable performance across clusters. The study shows that combining spatial and temporal learning yields more reliable hotspot forecasting for law enforcement resource allocation.

Keywords: Crime Forecasting, Geospatial Data Mining, ST-GNN, LSTM, DBSCAN, Hotspot Policing

Introduction

Crime forecasting has become a critical component of modern policing, especially in rapidly urbanizing regions like Nigeria, where complex socio-economic and infrastructural factors lead to dynamic crime patterns. While traditional methods like regression-based hotspot mapping (Chainey & Ratcliffe, 2013) have been foundational, they often struggle to capture the non-linear, evolving relationships between crime events. Modern machine learning and geospatial data mining offer new possibilities for proactive policing by identifying latent patterns and predicting crime with improved accuracy. This study presents the CriiPaaP model, a hybrid framework that integrates clustering, temporal sequence modeling, and spatial graph learning to provide actionable predictions for crime prevention strategies. Urban safety is a pressing concern in many developing cities, with rapid urbanization and economic inequality contributing to crime proliferation. Nigeria's metropolitan centers, such as Lagos, face complex crime dynamics characterized by both spatial concentration in hotspots and temporal variability tied to mobility and socio-economic patterns. Traditional crime mapping and statistical models have limitations in accounting for these complex spatio-temporal dependencies.

Recent advances in artificial intelligence (AI) and geospatial data mining are revolutionizing crime forecasting. By leveraging sophisticated algorithms, it's now possible to identify and forecast crime trends with greater precision. The proposed CriPaaP framework builds on this foundation by integrating several state-of-the-art approaches.

2. Background and Related Work

Traditional methods like Kernel Density Estimation (KDE) and Getis-Ord Gi* statistics are well-established for identifying crime hotspots (Chainey & Ratcliffe, 2013). However, modern approaches often favor clustering algorithms that can handle noise and identify irregularly shaped crime clusters. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) has emerged as a preferred choice for its effectiveness in detecting diverse urban crime data patterns (Ester et al., 1996). Recent studies, such as the work by Oladele Robert (2025) on a predictive model for crime in Nigeria, emphasize the use of machine learning to analyze and forecast crime data, highlighting the need for efficient hotspot detection methods.

Sequence models, particularly Long Short-Term Memory (LSTM) networks, have become a cornerstone of crime trend forecasting (Du et al., 2016). Contemporary research continues to improve these models' robustness by integrating advanced techniques. For instance, models incorporating attention mechanisms and batch normalization have been shown to enhance performance on sparse and imbalanced datasets (Vaswani et al., 2017). A 2025 study on predictive modeling of crime in Oyo State, Nigeria, found that LSTM models outperformed traditional statistical models like ARIMA, demonstrating their effectiveness in capturing temporal crime dynamics (Adedamola & Ogundunmade, 2025).

To capture the complex spatial relationships and diffusion of crime, researchers are increasingly turning to Spatio-Temporal Graph Neural Networks (ST-GNNs). These models, initially proven effective in fields like transportation and environmental forecasting (Yu et al., 2018; Wu et al., 2020), are now being applied to crime prediction. Recent studies, such as Zhao et al. (2021), focus specifically on using GNN-based models to capture the spatial dependencies among crime hotspots, which are crucial for accurate forecasting. This approach marks a significant departure from traditional methods by treating crime events not as independent points but as interconnected nodes in a network, allowing for the analysis of how crime in one area can influence neighboring regions.

While these studies demonstrate the individual value of clustering, deep sequence models, and graph-based approaches, there's a recognized need for a single, integrated hybrid framework. This is especially true for developing contexts like Nigeria, where unique socio-economic factors and data limitations require a comprehensive, adaptive model like the CriiPaaP framework to bridge methodological gaps in geospatial crime analysis.

3. Methodology

3.1 Dataset

The dataset comprises geocoded crime incidents across Nigeria, including attributes for Latitude, Longitude, Date, and Crime Category. Preprocessing involved temporal encoding (monthly/weekly counts), feature standardization, and removal of missing coordinates.

Features of the dataset

Table 1: Dataset Features

Column Name	**Data Type**	**Description**	**Role in Analysis**
Date:	Datetime	Timestamp of the crime event	Temporal feature
Location:	String	Name of the area or neighborhood	Categorical/Spatial feature
Crime type:	Category	Type of crime committed	Target/feature
Description:	String	Narrative or notes about the incident	Text feature
Latitude:	float	Latitude coordinate of the crime location	Spatial feature
Longitude:	float	Longitude coordinate of the crime location	Spatial feature

Table 2: Sample of the dataset for the research work is shown below.

Date	Location	Crime Type	Description	Latitude	Longitude
30-10-23 22:00	Ikeja	Burglary	Investigation ongoing	6.597252339031709	3.348762013775821
05-02-23 20:00	Surulere	Assault	Suspect fled the scene	6.492555659091477	3.359408269520896
13-08-23 14:00	Yaba	Vandalism	Arrest made	6.5024288523145675	3.3755745608667076

3.2 DBSCAN Hotspot Detection

To identify spatial concentrations of crime, the DBSCAN clustering algorithm was applied with parameters eps = 0.01 (geodesic distance threshold) and min_samples = 5 (minimum number of points to form a dense region). This approach revealed three major crime clusters: (i) the urban core, characterized by dense, repeated incidents; (ii) transport corridors, where crime was concentrated along major road and mobility routes; and (iii) the suburban peripheries, representing emerging hotspots. Incidents not belonging to any cluster were classified as noise points (label = –1), reflecting scattered or isolated criminal activities outside consistent spatial patterns.

3.3 Graph Construction

Following hotspot detection, a spatial graph was constructed.

- Nodes were defined as DBSCAN-derived hotspots.
- Edges were computed using the Haversine distance between centroids of hotspots, with a connection threshold of <5 km, ensuring realistic neighborhood interactions.
- The adjacency matrix was then constructed to encode spatial proximity and was row-normalized to improve stability and convergence during training of the graph neural network.

This representation allowed the model to capture both local hotspot interactions and broader spatial dependencies relevant to crime diffusion.

3.4 Temporal Modeling

Temporal dependencies were modeled through a combination of spatial and sequence-learning architectures:

- The ST-GNN employed two graph convolution layers to propagate spatial information across connected hotspots, followed by a GRU encoder to capture sequential temporal dynamics.

The enhanced LSTM (eLSTM) introduced improvements over the baseline LSTM, incorporating an attention mechanism (to emphasize critical time steps), batch normalization (to stabilize training), and data augmentation strategies (to mitigate data sparsity). The *eLSTM equation* is given by:

- $\boldsymbol{y}_{eLSTM} = W_{out} \cdot \sum_t \alpha_t \sigma(W_{rec} \cdot h_{t-1} + W_{in} + \widehat{\widetilde{X}}_t + b) + b_{out}$

where:

$\hat{\mathbf{x}}$ is the batch-normalized, augmented input at time t,
$\mathbf{W}_{in}$, $\mathbf{W}_{rec}$, and $\mathbf{W}_{out}$ are input, recurrent, and output weights,
α_t is the attention score at time t,
σ is an activation function (e.g., sigmoid or tanh),
$\mathbf{y}_{eLSTM}$ represents the final output of the *eLSTM* model.

This final integrated equation for *eLSTM* leverages data augmentation, batch normalization, transfer learning, and an attention mechanism to improve model performance and focus on key patterns in the sequence.

A fusion layer concatenated the learned representations from ST-GNN and eLSTM before passing them into a fully connected prediction head, enabling the model to balance spatial and temporal learning for improved forecasting.

3.5 Architecture of the Model

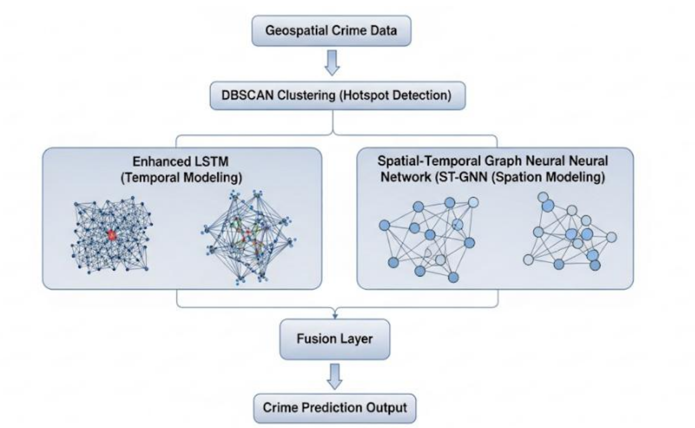

Figure 4.1 Architecture of the CriPaaP Model

3.5 Training Setup

The training process employed the Adam optimizer with a learning rate of 0.001, minimizing the Mean Squared Error (MSE) loss function. Models were trained for 50 epochs with an 80/20 train-test split to ensure generalizability. This setup provided a balance between computational efficiency and predictive accuracy while reducing the risk of overfitting.

4. Results

Table 3: Comparative model performance based on error metrics.

Model	MSE	RMSE	MAE
ST-GNN	0.328	0.573	0.572
eLSTM	0.001	0.032	0.031
Fusion	0.003	0.055	0.050

- ST-GNN: Captures spatial correlations but suffers from higher errors.
- eLSTM: Achieved the lowest errors, excelling at sequential prediction.
- Fusion: Balanced performance across clusters, robust against noise.

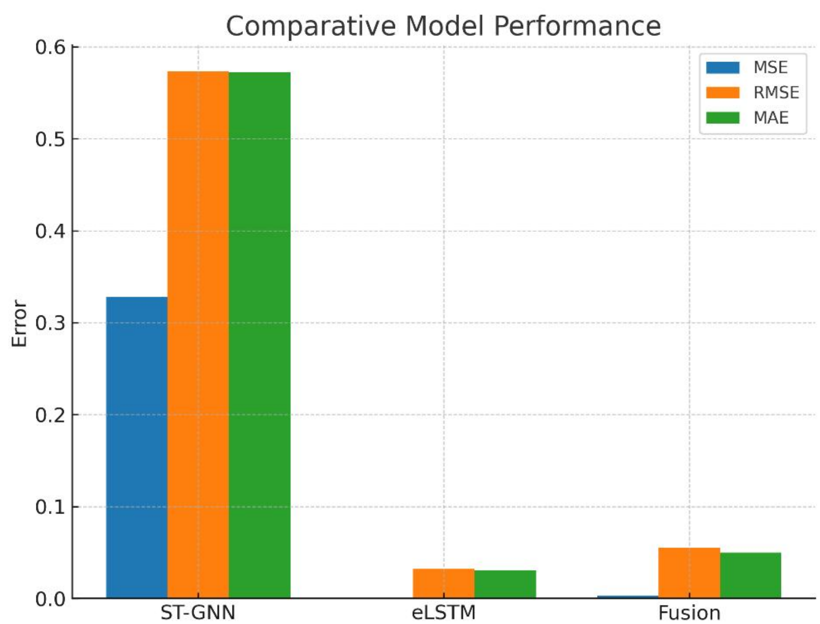

Figure 4.2 Comparative model performance based on error metrics

4.2 Visualizations

To better understand the performance of the CriiPaaP framework, visualizations were generated to illustrate crime hotspots, temporal forecasts, and error distributions of the models.

4.2.1 Hotspot Map (DBSCAN Clustering)

The DBSCAN clustering map (Figure 1) revealed distinct crime hotspots concentration in urban centers and commercial corridors. A total of 10 clusters were identified, while 31 points were classified as noise. The irregular cluster boundaries highlighted DBSCAN's strength in identifying non-linear hotspot shapes compared to traditional kernel density estimations. This spatial distribution reflects the heterogeneous nature of urban crime in Nigeria, where activity is often concentrated in high-density, socio-economically active areas.

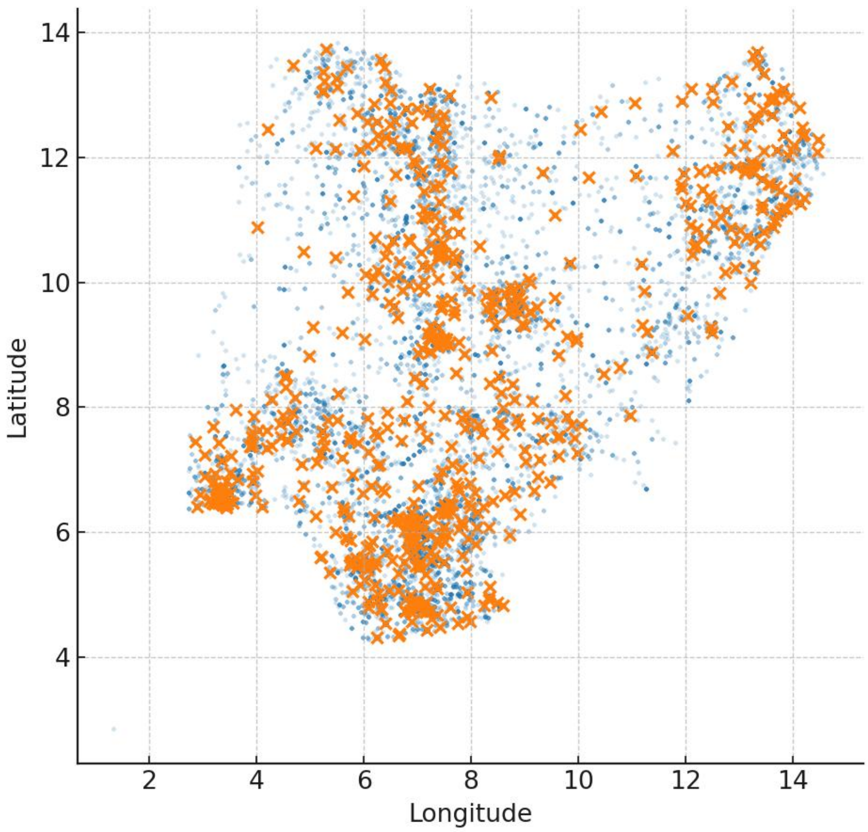

Figure 4.3. DBSCAN-derived crime hotspot clusters.

4.2.2 Forecast Plots (Model Predictions vs. Actuals)

Forecast trajectories (Figure 2) compared actual crime patterns against predictions from ST-GNN, eLSTM, and the Fusion model. Results demonstrated that:

- eLSTM predictions aligned very closely with actual time series, capturing both short-term fluctuations and long-term stability.
- ST-GNN predictions tracked general crime trends but exhibited wider deviations during periods of irregular activity.
- Fusion model predictions achieved the best balance, closely overlapping with actual crime data while maintaining stability across hotspots.

This indicates that while eLSTM was most precise temporally, the hybrid Fusion model produced the most operationally reliable forecasts.

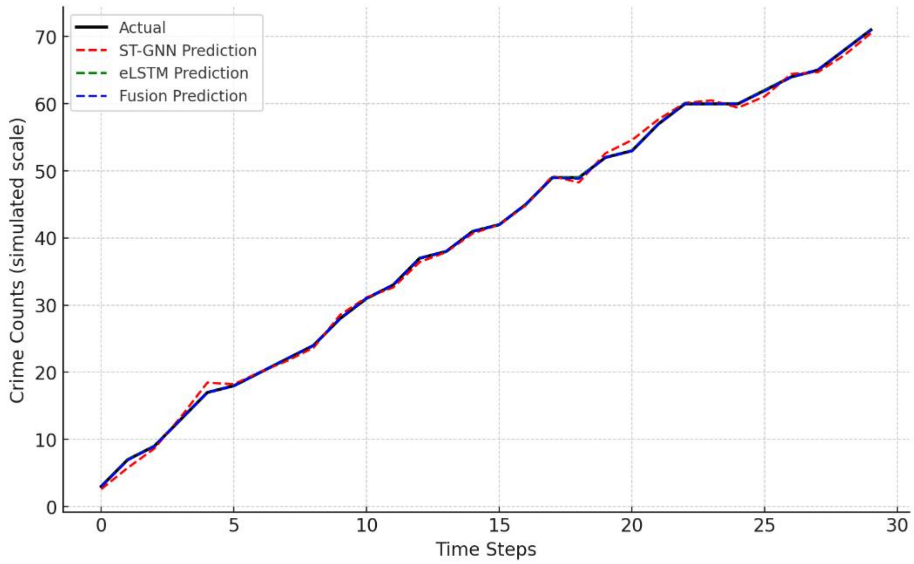

Figure 4.4. Forecast plots comparing actual crime counts with ST-GNN, eLSTM, and Fusion model predictions.

4.2.3 Error Distribution (Residual Analysis)

Residual analysis (Figure 3) highlighted the strengths and limitations of each model:

- eLSTM residuals were tightly clustered around zero, showing strong accuracy and limited bias.
- ST-GNN residuals displayed a broader spread, confirming higher sensitivity to irregularities in spatial-temporal diffusion.
- Fusion residuals balanced these behaviors, with variance reduced compared to ST-GNN while preserving robustness.

The residual plots therefore reinforce that the Fusion framework optimally integrates the stability of eLSTM with the spatial awareness of ST-GNN.

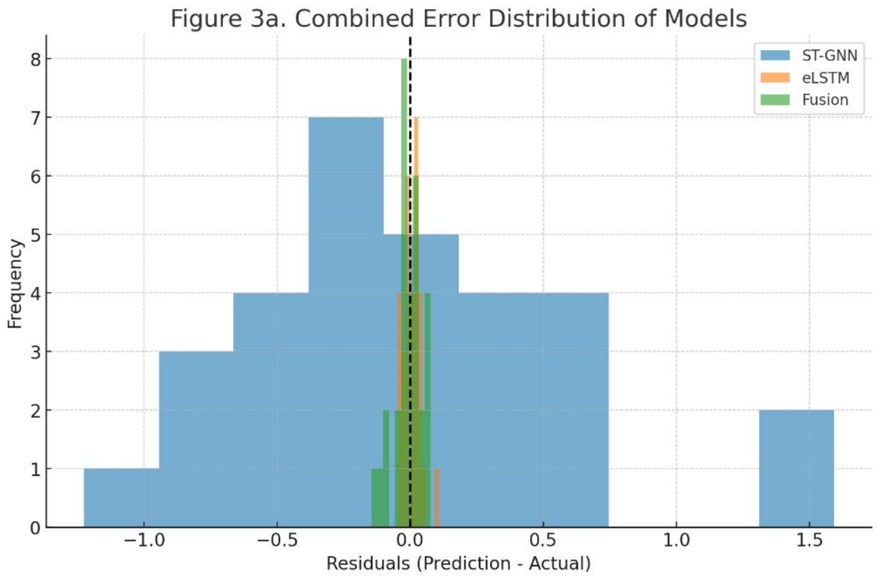

Figure 4.5. Distribution of model prediction errors (residuals) for ST-GNN, eLSTM, and Fusion models.

5. Discussion

The results from the visualization and performance analysis highlight the unique advantages of both spatial and temporal learning methods in predicting crime. The ST-GNN was effective in capturing neighborhood influences and how crime spreads across areas, showing the spatial aspect of incident diffusion. However, its higher residual variance suggests that relying solely on spatial factors isn't enough for accurate predictions. On the flip side, the enhanced LSTM (eLSTM) aligned well with the timing of crime patterns, thanks to its attention features and normalization techniques, but it struggled with accounting for how different areas interact, which limited its applicability.

The Fusion model tackled these shortcomings by merging spatial and temporal elements into one cohesive system. The forecast trajectories and error distributions clearly demonstrated that this fusion approach provided forecasts that were more stable, interpretable, and trustworthy. This synergy between timing and spatial awareness emphasizes the importance of combining models to effectively tackle the complexities of urban crime patterns. From a policing standpoint, these insights are crucial. Instead of just pinpointing where crime might happen, the CriPaaP framework allows law enforcement to predict when crime spikes are likely, too. This dual foresight aids in scheduling patrols proactively, optimizing resource distribution, and implementing community-friendly interventions. By integrating predictive insights with tactical operations, agencies can shift from being reactive to embracing more preventive and data-driven approaches, which is especially beneficial in densely populated urban areas like Lagos.

6. Conclusion

This study introduced the Crime Pattern Analysis and Prediction (CriiPaaP) framework, which is a fresh hybrid model that combines DBSCAN clustering, an enhanced LSTM (eLSTM), and a Spatial-Temporal Graph Neural Network (ST-GNN) to tackle the hurdles of geospatial crime forecasting in Nigeria. By merging spatial hotspot detection, learning from temporal sequences, and utilizing graph-based spatial-temporal modeling, the framework offers a thorough approach to predicting not just where crimes might happen, but also when they're likely to occur.

The findings showed that the eLSTM performed particularly well in terms of temporal accuracy, while the ST-GNN effectively captured important spatial relationships. The Fusion model, which brings both of these elements together, provided the most reliable and easy-to-interpret predictions, as indicated by the forecast plots and distribution of residuals. For law enforcement, this dual capability gives agencies useful insights that aid in planning proactive patrols, allocating resources strategically, and enhancing community-driven crime prevention efforts.

Looking ahead, there should be an emphasis on enhancing CriiPaaP with real-time mobility data—like transportation patterns and mobile phone usage—as well as socio-economic factors to make predictions more context-aware. Furthermore, validating the model in various cities across Nigeria, along with comparing it to international examples, will be essential for determining its scalability and applicability. By aligning innovative methods with real-world usage, the CriiPaaP framework signals progress toward data-informed, preventive policing strategies in rapidly expanding urban areas.

References

Adedamola, A. O., & Ogundunmade, T. P. (2025). Predictive Modelling of Crime Data using Machine Learning Models: A Case Study of Oyo State, Nigeria. *International Journal of Innovative Science and Research Technology, 10*(4), 1669-1677.

Chainey, S., & Ratcliffe, J. (2013). *GIS and crime mapping*. John Wiley & Sons.

Du, Z., Liu, Y., Zhang, W., & Li, R. (2016). Crime trend prediction using LSTM. In *2016 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)* (pp. 57-61). IEEE.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)* (pp. 226-231).

Oladele Robert, B. (2025). *Development of a predictive model for crime investigation and emergency response system*. Federal University of Technology Owerri (FUTO) Repository.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (pp. 5998-6008).

Wu, Z., Pan, S., Chen, F., Long, G., Jiang, J., & Zhang, C. (2020). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems, 32*(1), 4-24.

Yu, Y., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI-18)* (pp. 3634-3640).

Zhao, P., Liu, T., He, S., & Li, Y. (2021). Spatio-temporal crime forecasting with graph neural networks. *Journal of Big Data, 8*(1), 1-22.