
Spatial Pattern Mining Technique for Identifying Urban Traffic Congestion Hotspots Using GPS Trajectory Data

Omar Ali¹, Sara Khan²

¹Department of Computer Science, College of Computing and Information Technology, Emirates International University, Abu Dhabi, United Arab Emirates

²Department of Information Systems, College of Engineering and Information Technology, Gulf Emirates University, Dubai, United Arab Emirates
omarali2008phd@gmailcom

ABSTRACT

Autonomous mobility, route planning, emergency response, and intelligent transportation system (ITS) have been impacted by urban traffic congestion. Road-network connectivity and recurring congested relationships have been considered in existing GPS-based methods such as point-based clustering, density estimation, and spatial analysis, causing hotspot detection to be fragmented or unstable. This paper has proposed an Adaptive Spatial Trajectory Pattern Mining (ASTPM) technique to discover connected and persistent congestion hotspots from GPS trajectory data. The method has been applied to the trajectory preprocessing, segmentation, road-network map matching, and Spatio-Temporal feature extraction based on speed, density, travel-time deviation and temporal recurrence. Congested road segments have been identified using the Adaptive Congestion Intensity (ACI) model and patterns of connected congestion have been mined using Spatial Connectivity-Aware Pattern Mining (SC-PM) algorithm using spatial adjacency, co-occurrence, support and persistence. Intensity, spatial support, persistence and network connectivity have been subsequently evaluated with hotspot validation. ASTPM has been able to attain a HSI of 91.8% with 93.2% spatial consistency, 90.7% temporal persistence, and 92.1% network connectivity, thus showing stable hotspot identification. The network-aware congestion hotspot detection has been provided in the proposed framework. Future research is planned to investigate real time trajectories, multimodal data, online learning and cross city validation.

Keywords: GPS trajectory data, urban traffic congestion, spatial pattern mining, congestion hotspots, road-network connectivity, spatiotemporal analysis, trajectory mining, ASTPM, SC-PM.

1. Introduction

The problem of urban traffic congestion has emerged as a persistent challenge for modern cities, with the demand for vehicles increasing and putting significant strain on road infrastructure and transportation systems [1]. A less efficient travel environment has impacted travel efficiency, fuel use, travel time, and urban mobility, emergency response and intelligent transportation services, due to traffic congestion. The GPS trajectory data have proved to be a useful resource for the observation of vehicle movements, as they can record spatial locations, time-stamps, speeds, and travel patterns across vast road networks [2]. But raw trajectories have included positioning noise, irregular sample intervals, and diverse movement patterns [3]. Traditionally, congestion analysis has been done based on the average speed or single GPS observations leading to difficulties in identifying spatial and recurring congestion areas [4]. Thus, an identification method that combines trajectory features and spatial associations and road-network connectivity is needed to reliably identify the urban congestion hotspots [5].

The approaches that have been used so far are statistical traffic analysis, density-based clustering, kernel-based hotspot estimation, spatial autocorrelation-based approaches, and trajectory mining [6]. Spatial statistical and clustering analyses have been performed to identify dense traffic areas and to measure geographical concentration and local dependence by using clustering methods like DBSCAN [7]. Road connectivity has been added through graph-based approaches and recurring congestion patterns have been captured through temporal analysis [8], [9]. However, previous approaches have often considered GPS observations in isolation or applied static congestion thresholding, or focused on either spatial or temporal analysis [10]. These limitations have decreased their ability to identify persistent connected congestion from individual slow-moving observations. In this study this gap has been filled by proposing Adaptive Spatial Trajectory Pattern Mining (ASTPM), which integrates adaptive congestion intensity modeling with Spatial Connectivity-Aware Pattern Mining (SC-PM). The novelty has been focused on mining recurring congestion patterns directly across connected road segments, and not just clustering of nearby GPS points.

The objectives of this research are to:

- Implement a trajectory processing mechanism to clean, segment, match GPS data with maps, and extract spatiotemporal features to accurately represent the congestion on road segments.
- Work out an Adaptive Congestion Intensity model which combines speed reduction, trajectory density, travel-time deviation and temporal recurrence to determine the segments where congestion is suspected.
- A Spatial Connectivity-Aware Pattern Mining mechanism to identify patterns of recurring congestion and check for their meaningfulness as urban hotspots by applying four criteria: spatial support, intensity, persistence, and network connectivity.

In Section 2 the literature survey is discussed. The proposed ASTPM methodology and its processing stages are presented in Section 3 of this paper. Section 4 presents the experimental setup, dataset and experimental environment. The experimental results and comparative analysis are presented as seen in the Section 5. The research is summarized and future directions are provided in Section 6.

2. Literature Survey

Lyuheng Yuan et al. [11] designed a scalable framework for frequent subgraph pattern mining based on distribution, implemented parallel graph processing, accelerated pattern discovery and proved efficient big-graph mining. Yuan Gao et al. [12] studied hikers' GPS trajectories by spatial behavior pattern recognition, conducted trajectory based clustering and spatial analysis, discovered frequent trajectory structures, and showed the national park behavior characterization. For big GPS trajectories LSTM-autoencoder deep clustering was proposed by P. Zhang et al. [13] which is based on sequential representation learning and unsupervised clustering and showed effective mobility-pattern identification and enhanced parking behavior classification. Marek Lis, Maksymilian Mądział [14] created an IT-based solution for green transportation, which allowed for real-time optimization of the urban transportation network, enhanced transportation planning and included support for sustainable management of smart-city mobility.

Daniyal Hussain et al. [15] combined entropy weighting with TOPSIS to evaluate traffic congestion mitigation, applied multi-criteria decision analysis, distinguished mitigation techniques, and presented a systematic evaluation of traffic congestion management. Zhu et al. [16] developed a geomorphology-informed habitat connectivity analysis, applied spatial

restoration prioritization to semi-arid mining landscapes, calculated restoration priorities, and provided a better assessment of ecosystem connectivity. Bian et al. [17] created the spatiotemporal cluster analysis and dynamic evaluation model, applied instability-risk assessment of deep metal mining, presented the changing trends of risk patterns and showed the dynamic monitoring of rock mass. Munif et al. [18] have created network-constrained spatiotemporal analysis with predictive modeling and implemented road-accident hotspot detection in the city of Faisalabad, identified spatially connected risk patterns, and showed better urban traffic-safety assessment. Kazapoe [19] developed a probabilistic-deterministic approach for spatially varying toxic elements, applied soil-risk assessment to mining communities in Ghana, quantified health risks, and demonstrated spatially integrated exposure assessment.

A fractional ensemble learning approach for industrial-controller fault tolerance was developed by V. Rathinam et al. [20] and applied to the fault-resilient decision processing of the controller, enhanced the reliability of the controller and showed that the controller has strong fault tolerance to the industrial control failures. D. Zhang et al. [21] created an adaptive cognitive framework for data-management based on AI, had real-time task optimization in industry, and enhanced the adaptability of task processing and intelligent resource management under dynamic processing condition. Pan et al. [22] built the continuous variation detection based on intensive observation of Landsat time-series, used spatiotemporal disturbance analysis in Yuzhong mining areas, recognized the evolving disturbance patterns, and showed the detailed environmental monitoring. Dong et al. [23] constructed a reinforcement learning autonomous exploration method under connectivity graphs, carried out graph-constrained exploration decisions, enhanced connected environmental traversal and illustrated effective autonomous exploration maintaining connectivity. R. Gopi et al. [24] proposed DTLN (Deep Transformer Learning network) for strain classification with multiple wavelengths and adopted the transformer-based feature learning, enhanced strain classification, and wearable-signal classification.

Research Gap: The previous GPS-based congestion detection methods have been mainly point-based clustering, fixed thresholds and spatial or temporal analysis separately. These have failed to capture the connectivity of the road-network and the recurring congestion relationships. In addition, there are transient slowdowns that can be mistaken as hotspots due to fragmented detections. Thus, the need for adaptive spatial pattern mining framework for connected and persistent congestion hotspots still exists.

3. Proposed Methodology

Proposed Overview: Adaptive Spatial Trajectory Pattern Mining (ASTPM) is a novel approach that aims to discover the urban traffic congestion hotspots from GPS trajectory data in an integrated spatial and temporal analysis framework. The method includes raw trajectories processing, quality control, segmentation and road-network map matching. Then it generates traffic and spatial features and predicts the degree of congestion through adaptive traffic indicators. The road segments identified as congested are then given as input to Spatial Connectivity-Aware Pattern Mining (SC-PM), which finds recurring connected patterns. Finally, hotspot validation assesses spatial support, persistence, intensity and network connectivity, to generate robust urban congestion hotspot maps.

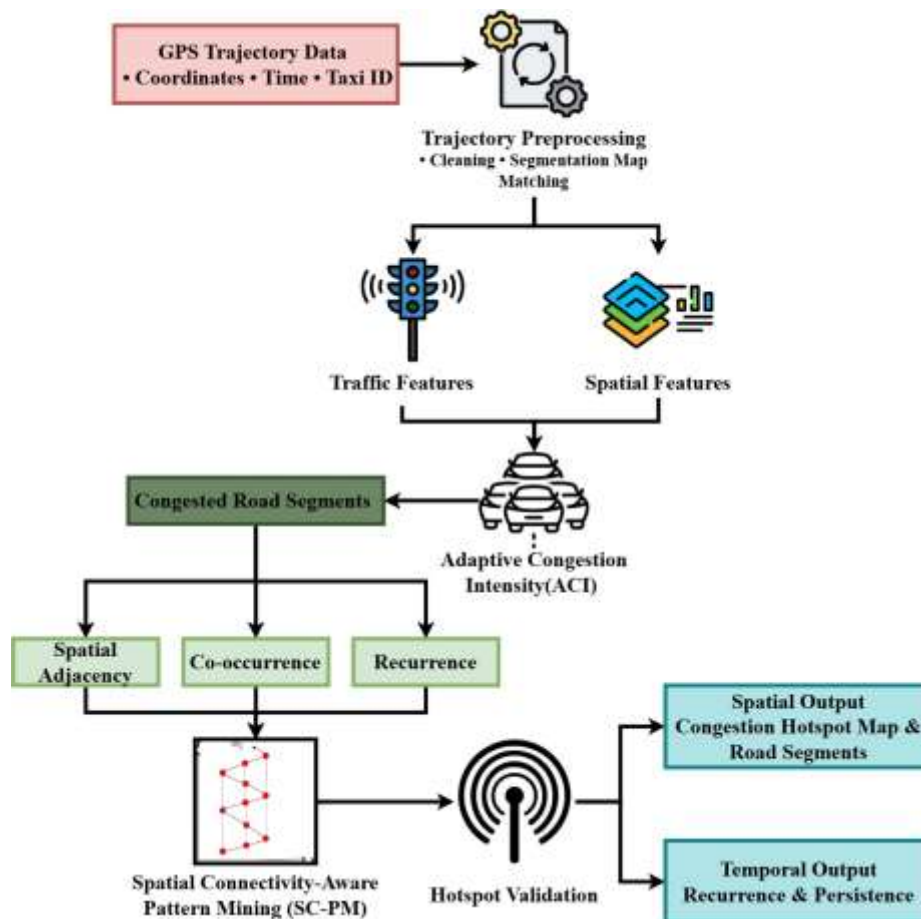


Figure 1: Overall Proposed ASTPM Framework Architecture

The proposed framework Adaptive Spatial Trajectory Pattern Mining (ASTPM) is shown in Figure 1. GPS trajectories are cleaned, segmented, and map-matched before generating traffic and spatial features. The Adaptive Congestion Intensity (ACI) model uses speed, delay, density, and recurrence to identify congested road segments. Then, SC-PM uses spatial adjacency, co-occurrence, and recurrence to identify connected congestion patterns. Finally, hotspot validation checks for intensity, support, persistence and connectivity and creates spatial hotspot maps and temporal congestion characteristics.

a. GPS Trajectory Preprocessing and Road-Network Mapping

The first stage is dedicated to preparing raw GPS trajectories so as to ensure reliable spatial analysis. Trajectory quality-control procedures are applied to the data to exclude duplicate observations, missing coordinates, abnormal time stamps and physically implausible movements. Observations are sorted by vehicle identifier and time to recreate each trip. Trajectory Segmentation - break continuous movement of vehicles into meaningful trips. Road-network map matching is then used to correlate the GPS data and road network segments based on the spatial proximity, direction of travel, and connectivity of the road network. The aim of this process is to transform the GPS observations, which are irregular, into structured road level trajectories that can be then used for congestion analysis.

Equation (1): GPS Trajectory Quality Filtering: GPS trajectories contain missing, duplicated, temporally inconsistent, and physically unrealistic observations. This equation filters such observations before spatial pattern mining so that unreliable GPS points do not propagate into congestion detection. To construct a reliable trajectory set by applying temporal consistency, maximum-movement, and validity constraints.

$$G_i^* = \{p_{i,j} \in G_i \mid \Delta t_j > 0, d(p_{i,j-1}, p_{i,j}) \leq v_{max} \Delta t_j, p_{i,j} \notin M\} \quad (1)$$

In equation (1), G_i denotes the original trajectory, G_i^* the filtered trajectory, and $p_{i,j}$ the j -th GPS observation containing latitude, longitude, and timestamp. Δt_j is the elapsed time, $d(\cdot)$ is geographic distance, v_{max} is the maximum acceptable speed, and M represents invalid or missing observations.

Equation (2): Road-Segment Speed Estimation: ASTPM operates at the road-segment level rather than directly on isolated GPS points. This equation converts sequential trajectory observations into representative traffic speeds for each segment and time interval. To estimate the average movement speed of vehicles traversing each mapped road segment.

$$v_{r,t} = \frac{1}{|E_{r,t}|} \sum_{(p_a, p_b) \in E_{r,t}} d(p_a, p_b) t_b - t_a \quad (2)$$

In equation (2), $v_{r,t}$ is the estimated speed of road segment r during interval t , while $E_{r,t}$ contains valid consecutive GPS transitions assigned to that segment. $d(p_a, p_b)$ denotes traveled distance and $t_b - t_a$ represents elapsed travel time.

b. Spatiotemporal Traffic Feature Extraction

The second stage converts map matched trajectories and outputs the traffic characteristics at the road-segment level. The speed estimation is based on sequential GPS observations, whereas the trajectory density analysis quantifies the density of vehicle movements in road segments and time windows. Travel-time analysis can be used to highlight abnormal conditions of travel, while temporal recurrence analysis can be used to capture recurring travel congestion behavior. Relationships between neighboring segments are also represented by spatial characteristics such as road adjacency and network connectivity. These additional information offers the traffic and spatial evidence for adaptive congestion intensity estimation and subsequent spatial pattern mining.

Equation (3): Normalized Spatiotemporal Congestion Features: Speed, density, travel-time deviation, and recurrence have different numerical scales and traffic interpretations. Normalization is therefore required before combining them into a unified congestion representation. To transform heterogeneous traffic indicators into comparable congestion-oriented features.

$$z_{r,t} = [1 - \frac{v_{r,t}}{v_r^{free}}, \frac{D_{r,t}}{\max(D_r)}, \frac{\tau_{r,t} - \tau_r^{free}}{\tau_r^{free}}, \frac{Q_{r,t}}{\max(Q_r)}] \quad (3)$$

In equation (3), $v_{r,t}$ and v_r^{free} denote observed and free-flow speeds, while $D_{r,t}$ represents trajectory density. $\tau_{r,t}$ and τ_r^{free} denote observed and free-flow travel times, and $Q_{r,t}$ represents temporal congestion recurrence.

c. Adaptive Congestion Intensity Estimation

The third stage is an Adaptive Congestion Intensity (ACI) mechanism to identify candidate congested road segments. The method is based on combining various indicators such as speed reduction, trajectory density and travel-time deviation and temporal recurrence over the urban network as a whole, rather than using a single fixed speed threshold. Local traffic characteristics are taken into account to help determine if the congestion is real or just normal variations or a single slow moving vehicle. The intensity representation classifies road segments based on the congestion levels and removes weak observations. Thus, only those congested segments that are well supported are passed to the spatial pattern mining stage.

Equation (4): Adaptive Congestion Intensity: Urban roads have different normal operating conditions, so fixed feature weights can produce inconsistent congestion detection. This

equation (4) adaptively weights the normalized indicators according to their temporal stability. To generate a road-segment-specific congestion intensity that integrates multiple traffic conditions.

$$ACI_{r,t} = \frac{\sum_{k=1}^4 w_{r,k} z_{r,t}^k}{\sum_{k=1}^4 w_{r,k}}, w_{r,k} = \frac{1}{\epsilon + Var_t(z_{r,t}^k)} \quad (4)$$

Here, $ACI_{r,t}$ is adaptive congestion intensity, $z_{r,t}^k$ is the k -th normalized traffic feature, and $w_{r,k}$ is its adaptive weight. Var_t measures temporal variability, while ϵ prevents numerical instability when feature variance approaches zero.

Equation (5): Adaptive Congestion-State Decision: Different Road segments exhibit different normal traffic characteristics. A segment-specific threshold allows ASTPM to identify abnormal congestion relative to each segment instead of applying one global threshold. To classify road segments as candidate congested or non-congested locations using their historical intensity distribution.

$$C_{r,t} = \begin{cases} 1, & ACI_{r,t} \geq \theta_r \\ 0, & ACI_{r,t} < \theta_r \end{cases}, \theta_r = Q_\gamma(\{ACI_{r,t}\}_{t=1}^T) \quad (5)$$

In equation (5), $C_{r,t}$ indicates whether segment r is congested at time t , while $ACI_{r,t}$ is its congestion intensity. θ_r is the adaptive segment threshold, Q_γ is the selected upper quantile, and T represents the analyzed time periods.

Algorithm 1: Adaptive Congestion Intensity Estimation

1: Input G, R, T
2: for each $r \in R$ and $t \in T$ do
3: Extract speed $S_{r,t}$, density $D_{r,t}$, delay $L_{r,t}$, recurrence $Q_{r,t}$
4: Normalize $S, D, L, Q \rightarrow [0,1]$
5: $\Delta S_{r,t} \leftarrow 1 - S_{r,t}/S_{free}^r$
6: $L_{r,t} \leftarrow \max(0, (\tau_{r,t} - \tau_{free}^r)/\tau_{free}^r)$
7: $CI_{r,t} \leftarrow w_s \Delta S + w_d D + w_l L + w_q Q$
8: if $CI_{r,t} \geq \theta_{high}$ then
9: $State_{r,t} \leftarrow Severe$
10: else if $CI_{r,t} \geq \theta_{low}$ then
11: $State_{r,t} \leftarrow Congested$
12: else
13: $State_{r,t} \leftarrow Normal$
14: end if
15: end for
16: return $C = \{(r,t) State_{r,t} \in \{Congested, Severe\}\}$

The algorithm 1 transforms the obtained speed, density, delay and recurrence values derived from GPS data to the adaptive congestion intensity values, discards normal segments, and keeps road segments with sufficiently strong congestion conditions.

d. Spatial Connectivity-Aware Pattern Mining

The fourth stage is the main contribution of ASTPM - Spatial Connectivity-Aware Pattern Mining (SC-PM). Candidate congested road segments are modeled as a road-network adjacency structure, not as independent points from GPS observations. The method is based on the observation of spatial adjacency, congestion co-occurrence, pattern support and temporal recurrence to find connected congestion patterns.

Adjacent segments that occur often are increasingly merged into “spatial patterns” while those that occur rarely and/or do not fit well are either deleted or merged with other segments. This network-aware mining process finds congestion structures which are based on connected road regions/corridors and not only GPS observations which are geographically dense.

Equation (6): Road-Network Connectivity Graph: The proposed method must distinguish physically connected congestion from geographically close but unrelated observations. This equation (6) explicitly represents congested road segments as a network graph. To construct a connectivity graph in which edges represent spatially adjacent and simultaneously congested road segments.

$$G_c = (V_c, E_c), A_{ij} = \begin{cases} 1, r_i \sim r_j, C_{i,t} C_{j,t} = 1, \\ 0, \text{otherwise} \end{cases} \quad (6)$$

Here, G_c is the congested road-network graph, V_c contains candidate congested segments, and E_c represents their connections. A_{ij} is the adjacency value, $r_i \sim r_j$ denotes spatial/network adjacency, and $C_{i,t} C_{j,t}$ indicate congestion states.

Equation (7): Spatial Connectivity-Aware Pattern Support: A connected group of congested segments should occur repeatedly rather than only once. This equation (7) measures both simultaneous occurrence and congestion intensity to identify meaningful recurring spatial patterns. To quantify the spatiotemporal support of each candidate connected congestion pattern.

$$Supp(P) = \frac{1}{T} \sum_{t=1}^T 1[\min_{r \in P} C_{r,t} = 1] \cdot \frac{1}{|P|} \sum_{r \in P} ACI_{r,t} \quad (7)$$

Here, P denotes a candidate connected congestion pattern and $Supp(P)$ measures its spatiotemporal support. $C_{r,t}$ is the congestion state, $ACI_{r,t}$ is congestion intensity, $1[\cdot]$ is the indicator function, and $|P|$ is the number of road segments in the pattern.

Equation (8): Persistent Hotspot Formation: Individual mined patterns may be fragmented even when they belong to the same congestion corridor. This equation (8) merges sufficiently supported, persistent, and connected patterns into coherent hotspots. To transform individual spatial patterns into persistent urban congestion hotspot regions.

$$H_m = \cup_{P_i \in P} P_i \text{ subject to } Supp(P_i) \geq \sigma, Pers(P_i) \geq \rho, P_i \leftrightarrow P_j \quad (8)$$

Here, H_m denotes the resulting hotspot, P is the set of mined patterns, and $Supp(P_i)$ measures pattern support. $Pers(P_i)$ represents temporal persistence, σ is the minimum support threshold, ρ is the persistence threshold, and $P_i \leftrightarrow P_j$ denotes network connectivity.

Algorithm 2: Spatial Connectivity-Aware Pattern Mining	
1:	Input C, R, σ, ρ
2:	Construct adjacency graph $A(R)$
3:	$P \leftarrow \{r \mid r \in C\}$
4:	repeat
	5: for each $p \in P$ do
6:	$N_p \leftarrow \{q \mid A(p, q) = 1 \wedge q \in C\}$
7:	$X_p \leftarrow p \cup N_p$
8:	$sup(X_p) \leftarrow occ(X_p) / T $
9:	if $sup(X_p) \geq \sigma$ then
10:	$P \leftarrow P \cup X_p$
11:	end if
12:	end for

13: until P converges
14: $H \leftarrow$ merge connected P with $\text{persistence}(P) \geq \rho$
15: $H \leftarrow \text{validate}(H, CI, \text{support}, \text{connectivity})$
16: return H

Algorithm 2 builds the graph of the road network, extracts frequent connected congestion segments based on spatial support, aggregates persistent congestion patterns, and filters out extracted patterns as meaningful urban congestion hotspots.

e. Hotspot Formation, Validation and Output Generation

The last step transforms the spatial patterns found into validated urban congestion hotspots. Adjacent patterns having similar congestion features are consolidated to create contiguous hotspot areas. Congestion intensity, spatial support, temporal persistence and network connectivity are used to evaluate each candidate hotspot. Patterns that are short-lived, incomplete or poorly supported are filtered out by Validation. The rest of the hotspots are described based on their spatial dimensions and periodic behavior and are expressed as road level hotspot maps. The final products deliver both spatial data to identify where congestion occurs and temporal data to show when congestion is occurring and if it is recurring.

4. Dataset

The Porto Taxi GPS Trajectory Dataset provides real-world taxi movements trajectories from Porto, Portugal, covering some 12 months [25]. This dataset contains the trajectories of 442 taxis where the GPS position of taxis are reported at regular 15 seconds intervals throughout each trip. Each trajectory includes a sequence of longitude and latitude positions as well as trip timestamps and identifiers. These observations allow reconstructing the movement of vehicles in the urban road network. The trajectories are cleaned, segmented and map matched to road segments, then speed, deviation in travel-time, density of trajectories and temporal recurrence are extracted for this study. These features are essential in Adaptive Spatial Trajectory Pattern Mining (ASTPM) and finding congestion hot spots.

Table 1: *Environmental Parameter Settings*

Parameter	Setting
Dataset	Porto Taxi GPS Trajectory Dataset
Study region	Porto, Portugal
Data period	July 2013 – June 2014
GPS sampling interval	15 seconds
Number of taxis	442
Coordinate system	WGS84
Spatial unit	Road segment
Temporal aggregation	Fixed time intervals
Map matching	Road-network-based
Congestion indicators	Speed, density, travel-time deviation, recurrence
Pattern mining	Spatial Connectivity-Aware Pattern Mining
Programming environment	Python
Data processing	Pandas, NumPy
Spatial processing	GeoPandas
Visualization	Matplotlib
Hardware	Multi-core CPU, ≥ 16 GB RAM
Evaluation	Spatial support, intensity, persistence, connectivity

Table 1 specifies the experimental settings that were used for the implementation and validation of ASTPM: the dataset, spatial-temporal representation, preprocessing environment, congestion indicators, mining configuration, computation tools and evaluation factors.

5. Results and Discussion

a. Precision

Precision is the percentage of identified congestion hotspots which are true congestion hotspots in the reference data. This is especially important for ASTPM, as a framework is needed that does not classify temporarily slow or normal road segments as congestion hotspots. A better precision suggests that the number of hotspot detections in the proposed spatial pattern mining process is very small in the presence of noises.

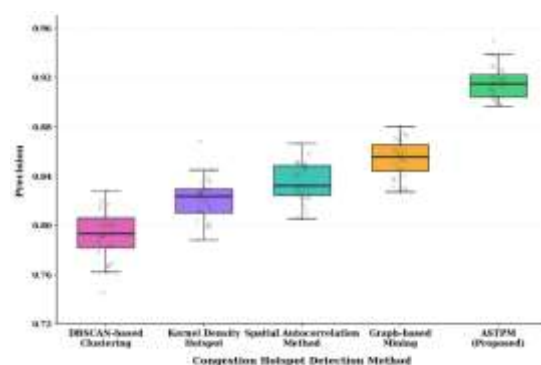


Figure 2. Precision Analysis

The precision of precision of precision of congestion hotspot detection methods is compared in Figure 2, which displays the percentage of congestion hotspots detected in relation to the correct congestion regions and thus represents the reliability of hotspot detection.

b. Recall

Recall is the ratio of the number of congestion hotspots identified by ASTPM to the number of congestion hotspots that exist in the reference. This measure determines if the proposed method can be used to identify a variety of congestion areas like persistent congestion, recurring congestion, and spatially connected hot spots. Higher recall suggests that the spatial pattern mining mechanism is effective to identify the relevant congested road segments and important congested pattern locations in the urban road network are not missed.

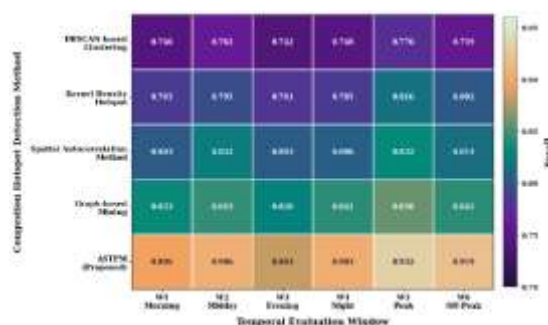


Figure 3: Recall Analysis

Figure 3 shows the recall of each method, which indicates how well each method can find congestion hotspots in the reference data and how well each method can retain the congestion patterns that are spatially distributed while not missing out important congested road segments.

c. F1-Score

F1-score is a balanced measure that combines precision and recall into one. It is suitable for ASTPM, since the identification of the hotspot needs to be accurate as well as covering enough actual hotspot regions. The higher the score the more balanced the proposed technique is in terms of false positive hot spot detection and missed congestion patterns in the road network evaluated.

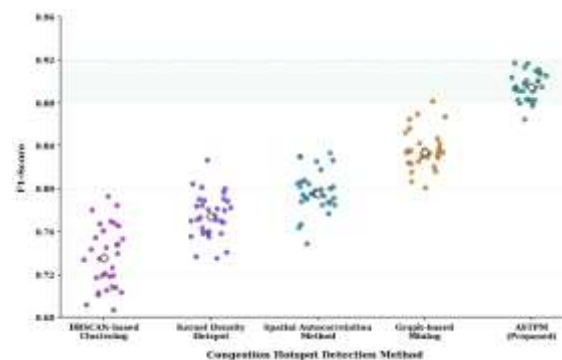


Figure 4: F1-Score Analysis

F1-scores distributions are plotted in Fig. 4, which provide an overall measure of precision and recall, or the balance between successful identification of hotspot regions and failure to detect or failure to discard congestion regions.

d. Intersection over Union (IoU)

Intersection over Union measures spatial intersection/overlap of detected congestion hotspots with reference hotspot regions. This is based on the intersection of the predicted and reference spatial areas and the union of the predicted and reference spatial areas. IoU is especially relevant to ASTPM as the method proposed identifies regions of connected road-segments as opposed to points. The greater the IoU value, the more similar the spatial correspondence between congestion hotspots detected and congestion hotspots in the reference.

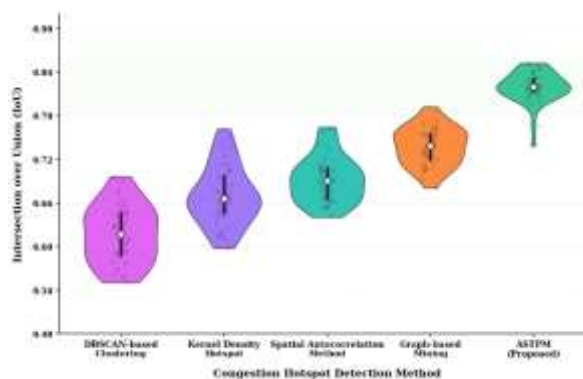


Figure 5: Intersection over Union Analysis

IoU distributions shown in Figure 5 represent the spatial overlap between the detected and reference road-network congestion hotspots to show how well each method captures the congruent congestion hotspots.

e. Hotspot Stability Index (HSI)

Hotspot Stability Index measures temporal stability of hotspots that are observed in multiple temporal observation windows. It tests the geographical and structural stability of the identified spatial patterns in analysing the trajectories of periods. This is an important ASTPM metric since the framework is designed to specifically extract recurring congestion patterns. Increased stability means that hotspots are not only picked up when a specific trajectory has some spikes of traffic but also when it has a specific traffic pattern.

Table 2: Comparison of Hotspot Stability Index (HSI)

Method	HSI (%)	Spatial Consistency (%)	Temporal Persistence (%)	Network Connectivity (%)
DBSCAN-based Clustering	72.4	74.1	70.8	68.9
Kernel Density Hotspot	75.8	78.2	72.6	70.5
Spatial Autocorrelation Method	79.6	81.4	77.9	75.3
Graph-based Mining	84.7	86.1	83.5	82.8
ASTPM (Proposed)	91.8	93.2	90.7	92.1

Table 2 shows the results of the comparisons between HSI, spatial consistency, temporal persistence and network connectivity, giving a stability based view of whether detected congestion hotspots have spatial coherence and temporal persistence.

Table 3: Ablation Study of ASTPM Components

Model Configuration	ACI	SC-PM	Hotspot Validation	Precision (%)	Recall (%)	F1-Score (%)	IoU (%)	HSI (%)
ASTPM w/o ACI	✗	✓	✓	82.6	79.8	81.2	69.4	76.8
ASTPM w/o SC-PM	✓	✗	✓	85.1	83.7	84.4	72.8	80.5
ASTPM w/o Hotspot Validation	✓	✓	✗	88.3	86.9	87.6	77.5	85.2
Full ASTPM	✓	✓	✓	91.4	89.8	90.6	82.3	91.8

Table 4: Overall Metric Comparison

Method	Precision (%)	Recall (%)	F1-Score (%)	IoU (%)	HSI (%)
DBSCAN-based Clustering	79.2	74.8	74.2	62.1	72.4
Kernel Density Hotspot	81.8	78.1	77.1	65.7	75.8
Spatial Autocorrelation Method	83.5	80.6	80.1	69.4	79.6
Graph-based Mining	85.8	83.4	83.3	73.8	84.7
ASTPM (Proposed)	91.4	89.8	90.6	82.3	91.8

Table 3 shows that if we remove ACI, SC-PM, or hotspot validation, then the detection performance is lowered. The full ASTPM setup provides the highest precision, recall, F1, IoU and hotspot stability. Table 4 provides a comparison of ASTPM and four current approaches, based upon five complementary metrics. ASTPM exhibits higher overall hotspot detection, spatial overlap, classification balance, and temporal-spatial stability when compared to evaluation criteria.

6. Conclusion

The proposed Adaptive Spatial Trajectory Pattern Mining (ASTPM) technique has provided a network-aware framework for identifying urban traffic congestion hotspots from GPS trajectory data. The method combines trajectory preprocessing, road-network map matching, spatiotemporal feature extraction, Adaptive Congestion Intensity (ACI), and Spatial Connectivity-Aware Pattern Mining (SC-PM) to recognize congestion areas by considering congestion intensity, spatial connectivity, and temporal recurrence. Experimental comparison has demonstrated that ASTPM had 91.8% Hotspot Stability Index (HSI), 93.2% spatial consistency, 90.7% temporal persistence and 92.1% network connectivity, whereas the existing approaches had less. The outcome has shown that taking into account the road-network relationships has helped to decrease fragmented detected cases and to enhance the identification of persistent congestion structures. Future work will be focused on expanding ASTPM towards real-time GPS trajectory streams for dynamically monitoring hotspots and early congestion detection. Spatial coverage will be further improved by integrating heterogeneous mobility sources like connected vehicles, public-transport trajectories, roadside sensors and traffic-camera observations. The research will also explore adaptive online learning, cross-city validation, uncertainty-aware hotspot detection and computational optimization for large-scale intelligent transportation systems in the future.

REFERENCES

- [1]. He, K., Carhart, N., Pregolato, M., & De Risi, R. (2026). Flood-induced traffic congestion and accessibility loss for urban road networks using agent-based simulation: The case study of Bristol, UK. *International Journal of Disaster Risk Reduction*, 106053.
- [2]. Garrido, R. A. (2026). Spatio-temporal patterns of urban demand: Statistical insights on hotspots and reallocation. *Spatial Statistics*, 101026.
- [3]. Huang, D., Xu, Y., Liu, W., & Yuan, Q. (2026). Assessing supply and demand dynamics of urban green spaces for recreational activity: A massive GPS trajectory data analysis. *Applied Geography*, 186, 103868.
- [4]. El Amrani, A. M., Fri, M., Benmoussa, O., & Rouky, N. (2026). Urban Freight in Casablanca: Congestion, Emissions, and Welfare Losses from Large-Scale Simulation-Based Dynamic Assignment. *Smart Cities*, 9(3), 48.
- [5]. Taran, A., Al-Khlaief, A., Bani-Khaled, H., & Al-Masaid, B. (2026). GIS Application in Urban Road Network Analysis: The Case of Panchkula City, Haryana, India. *International Journal of Geoinformatics*, 22(4), 1-16.
- [6]. Deng, C., Zhu, L., Liu, C., Zhai, H., Tian, B., Zhu, Z., ... & Zhang, C. (2026). Adaptive spatial-temporal graph learning-enabled short-term voltage stability assessment against time-varying topological conditions. *IEEE Transactions on Power Systems*.
- [7]. Oh, J., Yang, H., Raza, S. M., Ali, S. J., Kim, M., Chung, M. Y., & Choo, H. (2026). Connectivity-Aware UAV Path Planning with Continuous Action Modeling in Variable Altitude Urban Scenarios. *IEEE Transactions on Vehicular Technology*.
- [8]. Hui, J., & Cheng, Y. (2026). Trajectory-Guided Weakly Supervised Learning for Spatiotemporal Mapping of Vegetation Degradation and Restoration in Mining Areas. *Remote Sensing*, 18(16), 2734.
- [9]. Zhang, Q., Wang, H., Wen, H., Long, C., Su, L., He, X., ... & Yin, H. (2026). A survey of generative techniques for spatial-temporal data mining. *Data Science and Engineering*, 1-25.
- [10]. Roy, S., Mohammadi, A., & Roy, R. (2026). Road Traffic Accident Hotspot Detection: A GIS-Based Machine Learning Approach Using HDBSCAN and Spatial Clustering Techniques. *Geographies*, 6(2), 55.

- [11]. Yuan, L., Yan, D., Tao, D., Han, J., Adhikari, S., Long, C., & Zhou, Y. (2026). T-fsm: A scalable distributed task-based system for frequent subgraph pattern mining from a big graph. *ACM Transactions on Database Systems*, 51(2), 1-54.
- [12]. Gao, Y., Xing, T., Qu, B., Jiao, M., & Yuan, X. (2026). Spatial behavioural characteristics and pattern recognition of hikers in national parks: evidence from GPS trajectory data. *Tourism Recreation Research*, 1-17.
- [13]. Zhang, P., Stewart, K., Li, Y., & Zhu, G. (2026). LSTM-autoencoder deep clustering of big GPS trajectory data for parking behavior classification. *Journal of Intelligent Transportation Systems*, 1-21.
- [14]. Lis, M., & Mądziński, M. (2026). Green Transportation Planning for Smart Cities: Digital Twins and Real-Time Traffic Optimization in Urban Mobility Networks. *Applied Sciences*, 16(2), 678.
- [15]. Hussain, D., Jamal, A., Farooq, A., Almoshaogeh, M., Alharbi, F., & Farooq, D. (2026). Evaluation of traffic congestion mitigation techniques using an entropy-TOPSIS integrated method. *Scientific Reports*, 16(1), 5036.
- [16]. Zhu, Z., He, Y., Cao, H., Fang, S., Li, K., Du, H., & Yu, C. (2026). Geomorphology-informed habitat quality dynamics and connectivity-based restoration prioritization in semi-arid coal-mining landscapes. *Frontiers in Environmental Science*, 14, 1793022.
- [17]. Bian, Y., Zhu, W., Yan, F., & Huang, X. (2026). A Spatiotemporal Cluster Analysis and Dynamic Evaluation Model for the Rock Mass Instability Risk During Deep Mining of Metal Mine. *Mathematics*, 14(8), 1261.
- [18]. Munif, A., Khalid, S., Zameer, F., & Naqvi, S. A. A. (2026). Network-constrained spatio-temporal analysis and predictive modeling of road traffic accident hotspots for enhanced urban safety: a case study of Faisalabad, Pakistan. *Traffic Injury Prevention*, 1-9.
- [19]. Kazapoe, R. W. (2026). Spatial variability and human health risk assessment of potentially toxic elements in soils of Ghanaian mining communities: a probabilistic-deterministic framework. *Next Research*, 101474.
- [20]. V. Rathinam, R. Rajeswari, M. S. Pethuraj, and S. Baskar, "Design of a Fault Tolerance System for Industrial Controllers Utilizing Fractional Ensemble Learning," *International Journal of Information Technology & Decision Making*, 2026.
- [21]. D. Zhang, Z. Wang, T. Shi, S. Li, D. Akila, S. Baskar, and M. S. Pethuraj, "An Adaptive AI-Enabled Framework for Cognitive Data Management and Real-Time Optimization of Industrial Task Processing," *Scientific Reports*, 2026.
- [22]. Pan, J., Chen, T., Peng, L., & Plaza, A. (2026). Analyzing the Spatio-temporal Pattern of Disturbances in Yuzhong Mining Areas Using Continuous Variation Detection and Intensive Time-Series Landsat Observation. *IEEE Transactions on Geoscience and Remote Sensing*.
- [23]. Dong, Y., Lin, Z., Guan, X., Ren, S., Pan, P., & Li, W. (2026). An autonomous exploration method based on reinforcement learning under connectivity graph. *Knowledge-Based Systems*, 115457.
- [24]. R. Gopi, S. Manikandan, S. N. Dhote, S. Doss, S. Baskar, R. Sampath, et al., "DTLN: Deep Transformer Learning Network for Multi-Wavelength Strain Classification in Sports Wearables," *IEEE Sensors Journal*, 2026.
- [25]. <https://www.kaggle.com/datasets/crailtap/taxi-trajectory>