Neural Computing and Applications

Crime Frequency Forecasting Using an Optimized Deep Learning Approach

--Manuscript Draft--

Manuscript Number:	NCAA-D-24-01345
Full Title:	Crime Frequency Forecasting Using an Optimized Deep Learning Approach
Article Type:	Original Article
Keywords:	Crime analysis; Frequency forecasting; Deep leaning; Chicago
Abstract:	Predicting crime activity and forecasting occurrence frequency has proven to be a crucial and advantageous tool. The ability to predict changes in crime activity empowers law enforcement agencies to proactively plan and allocate resources efficiently. This challenge has been approached through diverse methodologies in prior studies. In our research, we aim to utilize a Deep Learning framework supplemented with additional refinements to forecast future crime occurrence counts. In this study, we analyzed the criminal activity dataset from Chicago, known as one of the most crime-ridden cities in the US. By leveraging the Recurrent Neural Network (RNN) model and incorporating temporal context data, weather conditions, and a multi-objective hyperparameter optimization process, we achieved highly accurate predictions for a 7-day timeframe of potential future activities. The visualizations generated by our model exhibit a strong alignment with actual recorded activity trends, surpassing the performance of previous studies. Hence, our model's outputs can be deemed dependable for practical applications in real-world scenarios.

Crime Frequency Forecasting Using an Optimized Deep Learning Approach

Armin Ghayur Sadigh [1], Ali Asghar Alesheikh [1*], Aynaz Lotfata[2]

[1]*Department of Geospatial Information Systems, K. N. Toosi University of Technology, Tehran, Iran – Email: a.ghayursadigh@email.kntu.ac.ir - alesheikh@kntu.ac.ir*

[2]*Department of Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California, Davis, USA - Email: alotfata@ucdavis.edu*

***Corresponding Author:**

Ali Asghar Alesheikh

Email: alesheikh@kntu.ac.ir

Tel: +98 (21) 88786212

Fax: +98 (21) 88786213

Abstract

Predicting crime activity and forecasting occurrence frequency has proven to be a crucial and advantageous tool. The ability to predict changes in crime activity empowers law enforcement agencies to proactively plan and allocate resources efficiently. This challenge has been approached through diverse methodologies in prior studies. In our research, we aim to utilize a Deep Learning framework supplemented with additional refinements to forecast future crime occurrence counts. In this study, we analyzed the criminal activity dataset from Chicago, known as one of the most crime-ridden cities in the US. By leveraging the Recurrent Neural Network (RNN) model and incorporating temporal context data, weather conditions, and a multi-objective hyperparameter optimization process, we achieved highly accurate predictions for a 7-day timeframe of potential future activities. The visualizations generated by our model exhibit a strong alignment with actual recorded activity trends, surpassing the performance of previous studies. Hence, our model's outputs can be deemed dependable for practical applications in real-world scenarios.

Keywords: Crime analysis, Frequency forecasting, Deep leaning, Chicago

Acknowledgments

We thank the editor and the anonymous reviewers for their comments and suggestions to improve the manuscript.

Statements and Declarations

None

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Armin Ghayur Sadigh performed the computations and analysis; He drafted the manuscript and designed the figures. **Ali Asghar Alesheikh** has made contributions to interpreting results and revising the final manuscripts. **Aynaz Lotfata** has contributed to Original Draft reviewing and editing.

Ethics approval and consent to participate

This article does not contain any studies with human participants or animals performed by any of the authors.

1. Introduction

Predicting criminal activity and its frequency has long been an intriguing topic since it can offer us tools for planning, optimal resource use, being proactive, and preventing damage to lives and property. Police departments attempt to discover patterns in criminal activities; Such a problem can be correlated to many factors such as location [1], [2], time and date [3], neighborhood [4], [5], [6], regional variations [7], and crime attractors and generators [8]. In the USA, it has been estimated that the average rate of violent crime offenses by population is approximately 382 per 100,000 [9].

Previous research has uncovered diverse limitations depending on the chosen methodology. The failure of classic methods used in forecasting crime has been assumed to be due to ignoring established important factors such as weather and socio-economic variables [10], [11]. They also suffered from inability to be applicable in short time scales and ignoring the spatial aspect of the data [3]. Usage of basic regression combined with kernel density and Geographic Information System (GIS) techniques (crime hotspot detection) exhibited sub-optimal forecasting performance for sub-monthly timeframes [12], [13], [14]. Multivariate regression and careful factor selection improved results but limited sample size and coverage grid size were recognized as additional possible challenges [15]. Incorporating such datasets and additional details into the modeling process necessitates methods capable of extracting complex features and patterns.

Recent studies started developing predictive tools using machine learning (ML) approaches, as their computational costs had lowered. Among the ones that employ ML, random forest (RF), multi-layer perceptron (MLP), support vector machine (SVM), and logistic regression (LR) have been the most used algorithms [16]. The performance of artificial neural networks (ANN) was assessed for crime prediction in different timeframes [7]. The dataset was then grouped by time, and 70% of the data was used to train the model while the remaining 30% was kept for testing and validation. The results of short-term prediction in crime occurrence demonstrated low error and good performance. Some studies compared the performance of multiple methods using the same dataset to compare their predictive capabilities. The performance of numerous algorithms such as LR, SVM, Naïve Bayes, decision tree (DT), MLP, RF, and eXtreme Gradient Boosting (XGBoost) were compared using Chicago and Los Angeles crime datasets [17]. The results concluded XGBoost to be the best classifier with highest average accuracy over both datasets. Another study attempted to extract patterns in criminal data using multiple ML algorithms and histogram analysis

by dividing the samples based on their district [4]. The results exhibited that this method can be successfully used in criminal activity modeling.

Deep Learning (DL) methods, which are relatively recent compared to other commonly employed ML methods, have exhibited outstanding results in many different fields of classification problems. They are also a good choice for dealing with spatial and temporal data. As such, they have become an interesting choice for dealing with complex and multi-variate problems such as criminal activity modeling and have been used in some recent studies. One study assessed the performance of multiple DL models (VGGNet, ResNet, FastMask and FastResidualMask) using multiple crime datasets (Philadelphia, Seattle, Minneapolis, DC Metro and San Francisco) [18]. Performance results were compared with commonly used ML methods, and they concluded that DL methods consistently outperform previous methods in hotspot prediction. Another study employed the Long Short Term Memory (LSTM) model to predict emergency events using Guatemala's spatial dataset of crime occurrences [5]. The study grouped crime samples in 16 spatial clusters and used regression and binary classification to indicate count and possibility of criminal activity in each cluster. They achieved good model accuracy for the binary LSTM algorithm. A recent study examined the performance of feed forward, CNN, RNN, and RNN+CNN models to predict crime count in different sections of Chicago and Portland [3]. Their database also included census, weather, and public transportation data to improve results. The Chicago dataset was also simplified into three main crime types. Accordingly, 9 months of aggregated data was used for training, and 3 months was used for assessment. This process was repeated for multiple timeframes. They concluded that RNN+CNN had the best overall classification accuracy in all datasets and provides maximum improvement when dealing with theft crimes. A similar study used this model combination for Baltimore [6]. The crime dataset was reclassified into two main classes and the study area was divided into a $k \times k$ grid and the count of each crime type was calculated for each grid in a timeframe of multiple days. They attempted to determine the optimal grid size and timeframe length by conducting multiple tests and managed to achieve 86% classification accuracy and 77% AUC.

Considering the inherent complexity of crime forecasting, the usage of the DL models would be suitable [3]. ML methods have been the most common ones, while DL has been the least explored one [16]. This study aims to further examine the usage of optimized DL algorithms in forecasting crime occurrence frequency in a comprehensive and novel way. It will add to the

current body of research in this field by: (1) Conducting Multi-objective Optimization to ensure maximum performance, (2) Employ meteorological data to provide a more realistic model, (3) Using a novel time embedding method and adding a temporal aspect to the data, (4) Using an extended 7-day prediction timeframe, and (5) Providing daily time resolution. We have rigorously crafted our model development process, adhering to strict standards, and achieved improved results compared to previous studies conducted in the same area. We have selected Chicago for this study as it is one of the most crime-ridden cities in the USA. In 2020 its crime rate was higher than 94.7% U.S. cities, ranking it at 3rd place among major American cities [19]. Using Chicago's public criminal activity dataset also allows us to compare model performance and results with previous studies and conclude whether the implemented data and methodology of this research provides more accurate and reliable results.

2. Materials and Methods

The generated and pre-processed dataset is utilized for constructing a model with the objective of forecasting the expected number of future crime occurrences. Detailed explanations of each phase and its components will be provided in the following subsections.

2.1. Datasets

The data used in this study has been gathered from multiple public sources, provided by different agencies. Our database is comprised of two crucial datasets: (1) Crime occurrence dataset, which is the outcome of the model, and (2) Factors which have been shown to affect crime activity. We have chosen Chicago as our study area, and its crime activity dataset is publicly available online on City of Chicago's Data Portal [20]. It includes detailed reports of criminal activity between 2001 and 2022 and has approximately 7.8 million records. This dataset will be used to calculate the total counts of crime instances recorded each day.

The second important part of our database is a collection of datasets for factors that influence the likelihood and frequency of crime occurrence. These factors have been chosen based on previous studies' results, connecting crime likelihood with specific factors. Meteorological data has been shown to be an important factor in crime activity [10], [21], [22], [23]. Therefore, we have gathered temperature maximum and minimums, precipitation data, and snowfall data for the study area. Temperature and precipitation data has been gathered from the PRISM dataset [24],

and the snowfall data has been sourced from The Global Historical Climatology Network daily (GHCNd) dataset [25]. In this study, an additional factor introduced for the first time as raw input data is the date variable, which is being incorporated into the analysis. The year, month, and day of each factor measurement is used as input in the model to provide additional context and a way to simulate the trend of crime activity over the years.

The data is then pre-processed. Total crime occurrence count is aggregated by day and used as the outcome result of the model. Temperature, precipitation, and snowfall are then normalized across the whole dataset's range of values, and finally, the time data is encoded using a sin-cos method. The year parameter is normalized and stored separately, but month and day values are normalized to the range of 0-2π, and sin and cos values of this normalized result are stored in the final dataset. This allows us to employ the date in our modeling process, enabling the model to understand similarities between specific timeframes and dates (e.g., beginning days of month, last months of year). It preserves the cyclic nature of months and days, providing a continuous and sequential data factor to the model.

The generated dataset is then divided into three subsets, namely the training, validation, and testing sets. In this study, we used the 2015-2019 timeframe's data as the training subset, 2020 as the validation subset, and 2021 as the testing dataset. The usage of a testing dataset alongside a validation dataset is mainly to conduct multiple-objective optimization and assess model performance using two datasets to ensure that the model is not only robust, but also has not overfitted to the data.

2.2. Methods

In deep learning applications, the RNN model has been widely used for sequence processing and forecasting. Using a sequence of input data, it can predict the most likely values that follow this sequence. In this study, we generate sequences of 30-day input and 7-day output timeframes. The values of impact factors are used as input values, and the output is the crime occurrence count of the next 7 days. This data is then used to develop a model. To ensure maximum performance, stability, and robustness of the final model, we combine this phase with an optimization process. To account for model performance using validation and testing datasets at the same time, multi-objective optimization method is conducted and a model with optimal performance using both datasets is selected as the final model.

2.2.1. Numerical modeling with RNN

The RNN model is a network comprised of multiple layers of artificial neurons (nodes) which passes data from one layer to another, altering its representation in a way that establishes a connection between the input and output values. RNN nodes also have recursive connections inside them which act similar to memory or feedback. As a result, each calculation performed in each node will be affected by previous calculations; This functionality allows this model to deal with temporal or sequential data better than other deep learning algorithms. Since the temporal aspect of our database is expected to play a significant role in testing performance improvements, RNN became the model selected as the basis of our model. This article briefly explains this model, and the readers can refer to other academic literature which have explained this model with more detail and provide its equations [26], [27], [28].

2.2.2. Hyper parameter optimization

The RNN model can be constructed in different architectures, comprised of different number of layers, nodes in each layer, and node behavior parameters. Additionally, it can be trained using different strategies which will lead to various levels of performance. As there are vast numbers of possible combinations of these parameters and testing this large number of instances is not time-efficient, a meta-heuristic method of searching for an optimal combination should be used instead. In this study, we have employed an optimization tool to test different combinations of these parameters in a proper manner and find an optimal model architecture and learning strategy in an efficient manner. This ensures that the used algorithm to construct the model reaches its full potential and a reliable model and result will be achieved.

For this purpose, a Python library named ‘Optuna’ has been used to perform a set of trials for the model [29]. Tree-structured Parzen Estimator Approach is employed using this library to explore the search space [30]. By recording performance results of different hyper-parameter combinations, we can also analyze the importance of different hyper-parameters in the model which helps us increase our understanding of the RNN model and its behavior. For this purpose, we use an improved version of Analysis of variance (ANOVA) framework, called functional ANOVA (fANOVA) framework [31]. It fits a random forest regression model that predicts the objective values of finished trials given their parameter configurations.

The optimal model we aim to develop must exhibit optimal performance using two datasets, namely the validation and test datasets. Goodness of fit is the goal of the model, which will be calculated using the Mean Absolute Percentage Error (MAPE). The performance metrics calculated for these trials usually conflict with each other. When comparing two trials, one trial may have higher MAPE using the validation set, but lower MAPE using the testing set. As such, no single absolute optimal solution can be determined in such cases with respect to the MAPE objective value. Instead, a set of optimal solutions can be determined for this problem. This set of optimal solutions is known as Pareto optimal solutions or Pareto front [32]. After determining this optimal set of models, they can be manually reviewed, and one trial can be selected as the optimal solution.

2.2.3. Performance assessment

After developing the model, its predictive performance using the validation and testing datasets will be further assessed using regression assessment metrics. For a comprehensive assessment, Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) will be used as they're commonly used in similar research.

MAPE is used to calculate an average percentage-based error and can be calculated using the formula below:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - P_i}{A_i}\right|$$

where A_i is the actual value, P_i is the predicted value, and n is the sample count. This formula calculates the relative error for each sample and calculates their averages. Multiplying it by 100 provides a percent number.

RMSE is also another commonly used metric in regression model assessment. This metric aggregates the magnitude of prediction errors and can be interpreted as a metric of accuracy as well by showing a better goodness of fit. This metric can be calculated using this equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - P_i)^2}$$

where A_i is the actual value, P_i is the predicted value, and n is the sample count.

The final metric to be calculated for assessment is the MAE. The equation for this metric is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|A_i - P_i|$$

where A_i is the actual value, P_i is the predicted value, and n is the sample count. While it's not as commonly used as the other metrics mentioned before, some researchers have recommended the use of MAE instead of the RMSE; It has more interpretability than RMSE as it's the average of absolute errors and easier to understand. Additionally, while prediction errors affect MAE in direct proportion to their absolute value, RMSE is affected by large errors more easily. This means that outlier samples and incorrect observations affect MAE less than RMSE.

3. Results

3.1. Multi-objective optimization results

The prediction model's optimization aimed to minimize validation and testing MAPE error, and with 100 iterations of model development, the results of their evaluation metric can be plotted. These results can be seen in Figure 1, where each dot represents one of the fully developed models using a specific combination of hyperparameters. A subset of these models creates the Pareto front in this plot and can be considered as the better models. These samples are shown as dots with a red accent and color intensity indicates superiority in terms of optimization metric criteria defined for the process.

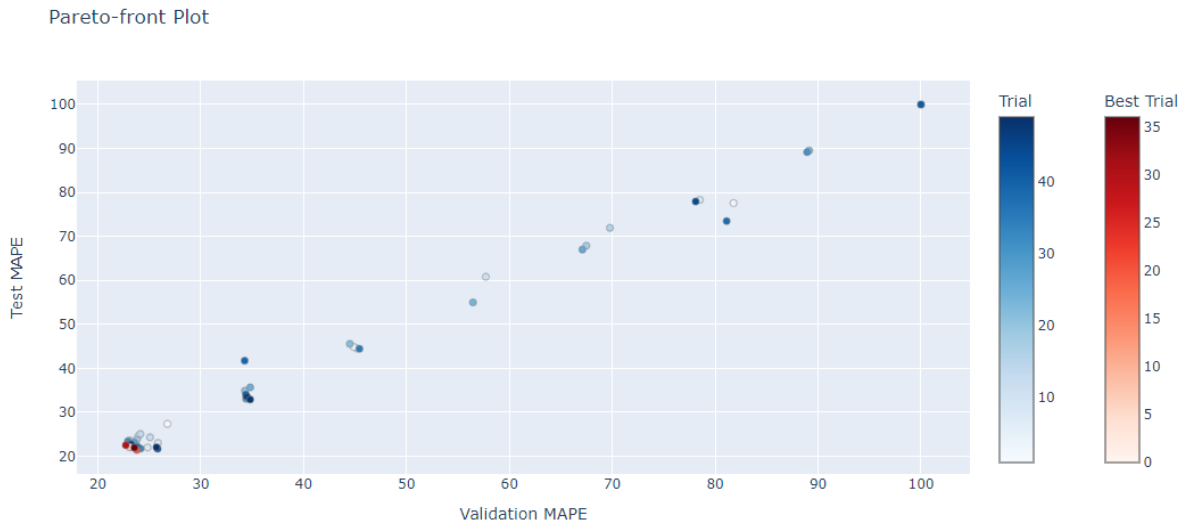

Figure 1 Pareto Front Multi-Objective optimization results

Upon further inspection of this subset, the model that meets our expectations in terms of goodness of fit to both objectives is selected. The detailed combination of hyperparameters used for this model has been summarized in **Table *1***.

Table 1 Hyper parameter list and their values for the optimal model

Hyperparameter	Optimal model value
LSTM layer count	3
Neuron count in LSTM layers	57-71-56
Dense layer count	1
Neuron count in Dense layers	10
Dense layer activation functions	Linear
Final dense layer activation function	Linear
Learning rate	1.714e-5
Batch size	35

3.2. Performance assessment results

To ensure the stability of the selected model and its convergence, the changes in its loss value must be reviewed after the training process has concluded. Recorded values per epoch for all metrics are used to create a plot of their progression, which can be seen in Figures 2-4. The graphs

show minimal magnitudes of fluctuation after convergence which indicates stable convergence and no risk of overfitting to any of the datasets.

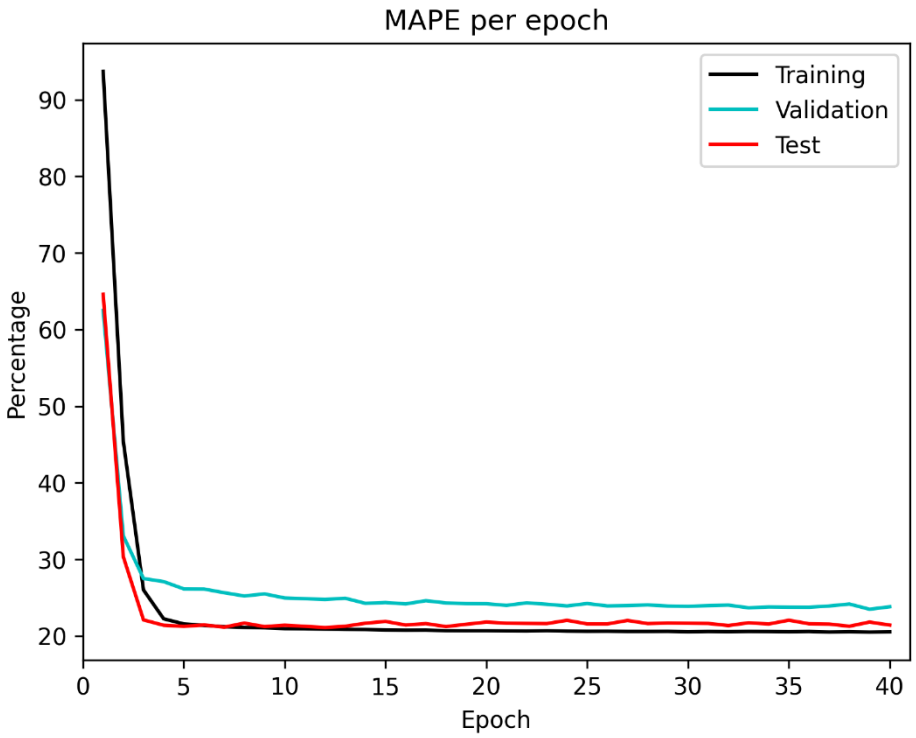

Figure 2 MAPE metric progression per epoch

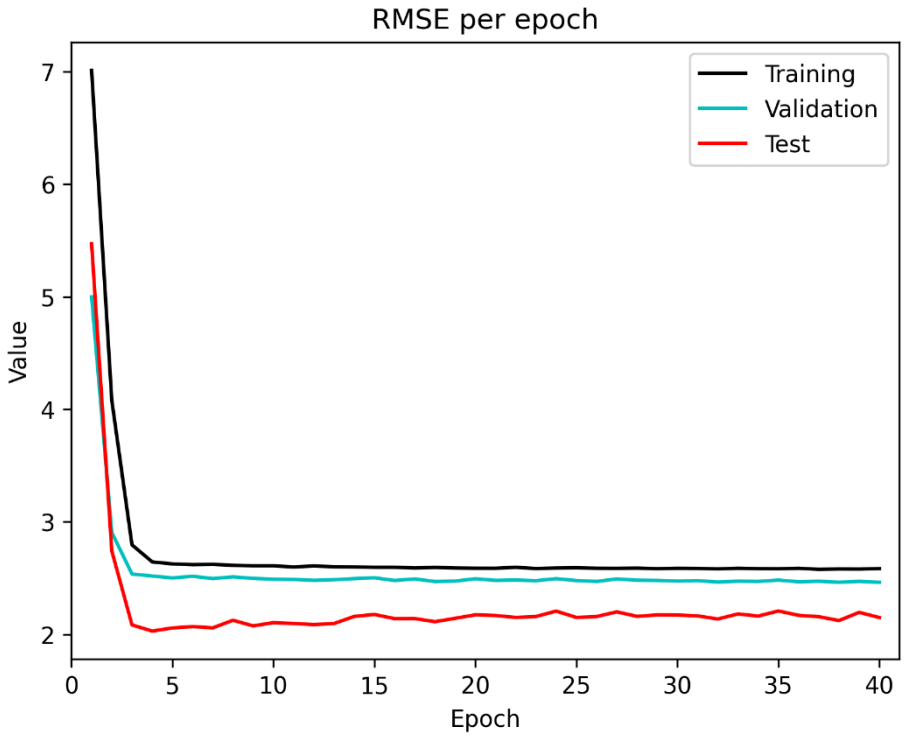

Figure 3 RMSE metric progression per epoch

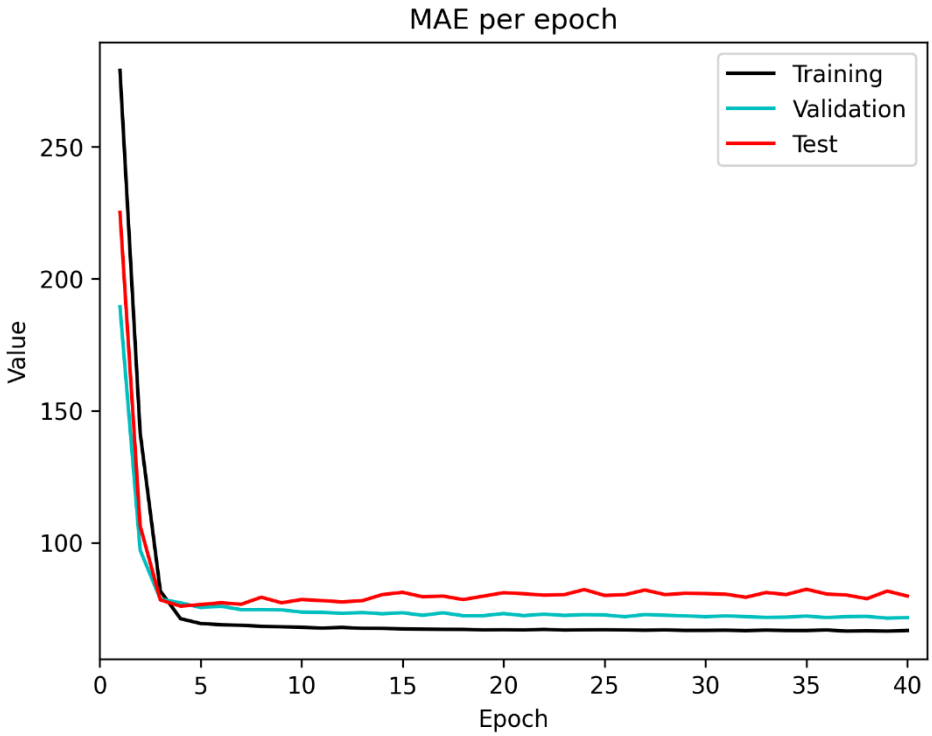

Figure 4 MAE metric progression per epoch

After being ensured of model stability, its performance is evaluated using the training, validation, and testing datasets. For each dataset, all three metrics mentioned before (i.e., MAPE, MAE, RMSE) are calculated for a comprehensive assessment. **Table *2*** summarizes the results of performance metric assessments.

Table 2 Final metric values for each dataset

Metric	***Dataset***		
	Training	***Validation***	***Testing***
MAPE	*20.54*	*23.80*	*21.41*
RMSE	*2.58*	*2.46*	*2.15*
MAE	*66.72*	*71.58*	*79.77*

3.3. Prediction Visualization

With optimal predictive performance being achieved, the model can now be used to predict future crime occurrence count. Using 30 days of historical data, the next 7 days are forecasted, resulting in at least 1 and at most 7 predicted values for each day in the validation and test timeframes; The average value for each day is calculated using the predicted values, resulting in one final mean predicted occurrence count for each day. **Figure *5*** (the complete timeframe) and **Figure *6*** (validation and testing timeframes only) exhibit the result of this process; The red line indicates predicted occurrence, and the blue line indicates true values.

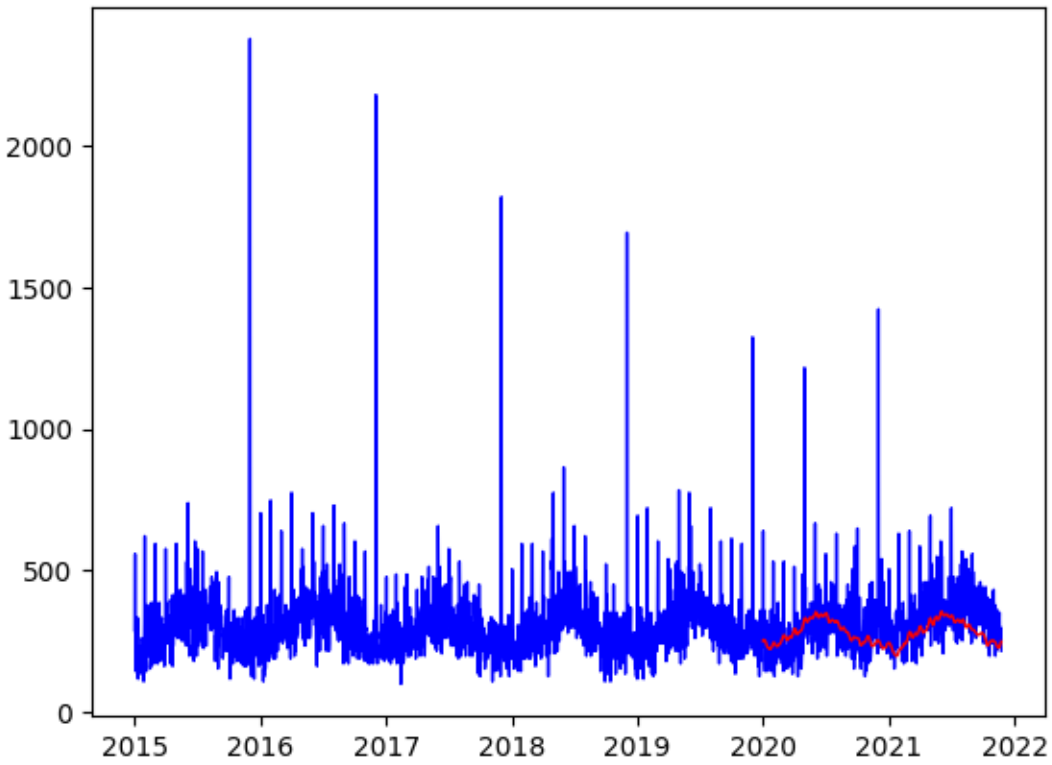

Figure 5 Model prediction results (entire data timeframe)

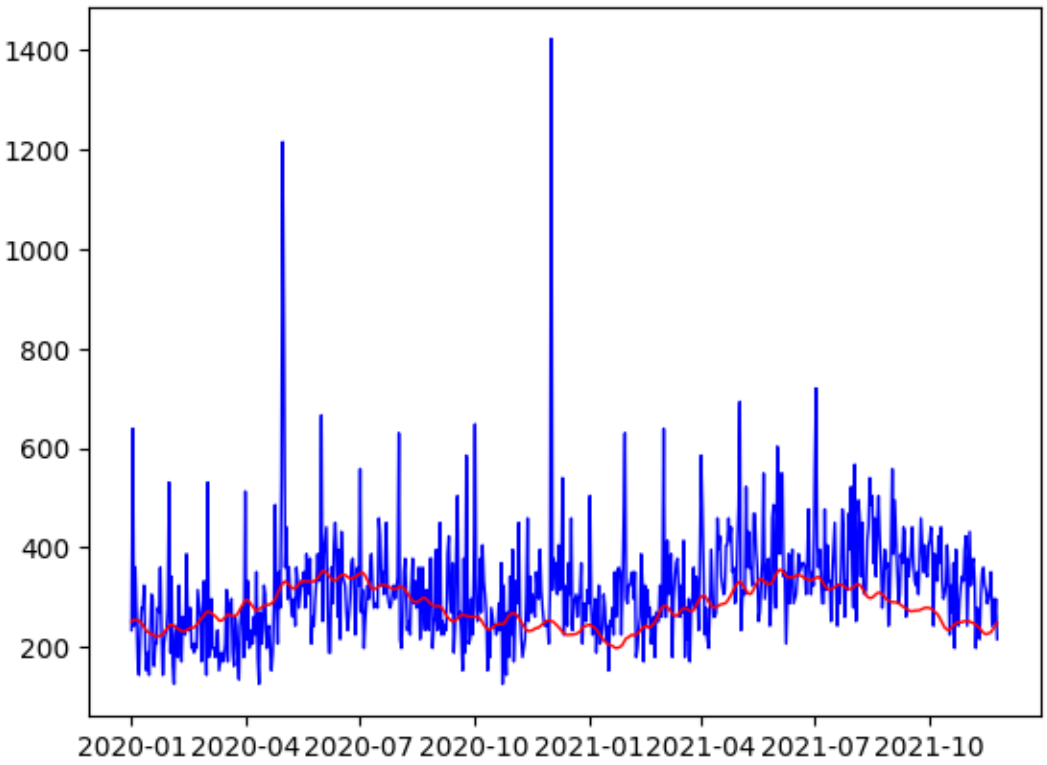

Figure 6 Model prediction results (validation and testing timeframe only)

4. Discussion

4.1. Results

This study employs RNN to forecast crime frequency in the city of Chicago. In addition to previous methodologies used in similar studies, the training process in this study aimed to maintain a more reliable training phase. This resulted in achieving accurate results while ensuring that the model does not suffer from possible shortcomings such as overfitting or loss of performance potential. Comparison of results with similar studies show good predictive performance metrics considering the output timeframe and the dataset's characteristics.

Analysis of hyperparameter importance using fANOVA algorithm also indicated that the hyperparameters with the most significance in model's predictive performance were the activation function of the final dense layer, and the learning rate. This held true for both objectives of the optimization process (i.e., validation MAPE and testing MAPE). The rest of the hyperparameters such as neuron count in each LSTM layer, neuron count of the hidden dense layers, and batch size

were not significantly impactful. The importance of each hyperparameter determined during this process can be seen in **Table *3***.

Table 3 Importance analysis for hyperparameters used in Multi-objective optimization

Hyperparameter	***Importance factor for optimization objective***	
	Validation dataset MAPE	***Testing dataset MAPE***
Final Dense layer activation function	*0.48*	*0.65*
Learning rate	*0.33*	*0.21*
Neuron count for LSTM layers (average)	*0.04*	*0.02*
All other hyperparameter	*=<0.02*	*=<0.02*

Our model simulates the behavior of Chicago's crime occurrence to a very good extent compared to some previous studies. Modeling Chicago's crime dataset has proven to be more difficult compared to other cities (e.g., San Francisco, Philadelphia) as studies using the same methodology across multiple datasets have all achieved lower performance modeling Chicago than other cities [6], [17], [33]. The predicted graph, however, cannot properly model the extreme peaks that occasionally occur. Nonetheless, as every study using this dataset to this date has shown minimal capability of doing said task, this can be considered a negligible shortcoming and a limitation to the fundamental nature of statistical modeling.

Performance metric comparison is limited by the fact that most studies develop classification models while we have developed a regression model. MAPE value achieved using our model is superior to results achieved by Borowik [34], exhibiting lower MAPE error compared to the majority of crime categories modeled in their study (metrics ranging from 10% to 60%). RMSE achieved here also shows similar relation with Jin's study conducted in San Francisco, as their study simplifies the dataset by reclassifying crime events into four major groups [35]. The same comparison can be seen with Wang's study which seems to use the exact same dataset and methodology as Jin's [36]. They follow a very similar methodology to Jin and achieve similar RMSE metrics, which are better than ours in some cases, and worse in others. The RMSE and MAE results achieved in this study surpass those of Safat's to a good extent as well [17] and

indicate improvements over their usage of the LSTM model. Stec also used multiple DL models to predict Chicago's crime activity, both in general and with crime classification.

4.2. Limitations and recommendations for future studies

Future attempts at improving the current scientific body in criminal activity forecasting could provide improvements by combining this study's novelties with recommendations provided in this section. Firstly, comprehensive pre-processing and filtering of the data could possibly provide better results and could be explored. This process could include reclassification of crime types, smoothing frequency graphs, or any similar technique used in previous studies. Considering the large timeframe and high resolution of the data employed in this study, this improvement was not explored in this study. With all current optimization steps introduced in our current methodology, an additional optimization phase would have exponentially increased the amount of time, memory, and power required for model development and assessment. Additionally, implementation of a spatial aspect in the input and output of our process could improve the clarity of the data provided by indicating the most likely regions to experience criminal activity at a specific time in the future. It's also possible that it could help with performance improvement, as some additional datasets commonly used in previous studies had a spatial nature (e.g., census information, proximity to public transportation).

5. Conclusion

In this study, the frequency of crime occurrence in Chicago was modeled using the RNN model. To improve the process compared to previous studies, multiple additional steps were introduced to the methodology: (1) A multi-objective optimization process for hyperparameter optimization, (2) Addition of encoded date variable to the input data, and (3) Usage of meteorological data as additional covariates. Task complexity was also elevated by: (1) Extending the prediction timeframe to 7 days, and (2) Usage of daily data resolution. Despite the added challenges and increased difficulty of the modeling task, we managed to achieve similar or superior performance results compared to similar previous studies. Based on the results and conclusions made in this study, it can be said with high confidence that the basis of our study provides an accurate and reliable framework for crime frequency forecasting. With further investigation in

future studies, this method can possibly improve even further, and provide decision makers with an impactful tool in optimal planning and resource management.

Conflict of Interest

The authors submitting this manuscript have no conflict of interest to declare.

Data Availability

The Crime incidents data used in this study is publicly available at the URL: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2/about_data [20]. The snowfall dataset has been resourced from the Global Historical Climatology Network daily (GHCNd) public dataset available at the URL: https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily [25]. Temperature data has been obtained from the PRISM dataset publicly available at the URL: https://prism.oregonstate.edu/ [24].

References

[1] K. Kianmehr and R. Alhajj, "Crime Hot-Spots Prediction Using Support Vector Machine," in *IEEE International Conference on Computer Systems and Applications, 2006.*, IEEE, 2006, pp. 952–959. doi: 10.1109/AICCSA.2006.205203.

[2] C.-H. Yu, M. W. Ward, M. Morabito, and W. Ding, "Crime Forecasting Using Data Mining Techniques," in *2011 IEEE 11th International Conference on Data Mining Workshops*, IEEE, Dec. 2011, pp. 779–786. doi: 10.1109/ICDMW.2011.56.

[3] A. Stec and D. Klabjan, "Forecasting Crime with Deep Learning," Jun. 2018.

[4] P. Fandino and J. Tan Jr, "Crime Analytics: Exploring Analysis of Crimes Through R Programming Language," *International journal of simulation: systems, science & technology*, Jul. 2019, doi: 10.5013/IJSSST.a.20.S2.29.

[5] B. Cortez, B. Carrera, Y.-J. Kim, and J.-Y. Jung, "An architecture for emergency event prediction using LSTM recurrent neural networks," *Expert Systems with Applications*, vol. 97, pp. 315–324, May 2018, doi: 10.1016/j.eswa.2017.12.037.

[6] N. Esquivel, O. Nicolis, B. Peralta, and J. Mateu, "Spatio-Temporal Prediction of Baltimore Crime Events Using CLSTM Neural Networks," *IEEE Access*, vol. 8, pp. 209101–209112, 2020, doi: 10.1109/ACCESS.2020.3036715.

[7] Z. M. Wawrzyniak *et al.*, "Data-driven models in machine learning for crime prediction," in *2018 26th International Conference on Systems Engineering (ICSEng)*, IEEE, Dec. 2018, pp. 1–8. doi: 10.1109/ICSENG.2018.8638230.

[8] J. B. Kinney, P. L. Brantingham, K. Wuschke, M. G. Kirk, and P. J. Brantingham, "Crime Attractors, Generators and Detractors: Land Use and Urban Crime Opportunities," *Built Environment*, vol. 34, no. 1, pp. 62–74, Apr. 2008, doi: 10.2148/benv.34.1.62.

[9] Federal Bureau of Investigation Crime Data Explorer, "National crime trend." Accessed: Oct. 10, 2023. [Online]. Available: https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/explorer/crime/crime-trend

[10] E. G. COHN, "WEATHER AND CRIME," *The British Journal of Criminology*, vol. 30, no. 1, pp. 51–64, 1990, doi: 10.1093/oxfordjournals.bjc.a047980.

[11] R. Allan, "Reviews : Women, Crime and Poverty Pat Carlen Open University Press, 1988," *Probation Journal*, vol. 36, no. 1, pp. 29–30, Mar. 1989, doi: 10.1177/026455058903600112.

[12] John Eck, Spencer Chainey, James Cameron, and Ronald Wilson, "Mapping Crime: Understanding Hot Spots.," *National Institute of Justice special report*, 2005.

[13] S. Chainey, L. Tompson, and S. Uhlig, "The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime," *Security Journal*, vol. 21, no. 1–2, pp. 4–28, Feb. 2008, doi: 10.1057/palgrave.sj.8350066.

[14] J. Chae *et al.*, "Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition," in *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, IEEE, Oct. 2012, pp. 143–152. doi: 10.1109/VAST.2012.6400557.

[15] W. L. Gorr and A. Olligschlaeger, "Crime Hot Spot Forecasting with Data from the Pittsburgh [Pennsylvania] Bureau of Police, 1990-1998," *Inter-university Consortium for Political and Social Research [distributor]*, 2015, doi: 10.3886/ICPSR03469.v1.

[16] O. Kounadi, A. Ristea, A. Araujo, and M. Leitner, "A systematic review on spatial crime forecasting," *Crime Science*, vol. 9, no. 1, p. 7, Dec. 2020, doi: 10.1186/s40163-020-00116-7.

[17] W. Safat, S. Asghar, and S. A. Gillani, "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques," *IEEE Access*, vol. 9, pp. 70080–70094, 2021, doi: 10.1109/ACCESS.2021.3078117.

[18] P. Stalidis, T. Semertzidis, and P. Daras, "Examining Deep Learning Architectures for Crime Classification and Prediction," *Forecasting*, vol. 3, no. 4, pp. 741–762, Oct. 2021, doi: 10.3390/forecast3040046.

[19] City-Data, "Crime in Chicago, Illinois (IL)." Accessed: Oct. 10, 2023. [Online]. Available: https://www.city-data.com/crime/crime-Chicago-Illinois.html

[20] Chicago Data Portal, "Crimes - 2001 to Present | City of Chicago." Accessed: Oct. 13, 2023. [Online]. Available: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2/about_data

[21] J. Corcoran and R. Zahnow, "Weather and crime: a systematic review of the empirical literature," *Crime Science*, vol. 11, no. 1, p. 16, Dec. 2022, doi: 10.1186/s40163-022-00179-8.

[22] J. Tiihonen, P. Halonen, L. Tiihonen, H. Kautiainen, M. Storvik, and J. Callaway, "The Association of Ambient Temperature and Violent Crime," *Sci Rep*, vol. 7, p. 6543, Jul. 2017, doi: 10.1038/s41598-017-06720-z.

[23] A. J. Sommer, M. Lee, and M.-A. C. Bind, "Comparing apples to apples: an environmental criminology analysis of the effects of heat and rain on violent crimes in Boston," *Palgrave Commun*, vol. 4, no. 1, Art. no. 1, Nov. 2018, doi: 10.1057/s41599-018-0188-3.

[24] PRISM Climate Group at Oregon State University, "PRISM dataset." Accessed: Oct. 13, 2023. [Online]. Available: https://prism.oregonstate.edu/

[25] "Global Historical Climatology Network daily (GHCNd)." Mar. 01, 2021. Accessed: Oct. 14, 2023. [Online]. Available: https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily

[26] P. Kim, "Deep Learning," in *MATLAB Deep Learning*, Berkeley, CA: Apress, 2017, pp. 103–120. doi: 10.1007/978-1-4842-2845-6_5.

[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[28] S. Mahdavifar and A. A. Ghorbani, "Application of deep learning to cybersecurity: A survey," *Neurocomputing*, vol. 347, pp. 149–176, Jun. 2019, doi: 10.1016/j.neucom.2019.02.056.

[29] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2623–2631, Jul. 2019.

[30] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for Hyper-Parameter Optimization," in *Advances in Neural Information Processing Systems*, J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2011. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf

[31] F. Hutter, H. Hoos, and K. Leyton-Brown, "An Efficient Approach for Assessing Hyperparameter Importance," in *Proceedings of the 31st International Conference on Machine Learning*, PMLR, Jan. 2014, pp. 754–762. Accessed: Oct. 14, 2023. [Online]. Available: https://proceedings.mlr.press/v32/hutter14.html

[32] M. Akbari, P. Asadi, M. K. Besharati Givi, and G. Khodabandehlouie, "13 - Artificial neural network and optimization," in *Advances in Friction-Stir Welding and Processing*, 2014, pp. 543–599. doi: 10.1533/9780857094551.543.

[33] M. Feng *et al.*, "Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data," *IEEE Access*, vol. 7, pp. 106111–106123, 2019, doi: 10.1109/ACCESS.2019.2930410.

[34] G. Borowik, Z. M. Wawrzyniak, and P. Cichosz, "Time series analysis for crime forecasting," in *2018 26th International Conference on Systems Engineering (ICSEng)*, IEEE, Dec. 2018, pp. 1–10. doi: 10.1109/ICSENG.2018.8638179.

[35] G. Jin, Q. Wang, X. Zhao, Y. Feng, Q. Cheng, and J. Huang, "Crime-GAN: A Context-based Sequence Generative Network for Crime Forecasting with Adversarial Loss," in *2019 IEEE International Conference on Big Data (Big Data)*, IEEE, Dec. 2019, pp. 1460–1469. doi: 10.1109/BigData47090.2019.9006388.

[36] Q. Wang, G. Jin, X. Zhao, Y. Feng, and J. Huang, "CSAN: A neural network benchmark model for crime forecasting in spatio-temporal scale," *Knowledge-Based Systems*, vol. 189, p. 105120, Feb. 2020, doi: 10.1016/j.knosys.2019.105120.

Cover letter

Click here to access/download
Supplementary Material
coverletter.docx