Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-026-05096-1

ORIGINAL RESEARCH

A novel fuzzy clustering urban hot spot detection method—an application in crime analysis

Rosa Cafaro[1] · Barbara Cardone[1] · Ferdinando Di Martino[1,2] · Cristiano Mauriello[1] · Vittorio Miraglia[1]

Received: 11 July 2025 / Accepted: 18 May 2026

Abstract

Urban hotspot detection methods represent a significant feature in many urban analysis problems because they allow to detect where a phenomenon has intensified and how it evolves over time. K-means and Fuzzy C-means clustering algorithms are frequently used in urban hotspot detection because they have the advantage of being computationally fast; however, they are not very robust to the presence of noise and outliers and fail to capture hotspots that are not well separable. In order to increase the accuracy of hotspot detection methods based on the Fuzzy C-means algorithm, a novel hotspot detection method is proposed that uses the weighted Fuzzy C-means algorithm to evaluate the impact of the data points density. The proposed algorithm has the advantage of improving the performance of Fuzzy C-means-based hotspot detection algorithms, increasing robustness to data noise and capturing overlapping and elongated hotspots, without reducing computational speed. Experimental tests were carried out on crime analysis events that occurred in the District of Columbia (USA) between 2019 and 2024. They highlighted the ability of the proposed hotspot detection method to capture intersecting and elongated hotspots and to provide a valid tool for localization and analysis of the temporal evolution of urban hotspots.

Keywords FCM · wFCM · Hotspot · Urban hotspot · GIS

1 Introduction

Urban hotspots are defined as urban areas with higher concentration of events/objects compared to the expected number given a random distribution of events/objects (Zeng and Xiang 2021). Each event is registered specifying its spatial reference, the time it occurred and all the information that characterizes it. It is displayed as a point on the map.

Clustering methods are generally used to detect hotspots. The data points are made up of events assigned as elements with point geometry on the map. Clustering algorithms are used to locate and construct hot spots as elements with polygonal geometry on a map, corresponding to regions of the study area where the phenomenon is most insistent. Moreover, by analyzing the location and extension of hot spots detected in successive time frames, it is possible to study their evolution over time.

Several cluster-based hotspot detection algorithms have been proposed in the last decades in various problems. K-means, due to its high computational speed, has been used in numerous applications to detect hotspots from localized events and analyze their changes over time. Agarwal et al. (2013), Sing and Manimannan (2013), Hajela et al. (2020) and Dubey et al. (2024) use K-means to detect hotspots in crime analysis. In Vadrevu et al. (2013) and Khairani Nabila and Sutoyo (2020) K-means is applied to detect hotspots in fire analysis. In Wahyono et al. (2024) K-means is applied to detect traffic accident hotspots.

✉ Ferdinando Di Martino
fdimarti@unina.it

Rosa Cafaro
rosa.cafaro@unina.it

Barbara Cardone
b.cardone@unina.it

Cristiano Mauriello
cristiano.mauriello@unina.it

Vittorio Miraglia
vittorio.miraglia@unina.it

1 Department of Architecture, University of Naples Federico II, 80134 Naples, Italy

2 Interdepartmental Center "Alberto Calza Bini", University of Naples Federico II, 80134 Naples, Italy

Published online: 02 June 2026

A widely used fuzzy clustering algorithm in hotspot detection is Fuzzy C-means, Bedzek (1981), (FCM). FCM is well known for its ability to handle fuzziness in data by allowing data points to belong to multiple clusters with different membership degrees; this allows for greater flexibility and more accurate clustering results interpretation. Like K-means, FCM provides good computational speed. FCM is applied in crime analysis (Kaur and Sehra 2014) and (Ansari et al. 2018), in disease analysis (Bandyopadhyaya and Mitra 2015) in road traffic crashes (Win et al. 2019)

Although they are very fast and easy to implement, K-means and FCM require a priori setting of the number of clusters and are sensitive to the choice of initial values, which can lead to local minimum.

A validity index can be used to set the best value of the number of clusters. To optimize the cluster initialization some authors developed hybrid models in which meta-heuristic techniques were used to optimize the setting of the initial cluster centers. A hybrid variation of K-means, in which the Particle Swarm Optimization algorithm is executed to find the optimal cluster centers, is tested in Li et al (2022) to detect hotspots from GPS data. In Ran et al (2021) a noise algorithm is employed in a K-means urban hotspot detection method to optimize the initialization of the clusters. In Cardone and Di Martino (2020) the De Luca and Termini fuzzy entropy measure, De Luca and Termini (1972), is applied in FCM to measure the fuzziness of clusters and optimize their initialization.

The critical point of these approaches consists of the increase in computational complexity, which affects their use in the presence of massive data. Moreover, K-means and FCM produce point-based cluster prototypes that provide the location of cluster centers. One of their limitations is the inability to capture the geometric shape of the hotspot; for this reason, some researchers proposed the use of density-based cluster methods, which are able to capture the geometric shape of clusters.

Density-based hotspot detection algorithms are proposed in traffic crashes analysis (Lin et al. 2011), in soil pollution (Chaney and Ratcliffe 2005) and in crime analysis Kalinic and Crisp (2018). DBSCAN is used in Santos et al (2021) for detecting traffic accident hotspots. Scalable density-based hotspot detection methods are applied in Cesario et al (2021) and Cesario et al. (2024), where DBSCAN is organized in a parallelized structure. In Huang et al. (2021) a fast version of DBSCAN is applied to GPS trajectory data to extract taxi passenger hot spots. The Kernel Density Estimation algorithm (KDE) is tested in Memisoglu Baykal (2025) for forest fire hotspot detection; comparison test with the traditional Getis Ord Gi* and Anselin Local Moran's I detection algorithm; comparative tests showed that KDE does not improve the accuracy of detected clusters compared to the other two methods. An urban hotspot detection method based on the spatiotemporal hierarchical density-based cluster algorithm ST-HDBSCAN was proposed in Li et al. (2021) to capture taxi trajectory hotspots and analyze its temporal evolution; this approach has the disadvantage of being too expensive for massive data and is particularly sensitive to the choice of parameters.

Density-based algorithms detect the geometric shape of hotspots more accurately than K-means and FCM; however, they can be too expensive as the cardinality of the dataset increases; furthermore, they are sensitive to the data scale, and it is necessary to appropriately fix the maximum distance between two points so that they are considered close. Approximation of the shape of hotspots with circular areas can represent a trade-off between computational speed and hotspot accuracy (Shen et al. 2019).

In Di Martino and Sessa (2011) and Di Martino et al. (2014, 2020) a hotspot detection method based on an extension of the FCM algorithm was experimented, which provides circular-shaped hotspots it proposed. In Cardone and Di Martino (2022) this method was tested to detect the temporal variation of crime hotspots in the city of London.

However, these methods have the defect of not sufficiently modeling hotspots with elongated geometries; furthermore, they are not able to detect compact and small hotspots.

In this paper, we propose an FCM-based hotspot detection method that uses the weighted FCM algorithm (Wang et al. 2004), (wFCM) to account for the densities of data points. The K Nearest Neighbors algorithm (KNN) is included in FCM to evaluate the weights to assign to the data points and the covariance matrix of each cluster is calculated to assess their geometric structure as an ellipse on the map.

The proposed method has the advantage of detecting better real hotspots than FCM, taking into account the density of data points. In fact, FCM fails to capture hotspots where data point densities have significant variations and clusters are not well separable and is not very robust to the presence of noise and outliers in the data.

In particular:

- points in dense regions will have high weights. FCM centroids will be more strongly attracted to dense points and clusters are better centered on hot spots;
- isolated points and outliers will have low weights and will not affect clustering;
- wFCM is better than FCM at detecting hotspots in real datasets where clusters are not clearly separated, and data point densities are variable.

In a nutshell the proposed local density-based Weighted FCM improves the performance of FCM-based hotspot detection algorithms, increasing robustness to noise in the data and capturing elongated hotspots, without, on the other hand, increasing the computational complexity of the algorithm. It assigns more influence to points in densely populated areas, improving the location, shape and significance of clusters compared to a standard FCM.

The remainder of this paper is organized as follows. In Sect. 2 the FCM and wFCM algorithms are presented; Sect. 3 describes introduces the proposed wFCM-based hotspot detection method In Sect. 4 are illustrates and discussed the results of comparison with other FCM-based algorithms using well-known classification data sets. Section 5 concludes this paper with final discussions.

2 Preliminaries

In this section the FCM and wFCM algorithms are briefly introduced.

2.1 The FCM algorithm

Let $X=\{\mathbf{x_1},\ldots,\mathbf{x_N}\}$ be a set of N data points where each point $\mathbf{x_j}$ is a vector in the feature space R^n. FCM applies the Euclidean metrics in the feature space to partition the data points in C clusters by minimizing the following objective function [2]:

$$J(\mathbf{U},\mathbf{V}) = \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^{m} d_{ij}^{2} \tag{1}$$

where $V=\{\mathbf{v}_1, \mathbf{v}_2,\ldots,\mathbf{v}_C\}$ is the set of the cluster centers and the matrix $\mathbf{U}$ is the $C \times N$ partition matrix, whose elements u_{ij} provide the membership degree of the jth data point to the ith cluster.

The value d_{ij} represents the Euclidean distance between the ith cluster center and the jth data point. The fuzzifier parameter m, is a value greater than 1 that controls the fuzziness of the clustering. Generally, an accepted value for m is m = 2.

The solutions for V and $\mathbf{U}$ that minimize J are obtained via the Lagrange multiplicators. They are given by:

$$\mathbf{v}_i = \frac{\sum_{j=1}^{N} u_{ij}^{m} \boldsymbol{x}_j}{\sum_{j=1}^{N} u_{ij}^{m}} \tag{2}$$

and

$$u_{ij} = \frac{1}{\sum_{h=1}^{c}\left(\frac{d_{ij}}{d_{hj}}\right)^{\frac{2}{m-1}}} \tag{3}$$

FCM calculates iteratively the cluster centers and the partition matrix via (2) and (3) until the following condition is reached:

$$\left|\mathbf{U}^{(t)} - \mathbf{U}^{(t-1)}\right| = \max\left\{\left|u_{ij}^{(t)} - u_{ij}^{(t-1)}\right|\right\} < \varepsilon i = 1,\ldots,C j = 1,\ldots,N \tag{4}$$

where $u_{ij}^{(t)}$ and $u_{ij}^{(t-1)}$ are, respectively, the membership degree of the j*th* data point to the i*th* cluster in the current cycle and in the previous cycle and the parameter ε, called iteration error, is a prefixed value.

Below the FCM algorithm is shown in pseudocode.

	Input:	X, C, m, ε
	Output:	V, **U**
1.	Initialize randomly $\mathbf{v}_i$	
2.	Calculate u_{ij} i=1,…, C using (3)	
3.	$u_{ij}^{prev} := 0$ i=1,…, C	
4.	*While* $\lvert\mathbf{U} - \mathbf{U}^{prev}\rvert > \varepsilon$	
5.	Calculate v_{ij} using (2)	
6.	$u_{ij}^{prev} := u_{ij}$	
7.	Calculate u_{ij} using (3)	
8.	*End while*	
9.	*Return* **V, U**	

Algorithm 1: FCM

2.2 The weighted FCM algorithm

wFCM is a robust clustering algorithm that extends the classical FCM method by integrating feature weighting into the clustering process. In wFCM a weight is assigned to each data point following a particular strategy; it is applied in many problems in which is useful assign different relevance to the data points, for example, if a data point is given by the average of a set of measure, to it can be assigned a weight proportional to the inverse of the standard deviation, or when it is necessary to mitigate the presence of outliers, assigning them a very low weight.

In wFCM the objective function to minimize is given by:

$$J_w(\mathbf{U},\mathbf{V}) = \sum_{i=1}^{C}\sum_{j=1}^{N} w_j u_{ij}^{m} d_{ij}^{2} \tag{5}$$

where w_j is the weight assigned to the jth data point.

Applying the Lagrange multipliers, are obtained the following solutions for the components of $\mathbf{V}$.

$$\mathbf{v}_i = \frac{\sum_{j=1}^{N} w_j u_{ij}^{m} \mathbf{x}_j}{\sum_{j=1}^{N} w_j u_{ij}^{m}} \tag{6}$$

The solution for the partition matrix $\mathbf{U}$ is given by (3).

The algorithm wFCM is schematized in pseudocode below:

3 The proposed wFCM-based hotspot detection algorithm

Let $X = \{\mathbf{x_1},\ldots,\mathbf{x_N}\} \subset R^n$ be a set of N data points where each point $\mathbf{x_j} \in R^2$ is given by a couple of geographic coordinates.

The proposed hotspot detection method uses a wFCM algorithm, where the weight assigned to the jth data point is given by:

$$w_j = \frac{1}{\frac{1}{K}\sum_{h=1}^{K} d_{hj} + \delta} \tag{7}$$

The first term in the denominator of (7) is the mean distance of the closest K data points and δ is a positive value close to zero, added to avoid divisions by zero, in case a point has a mean distance close to zero (e.g. if it is very close to its K-nearest neighbors) and to stabilize the weight values to prevent them from becoming too large in ultra-dense areas.

The Xie-Beni validity index, Xie and Beni (1991), (XB) is used to assign the optimal number of clusters. Compared to other well-known validity indices, such as Silhouette, Davies-Bouldin and Dunn, which are designed for hard clustering, XB reflects the uncertainty of clustering in FCM. Furthermore, it is computationally lighter than Silhouette or Dunn especially for massive datasets, as it does not perform complex intra-cluster averaging calculations and does not require recursive operations.

The formula for XB is:

	Input:	$\mathbf{X}$, $\mathbf{w}$, C, m, ε
	Output:	**V, U**
1.	Initialize randomly $\mathbf{v}_i$	
2.	Calculate u_{ij} using (3)	
3.	$u_{ij}^{prev} \coloneqq 0$	
4.	*While* $\lvert \mathbf{U} - \mathbf{U}^{prev} \rvert > \varepsilon$	
5.	Calculate v_{ij} using (6)	
6.	$u_{ij}^{prev} \coloneqq u_{ij}$	
7.	Calculate u_{ij} using (3)	
8.	*End while*	
9.	*Return* **V, U**	

Algorithm 2: wFCM

Table 1 Performance measures of FCM, DBSCAN and wFCM on the first data sample

Index	FCM	DBSCAN	wFCM
Xie-Beni	**0.0055**	-	0.0055
Silhouette	**0.9196**	0.8524	**0.9196**

The best values of the two indices are shown in bold

$$\mathrm{XB(c)} = \frac{\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m}\left\|x_j - v_i\right\|^2}{\min\limits_{r\neq i}\left\|v_r - v_i\right\|^2} \tag{8}$$

where the numerator measures the inter-cluster compactness and the denominator the intra-cluster separation. The intra-cluster compactness is given by the mean square distance between each data point and its cluster centers, the intra-cluster separation is given by the smallest distance between clusters.

The best number of clusters is given by the number C of clusters that maximizes XBI.

After assigning the weight to each data point by (7) and determining the optimal number of clusters, wFCM is executed to detect the C cluster.

The hotspots are assessed as ellipses considering the centroid of each cluster $\mathbf{v}_i$; the correspondent ellipse is created assigning its minor and major semiaxes given by the eigenvalues of the covariance matrix:

$$\boldsymbol{\Sigma}_i = \frac{\sum_{j=1}^{n} u_{ij}^{m}(\mathbf{x}_j - \mathbf{v}_i)(\mathbf{x}_j - \mathbf{v}_i)^T}{\sum_{j=1}^{n} u_{ij}^{m}} \tag{9}$$

Finally, the hotspots are extracted and displayed on the map.

To overcome this problem, we propose a fuzzy entropy-based variation of oFCM called hoFCM, in which the weight assigned to the clusters detected by executing FCM on each chunk, in addition to measuring the extension of the cluster, also measures its compactness. To assess the compactness of a cluster, the De Luca and Termini fuzzy entropy index is used; it is a measure of the degree of fuzziness of a fuzzy set.

The wFCM hotspot detection method is shown in pseudocode in Algorithm 3.

```
Input:    X, w, C, m, ε
Output:   Final hotspots as ellipses

1.  Compute the weight of the data points by (7)
2.  c:=2
3.  Execute wFCM(X,w,c,m, ε)
4.  Compute XB(c) by (8)
5.  For c := 3 to 10
6.      Execute wFCM(X,w,c,m, ε)
7.      Compute XB(c) by (8)
8.      If XB(c) < XB Then
9.          XB = XB(c)
            C = c
10.     End if
11. Next c
12. Execute wFCM(X,w,C,m, ε)
13. For i;= 1 to C
14.     Compute the ellipse axes by (9)
15.     Calculate the density of points in each hotspot
16. Next i
17. Plot the hotspots as ellipses on the map
18. Return the final hotspots
```

Algorithm 3: wFCM hotpot detection

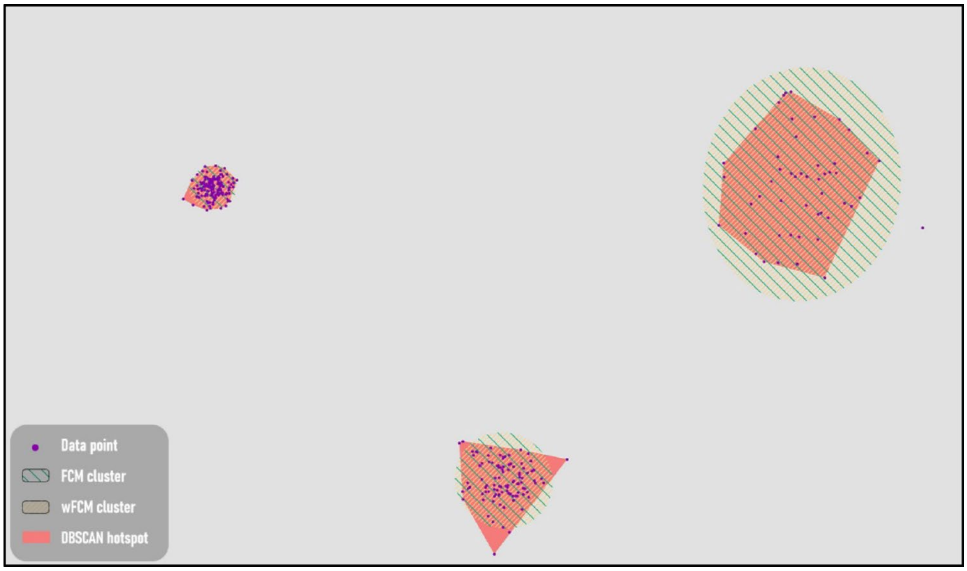

Fig. 1 Hotspots obtained on the first data sample executing he three clustering algorithms

Table 2 Performance measures of FCM, DBSCAN and wFCM on the second data sample

Index	FCM	DBSCAN	wFCM
Xie-Beni	0.0755	–	**0.0707**
Silhouette	0.6429	0.6366	**0.6450**

The best values of the two indices are shown in bold

The algorithm was implemented in the tool GIS Suite ARC-GIS Pro 3.4, using the python library ArcPy for ArcGIS Pro.

4 Results and discussion

To evaluate the performance of wFCM, initial comparative tests were performed with FCM and the density-based clustering algorithm DBSCAN, Ester et al. (1996), Schubert et al. (2017).

4.1 Comparison tests

The responses of the three algorithms are measured considering two point datasets artificially prepared. The two datasets are given by point georeferenced in projected coordinates UTM WGS84.

In the first experiment a dataset composed of three well-separable clusters with high and medium point density is created.

DBSCAN was run several times varying the eps parameter, which defines the maximum distance between two data points for them to be considered close. By varying the eps parameter from 5 to 500, the correct number of clusters, equal to 3, was obtained only for eps = 5; in all other cases the number of clusters was always 1. This result shows that DBSCAN is sensitive to the data scale and needs

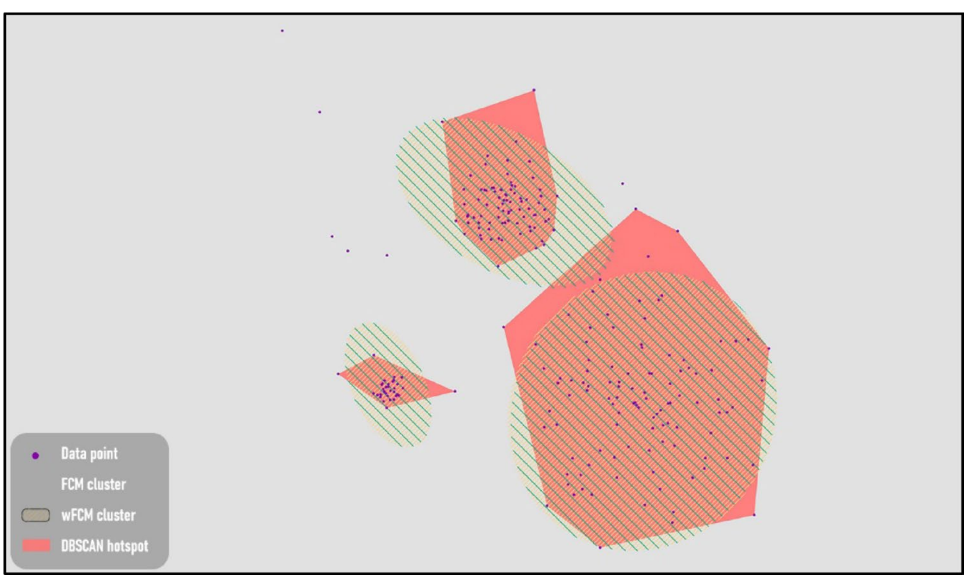

Fig. 2 Hotspots obtained on the second data sample executing the three clustering algorithms

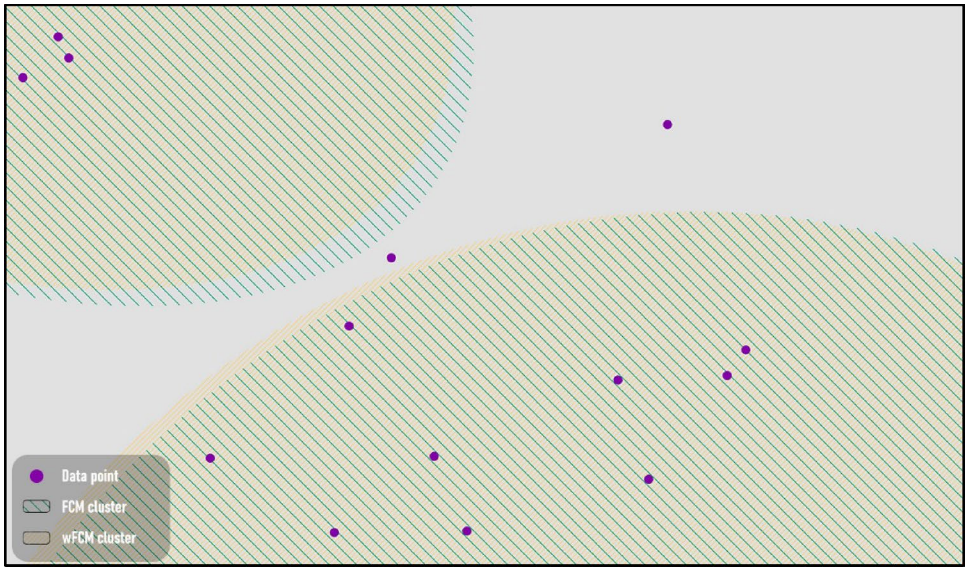

Fig. 3 Differences between hotspots generated by executing FCM and wFCM obtained from the second data sample

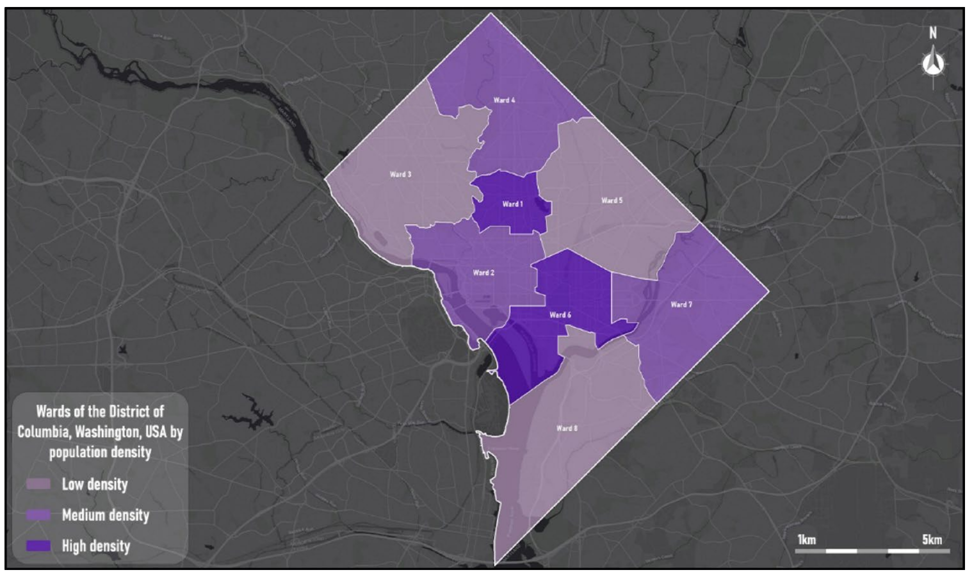

Fig. 4 Study area of the Wards of the District of Columbia, Washington, USA

Table 3 Extension and demographic data of the Wards of the District of Columbia, Washington, USA

Ward	Area [KM^2]	Population	Population density [INHAB. /KM^2]
1	6.57	79,373	12,076.49
2	18.36	78,878	4,296.90
3	28.32	78,404	2,768.43
4	23.31	85,587	3,670.90
5	26.93	88,426	3,283.48
6	17.62	84,004	4,768.64
7	24.00	90,898	3,788.16
8	32.35	86,509	2,674.22

preprocessing elaborations to find the correct value of the eps parameters.

The Xie-Beni index is measured to compare the performances of FCM and wFCM in terms of compactness of the cluster and separability among clusters; the Silouhette index is measured to compare wFCM with DBSCAN. The results, shown in Table 1, highlight that FCM and wFCM provide the same performance; DBSCAN provides a lower value of the Silhouette index index than FCM and wFCM.

In Fig. 1 are shown on the map the clusters obtained using the three algorithms. FCM and wFCM provide exactly the same hotspots, approximated as ellipses, DBSCAN provides hotspots as polygons.

In the second experiment a dataset composed of three not completely well-separated clusters is created; the first cluster has a high point density, the second medium density and the third low density.

DBSCAN was performed several times, varying the eps parameter from 5 to 500. As in the previous case, the correct number of clusters, equal to 3, was obtained only for eps = 5; in all other cases the number of clusters was always 1. This

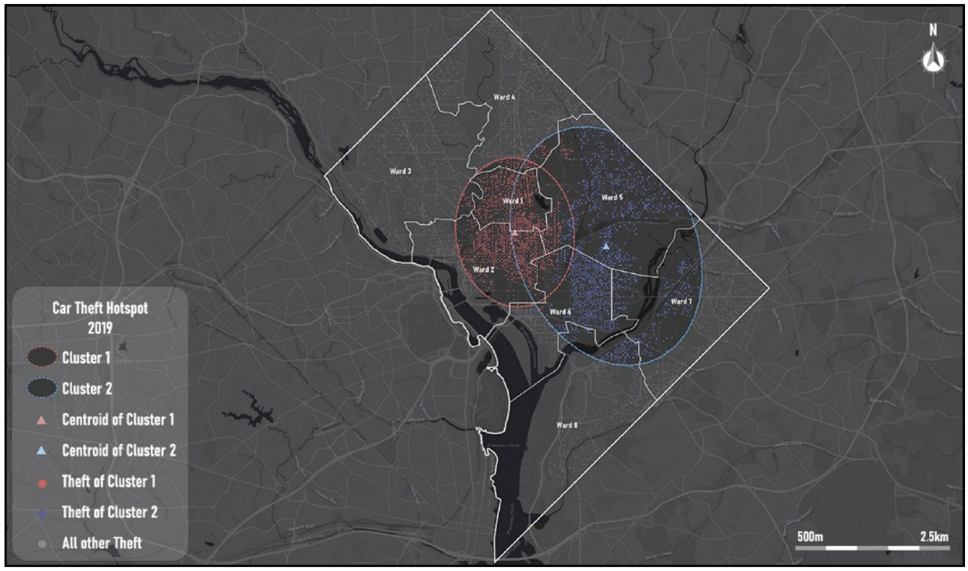

Fig. 5 Car theft events in 2019 in the District of Columbia, Washington (USA)

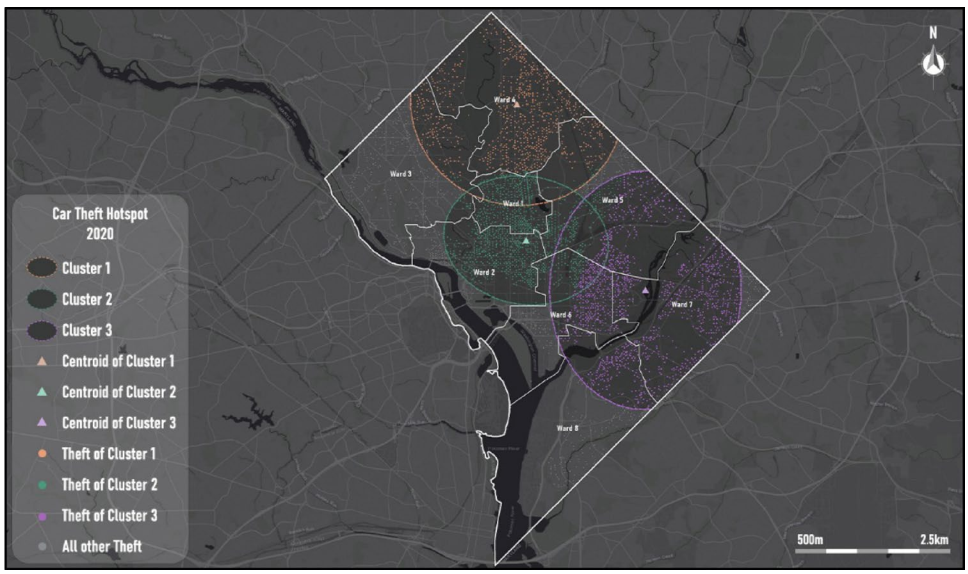

Fig. 6 Car theft events in 2020 in the District of Columbia, Washington (USA)

confirms that DBSCAN is sensitive to the data scale and a preprocessing step is performed to determine the correct value of the eps parameters.

In this second experiment wFCM shows better performance than FCM and DBSCAN, as shown in Table 2. In fact, wFCM provides lower XB index values and higher Silhouette index values than those provided by FCM. This indicates that, wFCM provides better compactness and separation between clusters than FCM. In addition, wFCM also shows better robustness to the presence of outliers in the data.

Figure 2 shows the hotspots detected using the three algorithms.

Unlike the first case, now the hotspots detected using FCM and wFCM are not perfectly coincident. While the two hotspots representing the compact cluster, well separated from the other two, are perfectly coincident, the hotspots related to the other two clusters are not, as shown by the zoom-in of the map in Fig. 3, mainly due to the tendency of FCM to not be very robust with respect to the presence of outliers.

The results in Fig. 2 also show that DBSCAN provides hotspots with a higher density of data points than FCM and wFCM. In fact, the average density of hotspots generated with DBSCAN is 832.35 data points per square kilometer, while that of hotspots generated with FCM and wFCM is, respectively, 446.18 and 450.52 data points per square kilometer. This is an indication of the fact that DBSCAN tends to underestimate the spatial extent and overestimate the apparent density of hotspots.

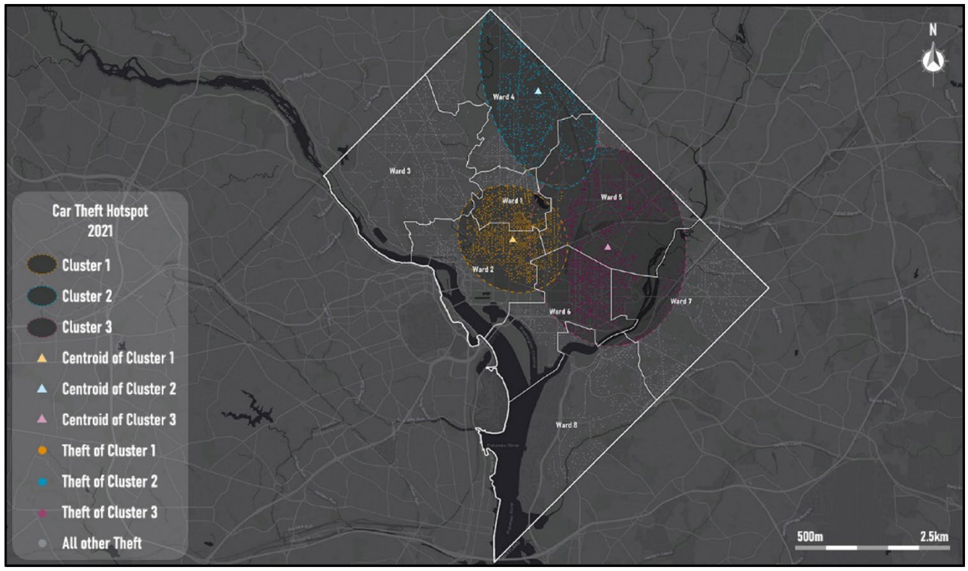

Fig. 7 Car theft events in 2021 in the District of Columbia, Washington (USA)

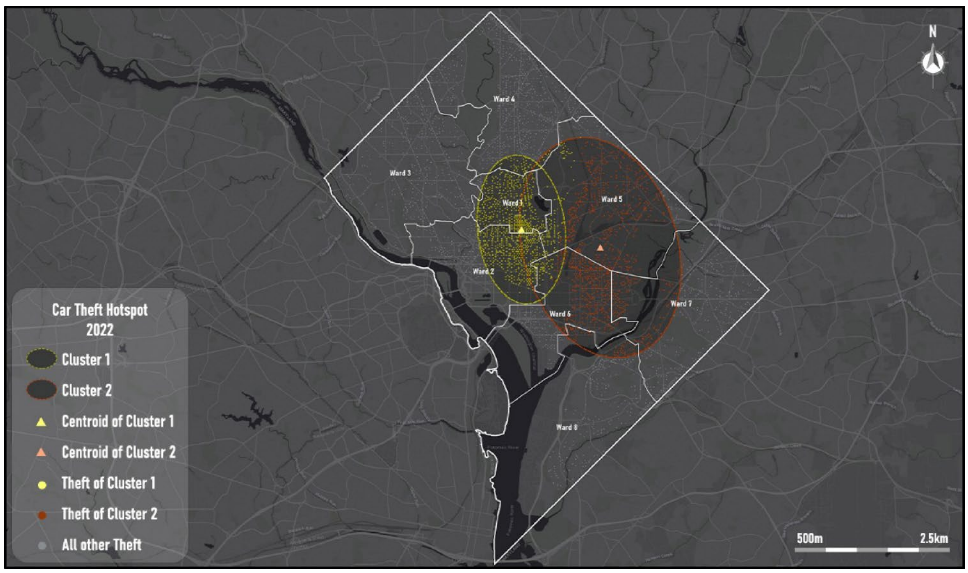

Fig. 8 Car theft events in 2022 in the District of Columbia, Washington (USA)

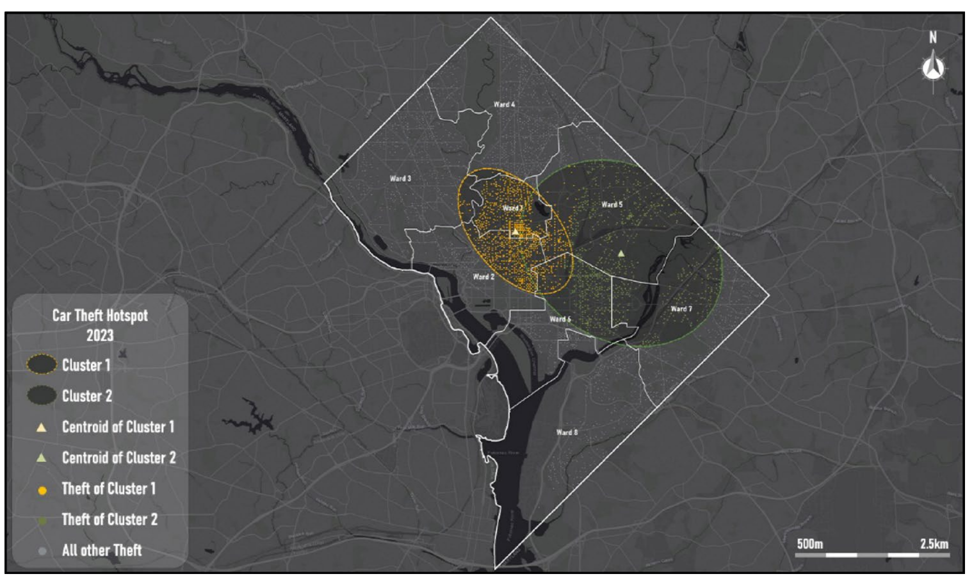

Fig. 9 Car theft events in 2023 in the District of Columbia, Washington (USA)

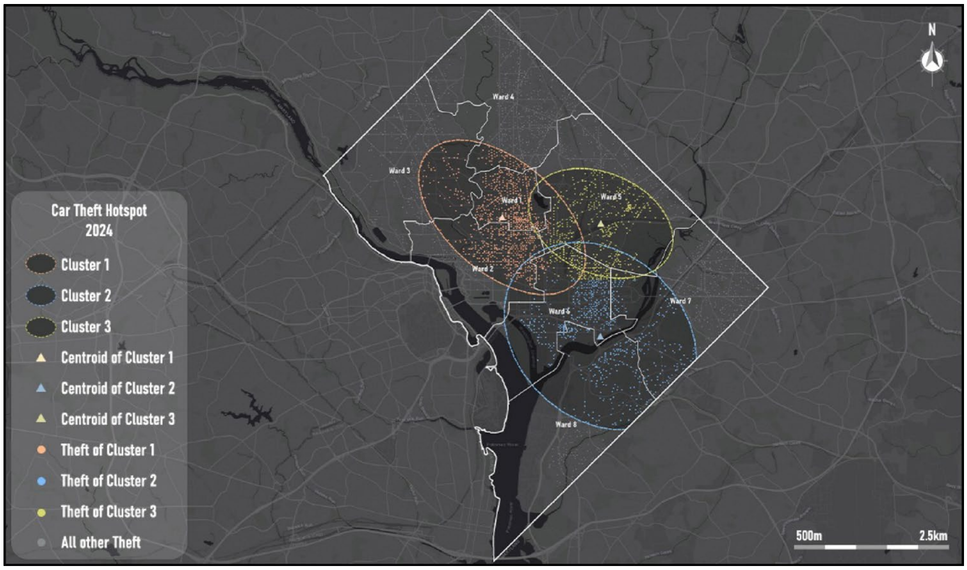

Fig. 10 Car theft events in 2024 in the District of Columbia, Washington (USA)

Table 4 Hotspot overlap year by year 2019–2024

Period	Hotspot's surface [km^2]	Coverage percentage
2019	70.01	–
2020	113.55	95.62%
2021	70.72	60.54%
2022	55.76	72.11%
2023	57.54	84.42%
2024	81.13	78.59%

4.2 An application in crime analysis

The wFCM hotspot detection method has been tested in a crime analysis application to analyze the location and evolution of hotspots generated by criminal events that occurred from 2019 to 2024 in the District of Columbia, Washington (USA). The study area, shown in Fig. 4.

It is divided into wards, which represent geographic subdivisions fundamental to the organization of local government and the distribution of public services.

The District of Columbia is currently divided into eight wards, each with clearly defined boundaries that reflect demographic, geographic, and socioeconomic criteria. These subdivisions do not constitute independent municipal entities, but rather representative units through which urban governance is structured and citizen participation is facilitated. From a socio-demographic perspective, the wards exhibit marked heterogeneity: each presents a distinctive profile in terms of ethnic composition, income levels, access to education, public health status, and urban infrastructure. This diverse configuration makes the wards a privileged object of study for disciplines such as urban planning, urban sociology, and public health, offering a representative microcosm of the tensions and potentials typical of contemporary American metropolises.

Table 3 provides data on the size and demographics of each ward.

Datasets of events were extracted for various types of crimes, including robberies, burglaries, car thefts, motor vehicle thefts, assaults with dangerous weapons, and sexual abuse. For the sake of brevity, the results obtained by applying wFCM on the car test events dataset are shown.

The thematic map illustrated in Fig. 5 shows the car theft events that occurred in 2019 in the study area.

The clustering process identified two distinct hotspots; the first is located in the central area of the district that records a smaller extension but a higher density than the other hotspot identified in the eastern outskirts.

The thematic map illustrated in Fig. 6 shows the car theft events that occurred in 2020 in the district of Columbia.

For the year 2020, the clustering process identified three different hotspots: one in the central area, one in the northern area and one in the eastern area. Cluster number 2, located in the city center, has a smaller extension than the other two but a higher density.

In Fig. 7 is shown the thematic map of car theft hotspots detected in 2021 in the study area.

As in the previous year, 3 different clusters were identified in 2021. Their location coincides with those of 2020: cluster 1, the smallest but densest, is in the city center, cluster 2 is located to the north of the study area and has a larger extension than cluster 1 but smaller than cluster 3, located in the eastern part of the city.

The thematic map illustrated in Fig. 8 shows the car theft events that occurred in 2022 in the study area.

The clusters identified in the year 2022 are 2, as already happened for the year 2019. Cluster 1 is in the city center and has a high density of car thefts; cluster 2, located in

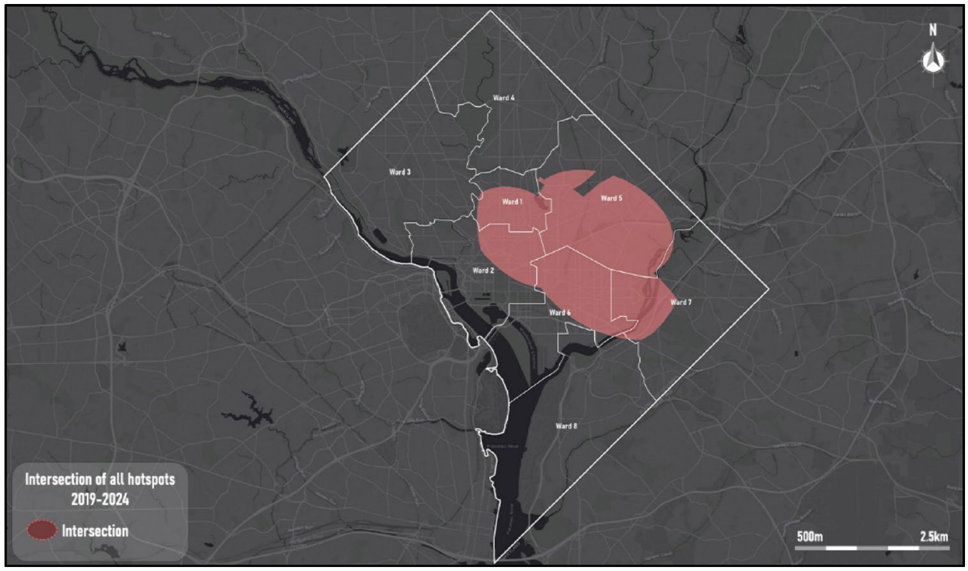

Fig. 11 Intersection area of individual annual hotspots (2019–2024) in the District of Columbia, Washington (USA)

Table 5 Critical area of the Wards of the District of Columbia with respect to car theft events occurred in the period 2019–2024

Ward	Ward critical area [km^2]	Hotspot cover ratio
1	6.57	68.17%
2	18.36	24.78%
3	28.32	0.11%
4	23.31	0.00%
5	26.93	61.53%
6	17.62	46.52%
7	24.00	22.45%
8	32.35	0.23%

the eastern area of the district, has a greater extension but a lower density.

Figure 9 shows the thematic map of the car theft events that occurred in 2023 in the District of Columbia.

As in the previous year, the clusters identified in the year 2023 are 2. The areas involved are the same, the city center with cluster 1, and the eastern area with cluster 2. Cluster 1 records a higher density while cluster 2 has a higher extension.

In Fig. 10 is shown the thematic map of car theft hotspots detected in 2024 in the study area.

Compared to the previous year, the clusters identified in the year 2024 are 3, as already occurred in 2020 and 2021. Cluster 1 is in the city center and, for the first time, records a lower density than the other clusters; cluster 2 located to the southeast of the study area, records the largest extension and the lowest density; and cluster 3, located to the northeast of the district, is the least extended and records the highest density.

In Table 4 are shown the surface covered by hotspots in every year (Hotspot's surface) and, in percentage terms, how much of the surface covered by hotspots in the previous year continues to be covered by hotspots in the year being analyzed (Coverage percentage).

The results in Table 4 highlight that, while the extension of hotspots detected in 2019 was almost completely covered by hotspots detected in 2020, only 60% of the extension of hotspots detected in 2020 was covered by hotspots detected in 2021. This is because the detected hotspots in 2020 collectively cover a much larger surface area than those covered by the hotspots detected in other years. Car thefts were likely more widespread and less concentrated in specific areas of the district in 2020.

Through spatial intersection processes, the surface on which hotspots were recorded during the entire period was identified and calculated. The aim is to identify the area that recorded the greatest criticality during the entire period analyzed, from 2019 to 2024.

Figure 11 shows the thematic map that illustrates the area resulting from the intersection of all the hotspots recorded in the five years.

In Table 5 are shown the extension of the critical area of the Ward and, in percentage, the ratio between the critical area and the area of the Ward (hotspot cover ratio).

The Wards 1 and 5 are the city areas most affected by hotspots. Specifically, over 70% of Ward 1 and over 60% of Ward 5 were covered by hotspots during all the six years from 2019 to 2024; this result highlights that the frequency of car thefts in these wards remained high throughout the investigation period. On the contrary, in Wards 3, 4 and 8 car thefts were much less frequent and did not occur consistently throughout this period. Further observation concerns the relationship between hotspot coverage and population density: although Wards 2 and 6 have high population densities, their hotspot incidence is significantly lower than

Ward 1. This suggests that population density alone is not sufficient to explain the presence of hotspots, and that other factors, such as the socioeconomic fabric, urban morphology, or the functional distribution of the territory, may play a decisive role. Furthermore, the interannual variation in coverage suggests that crime, while maintaining a certain spatial stationarity, may exhibit temporal dynamics linked to seasonal, pandemic, or economic factors.

5 Conclusion

This work tested a novel urban hotspot detection method based on a weighted FCM algorithm, where the weighting takes into account the density of data points. Comparative tests demonstrated that the proposed method improves the accuracy of the FCM hotspot detection algorithm, increasing its robustness to noise and its ability to capture overlapping and non-compact hotspots. In addition, compared to density-based clustering algorithms such as DBSCAN, it has the ability to capture even non-dense hotspots.

Further tests conducted on crime analysis events that occurred in the District of Columbia (USA) between 2019 and 2024 highlighted that the proposed method represents a robust and valid tool to support decision makers in analyzing the spatial distribution and temporal evolution of urban hotspots.

Looking ahead, we intend to further improve the accuracy of the proposed wFCM-based urban hotspot detection method by incorporating more dimensions into the weights, taking into account not only spatial density but also additional characteristics, such as event intensity and local vulnerability characteristics, related to an area's susceptibility to damage in the event of a given event.

Author contributions All authors wrote the original draft of the manuscript, conducted formal analysis and reviewed and revised the manuscript.

Funding Open access funding provided by Università degli Studi di Napoli Federico II within the CRUI-CARE Agreement. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval and consent to participate Not applicable.

Consent for publication Not applicable.

References

Agarwal J, Nagpal R, Sehgal R (2013) Crime analysis using K-Means clustering. Int J Comput Appl 83(4):1–4. https://doi.org/10.5120/14433-2579

Ansari M, Prakash A, Siddique M (2018) Application of spatiotemporal fuzzy C-Means clustering for crime spot detection. Def Sci J 68:374–380. https://doi.org/10.14429/dsj.68.12518

Bandyopadhyaya R, Mitra S (2015) Fuzzy cluster–based method of hotspot detection with limited information. J Transp Saf Secur. https://doi.org/10.1080/19439962.2014.959583

Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Kluwer Academic Publishers, Norwell, p 256. https://doi.org/10.1007/978-1-4757-0450-1

Cardone B, Di Martino F (2020) A novel fuzzy entropy-based method to improve the performance of the fuzzy C-Means algorithm. Electronics 9(4):554. https://doi.org/10.3390/electronics9040554

Cardone B, Di Martino F (2022) Fuzzy-based spatiotemporal hot spot intensity and propagation—an application in crime analysis. Electronics 11(3):370. https://doi.org/10.3390/electronics11030370

Cesario E, Lindia P, Vinci A (2024) A scalable multi-density clustering approach to detect city hotspots in a smart city. Future Gener Comput Syst 157:226–236. https://doi.org/10.1016/j.future.2024.03.042

Cesario E, Vinci A, Zarin S (2021) Towards parallel multi-density clustering for urban hotspots detection. In: 29th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, PDP, IEEE pp. 245–248. https://doi.org/10.1109/PDP52278.2021.00046

Chaney S, Ratcliffe J (2005) GIS and crime mapping Chap. 6, Identifying crime hot spots. Wiley, p 402. https://doi.org/10.1002/9781118685181

De Luca A, Termini S (1972) A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. Inf Control 20:301–312. https://doi.org/10.1016/S0019-9958(72)90199-4

Di Martino F, Sessa S (2011) The extended fuzzy C-means algorithm for hot spots in spatio-temporal GIS. Expert Syst Appl 38:11829–11836. https://doi.org/10.1016/j.eswa.2011.03.071

Di Martino F, Sessa S, Barillari U, Barillari M (2014) Spatio-temporal hot spots and application on a disease analysis case via GIS. Soft Comput 18:2377–2384. https://doi.org/10.1007/s00500-013-1211-7

Di Martino F, Pedrycz W, Sessa S (2020) Hierarchical granular hotspots detection. Soft Comput 24(2):1357–1376. https://doi.org/10.1007/s00500-019-03971-y

Dubey A, Venu N, Bisen D, Garg P, Srinivas K, Arya H (2024) A clustering-based method for hotspot recognition in crime forecasting. In: 2024 IEEE 8th International Conference on Information and Communication Technology (CICT), Prayagraj UP, India, pp. 1–6. https://doi.org/10.1109/CICT64037.2024.10899713

Ester M, Kriegel H-P, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Simoudis E, Han J, Fayyad UM (eds) Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press, pp 226–231

Hajela G, Chawla M, Rasool A (2020) A clustering based hot spot identification approach for crime prediction. Procedia Comput Sci 167:1462–1470. https://doi.org/10.1016/j.procs.2020.03.357

Huang Z, Gao S, Cai C, Zheng H, Pan Z, Li W (2021) A rapid density method for taxi passengers hot spot recognition and visualization based on DBSCAN+. Sci Rep 11:9420. https://doi.org/10.1038/s41598-021-88822-3

Kalinic M, Crisp JM (2018) Kernel density estimation (KDE) vs. Hot-spot analysis-detecting criminal hot spots in the City of San Francisco. In: Conference: Agile - 21st Conference on Geo-information Science, Lund, Sweden

Kaur R, Sehra S (2014) Analyzing and displaying of crime hotspots using fuzzy mapping method. Int J Comput Appl 103:25–28. https://doi.org/10.5120/18039-8914

Khairani Nabila A, Sutoyo E (2020) Application of K-Means clustering algorithm for determination of fire-prone areas utilizing hotspots in West Kalimantan Province. Int J Adv Data and Inform Syst 1(1):9–16. https://doi.org/10.25008/ijadis.v1i1.13

Li F, Shi W, Zhang H (2021) A two-phase clustering approach for urban hotspot detection with spatiotemporal and network constraints. IEEE J Sel Top Appl Earth Obs Remote Sens 14:3695–3705. https://doi.org/10.1109/JSTARS.2021.3068308

Li Y, Zhou X, Gu J, Guo K, Deng W (2022) A novel K-Means clustering method for locating urban hotspots based on hybrid heuristic initialization. Appl Sci 12:8047. https://doi.org/10.3390/app12168047

Lin Y-P, Chu H-J, Wu C-F, Chang T-K, Chen C-Y (2011) Hot spot analysis of spatial environmental pollutants using kernel density estimation and geostatistical techniques. Int J Environ Res Public Health 8:75–88. https://doi.org/10.3390/ijerph8010075

Memisoglu Baykal T (2025) Performance assessment of GIS-based spatial clustering methods in forest fire data. Nat Hazards 121:8445–8477. https://doi.org/10.1007/s11069-025-07135-0

Ran X, Zhou X, Lei M, Tepsan W, Deng W (2021) A novel K-Means clustering algorithm with a noise algorithm for capturing urban hotspots. Appl Sci 11:11202. https://doi.org/10.3390/app112311202

Santos D, Saias J, Quaresma P, Nogueira VB (2021) Machine learning approaches to traffic accident analysis and hotspot prediction. Computers 10:157. https://doi.org/10.3390/computers10120157

Schubert E, Sander J, Ester M, Kriegel H-P, Xu X (2017) DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. ACM Trans Database Syst 42(3):19:1–19:21. https://doi.org/10.1145/3068335

Shen F, Pedrycz W, De Moraes R M, Hu X, Wang X, Gacek A (2019) Clustering of information granules in hotspot identification. In: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), New Orleans, LA, USA, 2019, pp. 1–6. https://doi.org/10.1109/FUZZ-IEEE.2019.8858904

Sing AK, Manimannan G (2013) Detecting hot spots on crime data using data mining and geographical information system. Int J Statistika Mathematika 8:5–9

Vadrevu K, Csiszar I, Ellicott E, Giglio L, Badarinath KVS, Vermote E, Justice C (2013) Hotspot analysis of vegetation fires and intensity in the Indian region. IEEE J Sel Top Appl Earth Obs Remote Sens 6(1):224–238. https://doi.org/10.1109/JSTARS.2012.2210699

Wahyono H, Setiaji H, Hartati T, Wiliani N (2024) K-Means clustering for identifying traffic accident hotspots in Depok City. J Appl Res Comput Sci Inf Syst 2(1):159–170. https://doi.org/10.61098/jarcis.v2i1.182

Wang X, Wang Y, Wang L (2004) Improving fuzzy c-means clustering based on feature weight learning. Pattern Recognit Lett. https://doi.org/10.1016/j.patrec.2004.03.008

Win KN, Chen J, Chen Y, Fournier-Viger P (2019) PCPD: a parallel crime pattern discovery system for large-scale spatiotemporal data based on fuzzy clustering. Int J Fuzzy Syst 21:1961–1974. https://doi.org/10.1007/s40815-019-00673-3

Xie XL, Beni IG (1991) A validity measure for fuzzy clustering. IEEE Trans Pattern Anal Mach Intell 13:841–847. https://doi.org/10.1109/34.85677

Zeng Y, Xiang K (2021) Edge oriented urban hotspot prediction for human-centric Internet of Things. IEEE Access 9:71435–71445. https://doi.org/10.1109/ACCESS.2021.3078479