A comparative analysis of Deep Learning Models for Spatio-Temporal Crime Prediction in Chicago

MSc Research Project
Data Analytics

Yug Kathiriya
Student ID: 22187839

School of Computing
National College of Ireland

Supervisor: Jorge Basilio

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	YUG KATHIRIYA
Student ID:	x22187839
Programme:	Data Analytics
Year:	2023
Module:	Msc Research Project
Supervisor:	Jorge Basilio
Submission Due Date:	12/08/2024
Project Title:	A comparative analysis of Deep Learning Models for Spatio-Temporal Crime Prediction in Chicago
Word Count:	6944
Page Count:	21

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:	YUG KATHIRIYA
Date:	11th August 2024

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):	

A comparative analysis of Deep Learning Models for Spatio-Temporal Crime Prediction in Chicago

YUG KATHIRIYA
x22187839

Abstract

Chicago's complicated crime patterns provide a serious public challenge.While crime is a widespread issue, Chicago's high rates have received special attention. The city's varied neighbourhoods experience varying degrees of crime, emphasising the importance of localised crime prevention initiatives. This study tries to address this difficulty by studying the efficacy of deep learning models.ATTN-BILSTM, LSTM, and CNN-LSTM for predicting crime patterns. Additionally, eospatial clustering approaches, such as HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) and PCA-aided clustering, are used to detect and visualise crime hotspots.Analysing crime datasets and using these models can reveal trends in crime patterns, identify high-crime areas, and influence data-driven crime prevention efforts.

1 Introduction

Compared with some significant American cities, Chicago faces a greater violent crime rate, and one of the most alarming elements is how unsafe it is to venture outside. It is surprising to learn that crime is a worldwide issue that puts people's safety in danger in many nations. The crime rates in Chicago vary greatly by neighbourhood. A more accurate image of the criminal dynamic in the city is now possible because to the transformation of crime data displayed, prediction, and analysis brought about by deep learning.

This study will investigate a range of crimes in various Chicago neighbourhoods. Liang et al. (2024) used spatial correlation to forecast crime times and locations in the city. The writers highlighted a few significant points: Co-occurrence of Crime: Various forms of crimes, such as robbery and assault, are frequently perpetrated in the same block. This phenomena can be ascribed to two primary factors: (i) Criminals commonly commit many crimes concurrently due to low moral standards and a lack of legal awareness. (ii) Illegal businesses thrive in unregulated situations. Government Warnings: When crimes occur within a block, the government routinely issues public warnings, asking inhabitants to be cautious in nearby communities. This illustrates a typical feature of crime data: Proximity to crime scenes may increase both safety and risk.

Mandalapu et al. (2023), the task of crime prediction is intricate and requires sophisticated analytical tools to bridge the gaps in the existing detection methods. With the introduction of new technologies and an increasing amount of crime data, scholars now have a rare chance to study and research crime detection through the use of machines learning and deep learning techniques. Crime prediction has showed promise using deep learning techniques like recurrent neural networks and convolution. These algorithms may be able to accurately predict crime patterns in certain locations because they have been developed on data on crimes that includes both geographical and time dimensions. Algorithms that use deep learning have been applied to evaluate criminal data, such as the type of crime committed, date, and place.Utilising these data, a prediction model is developed that can identify potential crime hotspots and forecast the frequency of new crimes.

Additionally, they stated that the machine learning (ML), neural networks (DL), machine learning combined with natural language processing (NLP), as well as DL plus NLP were the most often utilised approach types in neighbourhood crime investigation. Of the articles on community crime, 68% used ML, 22% used DL, 7% used ML and DL together, a small number of papers used DL and NLP together, and a small number of research papers used ML and NLP together. Local crime scenes have demonstrated the regular application of machine learning techniques.

Classification tasks comprised 64% of research in the nearby crime domain; these were followed by a regression analysis (30%), segmentation (7%), and mixed tasks in a small number of papers. Classification with local crime data is the main focus.

A set of deep learning algorithms is demonstrated to be able to forecast different kinds of violence in the area regions. The city will categorise different types of crime in various places based on these facts. Additionally, when analysing trends in various scenarios, the precise spot of a the town's crime cluster dictates to what extent it is a safer region to commute. There are chronological and spatial categories in the dataset. We then use performance measurements against the state-of-the-art to assess and train our suggested model. The goal of our study is to generate estimates that are more precise than previous ones that consider crime categories and geography.

Predicting criminal behaviour is critical for improving public safety and allocating resources efficiently. Given that criminal behaviour varies with time and place, it is difficult to combine diverse types of crime (such as theft, robbery, assault, and damage) in a way that presents a comprehensive picture of criminal behaviour. To identify areas with elevated risks for crime prevention and to expedite criminal investigations, logical, numerical, or analytical prediction models that are able to forecast unlawful activity can be employed.In general, much research has been conducted, with data sourced from a variety of cities. The notion that hotspot maps alone can predict crime is refuted by evidence that combines geographical and temporal data to considerably enhance prediction accuracy.

Jenga et al. (2023) analysis attempts to improve our understanding of crime prediction methods and their implementation by creating more powerful predictive models.Proper identification and implementation of advanced models could usher in a new era in the use of predictive analytics to tackle safety hazards. Crime prediction is attracting more and more attention from researchers.

Furthermore, it states that extensive study has been conducted in the fields of criminal investigation and prediction. In comparison to the majority of industries, such as banking, healthcare, retail, agriculture, transportation, and customer service, there are few comprehensive and systematic literature reviews that focus only on crime prediction. Such evaluations can aid in the structure and synthesis of the body of knowledge, supporting information, and sector issues.

The project's concentration is on **"How does the Deep Learning model's efficacy and speed improve when ATTN-BiLSTM and HDBSCAN are used, and**

what is the difference between these algorithms in terms of location accuracy and criminal detection?". To answer the question, the suggested approach evaluates the ATTN-BiLSTM algorithm's performance in crime prediction. The approach will be developed and evaluated for prediction in a dataset. The ATTN-BiLSTM method will be compared to other deep learning models, such as CNN-LSTM and LSTM, to assess its effectiveness in terms of evaluation, statistical testing, and resilience.

2 Related Work

The Deep Learning approach has shown excellent performance in a variety of domains, particularly in spatial and temporal prediction. The present research will examine at the different versions as combinations of two networks for spatiotemporal prediction of crime events.

2.1 Spatial-Temporal Neural Network

The repercussions of early crime prediction, Alghamdi and Al-Dala'in (2024) traditional machine learning algorithms perform reasonably well but frequently lack to recognise critical information. They recommend training deep neural networks using real police log data from the Chicago open data platform to allow algorithms to find important elements without human interaction.

2.1.1 Advancements in Spatiotemporal Approach

Matereke et al. (2021) introduced deep learning approaches, notably the spatiotemporal residual network (ST-ResNet), deep multiview spatiotemporal network (MV-DilatedCNN), and Spatio-Tepmopral Dynamic Network, to measure crime mapping accuracy. Researchers found that their spatiotemporal dynamic network surpassed the others in terms of mean absolute error and root mean square error.

Wei et al. (2021) constructed a deep spatial-temporal-categorical neural network, CrimeSTC, to correctly forecast future crime episodes count for each city region and in different time windows. Subsequently consists of four major components: category, collaborative training, static, and dynamic. Static Module: Designed to characterise the ones connected to static data, which remains unchanged over time. Dynamic modules serve for day-to-day adjustments. It refers to a category class that defines relationships involving criminal categories. In order to verify the suggested strategy, they conducted extensive tests with real datasets.

In Esquivel et al. (2020), becomes clear the fact that the performance of the model depends on Temporal and Spatial Resolution, resulting in ultimately impacting classification accuracy. The neural network's outcome may change regardless of the type of crime, and succeeding timely preprocessing with architectural selection for Neural Network would dramatically affect crime prediction. Anticipating when and where crimes might occur would assist police to better utilise their limited resources. On the other hand, as indicated by sparse spatio-temporal matrices, the biggest drawback of CLSTM-NN is worse performance at low criminal event percentages, which has an impact on predicted spatial resolutions.

The author also suggests integrating the graph algorithms with other deep learning techniques to achieve greater accuracy in predictions. Furthermore, attentional neural networks will be utilised for enhancing crime prediction through absorbing extrinsic events related with crimes (such as weather and sociodemographic factors). The spatio-temporal forecast will additionally encompass new crime categories, such as common assault and burglary.

2.1.2 Graph Neural Network Approach

Sun et al. (2020) used a distance-based region graph to explore the spatial and temporal patterns of crime trends. It is based on the assumption that crime patterns closer in space or with past community criminal records over time are comparable. Furthermore, a Graph Neural Network (GNN) was employed to identify spatial relationships between regions.

Tekin and Kozat (2021) suggested an architecture to forecast high resolution on any spatiotemporal data using graph convolutional networks (GCN) and multivariate Gaussian distributions. Implement a subdivision method to offset sparsity in high-resolution GCNs. Researchers use graph convolutional gated recurrent units (Graph-ConvGRU) to describe category, temporal, and spatial correlations. In addition to the GCN attributes acquired, a multivariate probability distribution for each node in the network may be calculated. The researcher addresses the problem of limited high-resolution data.

Hu et al. (2021) addresses anomalies and irregular patterns in crime statistics. To circumvent this, they employed the Dual-Robust Enhanced Spatial-Temporal Learning Network (DuroNet), which has an encoder-decoder architecture. The game relies on two major aspects to boost its strength: DuroNet's Locality Enhanced Module uses both dynamic geographical and temporal information to reduce outliers and distinguish between normal and abnormal data points.The Pattern Representation Module uses self-attention strategies to avoid unanticipated data from being generated.

Spatio-Temporal Deep Fusion Graph Convolutional Networks (STDGCN) is the name of the suggested model. Making a spatiotemporal graph that includes every location at any given time and is connected to every other node both temporally and geographically is the first step in eradicating spatiotemporal borders. Secondly, the features produced by them accurately depict the spatial and temporal characteristics of the edges and nodes in this spatio-temporal graph. Lastly, they develop a spatiotemporal synchronous network convolution method to extract characteristics from the spatiotemporal neighbours of the prediction target.Furthermore, in order to improve the accuracy of crime forecasts, the author intends to look into the connections that exist between different categories of crimes and develop modules that can capture them.

2.1.3 Multi-Modal Neural Network

Wang et al. (2022) discusses geographical differences in crime rates. The HAGEN inventor built a graph layer that offers network structure.To boost the graph network's performance, CNN and RNN are combined. A homophily-aware loss function is developed to constrain the graph's shape.

They used two practical standards to evaluate HAGEN. HAGEN routinely outperforms the majority of powerful crime prediction algorithms and spatial-temporal graph neural networks. Through the next phase, they plan to enhance the technique and analyse traffic benchmarks to show that HAGEN may be used to anticipate spatially multivariate time series.

Liu et al. (2020) proposed a diverse fusion model that employs Graph Convolution Networks and Recurrent Units to combine information extraction with POTID, anomaly likelihoods, and other attributes. Although the city is represented as a graph, CNN performs lesser better than graph convolution networks. Once evaluated on real-world datasets, the proposed model performed much better than current state-of-the-art techniques.

Sun et al. (2020) designed and built CrimeForecaster, which aims to capture temporal repetition and spatial dependency across regions. Applying a Gated Recurrent Network with Diffusion Convolution modules enhanced accuracy by 21% contrasted with conventional techniques to convey relationships. Despite only St-ResNet has been used for crime prediction, the author evaluated the performance of three spatiotemporal deep learning systems: DMVST-Net, STD-Net, and St-ResNet. The outcome findings showed that STD-Net had the highest accuracy, MAE, and RMSE.Matereke et al. (2021)

Zhao et al. (2023) developed a Cross-Domain Attention Network, an Attention-Based fusion layer that collects and integrates features from various datasets with identical time periods. This was evaluated on real-world datasets and outperformed cutting-edge techniques, with each network's efficacy independently validated.

Butt et al. (2020) claims that in order to uncover the dominating technique, one ought to highlight the hotspot area, analyse it, and illustrate the changes over a ten-year period using performance metrics. Researchers also mention that a number of clustering and classification methods were utilised, including Random Forest and DBSCAN, which were recently developed, evaluated against cutting-edge techniques, and found to be beneficial and effective. Novel advances in crime hotspot detection have the potential to reduce the accuracy gap. The mean absolute error (MAE), mean absolute percentage error (MAPE), mean error (ME), and root mean square error (RMSE) are the four most commonly utilised performance indicators.

Particularly, time-series analysis approaches are becoming more and more used in machine learning and data mining programs for crime prediction. Algorithms for classification and clustering have been widely used in recent years to predict criminal activity. However, it has come to light that relying solely on these processes is not as reliable nor practical. Recently, the use of time series analysis to improve the prediction process led to a breakthrough in the field of crime prediction.

Enhancing the crime hotspot detection approach (DBSCAN) to improve detection accuracy, employing exponential smoothing to Long Short Term Memory (LSTM), and improving the crime forecasting algorithm (ARIMA) to improve prediction precision are some possible study topics.

2.2 Time Series Regression

To improve the prediction of crime sequences, Ayele et al. (2020) suggested using an LSTM model with a single hidden and output layer. The model is assessed using the R-squared value. With low error rates and great accuracy, the LSTM model makes predictions. above a certain threshold, data mining methods such as CNNs and traditional neural networks are unable to generate reliable results Thomas and Sobhana (2022). It can, however, be paired with different models. According to the study, combining LSTM with regression modelling improves prediction results and yields matches to real-world events with 90% accuracy. To get better results, time-series models and LSTMs should be contrasted. Many of the previously listed methods use several models to increase forecast accuracy.

Tasnim et al. (2022) maintains its reliability even when subjected to data that spans more than ten years. To tackle this issue, they used the ATTN-LSTM model in conjunction with the St-Bi-LSTM model's transfer learning technique.Although the model performed well, there were significant limitations due to the insufficient data set.

ConvBiLSTM is a model developed by Tam and Tanriöver (2023) that combines CNN with Bi-LSTM.The results of the experiment demonstrated that the model outperforms other traditional deep-learning and BERT models.Since this study solely considered stealing as a category of crime, it had several disadvantages. Consequently, in order to improve model performance, new crime types need be added.

2.3 Findings

The best technique for predicting crime using Deep Learning Neural Networks has been demonstrated to be time series analysis. For example, by focussing on temporal patterns, spatiotemporal approaches outperform conventional methodologies. CNNs, RNNs, GNNs, and GCNs are used to find geographic and temporal linkages in crime data, which greatly improves model performance. Traditionally, CNNs and LSTMs were used to identify patterns; however, merging several models further improved prediction accuracy. While sequential modelling is similar to text classification tasks, CNNs and RNNs perform well in sequential modelling even though they are not designed for parallel feature extraction. Due to its ability to access previous and subsequent contexts, BiLSTM outperforms LSTM in sequential modelling tasks, demonstrating its effectiveness in forecasting criminal activity attempts.

3 Methodology

The research focuses on the data collection,data pre-processing and applying algorithm to evaluate results

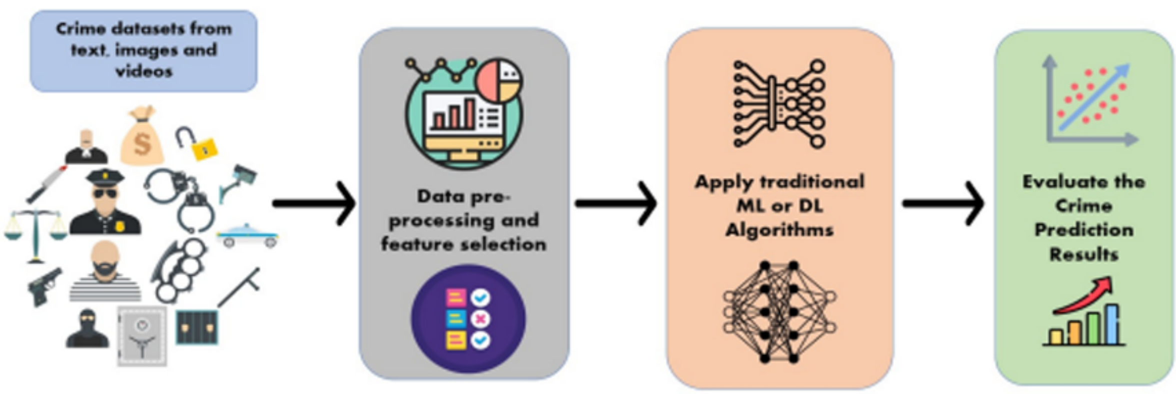

Figure 1: Flowchart for Crime Prediction Mandalapu et al. (2023)

3.1 Data Source

In order to optimise computational performance during early review, a consultant pattern of five million items can be extracted from the total dataset, which has approximately eleven million records. This pattern length is chosen to ensure a balance between statistical comprehensiveness and processing speed.

Following the preliminary evaluation, to focus on the most recent crime patterns, the statistics can be similarly filtered to include only the most useful rows from 2018 onwards. This centred choice will yield the final dataset of around 8110967 rows, suitable for in-depth evaluation and version training.

The data set contains information about recorded crime incidences in Chicago, as well as different attributes associated with each crime record.(Crimes 2001 to Present - Chicago Data Portal). ID: A unique key is an identifier for a criminal record. Case-number: The number assigned to each crime case. Date: The day and time when the crime occurred. Block: The block address where the crime took place. IUCR: Illinois Uniform offence Reporting Code, a number code that denotes the offence category. Primary Type: The fundamental classification of a crime (such as theft, assault, or robbery). Description: A thorough account of the crime. Location description: The sort of location where the incident occurred (e.g., street, home, park). Arrest: Whether or not an arrest was made for the crime. Domestic: Indicates whether the crime was committed in a domestic setting. Beat: The police beat in which the offence occurred. District: The police district where the crime took place. Ward: The city ward in which the crime happened. Community Area: The communal area where the crime occured. FBI-CODE: The FBI's classification of the offence. X-coordinate: The place where the crime happened (usually in a projected coordinate system such as UTM). Y-coordinate: The place where the crime happened (often in a projected coordinate system such as UTM). Year: The year in which the offence occurred. Updated-on: The date and time when the record was last updated. Latitude: The latitude coordinates of the crime scene. Longitude: The longitude coordinate of the crime scene. Location is a combination of latitude and longitude that is commonly depicted as a point on a map.

3.2 Data Preprocessing and Transformation

Following a variety of preprocessing tasks, a Python programming language is used to apply the model and process the data set. Before adding real data to the model, it's important to clean the raw data to remove irrelevant data like missing values (NA) and outliers. This ensures the model works properly. To cleanse data for this investigation, we used Google Colab The particular steps taken for data pre-processing are described below.

- Null Values are dropped
- Data before 2018 was discarded
- Date column is converted to datetime format
- Date components (Year, Month, Day, Hour, Minute, and Weekday) are extracted for further analysis.
- The dataset is divided into three time periods: before COVID (2018-2020), during COVID (2020-2022), and after COVID (2022-2024).

- Functions to identify the highest and lowest crime days throughout three periods.
- Fuction to find the top 10 crimes during three periods.
- Outliers are detected and removed using winsorize.
- Dividing the dataset into monthly groupings and further determine the seasons.
- Resampling and Normalizing the data

3.2.1 Missing Data

Some fields in the crime dataset, such as Location Description, Ward, Community Area, Latitude, Longitude, and Location, contain missing data which are eliminated using drop na.

3.2.2 Feature Engineering

The crime dataset covers offenses from 2001 to the present, but for our estimation, we will use data from 2018 to 2024. In this study, we are changing the date column to DateTime format. The top ten main crimes are filtered out, providing a more accurate picture of the statistics, because time-series forecasting requires a target column for forecasting, the crime count column is constructed. Outliers in this study are significant and handled. This study uses the date column to extract features such as hour, month, week, and year for monthly forecast purposes based on seasonality.Seasons are categorized based on months for the predictions.

3.3 Model Training

3.3.1 Time Series

ATTN-BILSTM: This model employs bidirectional LSTM layers with an attention mechanism. The attention mechanism allows the model to focus on relevant parts of the sequence, enhancing prediction accuracy.LSTM: Standard LSTM layers used for time series prediction.CNN-LSTM: This model combines Convolutional Neural Networks (CNN) and LSTM layers.CNN layers help capture spatial properties, whereas LSTM layers help detect temporal relationships.Each model is trained using the training data for the specific season.The training method involves fitting the model to the training data using a validation split.

3.3.2 Crime Hotspot using HDBSCAN

- Data loading and cleaning: Google Drive was used to load crime data.Missing values are handled by eliminating rows with null values in columns like latitude and longitude.The date column was converted to datetime format, and the month and year were extracted.The dataset was filtered from 2018 to 2024. Latitude and longitude values were normalised such that they contributed equally to distance calculation during clustering.
- Parameter Tuning : To create clusters, use HDBSCAN with modified options (min-cluster-size=100, min-samples=10) on filtered data.

- Visualization: Markers are colour-coded by crime type, creating a visual cluster highlighting crime hotspots.Clusters help to identify high-crime locations.

3.4 Evaluation Metrics

The productiveness of each model in predicting crimes will be evaluated using RMSE, MAE, R2 and Adjusted R2 performance metrics.

3.4.1 RMSE: Root Mean Squared Error

RMSE is the square root of the average square difference between anticipated and actual values; it evaluates how well the model's predictions fit the actual values. RMSE measures how well our models (ATTN-BILSTM, LSTM, and CNN-LSTM) predict monthly crime numbers. Lower RMSE values correspond to more accurate predictions, allowing us to compare the performance of different models.

3.4.2 MAE: Mean Absoulte Error

MAE is the mean absolute difference between expected and actual values. MAE allows us to understand the average magnitude of inaccuracy in our model's predictions. High MAE values suggest significant prediction errors, necessitating additional model tuning.

3.4.3 R-Squared

R-squared is a statistical measure that reflects how much of the dependent variable's variance can be predicted based on the independent variables. R^2 evaluates how accurately our models explain monthly crime numbers. Higher R^2 values indicate the model accurately represents the data's underlying trends.

3.4.4 Adjusted R-squared

Adjusted R^2 assesses model performance according to the number of features employed. It ensures that models with more predictors are chosen only if they improve the model's explanatory power.

Why these metrics: The RMSE and MAE are direct metrics of prediction error magnitudes that are essential for understanding the practical consequences of our model's results. R^2 and Adjusted R^2 measure the model's explanatory ability by providing the proportion of variation explained.

Lower RMSE and MAE values indicate more accurate projections. Models with higher R^2 and Adjusted R^2 values are more accurate in capturing data trends.

3.4.5 Silhouette Score

Silhouette Score: Measures how similar an object is to its own cluster compared to other clusters. It ranges from -1 to 1, with higher values suggesting a stronger grouping.

3.4.6 Davies-Bouldin Score

Davies-Bouldin Score: Determines the average similarity ratio of each cluster to its most similar cluster. Lower numbers indicate better categorisation.

3.4.7 Calinski-Harabasz Index:

The ratio of dispersion between clusters to dispersion within clusters. Higher numbers signify better classification.

4 Design Specification

This section will go over the architecture and specifications of the models being used.

4.1 ATTN-BILSTM

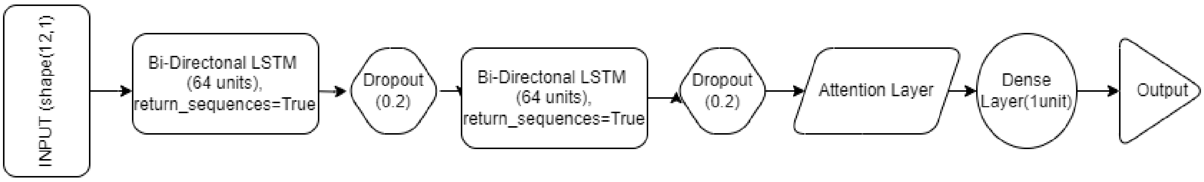

Figure 2: Architecture of ATTN-LSTM

Input Layer: The input shape (12, 1) represents 12 months of data and one feature (monthly crime count). Bidirectional LSTM: Two layers of 64-unit bidirectional LSTMs with return sequences are used to preserve the temporal structure of the data. To avoid overfitting, utilise dropout layers with a rate of 0.2. Attention Layer: A bespoke attention layer that emphasises on critical time points in the sequence. Dense Layer: A fully connected layer with a single unit that produces the final prediction.

4.2 LSTM

Input Layer: The input shape (12, 1) represents 12 months of data and one feature (monthly crime count). LSTM: Two layers, each with 64 units. Return sequences are activated in the first LSTM layer, which transfers the temporal structure to the next LSTM layer. To avoid overfitting, utilise dropout layers with a rate of 0.2. Dense Layer: A fully connected layer with a single unit that produces the final prediction.

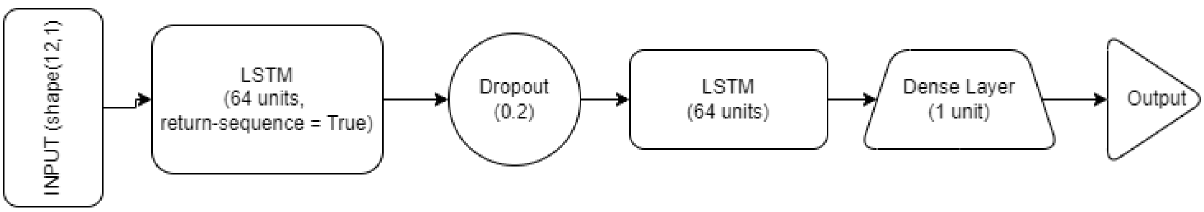

Figure 3: Architecture of LSTM

4.3 CNN-LSTM

Input Layer: The input shape (12, 1) represents 12 months of data and one feature (monthly crime count). Conv1D is a 1D convolutional layer with 64 filters, a 2 kernel size, and ReLU activation for extracting local information from a sequence. MaxPooling1D: To reduce dimensionality, use a max-pooling layer with a pool size of 2. LSTM: A typical LSTM layer with 64 units for temporal dependencies. Dense Layer: A fully connected layer with a single unit that produces the final prediction.

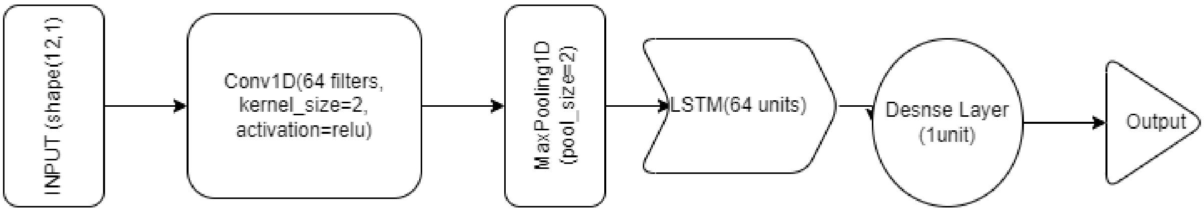

Figure 4: Architecture of CNN-LSTM

5 Implementation

- For the implementation of the project, the use of Jupyter Notebook on Google Collab has been utilized, leveraging high RAM environment to handle the extensive data processing requirements.The crime data from City of Chicago dataset spans from 2018 to 2024 and includes details such as Date,Primary Type of Crime, Latitude and Longitude.Dataset was stored in Google Drive for easy access and processing.The details are mentioned in the table 1

- Resampling and Normalizing the data: Crime data is resampled to get monthly crime counts, then data is normalized using MinMaxScaler to scale values between 0 and 1.Creating Sequence: Data is divided into sequences for preparation of time series prediction.Each sequence consists of 12 months of data to predict the next months crime count.Train-Test split: Sequences are split into training(80%) and testing (20%). Categorize Data by Seasons: Data is categorized into different seasons(Autumn,Spring,Summer,Winter) based on the months.

- Model Definitions:ATTN-BILSTM: This model uses Bidirectional LSTM layers with an attention mechanism , attention mechanism helps the model focus on relevant part of sequence improving the prediction accuracy.LSTM: Standard LSTM layers for time series prediction.

- CNN-LSTM: This model combines Convolutional Neural Networks (CNN) and LSTM layers.CNN layers assist capture spatial characteristics, whereas LSTM layers help catch temporal connections.Each model is trained employing the training data for the respective season.The training method consists of fitting the model to the training data using a validation split.

- Hyperparameter Tuning : Grid search was used to find the best hyperparameters for each model.Models were trained and evaluated using RMSE, MAE, R^2, and Adjusted R^2. The parameters used are

- Attn-bilstm-params ='units-1': 64, 'units-2': 64, 'dropout-1': 0.2, 'dropout-2': 0.2,'units-1': 128, 'units-2': 128, 'dropout-1': 0.3, 'dropout-2': 0.3.These parameters were used to evaluate model performance at various levels of complexity and regularisation, with the goal of balancing learning capacity and overfitting.

- Lstm-params = ['units-1': 64, 'units-2': 64, 'dropout-1': 0.2, 'units-1': 128, 'units-2': 128, 'dropout-1': 0.3].These configurations were chosen to compare the performance of simple LSTM to BiLSTM with Attention, with the goal of identifying the best-performing architecture.

- Cnn-lstm-params = ['filters': 64, 'units': 64, 'dropout': 0.2, 'filters': 128, 'units': 128, 'dropout': 0.3].It was examined at various degrees of complexity to find the optimal balance of local feature extraction (by CNN) and temporal modelling (via LSTM).

Table 1: Project Technical Details

Category	**Details**
IDE	Google Colab
Programming Language	Python
RAM	High RAM Environment
Disk	Google Drive Storage

Time Series Models	`tensorflow.keras.models.Sequential`
	`tensorflow.keras.layers.LSTM`
	`tensorflow.keras.layers.Bidirectional`
	`tensorflow.keras.layers.Conv1D`
	`tensorflow.keras.layers.MaxPooling1D`
	`tensorflow.keras.layers.Dense`
	`tensorflow.keras.layers.Input`
	Custom Attention Layer using `tensorflow.keras.layers.Layer`

Scikit-Learn	`sklearn.metrics.mean_squared_error`
	`sklearn.metrics.mean_absolute_error`
	`sklearn.metrics.r2_score`
	`sklearn.preprocessing.StandardScaler`
	`sklearn.metrics.silhouette_score`
	`sklearn.metrics.davies_bouldin_score`
	`sklearn.metrics.calinski_harabasz_score`

SciPy	`scipy.stats.ttest_rel`

Map Generation	`folium.Map`
	`folium.Marker`
	`folium.plugins.MarkerCluster`
	`folium.Icon`

Google Colab Specific	`google.colab.drive`

- Units(64,128): More units can capture more complex patterns in data but it increases cost and risk of overfitting.Dropout(0.2,0.3): It is used to prevent overfitting, higher dropout rate means more regularization which helps in improving generalization.

- Fiters: It determines how many filters are applied to input sequence to extract local feature before feeding them into LSTM.

- HDBSCAN: The data is separated into Year, Month, Primary Type, Logitude, and Latitude, and then the cluster model is applied.Importing Folium for producing web maps is used for visualisation 1MarkerCLuster is used to group several markers on a map.Once the cluster is organised, define the colours for the crimes and create the map.12.As longitude and latitude are employed, the visualisation aids in in-depth cluster development and pinpoints the actual location of crime in the region.

6 Evaluation

This section will demonstrate the results of the experiments performed.

6.1 Data Exploration

This experiment aims to study the patterns of crime data to determine which weekday has the highest or lowest crime rates in Pre-Covid, During-Covid and Post-Covid.First custom function is made consisting of Day counts, Highest and Lowest days according to each period(Pre-Covid,During-Covid and Post-Covid).

- The bar chart shows 5 crime occurences acroos different days of week from the Pre-COvid period spanning from 2018 to 2020. Monday, the crime occurred were approximately 76000 then on Tuesday, Wednesday, and Thursday the crimes decreased slowly.Friday is the most prevalent day for crimes (approx 79000), no day is more crime prone than others, as demonstrated in the figure. Furthermore it shows declining trend from Friday to Monday,this pattern suggest a potential link between workdays and reduced crimes. Sunday is the lowest crime approx 73000.

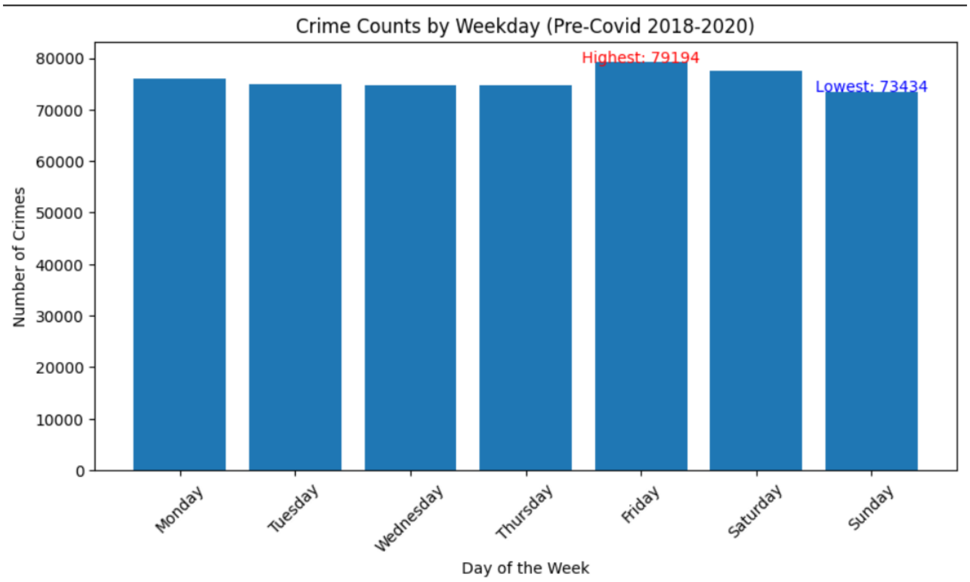

Figure 5: Pre-Covid

- The steady crime rate may reflect a shift in people's conduct as a result of lockdowns, limitations, and disrupted routines during the pandemic.In comparison to the pre-Covid chart, which showed higher crime trends on weekends, the post-Covid chart levelled the crime distributions across weekdays. The graph shows 6 a steady crime rate across all days, with small fluctuations and no notable peaks. As shown, Friday has the highest value of 62000, while Tuesday has the lowest value of 58000.

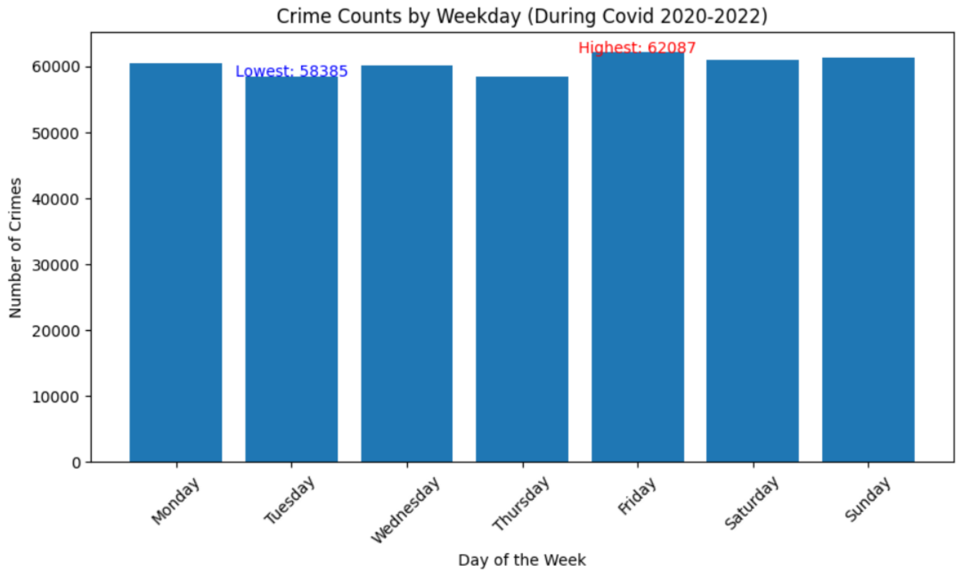

Figure 6: During-Covid

- In comparison to Pre-COvid and During-Covid, the trend in Post-Covid(2022-2024) is slightly different; when the week begins, the crime rate remains consistent from Monday to Thursday, with only minimal fluctuations.Crime rises over the weekend, Friday through Sunday. The highest number is Friday 73771 and the lowest is Thursday 705257.

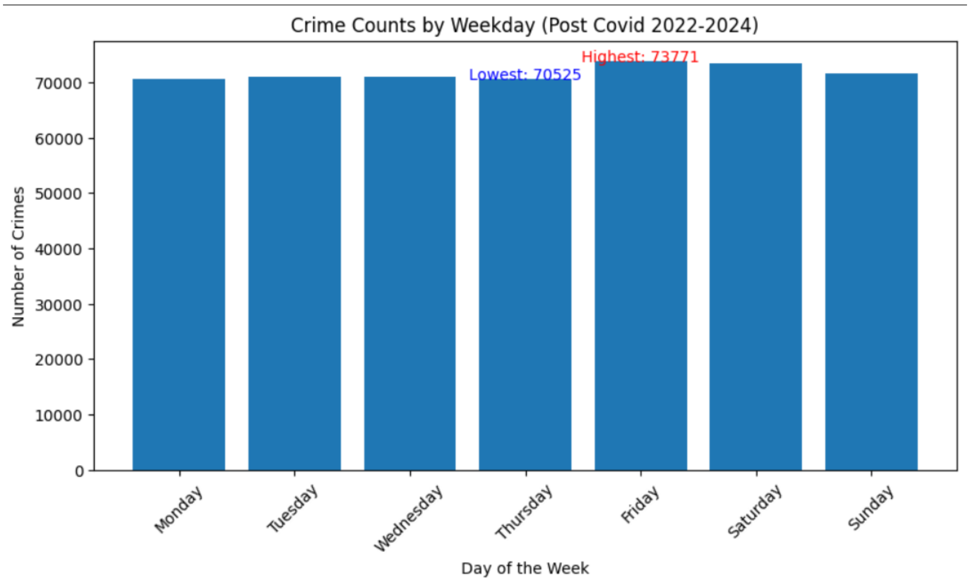

Figure 7: Post-Covid

So as and when years past wise crime also increases and decreases some crimes were Highest from 2018-2022 but Post-Covid new crimes are reported.

6.2 Experiment 2: Time Series on Seasonality Prediction

Model	Season	RMSE	MAE	R^2	Adjusted R^2
ATTN-BILSTM	Winter	4974.65	3683.24	0.560272	0.516300
LSTM	Winter	4537.46	3514.70	0.634166	0.597583
CNN-LSTM	Winter	4075.99	3229.62	0.704795	0.675275
ATTN-BILSTM	Spring	4830.35	3715.44	0.685296	0.653826
LSTM	Spring	4478.08	3507.49	0.729525	0.702478
CNN-LSTM	Spring	4480.98	3581.56	0.729174	0.702091
ATTN-BILSTM	Summer	3687.59	2825.34	0.805978	0.786576
LSTM	Summer	3291.35	2578.24	0.845435	0.829978
CNN-LSTM	Summer	3031.93	2473.38	0.868840	0.855724
ATTN-BILSTM	Autumn	3793.43	2746.10	0.751191	0.726310
LSTM	Autumn	3392.15	2605.10	0.801046	0.781150
CNN-LSTM	Autumn	2936.89	2334.95	0.850866	0.835952

Table 2: Performance Analysis Results

- Shi et al. (2024) mentioned that evaluation metrics for time series are RMSE,MAE,R2 to evaluate the best performing model Winter: CNN-LSTM has the lowest RMSE, indicating improved performance. In comparison to other models, the CNN-LSTM model better captures data variance. Spring: LSTM has lower RMSE and higher R2 values, indicating better capture of data variance. Summer: CNN-LSTM consistently outperforms with the lowest RMSE and highest R2 values. Autumn: CNN-LSTM outperforms other models with lower RMSE and greater R2 scores.

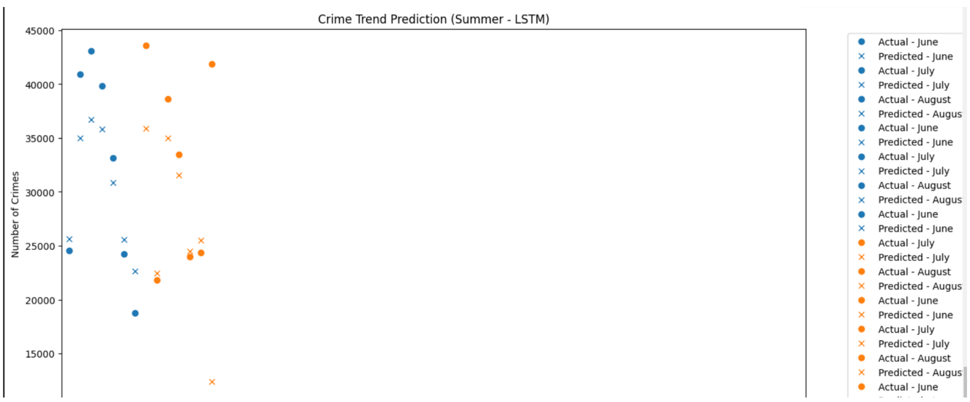

Figure 8: Summer-LSTM

- LSTM: Observation: The model displays similar trends to the actual data, although with some apparent gaps.8Strengths: Effective at detecting gradual changes.Weaknesses: There are more disparities at times of fast change, implying that it may not manage unexpected surges as well as the other models.

Figure 9: Summer-ATTN-BILSTM

ATTN-BILSTM: Observation: The projected values (orange) closely match the actual values (blue), with a few variances.9 Strengths include improved performance in capturing strong trends and peaks. Weaknesses: There are little discrepancies at several points, indicating that accuracy can be improved.

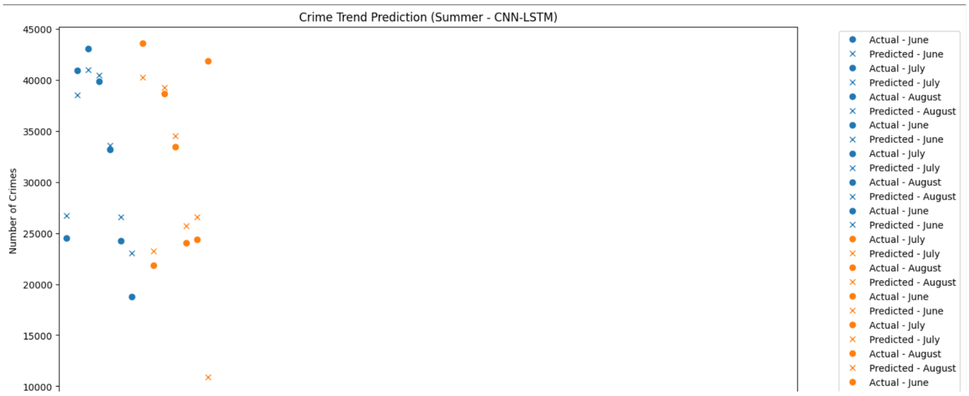

Figure 10: Summer-CNN-LSTM

CNN-LSTM: Observation: The orange and blue dots are generally close, but some differences are visible.10 Strengths: Effective at catching the general trend. Weaknesses: Slightly less accurate during sudden changes in crime counts than ATTN-BILSTM.

CNN-LSTM appears to be the best performing model, owing to its low RMSE and high R2 values.The model's performance varies with season, demonstrating that seasonality is an important component in the dataset; alternate models may be better suited to different seasons, and each statistic emphasises different aspects of performance.

Performance:CNN-LSTM consistently performed well throughout numerous seasons, giving it an excellent choice for crime prediction tasks. ATTN-BiLSTM produced competitive results and may be more useful with additional tuning and larger datasets due to its ability to focus on critical temporal patterns. LSTM performed well in the spring and summer, suggesting its ability to capture temporal dependencies.

Efficiency: While ATTN-BiLSTM increases computational complexity, its potential to capture nuanced temporal patterns makes it useful for deeper studies. CNN-LSTM and LSTM models establish a mix between performance and computational efficiency, making them appropriate for practical use.

6.2.1 Paired T test

Comparison	T-statistic	P-value
ATTN-BILSTM vs. LSTM	2.6395	0.0777
ATTN-BILSTM vs. CNN-LSTM	13.1618	0.0009
LSTM vs. CNN-LSTM	4.0492	0.0271

Table 3: Paired t-test Results

- ATTN-BILSTM against LSTM:The T-statistic of 2.639 indicates a moderate difference in performance between the ATTN-BILSTM and LSTM models. However, the p-value of 0.077 suggests that the difference is not statistically significant at the traditional 0.05 level. This indicates that we cannot reliably claim that ATTN-BILSTM outperforms LSTM using the RMSE metric.

- ATTN-BILSTM against CNN-LSTM: The T-statistic of 13.162 shows a significant performance difference between the ATTN-BILSTM and CNN-LSTM models. The extremely low p-value (0.0009) indicates that the difference is highly statistically significant. As a result, we can confidently state that ATTN-BILSTM outperforms CNN-LSTM using the RMSE metric.

- LSTM versus CNN-LSTM: The T-statistic of 4.049 indicates a significant performance difference between the LSTM and CNN-LSTM models. The p-value of 0.027 indicates a statistically significant difference at the 0.05 level. As a result, we may conclude that the RMSE measure distinguishes the LSTM model from the CNN-LSTM model. Sravani and Suguna (2022) suggested to compare the model using paired T-test

6.3 Experiment 3: Hotspot Prediction

Silhouette Score: 0.0612242951359238 The Silhouette Score assesses how similar an object is to its own cluster vs other clusters. The value ranges from -1 to one. Close to 1 indicates that the object is well matched to its own cluster but badly matched to neighbouring clusters. Close to 0: Denotes that the object is on or very close to the decision border between two adjacent clusters.

Davies-Bouldin Score: 1.338. A score of 1.338 denotes substantial cluster separation. While not extraordinarily high, this figure indicates that there is some degree of overlap across clusters and that the clusters are not very compact.

Calinski-Harabasz Index: 3.0803. 3.0803 is a low value, showing that the dispersion within clusters is not considerably smaller than the dispersion between clusters. This shows that the clusters aren't particularly distinct.

The maps illustrate criminal incidence clusters, with each colour-coded marker indicating a unique crime type. Hotspots are places with high crime rates, as indicated by groupings on the map. The red circles indicate areas of high crime activity, allowing for focused actions and resource allocation.

Hotspot Analysis: High-Density Clusters: These regions have a high concentration of crime, as indicated by larger and darker markers.11 Low-Density Clusters: Smaller, lighter indicators represent areas with fewer crimes.

As zoomed in more it can be seen that there are high crime cluster with large number and low dense areas with few crimes.12

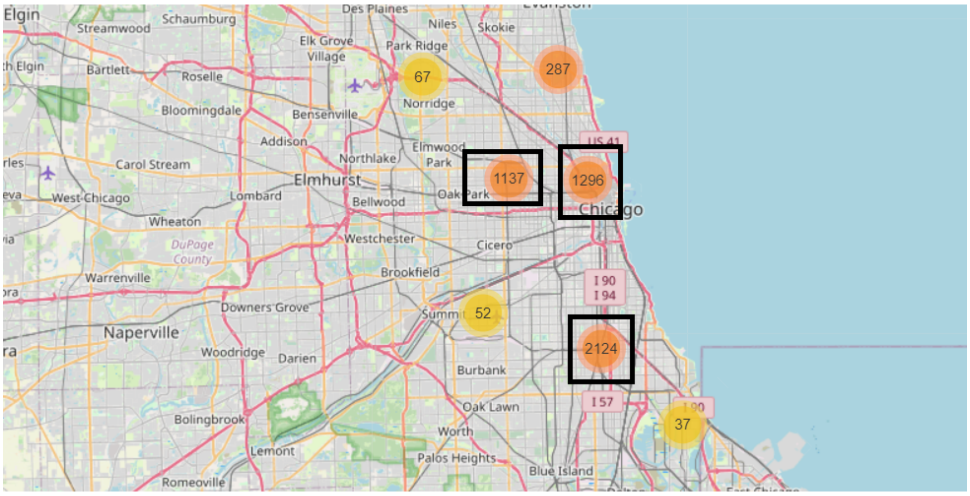

Figure 11: Cluster formation

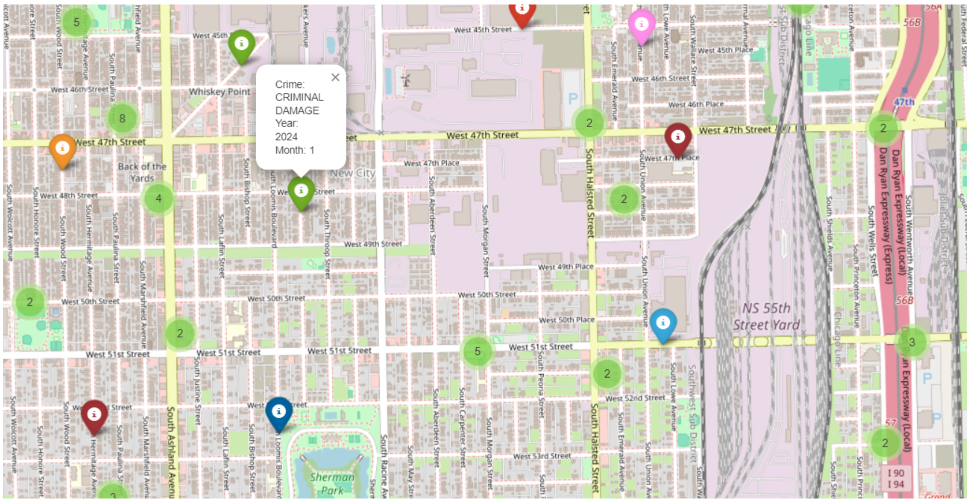

Figure 12: In-depth cluster

6.4 Discussion

To answer the study question, deep learning models, particularly ATTN-BiLSTM, have demonstrated considerable gains in detecting crime trends and forecasting locations. The comparison with LSTM and CNN-LSTM models shows that, while LSTM is useful in some cases, ATTN-BiLSTM and CNN-LSTM models are generally more accurate and resilient in identifying and predicting crime types. This is due to their ability to manage complicated spatial-temporal interactions in the data.

From the above experiments all the models performed differently depending on the seasons.According to the evaluation measures, the CNN-LSTM model outperformed the ATTN-BiLSTM and LSTM models in terms of RMSE, MAE, R2, and adjusted R2 scores across seasons. Our paired t-test results revealed statistically significant differences across the models, demonstrating that combining CNN and LSTM produces a more accurate representation of temporal patterns in crime data.

For cluster analysis, HDBSCAN was implemented to identify crime hotspots. The data set was filtered for relevant columns, and the HDBSCAN model was fine-tuned using the min-cluster size and min-samples parameters. The clustering results were visualised using Folium maps, which depict the spatial distribution of crime groups.

7 Conclusion and Future Work

In this research, the suggested ATTN-BiLSTM model was compared to other models such as CNN-LSTM and LSTM. CNN-LSTM had the highest R^2 and lowest RMSE, but a paired t-test based on RMSE showed that ATTN-BiLSTM outperformed CNN-LSTM with a very low p-value, indicating a highly significant difference. ATTN-BiLSTM is computationally intensive, yet its ability to capture detailed temporal patterns makes it useful for future research. On the other hand, CNN-LSTM and LSTM models strike a balance between performance and computational efficiency, making them appropriate for practical applications. Furthermore, the clustering analysis revealed a minor improvement in the cluster model, allowing it to recognise some spatial patterns.

7.1 Future Work

There are several avenues for future work Ensemble Methods for Time Series:

- Ensemble Method for Time Series: Combining ATTN-BILSTM, CNN-LSTM, and LSTM to maximise their respective strengths.Include external aspects like weather, holidays, and economic indicators to improve accuracy.Handling unbalanced data with SMOTE and investigating dimensionality reduction strategies.
- Real-time prediction and visualisation: Creating a dashboard with streaming data to provide timely insights and patterns of crime.You can also incorporate the models to provide more effective crime prevention measures.
- Advanced Architecture Investigate novel deep learning models, such as the Transformer model, or combine attention mechanisms with CNN-LSTM for future breakthroughs.

References

Alghamdi, J. and Al-Dala'in, T. (2024). Towards spatio-temporal crime events prediction, *Multimedia Tools and Applications* **83**(7): 18721–18737.
URL: *https://doi.org/10.1007/s11042-023-16188-x*

Ayele, N., Kibru, Y., Meskela, T., Mengist, T. and Teferi, M. (2020). Designing time series crime prediction model using long short-term memory recurrent neural network, *International Journal of Recent Technology and Engineering* **9**: 402–405.

Butt, U. M., Letchmunan, S., Hassan, F. H., Ali, M., Baqir, A. and Sherazi, H. H. R. (2020). Spatio-temporal crime hotspot detection and prediction: A systematic literature review, *IEEE Access* **8**: 166553–166574.

Esquivel, N., Nicolis, O., Peralta, B. and Mateu, J. (2020). Spatio-temporal prediction of baltimore crime events using clstm neural networks, *IEEE Access* **8**: 209101–209112.

Hu, K., Li, L., Liu, J. and Sun, D. (2021). Duronet: A dual-robust enhanced spatial-temporal learning network for urban crime prediction, *ACM Trans. Internet Technol.* **21**(1).
URL: *https://doi.org/10.1145/3432249*

Jenga, K., Catal, C. and Kar, G. (2023). Machine learning in crime prediction, *Journal of Ambient Intelligence and Humanized Computing* **14**: 1–27.

Liang, K., Zhou, S., Liu, M., Liu, Y., Tu, W., Zhang, Y., Fang, L., Liu, Z. and Liu, X. (2024). Hawkes-enhanced spatial-temporal hypergraph contrastive learning based on criminal correlations, *Proceedings of the AAAI Conference on Artificial Intelligence* **38**(8): 8733–8741.
URL: *https://ojs.aaai.org/index.php/AAAI/article/view/28719*

Liu, R., Zhao, S., Cheng, B., Yang, H., Tang, H. and Yang, F. (2020). St-mfm: A spatiotemporal multi-modal fusion model for urban anomalies prediction, *European Conference on Artificial Intelligence*.
URL: *https://api.semanticscholar.org/CorpusID:221714634*

Mandalapu, V., Elluri, L., Vyas, P. and Roy, N. (2023). Crime prediction using machine learning and deep learning: A systematic review and future directions, *IEEE Access* **11**: 60153–60170.
URL: *http://dx.doi.org/10.1109/ACCESS.2023.3286344*

Matereke, T., Nyirenda, C. N. and Ghaziasgar, M. (2021). A comparative evaluation of spatio temporal deep learning techniques for crime prediction, *2021 IEEE AFRICON*, pp. 1–6.

Shi, H., Wei, A., Xu, X., Zhu, Y., Hu, H. and Tang, S. (2024). A cnn-lstm based deep learning model with high accuracy and robustness for carbon price forecasting: A case of shenzhen's carbon market in china, *Journal of Environmental Management* **352**: 120131.
URL: *https://www.sciencedirect.com/science/article/pii/S0301479724001178*

Sravani, T. and Suguna, M. (2022). Comparative analysis of crime hotspot detection and prediction using convolutional neural network over support vector machine with engineered spatial features towards increase in classifier accuracy, *2022 International Conference on Business Analytics for Technology and Security (ICBATS)*, pp. 1–5.

Sun, J., Yue, M., Lin, Z., Yang, X., Nocera, L., Kahn, G. and Shahabi, C. (2020). Crimeforecaster: Crime prediction by exploiting the geographical neighborhoods' spatiotemporal dependencies, *ECML/PKDD*.
URL: *https://api.semanticscholar.org/CorpusID:232082537*

Tam, S. and Tanriöver, (2023). Multimodal deep learning crime prediction using tweets, *IEEE Access* **11**: 93204–93214.

Tasnim, N., Imam, I. T. and 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.

Tekin, S. F. and Kozat, S. S. (2021). Crime prediction with graph neural networks and multivariate normal distributions.

Thomas, A. and Sobhana, N. (2022). A survey on crime analysis and prediction, *Materials Today: Proceedings* **58**: 310–315. International Conference on Artificial Intelligence Energy Systems.
URL: *https://www.sciencedirect.com/science/article/pii/S2214785322007489*

Wang, C., Lin, Z., Yang, X., Sun, J., Yue, M. and Shahabi, C. (2022). Hagen: Homophily-aware graph convolutional recurrent network for crime forecasting, *Proceedings of the AAAI Conference on Artificial Intelligence* **36**(4): 4193–4200.
URL: *https://ojs.aaai.org/index.php/AAAI/article/view/20338*

Wei, Y., Liang, W., Wang, Y. and Cao, J. (2021). Crimestc: A deep spatial-temporal-categorical network for citywide crime prediction, *Proceedings of the 2020 3rd International Conference on Computational Intelligence and Intelligent Systems*, CIIS '20, Association for Computing Machinery, New York, NY, USA, p. 75–79.
URL: *https://doi.org/10.1145/3440840.3440850*

Zhao, S., Liu, R., Cheng, B. and Zhao, D. (2023). Classification-labeled continuousization and multi-domain spatio-temporal fusion for fine-grained urban crime prediction, *IEEE Transactions on Knowledge and Data Engineering* **35**(7): 6725–6738.