Applied Spatial Analysis and Policy (2024) 17:963–982
https://doi.org/10.1007/s12061-024-09572-z

An Application of Scan Statistics in Identification and Analysis of Hotspot of Crime against Women in Rajasthan, India

Poonam K. Saravag [1] · Rushi Kumar B. [1]

Received: 20 September 2022 / Accepted: 22 February 2024 / Published online: 14 March 2024

Abstract
Crime against women (CAW) is not a present-day problem but has been prevalent in the world through the ages and since the beginning of civilizations. The cases of CAW have been increasing in almost all parts of the world and India is no exception. The distribution of CAW cases has not been found uniform across the country. The evidence of heterogeneity of cases has been a concern. Rajasthan, the largest state in India, has witnessed a very high surge in CAW in recent years. Therefore, there arises a need to study and analyze the pattern of CAW to identify the areas with high intensity for prevention and control. The CAW data from the National Crime Records Bureau (NCRB) website for the period 2014 to 2021 and the population census data of 2011 are used for the analysis. The Statistical analysis software, SaTScan, is employed for hotspot (areas with a high concentration of crimes) detection. Python programming is used to compute the data's trend or pattern through visualization and descriptive statistics. In addition, the simple exponential smoothing method is applied for predicting the CAW for the year 2021. Our work elucidates Jhalawar, Baran, Kota, Bundi, Sawai Madhopur, and Chittorgarh districts as consistently occurring hotspots of CAW in the state. A comparative study of the hotspots found is made with the result obtained from the descriptive analysis. The trend in the data explains the years 2017 and 2019 as trough and crest of CAW cases. The hotspot detected using the forecast value of 2021 appears to be the same districts as for the period 2014 to 2020. Our work concludes that the consistency and the most likely cluster of CAW are distributed distinctly. We also found that the hotspot of CAW is not by chance but has certain man-made reasons. Most of the clusters have been identified as districts sharing boundaries with adjacent states. This further implies that if sincere efforts to collaborate with the government of the adjacent states like Madhya Pradesh, Haryana, Punjab, Uttar Pradesh, and Gujarat, the incidences resulting in detrimental

✉ Rushi Kumar B.
rushikumar@vit.ac.in

[1] Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India

effects due to CAW could reduce effectively and significantly. Thus, our study may help the government, law enforcement agencies, police organizations, judiciaries, and other stakeholders to optimize their scarce resources most effectively to curb such incidents.

Keywords Crime · Scan statistics · SaTScan · Hotspot · Time series

Abbreviations

Abbreviations	Description
CAW	Crime Against Women
NCRB	National Crime Records Bureau
SCRB	State Crime Records Bureau
DCRB	District Crime Records Bureau
IPC	Indian Penal Code
SLL	Special and Local Laws
MLC	Most Likely cluster

Introduction

Crime Against Women

Crime is a detrimental activity that desecrates someone physically, and mentally, mostly related to the assets and respect for people (Morrison, 2006). The meaning of crime against women can be defined as physical, sexual, or psychological cruelty of any form (Dayal & Singh, 2014; Vicente et al., 2019). Though crime against women has been prevalent in our society since ancient times, there is a considerable rise over the past few years. Society is made up of men and women equally (Khadke et al., 2019), which means both should enjoy equal rights in terms of all the factors affecting the growth of an individual as well as the society. No society can witness social, economic, political, and cultural progress without the participation of both genders. The gender system be it in India or the world has always been biased against women compelling them to always live in an inferior environment as compared to the male gender (Patel, 2015; Sharma et al., 2020). In India, a woman is glorified and at the same is vulnerable to various adversities of society. Male superiority existing in a patriarchal society does not allow women to grow and flourish. Though ancient Indian culture witnessed women exercising their right to education, administrative decisions, widow remarriages, etc., the degradation of women started in medieval times. Sati pratha, parda pratha, child marriages, restrictions on widow remarriage, widow exploitation, the devadasi system, etc., were some of the practices faced by women in medieval times. Women have been objectified and treated unequally to the opposite gender, which leads to violence against them (Lolayekar et al., 2020; Oliveira et al., 2019).

With time reforms have been made to abolish such systems and let women be free from exploitation. Though many of these problems have been abolished, still there are many existing and upcoming issues regarding the problems faced by women. According to the 'Women, Peace and Security Index 2021', India occupies 148th position among 170 countries in the world (WPS, 2021). A report by World Health Organization states that one-third of women worldwide have experienced physical or sexual violence in their lifetime. Also, every day around 137 women are losing their lives at the hands of their partner or a family member, or someone known.

According to the data released by NCRB for the year 2021, Rajasthan ranked second in CAW after Uttar Pradesh and placed third position in CAW rate in India. Rajasthan reported 6337 rape cases which is the highest number in the country in 2021 (NCRB, 2021). The news article published in 'The Statesman' highlights that out of the 6337 rape cases, in only 4% cases the rapist was not known to the victim whereas the remaining 96% cases were registered against the person who was directly or indirectly known to the victim. This is the biggest challenge to society. The rise in incidents of crimes against women can be attributed to the state's low rate of convictions and insufficient police personnel per hundred thousand people (NCRB, 2021). The low conviction rate provides the offenders a hope to get away easily after committing the crimes. The poor conviction rate reflects the inadequacy of government and corruption prevailing in the entire system. This leads to an increase in CAW in the state as well as the country. Furthermore, the distribution of CAW is not uniform across the nation. It relies on biological, psychological, societal, and financial elements. With the persistent rise in crime rates, there emerges a necessity to identify areas where the likelihood or frequency of criminal incidents surpasses that of other regions. Hence, there is a requirement to detect and examine areas that are at a higher risk of experiencing such events. To address this issue, a hotspot analysis has been conducted. Hotspot analysis is a geographical approach used to identify areas with higher concentrations (Lee & Eck, 2023). It involves analyzing crime data to pinpoint specific geographic locations where incidents of crime against women are more frequent than expected based on random distribution. This information can be extremely useful for law enforcement agencies, policymakers, and social organizations to allocate resources, implement targeted interventions, and develop strategies to address and prevent crime against women effectively.

Literature Review

Almanie et al. (2015) applied the apriori algorithm to determine the frequent crime patterns in Denver & Los Angeles cities and applied Decision Tree and Naïve Bayesian Classifier for predicting future crimes in a particular location during a particular time. Hackett (2011) used the multivariate linear regression method to predict a relationship between domestic violence and factors affecting it in India with the conclusion that domestic violence in India depends mainly on the factors like human development, gender development, and urbanization. The higher levels of overall development in the state lead to lowering domestic violence against women. Sukhija et al. (2017) discussed a spatial visualization of total cognizable crimes during the period 2010 to 2014. They presented the crime hotspot using scan statistics in the state of Haryana and found that Gurugram appeared consistently as the primary hotspot whereas the secondary hotspots keep on changing between Rewari, Faridabad, Ambala, and Hissar. Kuralarasan and Bernasco (2022) used chain snatching data from police records to identify the locations where the crime occurrence is highest in Chennai city. They employed ArcGIS and discrete crime location choice models to find the locations and concluded that marriage halls and religious places are the most favorable locations for offenders.

Maciejewski et al. (2011) employed varieties of tools for the analysis of spatio-temporal data and alert detection algorithms. The importance of visual analysis is emphasized for a better understanding of hotspots. Townsley (2008) proposed an approach 'Hotspot plot' for visualization of crime patterns in both space and time. Further, discussed that the temporal distribution of hotspots shows considerable variation daily as well as yearly. Anjali et al. (2021) identified suicide hotspots in Tamil Nadu for the data during the time span of 2011 to 2019. They employed a multiple linear regression model to analyze the relationship between the suicide cases as dependent variables and the factors like socio-economic, personal or professional, health-related, etc. as independent variables. Their work found Coimbatore, Madurai, and Tiruchirappalli as hotspot locations. Bhardwaj et al. (2019) surveyed different articles on crime detection and prediction techniques. They concluded that data pre-processing plays a vital role in predicting crimes using historical data. Also showed that the data mining algorithms proved better for supervised learning, and deep learning techniques for multimodal, huge, and unsupervised data.

Al-Sabbagh et al. (2023) utilized Harass-Map to study the incidents of sexual harassment crime in Egypt. The primary objectives included the application of the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm to identify the agglomeration of reported SHCs and evaluate their spatial relation with various land use types. The outcomes of the HDBSCAN algorithm indicated the presence of four crime clusters in the research area, primarily situated in Greater Cairo, Alexandria, and Behaira. Tayal et al. (2014) proposed an integrated technology of crime detection and criminal identification (CDCI) using data mining techniques (DMT) for the seven Indian cities to highlight crime and identify criminals. CDCI extracted unstructured data, pre-process it and applied k-means clustering techniques to the data during 2000-2012 through four cases. The k-nearest neighbour classification technique is applied for criminal identification and prediction. Lastly, WEKA (Waikato Environment for Knowledge Analysis) is used for the verification of k-means clustering. Ahmad et al. (2017) analyzed the temporal crime data during the period 2001-2015 in India for highlighting the pattern of general crime. For generating a crime density map and percent change used crime data from the year 2011 to 2015 for ten major cities. Crime data of the year 2015 is analyzed to understand the crime pattern and concluded that Delhi was found to have the highest crime density whereas Nagaland recorded the lowest, Indore topped in city crime whereas Chennai witnessed the minimum, Patna registered the maximum murder cases, Daman and Diu had the largest increase in crime followed by Delhi and Kerala. Hipp and Kim (2017) studied the variability in the level of crime concentration across the forty-two cities in Southern California from the years 2005-2012. They demonstrated that the traditional measure of crime concentration located spatially is due to random chance and employed different temporal assumptions, a historically adjusted crime concentration, and a temporally adjusted crime concentration measure. They concluded that there exists variability in the level of crime concentration. Runadi et al. (2017) detected hotspots of crime in Indonesia using spatial scan statistics which showed that most of the crimes occurred in the western part of Indonesia. Hotspots were obtained based on different crimes separately. Jambi, South Sumatra, and Riau Island appeared as the hotspots of crimes.

Mondal et al. (2022) detected hotspots in Pune city using three different crime hotspot detection methods, namely Kernel Density Estimation (KDE), Getis-OrdGi*, and SaTScan. KDE and Getis-OrdGi* are inbuilt into ArcGIS which is a very costly software whereas SPTM uses an open-source software SaTScan. They concluded that SPTM provides results similar to the two GIS-based (Geographic Information System) methods in identifying the hotspots of crime regions in Pune and is also cost-effective. Bowers et al. (2004) explored mapping strategies to gain prospective hotspots like the developing search efficiency rates and area-to-perimeter ratios. Also, standardized measures were used for the maps obtained by various methodologies so that significant comparisons can be made. Groff et al. (2010) examined both spatial and temporal variations in crime in Seattle city of Washington. They analyzed the block level variability by applying trajectory analysis for a time period of sixteen years. Further used quantitative spatial statistics to conclude that individual street segments trajectories were unrelated to their adjacent streets. Garfias et al. (2020) generated heat maps on household survey data to obtain locations with high intensity and a clear visual or trend analysis is performed using the factors affecting violence against women.

Seid et al. (2020) analyzed 2016 data from the Ethiopian health survey demographic to detect the spatial distribution of violence against women in Ethiopia using SaTScan. This gives the primary hotspot in Oromia and SNNP (Southern Nations, Nationalities, and People's Region) regions and the secondary in the Amhara region. Further, the factors of low economic status, and spouse alcohol consumption witnessed violence has been shown using a regression model. Gorr and Lee (2014) used kernel density smoothing to obtain chronic hotspots and simple rules for temporary hotspots for the dataset of the years 2000 to 2010. Further, considered a combination program for deploying police at locations with the hottest chronic hotspots. They concluded that temporary hotspots occurred outside chronic hotspots and recommended the use of the combination program of both these hotspots for better prevention. Lama and Singh (2018) applied GIS techniques to determine the hotspots of property crimes such as housebreaking, auto theft, and chain snatching, in Jodhpur city of Rajasthan. They found that house-breaking hotspots were obtained in areas with a huge difference in social profile, like affluent colonies, slum areas, densely populated areas, factories, etc. and auto theft hotspots emerged in public places. There were no particular hotspots of chain snatching.

After exploring the review of existing literature, we found that hotspots have been identified in different fields like criminal activity, suicides, robberies, etc. However, the geographical distribution of Crime Against Women (CAW), particularly within the region of Rajasthan, remains uninvestigated. The main objectives of the present study are:

- To summarize the CAW cases using descriptive statistics
- To examine the pattern analysis of CAW in the districts of Rajasthan
- To detect the CAW hotspots throughout the state by utilizing SaTScan software
- To forecast the crime incidences and hotspot analysis of the predicted values

Study Area

Rajasthan (Fig. 1) is a state in northern India. The state was formed on 30 March 1949. Rajasthan, 'the Land of Kingdoms', area-wise India's largest state, is carved up into 33 districts which are grouped into seven divisions namely Jaipur, Jodhpur, Ajmer, Bikaner, Kota, Bharatpur, and Udaipur. In the north, it shares a boundary with Punjab and Haryana, Uttar Pradesh and Madhya Pradesh lie to the east and southeast, and Gujarat in the southwest. It shares an international border with Pakistan in the west. The state is socially and culturally diverse. The vast area has different topographic regions including the great Thar desert in the west to the S-shaped Aravalli Mountain ranges towards the eastern part, fertile plains to rocky undulating land (Debnath & Ray, 2019).

According to the population census of the year 2011, it is the 8th most populated state in India with a 928-sex ratio and 66.11 percent literacy rate. The male and female

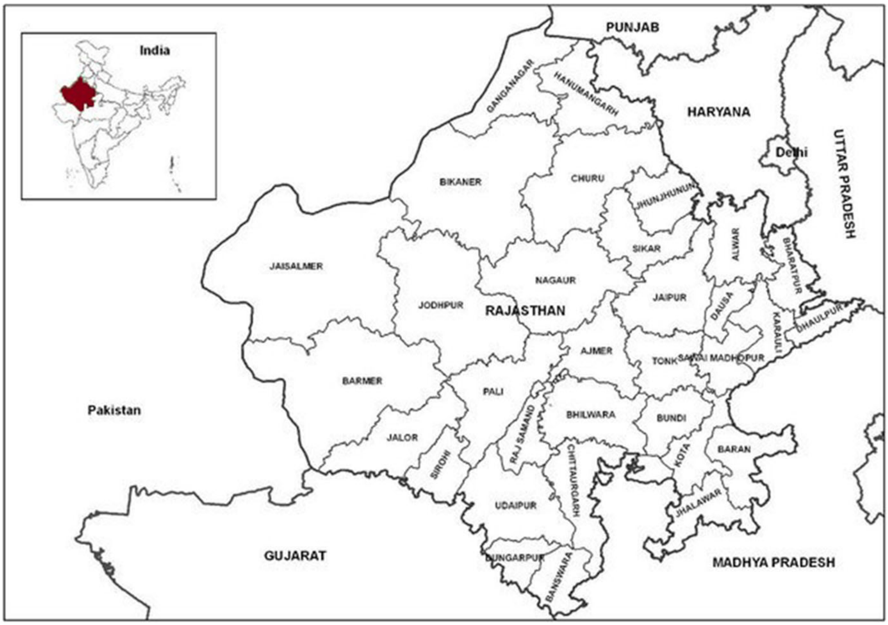

Fig. 1 Rajasthan State Map

literacy rates are 79.19 and 52.12 percent respectively recording the lowest female literacy rate as per the census of 2011. The literacy rate shows a wide gap between the two genders. According to the National Crime Records Bureau data, Rajasthan has always been in the top ten states for the last two decades. It occupied the sixth rank in 2000 and 2005 with 12942 and 11657 cases of CAW respectively. In 2010 and 2015, it ranked fourth with 18182 and 28165 cases respectively and there was a high jump in the year 2019 as it recorded the second highest number of CAW cases, 41550, after Uttar Pradesh. In 2020 the rank dropped from second to third with 34535 cases. The data for the year 2021 released by NCRB on 31st August 2022 placed Rajasthan in 2nd position in CAW after Uttar Pradesh.

Methodology

Data Sources and Processing

Since crime is a state subject and highly sensitive in nature, it is advisable to rely upon government sources for analysis and research. Unauthentic data sources can mislead by providing fabricated information and can be challenged by anyone, so data has been gathered from government portals for reliability. As such, we have retrieved needed data from the government portals and structured reports. The raw data of CAW was obtained from the National Crime Records Bureau (NCRB) website and the 2011 population census from government sources. The CAW case data files were obtained for a time period of seven years from 2014 to 2020. The data file contains the CAW cases of all the thirty-three districts of Rajasthan under the Indian Penal Code (IPC) and the Special and Local Laws (SLL) crime cases. Our study concentrated on IPC excluding the SLL cases.

Figure 2 shows the hierarchy of data percolation methods from the district level to the national level.

Various trends in the data were highlighted using Python programming. Graphs were obtained to provide a visual representation of CAW cases in all districts for better insight compared to the numeric data set. To calculate, summarize and describe, descriptive statistical analysis was performed to obtain logical and meaningful infor-

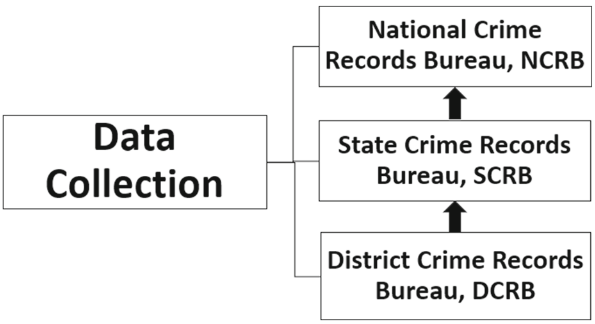

Fig. 2 Methodology of Data Collection

mation from the data set. The measure of location, dispersion, shape, and association was attained by descriptive statistics.

Hotspot Analysis

Crime, as an event, is Spatio-temporal in nature, it depends on both location and time. Scan statistic is an important tool used for the detection and evaluation of the statistical significance of clusters formed at a certain location (Jung, 2019). Detection of regions with clusters is traditionally performed by maximizing a likelihood ratio (Abolhassani & Prates, 2021). SaTScan software was employed to detect the criminal hotspots of the CAW in Rajasthan. SaTScan uses scan statistics to analyze purely spatial, purely temporal, and space-time datasets (Kulldorff, 1997) It can execute various tasks including space and time parameters. It helps in geographical surveillance, locating spatial or space-time clusters, examining the randomness of the hotspot distribution over space, time, or both, evaluating the significance of cluster alarms, and analysis of potential real-time or time-periodic surveillance for early-stage detection. SaTScan can be applied to both discrete and continuous scan statistics. In discrete scan statistics, the user fixes the non-random geographical locations where data is observed. These geographical locations represent larger areas such as the population-weighted centroids of areas like cities, districts, states, etc., and also the location can be specific such as houses, schools, offices, and other premises of social and economic importance. In continuous scan statistics, the observed locations are arbitrary and can be found anywhere in the user-defined study area, such as a rectangle (Kulldorff, 2021). We applied purely spatial scan statistics to identify the hotspot district of CAW in Rajasthan. The present work used the discrete Poisson model to obtain and analyze the hotspots of CAW. For every location and size of the scanning window, we frame the suitable null hypothesis H_0 as there is no significant risk inside the window and alternate hypothesis H_a can be defined as there exists a significant risk inside the window when compared with the outer region. For the assumption under the Poisson model, the likelihood function for a specific window is proportional to the equation Eq. 1

$$\left(\frac{w}{E[w]}\right)^{w}\left(\frac{W-w}{W-E[w]}\right)^{W-w} I(). \tag{1}$$

where the total number of crime instances is W, the observed number of crime instances inside the window is w and the covariate-adjusted expected number of instances inside the window under the null hypothesis is $E[w]$. The factor $W - E[w]$ is the anticipated number of instances in the outer of the window as the constraint for the analysis is the total number of cases observed. $I()$ in Eq. 1 refers to indicator function. It takes value 1 when the recorded instances from the window exceed as expected under the null hypothesis, and 0 otherwise. Equation 2 gives the Relative Risk (RR) which is the ratio of the estimated risk in the interior of the cluster to the estimated risk in the exterior of the cluster.

$$RR = \frac{w/E[w]}{(W-w)/(E[W]-E[w])} = \frac{w/E[w]}{(W-w)/(W-E[w]}. \tag{2}$$

The number of instances observed inside the cluster and the total number of instances are w and W respectively. Since the analysis is constrained to the total number of observed instances W, which is a constant, therefore $E[W] = W$.

Model Validation

The validation of hotspots generated using SaTScan was performed by hypothesis testing.
Step 1: The null (H_0) and alternate hypothesis (H_a) can be defined as H_0: There is no significant difference in CAW in hotspot and coldspot districts H_a: There is a significant difference between the two groups
Step 2: The level of significance α_1=0.05
Step 3: t-statistic
The $t_{calculated}$ is obtained from Eq. 3 and the pooled variance is computed using Eq. 4.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{{s_p}^2(\frac{1}{m_1} + \frac{1}{m_2})}}. \tag{3}$$

where $\bar{x}_1$ and $\bar{x}_2$ are the arithmetic means of CAW for Hotspot and Coldspot districts respectively, m_1 and m_2 are the sample size, ${s_p}^2$ is the pooled sample variance.

$${s_p}^2 = \frac{(m_1 - 1){s_1}^2 + (m_2 - 1){s_2}^2}{(m_1 + m_2 - 2)}. \tag{4}$$

Step 4: If $t_{calculated} > t_{critical}$, we reject null hypothesis
Step 5: Implications
The rejection of the null hypothesis indicates that sufficient evidence is present in the sample observations against the null hypothesis. Further, it implies that there is a significant difference between hotspot and coldspot districts concerning CAW cases. This leads to the conclusion that the hotspot formation is not due to chance rather than it owes to certain man-made specific reasons which need to be controlled, monitored, and surveilled for reducing such criminal activity in due course of time.

Time Series Forecasting

We used the simple exponential smoothing (SES) model for further fitting and prediction. It is also known as the exponentially weighted moving average method (Ostertagova & Ostertag, 2011). A simple exponential smoothing model is used to forecast future values when the data do not follow any trend as well as it is not cyclic. This method of forecasting uses a weighted average of past observations to predict future value (Shastri et al., 2015).

$$\tau(t+1) = \alpha\tau(t) + (1-\alpha)\tau(t). \tag{5}$$

where $\tau(t+1)$ is the forecast value for $(t+1)th$ point of time, α is the smoothing constant, $\tau(t)$ is the actual observation of the series. The value of α lies between 0 and 1 ($0 < \alpha < 1$). The error can be minimized by adjusting the value of smoothing constant α. As we go on increasing the value of α from 0 to 1, the smoother the series appears,which means the forecast values tend to get closer and closer to the original observations with a higher value of smoothing constant. When α is closer to unity, the latest forecast value will substantially adjust for any error occurring in the previous forecast value. When α is closer to 0, the forecast value obtained is almost similar to the previous forecast (Ostertagova & Ostertag, 2012).

Results and Discussion

Trend and Descriptive Analysis

Table 1 is prepared to describe and summarize the CAW cases which enabled us to get an insight into the data to highlight the arithmetic mean, standard deviation, range, and quartiles. The data summary indicates a substantial disparity between the minimum and maximum case numbers throughout the study period. Consequently, the high standard deviation suggests a significant dispersion of the data, highlighting the non-uniform distribution of crime against women (CAW) cases across the state. Specifically, the years 2017 and 2019 recorded the lowest and highest mean crime against women cases, respectively.

The top five most consistent (Table 2) and bottom five least consistent (Table 3) districts in CAW cases were obtained with the help of coefficient of variation (CV). Chittorgarh, Baran, Bhilwara, Kota, and Ganganagar have been found the top five districts which proved to be most consistent in CAW cases during the period of study from 2014 to 2020.

This implies further that among these five districts of Rajasthan, the criminals are deeply rooted with a more systematic arrangement of crime infrastructure and network system than in other districts of the state. Though crime occurs in other districts too it is not so consistent to be at an alert level. The consistency of crime in these districts indicates that they are at high risk in the future too. Thus, it is of immediate concern and needs to be addressed as quickly as possible. In contrast to the above districts, Dausa, Banswara, Barmer, Jalore, and Jaisalmer, the least five consistent districts,

Table 1 Descriptive Statistics

	2014	2015	2016	2017	2018	2019	2020
Count	33	33	33	33	33	33	33
Mean	940.1818	849.9091	827.5152	784.5758	842.0909	1257.545	1044.364
S.D	543.049	496.862	465.2442	439.4528	478.8	750.9179	567.8575
Min	167	118	146	189	177	301	247
25 percent	599	518	499	458	527	746	652
50 percent	806	737	731	713	714	1137	955
75 percent	1183	1187	1089	1042	1149	1537	1244
Max	3033	2747	2599	2306	2579	4417	3101

Mean-Arithmetic mean, S.D-Standard Deviation, Min-Minimum, Max-Maximum

Table 2 Top five most consistent districts of CAW

District	Mean	S.D	C.V	Rank
Chittorgarh	959.43	118.01	12.30	1
Baran	880.14	121.51	13.80	2
Bhilwara	1315.28	185.64	14.11	3
Kota	1331.43	195.96	14.718	4
Ganganagar	1316.57	200.93	15.26	5

Mean-Arithmetic mean of CAW cases from 2014 to 2020, S.D-Standard Deviation, C.V-Coefficient of Variation

predominantly lie in the tribal or less populated belt have witnessed the CAW cases as volatile under large variability. This volatility indicates that though crimes exist, there have been no specific roots and it seems to be erratic in nature which can be easy to control and monitor. It has been observed that the districts having higher population densities are more consistent in CAW as compared to the districts with lower population densities.

The trend in the CAW cases highlighted that Rajasthan recorded the minimum number of cases in the year 2017 and the maximum number of cases in the year 2019 shown in Fig. 3. Therefore, the years 2017 and 2019 can be considered the trough and crest of the CAW in the state of Rajasthan. Figure 3 also shows that there had been a constant decline in CAW from the year 2014 to 2017 after which a sudden jump was observed in 2019. Again, a dip can be noticed in the year 2020. Therefore, there is a need for in-depth analysis especially CAW in the year 2019. The dataset spanning from 2014 to 2020 is used to create the boxplot for the state (Fig. 4). This boxplot showcases summary statistics and reveals anomalies within the dataset of crime incidences. $(Q_1 - 1.5IQR)$ to $(Q_3 + 1.5IQR)$ is the range to demarcate the outliers, where Q_1 is the quartile 1, Q_3 is the quartile 3 and IQR is the interquartile range. CAW incidences in all the districts appear to be inliers except the Jaipur district. It lies outside the range and hence appears as an outlier. Additionally, there is evident skewness in the data, with a rightward skew observed consistently across all years.

Hotspots

SaTScan used to detect hotspot districts found that Jhalawar, Baran, Kota, Bundi, Sawai Madhopur, and Chittorgarh emerged as the first most likely cluster (MLC),

Table 3 Bottom five least consistent districts of CAW

District	Mean	S.D	C.V	Rank
Jaisalmer	192.14	62.39	32.47	1
Jalore	432.43	132.91	30.74	2
Barmer	827.86	248.41	30.00	3
Banswara	580.43	172.38	29.70	4
Dausa	461.43	123.08	26.67	5

Mean-Arithmetic mean of CAW cases from 2014 to 2020, S.D-Standard Deviation, C.V-Coefficient of Variation

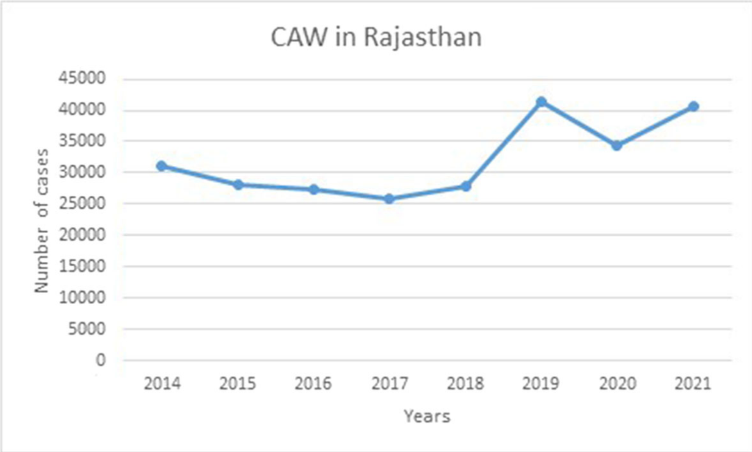

Fig. 3 Trend analysis of CAW

presented in Table 4, and Pratapgarh as the second most likely cluster (Fig. 5). We used the mean of the rate of CAW cases for the year 2014 to 2020 as the case file. The population file used was the total population of each district obtained from the population census of 2011. The coordinate file containing the latitude and longitude information for all the districts was utilized during the analysis of hotspots. Purely spatial analysis with the discrete Poisson model was deployed to get the locations with a high concentration of crime rate. The percentage of the population risk was adjusted to 20.0 percent from the default value of 50.0 percent to highlight the cluster formation properly. Google maps HTML files as inbuilt-functions were utilized to obtain the visualization of hotspots identified and detected. The minimum number of cases restricted to two was imposed as an initial condition to avoid degeneracy for the

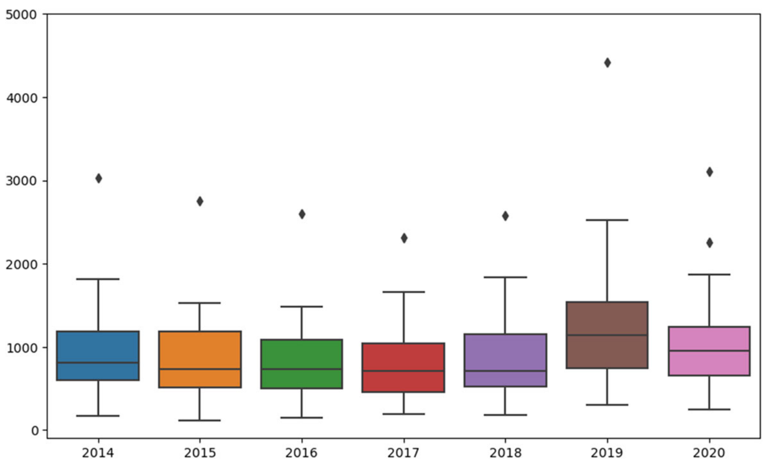

Fig. 4 Boxplot of CAW cases from 2014 to 2020

Table 4 Primary and Secondary Hotspots districts

	Hotspot	
Year	Primary Hotspot	Secondary Hotspot
2014	Pratapgarh	Jhalawar
		Baran
		Kota
		Bundi
		Sawai Madhopur
		Chittorgarg
	Jhalawar	Pratapgarh
2015	Baran	
to	Kota	
2020	Bundi	
	Sawai Madhopur	
	Chittorgarg	

Primary and Secondary Hotspot Districts from 2014 to 2020

generation of high-rate clusters. Other adjustments were used as default settings. The statistic of SaTScan analysis for the hotspots is provided in Table 5.

The primary hotspots were mainly found on the border regions adjacent to the Madhya Pradesh state. The primary hotspot districts are sharing either one or three side boundaries with Madhya Pradesh (Fig. 6). The secondary hotspot highlighted the same districts Baran, Kota, Jhalawar, Bundi, and Sawai Madhopur whereas the Pratapgarh district also sharing a border with Madhya Pradesh state emerged as a tertiary hotspot. Because crime is a state subject, criminal in the border areas has a

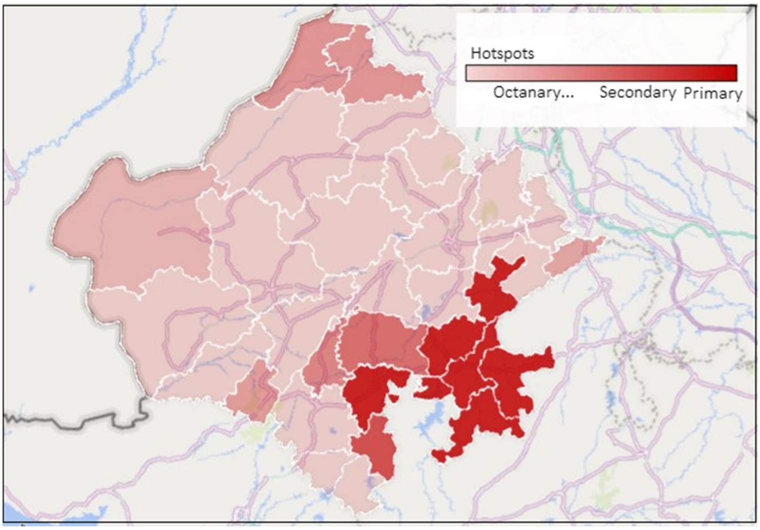

Fig. 5 Choropleth map of CAW Hotspot Districts of Rajasthan

Table 5 Hotspot Analysis

Parameters	Primary Hotspot districts Jhalawar, Baran, Kota, Bundi, Sawai Madhopur, Chittorgarh
Gini Cluster	No
Population	8575693
Number of Cases	270
Expected Cases	134.49
Annual Cases/10000	3.1
Observed/Expected	2.01
RR	2.35
LLR	62.933284
p-value	<0.00000000000000001

LLR-Log Likelihood Ratio, RR-Relative Risk

high probability to escape from the crime spot to shelter in another state which can avoid immediate arrest and uses this time to influence the corrupt system to get away from punishment.

A comparative study of hotspots and the districts that emerged to be consistent in CAW during descriptive statistical analysis was carried out. Baran, Kota, and Chittorgarh which emerged as hotspots were also the 1st, 2nd, and 4th top most consistent districts. This shows that the districts Baran, Kota, and Chittorgarh followed by Jhalawar, Bundi, and Sawai Madhopur are of greater concern for the purpose of highlighting the issues of CAW at an alarming stage during the period under study (2014-2020).

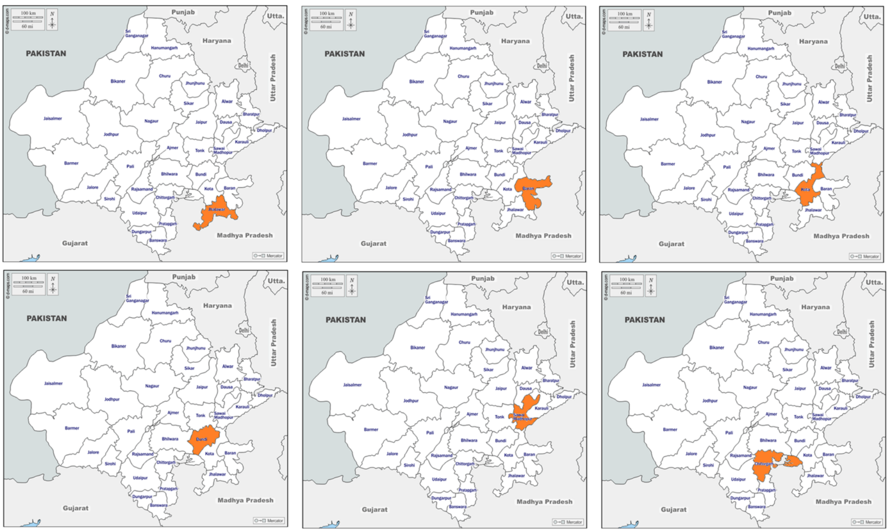

Fig. 6 Primary Hotspot Districts of CAW in Rajasthan from the year 2015 to 2020

Table 6 t-test statistics

	Hotspot	Coldspot
Mean	40.35714286	26.84210526
Variance	103.0164835	44.91812865
Observations	14	19
Pooled Variance	69.28195489	
Hypothesized Mean Difference	0	
df	31	
t Stat	4.609895066	
P(T<=t) one-tail	0.0000327478190464845	
t Critical one-tail	1.69551878254587	
P(T<=t) two-tail	0.0000654956380929689	
t Critical two-tail	2.03951344639641	

df - degrees of freedom, t Stat - t-test statistics

Hotspot Validation

The validity of the hotspots obtained needs to be checked through the procedure of hypothesis testing. The hypothesis framed was tested using the t-statistic. The statement of the hypothesis regarding the population parameters is provided in subsection (2.3). The data was divided into two samples, hotspot districts with a sample size of 14 and coldspot districts with a sample size of 19. The t-test with two samples assuming equal variances was performed (Table 6). The t-test statistic, $t_{calculated}$, was obtained as 4.609895066 which is greater than the tabular value, $t_{critical} = 2.039513446$. Since $t_{calculated} > t_{critical}$, we reject the null hypothesis (H_0). This implies that there is a significant difference between the hotspot and coldspot districts, which supports that the hotspot of CAW is exclusively man-made and not due to chance. Hence, we recommend framing and adopting effective strategic policing to combat the threat of CAW.

Forecast Values and Analysis

The simple exponential smoothing method was used to forecast the CAW cases for the year 2021. District wise forecast values were predicted using two different values

Table 7 Forecast Accuracy of Hotspot Districts

Districts	Forecast Accuracy-2021	
	$\alpha = 0.3$	$\alpha = 0.7$
Jhalawar	256.8940611	262.2069723
Baran	134.0252509	146.7783682
Kota	236.5561081	247.1705753
Bundi	124.4159026	142.3952105
Sawai Madhopur	99.99951499	85.5092484
Chittorgarh	144.1011224	140.3411496

α-smoothing constant

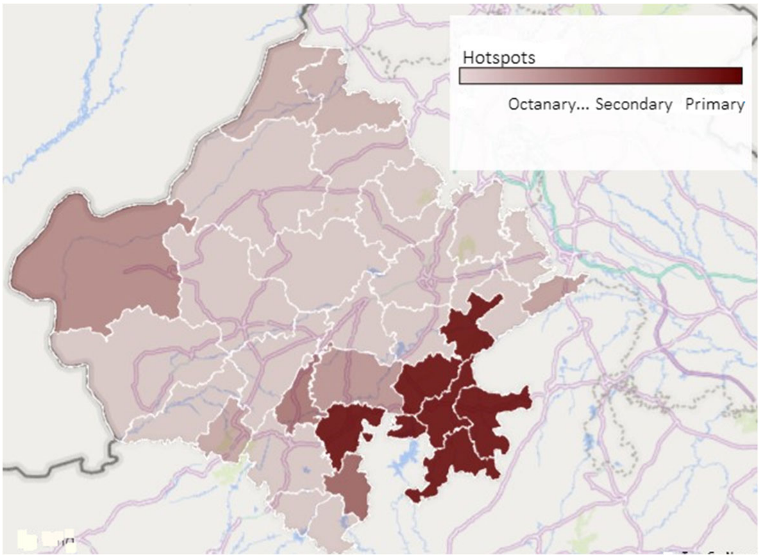

Fig. 7 Choropleth map of CAW Hotspot districts of forecast values

of smoothing constant, i.e., $\alpha = 0.3$ and $\alpha = 0.7$ (Table 7). As the smoothing constant α lies between 0 and 1, the α-value tends to unity, represent better forecast accuracy. This means that the future values predicted will have a larger probability to explain and represent the original data. Further, hotspot analysis was performed with the predicted values. Jhalawar, Baran, Kota, Bundi, Sawai Madhopur, and Chittorgarh recurred as the hotspots for the predicted values of the year 2021 (Fig. 7), which have also been the consistent hotspots from the years 2014 to 2020. Validation of the hotspots of 2021 was again carried out using hypothesis testing and found a significant difference in the rate of cases in hotspot and coldspot locations. Also, when compared with the actual data from the NCRB website released for the year 2021, the percentage error between the actual and forecast value turns out to be less with a higher value of α (Table 8). Thus, with an increase in the value of smoothing constant α, the percentage error between the actual and forecast values is minimum. Thus, we may say that the forecast value obtained with $\alpha = 0.7$ provided us more accurate forecasted value as compared to $\alpha = 0.3$. Table 9 shows that the root mean squared error (RMSE) is comparatively less for $\alpha = 0.7$ than $\alpha = 0.3$. This is to note that with the increase in the value of α the RMSE gets minimized.

Table 8 Percentage Error between Actual and Forecast values

	2021	
	$\alpha = 0.3$	$\alpha = 0.7$
Actual	40056	40056
Forecast	32834.4878	35304.79916
Percentage Error	18.02854054	11.86139614

α-smoothing constant

Table 9 Root Mean Squared Error

	Forecast and RMSE-2021	
	$\alpha = 0.3$	$\alpha = 0.7$
Forecast	32834.4878	35304.79916
RMSE	5957.105759	6107.629358

α-smoothing constant

Conclusion

This work highlights the areas with MLC of CAW during the time span 2014 to 2020 in Rajasthan state. We found that the hotspot districts are consistent over the time period under study. Pratapgarh district was found to be primary hotspsot for the year 2014 and secondary hotspot from 2015 to 2020. Also, Jhalawar, Baran, Kota, Bundi, Sawai Madhopur, and Chittorgarh as the secondary hotspot districts for the year 2014 and primary hotspot from the years 2015 to 2020. It has been observed that though Jaipur, the capital city of Rajasthan, characterized as the most urbanized, densely populated district with highest number of CAW cases did not emerge as MLC in any of the years. The hotspot districts are mainly those lying in the eastern part of Rajasthan sharing a boundary with Madhya Pradesh state. Even the descriptive analysis also ranked the most consistent districts mainly in the boundary area of Rajasthan. Therefore, we may conclude that as crime is a state subject the bordering districts are at higher risk, and more vulnerable to cases like CAW. Concerning these facts, our work has been successful for identifying the hotspots as mentioned above but excluded Jaipur which otherwise appears to be the hotspot.

Further, the graphical representation of the CAW cases for the time span of 2014 to 2020 presented a deep insight into variational patterns. The criminal cases show a minimum level of fluctuations from the years 2014 to 2018 but the year 2019 is a paradigm shift toward a sudden spurt of CAW cases which is very depressing especially when we aim toward the world leader of the 21st century. Also, due to the Covid-19 pandemic, we observed a decline in the number of cases in the year 2020. At the same time, the CAW cases show a hike in the year 2021 at an alarming level. This could be cited as strong evidence of the failure of policy effectiveness to combat and handle such a heinous crime in the state of Rajasthan. The government action seems to be inadequate in handling the state situation.

The CAW for the year 2021 predicted using the simple exponential forecasting method shows approximately about 2.5 percent rise in the CAW from the previous year whereas the original data of the year 2021 gives more than 16 percent rise. The hotspot analysis of the predicted value of 2021 ends up giving the same regions again as hotspots of CAW. Further, it appears that the hotspot districts are static in nature rather than dynamic in the locational aspect during the period of study.

Limitations

The first limitation of our work is the nonavailability of the CAW data at the district level for the year 2021 which restricts us to identify the hotspot districts of the original cases. The second limitation of our work is that the data used in this study were collected as a secondary source from the NCRB website. Thus, a primary data collection need arises for more precise analysis and prediction of CAW. But since CAW is a very sensitive issue it is difficult to gather information on an individual level and a huge number of cases go unreported. Also, the collection of primary-level data from all the thirty-three districts of Rajasthan is again a mammoth task.

Future Scope

As an extension to our work, we plan to study the nature and factors behind the CAW in the districts of Rajasthan. Our work will lead to further opportunities for new researchers to find out the root cause analysis of the hotspot region. This paper not only gives a road map to the Police department and administration rather than it gives an opportunity to other researchers too to investigate further in the area.

Acknowledgements The authors are thankful to the Vellore Institute of Technology, Vellore, for providing us with the required facilities to successfully carry out this research work.

Author Contributions All the authors contributed equally.

Funding The authors did not receive any funding for this research from any organization.

Data Availability The datasets generated and analysed during the study are freely applicable on the National Crime Records Bureau (NCRB) and Census of India Website.

Declarations

Conflicting interests The authors declared no conflict of interest.

References

Abolhassani, A., & Prates, M. O. (2021). An up-to-date review of scan statistics. *Statistics Surveys, 15*, 111–153.

Ahmad, F., Uddin, M. M., & Goparaju, L. (2017). Role of geospatial technology in crime mapping: A perspective view of India. *World Scientific News, 88*(2), 211–226.

Almanie, T., Mirza, R., & Lor, E. (2015). Crime prediction based on crime types and using spatial and temporal criminal hotspots. *International Journal of Data Mining & Knowledge Management Process, 5*(4), 01–19.

Al-Sabbagh, T. A., Li, Y., Lee, Y. J., & El Kenawy, A. M. (2023). Land use and sexual harassment: A geospatial analysis based on the volunteer HarassMap-Egypt. *Geographical Research,* 01–18. https://doi.org/10.1111/1745-5871.12619

Anjali, Kumar, B. R., & Kumar, J. (2021). Spatio-temporal aspect of suicide and suicidal ideation: An application of SaTScan to detect hotspots in four major cities of Tamil Nadu. *Journal of Scientific Research, 65*(9), 07–18.

Bhardwaj, A. S., Divakar, K. M., Ashinik, A., Devishree, D. S., & Younis, S. M. (2019). Deep learning architectures for crime occurrence detection and prediction. *International Journal of Advance Research, Ideas and Innovations in Technology, 5*(2), 822–824.

Bowers, K., Johnson, S., & Pease, K. (2004). Prospective Hot-Spotting: The Future of Crime Mapping? *British Journal of Criminology, 44*(5), 641–658.

Crime against women in India. (2021). *National Crime Record Bureau*. Government of India: Ministry of Home Affairs.

Dayal, B., & Singh, N. (2014). Crimes against women and societal ills: an overview. *International Journal of Advanced Scientific and Technical Research, 3*(4), 44–56.

Debnath, M., & Ray, S. (2019). Population move on Rajasthan: Regional analysis. *Journal of Geography and Regional Planning, 12*(3), 43–51.

Garfias Royo, M., Parikh, P., & Belur, J. (2020). Using heat maps to identify areas prone to violence against women in the public sphere. *Crime Science, 9*(15), 1–15.

Gorr, W. L., & Lee, Y. (2014). Early Warning System for Temporary Crime Hot Spots. *Journal of Quantitative Criminology, 31*(1), 1–23.

Groff, E. R., Weisburd, D., & Yang, S. M. (2010). Is it Important to Examine Crime Trends at a Local "Micro" Level?: A Longitudinal Analysis of Street to Street Variability in Crime Trajectories. *Journal of Quantitative Criminology, 26*, 7–32.

Hackett, M. (2011). Domestic Violence against Women: Statistical analysis of crimes across India. *Journal of Comparative Family Studies, 42*(2), 267–288.

Hipp, J. R., & Kim, Y. A. (2017). Measuring Crime Concentration Across Cities of Varying Sizes: Complications Based on the Spatial and Temporal Scale Employed. *Journal of Quantitative Criminology, 33*(3), 595–632.

Jung, I. (2019). Spatial scan statistics for matched case-control data. *PLoS ONE, 14*(8), 1–10.

Khadke, P. A., Kamble, K. R., & Kharat, R. U. (2019). Spatial and temporal analysis of crime against women in Nanded city. *International Journal of Research in Social Sciences, 9*(3), 678–692.

Kuralarasan, K., & Bernasco, W. (2022). Location Choice of Snatching Offenders in Chennai City. *Journal of Quantitative Criminology, 38*(3), 673–696.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics: Theory and Methods, 26*, 1481–1496.

Kulldorff, M. (2021). *SaTScan v10.0: Software for the spatial and space-time scan statistics*. Information Management Services Inc., 2021. Studies

Lama, S., & Rathore, S. (2018). Crime Mapping and Crime Analysis of Property Crimes in Jodhpur. *International Annals of Criminology, 55*(2), 01–15.

Lee, Y., & Eck, J. E. (2023). *Measures of crime concentration for places and other units* (pp. 187–196). A Methods Handbook: Understanding Crime and Place.

Lolayekar, A. P., Desouza, S., & Mukhopadhyay, P. (2020). Crimes Against Women in India: A District-Level Analysis (1991–2011). *Journal of Interpersonal Violence, 37*, 7289–7314.

Maciejewski, R., Hafen, R., Rudolph, S., Larew, S. G., Mitchell, M. A., Cleveland, W. S., & Ebert, D. S. (2011). Forecasting hotspots-a predictive analytics approach. *IEEE Transactions on Visualization and Computer Graphics, 17*(4), 440–453.

Mondal, S., Singh, D., & Kumar, R. (2022). Crime hotspot detection using statistical and geospatial methods: A case study of Pune city, Maharashtra. *India. GeoJournal, 87*(1), 1–17.

Morrison, B. (2006). Restorative Justice and Civil Society: Emerging Practice, Theory and Evidence. *Journal of Social Issues, 62*(2), 209–215.

Oliveira, B. M., Lucena, K. D. T., Gomes, R. G. S., Coêlho, H. F. C., Vianna, R. P. T., & Meira, R. M. B. (2019). Spatial distribution of domestic violence against women. *Journal of Human Growth and Development, 29*(1), 102–9.

Ostertagova, E., & Ostertag, O. (2011). *2011*. Modeling of Mechanical and Mechatronic Systems: The Simple Exponential Smoothing Model.

Ostertagova, E., & Ostertag, O. (2012). Forecasting Using Simple Exponential Smoothing Method. *Acta Electrotechnica et Informatica, 12*(3), 62–66.

Patel, A. B. (2015). Crime against the women in India. *Forensic Research & Criminology International Journal, 1*(4), 1–5.

Runadi, T., & Widyaningsih, Y. (2017). Application of hotspot detection using spatial scan statistic: Study of criminality in Indonesia. *AIP conference proceedings, 1827*(1), 020011.

Seid, E., Melese, T., & Alemu, K. (2021). Spatial distribution and predictors of domestic violence against women: evidence from analysis of Ethiopian demographic health survey 2016. *BMC Women's Health, 21*(334). https://doi.org/10.1186/s12905-021-01465-4

Sharma, V., Ishtiaque, M., & Kumar, D. (2020). Spatio-Temporal Analysis of Crimes Against Women in Uttar Pradesh (2001–2018). *National Geographical Journal of India, 66*(4), 371–386.

Shastri, S., Sharma, A., & Mansotra, V. (2015). A Model for Forecasting Tourists Arrival in J&K. *India. International Journal of Computer Applications, 129*(15), 32–36.

Sukhija, K., Singh, S. and Kumar, J. (2017). Spatial visualization approach for detecting criminal hotspots: An analysis of total cognizable crimes in the state of Haryana. 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). https://doi.org/10.1109/RTEICT.2017.8256761

Tayal, D., Jain, A., Arora, S., Agarwal, S., Gupta, T., & Tyagi, N. (2014). Crime detection and criminal identification in India using data mining techniques. *AI & Society, 30*(1), 117–127.

Townsley, M. (2008). Visualizing space time patterns in crime: The hotspot plot. *Crime Pattern and Analysis, 1*(1), 61–74.

Vicente, G., Goicoa, T., Fernandez-Rasines, P., & Ugarte, M. (2019). Crime against women in India: unveiling spatial patterns and temporal trends of dowry deaths in the districts of Uttar Pradesh. *Journal of the Royal Statistical Society, 183*(2), 655–679.

Women Peace and Security Index (2021/22). Report, Georgetown Institute for Women, Peace and Security and Peace Research Institute Oslo, Washington DC.