Validating Cyclist Platoon Detection Using Fuzzy C-Means and the DLR-UT Trajectory Dataset

Sarah Salem[1] and Axel Leonhardt[1]

[1]Institute of Transport and Spatial Planning, University of the Bundeswehr Munich, Germany

Short summary

This study presents a trajectory-based validation framework for cyclist platoon detection using high-resolution data from the DLR Urban Traffic Dataset. Cyclist trajectories are analyzed to capture arrival patterns at an urban signalized intersection. A fuzzy C-means (FCM) clustering algorithm is applied to identify cyclist platoons based on spatial and temporal proximity.
To enable quantitative evaluation, a reference dataset of cyclist platoons is constructed using deterministic geometric and temporal criteria derived from upstream crossing events. The detected clusters are compared with the reference platoons to assess the accuracy and robustness of the clustering approach.
The results provide insights into cyclist group formation and highlight the relationship between upstream spatial clustering and platoon formation at the intersection. The study establishes a practical framework for validating clustering-based platoon detection using empirical trajectory data.
Keywords: cyclist platoon detection; trajectory data; clustering; validation framework; urban traffic

1 Introduction

Cycling plays an increasingly important role in urban transport systems due to its environmental and societal benefits. However, signalized intersections remain a major source of delay and discomfort for cyclists. Previous studies show that cyclists overestimate waiting time at signals, making the reduction of stops a key objective in intersection control design Börjesson & Eliasson (2012); Fioreze (2019).
Existing cyclist-priority strategies typically operate at the level of individual cyclists. In contrast, cyclists often travel in small groups or platoons, which can be utilized for more efficient traffic management. While platoon detection is well established in vehicular traffic, methods for identifying cyclist platoons remain limited due to the higher variability in speed, spacing, and lateral behavior Dong et al. (2021); Li et al. (2023).
Recent advances in trajectory data collection enable detailed analysis of cyclist movements. Clustering methods provide a promising approach to identify groups without relying on fixed headway thresholds Xu & Wunsch (2005). In particular, fuzzy clustering allows gradual group membership, which is suitable for representing the continuous nature of cyclist interactions Bezdek et al. (1999).
However, a key challenge is the validation of clustering-based platoon detection. Many existing studies rely on heuristic or simulated definitions of platoons, limiting the ability to assess detection accuracy using empirical data.
This study addresses this gap by proposing a trajectory-based validation framework for cyclist platoon detection. The main contributions are:

- the development of two complementary reference definitions capturing upstream spatial clustering and intersection-based platoon formation,
- the application of a fuzzy C-means clustering approach for detecting cyclist groups,
- and a quantitative evaluation of clustering performance relative to empirically derived platoon structures.

2 Methodology

The methodology consists of three main components. First, cyclist trajectories are extracted and preprocessed from the DLR Urban Traffic Dataset. Second, reference platoon datasets are constructed using deterministic geometric and temporal rules derived from trajectory information. Finally, a fuzzy C-means (FCM) clustering algorithm is applied to detect cyclist groups upstream of the intersection, and the detected clusters are compared with the reference datasets.

Trajectory Data and Preprocessing

The trajectory data used in this study come from the DLR Urban Traffic Dataset (DLR-UT), collected at the AIM research intersection in Braunschweig, Germany Schicktanz et al. (2025). The dataset contains high-resolution road-user trajectories derived from synchronized video recordings using a multi-sensor system consisting of stereo cameras and infrared illumination. Moving objects are detected using optical-flow–based motion analysis and subsequently tracked to generate trajectories containing spatial coordinates, timestamps, and classification probabilities for different road-user types.
The raw dataset consists of multiple trajectory files corresponding to consecutive 15-minute observation intervals. These files were concatenated to form a continuous dataset of cyclist trajectories. Each trajectory record contains the following attributes:

- unique trajectory identifier,
- timestamp,
- spatial coordinates in UTM projection (Easting and Northing),
- classification probabilities for road-user types.

To isolate cyclist trajectories, observations were filtered using the probabilistic classification scores provided in the dataset. A trajectory point was classified as a cyclist when the probability associated with the cyclist class exceeded the probabilities of all other classes. This filtering step ensured that only trajectories with a high likelihood of representing cyclists were retained. After processing, the dataset contained 183,463 trajectory points representing 1,067 individual cyclists. The spatial coverage of the trajectory data extends approximately from $-38.6\,\mathrm{m}$ upstream to $44.3\,\mathrm{m}$ downstream relative to the intersection center. This coverage provides sufficient observation range to analyze cyclist behavior both before and after the intersection crossing.

Reference Platoon Definitions

To evaluate the clustering-based platoon detection algorithm, two deterministic reference definitions of cyclist platoons were constructed from the trajectory data. These definitions capture different aspects of cyclist group behavior along the intersection approaches.

Crossing-Based Reference

The first reference definition identifies cyclist platoons based on upstream crossing events near the intersection. For each trajectory, a representative movement direction was estimated using the displacement between the 20% and 80% temporal quantiles of trajectory points within a $60\,\mathrm{m}$ radius of the intersection center. This reduces the influence of local curvature and measurement noise. Cyclists were assigned to intersection approaches using predefined geometry-based reference directions derived from upstream and stop-line coordinates. Each trajectory was assigned to the approach with the highest directional similarity, computed via the dot product between trajectory and reference vectors.
Within each approach, trajectories were projected onto a longitudinal axis aligned with the mean movement direction. The axis orientation was adjusted to ensure that longitudinal coordinates increase toward the intersection. Cyclist crossing events were defined as the first intersection of a trajectory with a virtual detection line located D_{upstream} meters upstream of the reference point. Crossing times were estimated using linear interpolation between trajectory samples.
Platoons were constructed by grouping consecutive cyclists based on a critical headway threshold H_{crit}. Cyclists with headways below this threshold were assigned to the same platoon, while larger gaps initiated new platoons. A sensitivity analysis was conducted for $D_{\mathrm{upstream}} \in \{10, 12.5, 15, 17.5, 20\}\,\mathrm{m}$, and the final dataset was generated using $D_{\mathrm{upstream}} = 12.5\,\mathrm{m}$.

This procedure provides a deterministic ground-truth representation of cyclist platoons for validating clustering-based detection methods.

Upstream Spatial–Temporal Reference

While the crossing-based reference captures platoon formation at the intersection, cyclist groups may already form upstream before reaching the stop line. To capture this behavior, a second reference definition was developed based on spatial and temporal proximity upstream of the intersection.
Cyclists were considered within an upstream detection region defined by

$$-30\,\mathrm{m} \leq s_{\mathrm{long}} \leq -2.5\,\mathrm{m}.$$

Within this region, cyclists were grouped at each time step using two criteria:

- small longitudinal distance between consecutive cyclists,
- similar cyclist speeds.

Only groups containing at least two cyclists and persisting over consecutive time steps were considered valid upstream platoons. This ensures that detected groups represent stable and cohesive cyclist clusters rather than short-term fluctuations.
The two reference definitions capture different stages of cyclist grouping: the upstream spatial clustering of cyclists before the intersection, and the temporal formation of platoons at the stop line.
This approach identifies cyclists who ride close together upstream and maintain similar speeds over time, representing naturally formed groups before reaching the intersection.

Fuzzy Clustering for Cyclist Platoon Detection

To identify cyclist platoons upstream of the intersection, a fuzzy C-means (FCM) clustering algorithm was applied to the trajectory data, building on the clustering approach proposed in (Salem & Leonhardt, 2026). Unlike hard clustering methods, fuzzy clustering allows each cyclist to belong to multiple clusters with varying membership degrees, enabling a more flexible representation of group boundaries. The FCM algorithm was configured with a fixed number of clusters and standard parameter settings, and cyclists were assigned to clusters based on maximum membership values.
For each cyclist approaching the intersection, feature vectors were constructed using trajectory-derived attributes including

- projected longitudinal position relative to the intersection,
- cyclist speed,
- temporal proximity within the upstream detection region.

Cyclists were assigned to clusters according to their highest membership value. Clusters containing at least two cyclists and satisfying spatial cohesion criteria were interpreted as candidate cyclist platoons. For each detected cluster, attributes such as platoon size, mean speed, and estimated arrival time at the intersection were computed.

Validation Framework

The performance of the fuzzy clustering algorithm was evaluated by comparing detected cyclist clusters with the two deterministic reference datasets described above.
Evaluation was conducted at the event level by comparing detected cluster intervals with reference platoon intervals. A detection was considered correct when the predicted cluster overlapped temporally with a reference platoon on the same approach. This interval-overlap criterion accounts for the spatial offset between upstream cluster detection and intersection crossing events. The evaluation focuses on consistency between detected clusters and empirically observed grouping behavior rather than strict classification accuracy.
In addition to event matching, the comparison focuses on differences in the number of detected groups and their characteristics relative to the reference definitions. This includes the number of clusters, group sizes, and temporal alignment with observed platoon formations.

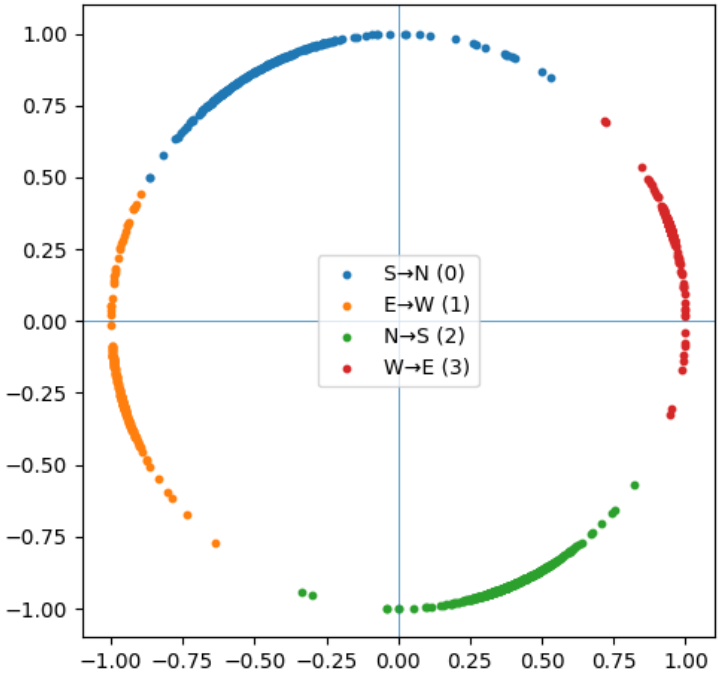


Figure 1: Per-track direction vectors near the intersection projected onto the unit circle. Each color corresponds to one intersection approach derived from geometry-based reference directions.

Furthermore, an arrival-time prediction analysis was performed to assess the temporal consistency of the detected clusters. The time difference between predicted and observed arrivals at the intersection was used to evaluate the short-term predictive capability of the clustering approach.

3 Results and Discussion

Trajectory Dataset and Spatial Coverage

The final trajectory dataset contains 1,067 cyclist trajectories comprising 183,463 trajectory points. After preprocessing and classification filtering, cyclist movements were analyzed across the four intersection approaches. The spatial extent of the trajectories relative to the intersection center varies across approaches. The observed longitudinal coverage ranges from approximately $-38.6\,\mathrm{m}$ upstream to $44.3\,\mathrm{m}$ downstream. This coverage allows the observation of cyclist behavior both before and after reaching the intersection.
The upstream region used for clustering analysis was defined within the interval $-30\,\mathrm{m} \leq s_{\mathrm{long}} \leq -2.5\,\mathrm{m}$, ensuring that cyclists were still approaching the intersection while remaining within the reliable observation range. Figure 1 illustrates the distribution of trajectory direction vectors used to assign cyclists to intersection approaches.

Reference Definitions of Cyclist Platoons

Two deterministic reference definitions were constructed to characterize cyclist grouping behavior at different spatial locations along the intersection approaches.

Crossing-Based Reference

Cyclist crossing events were extracted at a detection line located $12.5\,\mathrm{m}$ upstream of the intersection. Using a critical headway threshold of $H_{\mathrm{crit}} = 1.5\,\mathrm{s}$ and excluding *start_ inside* trajectories, 17 crossing-based platoons were identified. The mean platoon size was 2.12 cyclists, with a maximum observed size of four cyclists. The headway threshold was reduced from $2.5\,\mathrm{s}$ to $1.5\,\mathrm{s}$ to avoid grouping loosely spaced cyclists into the same platoon. This value better reflects closely spaced cyclist arrivals observed in the trajectory data. The selected threshold reflects observed cyclist

Table 1: **Summary statistics of the crossing-based reference dataset.**

Metric	Value
Total cyclist trajectories	1067
Valid upstream crossings	447
Detected platoons	17
Mean platoon size	2.12 cyclists
Maximum platoon size	4 cyclists
Critical headway threshold	1.5 s
Upstream detection distance	12.5 m

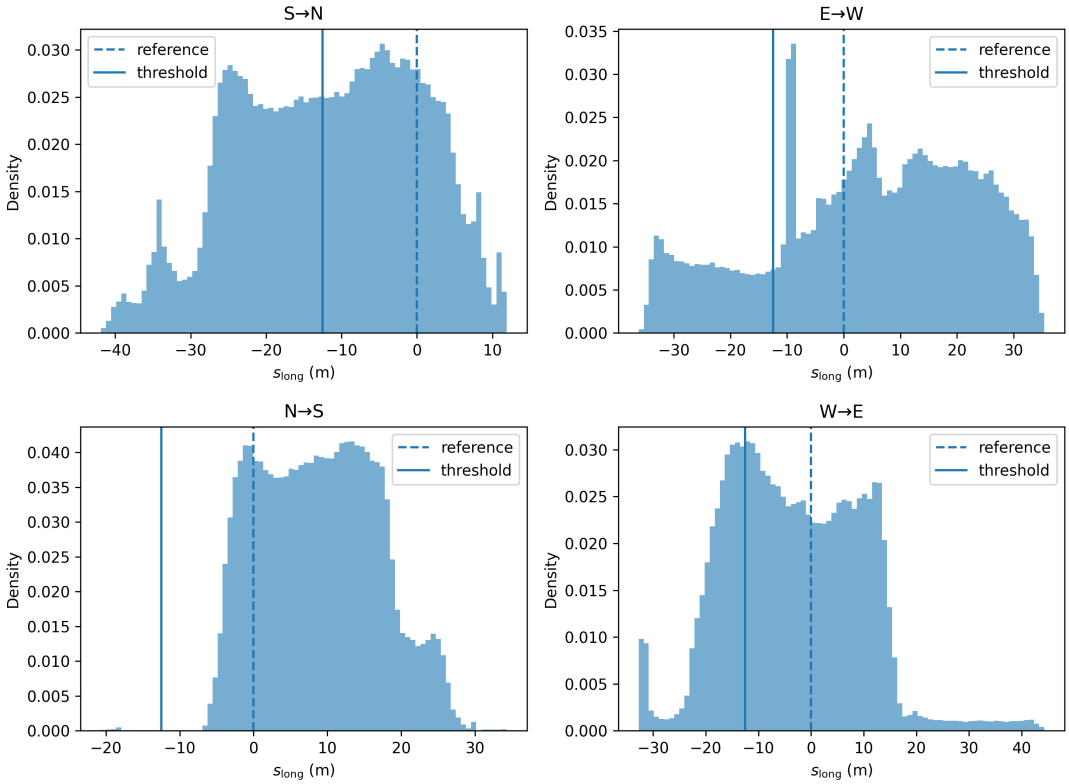


Figure 2: **Distribution of projected longitudinal coordinate s_{long} for each approach.**

arrival patterns and is not tuned to the clustering results. Table 1 summarizes the characteristics of this reference dataset. Figure 2 shows the distribution of projected longitudinal coordinates used to define crossing events. The dashed line indicates the intersection reference point ($s_{\text{long}} = 0$), while the solid line represents the upstream detection threshold ($s_{\text{long}} = -12.5\,\text{m}$) used to define cyclist crossing events and platoon formation.

The distribution of headways between consecutive cyclist crossings is shown in Fig. 3. The dashed line indicates the critical headway threshold used for platoon formation. Headways below this threshold correspond to cyclists belonging to the same platoon. A substantial proportion of observed headways fall below the 1.5 s threshold, indicating the presence of closely spaced cyclist arrivals that form platoons. Larger headways correspond to temporal gaps between groups, separating consecutive platoons.

The resulting platoon size distribution is presented in Fig. 4. Most detected platoons consist of two cyclists, indicating that cyclist groups are typically small and loosely formed. Larger platoons occur less frequently, reflecting the stochastic nature of cyclist arrivals.

Upstream Spatial–Temporal Reference

To capture cyclist grouping behavior before reaching the intersection, a second reference dataset was constructed based on spatial and temporal proximity within the upstream detection region.

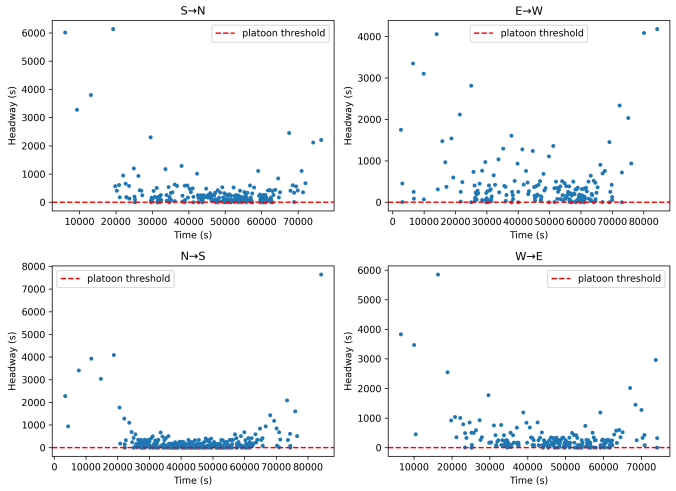


Figure 3: **Headway distribution between consecutive cyclist crossings.**

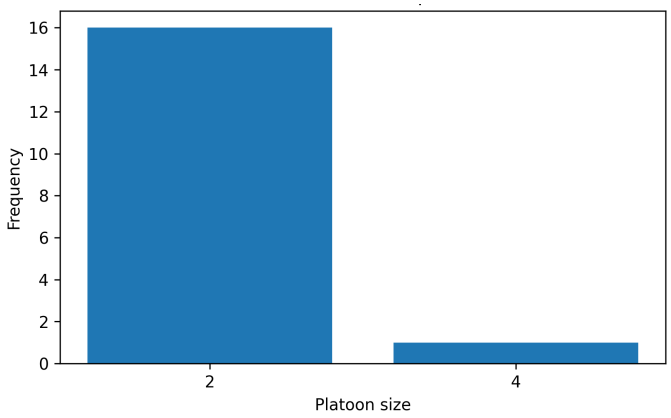


Figure 4: **Distribution of detected cyclist platoon sizes.**

Using deterministic rules based on inter-cyclist spacing and speed similarity, a total of 48 upstream clusters were identified. The mean cluster size was 2.06 cyclists, with a maximum size of three cyclists, indicating that upstream cyclist groups are generally small and loosely formed.
Approach-level results are summarized in Table 2. The distribution of clusters varies across approaches, reflecting differences in cyclist demand and trajectory coverage. Together, the two reference definitions capture complementary stages of cyclist grouping: spatial clustering upstream of the intersection and temporal platoon formation at the stop line.

FCM Detection and Comparison with Reference Definitions

The fuzzy C-means clustering algorithm detected 17 cyclist clusters within the upstream observation region. These clusters correspond to spatially cohesive groups of cyclists approaching the intersection. Compared with the upstream spatial–temporal reference, which identifies 48 clusters, the clustering algorithm detects a subset of the most cohesive upstream groups. This indicates that the FCM method primarily captures the most compact and cohesive cyclist formations, characterized by small inter-cyclist spacing and similar speeds.
When compared with the crossing-based reference, the clustering results show moderate agreement. Analysis of space–time trajectories indicates that cyclist groups tend to form approximately 15–25 m upstream of the intersection. As cyclists approach the stop line, these groups may expand or disperse due to speed variability and interactions with surrounding traffic. As a result, only a portion of the upstream clusters remain sufficiently compact to satisfy the temporal headway criterion used in the crossing-based platoon definition. This explains the difference between the number of upstream clusters (48) and the number of detected clusters (17).
These results suggest that upstream spatial clustering and intersection-based platoon formation represent related but distinct stages of cyclist group dynamics. This indicates that cyclist grouping

Table 2: **Approach-level statistics of upstream cyclist clusters.**

Approach	Cyclist trajectories	Upstream clusters
S→N	221	27
E→W	198	2
N→S	430	3
W→E	218	16
Total	1067	48

is dynamic, with upstream spatial clusters evolving into, but not always resulting in, compact platoons at the intersection.

Arrival Time Prediction and Implications

To assess the temporal consistency of the detected clusters, an arrival-time prediction analysis was conducted. For each detected cluster, the expected arrival time at the intersection was estimated using the remaining distance to the stop line and an assumed average cyclist speed of 5 m/s. Among the 17 detected clusters, 14 corresponded to a crossing-based platoon within a matching window of ±30 s. For these matched cases, the mean prediction error was 3.11 s, the median error was 3.28 s, and the RMSE was 3.84 s. Additionally, 42.9% of predictions fell within ±3 s of the observed arrival time.

These results indicate that the clustering algorithm provides moderately accurate short-term predictions of cyclist arrivals. Even when upstream clusters partially disperse before reaching the intersection, they still represent groups of cyclists arriving within a limited time window. From an operational perspective, this early detection provides anticipatory information on cyclist arrivals. Such information could support future development of adaptive control strategies, for example by enabling earlier responses to approaching cyclist groups. The results suggest that upstream fuzzy clustering captures meaningful temporal structure in cyclist flows and provides short-term arrival predictions that may be useful for real-time traffic management applications.

4 Conclusion

This study presented a validation framework for cyclist platoon detection based on high-resolution trajectory data. By combining fuzzy C-means clustering with two complementary reference definitions: spatial clustering upstream and temporal platoon formation at the intersection, the analysis provides a detailed assessment of cyclist group dynamics and detection performance.

The results show that fuzzy clustering effectively identifies cohesive cyclist groups upstream and captures the dominant structure of cyclist arrivals. However, not all upstream clusters persist as compact platoons at the intersection, highlighting the distinction between spatial grouping and temporal platoon formation. This confirms that cyclist grouping is a dynamic process, with upstream clusters evolving into, but not always resulting in, compact platoons at the stop line.

The proposed framework enables systematic evaluation of clustering-based platoon detection methods and demonstrates the value of trajectory data for understanding cyclist flow behavior. In addition, the clustering approach provides short-term arrival predictions with moderate accuracy, indicating its potential for anticipatory traffic management applications. Also, the analysis is limited by trajectory coverage and potential classification uncertainties in the dataset, which may affect the detection of loosely formed cyclist groups.

Future work should explore alternative clustering methods and incorporate additional contextual factors such as signal states and interactions with other road users. Extending the framework to real-time applications remains a promising direction for cyclist-oriented traffic management.

Acknowledgements

This research is part of the project MORE – Munich Mobility Research Campus. The project is funded by dtec.bw – Digitalization and Technology Research Center of the Bundeswehr. dtec.bw is funded by the European Union – NextGenerationEU.

References

Bezdek, J. C., Keller, J., Krisnapuram, R., & Pal, N. R. (1999). *Fuzzy models and algorithms for pattern recognition and image processing.* New York: Springer.

Börjesson, M., & Eliasson, J. (2012). The value of time and external benefits in bicycle appraisal. *Transportation Research Part A: Policy and Practice*, *46*(4), 673–683. doi: 10.1016/j.tra.2012.01.006

Dong, F., Zhao, X., & Chen, Y.-H. (2021). Optimal longitudinal control for vehicular platoon systems: Adaptiveness, determinacy, and fuzzy. *IEEE Transactions on Fuzzy Systems*, *29*(4), 889–903. doi: 10.1109/TFUZZ.2020.2966176

Fioreze, T. (2019). Perceived waiting time versus actual waiting time: A case study among cyclists in enschede, the netherlands. In (Vol. 41). (Presented at the 8th International Cycling Safety Conference (ICSC), 2019)

Li, J., Zhao, P., & Chen, Y. (2023). A review of platoon-based traffic signal control strategies in urban networks. *Transportation Research Part C: Emerging Technologies*, *146*, 103896. doi: 10.1016/j.trc.2023.103896

Salem, S., & Leonhardt, A. (2026). Real-time cyclist platoon detection using fuzzy clustering. In *Transportation research board (trb) 105th annual meeting.* Washington, DC, USA. (Poster presentation)

Schicktanz, C., Klitzke, L., Gimm, K., Rizzo, G., Liesner, K., Mosebach, H. H., & Knake-Langhorst, S. (2025). The dlr urban traffic dataset (dlr-ut): A comprehensive traffic dataset from an urban research intersection. In *2025 ieee intelligent vehicles symposium (iv)* (p. 398-405). doi: 10.1109/IV64158.2025.11097400

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, *16*(3), 645-678. doi: 10.1109/TNN.2005.845141