Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

CROSS-CITY CRIME RISK PREDICTION USING UNSUPERVISED DOMAIN ADAPTATION

[1]Mittameedi Indra Kumar, MCA Student, Department of MCA
[2] B. Harish Kumar Reddy, M.Tech, (Ph.D), Assistant Professor, Department of MCA
[12]Dr KV Subba Reddy Institute of Technology, Dupadu, Kurnool

ABSTRACT
Predicting crime risk is essential for both city safety and the standard of living for locals. However, it is difficult to forecast the danger of crime in cities without labelled data. Collecting high-quality labelled crime data is not easy for many places because of maintenance expenses and municipal rules. Specifically, some cities may contain a large amount of labelled data, while others may have a little amount. By learning from a city with a wealth of data, a crime prediction model for a place without labelled crime data has been developed. However, this prediction job is made more challenging by the disparity in pertinent background data across cities. In order to solve the problem of context inconsistency, this study suggests an efficient unsupervised domain adaptation model (UDAC) for crime risk prediction across cities. More precisely, for every target city grid, we first find a number of comparable source city grids. We then create auxiliary contexts for the target city based on these source city grids in order to maintain consistency between the two cities. The goal of a dense convolutional network with unsupervised domain adaptation is to concurrently learn domain-invariant features for domain adaptation and high-level representations for precise crime risk prediction. Three real-world datasets are used in comprehensive tests to confirm our model's efficacy.

A key element of the expanding discipline of data science is machine learning. Various algorithms are taught using statistical techniques to create predictions or classifications and to provide important information for this project. These insights then inform business and application decision-making, hopefully influencing important growth indicators.

This project data, sometimes referred to as training data, is used by machine learning algorithms to create a model that allows them to make judgements or predictions without explicit programming. When developing traditional algorithms to accomplish the required tasks is challenging or impractical, machine learning methods are used in a broad range of datasets.

I. INTRODUCTION

Crime consistently jeopardises urban safety and degrades the standard of living for its residents. There have been 435 mass shootings in the United States in 2019, according to [1], which have left 517 people dead, 1648 injured, caused significant property damage, and caused unimaginable sorrow. In order to avoid and lessen possible criminal incidents, it is crucial for both individuals and society as a whole to recognise crime risk. Thankfully, researchers have never-before-seen opportunities to investigate crime-related issues like crime hotspot detection [2], [3], crime classification [4], [5], [6], crime rate inference [7], [8], and crime count prediction [4], [5], [9], [10], [11] thanks to the availability of diverse urban data in some cities (like Chicago). 38 A certain

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

quantity of urban data has been shown to be useful 39 for enhancing the effectiveness of research pertaining to crime. Human movement, for instance, may have an impact on the incidence of criminal occurrences; a larger crowd may increase the likelihood of stealing.

However, since 44 communities have varying levels of development, some do not make data publicly available for a variety of reasons, including the high expense of data collection and upkeep, the lack of precise legislation, and growing privacy concerns. Residents must thus have enough experience to determine if there may be danger. However, not all locals have this kind of expertise, which makes things more difficult for visitors and other newcomers. A novel paradigm known as transfer learning [12], [13] has recently emerged that allows us to apply information acquired in a data-rich city (source city) to equivalent tasks in a data-scarce city (target city), such as crowd flow prediction and chain store site choice [14], [15]. Therefore, in order to investigate crime risk prediction in cities without labelled crime data, we try to use unsupervised transfer learning.

Nevertheless, a prediction model trained with sufficient labelled data from a source city may not be able to estimate crime risk in the target city in the absence of labelled crime data. The relevant context data that is accessible in various cities may vary depending on the data gathering capabilities. Assume that Los Angeles (LA) is the target city and New York City (NYC) is the source city. NYC has gathered multisource urban data throughout the years and continues to make information available to the public, such as point of interest (POI) distribution and cab trip records, thanks to its extensive use of detection equipment and long-running open data initiative. Additionally, some cities gather a lot of urban data. However, certain important and pertinent data, including taxi mobility statistics in locations like Los Angeles, are not made publicly accessible owing to privacy concerns or hefty data gathering expenses. The effectiveness of crime risk prediction in cities with unlabelled data is therefore hampered by the contexts' inconsistency problem.

Using just common context data to train a model from the source city and then fine-tuning this model to perform tasks in the target city, while leaving inconsistent city-specific context data alone, is an obvious way to address context inconsistencies. However, this could lose some important information for predicting crime risk, and it might become worse when there is a lack of context data. Since deep learning technologies are so common, more data might result in improved performance since they can extract more valuable information from networks. In order to solve the contexts' inconsistency issue, we therefore attempt to create an efficient model that can provide potential source-city specific context data for the target city. Then, using the information that has been acquired from the source city, we forecast crime in the target city.

In this paper, we solve the context inconsistency problem by proposing an unsupervised domain adaptation model (UDAC) for crime risk prediction across cities. By applying the knowledge gained from a source city with a wealth of labelled data to the target city, UDAC aims to tackle the problem of crime risk prediction in a target city without labelled crime data. Motivated by [15], we devise a technique to solve the problem of inconsistent context by building potential city-specific context data for the target 98 city grids using context data in comparable source city grids.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

Then, for the success of unsupervised domain adaptation, we offer a network that concurrently learns domain-invariant features and effective features for crime risk prediction in the source city. Three factors would be taken into account at the same time throughout the optimisation process: distribution discrepancy distance, domain categorisation error, and crime risk prediction error. Three real-world datasets from NYC, Chicago, and LA are used in extensive trials to confirm UDAC's efficacy.

The following are our work's primary contributions.

1) Our study addresses the problem of context inconsistency across cities and is a promising start towards unsupervised domain adaptation in crime prediction across cities.

2) Using information from a source city with a wealth of labelled data, we provide an efficient model for crime risk prediction in cities without labelled data that can support deep unsupervised domain adaptation. We first create city-specific contexts for the target city in order to solve the inconsistent contexts between the two cities. After that, we offer a dense convolution network that learns domain-invariant features for unsupervised domain adaptation and effective features for precise crime prediction. It may be possible to anticipate crime in the target city using the optimised network.

3) We use real-world datasets from three cities in our comprehensive tests to demonstrate the efficacy of our proposed UDAC model. The experimental findings demonstrate that our approach works better than the most advanced comparison techniques.

II. LITERATURE SURVEY

M. S. Gerber, "Using kernel density estimation and Twitter to predict crime"

Because Twitter is widely utilised both domestically and internationally, there are several potential to integrate Twitter-driven predictive analytics into decision support systems. Because its millions of users openly discuss events, emotions, and countless other topics, its content is created and shared in real time at no cost, and individual messages—also called tweets—are frequently tagged with exact spatial and temporal coordinates, Twitter is a perfect data source for decision support. Research on the use of spatiotemporally tagged tweets for crime prediction is presented in this article. We automatically identify debate topics in a large U.S. metropolis using statistical topic modelling and Twitter-specific language analysis. We next integrate these themes into a crime prediction model and demonstrate that the inclusion of Twitter data enhances crime prediction performance compared to a traditional technique based on kernel density estimation for 19 of the 25 crime categories we analysed. We pinpoint many performance snags that could affect Twitter's use in a real-world decision support system. We also highlight key directions for future study in this field, such as temporal modelling, including auxiliary data sources, and doing a more thorough semantic analysis of message content. Decision-makers in criminal justice who are in charge of allocating funds for crime prevention will be particularly impacted by this study. In a broader sense, this study has consequences for decision makers who are interested in the geographic areas where Twitter users reside.

Using a combination of social media and urban data, D. Yang, T. Heaney, A. Tonon, L. Wang, and P. Cudré-Mauroux, 710 "CrimeTelescope: Crime hotspot prediction

Crime is a complicated societal problem that affects a large portion of a society's population. In many nations, preventing and decreasing crime is of utmost importance. Finding efficient ways to use the limited resources available for

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

police and crime reduction is often essential. Crime hotspot prediction has been proposed as a means to achieve this objective. Crime hotspot prediction uses historical data to pinpoint regions that are likely to experience crime in the future. However, the majority of current methods for predicting crime hotspots ignore the predictive potential of additional data, such as urban or social media data, and instead only utilise past crime records to pinpoint crime hotspots. In this work, we offer CrimeTelescope, a platform that uses a combination of several data sources to anticipate and visualise crime hotspots. Our program regularly gathers online data on crime, metropolitan areas, and social media. After that, it uses both statistical and linguistic analysis to identify important aspects from the gathered data. Lastly, it uses the collected data to identify criminal hotspots and provides interactive map visualisations of the hotspots. Using real-world data from New York City, we demonstrate that, in contrast to traditional methods that rely only on historical crime records, integrating several data sources may significantly increase the accuracy of crime hotspot predictions (by as much as 5.2%). Furthermore, we use a System Usability Scale (SUS) survey on a complete CrimeTelescope prototype to show how user-friendly our platform is.

III. SYSTEM ANALYSIS

EXISTING SYSTEM

Our approach is connected to earlier research on unsupervised domain adaptation and crime prediction. We provide a quick overview of some relevant work from these two areas in this section.

In recent decades, there have been a few studies on the forecast of urban crime. It is important to identify relevant external factors for crime prediction studies. Ranson [16] examined meteorological data, such as temperature and weather information, and discovered that this data may be related to crime. 145

In order to better understand crime, Zhou et al. [17] used a variety of urban data, such as weather, the distribution of POI, and taxi trip data. They discovered a useful relationship between crime and these variables. It has been common practice to create efficient models using characteristics analysed in order to get precise prediction. Over the past few decades, numerous spatio-temporal prediction models have been proposed to capture spatial and temporal dependencies in order to solve a variety of tasks, including logistics management optimisation [23], social event prediction [21], traffic prediction and inference [18], [19], and [20], as well as air quality prediction [22]. One

However, when it came to prediction tasks without labelled data, these spatiotemporal prediction models put in little effort. Numerous approaches (such as linear models, count models, and machine learning models) have been applied to enhance prediction performance for crime prediction using a vast quantity of data, such as taxi trip, Twitter, demographic, and Foursquare data [9], [24], and [25]. Using crime data, POI, and 311 public service complaint data, Huang et al. [4] developed a hierarchical recurrent neural network with an attention layer to identify dynamic trends and learn temporal significance for predicting future crime occurrences. In order to forecast crime hotspots in New York City, Yang et al. [3] used Twitter and POI data in a variety of machine learning models (such as random forest and decision tree). 169

In order to extract spatiotemporal information and enhance future crime prediction ability, Yi et al. [26] suggested an integrated model employing the clustered continuous conditional random field (CCRF) approach. Additionally,

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

they added stacked denoising autoencoders to learn pairwise interactions between spatial areas and long short-term memory (LSTM) units to the previously stated CCRF technique to learn nonlinear relationships between the input and the output [10]. In their hierarchical approach for road-level crime prediction, Zhou et al. [27] first estimated prior knowledge about crimes using spatiotemporal data, and then they updated the findings by adding recurrent crime features. In order to forecast future road-level crime risk, they suggested a zero-inflated negative binomial regression model and looked at crime dynamics from the standpoint of influence propagation [28].

Disadvantages

- The system has not implemented an efficient model that can construct potential source-city specific context data for the target city to address the contexts' inconsistency problem, and then predict crime in the target city using learnt knowledge from the source city.
- The current methodology does not implement the Auxiliary Features' Construction method.

PROPOSED SYSTEM

The system addresses the contexts' inconsistency problem by proposing an unsupervised domain adaptation model (UDAC) for crime risk prediction across cities. By applying the knowledge gained from a source city with a wealth of labelled data to the target city, UDAC aims to tackle the problem of crime risk prediction in a target city without labelled crime data. Motivated by [15], in order to overcome the problem of inconsistent context, we develop a method97 to build potential city-specific context data for the target city grids based on context data in comparable source city grids. Then, for the success of unsupervised domain adaptation, we offer a network that concurrently learns domain-invariant features and effective features for crime risk prediction in the source city. The optimisation method would simultaneously take into account three factors: distribution discrepancy distance, domain categorisation error, and crime risk prediction error. Three real-world datasets from NYC, Chicago, and LA are used in extensive trials to confirm UDAC's efficacy.

1) Our study addresses the problem of context inconsistency across cities and is a promising step towards unsupervised domain adaptation in crime prediction across cities.

2) Using information from a source city with a wealth of labelled data, we provide an efficient model for crime risk prediction in cities without labelled data that can support deep unsupervised domain adaptation. We first create city-specific contexts for the target city in order to solve the inconsistent contexts between the two cities. After that, we offer a dense neural network that learns domain-invariant features for unsupervised domain adaptation and effective features for precise crime prediction. It may be possible to anticipate crime in the target city using the optimised network. 124

3) We use real-world datasets from three cities in our comprehensive tests to demonstrate the efficacy of our proposed UDAC model. The experimental findings demonstrate that our approach works better than the most advanced comparison techniques.

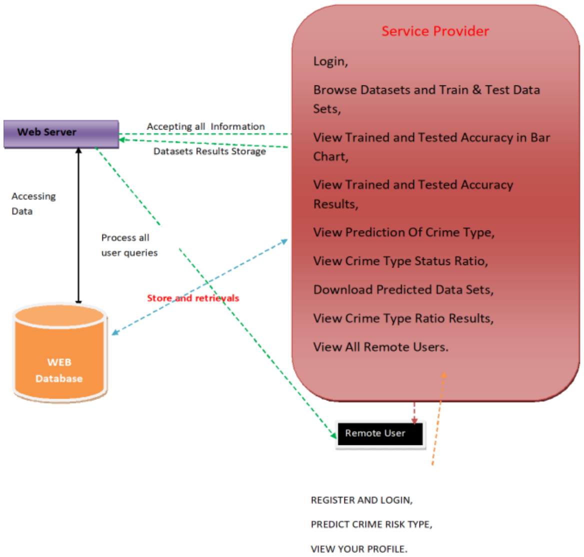
Advantages

- An efficient model that can build potential source-city specific context data for the target city in order to solve the issue of context inconsistency, and then use the information gained from the source city to forecast crime in the target city.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

- Dense Convolutional Network with Unsupervised Domain Adaptation was implemented by the system in this study.

SYSTEM ARCHITECTURE



IV. SYSTEM IMPLEMENTATION

Modules

Service Provider

The Service Provider must use a working user name and password to log in to this module. Following a successful login, one may do many tasks such Examine datasets, train and test datasets, and See the results of the trained and tested accuracy, the bar chart showing the accuracy, the crime type prediction, the crime type status ratio, the predicted data sets that were downloaded, the results of the crime type ratio, and the view of all remote users.

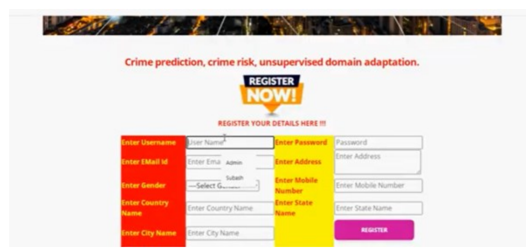
View and Authorize Users

A list of every user registered in this module is visible to the administrator. Here, the administrator may see user data such as name, email address, and address. They can also authorise users.

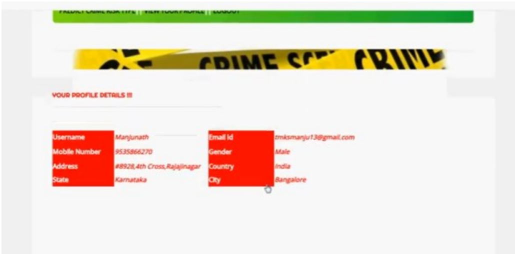
Remote User

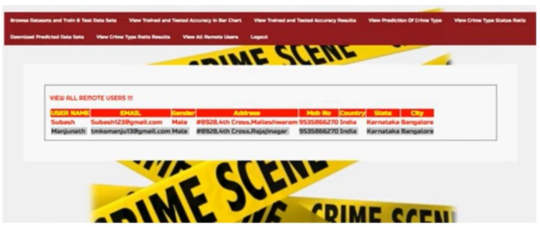
A total of n users are present in this module. Before beginning any actions, the user needs register. Following registration, the user's information will be entered into the database. Following a successful registration, he must use his password and authorised user name to log in. Following a successful login, the user may do tasks including registering and logging in, predicting the kind of crime risk, and seeing their profile.

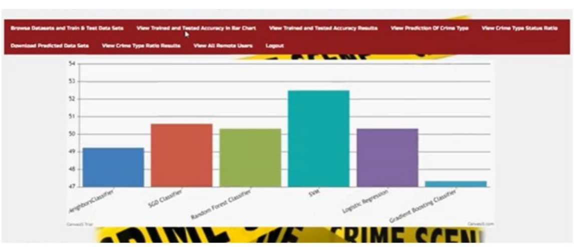
V. RESULTS



International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025







VI. CONCLUSION

In this paper, we offer an unsupervised domain adaption approach to perform the prediction problems using unlabelled data. This approach may estimate crime risk by using labelled data from a source city to transmit knowledge to the target city while resolving the

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

problem of context inconsistency across cities. For each target grid, we locate many comparable source city grids and divide cities into numerous equal-sized grids. To solve the issue of context inconsistency between cities, we build auxiliary features for the target city based on these pairings. Then, in order to learn from the source city and apply it to the target city for future crime risk prediction, we suggest an unsupervised domain adaptation module based on dense-convolutional6networks. To make knowledge transfer easier, domain-invariant characteristics are learnt. Our method's efficacy is confirmed by extensive trials utilising real-world data from NYC, Chicago, and LA. The experimental findings demonstrate that our suggested approach outperforms a number of cutting-edge comparison techniques.

We want to enhance our work in the future from a number of angles. Initially, we want to investigate prediction performance in cases when there is a more significant discrepancy between the source and destination cities. Next, we want to look at fine-grained unsupervised criminal risk prediction, including forecasting the likelihood of crime on roadways. Due to a serious data shortage issue, this would be more difficult.

REFERENCES

[1] (2020). GunViolenceArchive. List of Mass Shootings in the United States in 2019. [Online]. Available: https://www.gunviol encearchive.org/reports/mass-shooting

[2] M. S. Gerber, "Predicting crime using Twitter and kernel density estimation," Decision Support Syst., vol. 61, pp. 115–125, May 2014.

[3] D. Yang, T. Heaney, A. Tonon, L. Wang, and P. Cudré-Mauroux,"CrimeTelescope: Crime hotspot prediction based on urban and social media data fusion," World Wide Web, vol. 21, pp. 1323–1347, Sep. 2018.

[4] C. Huang, J. Zhang, Y. Zheng, and N. V. Chawla, "DeepCrime: Attentive hierarchical recurrent networks for crime prediction," in Proc. 27th ACM Int. Conf. Inf. Knowl. Manage., 2018, pp. 1423–1432.

[5] L. Xia et al., "Spatial–temporal sequential hypergraph network for crime prediction with dynamic multiplex relation learning," in Proc. 30th Int.Joint Conf. Artif. Intell., Aug. 2021, pp. 1631–1637.

[6] P. Das, A. K. Das, J. Nayak, D. Pelusi, and W. Ding, "Incremental classifier in crime prediction using bi-objective particle swarm optimization,"Inf. Sci., vol. 562, pp. 279–303, Jul. 2021.

[7] H. Wang, D. Kifer, C. Graif, and Z. Li, "Crime rate inference with big data," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2016, pp. 635–644.

[8] H. Wang, H. Yao, D. Kifer, C. Graif, and Z. Li, "Non-stationary model for crime rate inference using modern urban data," IEEE Trans. Big Data, vol. 5, no. 2, pp. 180–194, Jun. 2019.

[9] L. Vomfell, W. K. Härdle, and S. Lessmann, "Improving crime count forecasts using Twitter and taxi data," Decis. Support Syst., vol. 113,pp. 73–85, Sep. 2018.

[10] F. Yi, Z. Yu, F. Zhuang, and B. Guo, "Neural network based continuous conditional random field for fine-grained crime prediction," in Proc. 28th Int. Joint Conf. Artif. Intell., Aug. 2019, pp. 10–16.