International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03400
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

City Crime Hotspot Identification and Visualization Development of a Web Application for Improving Urban Safety and Decision Making.

Dr Kavitha Soppari[1], Pakkurthi Ravi Kiran[2] and Junuthula Shashidar[3].

[1]Head of Department of CSE (AI & ML), ACE Engineering College, Hyderabad, Telangana, India.

[2]Student of Department of CSE (AI & ML), ACE Engineering College, Hyderabad, Telangana, India.

[3]Student of Department of CSE (AI & ML), ACE Engineering College, Hyderabad, Telangana, India.

Abstract

Urban crime hotspots present significant challenges to public safety and effective city planning. This study proposes a web-based crime hotspot identification and visualization system that utilizes geospatial mapping, machine learning, and data analytics. The application empowers law enforcement agencies and policymakers by classifying high-risk areas, tracking evolving crime trends, and predicting future crime occurrences. The system's interactive dashboard delivers real-time, data-driven insights to support strategic decision-making and improve urban safety outcomes.

Keywords: Crime Mapping, Hotspot Identification, Geospatial Analysis, Data Visualization, Machine Learning, Predictive Analytics, Data Analysis, Crime Trend Analysis, Risk Classification.

1. Introduction

1.1 Background and Motivation

City life across the globe is constantly burdened by crime, which poses not only threats to public security but also retards proper urban planning and urban development. Whether or not the police can curb and prevent crime is essential for the welfare of citizens and for the overall urban development of a city. Conventional techniques of crime prevention and analysis tend to be restricted by the very magnitude and nature of crime data, rendering pattern detection, anticipation of future instances, and effective resource allocation a challenge. Improved access to big data and advancements in technology, including Geographical Information Systems (GIS), machine learning, and data analytics, have created new possibilities for analyzing and forecasting crimes. Through the application of these technologies, it is now feasible to create more advanced systems that are capable of examining past crime records, detecting hotspots or areas of high crime risk, monitoring changing crime patterns, and predicting future crime events. Yet, in spite of extensive research activities, there are still some challenges. These include the requirement for more precise and credible crime information, the complexity of processing Spatio-Temporal crime patterns, and the challenge of creating resilient predictive algorithms. Additionally, there is a requirement for systems that are able to effectively present and convey crime-related information to decision-makers in a clear and actionable format.

This research is driven by the necessity to resolve these challenges through the creation of a web-based crime hotspot identification and visualization system. Through the combination of geospatial mapping, machine learning, and data analytics, the suggested system will offer an effective means for enhancing urban safety, better decision-making, and ultimately more liveable and secure cities.

1.2 Introduction

Major cities around the world struggle with the ongoing problem of crime, which is a serious threat to public safety and interferes with efficient city planning and economic development. The efficiency of law enforcement organizations in controlling and preventing crime is important for guaranteeing the health of local citizens and promoting the healthy development of cities. Conventional approaches to crime prevention and analysis are usually unable to deal with the sheer amount and intricacy of crime data, making it difficult to identify patterns, predict crime events accurately, and allocate resources effectively. Nevertheless, the growing volume of large datasets, along with the evolution in technologies like Geographical Information Systems (GIS), machine learning, and data analytics, has opened up opportunities for bolstering crime analysis and prediction. Through these technologies, it is now feasible to create more evolved systems for scrutinizing past crime data, pinpointing high-risk locations or "hotspots", monitoring changing crime patterns, and predicting future crime occurrences. These systems have the potential to offer useful insights to policymakers and law enforcement, enabling them to make informed decisions based on data to enhance urban safety and maximize resource allocation. This research suggests a web-based crime hotspot identification and visualization system that can help solve these problems. Through the use of geospatial mapping, machine learning, and data analytics, the system seeks to create an effective aid for enhancing urban security, aiding strategic planning, and eventually helping in making cities safer and more habitable.

Important features of this study involve the application of machine learning algorithms for hotspot detection of crimes and the creation of an interactive web application to display crime information and patterns. The system's capacity to map high-risk locations, monitor changing crime trends, and forecast upcoming crime incidents provides important promise for maximizing the efficacy of crime prevention and intervention programs.

2. Literature Review

2.1. [1] G. Maji, S. Mandal, and S. Sen. (2023): Identification of city hotspots by analyzing telecom call detail records using complex network modelling.

Examining telecom big data (CDR) provides useful information about city dynamics because it is inexpensive, real-time, and has a clean structure. Yet, its size is so large that processing is difficult. To cope with this, we employed graph-based network analysis. We represented CDR data in Milan and Trento as weighted networks, where regions with greater communication were regarded as more central. To simplify, we removed weaker links, forming smaller, less complex networks. We used conventional centrality indices, such as weighted degree and weighted k-shell, and a new one we designed—the modified Edge Weight Degree Neighborhood (EwDN)—to find top hotspots. Through the examination of five days of data, we monitored hotspot evolution and cross-checked our findings against real-world data with QGIS. For lack of ground truth, we compared our findings against the SIR model as well using Kendall's rank correlation. Our results indicate that careful network modeling, intelligent data reduction, and sophisticated techniques such as EwDN facilitate rapid and accurate detection of significant city regions at low computational expense.

2.1.1. Methodologies and Algorithms:

The research applied a graph-based model to represent call detail records (CDR) in the form of weighted networks, with nodes indicating regions or time frames and edges representing the intensity of communication between them. The original networks were extremely dense, so one of the initial steps was to remove weaker ties through edge-weight thresholding, forming smaller, more lightweight networks. To identify significant locations (hotspots), standard centrality measures such as weighted degree and weighted k-shell were employed, and a new approach known as modified Edge Weight Degree Neighborhood (EwDN). The approaches were evaluated on five days of CDR data and verified against real-world data with QGIS. Since there was restricted

ground truth, the Susceptible-Infectious-Recovered (SIR) model was taken as a benchmark, and outcomes were compared based on Kendall's rank correlation to test for consistency.

2.2. [2] X. Ran et al. (2021): A Novel K-Means Clustering Algorithm with a Noise Algorithm for Capturing Urban Hotspots

Large cities all over the world experience traffic congestion as they expand, and clever planning of roads can alleviate this, but it's a challenging task for computers. Scientists have employed an algorithm known as K-means clustering to cluster similar traffic patterns, but it is not without issues: it's difficult to determine how many traffic clusters there are, and the initial positions of these clusters can ruin the results. To address this, we inserted a "noise" trick into K-means. This approach assists in automatically determining the optimal number of traffic groups and identifying improved starting points. We used typical checks to test our new method and even applied a statistical test to ensure our results were valid. We then tried out our enhanced K-means algorithm with actual taxi GPS traces for five cities: Aracaju, San Francisco, Rome, Chongqing, and Beijing. We also compared it against other clustering approaches, and results indicated that our noise trick resulted in better accuracy in detecting the correct number of traffic clusters, discovering suitable initial points, and correctly marking heavy "hotspot" regions of the cities.

2.2.1. Methodologies and Algorithms:

The authors approached the task of discovering urban hotspots from taxi GPS traces by designing a new K-means clustering algorithm based on a noise-based strategy. The noise technique itself identifies automatically the optimal number of clusters and selects the initial cluster centers, bypassing issues with typical K-means. To evaluate the performance of their algorithm, they applied four clustering quality metrics to assess the quality and separation of the clusters. They also used a nonparametric Wilcoxon test to verify the reliability of their findings by comparing various algorithms. Lastly, they experimentally confirmed the efficiency of this new approach using real-world taxi GPS traces from five cities and demonstrated that it was more efficient in detecting urban hotspots.

2.3. [3] L. Cai, H. Wang, C. Sha, F. Jiang, Y. Zhang, and W. Zhou. (2023): The mining of urban hotspots based on multi-source location data fusion.

Urban hotspots are crowded locations where many people congregate, and having a sense of where they are is useful for city planning and security. Most research relies on a single source of location data to identify these hotspots, but what if we could use multiple sources to get a better view? The catch is, different sources tend to have lopsided amounts of information, which can complicate the identification of hotspots. To fix this, we invented a new solution. We first bridged the gaps in the less detailed data so that everything is more balanced. Then, we designed a special technique to locate hotspots even if the data is unbalanced. We also invented a new method to determine whether our outcomes are valid. When we tested with real data such as points of interest, check-ins, and GPS tracks, the result demonstrated that our algorithm identified all major hotspots sooner and with more precision compared to other algorithms.

2.3.1. Methodologies and Algorithms:

To determine urban hotspots from skewed location data, the researchers built a two-stage framework. In the first step, they built a model to impute missing data from a single source so that the information is balanced across all sources. In the second step, they built a customized clustering algorithm that is effective with this unbalanced data so that hotspots are identified correctly even with missing data. To verify the quality of their findings, they added a new measure of evaluation to determine how accurately the hotspots were determined. This methodology, which incorporates data balancing, clustering, and assessment, is the basis of their study.

2.4 [4] J. Wu, E. Frias-Martinez, and V. Frias-Martinez. (2021): Spatial sensitivity analysis for urban hotspots using cell phone traces.

As cities expand, it is important to know about busy places, or urban hotspots, in order to understand city life. With technologies such as cell phones monitoring our locations (such as Call Detail Records), we have plenty of data to locate these hotspots. But the catch is, everyone has their own method of defining them! Important choices such as where the city limits are, how to split the city up into smaller pieces, how to deal with missing data, and what constitutes a 'hotspot' can all differ. Therefore, we used a big sample of cell phone data from Mexico to investigate how these decisions influence the hotspots we discover. We wanted to know whether hotspots stay fixed or shift according to these choices, both within a given moment and over several days. Our aim was to learn more about how these approaches define our perspective of a city's busiest places.

2.4.1. Methodologies and Algorithms:

This study examines the spatial sensitivity of city hotspots from mobile phone data. It quantifies how stable hotspot indices are if computed under varying definitions of city boundaries, spatial units, and unit memberships of people. The study establishes methodological combinations that yield consistent and very sensitive hotspot measures.

2.5 [5] XU ZHANG 1,2, LIN LIU2,3, LUZI XIAO2. (2020): Comparison of Machine Learning Algorithms for Predicting Crime Hotspots

The research compares machine learning algorithms for crime prediction based on past public property crime data of a Southeast China city (2015-2018). LSTM outperformed the others based on only historical data, and including built environment data (road density, POIs) also improved LSTM's accuracy, implying the significance of combining both types of data in future prediction.

2.5.1. Methodologies and Algorithms:

The research employs machine learning and time series analysis to analyze crime. Machine learning techniques such as KNN, SVM, Naïve Bayes, Random Forest, and deep models are employed for crime prediction. Time series analysis such as ARIMA and LSTM is employed for predicting trends in crime, which is frequently performed with spatio-temporal data for hotspot identification. A few studies also incorporate variables such as POIs and road networks to enhance predictions.

2.6. [6] FARHAN AMIN, (Graduate Student Member, IEEE), AND GYU SANG CHOI (2020): Hotspots Analysis Using Cyber-Physical-Social System for a Smart City

IoT creates a lot of data in intelligent cities. The paper discusses leveraging complex networks for predicting communication hotspots of traffic. A model of cyber-physical-social system (CPSS) builds a graph and performs social network analysis to extract and quantify hotspot importance for rapid detection and prioritization. The model is efficient and should be favourable to urban/telecom development.

2.6.1. Methodologies and Algorithms:

The research suggests a cyber-physical-social system (CPSS) model to examine high communication traffic hotspots based on telecom data. The model builds a graph and conducts social network analysis. The procedure includes hotspot extraction and then social network analysis, where the significance of each hotspot is measured in terms of network metrics. These metrics are used to determine the importance of hotspots in a telecom data network.

2.7. [7] UMAIR MUNEER BUTT, SUKUMAR LETCHMUNAN, FADRATUL HAFINAZ HASSAN, MUBASHIR ALI, ANEES BAQIR4 AND HAFIZ HUSNAIN RAZA SHERAZI (2020): Spatio-Temporal Crime Hotspot Detection and Prediction: A Systematic Literature Review

The paper is on recent methods for identifying and forecasting crime hotspots from Spatio-Temporal data. The authors perceived that there was no thorough review in this field, so they carried out one. The review examines the application of data mining and machine learning, the significance of time-oriented and location-oriented data, and the challenges involved. Finally, the review hopes to lay a groundwork for Spatio-Temporal crime prediction research in the future.

2.7.1. Methodologies and Algorithms:

It evaluates the application of data mining and machine learning methodologies, specifically clustering methods, for crime hotspot identification. It also investigates the application of time series analysis and deep learning methods crime trend prediction.

2.8. [8] WAJIHA SAFAT, SOHAILASGHAR, (Member, IEEE), AND SAIRA ANDLEEB GILLANI (2021): Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques

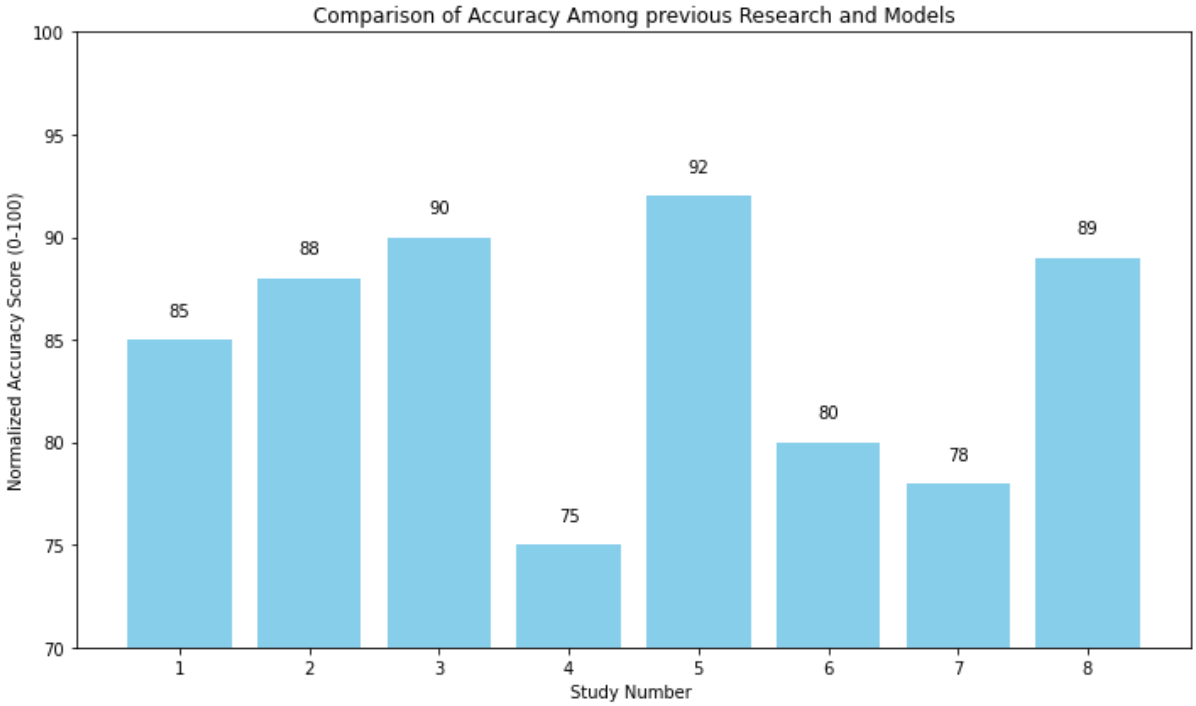
Precise crime forecasting with computational approaches is important to improve metropolitan safety since human interpretation of big data is limited. This research utilized different machine learning models (logistic regression, SVM, Naïve Bayes, KNN, decision tree, MLP, random forest, XGBoost) and time series techniques (LSTM, ARIMA) on crime data. LSTM demonstrated fairly good performance in time series analysis for both Chicago and Los Angeles datasets. Explanatory analysis demonstrated more than 35 crime types, an annual decrease in Chicago crime rate, a slight growth in Los Angeles, and less crime in February. ARIMA projected a steep fall in Los Angeles crime rates and that Chicago's rate will moderately rise before it might decrease. The research established future crime hotspots in the two cities, providing better predictive precision for informing police strategies.

2.8.1. Methodologies and Algorithms:

the research makes use of a variety of machine learning models such as logistic regression, support vector machine (SVM), Naïve Bayes, k-nearest neighbors (KNN), decision tree, multilayer perceptron (MLP), random forest, and extreme Gradient Boosting (XGBoost) for the prediction of crime. For time series forecasting, the research makes use of long-short term memory (LSTM) and autoregressive integrated moving average (ARIMA) models.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03400
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

2.9. Comparison of Accuracy of Existing Algorithms and Models



The above figure is a comparative figure illustrating the normalized accuracy scores of the reviewed studies and their corresponding models. The y-axis depicts the normalized accuracy score, measured from 70 to 100, giving an equal standard for comparison of the predictive capability of various methodologies. The x-axis counts the study number, referring to the order of works examined in this survey.

As presented in the graph, accuracy scores have a significant difference. Study 5 shows the best accuracy, with a score of 92, indicating a comparatively better predictive ability of its utilized methodologies. Conversely, Study 7 has the lowest accuracy of 78, implying possible shortcomings in its methodological process. The other studies offer intermediate accuracy measurements between 75 and 90, which describe the varied efficacy of different algorithms and models for urban hotspot prediction and detection.

This visualization efficiently encapsulates the comparative performance of various methodologies, providing a rapid glimpse into the state-of-the-art precision in the discussed literature. It emphasizes the predictability variability and the significance of methodology selection to attain reliable and strong results in urban hotspot studies. Analysis of these variabilities may proceed further to determine the most efficient methods and points for potential development improvement in the field.

2.10 Comparative Analysis

Table: Comparative analysis of existing research papers.

Name of the Paper	Year of Publication	Algorithms Used	Key Findings / Accuracy	Limitations
Identification of City Hotspots by Analyzing Telecom CDR Using Complex	2023	Graph Theory, Node Centrality, EwDN.	Real-time hotspot detection with reduced computational	Limited real-world ground truth data for verification.

Network Modeling.			load; validated using SIR model.	
A Novel K-Means Clustering Algorithm with a Noise Algorithm for Capturing Urban Hotspots.	2021	Modified K-Means with Noise.	Automatically finds cluster numbers and centers outperforms traditional K-means in hotspot detection.	Higher computational cost due to added complexity of noise algorithm.
The Mining of Urban Hotspots Based on Multi-Source Location Data Fusion.	2023	Data Imputation, Imbalance-Aware Clustering.	Accurately detected over 90% of real-world hotspots; fast and effective.	Sensitive to data balancing, lacks generalization beyond Kunming.
Spatial Sensitivity Analysis for Urban Hotspots Using Cell Phone Traces.	2021	Sensitivity Analysis on Spatial Units and Definitions.	Identifies how methodological choices affect hotspot stability and reliability.	Dataset limited to Mexico; not exhaustive in method combinations.
Comparison of Machine Learning Algorithms for Predicting Crime Hotspots.	2020	LSTM, SVM, KNN, RF, CNN.	LSTM achieved highest prediction accuracy; improved further with built environment data	Focused on one crime type in one city; lacks additional influencing factors
Hotspot Analysis Using Cyber-Physical-Social System for a Smart City.	2020	Graph Modeling, Social Network Analysis.	Efficiently prioritizes hotspots using telecom traffic data; beneficial for urban planning.	Requires high computational and storage capacity.
Spatio-Temporal Crime Hotspot Detection and Prediction: A Systematic Literature Review.	2020	Clustering, ML, DL, Time Series	Emphasizes importance of spatial-temporal data; surveys wide methods.	No universal benchmark; lacks standard evaluation framework.
Empirical Analysis for Crime Prediction	2021	Logistic Regression,	LSTM and ARIMA models effectively predict crime types and	Big data complexity poses processing challenges;

Using ML and DL Techniques.		SVM, KNN, RF, LSTM, ARIMA.	trends; analysis of >35 crime types.	requires stronger real-time guidance tools.

2.11. Research Gaps

Existing crime hotspot detection and prediction frameworks are faced with a number of key research loopholes. To begin with, the majority of models are based on mono-source datasets, which hinders them from obtaining an exhaustive overview of criminal activities. Secondly, there is limited real-time analysis owing to the computational intensity involved in handling spatio-temporal streams of data. There is also a significant absence of generalizability across urban areas, with models usually being developed for a specific city or dataset, such that they perform poorly in new settings. Ethical and privacy implications are posed by the employment of sensitive information like telecom records and GPS traces, with minimal application of privacy-preserving methods like federated learning. Additionally, the lack of standardized benchmarks and evaluation schemes hinders cross-study comparison of models. Finally, few are validated or tested in conjunction with urban planners and police forces, raising the need for real-world deployment and assessment.

Abbreviations

- GIS – Geographical Information System
- CDR – Call Detail Record
- ML – Machine Learning
- DL – Deep Learning
- SVM – Support Vector Machine
- LSTM – Long Short-Term Memory
- ARIMA – Autoregressive Integrated Moving Average
- CPSS – Cyber-Physical-Social System
- POI – Point of Interest
- RF – Random Forest
- KNN – K-Nearest Neighbors
- CNN – Convolutional Neural Network
- NLP – Natural Language Processing

3. Conclusion

The literature reviewed provides a rich selection of methods for identifying urban crime hotspots, ranging from traditional clustering techniques to cutting-edge predictive analytics powered by AI. While the techniques are useful in providing insights, they are also subject to typical limitations, which include dependence on single-source data, ethical issues surrounding the use of data, and difficulties in scaling to real-time. To overcome such constraints, one needs systems that blend multiple data streams, support privacy-preserving measures, and are tested in real-world environments with the involvement of urban planners and law enforcement agencies.

The creation of these systems would support more accurate and detailed crime hotspot identification and forecasting, ultimately resulting in better urban safety and more informed decision-making. Our system seeks to overcome the challenges posed by these issues by combining several streams of data into a web-based system. This system will yield real-time tracking, interactive dashboards, and predictive features to enable improved urban safety efforts.

References

[1] G. Maji, S. Mandal, and S. Sen, "Identification of city hotspots by analyzing telecom call detail records using complex network modeling," *Expert Systems with Applications*, vol. 215, 2023.

[2] X. Ran, Y. Liu, H. Zhang, and J. Wang, "A Novel K-Means Clustering Algorithm with a Noise Algorithm for Capturing Urban Hotspots," *Applied Sciences*, vol. 11, no. 23, 2021.

[3] L. Cai, H. Wang, C. Sha, F. Jiang, Y. Zhang, and W. Zhou, "The mining of urban hotspots based on multi-source location data fusion," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 2, 2023.

[4] J. Wu, E. Frias-Martinez, and V. Frias-Martinez, "Spatial sensitivity analysis for urban hotspots using cell phone traces," *Environment and Planning B: Urban Analytics and City Science*, vol. 48, no. 9, pp. 2499–2514, 2021.

[5] X. Zhang, L. Liu, and L. Xiao, "Comparison of machine learning algorithms for predicting crime hotspots," *IEEE Access*, vol. 8, 2020.

[6] F. Amin and G. S. Choi, "Hotspots analysis using cyber-physical-social system for a smart city," *IEEE Access*, vol. 8, 2020.

[7] U. M. Butt, S. Letchmunan, F. H. Hassan, M. Ali, A. Baqir, and H. H. R. Sherazi, "Spatio-Temporal crime hotspot detection and prediction: A systematic literature review," *IEEE Access*, vol. 8, 2020.

[8] W. Safat, S. Asghar, and S. A. Gillani, "Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques," *IEEE Access*, vol. 9, 2021.