Unsupervised Clustering of Smartphone Sensor Data: Structure Discovery and Post-Hoc Transport-Mode Analysis

Monika Stoilkovska
kti1122022@feit.ukim.edu.mk
Faculty of Electrical Engineering and Information Technologies
Ss. Cyril and Methodius University
Skopje, North Macedonia

Angela Nastovska
kti982022@feit.ukim.edu.mk
Faculty of Electrical Engineering and Information Technologies
Ss. Cyril and Methodius University
Skopje, North Macedonia

Rezearta Ljuzha
kti902022@feit.ukim.edu.mk
Faculty of Electrical Engineering and Information Technologies
Ss. Cyril and Methodius University
Skopje, North Macedonia

Emilija Kizevska
ekizevska@feit.ukim.edu.mk
Faculty of Electrical Engineering and Information Technologies
Ss. Cyril and Methodius University
Skopje, North Macedonia

Hristijan Gjoreski
hristijang@feit.ukim.edu.mk
Faculty of Electrical Engineering and Information Technologies
Ss. Cyril and Methodius University
Skopje, North Macedonia

Abstract

Most transport-mode recognition systems depend on costly transport-class labels, leaving unclear what structure is discoverable without them. We analyze the Sussex–Huawei Locomotion and Transportation (SHL) Preview v2 dataset independently at four phone positions. Multimodal smartphone signals are robustly scaled and segmented into overlapping 10 s windows; 135 extracted features are pruned and reduced with Principal Component Analysis (PCA). K-Means, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Ward agglomerative clustering, and a Gaussian Mixture Model (GMM) are fitted without transport-class labels; the Null annotation is used only to remove invalid or unannotated periods. Internal evaluation selects two K-Means clusters at every phone position and reaches a silhouette score of 0.730 for hips. Withheld transport-class labels are introduced only for post-hoc evaluation using the Adjusted Rand Index (ARI), whose best value is 0.193 for eight-cluster K-Means on torso data. The discovered structure therefore does not directly reproduce the eight transport modes; coarse movement intensity and phone-position effects are plausible explanations that we do not test here.

Keywords

smartphone sensing, transport mode, clustering, SHL dataset, multimodal smartphone sensing, unsupervised learning

1 Introduction

Smartphone sensors provide a low-cost basis for mobility monitoring and transport-mode recognition. Most studies map measurements to predefined labels using supervised classifiers [1, 6]; SHL benchmarks support reproducible evaluation with synchronized multimodal signals [5, 14]. However, journey annotations are costly, transitions are ambiguous, and labels may not transfer across participants, devices, or phone positions.

Unsupervised analysis asks what structure exists before imposing a transport taxonomy. Prior work has learned clustering representations for unlabeled GPS trajectories [11], but GPS omits body motion, orientation, vibration, and phone-position effects. Moreover, geometric cluster quality need not correspond to named transport modes. Few studies jointly compare standard clustering families on multimodal smartphone signals, examine phone position, and separate internal cluster quality from post-hoc semantic agreement.

This study addresses these gaps through two concrete research questions. *RQ1: What cluster structure, cluster resolution, and internal cluster quality emerge without transport-class labels independently for hand, bag, hips, and torso measurements? RQ2: After clustering and model selection are complete, how closely do the frozen cluster assignments correspond to the eight withheld transport-class labels?* The first question concerns unsupervised structure discovery, whereas the second is strictly a post-hoc semantic evaluation.

The objective of this study is to characterize this naturally occurring structure rather than to claim a label-free replacement for supervised classification. The work contributes: (1) an end-to-end feature pipeline constructed without transport-class labels; (2) a comparison of K-Means, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Ward agglomerative clustering, and Gaussian Mixture Models (GMMs) across four phone positions; and (3) internal evaluation followed by post-hoc semantic agreement with withheld transport-class labels.

2 Dataset

2.1 SHL Preview v2 Dataset

SHL Preview v2 contains three recording days for each of three participants, sampled at 100 Hz, with synchronized Huawei Mate 9 phones at hand, bag, hips, and torso positions [5]. It provides eight transport classes—Still, Walking, Run, Bike, Car, Bus, Train, and Subway—and a separate Null validity annotation. Null removes invalid or unannotated periods; the transport-class labels are withheld until post-hoc evaluation. The large sample count therefore represents only three people.

2.2 Sensors and Data Description

Each motion record contains a timestamp and the 22 raw or software-derived sensor channels in Table 1; these are distinct from the subsequently extracted window features.

3 Methodology

In *Level 1*, valid sensor samples are transformed, reduced with Principal Component Analysis (PCA), clustered, and evaluated internally without transport-class labels. In *Level 2*, fixed assignments are compared with the withheld labels using Adjusted Rand Index (ARI), without feedback to fitting or selection [9].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Information Society 2026, 5–9 October 2026, Ljubljana, Slovenia
© 2026 Copyright held by the owner/author(s).
https://doi.org/10.70314/is.2026.skui.73

Table 1: Smartphone signals used in the analysis.

Signal group	Components	Channels
Accelerometer	x, y, z	3
Gyroscope	x, y, z	3
Magnetometer	x, y, z	3
Orientation	quaternion w, x, y, z	4
Gravity	x, y, z	3
Linear acceleration	x, y, z	3
Ambient	pressure, altitude, temp.	3
Total		22

3.1 Preprocessing and Feature Engineering

For each phone position, recordings are matched with annotation files; timestamps and Null-marked samples are removed. Transport-class labels are stored separately and never enter imputation, scaling, windowing, feature construction or pruning, PCA, clustering, model selection, or internal evaluation. Processing then follows imputation → scaling → windowing → feature extraction → correlation pruning → PCA → clustering.

Missing values are first imputed with the median of their channel. Each raw channel is then robustly scaled:

$$\tilde{x} = \frac{x - \text{median}(x)}{Q_{0.75}(x) - Q_{0.25}(x)}, \tag{1}$$

where the denominator is the interquartile range, reducing sensitivity to isolated extremes.

The stream is segmented into 1,000-sample (10 s) windows with a 500-sample step (50% overlap), which provides context for repeated motion and vehicle vibration while reducing boundary sensitivity and event loss [13, 14]. Windows slide over the concatenated, Null-filtered stream of all nine recordings per position without segment boundaries, so a window can span a removed Null gap or a recording boundary; the latter is bounded at ≤16 of 40,968 windows (≤0.04%), while Null-gap crossings were not tracked (limitation, Section 5.1). Each window's majority transport-class label is stored separately only for post-hoc evaluation.

For each channel, mean, standard deviation, minimum, maximum, median, and signal magnitude area are calculated. The signal magnitude area of a channel is

$$\text{SMA} = \frac{1}{N} \sum_{i=1}^{N} |\tilde{x}_i|, \tag{2}$$

over the $N = 1{,}000$ samples of the window, i.e., it is normalized by window length. Three accelerometer cross-axis correlations (x–y, x–z, and y–z) capture coordinated movement, producing 22×6+3 = 135 features per window.

Features with absolute Pearson correlation above 0.90 to an already retained descriptor are removed. The retained features are *not* re-standardized before PCA, so features of larger natural magnitude (e.g., minima/maxima) can dominate the variance PCA optimizes for (limitation, Section 5.1). PCA [8] retains the minimum components explaining 95% of variance. Clustering uses the PCA coordinates.

3.2 Unsupervised Clustering

3.2.1 K-Means

K-Means [10] is evaluated for $k = 2, \ldots, 10$ with ten initializations. The data-driven count is

$$k^* = \underset{k \in \{2,\ldots,10\}}{\arg\max}\ \text{Silhouette}(k). \tag{3}$$

Silhouette is final; an automatically detected inertia knee is diagnostic only. A second, domain-informed configuration fixes $k = 8$ from the known class count but supplies no labels, identities, or distributions during fitting.

3.2.2 DBSCAN

DBSCAN [4] sets ε to the 90th percentile of fifth-nearest-neighbor distances and selects `min_samples` from {5, 10, 20} by silhouette; `min_samples`= 20 was selected at every position, with $\varepsilon = 11.491$ (hand), 7.823 (bag), 5.210 (hips), and 1.771 (torso) (noise 11.4%, 12.3%, 14.2%, 15.4%; also reported in Table 3). Noise is excluded from internal metrics and ARI because −1 marks rejected observations, not one coherent cluster.

3.2.3 Agglomerative Clustering

Ward agglomerative clustering [15] uses a sparse nearest-neighbor connectivity graph with 15 neighbors ($\min(15, n-1)$, always 15 in this dataset). Its cluster count is not selected separately: it is inherited from K-Means ($k^* = 2$).

3.2.4 GMM

GMM provides probabilistic membership. Like Ward, it reuses the K-Means count ($k^* = 2$) rather than selecting its own, so its scores reflect the method under a fixed k and not an independently optimized model. Covariance type: full (unconstrained per-component covariance matrices).

3.3 Cluster Evaluation

3.3.1 Internal Evaluation

Silhouette [12] is primary; Davies–Bouldin (DB) [3] and Calinski–Harabasz (CH) [2] are complementary. Higher silhouette and CH and lower DB are preferred. Silhouette is computed on a random subsample of up to 5,000 windows (fixed random state) for computational tractability; DB and CH are computed on the full window set (noise windows excluded for DBSCAN).

3.3.2 External/Post-hoc Evaluation

After assignments are fixed, ARI [7] compares partitions with withheld labels without matching identifiers. Zero indicates chance-level pairwise agreement and one identical partitions; ARI never feeds back into analysis.

4 Experiments and Results

4.1 Experimental Setup

Each phone position is evaluated with data-driven K-Means, DBSCAN, Ward agglomerative clustering, GMM, and K-Means with $k = 8$. Ward and GMM use the cluster count selected for K-Means. Silhouette, DB, and CH measure internal cluster quality; ARI measures post-hoc semantic agreement. Random state 42 is used where applicable.

4.2 Unsupervised Clustering Results

Transport-class labels remain withheld throughout feature processing, clustering, selection, and internal evaluation.

4.2.1 Feature Processing

The pipeline produces 40,968 overlapping windows per phone position; Table 2 shows 74–86 retained descriptors and four to nine PCA components.

Table 2: Feature reduction for each phone position.

Position	Windows	Kept	Dropped	PCA comp.
Hand	40,968	86	49	9
Bag	40,968	78	57	6
Hips	40,968	75	60	4
Torso	40,968	74	61	4

4.2.2 Internal Clustering Results

Table 3 reports the internal clustering results. Silhouette selects $k = 2$ everywhere, while the diagnostic elbow gives $k = 4$; thus $k = 2$ is used for data-driven K-Means, Ward agglomerative clustering, and

Table 3: Internal clustering results (no transport-class labels). DB: Davies–Bouldin; CH: Calinski–Harabasz; noise applies only to DBSCAN; Ward and GMM use K-Means's cluster count.

Position	Method	Clusters	Noise (%)	Sil.	DB	CH
Hand	K-Means	2	–	0.575	0.854	39,398.2
	DBSCAN	6	11.4	0.575	0.480	1,120.9
	Ward agglomerative clustering	2	–	0.476	0.864	30,186.8
	GMM	2	–	0.323	0.906	18,661.1
	K-Means ($k = 8$)	8	–	0.398	1.260	20,046.7
Bag	K-Means	2	–	0.697	0.807	41,025.4
	DBSCAN	8	12.3	0.620	0.682	1,665.7
	Ward agglomerative clustering	2	–	0.688	0.829	39,153.6
	GMM	2	–	0.459	0.932	19,461.2
	K-Means ($k = 8$)	8	–	0.484	1.114	24,604.8
Hips	K-Means	2	–	0.730	0.631	65,970.9
	DBSCAN	11	14.2	0.681	0.551	11,199.4
	Ward agglomerative clustering	2	–	0.723	0.641	60,303.9
	GMM	2	–	0.497	0.865	26,242.3
	K-Means ($k = 8$)	8	–	0.548	0.899	50,855.0
Torso	K-Means	2	–	0.713	0.703	41,441.8
	DBSCAN	6	15.4	0.584	0.410	1,956.9
	Ward agglomerative clustering	2	–	0.657	0.753	34,971.8
	GMM	2	–	0.489	0.887	20,357.3
	K-Means ($k = 8$)	8	–	0.462	0.884	52,860.2

GMM. DBSCAN finds 6, 8, 11, and 6 non-noise clusters for hand, bag, hips, and torso.

The highest silhouettes are 0.730 for hips K-Means, 0.723 for hips Ward agglomerative clustering, and 0.713 for torso K-Means; GMM is lowest at every phone position (Figure 1). DBSCAN's highest silhouette is 0.681, for hips; noise windows are excluded from its internal metrics.

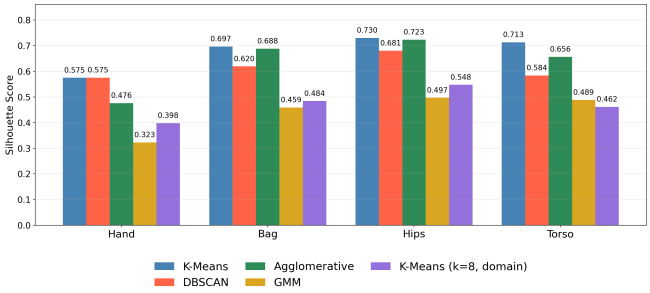


Figure 1: Internal silhouette scores obtained without transport-class labels. GMM denotes Gaussian Mixture Model.

To characterize this dominant solution, Table 4 profiles hips K-Means ($k = 2$): Cluster 0 (81.3%) is spread across Still, Car, Bus, Train, and Subway with almost no Walking or Run; Cluster 1 (18.7%) is dominated by Walking. Its five most discriminative features (Kruskal–Wallis, all $p < 10^{-300}$) are gyroscope-x std, linear-acceleration x/y maxima, gravity-z std, and gyroscope-y min—all rotation/acceleration features—suggesting the clusters separate on movement intensity rather than the eight transport modes, though this is one representative solution, not a tested claim.

4.3 Agreement with Transport Modes

Withheld transport-class labels are introduced only here; ARI does not alter assignments or fitting.

Table 5 reports post-hoc ARI: DBSCAN ranges from 0.004–0.057 and data-driven K-Means from 0.042–0.102. Fixing $k = 8$ raises ARI at all four positions (hand 0.091→0.136, bag 0.042→0.173, hips 0.102→0.141, torso 0.058→0.193); the gain is smallest for hips, and torso is highest at 0.193, followed by bag at 0.173. GMM reaches 0.158 for torso and 0.156 for bag and slightly exceeds $k = 8$ K-Means for hips (Figure 2).

Table 4: Hips K-Means ($k = 2$) solution: cluster sizes and transport-mode composition (each column sums to 1).

	Cluster 0	Cluster 1
n (%)	33,311 (81.3)	7,657 (18.7)
Still	0.23	0.01
Walking	0.03	0.75
Run	0.00	0.11
Bike	0.08	0.11
Car	0.18	0.00
Bus	0.11	0.01
Train	0.18	0.01
Subway	0.18	0.01

Table 5: Post-hoc agreement (ARI) with withheld transport-class labels. DBSCAN excludes noise windows (11.4–15.4%, Table 3), a different subset than the other methods; Ward and GMM use K-Means's cluster count.

Position	Method	Clusters	ARI
Hand	K-Means	2	0.091
	DBSCAN	6	0.004
	Ward agglomerative clustering	2	0.112
	GMM	2	0.080
	K-Means ($k = 8$)	8	0.136
Bag	K-Means	2	0.042
	DBSCAN	8	0.007
	Ward agglomerative clustering	2	0.045
	GMM	2	0.156
	K-Means ($k = 8$)	8	0.173
Hips	K-Means	2	0.102
	DBSCAN	11	0.057
	Ward agglomerative clustering	2	0.073
	GMM	2	0.143
	K-Means ($k = 8$)	8	0.141
Torso	K-Means	2	0.058
	DBSCAN	6	0.007
	Ward agglomerative clustering	2	0.124
	GMM	2	0.158
	K-Means ($k = 8$)	8	0.193

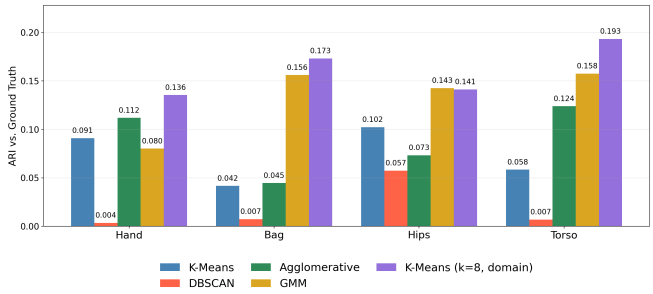


Figure 2: Post-hoc ARI against withheld transport-class labels. GMM denotes Gaussian Mixture Model.

5 Discussion and Conclusion

5.1 Discussion

For RQ1, silhouette selects two clusters everywhere and gives the highest internal cluster quality for hips, indicating coarse structure that does not correspond to eight natural groups; whether it is driven by movement intensity or phone position was not tested. For RQ2, the best ARI is only 0.193 even with domain-informed $k = 8$. Thus, high internal cluster quality does not imply post-hoc semantic agreement.

Phone position plausibly contributes: hips emphasize broad locomotion intensity, torso may preserve body dynamics while limiting hand motion, bag signals are mechanically filtered, and hand signals include unconstrained interaction. These interpretations are not cluster labels.

Supervised studies optimize transport recognition directly [1, 5, 6, 14], while GPS-based unsupervised work jointly learns reconstruction and clustering representations [11]. Here, PCA preserves variance and clustering rewards geometry, neither necessarily prioritizing transport semantics.

The 40,968 overlapping windows per phone position are dependent observations from only three participants, precluding population-level generalization. Further limitations are lost temporal order, majority-label masking of transitions, and windows spanning a removed Null gap or a recording boundary, including $\leq$0.04% spanning a recording boundary, whose exclusion effect was not directly measured. Window-level features also enter PCA without re-standardization, so larger-magnitude features (e.g., minima/maxima) can dominate the retained components disproportionately. DBSCAN's global density assumption may fit phone positions unevenly, potentially affecting comparability.

Future work should examine transition windows, frequency-domain and orientation-invariant features, sequence-aware or contrastive embeddings, variable-density clustering, and limited labels for semantic assignment.

5.2 Conclusion

This study separates structure discovery without transport-class labels from post-hoc semantic evaluation across four SHL phone positions. K-Means finds a strong two-cluster solution for hips (silhouette 0.730), yet the best agreement with the eight withheld transport-class labels is only ARI 0.193 for domain-informed K-Means on torso data. The dominant structure therefore does not align closely with specific transport modes; broader movement or phone-position effects are a plausible but untested explanation. Unsupervised clustering should not be presented here as a replacement for supervised recognition. Its value is diagnostic: it reveals what is naturally separable without transport-class labels and motivates sequence-aware representations or limited semantic supervision when named transport modes are required.

References

[1] H. I. Ashqar, M. H. Almannaa, M. Elhenawy, H. A. Rakha, and L. House. 2019. Smartphone transportation mode recognition using a hierarchical machine learning classifier and pooled features from time and frequency domains. *IEEE Transactions on Intelligent Transportation Systems* 20, 1 (2019), 244–252.

[2] T. Caliński and J. Harabasz. 1974. A dendrite method for cluster analysis. *Communications in Statistics* 3, 1 (1974), 1–27.

[3] D. L. Davies and D. W. Bouldin. 1979. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-1, 2 (1979), 224–227.

[4] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. 226–231.

[5] H. Gjoreski, M. Ciliberto, L. Wang, F. J. Ordóñez Morales, S. Mekki, S. Valentin, and D. Roggen. 2018. The University of Sussex–Huawei locomotion and transportation dataset for multimodal analytics with mobile devices. *IEEE Access* 6 (2018), 42592–42604.

[6] S. Hemminki, P. Nurmi, and S. Tarkoma. 2013. Accelerometer-based transportation mode detection on smartphones. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. Article 13, 1–14.

[7] L. Hubert and P. Arabie. 1985. Comparing partitions. *Journal of Classification* 2 (1985), 193–218.

[8] I. T. Jolliffe. 2002. *Principal Component Analysis* (2nd ed.). Springer, New York.

[9] E. Kizhevska, H. Gjoreski, and M. Luštrek. 2026. A two-stage methodology combining clustering and predictive models to estimate the levels of participants' empathy and other affective states during virtual reality sessions. *IEEE Access* 14 (2026), 38493–38502. DOI: https://doi.org/10.1109/ACCESS.2026.3668967.

[10] J. MacQueen. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1. 281–297.

[11] C. Markos and J. J. Q. Yu. 2020. Unsupervised deep learning for GPS-based transportation mode identification. In *Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems*. 1–6.

[12] P. J. Rousseeuw. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20 (1987), 53–65.

[13] A. Skretting, T.-M. Grønli, T. A. Majchrzak, C. Mateos, and M. Hirsch. 2024. Enhanced transport mode recognition on mobile devices. In *Proceedings of the 57th Hawaii International Conference on System Sciences*. 7740–7749.

[14] L. Wang, H. Gjoreski, M. Ciliberto, S. Mekki, S. Valentin, and D. Roggen. 2019. Enabling reproducible research in sensor-based transportation mode recognition with the Sussex–Huawei dataset. *IEEE Access* 7 (2019), 10870–10891.

[15] J. H. Ward, Jr. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58, 301 (1963), 236–244.