Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 257 (2025) 233–240

Procedia Computer Science

www.elsevier.com/locate/procedia

The 16th International Conference on Ambient Systems, Networks and Technologies (ANT)
April 22-24, 2025, Patras, Greece

Automating Composition of Origin-Destination Flows of Intersections Based on UAV Data

Abel Michael Betru[a], Thi M. D. Tran[a], Wim Ectors[a]*

[a]*UHasselt – Hasselt University, Transportation Research Institute (IMOB), Martelarenlaan 42, 3500 Hasselt, Belgium*

Abstract

With the exponential development rate of UAV and computer version technologies, vast and sophisticated data on traffic is now available. Spatial and temporal data, including speed and other parameters of trajectory data, can be captured. Likewise, there are unsupervised clustering algorithms in the domain of machine learning that can group data points into clusters based on their inherent similarities without using labeled data. Algorithms such as GMM, DBSCAN, and HDBSCAN identify patterns and structures within the dataset, allowing for the discovery of natural groupings. Leveraging data from UAVs and these clustering algorithms, this paper aims to develop an effective and efficient methodology to automate the extraction of Origin-Destination (OD) flows of different types of intersections. A new custom-made intersection OD flow automation method called IODF is introduced, along with the deployment of DBSCAN, HDBSCAN, and GMM algorithms for clustering intersection trajectories, leading to the automatic extraction of OD flows. The results demonstrate that all four methods performed effectively in extracting OD flow for various intersection types.

Keywords: OD flow; UAV; DBSCAN; HDBSCAN; GMM

* Corresponding author. Tel.: +32 11 26 91 14
E-mail address: wim.ectors@uhasselt.be

1877-0509
10.1016/j.procs.2025.03.032

1. Introduction

1.1. Background

Origin-destination (OD) flow analysis is essential for optimizing transportation systems, as it provides key insights into traffic movement and infrastructure needs. Traditional data collection methods are often inefficient and prone to inaccuracies, making the adoption of innovative approaches crucial. Advances in Unmanned Aerial Vehicles (UAVs) offer a promising solution, capturing real-time, high-resolution traffic data over large areas (Outay et al., 2020).), especially in locations where traditional methods are impractical or costly. Coupled with machine learning and AI, these technologies improve the accuracy of OD matrices by automating the analysis of traffic patterns (Abduljabbar et al., 2019). Unsupervised clustering algorithms, such as DBSCAN, HDBSCAN, and GMM, enhance OD flow analysis by identifying patterns and relationships within traffic data, with the choice of algorithm tailored to specific data characteristics. This paper introduces a novel methodology for automating OD flow extraction at intersections using UAV data, advanced clustering algorithms, and a custom Python script (IODF), offering an innovative approach to improve urban traffic management and planning by providing detailed insights into intersection traffic dynamics.

1.2. Problem statement

Effective urban traffic management relies on precise data regarding vehicle movements at intersections, but traditional methods such as manual counts and fixed sensors are labor-intensive and often yield incomplete or unreliable data, hindering the development of effective traffic strategies and infrastructure. The absence of affordable and flexible data collection techniques further limits the use of origin-destination (OD) data, resulting in suboptimal road designs and increased societal costs. Unmanned Aerial Vehicles (UAVs) offer a promising solution by providing high-resolution, real-time traffic data, though analyzing this data to accurately derive OD flows remains complex. Current methods, such as manual gate drawing, are error-prone, particularly in complex intersection layouts. While clustering algorithms like DBSCAN, HDBSCAN, and GMM show potential in clustering trajectories, their effectiveness in OD flow extraction at urban intersections is not tested and fully understood. This paper proposes leveraging UAV data and advanced clustering algorithms, along with a custom Python script, to automate OD flow extraction and provide more accurate insights into intersection traffic patterns and congestion, thereby improving urban traffic management.

2. Literature Review

2.1 UAV cameras and their application

UAVs have transformed traffic data collection by providing flexibility and dynamic views of traffic patterns. Unlike fixed cameras, UAVs can be repositioned for broader surveillance, offering coverage beyond the range of stationary sensors, especially in areas where installing infrastructure is impractical or too costly (Butilă & Boboc, 2022). This mobility also enhances traffic incident management and improves monitoring efficiency through technological advancements and robust regulatory frameworks. UAVs are increasingly used for infrastructure management, traffic monitoring, and road safety analysis (Outay et al., 2020), capturing vehicle movements in diverse settings, including heavy traffic and large events, where stationary sensors might fail (Jian et al., 2019; Khan et al., 2020).Their ability to analyze micro-level traffic behavior—such as interactions between pedestrians, cyclists, and vehicles at intersections or crosswalks—supports infrastructure design improvements like pedestrian walkways and bike lanes, enhancing safety(Fu et al., 2023). The integration of computer vision and machine learning algorithms further optimizes UAV data analysis, allowing for accurate identification of vehicle types and trajectories under varying conditions, driven by advancements in deep learning (Palmas & Andronico, 2022).

2.2 Unsupervised Clustering Algorithms

Unsupervised clustering algorithms are used to organize unlabeled datasets into meaningful clusters without requiring pre-assigned labels (Chander & Vijaya, 2021). These algorithms are particularly valuable in transportation, where they help analyze traffic flow by identifying patterns and grouping data points. In traffic flow analysis, unsupervised clustering is crucial for categorizing vehicles and traffic streams, especially at intersections or along corridors (Thomsen & Tomforde, 2022). Studies have shown the effectiveness of clustering algorithms in categorizing trajectories across various domains, such as civil flights, maritime traffic, and road traffic data. These algorithms excel at accurately sorting and analyzing movement patterns (Bandaragoda et al., 2019 ; Xu et al., 2022).The primary goal within this horizon is to segregate the vast amount of data recorded from UAVs into meaningful categories that reflect OD pairs and vehicle paths. To this end, several clustering algorithms, such as DBSCAN, HDBSCAN, and GMM, have been applied.

2.2.1 DBSCAN, HDBSACN and GMM

DBSCAN is a widely used algorithm that excels in identifying clusters of varying shapes and sizes while handling noise effectively (Beer et al., 2023; Singh et al., 2022).It uses two key parameters: minPts, the minimum number of points required to form a dense region, and eps (ε), which defines the proximity for clustering. While DBSCAN is effective in many cases, it struggles with datasets of varying densities, which can limit its performance in certain traffic flow applications.

HDBSCAN (Hierarchical DBSCAN) extends DBSCAN by addressing its limitations in handling clusters with varying densities. It employs a hierarchical approach, building a minimum spanning tree (MST) from distances between data points and subsequently deriving clusters from this tree (Campello et al., 2013; McInnes & Healy, 2017). This method allows for automatic cluster quantity determination and effective outlier management, making HDBSCAN a robust choice for traffic flow analysis (Campello et al., 2015). Its application has proven effective across various domains, including traffic, maritime, and air traffic data analysis (Basora et al., 2017; Ibrahim & Omair Shafiq, 2018; Wang et al., 2021).

Gaussian Mixture Models (GMM) assume that data points originate from a combination of multiple Gaussian distributions. GMM uses the Expectation-Maximization (EM) algorithm to identify clusters by iteratively determining the probability of each point belonging to a cluster and refining the Gaussian distribution parameters (Liu et al., 2010). This process continues until the optimal cluster definitions are found, making GMM a powerful method for modeling complex traffic patterns (Jiao et al., 2022; Lin et al., 2019; Liu et al., 2010). The elegant mathematics of the probability density function of GMM is widely discussed in various literature (Bouguila, N., & Fan, 2020; Jiao et al., 2022; Lin et al., 2019; Liu et al., 2010).

3. Methodology

This study employs a methodology to automate the extraction of Origin-Destination (OD) flows at intersections, utilizing UAV-collected trajectory data processed through various clustering algorithms. UAVs equipped with GPS and IMU systems capture precise geolocation, time, and motion data, ensuring stable flight and accurate maneuvering. The data collection process includes privacy protection through blurring techniques, followed by AI-driven analysis using the DataFromSky platform to track and analyze vehicle trajectories, speeds, and movement patterns. Georeferencing is applied to align the video data with real-world coordinates, ensuring spatial accuracy. The processed data undergoes an ETL (Extract, Transform, Load) procedure, where relevant information is extracted, cleaned, and structured into a CSV format, including key variables such as vehicle type, speed, trajectory ID, latitude, longitude, timestamps, and heading. This structured dataset forms the foundation for comprehensive and accurate traffic flow analysis, providing essential insights for urban traffic management and planning.

3.1. Study locations

To provide a thorough investigation, a range of intersection types were included in the intersections selected for this experiment in Limburg province of Belgium. These are a turbo roundabout in Dilsen-Stokkem, a four-legged roundabout in Lanaken, a three-legged intersection at ElfdeLiniestraat in Hasselt, and a standard four-legged intersection of Ring Hasselt (N71) and KempischeSteenweg. The study attempts to handle various traffic flow dynamics and obstacles by choosing these varied intersection types, offering a thorough assessment of the suggested OD flow automation techniques. This paper provides an analysis of the four-legged intersection and the turbo roundabout.

3.2. Data analysis

3.2.1. Peak Traffic Flow Extraction

Peak 15-minute traffic data is critical for transport planners to capture high-demand periods, enabling infrastructure design that handles peak traffic loads. Focusing on these intervals allows for efficient resource allocation and prioritized interventions, ensuring cost-effective solutions. Standardized 15-minute intervals support consistent data collection, comparisons, and the establishment of benchmarks for traffic management. Analyzing peak-period data provides insights into travel behaviors and congestion, aiding real-time management, signal optimization, and incident response. This analysis ensures transportation networks meet peak demand, improving efficiency and reducing congestion. Peak traffic flows are extracted through an iterative approach, selecting the 15-minute interval with the highest traffic flow to offer a comprehensive snapshot of the busiest conditions.

3.2.2. Data Filtering

Short trajectories, whether truncated due to the recording process or the extraction of peak data, should be filtered out as they may miss key destination points, leading to incomplete analysis. To address this, trajectories falling below the 10th percentile or the lower bound of the interquartile range (IQR) are excluded, ensuring a more complete dataset for accurate OD flow analysis.

Each trajectory typically consists of multiple spatial points, but only the first and last points are used for clustering. The trajectories are represented as:

$$T_i = \{(x_{i1}, y_{i1}, t_1), (x_{i2}, y_{i2}, t_2), (x_{i3}, y_{i3}, t_3) \ldots\ldots\ldots (x_{in}, y_{in}, t_n)\} \quad (1)$$

To ensure consistency across features, "StandardScaler" is applied to standardize the trajectory data, ensuring each feature contributes equally to clustering distance calculations. This step normalizes the data, preventing any single feature from disproportionately influencing the clustering algorithm (Pedregosa et al., 2011). Standardization improves the performance and stability of clustering algorithms such as GMM, HDBSCAN, and DBSCAN, resulting in more accurate OD flow analysis.

3.2.3. OD Flow Automation Methods

The primary technique for automating the extraction of OD flows is an in-house system called IODF, which clusters trajectories based on their initial and final positions, grouping them if the start and end points are within a specified Euclidean distance threshold, ensuring precise OD flow data.

Additionally, DBSCAN, HDBSCAN, and GMM algorithms are employed. DBSCAN clusters trajectories based on two parameters: minPts (minimum points required for a dense region) and eps (ε), the maximum distance between points considered part of the same neighborhood. While DBSCAN struggles with varying densities, it performs well in intersection data, where density is relatively consistent. HDBSCAN, on the other hand, is designed to handle varying densities and adjusts parameters such as min_cluster_size (minimum points required for a cluster) and min_samples (minimum points for a core point). GMM assumes data is generated from a mixture of Gaussian distributions and requires the number of clusters to be predefined, which is typically known from the intersection's infrastructure layout. By specifying this number, GMM can efficiently automate OD flow extraction. These clustering techniques, along with the ‘StandardScaler’ for feature normalization, are implemented using the Python scikit-learn library (Pedregosa et al., 2011). The combined use of these methods allows for precise and reliable OD flow analysis, contributing to enhanced traffic management and infrastructure planning.

4. Result

4.1. Four-legged intersection

The location selected for the experiment is the intersection of ring Hasselt (R71) and KempischeSteenweg (N74) roads, which is the main route to connect northern parts of Limburg province to the city of Hasselt, managed by a conventional traffic light system.

After recording the trajectories, a 15-minute peak hour flow was filtered for the clustering analysis. For this typical four-legged intersection, all the clustering algorithms performed well in grouping the trajectories and automating the OD flows of this intersection.

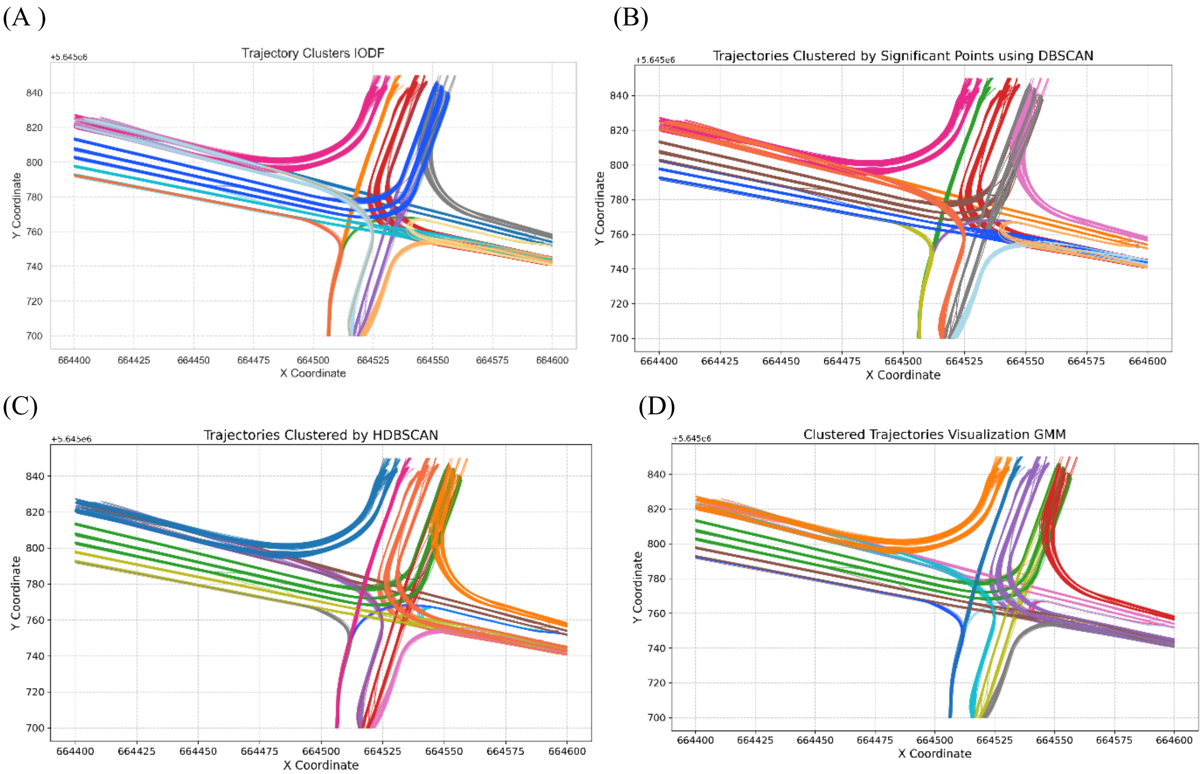

Figure 1 Clustered trajectories of ring Hasselt and KempischeSteenweg(A) IODF (B) DBSCAN (C) HDBSCAN (D) GMM

The custom IODF technique developed for this study effectively clustered trajectories at a typical-legged intersection by calculating the total length of each trajectory and filtering out shorter ones using dynamic thresholds based on interquartile range (IQR) and percentile values. Trajectories meeting the length criteria were then clustered based on the proximity of their start and end points, with each cluster visually distinguished by distinct colors to highlight traffic flow patterns. Despite the heavy traffic during data collection, some short and incomplete trajectories were grouped into a single cluster, offering insights into traffic dynamics and facilitating the inclusion of similar clusters in the analysis. Additionally, DBSCAN was applied to the dataset, performing well in handling dense traffic conditions and offering valuable insights. The clustering process was enhanced by StandardScaler, which standardized key trajectory features before DBSCAN was applied, and the results were visualized with pre-established color schemes, with noise points marked in black. The careful tuning of DBSCAN parameters, minPts and eps (ε), ensured meaningful clustering, such as identifying a U-turn trajectory as noise due to its distinct characteristics. HDBSCAN also performed effectively, with the min_cluster_size and min_samples parameters adjusted to enhance sensitivity to smaller clusters while reducing noise. The cluster_selection_epsilon parameter facilitated flexible merging of clusters, ensuring compactness and accommodating shape variations. Lastly, the Gaussian Mixture Model (GMM) was deployed, with the number of clusters estimated based on the intersection's infrastructure, which was set to 14. However, certain traffic conditions, such as truck restrictions, could complicate clustering outcomes, potentially leading to misleading results. For illustration, figure 2 shows decomposed and clustered individual trajectories by (H)DBSCAN

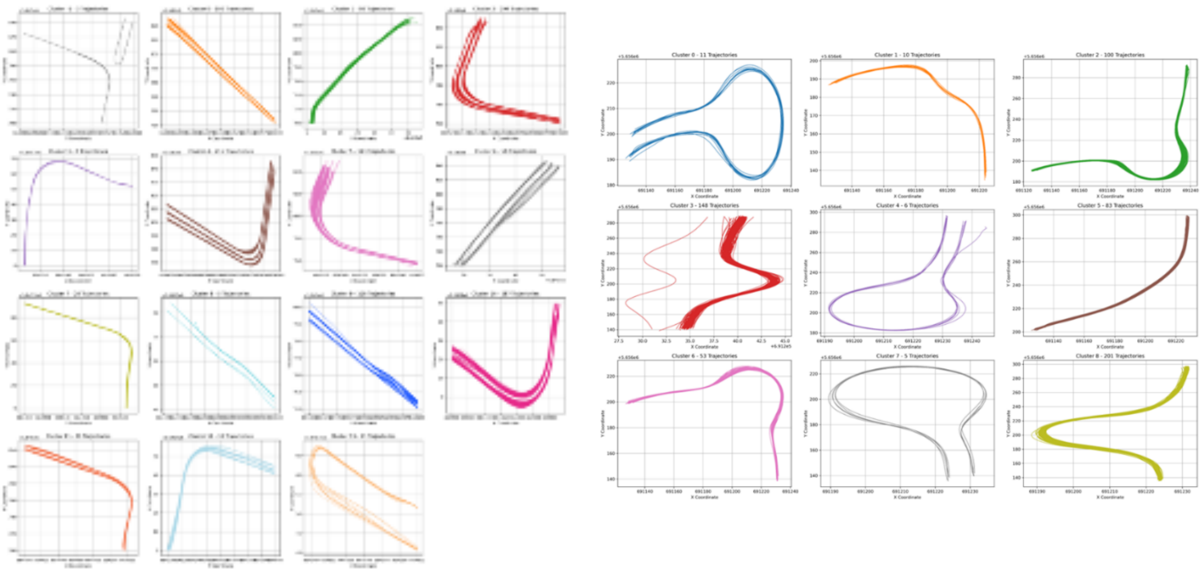

Figure 2 Decomposed and clustered individual trajectories by DBSCAN (left: four-legged intersection) and HDBSCAN (right: turbo roundabout)

4.2. Turbo Roundabout

A turbo roundabout enhances traffic flow and safety by directing vehicles into designated lanes for their exit, reducing lane changes with physical barriers and spiral markings, which minimizes collision risks and increases capacity (Guerrieri et al., 2012). A turbo roundabout at Dilsen-Stokkem, with a 15-minute peak traffic period selected for clustering analysis after recording the trajectories. The IODF method performed well overall, though it mistakenly identified a short trajectory as a separate cluster. DBSCAN corrected this by classifying the short trajectory as noise and efficiently clustering the remaining data by flow direction. HDBSCAN excelled in accurately clustering all trajectories, while GMM effectively grouped the data into nine predetermined clusters despite the challenge of specifying cluster numbers in advance. Each algorithm demonstrated strong capabilities in handling the complex traffic patterns at the turbo roundabout. While IODF, DBSCAN, and GMM were effective, HDBSCAN stood out for its precision, highlighting the adaptability and robustness of unsupervised clustering algorithms in real-world traffic analysis.

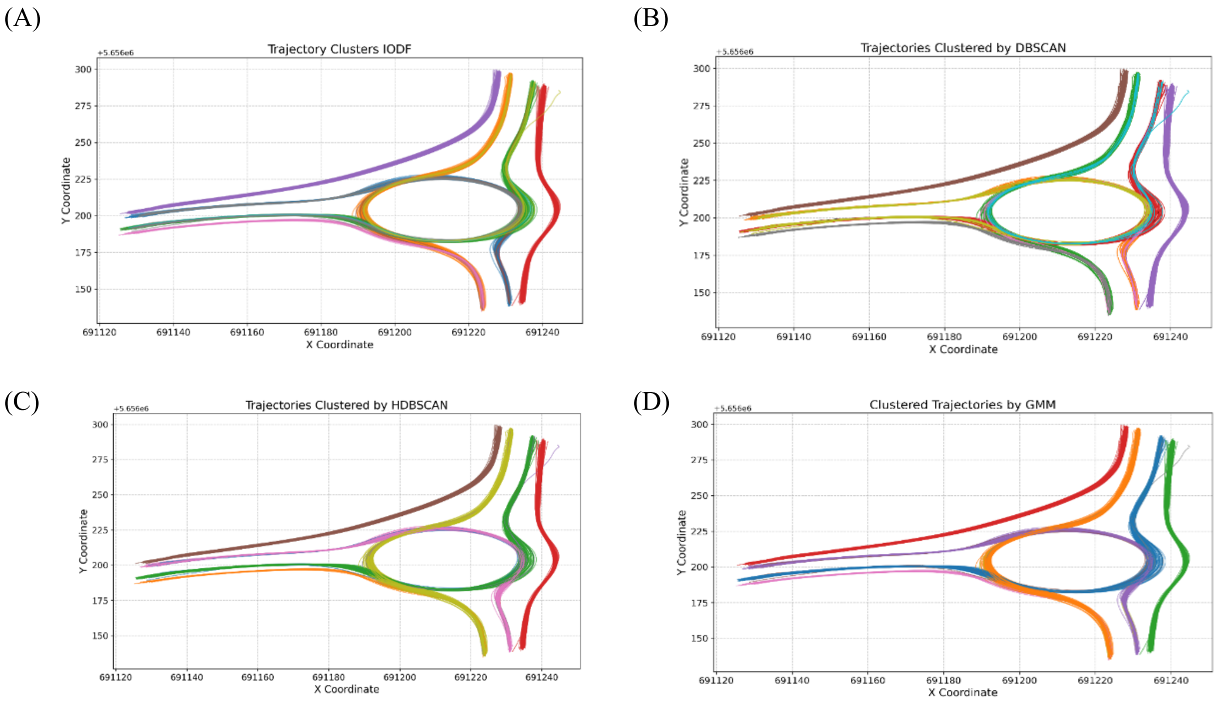

Figure 3 Clustered trajectories of turbo roundabout at Dilsen-Stokkem. (A) IODF (B) DBSCAN (C) HDBSCAN (D) GMM

5. Conclusion

OD flows at intersections are essential for transportation planning, offering critical data for designing efficient, safe, and sustainable systems. Accurate OD flow data helps optimize intersection infrastructure by identifying the need for additional lanes, traffic signals, and other control measures to enhance traffic efficiency and safety. Analyzing OD flows enables planners to forecast capacity requirements, supporting infrastructure expansion to accommodate growing traffic volumes. Additionally, OD data aids in traffic management strategies, including signal timing, signage placement, and lane configurations, and ensures public transportation services align with travel patterns. Safety improvements are also informed by OD flow analysis, identifying hazardous movements and guiding the implementation of safety measures such as turn lanes and pedestrian crossings. Optimizing traffic flow can also reduce idling, emissions, and fuel consumption, contributing to better air quality and lower carbon footprints. The methodologies presented in this paper for automating OD flow extraction using UAVs and unsupervised clustering algorithms show significant promises for advancing transportation science. Future research should focus on deploying UAV swarms for network-wide analysis, investigating driving behavior, and integrating spatial data with three-dimensional or four-dimensional models to provide deeper insights into traffic dynamics, enhancing the applicability of clustering algorithms in traffic analysis.

Acknowledgements

This paper is based on the project that has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 101037193.

References

[1] Abduljabbar, R., Dia, H., Liyanage, S., & Bagloee, S. A. (2019). Applications of artificial intelligence in transport: An overview. *Sustainability (Switzerland)*, *11*(1). https://doi.org/10.3390/su11010189

[2] Bandaragoda, T., De Silva, D., Kleyko, D., Osipov, E., Wiklund, U., & Alahakoon, D. (2019). Trajectory clustering of road traffic in urban environments using incremental machine learning in combination with hyperdimensional computing. *2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019*, *October*, 1664–1670. https://doi.org/10.1109/ITSC.2019.8917320

[3] Basora, L., Morio, J., Mailhot, C., Clustering, A. T., Air, A., & Basora, L. (2017). *Traffic Flows To cite this version : HAL Id : hal-01655747 A Trajectory Clustering Framework to Analyse Air Traffic Flows*.

[4] Beer, A., Draganov, A., Hohma, E., Jahn, P., Frey, C. M. M., & Assent, I. (2023). Connecting the Dots - Density-Connectivity Distance unifies DBSCAN, k-Center and Spectral Clustering. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 80–92. https://doi.org/10.1145/3580305.3599283

[5] Bouguila, N., & Fan, W. (Eds. . (2020). Mixtures - Models and Applications. In *The Exponential Distribution*. https://doi.org/10.1201/9780203756348-19

[6] Butilă, E. V., & Boboc, R. G. (2022). Urban Traffic Monitoring and Analysis Using Unmanned Aerial Vehicles (UAVs): A Systematic Literature Review. *Remote Sensing*, *14*(3). https://doi.org/10.3390/rs14030620

[7] Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7819 LNAI*(PART 2), 160–172. https://doi.org/10.1007/978-3-642-37456-2_14

[8] Campello, R. J. G. B., Moulavi, D., Zimek, A., & Sander, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data*, *10*(1), 1–51. https://doi.org/10.1145/2733381

[9] Chander, S., & Vijaya, P. (2021). Unsupervised learning methods for data clustering. In *Artificial Intelligence in Data Mining: Theories and Applications*. INC. https://doi.org/10.1016/B978-0-12-820601-0.00002-1

[10] Fu, T., Xie, S., Li, R., Wang, J., Zhang, L., Sobhani, A., & Fang, S. (2023). Investigating Pedestrian Crossing Patterns at Crossing Locations Based on Trajectory Data Collected by UAV. *Journal of Advanced Transportation*, *2023*(1). https://doi.org/10.1155/2023/5529970

[11] Guerrieri, M., Ticali, D., & Corriere, F. (2012). Turbo roundabouts: geometric design parameters and performance analysis. *GSTF Journal on Computing (JoC)*, *2*(1), 227–232.

[12] Ibrahim, R., & Omair Shafiq, M. (2018). Mining trajectory data and identifying patterns for taxi movement trips. *2018 13th International Conference on Digital Information Management, ICDIM 2018*, 130–135. https://doi.org/10.1109/ICDIM.2018.8847135

[13] Jian, L., Li, Z., Yang, X., Wu, W., Ahmad, A., & Jeon, G. (2019). Combining Unmanned Aerial Vehicles with Artificial-Intelligence Technology for Traffic-Congestion Recognition: Electronic Eyes in the Skies to Spot Clogged Roads. *IEEE Consumer Electronics Magazine*, *8*(3), 81–86. https://doi.org/10.1109/MCE.2019.2892286

[14] Jiao, L., Denœux, T., Liu, Z. ga, & Pan, Q. (2022). EGMM: An evidential version of the Gaussian mixture model for clustering. *Applied Soft Computing*, *129*, 109619. https://doi.org/10.1016/j.asoc.2022.109619

[15] Khan, N. A., Jhanjhi, N. Z., Brohi, S. N., Usmani, R. S. A., & Nayyar, A. (2020). Smart traffic monitoring system using Unmanned Aerial Vehicles (UAVs). *Computer Communications*, *157*(January), 434–443. https://doi.org/10.1016/j.comcom.2020.04.049

[16] Lin, X., Yang, X., & Li, Y. (2019). A Deep Clustering Algorithm based on Gaussian Mixture Model. *Journal of Physics: Conference Series*, *1302*(3). https://doi.org/10.1088/1742-6596/1302/3/032012

[17] Liu, J., Cai, D., & He, X. (2010). Gaussian Mixture Model with Local Consistency. *Proceedings of the 24th AAAI Conference on Artificial Intelligence, AAAI 2010*, 512–517. https://doi.org/10.1609/aaai.v24i1.7659

[18] McInnes, L., & Healy, J. (2017). Accelerated Hierarchical Density Based Clustering. *IEEE International Conference on Data Mining Workshops, ICDMW*, *2017-Novem*, 33–42. https://doi.org/10.1109/ICDMW.2017.12

[19] Outay, F., Mengash, H. A., & Adnan, M. (2020). Applications of unmanned aerial vehicle (UAV) in road safety, traffic and highway infrastructure management: Recent advances and challenges. *Transportation Research Part A: Policy and Practice*, *141*(October 2020), 116–129. https://doi.org/10.1016/j.tra.2020.09.018

[20] Palmas, A., & Andronico, P. (2022). *Deep Learning Computer Vision Algorithms for Real-time UAVs On-board Camera Image Processing*. http://arxiv.org/abs/2211.01037

[21] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

[22] Singh, H. V., Girdhar, A., & Dahiya, S. (2022). A Literature survey based on DBSCAN algorithms. *Proceedings - 2022 6th International Conference on Intelligent Computing and Control Systems, ICICCS 2022*, *Iciccs*, 751–758. https://doi.org/10.1109/ICICCS53718.2022.9788440

[23] Thomsen, I., & Tomforde, S. (2022). Intersection-centric Urban Traffic Flow Clustering for Incident Detection in Organic Traffic Control. *International Conference on Vehicle Technology and Intelligent Transport Systems, VEHITS - Proceedings*, *Vehits*, 410–417. https://doi.org/10.5220/0011085400003191

[24] Wang, L., Chen, P., Chen, L., & Mou, J. (2021). Ship ais trajectory clustering: An hdbscan-based approach. *Journal of Marine Science and Engineering*, *9*(6). https://doi.org/10.3390/jmse9060566

[25] Xu, R., Liu, R. W., Nie, J., Deng, X., Xiong, Z., Jiang, W., Yu, H., & Niyato, D. (2022). AIS Data-Based Efficient Ship Trajectory Clustering. *Proceedings - 20th IEEE International Symposium on Parallel and Distributed Processing with Applications, 12th IEEE International Conference on Big Data and Cloud Computing, 12th IEEE International Conference on Sustainable Computing and Communications An*, 581–588. https://doi.org/10.1109/ISPA-BDCloud-SocialCom-SustainCom57177.2022.00080