CLUSTERING ENSEMBLE AND HYBRID OF DEEP LEARNING FOR SPATIO-TEMPORAL CRIME PREDICTIONS

UMAIR MUNEER BUTT

UNIVERSITI SAINS MALAYSIA

2024

CLUSTERING ENSEMBLE AND HYBRID OF DEEP LEARNING FOR SPATIO-TEMPORAL CRIME PREDICTIONS

by

UMAIR MUNEER BUTT

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

June 2024

ACKNOWLEDGEMENT

In the name of Allah, the most gracious and merciful. I want to extend my most profound praise to Allah S.W.T for the blessings and guidance in completing this research and thesis. Also, I would like to extend special thanks and gratitude to Universiti Sains Malaysia and Fundamental Research Grant Scheme (FRGS) for the financial support throughout my studies and the resources and facilities they provided. My supervisors, Dr Sukumar Letchmunan and Dr Fadratul Hafinaz Hassan, for their kind supervision and guidance throughout this research.

I would also be thankful to my mother, wife, and my sons, Abdul Ahad, Muhammad Azlan, and Abdullah, for the love, patience, trust, and inspiration they gave me. My brothers were very supportive and were always there when I needed them most. And lastly, my friends, this research will not be completed without your support.

This thesis is dedicated to all my beloved, especially my father, who is no more in this world.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF ABBREVIATIONS

HDBSCAN	Hierarchical Density-Based Spatial Clustering of Applications with Noise
SARIMA	Seasonal Auto-Regressive Integrated Moving Average
NYC	New York City
USA	United States of America
GIS	Geographic Information System
SNA	Social Network Analysis
KDE	Kernel Density Estimation
LA	Los Angeles
GAM	General Additive Model
STNN	Spatio-Temporal Neural Network
STKDE	Spatio-Temporal Kernel Density Estimation
PAI	Predictive Accuracy Index
S-T GAM	Spatio-Temporal Generalized Additive Model
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
SVM	Support Vector Machine
1NN	One Nearest Neighbor
RF	Random Forest
EL	Ensemble Learning
CCRBoost	Cluster-Confidence-Rate- Boosting

SEPP	Self-Exciting Point Process
DSTP	Distributed Spatio-Temporal Pattern
ESTP	Ensemble Spatio-Temporal Pattern
ARIMA	Auto-Regressive Integrated Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ME	Mean Error
RMSE	Root Mean Square Error
STCP	Spatio-Temporal Crime Prediction
HAC	Hierarchical Agglomerative Clustering
AR	Auto Regressive
MA	Moving Average
ARMA	Auto Regressive Moving Average
MODA	Mayors Office of Data Analytics
DoITT	Department of Information Technology and Telecommunications
DCR	Dense Crime Region
IPS	Institut Pengajian Siswazah
PPSK	Pusat Pengajian Sains Komputer
USM	Universiti Sains Malaysia
UTMK	Unit Terjemahan Melalui Komputer

ENSEMBEL PENGELOMPOKAN DAN HIBRID PEMBELAJARAN MENDALAM UNTUK RAMALAN JENAYAH RUANG-MASA

ABSTRAK

Pertambahan penduduk bandar menimbulkan cabaran dalam mengurus perkhidmatan dan keselamatan daripada aktiviti jenayah. Pihak berkepentingan yang berkenaan berhasrat untuk meramalkan masa, lokasi, bilangan, dan jenis jenayah untuk mengambil langkah pencegahan yang sesuai. Pengenalpastian yang tepat dan ramalan tempat tumpuan jenayah boleh memberi manfaat yang ketara kepada pihak berkepentingan dalam mencegah jenayah dengan mencipta visualisasi ancaman yang tepat dan memperuntukkan sumber polis dengan cekap. Beberapa teknik telah dicadangkan untuk ramalan jenayah, tetapi ia terhad dalam ketepatan dan meramalkan jenayah mengikut jenis jenayah setiap jam, bulanan dan bermusim. Pendekatan pengesanan tempat liputan jenayah adalah terutamanya sensitif terhadap pemilihan parameter awal dan mencari kelompok pelbagai bentuk dan kepadatan. Begitu juga, pendekatan ramalan Jenayah sedia ada adalah terhad dalam menangkap data tidak pegun dan kebergantungan jangka panjang dengan memfokuskan pada jenis jenayah. Oleh itu, mekanisme pengesanan dan ramalan jenayah memerlukan penambahbaikan dalam bilangan jenayah, rentang jenayah, ketepatan, dan wilayah dan ramalan jenayah yang padat. Objektif teras kajian ini adalah dua kali ganda. Pertama, ia mencadangkan model pengesanan tempat liputan jenayah untuk meningkatkan ketepatan menggunakan Pengelompokan Ruang Berasaskan Ketumpatan Hierarki bagi Aplikasi dengan Bunyi (HDBSCAN) dan ensembel pengelompokannya untuk menangkap gugusan bentuk dan ketumpatan yang berbeza-beza serta meningkatkan ketepatan. HDBSCAN digunakan dengan pemulaan parameter yang berbeza-beza dalam mekanisme penjanaan di bawah para-

digma ensembel kelompok. Selain itu, enam ukuran jarak berbeza digunakan untuk memastikan kepelbagaian. Di samping itu, fungsi penilaian dicadangkan diparameterkan oleh skor siluet untuk memilih pengelompokan yang stabil di antara kumpulan penyelesaian pengelompokan untuk memastikan kualiti. Tambahan pula, kajian ini mencadangkan algoritma Edge Enhanced Hypergraph Construction (EEHC) sebagai fungsi konsensus dengan memanfaatkan kekuatan Hypergraph Partitioning Algorithm (HGPA) dan Kernel Density Estimation (KDE) untuk menangkap struktur hierarki dan melicinkan tepi dengan mengurangkan kesan. ketidaksempurnaan data dalam proses pembahagian. Tiga algoritma pembahagian hipergraf terkini telah dibandingkan untuk memilih kaedah dengan pembahagian kelompok yang berkualiti. Kedua, kajian ini mencadangkan Spatio-temporal Autoregressive Transformer (START) dengan memanfaatkan kekuatan spatio-temporal transformer dan Vector Autoregression (VAR) untuk meramal jenayah mengikut jenis jenayah dengan mengendalikan data tidak pegun dan menangkap kebergantungan jangka panjang. Satu pengekodan panas digunakan untuk membezakan jenis jenayah, diikuti oleh Penguraian Aliran Bermusim menggunakan LOESS (STL) untuk menguraikan siri masa kepada tiga komponen (trend, bermusim dan baki) untuk menangkap variasi baki yang diabaikan untuk komponen aliran dan bermusim dengan cekap. Selain itu, perhatian VAR boleh meningkatkan ketepatan ramalan jenayah dengan ketara dengan memfokuskan pada jenis jenayah tertentu, membolehkan model mengutamakan maklumat yang berkaitan semasa ramalan. Kaedah yang dicadangkan dinilai pada set data jenayah terkini New York, Chicago, Lahore dan Los Angeles yang tersedia secara terbuka. Kajian ini menilai metodologi pengesanan hotspot jenayah menggunakan skor Silhouette, Indeks Davies Bouldin (DBI), Indeks Calinski-Harabasz (CHI), Indeks Nisbah Penentu (DRI), kaedah Sisih-

an Piawai Purata Kuasa Akar (RMSSTD), dan kuasa dua r (RS) untuk memeriksa pemisahan, ketersambungan dan kekompakan kelompok. Selain itu, Ralat Min (ME), Ralat Peratusan Mutlak Min (MAPE), dan Ralat Min Kuasa Dua Akar (RMSE) digunakan untuk menilai kaedah ramalan jenayah. Keputusan eksperimen menunjukkan kepentingan kaedah yang dicadangkan dalam meningkatkan ketepatan pengesanan hotspot jenayah dan ketepatan ramalan bagi kebanyakan jenayah kritikal berbanding kaedah terkini. Peramal jenayah boleh membantu agensi penguatkuasaan undang-undang menggunakan sumber polis dengan cekap dan berkesan.

CLUSTERING ENSEMBLE AND HYBRID OF DEEP LEARNING FOR SPATIO-TEMPORAL CRIME PREDICTIONS

ABSTRACT

The increase in the urban population poses challenges in managing services and safety from criminal activities. The concerned stakeholders intend to predict the time, location, number, and types of crimes to take suitable preventive measures. Accurate identification and prediction of crime hotspots can significantly benefit the concerned stakeholders in preventing crime by creating accurate threat visualizations and allocating police resources efficiently. Several techniques have been proposed for crime prediction, but they are limited in accuracy and predicting crime according to crime type on an hourly, monthly, and seasonal basis. Crime hotspot detection approaches are primarily sensitive to initial parameter selection and finding clusters of varying shapes and densities. Similarly, existing Crime prediction approaches are limited in capturing non-stationary data and long-term dependencies by focusing on crime types. Thus, the crime detection and prediction mechanisms need improvement in the number of crimes, crime span, accuracy, and dense crime region and prediction. The core objective of this study is twofold. First, it proposes a crime hotspot detection model to improve accuracy using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) and its clustering ensemble to capture varying shapes and densities clusters and improve accuracy. HDBSCAN is used with varying parameter initialization in the generation mechanism under the cluster ensemble paradigm. Moreover, six different distance measures are used to ensure diversity. In addition, an evaluation function is proposed parameterized by silhouette score to select the stable clustering among a pool of clustering solutions to ensure quality. Furthermore, this study

proposes an Edge Enhanced Hypergraph Construction (EEHC) algorithm as a consensus function by leveraging the strengths of the Hypergraph Partitioning Algorithm (HGPA) and Kernel Density Estimation (KDE) to capture the hierarchical structure and smoothing the edges by mitigating the impact of data imperfections in partitioning process. Three state-of-the-art hypergraph partitioning algorithms were compared to select the method with quality cluster partitioning. Second, this study proposes a Spatio-temporal Autoregressive Transformer (START) by leveraging the strengths of spatio-temporal transformer and Vector Autoregression (VAR) to predict crime according to crime types by handling non-stationary data and capturing long-term dependencies. One hot encoding is applied to distinguish crime types, followed by Seasonal Trend Decomposition using LOESS (STL) to decompose time series into three components (trend, seasonality, and remainder) to efficiently capture the residual variations overlooked for trend and seasonality components. Moreover, VAR attention can significantly enhance crime prediction accuracy by focusing on specific crime types, allowing the model to prioritize the relevant information during prediction. The proposed methods are evaluated on publicly available state-of-the-art New York, Chicago, Lahore, and Los Angeles crime datasets. This study evaluates crime hotspot detection methodology using the Silhouette score, Davies Bouldin Index (DBI), Calinski-Harabasz Index (CHI), Determinant Ratio Index (DRI), Root Mean Square Standard Deviation (RMSSTD) method, and the r-squared (RS) to check separation, connectedness, and compactness of clusters. Moreover, Mean Error (ME), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) are used to evaluate the crime prediction method. The experimental results show the significance of the proposed methods in enhancing crime hotspot detection accuracy and prediction accuracy for most critical crimes compared to the state-of-the-art methods. The crime predictor can assist law enforcement agencies in using police resources efficiently and effectively.

CHAPTER 1

INTRODUCTION

1.1 Overview

Crime is a social drawback, compromising the standard of everyone's life and affecting the developmental growth rate and the economic process of any society (Bogomolov et al., 2014; Podoletz, 2023). Considering the crime trends of a society, it can be a thoughtful and important issue while making decisions by individuals regarding whether or not they should move to a new town and what places they should avoid while traveling (Arulanandam, Savarimuthu, & Purvis, 2014; H. Chen & Cheng, 2023). With the constant growth in crime trends, advanced geographic data systems and new data processing techniques are the needs of the hour for law enforcement agencies to protect their societies better and tackle the crimes committed against them (Buczak & Gifford, 2010).

Researchers are pursuing research in criminal investigations and crime because of the vast amount of data available in recent years. Priority has been placed on analyzing crime trends and patterns to develop effective policies using the data to prevent these crimes from occurring and promote safer and more tranquil neighborhoods. Additionally, based on historical data, forecasting crimes has been a topic of interest that has attracted a lot of attention in research, leading to the proposal of a sizable number of unique approaches for identifying various factors linked to crime prediction (Yi, Yu, Zhuang, Zhang, & Xiong, 2018).

Moreover, with the availability of spatial-temporal information, people's decision-making regarding their living spot choices in an exceptional town can be improved. On the other hand, law enforcement agencies can utilize this information to allocate resources effectively and efficiently, broadening the measure of wrongdoing forecasts and looking to stop it before it happens. It will encourage the circulation of police powers at the most apparent wrongdoing places for some random time to concede conservative and greatly productive utilization of police assets (Nath, 2006).

The current century is often referenced as the *"City's Century"*, which reflects the unrivaled migration from rural to urban areas globally (Spencer & Butler, 2010). Among other challenges in urbanization, crime spikes concerning seasonality are becoming a challenging social problem. That affects the public's safety, education, health, development of children, and economic status of adults (Tayebi, Ester, Glässer, & Brantingham, 2014; H. Wang, Kifer, Graif, & Li, 2016). Figure 1.1 illustrates the crime trend in New York City (NYCOpenData, 2019).

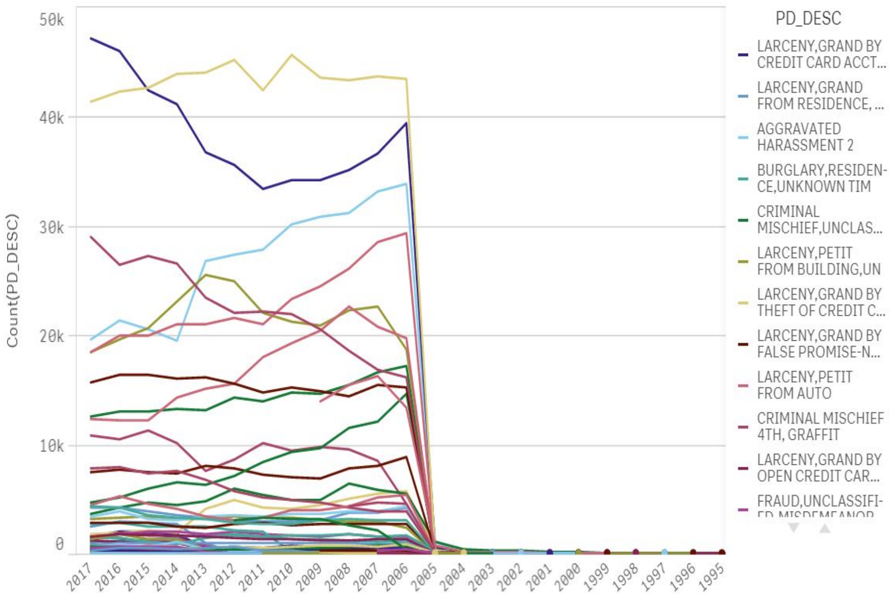

Figure 1.1: Crime patterns in New York across time (NYCOpenData, 2019).

Appropriate measures must be implemented to combat the rise in crime. One of the earliest statistical-based prediction tools for this purpose is predictive policing, which uses historical data to estimate the likelihood of future crime and the number of crimes and identify those responsible for previous crimes (Perry, 2013). A crime is based on the perpetrator, victim, time, and location where the crime occurred. Therefore, predictive policing techniques need to be able to answer the following questions:

- Who will be the perpetrator?
- Who will be the victim?
- What type of crime will be committed?
- What would be the location at which a crime would be committed?
- What would be the time a crime would be committed?

To study and comprehend crime trends and patterns, authorities must use modern technology as higher crime rates increase the complexities (Catlett, Cesario, Talia, & Vinci, 2019). The distribution of criminal episodes within a city is not uniform, as is well recognized in this context. Results retrieved from NYC crime data (NYCOpenData, 2019) are depicted in figure 1.2.

Crimes are viewed as spatial and temporal events. These occurrences are characterized by space, time, and attributes, where space identifies the crime's location, and time explains the moment it occurred (Wortley & Mazerolle, 2016). Understanding the relationship between crime, geography, and time helps us find spatio-temporal crime patterns. Machine learning methodologies examine single qualities, like robberies, or

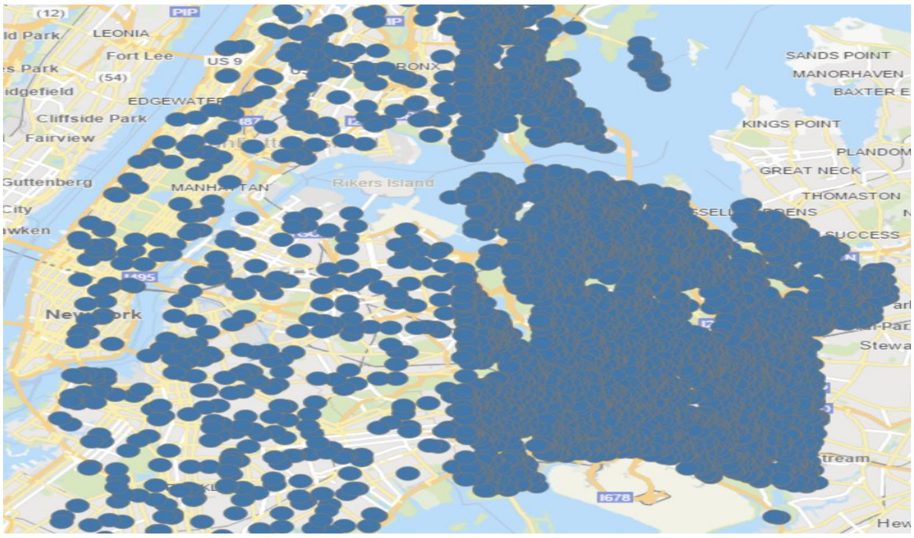

Figure 1.2: Distribution of criminal events over the years in NYC (NYCOpenData, 2019).

a particular pattern, like crime hotspots (Belesiotis, Papadakis, & Skoutas, 2018). In order to predict crime, spatio-temporal outliers, hotspots, coupling, and partitioning are significant. Finding unexpected occurrences requires finding spatio-temporal outliers, which are events with characteristics that differ dramatically from those of nearby objects (e.g., crime). Spatial and temporal clustering is the process of putting comparable crimes together. A concealed pattern of criminal activity can be found using this method. Locations with a high density of occurrences are known as spatio-temporal hotspots. It aids law enforcement agencies in identifying and reducing illegal activity.

Weisburd et al. (Weisburd, Eck, Braga, & Cave, 2016) identified five factors explaining crime concentration: (1) infeasible physical design, (2) insufficient and improper guardianship, (3) repeat victimization, (4) multiple targets, and (5) hot products. Some factors can be hard to control or impossible to change in residential areas. However, some measures can be taken to control and limit these factors to the maximum possible extent. By focusing on factors 1 and 2 and controlling them to the maximum extent possible, crime event risk can be reduced. Additionally, with knowledge of crime-affected areas, the possibility of taking suitable measures to counter the crimes can increase.

Criminology is the scientific study of crime, including its causes, responses by the criminal justice system, and prevention. The study of crime and criminal behavior is critical because it helps to understand and prevent criminal activity, promotes public safety, and improves the quality of life for citizens. Criminologists use data and research to understand why crimes occur, who commits them, and what can be done to prevent them. They can develop effective strategies to reduce criminal activity and protect communities by analyzing crime trends and patterns. Figure 1.3 depicts the numerous crimes committed in New York City over ten years.

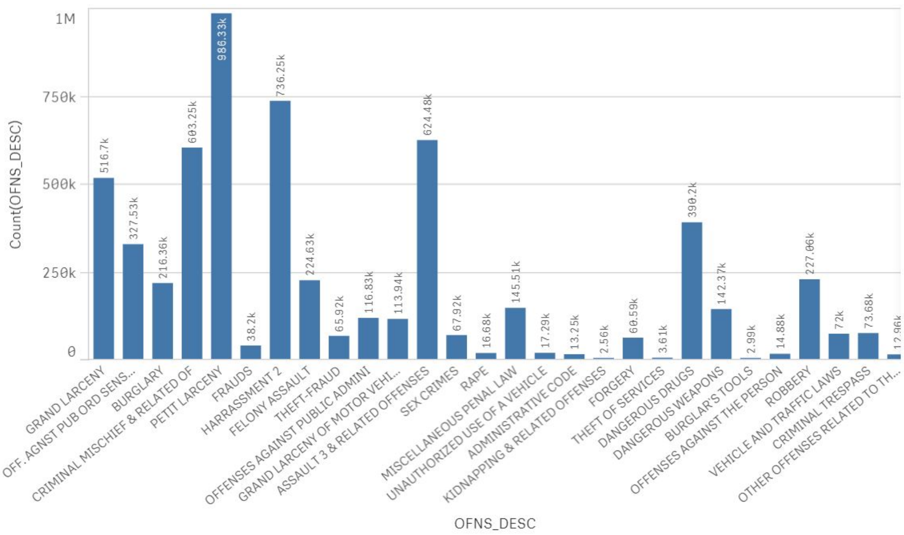

Figure 1.3: Distribution of various criminal incidents in NYC (NYCOpenData, 2019).

1.2 Motivation

Crime and criminal activity prevention are essential for a country's continued economic growth. Over the past few years, it has attracted researchers to contribute to crime identification and prediction techniques. Geographic Information System (GIS) technology allows police to visualize, analyze, and manage data geographically, helping to improve their understanding of crime patterns and hotspots. This enables them to allocate resources more effectively and respond to crime incidents more efficiently. In addition, GIS can help identify areas at a higher risk of crime and develop proactive strategies to prevent criminal activity. By incorporating crime data with other data sources, such as population demographics, land use, and transportation networks, police can develop more effective crime prevention strategies.

Informational and technological advancements allow police departments to collect, visualize, and track crime events. However, few publicly available datasets contain spatial and temporal information about crime events. GIS systems allow one to collect spatial and temporal information about a crime. The availability of spatial and temporal information urges researchers to apply data analytical methodologies to build predictive models for likely crime events. Figure 1.4 shows the significance of data analytics techniques in various aspects of crime analysis. This will help the forces use their resources efficiently and take preventive measures before a crime occurs.

The prospective analysis is a more proactive approach to crime prevention, allowing police to predict crime before it occurs and allocate resources accordingly. This approach is based on various factors such as historical crime data, demographics, land use, and environmental features. By using machine learning algorithms, GIS can analyze these factors to generate crime risk maps, helping police focus their efforts on areas that are most likely to experience crime. This results in the more effective use of resources and reduced crime rates (Wortley & Mazerolle, 2016).

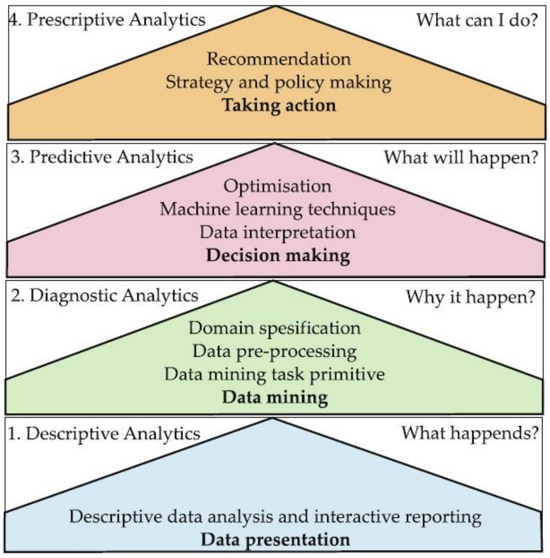

Figure 1.4: Crime analytics significance in various aspects of crime analysis (Farsi et al., 2018).

Thus, it is crucial to pinpoint places that pose more danger from crime than other regions. Mapping crime-dense zones has evolved into a highly sophisticated analytical tool to identify locations with a higher crime risk for efficient and effective allocation of police resources and suitable measures. A model must pinpoint these locations to stop crimes before they start and provide precise crime forecasts for each site.

Several efforts have been made in the literature for crime hotspot detection and prediction to help law enforcement agencies detect and predict crime (Butt, Letchmunan, Hassan, & Koh, 2022; Dakalbab et al., 2022). Clustering approaches provide promising results for crime hotspot detection compared to other state-of-the-art machine learning approaches (Gulati, Rajpoot, Gautam, & Seeja, 2023; R. Jain & Bhat, 2022). Mainly, Density-based approaches such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Hierarchical DBSCAN (HDBSCAN) detect clusters of varying shapes and densities (Catlett et al., 2019; Mandalapu, Elluri, Vyas, & Roy, 2023). Moreover, cluster ensemble further strengthens the cluster detection foundation by enhancing accuracy by including quality and diversity. However, generation and consensus mechanisms need the attention of the researchers to solve diversity, quality, stability, and accuracy issues (Mahmud, Huang, Ruby, Ngueilbaye, & Wu, 2023; M. Zhang, 2022). The graph-based consensus functions primarily show promising results with low computational complexity compared to others. In addition, Kernal Density Estimation (KDE) has also been used in literature for density estimation and detecting hotspots (Hinneburg & Keim, 2003; Kalinic & Krisp, 2018). It has also been used widely in the literature as a smoothing technique hybrid with other methods to make them robust to outliers and precisely capture density information (Scaldelai, Matioli, & Santos, 2022; Y. Wu, He, & Huang, 2020).

Similarly, for crime prediction, recent approaches like Autoregressive Integrated Moving Averages (ARIMA) (K. Islam & Raza, 2020), spatio-temporal Neural Network (STNN) (Y. Zhuang, Almeida, Morabito, & Ding, 2017), and deep learning (Jin, Sha, Feng, Cheng, & Huang, 2021), etc., work only on stationary data that does not possess clear trends and residuals. So, it may lead to unrealistic results and take more time to process. Moreover, ARIMA and Exponential Smoothing (ES) have been widely used for univariate analysis, but several researchers have focused on predicting crime according to crime type to enhance accuracy (Bonam et al., 2023; Dong, Ye, & Li, 2022). Vector Autoregression (VAR) has been widely employed in literature for multivariate analysis (Zivot & Wang, 2006). However, they are also limited in capturing long-term dependencies and focus mainly on temporal dimensions.

Spatio-temporal transformers have been introduced to handle both spatial and temporal dynamics mitigating the limitations of earlier methods (Grigsby, Wang, Nguyen,

& Qi, 2021; Wen et al., 2022). However, predicting crime by focusing on crime types is still unexplored. Moreover, capturing non-stationary seasonality and trends has also been a potential motivation for the researchers, as earlier research focuses on stationary data. Techniques like Seasonality Trend decomposition (STL) using Locally Estimated Scatter plot Smoothing (LOESS), Fourier transform, and wavelet transform have been used to preprocess time series to extract trends and seasonality for enhancing crime prediction accuracy (Butt et al., 2022; Ermshaus, Schäfer, & Leser, 2023; Theodosiou, 2011). Since no similar work has been reported throughout the literature, this investigation is a valuable reference for other researchers interested in time-series analysis to enhance crime prediction accuracy.

Moreover, generalizability is a significant problem in cutting-edge approaches that require substantial consideration by searchers. Given the differences in demographic trends, contributing causes of crime, criminal behavior, and regional cultures, the crime prediction algorithm should be globalized. A generic crime prediction model should be in place to account for variance in the boroughs or nearby locations. Due to its capacity to generalize a model, transfer learning has only recently become a significant field of research (Bashath et al., 2022; Iman, Arabnia, & Rasheed, 2023). It has been utilized to address various practical issues that need the dissemination of knowledge, like traffic prediction. (C. Zhang, Zhang, Qiao, Yuan, & Zhang, 2019), financial time forecasting (Q.-Q. He, Pang, & Si, 2019) and air quality prediction (Ma, Cheng, Lin, Tan, & Zhang, 2019).

1.3 Research Problem

Numerous studies on criminal justice show that crimes are not distributed equally among cities (Catlett et al., 2019). Geographical location can affect the crime rate; some places can be classified as low- or high-risk zones. Seasonal and yearly patterns might have an impact on crime rates. Over the past ten years, more investigation of their spatial features has been conducted. Analyzing crime hotspots is a crucial and well-liked strategy in this regard (Short et al., 2008), (Mohler, Short, Brantingham, Schoenberg, & Tita, 2011), (Short, D'Orsogna, Brantingham, & Tita, 2009), (Eck, Chainey, Cameron, & Wilson, 2005). In addition, there is a lack of research identifying crime according to crime types (Liang, Wang, Tao, & Cao, 2022; Safat, Asghar, & Gillani, 2021). Various studies have enlightened the significance of crime prediction by crime type (Albahli et al., 2021; AL Mansour & Lundy, 2019).

Several attempts have been made in the literature to detect crime hotspots using historical data (Kianmehr & Alhajj, 2006; Tayebi et al., 2014). The most well-known clustering k-means and agglomerative hierarchical clustering algorithms struggle with noisy data or nonconvex clusters (Agarwal, Nagpal, & Sehgal, 2013; Boongoen & Iam-On, 2018). Kernel Density Estimation (KDE) offers an alternative, excelling in capturing irregular shapes of crime hotspots (Pu, Yao, Li, & Alhudhaif, 2024). KDE has been used in literature with other methods as preprocess or post-process for providing smooth density gradients allowing a clear distinction between low and high density areas (Bortoloti, de Oliveira, & Ciarelli, 2021; K. Zhang & Chen, 2021). However, it is susceptible to bandwidth parameter selection that can lead to oversmooth or undersmooth gradients.

State-of-the-art hotspot detection methods such as DBSCAN (Catlett et al., 2019) and STKDE (Hu, Wang, Guin, & Zhu, 2018) suffer when clusters have different densities, arbitrary shapes and sizes. HDBSCAN addresses this by capturing varying shapes, sizes, and density clusters (Baqir, ul Rehman, Malik, ul Mustafa, & Ahmad, 2020). However, selecting optimal parameters remains a challenging task impacting clustering outcomes.

Cluster ensembles have emerged as a solution, combining multiple clustering solutions to enhance accuracy (Khalili, Rabbani, & Akbari, 2023; P. Zhou, Wang, Du, & Li, 2022). This involves building base clusters and the consensus outcome (T. Li, Rezaeipanah, & El Din, 2022). Diverse base clusterings are advantageous for ensembles to enhance performance. However, a reduction in ensemble quality may arise from utilizing clustering techniques that produce many base clustering (Koko, Yassine, Wahed, Madete, & Rushdi, 2023; P. Zhou, Du, & Li, 2023). Therefore, diversity and quality issues prevail that need careful selection of the quality and diversity tradeoff (Khan, Luo, Shaikh, & Hedjam, 2021; P. Zhou et al., 2022).

Graph-based consensus functions such as Hypergraph Partitioning algorithms (HGPA) offer low computational complexity and flexibility in handling varying shapes and sizes of clusters (Karypis & Kumar, 1998; Schlag et al., 2023). However, careful parameter tuning is necessary, primarily defining edge weights that can significantly impact clustering outcomes. Despite their strengths, mitigating edge effects remains challenging, requiring enhanced methods (Ü. Çatalyürek et al., 2023). In addition, hypergraph partitioning algorithms such as hMETIS, PaToH, and KaHYPaR play a crucial role by combining multiple clustering results to provide robust and quality results. However, selecting a tradeoff between quality clustering and efficiency remains challenging. Therefore, creating a consensus function less sensitive to ensemble size, quality, and diversity is challenging (C. Liu, Liu, & Osmani, 2023; J. Yan, Liu, et al., 2022).

Similarly, for crime prediction, traditional time-series analysis approaches like Autoregressive Integrated Moving Averages (ARIMA) models and Exponential smoothing (ES) have been widely used (Adenomon & Oyejola, 2014; Butt et al., 2021; Sirisha, Belavagi, & Attigeri, 2022) for short-term forecasting. However, they struggle with complex non-linear trends, multivariate relationships, assuming stationary data, and seasonal variations in the data. Conversely, Vector Autoregression (VAR) is suitable for multivariate analysis and can handle both stationary and non-stationary data by using lagged values to capture trends for short-term forecasting (Munkhdalai, Li, Theera-Umpon, Auephanwiriyakul, & Ryu, 2020; Siraj-Ud-Doulah, 2019). However, VAR models are limited in handling complex non-linear relationships and long-term dependencies(G. Li & Jung, 2023).

One possible solution is deep learning models such as LSTM (B. Wang, Zhang, Zhang, Brantingham, & Bertozzi, 2017), BiLSTM(K. Islam & Raza, 2020), and STNN (Y. Zhuang et al., 2017) are capable of handling complex non-linear, Intricate patterns, long-term dependencies, and can provide generalized representations. However, these traditional deep-learning models primarily focus on temporal dependencies, which are lacking in capturing spatial geographic distributions. Moreover, they also struggle to capture long-term dependencies over longer horizons due to vanishing gradients (Bi, Yuan, Liu, Wang, & Zhang, 2022; Tong, Limperis, Hamza-Lup, Xu, & Li, 2024).

Recently, spatio-temporal transformers have enriched enough to model spatial and temporal relationships in the data (Grigsby et al., 2021). In addition, they can capture long-term dependencies over longer horizons using attention mechanisms (Banda, Bhuiyan, Hasan, Zhang, & Song, 2021; H. Wang, Zhang, Liang, & Liu, 2023). However, they struggle to find interdependencies in crime types and non-stationary trends and patterns of data. Therefore, there is a need to develop enhanced spatio-temporal transformers that can focus on specific crime types and can handle both stationary and non-stationary data. One possible solution is to use VAR as an attention mechanism to focus on specific crime types.

1.4 Research Question

This study is designed to answer the following research questions.

1. How can we leverage clustering ensemble to achieve a better tradeoff between quality and diversity in identifying crime hotspots compared to individual clustering algorithms?
    - How effective is the HDBSCAN in generation mechanism for cluster ensemble with varying parameter initialization when dealing with clusters of varying shapes and densities to improve cluster quality and diversity?
    - Can the graph-based consensus function be enhanced with KDE to mitigate the edge effects and improve the robustness of the model?
2. How to enhance the architecture of spatio-temporal transformers to improve the crime prediction accuracy considering crime types?

- How effective is the VAR attention mechanism with spatio-temporal transformers to enhance its ability to predict crime according to crime types?
- Can STL decomposition as pre-processing enhance spatio-temporal transformers to capture non-stationary trends and seasonality, significantly leading to more accurate predictions?

1.5 Research Objective

This study aims to address the following research objectives to enhance crime hotspot detection and prediction accuracy:

1. To propose a cluster ensemble approach that balances quality and diversity in crime hotspot detection
    - Investigate the effectiveness of HDBSCAN with varying parameter initialization to identify clusters of varying shapes and sizes as a generation mechanism in cluster ensemble for ensuring clustering diversity.
    - Examining the enhancement of graph-based consensus function by incorporating KDE to mitigate the edge effects to improve the robustness of the crime prediction model.
2. To propose an enhanced spatio-temporal transformer architecture for crime prediction according to crime type
    - Investigate incorporating the VAR attention mechanism in spatio-temporal transformers to enhance crime prediction accuracy by focusing on crime types.

- Evaluate the impact of pre-processing crime data leveraging STL decomposition for spatio-temporal transformers to handle non-stationary trends and seasonality for providing more accurate crime predictions.

1.6 Contribution

This study contributed in the following ways:

1. Develop a cluster ensemble framework utilizing HDBSCAN as a generation mechanism with varying parameter initialization to improve cluster quality and diversity by capturing varying shapes and density clusters. Proposed an enhanced graph-based consensus function incorporating KDE to mitigate edge effects and improve the overall robustness of the model.
2. Evaluate and establish the effectiveness of the VAR attention mechanism in spatio-temporal transformers to enhance crime prediction accuracy considering crime types. Investigate and demonstrate the benefits of pre-processing crime data using STL to lead to more accurate crime predictions by capturing non-stationary trends and seasonality along shorter and longer horizons.

1.7 Significance of the Study

Crime and criminal activity detection and prevention strategies are essential to keep the ordinary public safe. It is a primary concern for a country's better economic growth. The results can significantly benefit the stakeholders if the crime hotspots are accurately identified. Creating real threat visualizations can help the public and law enforcement agencies effectively detect dense crime areas and allocate police resources more efficiently by forecasting the number of crimes in those areas. It will create a healthy and nourishing environment and improve economic growth substantially. Besides, it will help the communities by alerting the neighborhood watch or patrol departments during the high probability of a crime or suggesting students or business travelers plan their stay a bit safer.

Traditional clustering algorithms for crime hotspot detection are limited in capturing diverse data characteristics. This study leverages the cluster ensemble paradigm to generate diverse clustering solutions to mitigate the effect of parameter initialization and capture varying clusters' shapes and sizes. In addition, this study utilizes the strengths of KDE and graph-based consensus function to mitigate edge effects and capture more nuanced hierarchical relationships in the data to enhance the robustness of the model. This will give law enforcement agencies a more precise understanding of crime-dense regions to implement more targeted crime prevention strategies.

Similarly, the VAR attention mechanism has been incorporated in spatio-temporal transformers to focus on specific crime types, leading to more comprehensive insights and enhancing the accuracy of crime prediction. Processing crime data with STL decomposition helps the traditional model to focus on underlying trends, patterns, and seasonality effectively.

1.8 Research Scope

This study covers the following scope in various experiments. This study investigates state-of-the-art cluster ensemble techniques and their limitation in crime hotspot detection. Moreover, the research will develop a cluster ensemble framework that balances quality and diversity tradeoffs, potentially leveraging HDBSCAN and K-means as a generation mechanism with varying parameter initialization. The scope includes enhancing graph-based consensus function by leveraging KDE to mitigate edge effects and improve the overall robustness of crime hotspot detection. Furthermore, this study investigates the significance of incorporating VAR as an attention mechanism in spatio-temporal transformers to predict crime according to crime types to enhance prediction accuracy. In addition, the significance of pre-processing crime data using STL decomposition is investigated for spatio-temporal transformers to handle non-stationary trends and seasonality, which is within the scope.

1.9 Basic Preliminaries

This section discusses the fundamental notations utilised to discover and forecast crime hotspots for this study. There are four components: the location set (a pair of coordinates for the precise site of the crime), the area under study (coordinates in a specific area of interest), the hotspot (crime Hotspot Area), and the crime timestamp list (crime occurrence Ordered timestamp list). Let D be a dataset consisting of timestamp crime instances, $D = \left\{d_1, d_2, d_3, \ldots d_k\right\}$ where each d_i is a criminal record describing the following features.

1.9.1 Spatio-temporal Location Set

A location set, L, is the set of all geocoded points where crime has occurred. A point $l \varepsilon L$ is a point in a location specified by a coordinate system's combination of latitude and longitude. These points specify the crime location in the particular area under study $L = l_{1(lat,long)}, l_{2(lat,long)}, l_{3(lat,long)} \ldots l_{k(lat,long)}$. Let $T = \{t_1, t_2, t_3, \ldots t_k\}$

be an ordered set containing temporal information about a crime incident in hourly, weekly, monthly, and yearly time intervals such as $t_h < t_{h+1}$.

1.9.2 Area Uunder Study

The area under study A is the maximum area bounded by the location set L in Euclidean space.

1.9.3 Hotspot

A hotspot is an unusual type of cluster with more activity, bounding, danger, dense area, or violence than a typical cluster. A hotspot can be defined mathematically as:

Let HF be the hotspot function, $HF : DC \xrightarrow{T} HS$ where DC be the set of spatio-temporal dense cluster $\{c_1, c_2, \ldots c_k\}$ and HS be the set of hotspots $\{HS = h_1, h_2, h_3, \ldots h_k\}$. So, a hotspot function can be further elaborated as follows:

$HS(c_i) = h_i$, it means any cluster (C_i) may contain a hotspot (h_i), but the hotspot (h_i) may not be present in every cluster (C_i).

1.9.3(a) Mathematical Properties of Hotspot

1. $h_i \subseteq c_i$, i.e. every hotspot may be a part of a spatio-temporal cluster, but some clusters may not have a hotspot and $|h_i| < |c_i|$.

2. The density of a hotspot can be greater than or equal to the density of a typical cluster Density(h_i) >Density(c_i). In an exceptional case, the density will be equal if the whole cluster is a hotspot.

The goal of a crime prediction study is to predict the future timestamp of crime events such as $F = \{t_f, t_{f+1}, \ldots t_k\}$ with $f < k$. A study should aim to obtain a reliable and accurate number and location of a crime using a specified timestamp such as $t_f \varepsilon F$. More precisely, to accomplish the following goals:

1. Locate a set of Crime Hotspot Regions (CHR) or (Crime dense, blobs):

   $CHR = \{CHR_1, CHR_2, \ldots CHR_K\}$

   where CHR_K is a spatio-temporal crime hotspot region in which crime incidents occur more frequently or in higher density than other neighbourhoods in a city.

2. Predict the number of crimes likely to happen given a time stamp $t_f \varepsilon F$. The aim is to present a prediction function $P_{crime} : F \xrightarrow{A} (CHR, A)$ given an area A that predicts the crime hotspot regions $CHR_i \varepsilon CHR$ where a crime is likely to happen.

1.10 Organization of the Thesis

This study is divided into the following chapters: **Chapter 2: Literature Review:** This section discussed the cutting-edge crime hotspot identification and prediction methods and emphasised the difficulties encountered.

Chapter 3: Proposed Methodology: This section presents the Methodology with the experimental datasets used and evaluation measures to evaluate the performance of the proposed algorithm.

Chapter 4: Experimental Evaluation: This section covers the experimental and comparative analyses with the state-of-the-art approach.

Chapter 5: Conclusion: This section concludes the study by summarizing the findings with possible future directions.

CHAPTER 2

LITERATURE REVIEW

This section covers recent advances in detecting and predicting spatio-temporal crime hotspots. Moreover, section 2.1 presents the significance of this study. Section 2.2 summarizes the state-of-the-art data mining and machine learning exploited for crime analysis. Primarily, three significant aspects of this study are presented. First, section 2.3 presents crime hotspot detection techniques with advantages and disadvantages. Second, section 2.4 discusses crime hotspot prediction techniques with advantages and challenges. Finally, the significance of the spatial-temporal aspect of crime datasets is highlighted in section 2.5.

2.1 Introduction

Crime hotspot detection and prediction is a crucial research topic to ensure public safety worldwide. Crime hotspot detection and prediction play a significant role in empowering law enforcement agencies to allocate resources effectively and ultimately prevent criminal activities (Perry, 2013). However, predicting crime accurately is a crucial task due to its inherent complexities. It is to capture not only spatial dynamics but also temporal and intricate spatial relationships. Researchers have found different parts of crime prediction, such as historical data (Jenga, Catal, & Kar, 2023; Yu, Ding, Morabito, & Chen, 2016), information about geography, and demographics (Geldenhuys, 2023; X. Wang & Brown, 2011).

Existing crime prediction methods have relied on spatial statistics and basic time series analysis techniques (Buczak & Gifford, 2010; Tayebi et al., 2014; Yi et al., 2018). Although these techniques have provided significant insights, they often fall short of capturing the intricate relationships between spatial and temporal dynamics. Spatial methods such as SatScan, Getis-Ord Gi, and Kernel Density Estimation (KDE) are good at detecting hotspots but struggle to capture temporal patterns. Time series analysis techniques such as Autoregressive Integrated Moving Averages (ARIMA) and Exponential Smoothing (ES) effectively capture temporal distributions but neglect the spatial dynamics where crime unfolds (Buczak & Gifford, 2010; Tayebi et al., 2014; Yi et al., 2018). These gaps highlight the necessity of more nuanced methodologies that can capture rich spatio-temporal information in the crime data.

Recent technological advancements in machine learning offer promising directions to address these gaps (Hassani, Huang, Silva, & Ghodsi, 2016; Kajita & Kajita, 2019), as shown in figure 2.1. It is used to discover hidden clustering patterns and trends and find crime hotspots of all shapes and sizes. These include density-based algorithms like Hierarchical Density-Based Spatial Clustering of Application with Noise (HDBSCAN) and graph-based algorithms like the Hypergraph Partitioning Algorithm (HGPA) Catlett, Cesario, Talia, and Vinci (2018); Cecaj, Lippi, Mamei, and Zambonelli (2020). To make more reliable solutions, clustering ensemble improves the skills of clustering algorithms like supervised learning (J. Xu, Li, Zhang, & Wu, 2024). Moreover, it enhances the robustness and flexibility of hotspot detection by utilizing the strength of multiple clustering algorithms.

However, hotspot identification alone is not sufficient for effective crime preven-

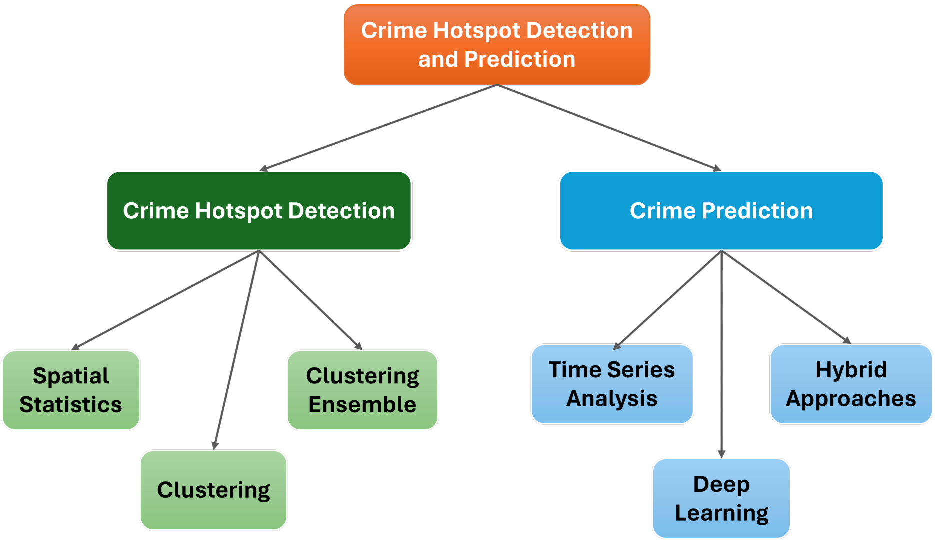

Figure 2.1: Literature classification for crime hotspot detection and prediction

tion. Predicting crime occurrences requires deeper knowledge of temporal and spatio-temporal dynamics. Deep learning models have shown promising results in capturing intricate patterns in spatio-temporal data. Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and Transformers have shown exceptional capabilities in capturing and learning these relationships with significant accuracy. Moreover, the attention mechanism, in conjunction with deep learning models such as Long Short Term Memory (LSTM) and Transformers, has also shown promising results by focusing on relevant data (Fu et al., 2024; X. Liu & Zhou, 2024). By focusing on specific regions and timestamps, they can provide valuable insights into potential future crime hotspots.

Yet, relying alone on past crime data can overlook the influence of external factors that have a significant impact on crime trends and patterns (Helbich & Leitner, 2017; Kapoor, Singh, & Cherukuri, 2020). Crime type, seasonality, and economic fluctuations can play a critical role in shaping criminal activities. Existing research primarily focuses on univariate data and utilizes ARIMA and ES variants. However, crime trends and patterns vary according to crime types, which need multivariate analysis. Vector Autoregression (VAR) with exogenous variables provides a robust framework by providing comprehensive and nuanced predictions (Gabauer, Gupta, Marfatia, & Miller, 2024; Hopfe, Lee, & Yu, 2024). Similarly, the transfer learning paradigm has also been adopted in the literature to solve data insufficiency problems and to solve and understand problems under the same umbrella. It can help to understand crime dynamics in the neighboring regions (Laurer, Van Atteveldt, Casas, & Welbers, 2024).

Through this comprehensive approach, this research endeavors to make a significant stride toward achieving more robust and accurate results. Primarily, this research aims to help law enforcement agencies enhance public safety and proactive crime prevention strategies.

2.2 Background for Crime Analysis

Crime can be categorized as a local crime in the street, like theft and assault, etc., or a well-planned international crime with a significant impact (Prathap, 2022; Wei, Xu, Qin, & Wang, 2014). As with traditional data mining, mining of crime events is concerned with privacy (Kargupta, Liu, & Ryan, 2003). In structured data, patterns or different trends and variations in crime events are discovered using diverse data mining techniques, like clustering, association, prediction, classification, and analysis of outliers (J. Han, Kamber, & Pei, 2011). Advanced data mining approaches can handle structured and unstructured data to discover different trends and patterns (H. Chen et al., 2004; Gupta, 2014; Tasnim, Imam, & Hashem, 2022). This section focuses on cutting-edge criminal data mining strategies, as Figure 2.2 demonstrates.