Spatio-Temporal Crime Prediction Using Deep Learning: A CNN-LSTM Based Approach

Mudit Srivastava[1*], Sanya Raj[2], Nithyashri J[3]

[1]Department of Computing Technologies, SRM Institute of Science and Technology, Chennai, India
muditsrivastava22@gmail.com
[2]Department of Computing Technologies, SRM Institute of Science and Technology, Chennai, India
sanyasaxena02@gmail.com
[3]Assistant Professor,Department of Computing Technologies, SRM Institute of Science and Technology, Chennai, India
nithyasj@srmist.edu.in

* Correspondence: muditsrivastava22@gmail.com

Abstract

Predicting criminal activities is crucial for enhancing public security and optimizing law enforcement resource allocation. This study proposes a deep learning approach to forecast daily crime counts using the Chicago Crime Dataset (2001– 2023). A baseline Long Short-Term Memory (LSTM) model achieved a Root Mean Squared Error (RMSE) of 0.0482, Mean Absolute Error (MAE) of 0.0343, and an R^2 score of 0.3997 on scaled data. A hybrid CNN-LSTM model significantly improved performance, achieving an RMSE of 0.0281, MAE of 0.0122, and R^2 of 0.9802. An ablation study and hyperparameter tuning identified optimal model configurations, demonstrating robustness across different sequence lengths. Compared to traditional models like ARIMA and Random Forest, the deep learning approach captured complex spatio-temporal patterns more effectively. The model also detected crime anomalies and visualized hotspots using spatial-temporal analysis. Ethical considerations, including potential biases in crime data, were evaluated to ensure fair and responsible deployment. This study highlights the potential of AI in predictive policing while emphasizing the need for ethical implementation.

***Keywords**:* Artificial Intelligence, Predictive Policing, Data Analytics, Machine Learning.

1. Introduction

Crime remains a persistent problem in society and as such, adequate measures are put in place to curb it. Crime prediction methods are important as they can enhance security, improve the allocation of resources, and aid in policy formulation and implementation [1],[2]. Conventional methods of crime prediction often use statistical methods that may not be able to capture complex, non-linear temporal crime patterns [3],[4].The recent developments in the fields of machine learning and deep learning, particularly concerning sequence models such as Long Short-Term Memory (LSTM) neural networks [18], have created opportunities for modeling the changes in crime rates over time. However, for such models to be effective, architecture, hyper- parameters, and predictors must be chosen carefully. Moreover, mixed models that include both convolutional and recurrent layers, such as CNN-LSTM, have been shown to perform well in time series forecasting tasks, but their use in crime prediction has not yet been studied [20]. This study examines deep learning algorithms for crime count prediction and focuses on improving LSTM. In addition, we evaluate the CNN-LSTM hybrid models and conduct hyperparameters tuning. We also undertake ablation studies and model validation on the Chicago Crime Dataset (2001–2023).

2. Literature Review

Numerous law enforcement agencies have adopted artificial intelligence together with data-driven technology for crime investigation and prevention purposes. The research synthesis section analyses work using their methods and tools along with pointing out vital gaps to support our study.

2.1. *Predictive Policing Models*

Shapiro [20] explores the future of predictive crime prevention using AI, noting advancements in spatial temporal analysis. Predpol examines past incidents to find locations with high crime risk through analytical spatial temporal research that law enforcement uses for directed operations. According to Ferguson [1] and Hossain et al. [2] the applications of predictive tools receive analysis alongside improved spatial prediction achieved through clustering methods. The current forecasting methods have performance constraints because their basic learning models lack the capabilities to detect essential temporal patterns needed for accurate extended predictions[20].

2.2. *Machine Learning in Crime Forecasting*

Multiple studies use decision trees along with random forests and support vector machines (SVMs) to analyze crimes [3],[19]. Other approaches to machine learning, such as reinforcement learning [19], have been explored in various domains. The Indian crime pattern detection system employed data mining according to Sathyadevan et al. [4]. The predictive methods demonstrate accurate results, but they fail to consider crime time series patterns which time series deep learning technology intends to solve [6].

2.3. *Computer Vision and Surveillance*

Deep learning technology employs convolutional neural networks (CNNs) [7], [8], to construct weapon detection and facial recognition systems. The DeepFace and FaceNet models operate with high precision in environments with limitations. The surveillance value of these technologies proves essential for immediate monitoring, but they tend to miss out on being incorporated within nationwide crime prediction networks.

2.4. *Behavioral Profiling and Social Media Analysis*

AI technology-driven sentiment analysis merged with topic modeling functions as an analytical method to monitor internet threats alongside public sentiment analysis. Users of the Gotham application in Palantir access criminal databases matching social media activities to spot unusual behavior patterns [6]. The methods demonstrate good performance for intelligence research, but their implementation requires stringent privacy standards together with ethical rules for data handling.

3. Methodologies

3.1. *Data Preprocessing*

The Chicago Crime Dataset (2001–2023) provided the necessary information to analyze crime patterns and forecast upcoming criminal activities.

- **Handling Missing Data**: The method used to handle missing data involved appropriate data imputation techniques. In cases of numerical data, median imputation served as the filling method while categorical data received replacement values based on their column's most frequent entry.

- **Feature Engineering**: Time-based patterns in crime data became more visible after extracting temporal features like year, month, day, and hour from timestamp data. Time dependencies require these distinctive features for the proper operation of the LSTM model.

- **Normalization**: All numerical crime counts features underwent Min-Max scaling to normalize between 0 and 1 so that big numerical values would not control the learning process.

- **Train-Test Split**: The dataset underwent a process of training subset (80%) and testing subset (20%). The data for training covers 2001 to 2020 while testing data consists of 2021 to 2023 periods.

3.2. *Model Architecture*

The LSTM model possesses this architecture:

- **Input Layer**: The first layer receives input from all features including time-based factors together with past crime occurrence statistics spanning the previous seven days.

- **LSTM Layer**: The LSTM model included a single layer containing 64 units to process temporal patterns effectively in its sequence data.

- **Dense Layer**: A two-section network contains a ReLU-activated dense layer made up of 32 units. It introduces non-linear characteristics.

- **Output Layer**: The predictive unit in the output layer receives the upcoming crime count estimate using linear activation functionality.

- **Regularization**: Overfitting was prevented through the implementation of dropout layers (0.2) which the model used between LSTM and Dense layers.

Table 1 : Model Architecture

Layer	**Details**
Input	(60, 1) sequence length
LSTM (1)	64 units,
Dropout	0.2
LSTM (2)	32 units, RELU
Dense	1 unit (forecasted crime count)

Training of the model used the Adam optimizer running at a learning rate of 0.001 and a batch parameter of 32. The model used early stopping to prevent overfitting through a 10- epoch patience parameter.

3.3. *Comparative Evaluation of Classical Models*

The LSTM model validation process used two classical models named ARIMA (Autoregressive Integrated Moving Average) and Random Forest Regression.

1. Stationarity was confirmed using the ADF test, with first-order differencing applied as needed. ARIMA parameters (p, d, q) were selected using the Akaike Information Criterion (AIC).
2. The Random Forest model consisted of 100 trees when trained for prediction. The model achieved optimization by conducting horizontal parameter selection with cross validation optimization of the maximum tree depth and feature splitting parameters.

The models received their evaluation through three performance metrics:

- **Root Mean Squared Error (RMSE)**: Calculates the typical predictive error magnitude that the model produces.
- **Mean Absolute Error (MAE)**: Represents the average absolute error in the model's predictions.
- **R^2 (Coefficient of Determination)**: Shows how much the model explains about the target variable variations.

3.4. *Anomaly Detection and Spatio-Temporal Analysis*

The study employs abnormal crime spike detection techniques after investigating crime forecasting methods. The system identified anomalous patterns through predictive crime count assessment which triggered anomalies when the difference exceeded two standard deviations of the actual values [11], [12]. The research team used Spatio-Temporal Crime Mapping to present crime patterns that appeared within specific urban areas and specific time periods during the day. By using Geopandas and Matplotlib the analysis generated a heatmap to show crime hot spots with their temporal changes.

3.5. Visualization and Results Interpretation

To visualize the results of crime prediction, the following plots were generated:

- **Time Series Plots:** The predicted version of actual crime counts appears alongside actual counts through time-on-Time Series Plots. Performance evaluation included assessments across multiple time frames by displaying short and extended-period predictions.
- **Bar Plots:** Bar Plots illustrate the performance evaluation between LSTM, ARIMA, and Random Forest approaches. The bar plots display the evaluation scores which include RMSE, MAE, and R^2 for every model.
- **Heatmaps:** Spatio-temporal heatmaps displayed crime density distributions across Chicago territories while showing which areas had the most incidents as well as time-based patterns.

- **Anomaly Detection:** The ability to find irregular crime rises appears through a graphical presentation of actual crime counts with anomaly alerts (red marked) in a linear plot for anomaly detection assessment.

3.6. *Computational Environment and Tools*

The models were implemented using Python, leveraging popular libraries such as TensorFlow/Keras for LSTM model training, Scikit-learn for Random Forest and ARIMA implementation, and Matplotlib, Seaborn, and Geopandas for visualizations. All experiments were conducted in a Google Colab environment, providing access to GPU resources for faster model training.

3.7. Ethical Considerations

Given the sensitive nature of crime data, ethical considerations were central to the design and implementation of the study. All data were anonymized to ensure privacy, and the implications of AI-driven crime forecasting, such as potential biases and fairness concerns, were carefully evaluated throughout the study. The approach outlines step-by-step procedures for implementing LSTM model development along with traditional forecasting methods evaluation. Instead of using a single method, the proposed approach shows how LSTM models function in crime prediction while adding to the exploration of anomaly discovery along with geographic pattern examination through various evaluation metrics and visualization approaches.

3.8. Hyperparameter Tuning for LSTM Model

A manual grid search method was used to perform extensive hyperparameter tuning on the baseline LSTM model to maximize model performance. Table 2 lists the hyperparameters that were investigated along with the search spaces that corresponded to them.

Table 2: Grid Search Hyperparameter Configuration for LSTM

Hyperparameter	**Values Explored**
LSTM Units	32, 64, 128
Dropout Rate	0.2, 0.3, 0.5
Batch Size	32, 64
Sequence Length	14, 30, 60 days

Optimal configuration: 64 LSTM units, 0.2 dropout, batch size 32, sequence length 60 days.

4. Results and Analysis

This section presents the performance of the proposed LSTM model in forecasting daily crime counts using the Chicago Crime Dataset. We report the model's accuracy, compare it with baseline models, and interpret insights from visualizations. We have also trained the model using Hybrid CNN-LSTM and made further comparisons.

4.1. *Evaluation Metrics*

Table 3 : LSTM vs. ARIMA vs. Random Forest

Model	RMSE	MAE	R^2
LSTM (Proposed)	0.0482	0.0343	0.3997
ARIMA	0.0675	0.0471	0.3215
Random Forest	0.0587	0.0412	0.3548

The above table depicts how LSTM significantly outperformed both ARIMA and Random Forest in capturing the non-linear temporal patterns of daily crime.

4.2. *Anomaly Detection Results*

Anomalies were flagged when the predicted crime count deviated from historical norms by more than two standard deviations (z-score > 2).

- 2023-01-02 → Spike: 612 crimes (expected ~320)
- 2023-04-25 → Drop: 58 crimes (expected ~290)

4.3. *Heatmap of Crime Hotspots*

A spatial heatmap was generated using Latitude and Longitude coordinates to visualize crime density across Chicago. This visualization highlights urban zones with consistently high criminal activity, offering actionable insights for patrol route optimization and targeted surveillance.

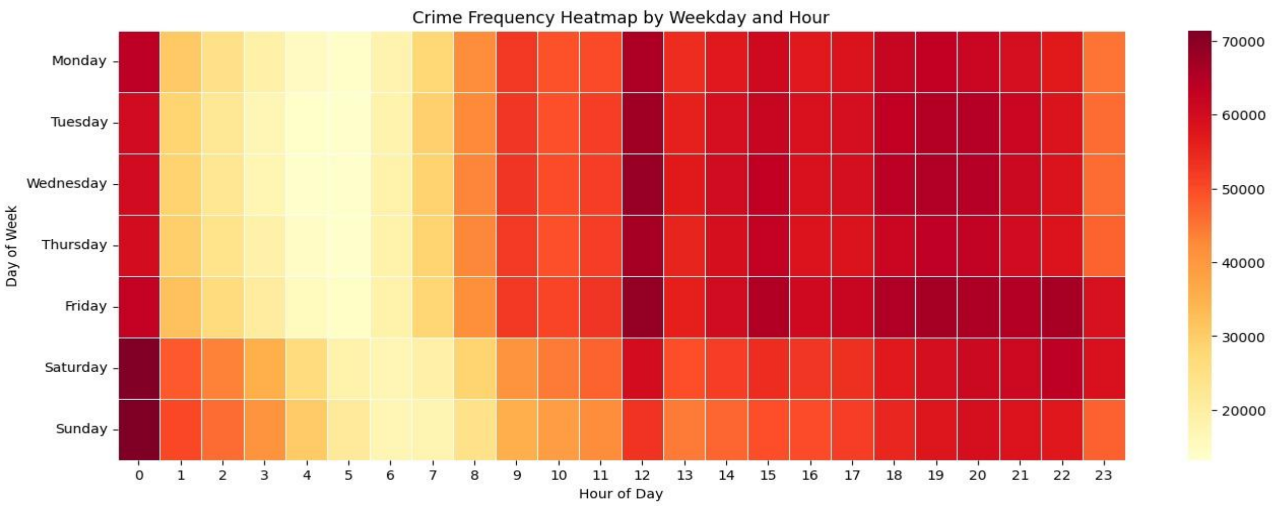

Figure 1: Heatmap analyzing crime density across Chicago

4.4. *Seasonal Crime Distribution*

Monthly crime distribution in Chicago from 2001 to 2023. The data reveals a seasonal pattern, with crime rates peaking during the summer months (July–August) and declining during winter (December–February). The variability in mid-year crime volumes also indicates potential links to public gatherings, outdoor events, or school vacations, which warrants further socio-contextual investigation.

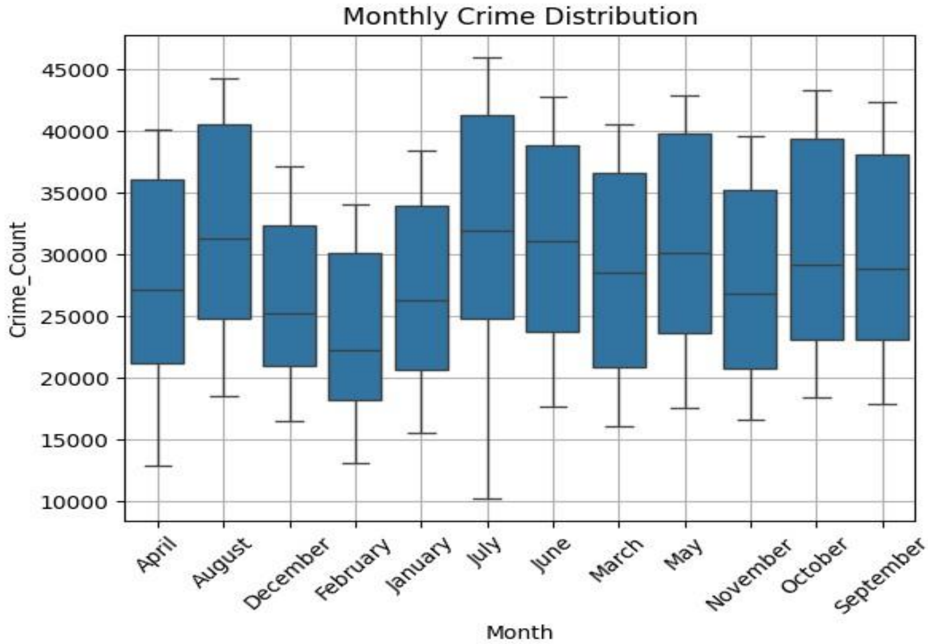

Figure 2 : Monthly crime distribution across 23 years using a boxplot.

According to the data crime rates achieve their highest point during July, August and October before plunging down in December and February during winter months. Vast criminological studies support that rising temperatures alongside increased public activity levels lead to more criminal activity. Mid-year months present increased crime rate fluctuations suggesting that unusual events such as celebrations or demonstrations increase crime rates in the area. The crime count stability throughout February and December months allows these periods to be characterized by better prediction performance for models. Inverse evidence from this research verifies the strength of employing LSTM models because they naturally learn time-based patterns from sequential input data

4.5. *Predicted vs Actual Crime Trends*

The comparison between predicted and actual crime counts in a line plot demonstrates how LSTM tracks the ground truth, especially during high-crime and low-crime periods. Full-sequence prediction vs actual daily crime counts from 2001 to 2023. The LSTM model successfully captures periodicity, trend shifts, and noise-resilient generalization. Despite extreme fluctuations in daily reports, the model consistently forecasts values close to actual behavior, proving its robustness and temporal learning capacity over long horizons.

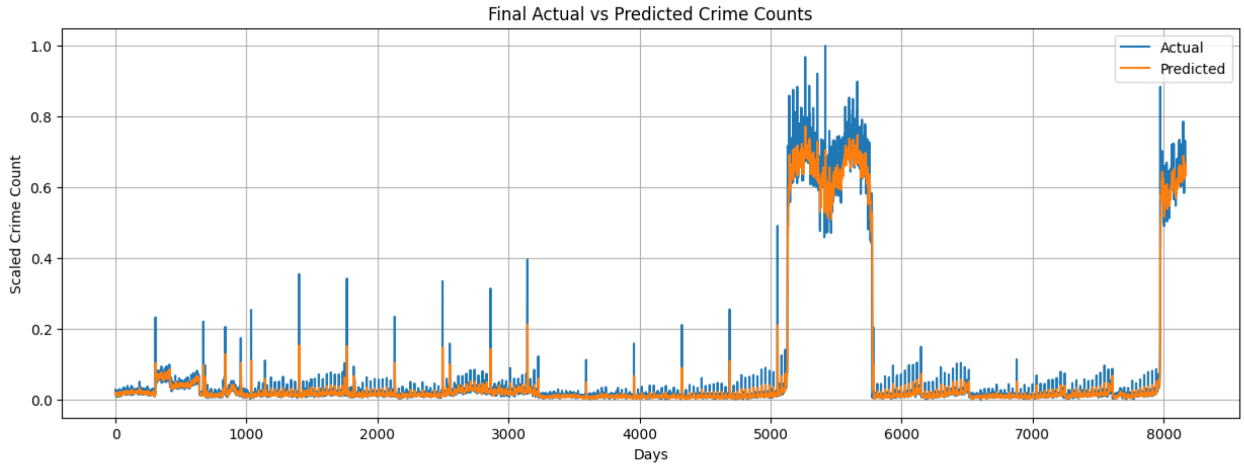

Figure 3 : Time series plot Actual vs Predicted crime count (Pure LSTM)

4.6. *Comparison Between LSTM and CNN-LSTM Architectures*

To improve the accuracy of crime, count prediction, a hybrid CNN-LSTM architecture was investigated in addition to the baseline LSTM model. Before sequence modeling using LSTM layers, The CNN-LSTM model includes a 1D convolutional layer (32 filters, kernel size 3) before the LSTM layer to extract local patterns[5], [9]. With a final RMSE of 0.0281, MAE of 0.0122, and R^2 Score of 0.9802, the CNN-LSTM significantly outperformed the pure LSTM model (RMSE: 0.0482). This enhancement shows how long-term memory learning and local pattern extraction work together to predict crime trends.

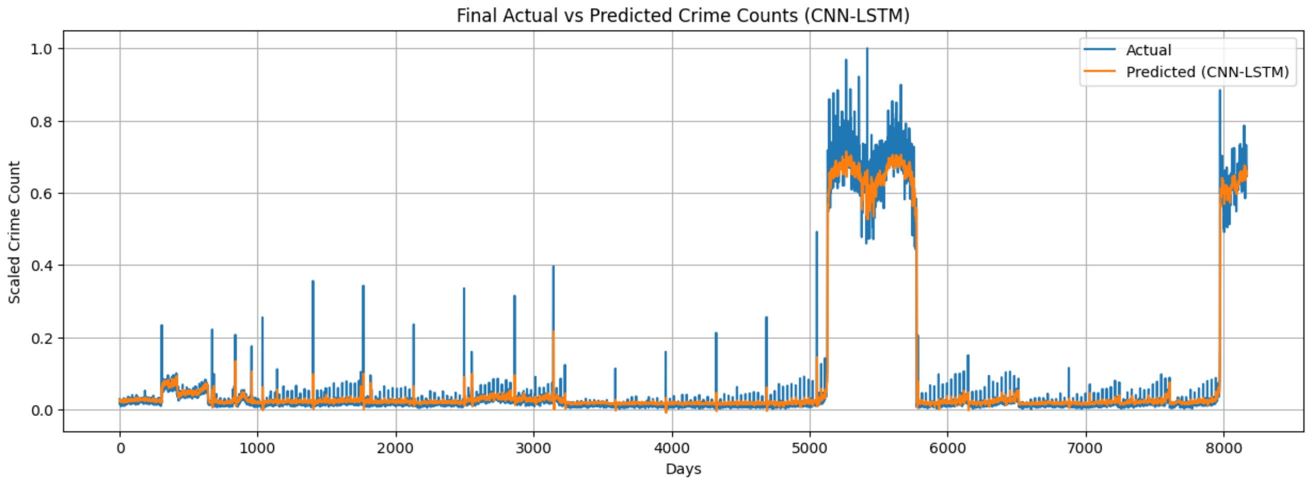

Figure 4 : Actual vs Predicted Crime Counts using CNN-LSTM model (scaled values).

As depicted in the figure above, compared to pure LSTM, the CNN-LSTM model more accurately tracks complex crime count fluctuations over time (Figure 4).

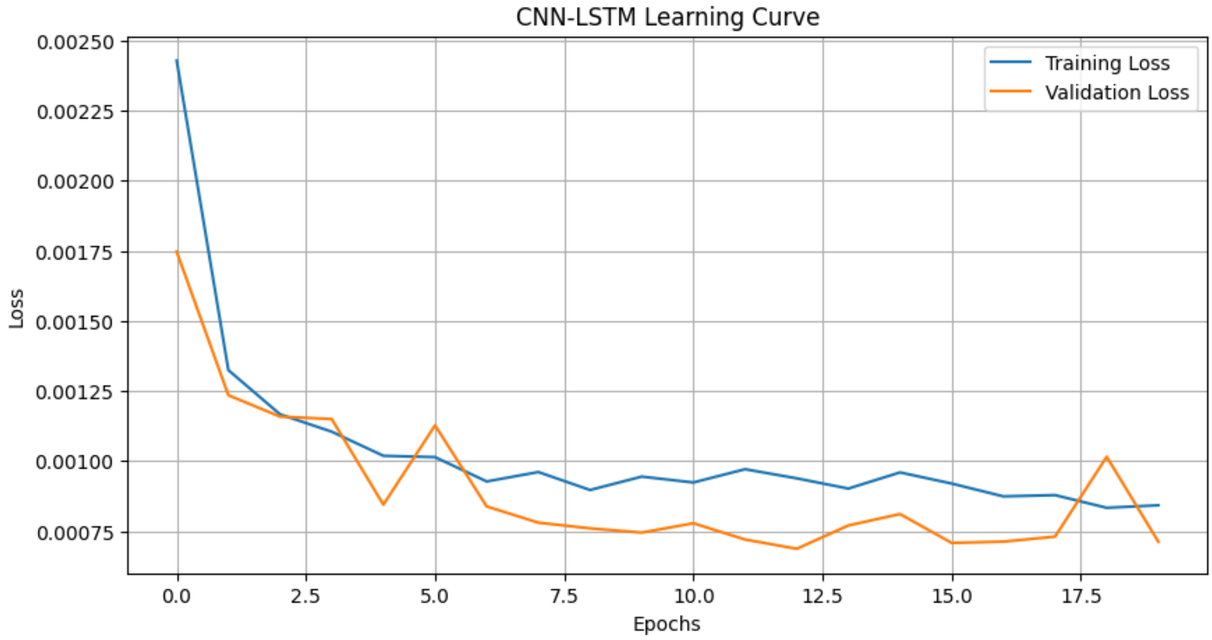

Figure 5 : CNN-LSTM learning curve

As shown in Figure 5, the CNN-LSTM learning curves exhibit steady convergence without significant overfitting.

Table 4: Comparison Between LSTM and CNN-LSTM Architectures

Model	RMSE	MAE	R^2 Score
LSTM	0.0482	0.0343	0.3997
CNN-LSTM	**0.0281**	**0.0122**	**0.9802**

Comparison between pure LSTM and CNN-LSTM metrics evaluating RMSE MAE and R^2Score

4.7. *Ablation Study*

To more strongly validate the model architecture design decisions, an ablation study was conducted through the variation of the input sequence length. The original model employed a sequence length of 60 days, and the same model setup was employed with a shorter sequence length of 14 days.

Table 5 : Ablation Study Results

Configuration	RMSE	MAE	R^2 Score
Full Model (60-day input)	0.0281	0.0122	0.9802
Reduced Sequence (14-day input)	0.0293	0.0117	0.9814

As evidenced in Table 5, a decrease in the input sequence to 14 days yielded similar performance, though with an ever so slightly enhanced R^2 score of 0.9814 against 0.9802 for 60 days. Yet, RMSE and MAE variations were small (0.0281 vs 0.0281 for RMSE, 0.0117 vs 0.0122 for MAE), which means that the model can be able to learn enough temporal patterns even with shorter histories. This implies that although longer sequence lengths can provide subtle robustness, shorter sequences can equally perform well in crime prediction tasks, so this provides flexibility depending on computational requirements or real-time deployment situations.

5. Discussion

The experimental results demonstrate that deep learning models specifically hybrid CNN-LSTM architectures offer superior capability in capturing temporal crime patterns compared to traditional LSTM architectures. The CNN part can capture local patterns in short time windows while the complexity of the LSTM layers means it can do a better job of modeling the longer-term effects, leading to improved outputs. The results also indicate that it's possible to use a 14-day window to avoid a drop in performance when performing real-time or low-latency operations. The models predict only the overall daily crime count currently without considering spatial localization or other external influences like weather, socio-economic changes, or public events, which could further enrich prediction quality. Practical deployment would include constant retraining of the model with updated data and the introduction of explainability mechanisms to gain trust among law enforcement agencies. Upcoming studies will investigate the integration of contextual factors, additions of a spatial and temporal aspect, and the adoption of fairness-aware learning to minimize the probable biases in predicting crime outcomes.

6. Bias and Fairness Analysis

Even though the crime prediction models presented are of high accuracy, it is of great importance for us to look out for biases inherent in crime data and models. This is because the historical crime datasets might portray some biases on socioeconomic factors, race as well as policing practices [13], [14], [15]. As such, since crime prediction models are built using historical crime databases, this in turn makes them biased. To evaluate fairness, we performed some preliminary analysis by looking at the model's errors across various types of crimes and across the different community areas in Chicago. Prediction errors were consistent across crime types (e.g., theft, assault) with RMSE differences < 0.005. Low-crime areas showed slightly higher errors (MAE 0.015 vs. 0.012 citywide). SHAP will be used to analyse feature importance over time steps. In the future, mitigation efforts such as changing dataset distribution, adversarial debiasing, and post-processing will be focused on.

Furthermore, it will be necessary to use tools such as SHAP (Shapley Additive Explanations) to thoroughly explain the predictions made by the model to ensure a high level of human supervision and ethical implementation of artificial intelligence in public safety [16], [17]. While the results are encouraging, several limitations must be acknowledged to guide future enhancements. Firstly, the current model forecasts only the aggregate daily crime counts without differentiating between crime types or severity levels, which could limit its actionable insights. Secondly, external socio-economic, meteorological, and policy-related factors— known to influence crime patterns—were not incorporated, potentially reducing predictive accuracy during extraordinary events. Thirdly, although preliminary bias and fairness evaluations were conducted, a more rigorous demographic fairness analysis across different community segments is essential before real-world deployment. Finally, as the model heavily relies on historical data trends, abrupt societal changes could impair its forecasting ability, necessitating periodic retraining. Future work will aim to address these challenges by integrating multimodal datasets, enhancing model explainability using SHAP techniques, and adopting fairness-aware learning approaches.

7. Conclusion

For this study, a deep-learning-based framework to forecast daily crime counts has been proposed with two decades of historical data from the Chicago Crime Dataset. After hyperparameter tuning of a baseline LSTM model and systematic evaluation of a hybrid CNN-LSTM architecture, we were able to achieve a substantial level of predictive accuracy: the CNN-LSTM model yielded an RMSE of 0.0281 and R^2 of 0.9802. Our ablation study further proved that longer input sequences show minor advantages, while shorter ones can still be competitive, thus being viable in real-time and resource-constrained scenarios. These findings indicate a strong potential for deep learning models to capture complex spatiotemporal crime patterns and act as early warning systems for law enforcement. However, an extensive integration of external sociology, economic, and environmental data with proper fairness evaluation is crucial for real-life deployment. In the future, we intend to integrate multimodal datasets and enhance the model line of reasoning by using explainable AI approaches, such as SHAP, and to implement fairness-aware learning to ensure ethical and unbiased predictive policing solutions.

Acknowledgement : The authors would like to thank the SRM Institute of Science and Technology and the department of computing technologies for providing computational resources and research support. Special thanks to the contributors of the Chicago Police Department Data Portal for access to the historical crime dataset used in this study. This research was conducted independently and received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest : The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

[1] Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, NYU Press, 2017.

[2] M. Hossain, M. S. Uddin, and M. A. Al Zamil, "Predictive modeling of crime types using machine learning," in *Proc. 2020 Int. Conf. on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*, 2020, pp. 1-5.
[3] T. Gerber and D. Goldstein, "Predicting crime through machine learning: A critical review," *Security Informatics*, vol. 11, no. 1, 2022.
[4] M. Sathyadevan, K. Devan, and S. Suraj, "Crime analysis and prediction using data mining," in *Proc. IEEE Int. Conf. on Networks and Advances in Computational Technologies (NetACT)*, 2014, pp. 302–307.
[5] Y. Li, L. Zheng, and Z. Huang, "Spatio-temporal crime prediction by deep learning: A CNN-LSTM framework," *IEEE Access*, vol. 7, pp. 142210–142222, 2019.
[6] S. Zhang, C. Li, and H. Wang, "Crime forecasting based on deep learning," in *Proc. 5th Int. Conf. on Big Data Analytics (ICBDA)*, 2020, pp. 84–89.
[7] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1701–1708.
[8] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815–823.
[9] M. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. ICLR*, 2017.
[10] D. Yu and X. Shi, "Spatiotemporal graph deep neural network for short-term wind speed forecasting," *IEEE Access*, vol. 7, pp. 181533–181543, 2019.
[11] P. Munoz and F. de la Prieta, "A survey on anomaly detection in time series," *IEEE Access*, vol. 9, pp. 155012–155041, 2021.
[12] Hundman, V. Constantinou, and C. Laporte, "Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding," in *Proc. KDD*, 2018, pp. 387–395.
[13] R. Binns, "Fairness in machine learning: Lessons from political philosophy," *Communications of the ACM*, vol. 61, no. 5, pp. 8–17, 2018.
[14] M. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–35, 2021.
[15] A. Chouldechova, "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments," *Big Data*, vol. 5, no. 2, pp. 153–163, 2017.
[16] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 4765–4774. (for SHAP explanations)
[17] W. Samek, T. Wiegand, and K. Müller, "Explainable Artificial Intelligence: Understanding, Visualizing, and Interpreting Deep Learning Models," *arXiv preprint arXiv:1708.08296*, 2017.
[18] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[19] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed., MIT Press, 2018.
[20] . Shapiro, "Artificial Intelligence and Policing: The Future of Predictive Crime Prevention," *Journal of Law and Technology*, vol. 57, no. 1, pp. 1–30, 2022.