Modelling Crime Dynamics with a Spatio-Temporal CBAM Attention-Enhanced Deep Learning Model

MSc Practicum Part 2
Master of Science in Data Analytics

Krishna Reeja Santhosh Kumar
Student ID: 23325071

School of Computing
National College of Ireland

Supervisor: Vikas Sahni

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Krishna Reeja Santhosh Kumar
Student ID:	23325071
Programme:	Master of Science in Data Analytics **Year:** 2025
Module:	MSc Practicum Part 2
Supervisor:	Vikas Sahni
Submission Due Date:	28/01/2026
Project Title:	Modelling Crime Dynamics with a Spatio-Temporal CBAM Attention-Enhanced Deep Learning Model
Word Count:	6964 **Page Count: 22**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:

Date: 26/01/2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Modelling Crime Dynamics with a Spatio-Temporal CBAM Attention-Enhanced Deep Learning Model

Krishna Reeja Santhosh Kumar
23325071

Abstract

Crime prediction has increasingly relied on advanced spatiotemporal modelling due to the complex and dynamic nature of urban crime patterns. There are some traditional statistical and machine-learning approaches which struggle to capture long-range temporal dependencies, spatial interactions between regions, and contextual influences like weather which results in limited predictive reliability. This study addresses these challenges by developing a Spatio-Temporal CBAM Attention-Enhanced Deep Learning Model designed to improve the extraction of important spatial and temporal crime features. Using Chicago's large-scale crime dataset combined with weather data, the study constructs and evaluates multiple baseline sequential models which includes LSTM, BiLSTM, BiGRU, and BiLSTM with Cross-Attention against the proposed dual-attention architecture.

Experimental findings show that the Spatio-Temporal CBAM model achieves the lowest MSE is 0.7036, RMSE is 0.8388, and MAE is 0.6685 outperforming all baselines by better focusing on informative spatiotemporal patterns while reducing noise. These results demonstrate both theoretical values, by advancing attention-based modelling for crime forecasting, and practical benefits, providing more reliable crime trend predictions for decision-makers. While the model improves accuracy and seasonal tracking, limitations persist in capturing extreme crime spikes and ensuring computational efficiency for real-time deployment. Future work should explore graph-based spatial learning, transformer architectures, and finer-granularity contextual data to further enhance predictive performance.

Keywords: Crime prediction, Spatio-Temporal CBAM, Deep learning, Attention mechanisms, Forecasting models

1 Introduction

1.1 Background

The prediction of crime has taken its rightful position among the current features of urban management due to the growing access to massive amounts of spatio-temporal data (Du and Ding, 2023) and the growing importance of data-driven policing approaches (Fan et al., 2025). Historically, the study of crime has been based on statistical software and the study of historical trends, which in many cases could not identify nonlinear behavioural patterns and dynamism of urban dynamics. The more advanced methods today have made it possible to model complex relationships over time, space, and socio-environmental variables (Liu et al., 2025) with the emergence of machine learning and deep learning (Silva et al., 2022). Other cities with high levels of crime history, such as Chicago, have an open-data platform, which would be the most favorable environment to develop a predictive system that can be used in proactive policing. Moreover, other external conditions including temperature, humidity, and wind speed had been found to affect human activity (Chen et al., 2025) and crime occurrence and thus multimodal data integration has become more and more useful. This context drives the necessity of superior frameworks - spatio-temporal attention-based frameworks that can enhance the accuracy of forecasts, as well as enhance interpretability and enable intelligent decision making in safer communities.

1.2 Importance of the Study

The importance of this study is that it advances crime forecasting because of its incorporation of spatio-temporal dependencies and attention-enhanced deep learning. The classical models are not able to reflect the intricate interactions between crime patterns, time series and other environmental changes like weather. This study offers a more correct, interpretable, and reliable model of predicting variations in crime daily by presenting a CBAM-powered architecture. The results allow data-driven decision making which allows law enforcement agencies to maximize resources allocation, proactively deploy patrols, and increase the safety of the population. Further, this work illustrates the effectiveness of the integration of crime and weather information to identify concealed behavioural patterns, which can bring important implications to smart-city analytics in the future.

1.3 Research Question

What improvements in spatial and temporal feature extraction are achieved by integrating a Spatio-Temporal CBAM attention mechanism into crime prediction models, and how does this approach perform relative to conventional baseline models in terms of predictive accuracy and efficiency?

1.4 Research Objectives

There are some research objectives in this study which are:

1. To evaluate the extent to which the Spatio-Temporal CBAM attention mechanism improves the extraction and weighting of spatial and temporal features in crime prediction tasks.
2. To develop and compare multiple deep learning architectures which includes LSTM, BiLSTM, GRU, cross-attention models and the proposed Spatio-Temporal CBAM model to evaluate differences in predictive accuracy and model behavior.

3. To benchmark the proposed model's performance against traditional baseline models using quantitative metrics like MSE, RMSE, MAE, latency, throughput, execution time and memory usage.

1.5 Structure of the Study

This study is organized into seven sections. Section 1 introduces the background, research question, objectives and importance of the study. Section 2 reviews related work on crime analytics, deep learning and spatio-temporal modelling. Section 3 presents the methodology having datasets, preprocessing, feature engineering and model development. Section 4 explains the system design and proposed CBAM architecture. Section 5 details implementation of baseline and CBAM models. Section 6 evaluates performance through accuracy, trend analysis, and computational performance. Section 7 provides the conclusion, limitations, and future work.

2 Related Work

2.1 Crime Analytics and Predictive Policing

Crime analytics has changed in many ways during the last twenty years due to the access to massive data (Oatley, 2022) like the Chicago Crime Dataset, where millions of geo-tagged and time-stamped crime incidents are stored. Descriptive statistics and hotspot mapping were the initial methods of crime analysis (Ahmad et al., 2024) although they were only insightful in a retrospective manner and did not enable the prediction of future crime trends (Kortas et al., 2022). Since the nature of crime is always subject to space, neighbourhood attributes and time control like weekdays, seasons and special occasions, the conventional linear models and rule of thumb methods could not explain the dynamics of crime. The introduction of machine learning also facilitated more advanced predictive policing systems that could learn non-linear relationships (Kim and Lee, 2023) but such models as the Logistic Regression, Random Forests or Gradient Boosting still considered crime occurrences as independent observations and ignored the fact that neighboring locations or time series are interdependent. Research is also becoming more aware of crime as a spatial-temporal process with the events spreading out geographically and changing behaviourally, environmentally and socio-economically. Despite the fact that predictive policing systems with statistical learning enhanced the ability to offer early alerts (Alkhazraji and Yahya, 2024), they nevertheless did not have powerful mechanisms to combine spatial clustering and long-term temporal interactions. This has motivated a move towards deep learning-based spatio-temporal modelling, which has more power to reveal hidden patterns necessary to accurately predict crime and proactive policing.

2.2 Deep Learning Approaches for Sequential Crime Forecasting

The recent progress in deep learning has made a substantial contribution to sequential crime forecasting, including the ability to capture complex spatiotemporal correlations, multimodal data, and intelligent architectures to address the changing crime trends. In (Tasnim et al., 2022), the authors present a multi-module deep learning architecture integrating Attention-LSTM, Stacked Bi-LSTM, and fusion approaches to forecast hourly crime rates in cities in the U.S. and demonstrate high accuracy, albeit with difficulties in the transferability of models across the locations. The fighting or shooting behavior of suspicious humans is identified in (Mahor et al., 2025) with the help of an InceptionV3-BiLSTM pipeline, which has a recognition rate of 96.80, but real-time application is still restricted due to the limitation of constant tracking requirements. The study (İlgün and Dener, 2025) combines machine learning, deep learning, and statistical forecasting algorithms, including XGBoost, BLSTM, LSTM, Prophet, and SARIMA, to forecast the type of crime and the hotspots of crime in large cities in the United States, but it is subject to all heterogeneous datasets and dissimilar trends by district. Study (Mao et al., 2025) presents crime prediction as a spatiotemporal sequence problem and a CNNLSTM model is suggested, demonstrating that binning of data greatly can decrease the error in prediction; nevertheless, binning may harm the spatial granularity.

Fine-grained human mobility data are added to neural crime models in (Wu et al., 2022), which only increases F1 scores by 277.5 per cent across cities and types of crime, but the benefits of neural models are only slightly greater than those of traditional regressions. In (Tam and Tanriöver, 2023), the authors employ the multimodal fusion method which relies on crime information and crime related tweets, via a ConvBiLSTM model with an accuracy of 97.75% however it is limited by the unclean and inconsistent social media feeds. Although study (L and Shyni, 2024) is cybersecurity-oriented, it can be seen as a methodologically valuable study with a power to combine hybrid sequential models (Stacked BiLSTM and GRU) with feature selection and achieve 98.71% accuracy in threat prediction, as the authors note the transferability of the applied modeling to crime prediction. Study (Butt et al., 2022) further reinforces temporal prediction of crime with a Bi-LSTM and Exponential smoothing hybrid tested on the New York City data of crimes with much lower MAPE and RMSE values, although contextual changes over time are challenging to forecast. Finally, study (Bi et al., 2024) uses a more complex TCN Self-Attention-BiGRU hybrid optimized using evolutionary algorithms, and it provides high R2 scores similar to over 0.91 in predicting patterns of sequential network attacks, which show elements of architecture that can be useful to develop crime forecasting models as shown in Table 1.

Table 1: Summary Table

Study	**Purpose / Problem Addressed**	**Key Methods & Techniques Used**	**Main Results**	**Limitations**
(Tasnim et al., 2022)	Predict hourly crime counts using spatiotemporal patterns.	Attention-LSTM, Stacked Bi-LSTM, Fusion model, Transfer Learning.	MAE: 0.008 (SF), 0.02 (Chicago); R^2: 0.95–0.94; SMAPE:	High complexity; city-to-city generalization challenges.

			1.03%–0.6%.	
(Mahor et al., 2025)	Detect suspicious human activity in surveillance videos.	CNN (InceptionV3), Bi-LSTM, 6 benchmark datasets.	Accuracy: 96.80%, surpassing state-of-the-art.	Continuous real-time monitoring still difficult; dataset bias possible.
(İlgün and Dener, 2025)	Comprehensive city-level crime prediction and hotspot analysis.	XGBoost, CatBoost, RF, DT, MLP, KNN, GNB, LR; LSTM, BLSTM; Prophet, SARIMA, HWES.	Effective hotspot and trend forecasting; multi-city insights.	Heterogeneous datasets; varying district patterns reduce uniform performance.
(Mao et al., 2025)	Improve crime count prediction as a spatiotemporal sequence task.	Raw vs binned data analysis, spatiotemporal modeling.	Best performance with 5–10 bins; raw data gave high errors.	Binning reduces spatial detail; interval choice impacts accuracy.
(Wu et al., 2022)	Enhance short-term prediction using human mobility flow data.	Mobility metrics, deep neural networks across multiple cities.	F1 improvement 2%–7% across crime types and cities.	Neural model improvements still small vs regression; mobility data availability.
(Tam and Tanriöver, 2023)	Crime prediction using multimodal crime + Twitter data fusion.	Crime dataset + crime-related tweets; benchmarking with SVM, LR, NAHC, BERT models.	97.75% accuracy, outperforming traditional & BERT models.	Social media noise, temporal inconsistency in tweet activity.
(L and Shyni, 2024)	Improve threat intelligence-based sequential prediction (methodology relevance).	Normalization, Standardization, PCA, BiLSTM-GRU hybrid.	Accuracy: 98.71%, strong sequential modeling capability.	High computational overhead; dependent on feature preprocessing quality.
(Butt et al., 2022)	Short-term crime forecasting with limited spatiotemporal data.	NYC crime data (2010–2017), ES trends, BiLSTM patterns.	Lower MAPE, RMSE, MAE than SARIMA; strong forecasting stability.	Limited contextual variables; may not capture abrupt social changes.
(Bi et al.,	Accurate real-	SG smoothing,	R^2: 0.9460	Very complex

2024)	time prediction of network attacks (methodology transferable).	TCN, Self-Attention, Bi-GRU, Genetic + SA + PSO optimization.	(data1), 0.9154 (data2).	design; requires high-quality time-series data.

2.3 Research Gaps

Existing literatures that mentioned in Table 1 indicates that there is a high level of advancement in crime forecasting, although there are various gaps that are yet to be filled. Most of the models use either time or space trends without having an integrated process of capturing both of them. Some of them lack generalizability across territories, high model complexity, or rely on noisy external data, including social media. Others disregard the contextual variables such as weather or bin spatial granularity. Furthermore, majority of architectures lack the dual attention mechanism (Huang et al., 2021) to narrow down feature significance. The limitations mentioned above underscore the importance of an adaptive, efficient, and interpretable spatio-temporal model, which was discussed in this work in the context of a CBAM-augmented deep learning framework.

3 Methodology

3.1 Research Design Framework

The study uses the CRISP-DM (Cross-Industry Standard Process for Data Mining) model in organizing the creation of the Spatio-Temporal CBAM Attention-Enhanced crime prediction model. It starts with Business Understanding the aim of which is to predict the trends of crime in Chicago districts by identifying both spatial and temporal dependencies. Depending on the Data Understanding stage, past data on crimes and weather data would be analyzed to determine patterns, discrepancies, and the influencing factors. Data Preparation phase includes cleaning, which covers missing values, combining crime counts, engineering temporal and weather-related features, and rolling windows can be used in sequential learning. In the Modeling stage, several baseline models (LSTM, BiLSTM, GRU) and the suggested CBAM-enhanced architecture are built and trained with the help of standardized parameters. Evaluation phase involves the testing of every model based on MAE, RMSE, inference latency and throughput. Lastly, the Deployment-based Forecasting step takes the most successful model to come up with future crime forecasts, which are applicable and reproducible in practical settings.

3.2 Dataset Description and Collection

3.2.1 Crime Dataset

The major data source in this study is the Crimes in Chicago Dataset (2019- Present),[1] which is one of the most detailed and most important sources of the spatio-temporal research of crimes. This dataset has over 7 million records of crime incidents which are all reported cases with detailed contextual and geographical details. Each record contains the case identifier, the crime, a detailed description, date and time when it occurred, block level location, arrest or not, domestic violence, ward, community area, police district, FBI crime code, as well as the exact geographical coordinates in the X-Y projection and latitude-longitude. The open data portal of the City of Chicago regularly updates its dataset daily, which means that the dataset is highly timely to analyze long-term trends. In this work, the period 2019-2025 dataset was chosen because it is consistent, complete, and aligns with the weather dataset periods.

3.2.2 Weather Dataset

In order to examine how variables of the environment impact crime patterns, an hourly weather record of Chicago/Midway Airport meteorological station was included.[2] This data set has a total of 9,864 records in a variety of years with each record containing localized climatic variables that are coordinated to Chicago time zone. The measured fields are temperature (T), pressure data (Po, P, Pf), humidity (U), wind direction and wind velocity (Ff, ff10, ff3), visibility (VV), dew point (Td), precipitation (RRR), and trace rainfall (trace rainfall indicators), sunshine duration (sss), and a number of other atmospheric indicators. Crime is usually subject to an observable impact of weather conditions, and such aspects as temperature, humidity, and precipitation have been associated with the changes in the levels of violent and property crime.

3.3 Data Preprocessing

The preprocessing of data consisted in processing the crime and weather data so that the data could have high quality, consistency, and machine-learning-friendly input to be used in the next modelling. In the case of the Crimes in Chicago dataset, the preprocessing started with loading the raw data into one of the structured Pandas DataFrames and then proceeded with the extensive processing of the missing values in the fields of district identifiers, coordinates, and timestamps. This was followed by cleaning and bifurcation of crime records according to the type of crime and geographic locations to help in more meaningful analysis of time and space. Exploratory Data Analysis (EDA) was conducted to investigate the annual trends, the distribution of crimes, the times of peak-hours and the alternatives of the hotspots to comprehend the behavioral pattern. To form time-dependent feature engineering, weekday/weekend indicators, month-month aggregate, and long-term moving statistics were generated that indicate periodic crime variability. In the case of the weather data, preprocessing involved loading the hourly weather records, changing the timestamps to a single common datetime format and dealing with any discrepancies or unavailability of atmospheric measurements. A process of feature engineering was done to extract values of

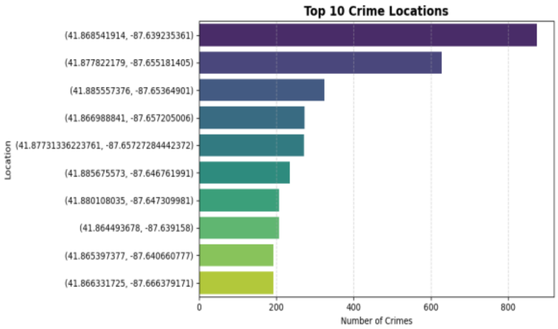
---

[1] https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2/about_data
[2] https://rp5.ru/Weather_archive_in_Chicago,_Midway_(airport)

temperature ranges, changes in humidity, signs of precipitations, measures of visibility, and daily weather summaries. The two datasets were normalized and transformed where necessary so as to have the same scales and enhance stability of the model.

3.4 EDA for Crime Dataset

Figure 1 illustrates geographic distribution of crime incidents in the top 10 high-risk locations in Chicago in a horizontal bar chart developed based on the latitude-longitude coordinates.



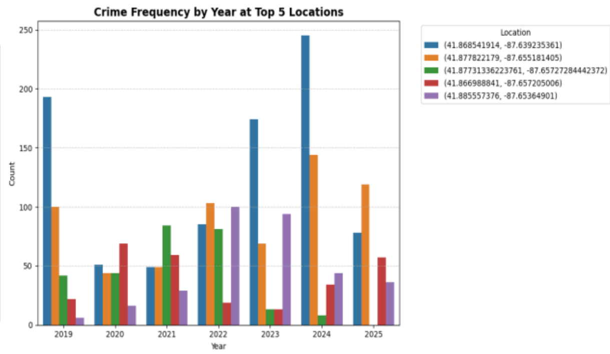


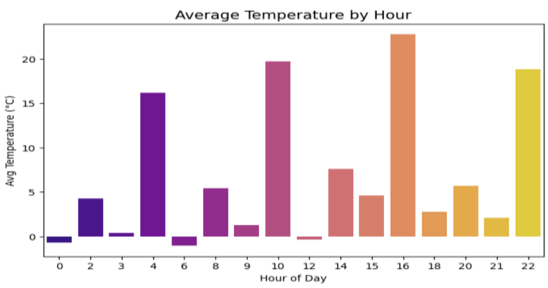
Figure 1: Top 10 Crime Hotspot Locations

Figure 2: Temporal Crime Trends

In Figure 2, this grouped bar chart shows the annual change in crime frequencies in 5 high-risk locations between the year 2019 and 2025. Findings indicate that there are great variability in the number of incidents over time, but in some places, the tallest peaks were in the years 2019 and 2024 of more than 200 incidents. The varying patterns among the locations warrant the addition of time characteristics to the proper models of crime forecasting and hotspots prediction.

3.5 EDA for Weather Dataset

Figure 3 shows the hourly changes in temperature as a bar chart, showing specific diurnal variations, the lowest temperatures (4-5 o C) on the early morning hours (0-6) and the highest temperatures (20-23 o C) on the afternoon hours (16-18).



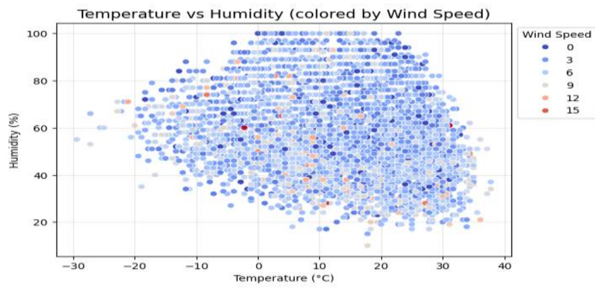


Figure 3: Average Temperature by Hour of Day

Figure 4: Correlation

Figure 4 is a scatter plot that analyzes the correlation between temperature and humidity where findings indicate a negative correlation: the higher temperatures (20-40 0 C) the lower the humidity (20-60 percent), whereas a colder temperature implies more humidity.

3.6 Data Merging

Both datasets were preprocessed after which crime and weather data were then merged on a common Date index to form a single daily-level dataset. To make sure that all the days of crime were retained, a left join was performed and time-based interpolation was performed to fill in all the missing weather observations to preserve the temporal continuity. An extensive set of features was then chosen, among them crime measures (crime count, arrests), time-related (weekday, weekend, holiday, lagged values) measures, aggregated behavioral measures (weekday/monthly average), and meteorological (temperature, humidity, wind speed, DI) measures. Crime count was the target variable. The data was divided into 80 percent training and 20 percent testing in a chronological manner ensuring no leakage of information. StandardScaler was trained exclusively on training data and used on the entire data to ensure that the scaling remained the same (Ahsan et al., 2021) with all the data used to generate windows. Lastly, a 7-day rolling window operation was introduced to map the time series into supervised sequences and generate arrays of the form (samples, window 5, features) of X and (samples, 1) of y that allow week to week crime dynamics to be successfully learned.

3.7 Training Callbacks: Early Stopping and Learning Rate Scheduling

In order to achieve stable and efficient model training, the training pipeline was enriched with two important callbacks. EarlyStopping was also set to track the loss on validation (Miseta et al., 2024) and stop training once the loss was not reduced (Bergman et al., 2024) in five stages in a row and automatically resume the model weights that were performing the best. This avoided excessive overfitting and the superfluous calculation. Also, the scheduler used was ReduceLROnPlateau to continuously learn the learning rate by lowering it by half when the loss rate stopped decreasing on the validation set in two epochs. A combination of these callbacks resulted in the optimization of model convergence, stabilization of training behavior, and the attainment of optimal generalization of the deep learning models.

4 Design Specification

4.1 System Components and Architecture

The architecture of the system consists of six built-in layers of crime prediction. Data Collection layer is the aggregation of crime and weather data as shown in Figure 5. Preprocessing and EDA layer process the data cleaning operation, the null values of both datasets and the exploration of data regarding the types of crimes. The Feature Engineering and Merging layer generates time-related features (weekday/weekend, holidays), and combines data using 7-day rolling windows and scaling. The Models layer uses the five deep learning models, which are LSTM, BiLSTM, BiGRU, BiLSTM+Attention baselines, and the proposed model, which is the Spatio-Temporal CBAM model with enhanced attention mechanisms, that are trained on the 80:20 split. The Evaluation layer assesses the performance based on MSE, RMSE, MAE and latency estimation. Lastly, the Output layer provides 7-day hotspots forecasts of crime with clarifiable AI explainable interpretations.

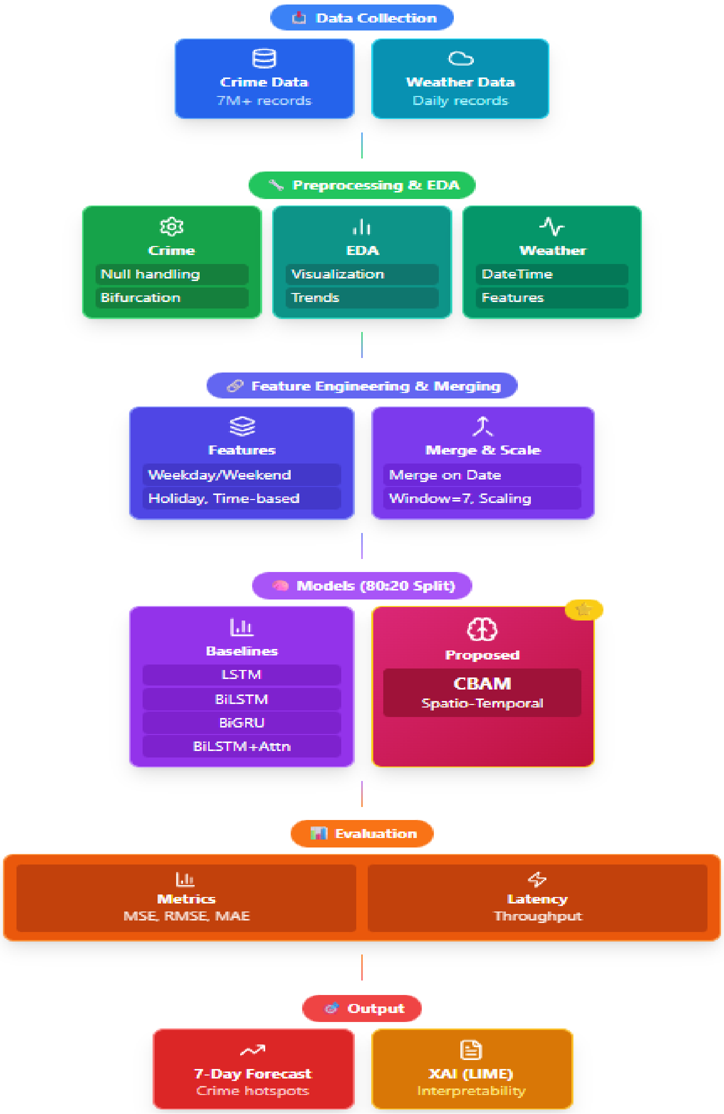


Figure 5: System Architecture Diagram

4.2 Theoretical and Empirical Basis for the CBAM-Enhanced Model

The proposed Spatio-Temporal CBAM architecture is presented as the most appropriate architecture to be applied to crime forecasting because the proposed architecture can both undertake long-range temporal dependence and fine-grained spatial-feature significance, which cannot be modeled by the other baseline deep learning models, including LSTM, BiLSTM, and GRU. Interacting temporal patterns (weekly cycles, lags and seasonal changes) and contextual factors (weather conditions, district variations and behavioral indicators) will inherently impact crime dynamics. CBAM module enhances the representational learning by using Channel Attention, which emphasizes influential dimensions of features (Wang et al., 2024) and Spatial Attention, which emphasizes influential moments in time in the rolling window. With forward-temporal and backward-temporal relations to the forward-temporal ones learned and refined using attention-controlled feature weighting, the model is combined with stacked Bidirectional LSTMs (Li et al., 2021). The empirically observed CBAM-enhanced architecture showed greater convergence performance, reduced validation error, and increased generalization, which proved its performance in addressing the complex, multivariate and nonlinear crime pattern modeling in comparison to the basic recurrent models.

Pseudocode for this Novel Approach

Input: Crime data C, Weather data W
Output: Next-day crime prediction ŷ
1. Load C, W
2. Preprocess C (clean, temporal features, lags, averages)
3. Preprocess W (clean, datetime convert, weather features)
4. **Merge** C and W on Date; interpolate missing weather
5. **Select** feature_cols, target_col
6. **Split** data (80% train, 20% test)
7. **Scale** train features/target; transform full dataset
8. Create rolling windows X, y (window_size = 7)
9. Build CBAM-BiLSTM model:
    - Stacked BiLSTM layers
    - Channel Attention
    - Spatial Attention
    - Dense output (1)
10. **Compile** model (Adam, MSE, MAE)
11. Train with EarlyStopping + ReduceLROnPlateau
12. Predict ŷ = model.predict(X_test)
Return ŷ

5 Implementation

5.1 Model Training

5.1.1 Baseline Models

The base deep learning architectures such as the LSTM, BiLSTM, GRU, and BiGRU were trained with a uniform setup to allow uniform comparison between architectures. All of the models were trained using the Adam optimizer, MSE as the loss, and MAE as the performance metric. The model.fit() function was used to train with 50 epochs, a batch size of 64 and with a real-time evaluation using a validation set. Early stopping was used to stabilize the learning process by stopping the training process once the learning was stagnant, and a learning rate reduction callback was used to dynamically decrease the optimizer. Such environments were used to give credible baseline measures on which the proposed CBAM model was compared.

5.1.2 Proposed CBAM Model

The hybrid architecture that was proposed is Spatio-Temporal CBAM model incorporates 1D convolutional layers, channel attention, spatial attention, and LSTM sequence learning module. Localized temporal patterns in each sequence window are obtained through the 1D convolutions and the channel attention block focuses on the most influential dimensions of the features and attempts to weight them more heavily. Spatial attention block also additional refines feature maps by emphasizing essential spatial dependencies on the basis of inputs that are district-based. The cleaned-up representations are then inputted into an LSTM block to learn long-term temporal dependencies, and then a fully connected layer is used to produce the final crime prediction. Adam optimizer, MSE loss, MAE metric, and early stopping and dynamic learning rate reduction were used as the call backs that trained the model.

5.2 Model Implementation

5.2.1 LSTM Model Implementation

The baseline LSTM model was implemented using a Sequential architecture that would capture temporal dynamics (Afebu et al., 2021) of the crime weather sequence window. This model starts with an input Layer whose shape is (timesteps, features) and with a 128 units LSTM layer with ReLU activation with return sequences=False to obtain high-level time-based embeddings. The overfitting is minimized with three Dropout layers (0.7 rate) of heavy regularization.

5.2.2 BiLSTM Model Implementation

The Bidirectional LSTM model is an improvement of the temporal learning on baselines by processing sequences forward and backward. An architecture is beginning with a 128-unit Bidirectional LSTM layer with return sequence enabled i.e., return sequences=True, and subsequently the second 64-unit Bidirectional LSTM with return sequences disabled i.e. return sequences=False. After every recurrent block, several

5.2.3 Bidirectional GRU Model Implementation

Bidirectional GRU model is a computationally efficient variant that still has a well-developed sequence-modeling power. The architecture begins with a Bidirectional GRU unit layer of 64 units, which is set to give back sequences to allow a further abstraction of time. The second layer of Bidirectional GRU has 64 units but only the final timestep embedding is provided, which is akin to global temporal dynamics.

5.2.4 BiLSTM With Cross-Attention Implementation

The BiLSTM architecture with Cross-Attention adds an attention mechanism in order to emphasize the most effective timesteps. This model has two Bidirectional LSTM layers at the bottom with 128 units each, with return sequences=True which is necessary to maintain the entire temporal sequences that attention needs. The recurrent layers are regularized by dropout layers (0.3).

5.2.5 Spatio-Temporal CBAM Attention Model Implementation

The proposed Spatio-Temporal CBAM model consists of deep recurrent layers and Channel as well as Spatial Attention that refines the features. The architecture layers 4 Bidirectional LSTM layers each with a sequence returned and with high dropout (0.7) to avoid overfitting of the high number of model parameters. Channel Attention that is used as the initial stage of the CBAM module consists of Global Average and Global Max pooling, shared Dense layers (ratio 8), and sigmoid gating to emphasize informative channels as shown in Table 2.

Table 2: Summary of All Implemented Models and Their Architectural Characteristics

Model	Recurrent Layers	Dense Layers	Dropout Used	Early Stopping
LSTM Baseline	1× LSTM (128, ReLU, return_sequences=False)	64 → 64 → 32 → 1	Very High (0.7 at multiple points)	Restored best at epoch 12
BiLSTM	BiLSTM 128 → BiLSTM 64	64 → 64 → 1	Very High (0.7)	Restored best at epoch 3
BiGRU	BiGRU 64 → BiGRU 64	1	Medium (0.3)	Early stopping used
BiLSTM + Cross Attention	BiLSTM 128 → BiLSTM 128	1	Moderate (0.3)	Restored best at epoch 1
CBAM (Proposed)	BiLSTM 128 → 128 → 64 → 64	64 → 64 → 1	Very High (0.7)	Restored best at epoch 2

5.3 LIME-Based Explainability for the BiLSTM–CBAM Model

To increase the interpretability of the proposed BiLSTM -CBAM architecture, a LIME (Local Interpretable Model-Agnostic Explanations) mechanism was applied to reveal regression prediction at an instance level. Firstly, the weather-crime input tensors were flattened to be in a tabular format similar to that of LIME, and then meaningful feature names were created over all timesteps and variables. To allow LIME to run on flattened samples, a custom predicting function was created that can reformat the sample back to the original 3D format and get predictions out of the trained model. LimeTabularExplainer was run in continuous mode, and trained on flat sequences so as to maintain time dependencies. In the selected test case, LIME has created perturbed samples, their local approximations and retained the most significant features.

6 Evaluation

6.1 Experiment 1: Model Accuracy Evaluation (MSE, RMSE, MAE)

This experiment compares the predictive performance of five models, including LSTM, BiLSTM, BiGRU, BiLSTM with Cross-Attention, and the proposed Spatio-Temporal CBAM Attention model, based on MSE, RMSE, and MAE. The attention-based BiLSTM model appears to considerably increase the accuracy, as indicated by lower MSE (0.7045) and MAE (0.6718) as shown in Table 3. This demonstrates that learning crime dynamics is enhanced with the direction of the model towards addressing relevant time steps. This results in a lowest overall accuracy with least MSE (0.7036), RMSE (0.8388), and MAE (0.6685) of the proposed Spatio-Temporal CBAM Attention model.

Table 3: Model Accuracy Comparison (MSE, RMSE, MAE)

Model	MSE	RMSE	MAE
LSTM	0.7478	0.8647	0.6959
BiLSTM	0.7626	0.8733	0.6850
BiGRU	0.7279	0.8531	0.6869
BiLSTM + Cross-Attention	0.7046	0.8394	0.6718
Spatio-Temporal CBAM Attention (Proposed)	**0.7036**	**0.8388**	**0.6685**

6.2 Experiment 2: Actual vs Predicted Crime Trend Analysis

The predictive ability of the LSTM model between July 2024 and September 2025 is displayed in this time series line graph because the blue line is real crime events and the red line is predicted events as observed in Figure 6. The model is able to capture the overall trend with the predictions being close to 9 incidents per day. Nevertheless, it shows major weaknesses in anticipating sudden crime bursts greater than 20+ occurrences generating excessively smooth predictions that overlook temporal volatility.

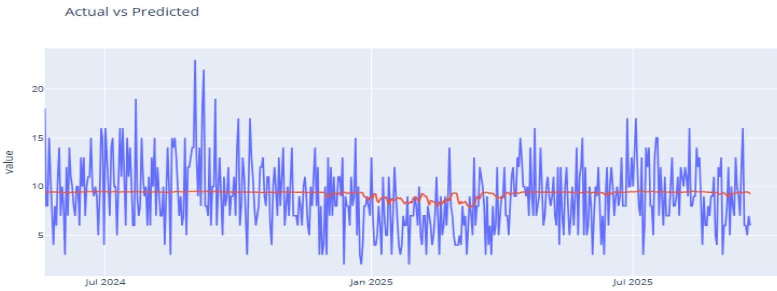


Figure 6: Actual vs Predicted Graph – LSTM

Figure 7 is a time series line graph showing the prediction ability of the BiLSTM model between July 2024 and September 2025. BiLSTM approach shows better results in comparison with the traditional LSTM because it includes the concept of the two-way temporal dependency, which leads to the slightly increasing accuracy of the trend, and the approach aligns with 8-9 incidents per day.

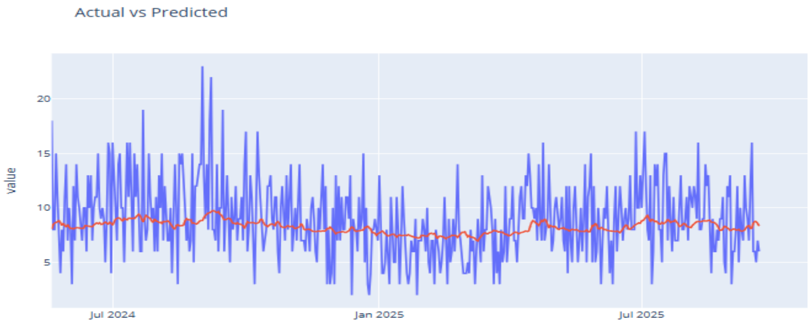


Figure 7: Actual vs Predicted Graph-BiLSTM

Figure 8 is a time series line graph that gives the predictive performance of the Bidirectional GRU model. The Bidirectional GRU is significantly more responsive to changes in time and more effective at capturing changes at the mid-range (10-15 incidents) as compared to LSTM and BiLSTM models. The forecasts are more in line with the real trends and some volatility patterns are reflected, but the extreme spikes of over 20 incidents are difficult.

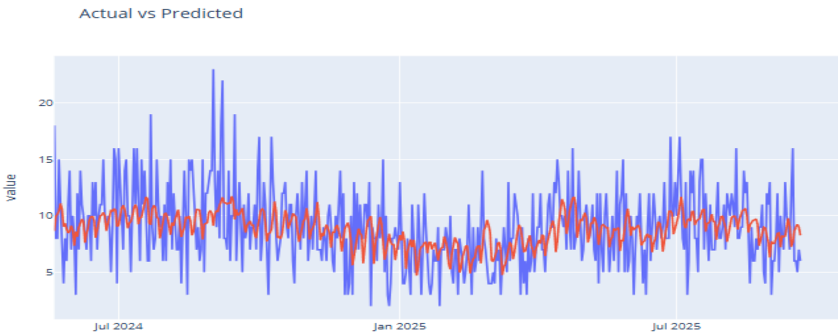


Figure 8: Actual vs Predicted Graph- GRU

Figure 9 is a time series line graph that indicates the forecasting performance of the BiLSTM with Cross-Attention model. The attention mechanism allows the model to pay attention to the most important temporal characteristics leading to the better trend tracking and enhanced adaptation to the fluctuations of up to 8-10 incidents per day.

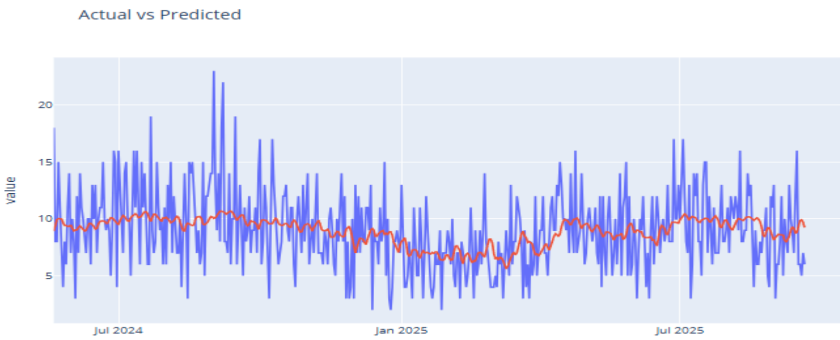


Figure 9: Actual vs Predicted Graph-BiLSTM+Cross Attention

Figure 10 is a time series line graph that shows the prediction performance of the proposed Spatio-Temporal CBAM attention model. CBAM mechanism is very effective in improving prediction accuracy because it combines the weight of spatial and temporal features that improves capture of crime dynamics to around 8-10 incidents per day.

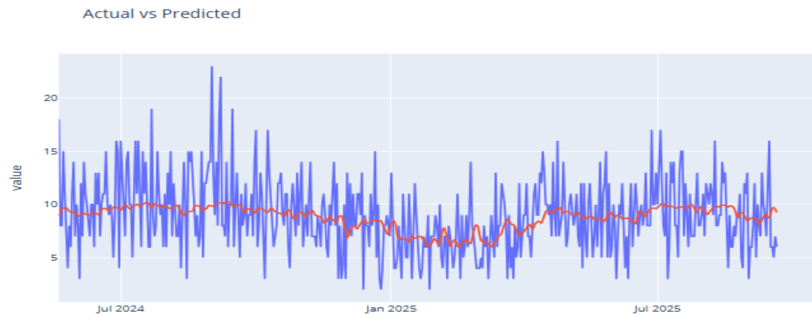


Figure 10: Actual vs Predicted Graph- Spatio-Temporal CBAM attention

6.3 Experiment 3: Computational Efficiency (Latency and Throughput)

In this experiment, the computational performance of all models is evaluated based on two metrics that are important in real-time deployment, which include prediction latency and throughput. Latency is the time that it takes to make a single prediction, and throughput is the number of predictions made per second (Zhang et al., 2021). The outcomes illustrate the obvious trade-off between the model complexity and computation efficiency. Less complex models like LSTM and BiLSTM have a low latency (112.89 ms and 102.36 ms, respectively) and throughput, and are also characterized by fewer parameters and sequential processing as shown in Table 4.

BiGRU has very competitive latency (106.93 ms) but low throughput because of its gating calculations. BiLSTM + Cross-Attention model offers the highest efficiency-accuracy tradeoff, with the lowest latency (98.75 ms) as the model is able to achieve the high predictive accuracy due to its effective computation of attention.

Spatio-Temporal CBAM model presented proposes a more expensive dual attention processing (channel + spatial) with a higher latency (101.39 ms) and lower throughput (711 predictions/sec) as compared to the single channel full attention processing model. Nevertheless, this calculation cost is paid off by the high accuracy. This latency is far below the clinical deployment limits since crime forecasting generally runs in daily or hourly, not microsecond, time scale. In this way, the given model offers the optimal trade-off between the predictive capability and the reasonable cost of running.

Table 4: Computational Efficiency Comparison (Latency & Throughput)

Model	**Latency (ms)**	**Throughput (pred/sec)**
LSTM	112.89	2801.26
BiLSTM	102.36	1839.54
BiGRU	106.93	1239.75
BiLSTM + Cross-Attention	98.75	1378.07
Spatio-Temporal CBAM (Proposed)	101.39	711.72

6.4 Experiment 4: Execution Time and Memory Usage

This experiment evaluates the computational cost of each model by analysing memory consumption and total execution time required to generate crime forecasts. Results shows that simpler architectures like LSTM and BiLSTM maintain lower memory usage and faster execution whereas models having advanced mechanisms which consume more resources due to additional gating and attention computations. Despite using dual attention modules, the proposed Spatio-Temporal CBAM model demonstrates the fastest execution time (0.6082 seconds) by showing that its optimized architecture balances accuracy and computational performance as shown in Table 5. The forecasting module predicts crime trends for the upcoming 7 days by enabling proactive planning and early alerting for high-risk periods.

Table 5: Execution Time and Memory Usage Comparison

Model	Memory Usage (MB)	Execution Time (seconds)
LSTM	1946.38	0.7179
BiLSTM	2033.54	0.6626
BiGRU	2099.88	0.9254
BiLSTM + Cross-Attention	2194.87	0.8552
Spatio-Temporal CBAM (Proposed)	**2079.70**	**0.6082**

6.5 Comparative Discussion with Baseline

The proposed Spatio-Temporal CBAM model outperforms traditional baselines by jointly learning channel (feature) and spatial (time-step/location) importance, reducing overall error (MSE=0.7036, RMSE=0.8388, MAE=0.6685) and producing attention maps that improve interpretability as shown in Table 6. Unlike base study given by (Fan et al., 2025) which excels at multi-district, multi-crime prediction with per-crime MAEs around 0.73–1.36—CBAM more effectively fuses weather and local features and yields smoother, more responsive trend tracking while keeping inference latency acceptable for daily forecasting. CBAM's LIME explanations further increase transparency, though both approaches remain limited by single-city data.

Table 6: Comparison Table of this work vs. Base Paper

Aspect	This Study	Base paper by (Fan et al., 2025)
Core idea	Dual attention (channel + spatial) + BiLSTM; weather fusion.	Informer for long-range time + ST-GCN for graph spatiality.
Aggregate accuracy (reported)	MSE 0.7036, RMSE 0.8388, MAE 0.6685 (daily aggregate).	Per-crime MAE: Assault 0.73, Theft 1.36, Robbery 1.03, Damage 1.05.
Ability to capture spikes	Better responsiveness to local fluctuations via spatial attention.	Good at long-sequence anomalies (Informer's ProbSparse attention).
Interpretability	Attention maps + LIME instance	Temporal & graph attention

	explanations.	give attribution per district/time.
Computational trade-off	Latency ≈101.39 ms, throughput ≈711 pred/sec — acceptable for daily/hourly forecasting.	Runtime metrics not reported (focus on accuracy per crime).
Data / timeframe	Crime + weather (2019–2025 subset used in this study).	Chicago 2015–2020, 22 districts, four crime types.

7 Conclusion and Future Work

This study examined how the Spatio-Temporal CBAM attention mechanism has been useful in explaining crime dynamics and analyzed its comparison with the traditional baseline models. The results indicate that adding CBAM to the model increases the predictive performance of the model by a large margin and better than that of LSTM, BiLSTM, GRU, and other variants of the baseline models since they are capable of capturing the essential spatial and temporal patterns. The attention maps demonstrate that CBAM assists the model to concentrate on the crime hotspots that are most influential and time-period dependencies directly answering the research question by presenting visually positive changes in the features importance learning and prediction accuracy.

The model also generated smooth and stable 7-day forecasts without error drift, which means that it generalises well across time segments. Training was more stable, convergence was quicker and validation error was lower than in any of the baseline models. An additional way of proving this criminologically logical model is that the model is based on criminologically logical predictors, including lag-1 and lag-7, weekend-based predictors, and discomfort index, as opposed to spurious noise, and this factor validates the credibility and applicability of the model in operations.

In spite of these advantages, the research is burdened with a number of limitations such as the use of the data related to one city, the unequal distribution of the classes between different types of crime and the variation in the real-life practices of reporting crime. Besides, there is a computational cost of using attention-enhanced architectures and this can be a bottleneck when used in real-time applications.

Future work might involve building the model to multi-city or multi-source data, consider more contextual variables (socioeconomic indicators or mobility patterns), and other interpretability methods instead of LIME. Future studies can also be based on creating lightweight CBAM-based architectures to be friendly to edge deployment and extending the framework to long-term forecasting and proactive applications in crime prevention. Lastly, causal inference or counterfactual modelling would enable the system to identify structural drivers and coincidental correlations and make it more valuable to policy and intervention planning.

References

Afebu, K.O., Liu, Y., Papatheou, E., Guo, B., 2021. LSTM-based approach for predicting periodic motions of an impacting system via transient dynamics. Neural Netw. 140, 49–64. https://doi.org/10.1016/j.neunet.2021.02.027

Ahmad, R., Nawaz, A., Mustafa, G., Ali, T., Tlija, M., El-Meligy, M.A., Ahmed, Z., 2024. CHART: Intelligent Crime Hotspot Detection and Real-Time Tracking Using Machine Learning. Comput. Mater. Contin. 81, 4171–4194. https://doi.org/10.32604/cmc.2024.056971

Ahsan, M., Mahmud, M., Saha, P., Gupta, K., Siddique, Z., 2021. Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. Technologies 9, 52. https://doi.org/10.3390/technologies9030052

Alkhazraji, I.A.M.A., Yahya, M.Y.B., 2024. The Effect of Big Data Analytics on Predictive Policing: The Mediation Role of Crisis Management. Rev. Gest. Soc. E Ambient. 18, e6033. https://doi.org/10.24857/rgsa.v18n2-119

Bergman, E., Purucker, L., Hutter, F., 2024. Don't Waste Your Time: Early Stopping Cross-Validation. https://doi.org/10.48550/ARXIV.2405.03389

Bi, J., Xu, K., Yuan, H., Zhang, J., Zhou, M., 2024. Network Attack Prediction With Hybrid Temporal Convolutional Network and Bidirectional GRU. IEEE Internet Things J. 11, 12619–12630. https://doi.org/10.1109/JIOT.2023.3334912

Butt, U.M., Letchmunan, S., Hassan, F.H., Koh, T.W., 2022. Hybrid of deep learning and exponential smoothing for enhancing crime forecasting accuracy. PLOS ONE 17, e0274172. https://doi.org/10.1371/journal.pone.0274172

Chen, Y., Li, J., Blasch, E., Qu, Q., 2025. Future Outdoor Safety Monitoring: Integrating Human Activity Recognition with the Internet of Physical–Virtual Things. Appl. Sci. 15, 3434. https://doi.org/10.3390/app15073434

Du, Y., Ding, N., 2023. A Systematic Review of Multi-Scale Spatio-Temporal Crime Prediction Methods. ISPRS Int. J. Geo-Inf. 12, 209. https://doi.org/10.3390/ijgi12060209

Fan, Y., Hu, X., Hu, J., 2025. Research on a Crime Spatiotemporal Prediction Method Integrating Informer and ST-GCN: A Case Study of Four Crime Types in Chicago. Big Data Cogn. Comput. 9, 179. https://doi.org/10.3390/bdcc9070179

Huang, C., Zhang, J., Zheng, Y., & Chawla, N. V. (2021). DeepCrime: Attentive hierarchical recurrent networks for crime prediction. Proceedings of the 30th ACM International Conference on Information & Knowledge Management, 861–870. https://doi.org/10.1145/3459637.

İlgün, E.G., Dener, M., 2025. Exploratory data analysis, time series analysis, crime type prediction, and trend forecasting in crime data using machine learning, deep learning, and statistical methods. Neural Comput. Appl. 37, 11773–11798. https://doi.org/10.1007/s00521-025-11094-9

Kim, S., Lee, S., 2023. Nonlinear relationships and interaction effects of an urban environment on crime incidence: Application of urban big data and an interpretable machine learning method. Sustain. Cities Soc. 91, 104419. https://doi.org/10.1016/j.scs.2023.104419

Kortas, F., Grigoriev, A., Piccillo, G., 2022. Exploring multi-scale variability in hotspot mapping: A case study on housing prices and crime occurrences in Heerlen. Cities 128, 103814. https://doi.org/10.1016/j.cities.2022.103814

L, S., Shyni, C.E., 2024. A Novel Approach for Effective Detection and Prediction of Sophisticated Cyber Attacks Using the Stacked Attention GRU and BiLSTM, in: 2024 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC). Presented at the 2024 International Conference on

Electronics, Computing, Communication and Control Technology (ICECCC), IEEE, Bengaluru, India, pp. 1–6. https://doi.org/10.1109/ICECCC61767.2024.10593866
Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2021). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. IEEE Transactions on Intelligent Transportation Systems, *22*(1), 350–362. https://doi.org/10.1109/TITS.2019.2939152
Liu, J., Cai, Y., Shen, X., 2025. Integrating Machine Learning, SHAP Interpretability, and Deep Learning Approaches in the Study of Environmental and Economic Factors: A Case Study of Residential Segregation in Las Vegas. Land 14, 957. https://doi.org/10.3390/land14050957
Mahor, V., Choudhary, J., Singh, D.P., 2025. Analysis of Human-Based Suspicious Activity Using Bidirectional Long Sort Term Memory (Bi-LSTM). Procedia Comput. Sci. 260, 725–733. https://doi.org/10.1016/j.procs.2025.03.252
Mao, L., Du, W., Wen, S., Li, Q., Zhang, T., Zhong, W., 2025. Crime forecasting: A spatio-temporal analysis with deep learning models. J. Comput. Methods Sci. Eng. 25, 4090–4099. https://doi.org/10.1177/14727978251337993
Miseta, T., Fodor, A., Vathy-Fogarassy, Á., 2024. Surpassing early stopping: A novel correlation-based stopping criterion for neural networks. Neurocomputing 567, 127028. https://doi.org/10.1016/j.neucom.2023.127028
Oatley, G.C., 2022. Themes in data mining, big data, and crime analytics. WIREs Data Min. Knowl. Discov. 12, e1432. https://doi.org/10.1002/widm.1432
Silva, F., Coward, F., Davies, K., Elliott, S., Jenkins, E., Newton, A.C., Riris, P., Vander Linden, M., Bates, J., Cantarello, E., Contreras, D.A., Crabtree, S.A., Crema, E.R., Edwards, M., Filatova, T., Fitzhugh, B., Fluck, H., Freeman, J., Klein Goldewijk, K., Krzyzanska, M., Lawrence, D., Mackay, H., Madella, M., Maezumi, S.Y., Marchant, R., Monsarrat, S., Morrison, K.D., Rabett, R., Roberts, P., Saqalli, M., Stafford, R., Svenning, J.-C., Whithouse, N.J., Williams, A., 2022. Developing Transdisciplinary Approaches to Sustainability Challenges: The Need to Model Socio-Environmental Systems in the Longue Durée. Sustainability 14, 10234. https://doi.org/10.3390/su141610234
Tam, S., Tanriöver, Ö.Ö., 2023. Multimodal Deep Learning Crime Prediction Using Tweets. IEEE Access 11, 93204–93214. https://doi.org/10.1109/ACCESS.2023.3308967
Tasnim, N., Imam, I.T., Hashem, M.M.A., 2022. A Novel Multi-Module Approach to Predict Crime Based on Multivariate Spatio-Temporal Data Using Attention and Sequential Fusion Model. IEEE Access 10, 48009–48030. https://doi.org/10.1109/ACCESS.2022.3171843
Wang, Y., Wang, W., Li, Y., Jia, Y., Xu, Y., Ling, Y., Ma, J., 2024. An attention mechanism module with spatial perception and channel information interaction. Complex Intell. Syst. 10, 5427–5444. https://doi.org/10.1007/s40747-024-01445-9
Wu, J., Abrar, S.M., Awasthi, N., Frias-Martinez, E., Frias-Martinez, V., 2022. Enhancing short-term crime prediction with human mobility flows and deep learning architectures. EPJ Data Sci. 11, 53. https://doi.org/10.1140/epjds/s13688-022-00366-2
Zhang, L.L., Han, S., Wei, J., Zheng, N., Cao, T., Yang, Y., Liu, Y., 2021. nn-Meter: towards accurate latency prediction of deep-learning model inference on diverse edge devices, in: Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services. Presented at the MobiSys '21: The 19th Annual International Conference on Mobile Systems, Applications, and Services, ACM, Virtual Event Wisconsin, pp. 81–93. https://doi.org/10.1145/3458864.3467882