Data Science for Transportation (2026) 8:13
https://doi.org/10.1007/s42421-026-00155-7

RESEARCH

A Functional Bus Stop Typology: Explainable Machine Learning on Transit Data to Inform Policymaking

Mahdi Azhdari[1] · Jimi Oke[1] · Eric J. Gonzales[1]

Received: 25 November 2025 / Revised: 3 March 2026 / Accepted: 7 April 2026

Abstract
Bus stops are critical elements of transit networks, shaping accessibility, reliability, and multimodal connectivity. While they simultaneously play different functional roles and are tightly connected within the wider system, there is a need for analyses that capture their functional relationships and roles in an integrated way. This paper develops a data-driven bus stop typology that combines operational performance, ridership and fare behavior, and multimodal context, and shows how it can support targeted, policy-relevant interventions. Using the Massachusetts Bay Transportation Authority (MBTA) bus network as a case study, we use Automated Fare Collection (AFC), Automated Passenger Counter (APC), and GIS data into a stop-level feature space that captures the diverse functional patterns of bus stops across operations, demand, and multimodal context. We apply a Gaussian Mixture Model (GMM) to standardized indicators to uncover four distinct stop types. To interpret and validate these roles, we use one-vs-rest Random Forest classifiers with SHAP explanations and a multiclass XGBoost model, which achieves high out-of-sample accuracy and near-perfect ROC–AUC, indicating that the clusters are robust and recoverable from their underlying indicators. The resulting typology distinguishes multi-route destinations, rail-adjacent distributors, high-demand hubs, and low-ridership locals, each with characteristic operational and spatial signatures. We then translate these types into actionable policy implications, providing a transferable framework to support evidence-based decisions and investment prioritization in bus stop planning and management.

Keywords Bus stop typology · Explainable machine learning · Automated transit data · Data-driven transit policy

Introduction

Bus stops serve distinct and context-dependent functions within a transit system. Their roles are shaped by geography, surrounding land use, passenger demand, and their position within the broader network (Fitzpatrick et al. 1996; NACTO 2016). Rather than operating uniformly, stops contribute differently to accessibility, connectivity, and overall performance. These differences manifest in variation of operational patterns, such as passenger boardings and alightings, dwell times, and localized crowding—all of which influence reliability and the rider experience. Recognizing this heterogeneity is critical for effective service planning, targeted investment, and design decisions that promote integrated, efficient travel across modes (TCRP 2006; Ge et al. 2021).

Despite growing recognition of these differences, many planning and evaluation frameworks still treat bus stops as uniform elements. Traditional design manuals and rule-based methods—checklists, spacing thresholds, and placement heuristics—emphasize geometry, spacing, and amenities while overlooking functional diversity and interaction effects within the broader network (Fitzpatrick et al. 1996; TCQSM 2013). Empirical analyses, meanwhile, often rely on coarse indicators (e.g., boardings/alightings, nominal spacing) that mask the interplay between passenger behavior, service performance, and multimodal context (Wong and Khani 2018; TCRP 2006). As a result, agencies lack systematic, data-integrated tools to distinguish stop roles, prioritize operational and capital improvements, and trace how stop-level dynamics scale to network-wide performance.

✉ Mahdi Azhdari
mazhdari@umass.edu

Jimi Oke
jimi@umass.edu

Eric J. Gonzales
gonzales@umass.edu

[1] Department of Civil and Environmental Engineering, University of Massachusetts, 130 Natural Resources Rd, Amherst, MA 01003, USA

With expanding access to large automated datasets—Automatic Passenger Counters (APC), Automatic Vehicle Location (AVL), Automated Fare Collection (AFC), and route and schedule information in General Transit Feed Specification (GTFS)—there is a growing need for a robust, fully data-driven framework at the stop (Welch and Widita 2019). This framework should draw indicators from diverse domains related to passenger behavior, service performance, and the multimodal context. Integrating these sources within a unified approach helps agencies see how stops actually function in the network and how their characteristics affect reliability, connectivity, and user experience. The resulting evidence can guide more targeted planning and operational management, as well as smarter design and investment decisions to prioritize improvements, where they yield the greatest system-wide benefits (TCRP 2006; Wang et al. 2022).

We address this need by developing a data-driven typology of bus stops built from integrated AFC, APC, and GIS data that combine stop-level operations, ridership behavior, and spatial context from bus stops and nearby train stations. We use the Massachusetts Bay Transportation Authority (MBTA) network as a case study. Our approach results in identification of novel stop types based on how they actually function in the network and how those roles shape reliability, connectivity, and rider experience. Methodologically, we assemble a unified feature set, apply an interpretable clustering framework to reveal distinct stop roles, and confirm that these roles are consistent and generalizable through supervised validation. Our key contributions are as follows: (1) we construct an integrated stop-level dataset that fuses behavioral, operational, and spatial indicators; (2) we identify a concise and interpretable set of stop roles; (3) we also demonstrate that these roles are stable across data samples; and (4) linking each role to actionable levers in operations, curb and stop design, spacing and assignments, and safety. The framework offers a transferable, evidence-based approach that helps agencies target interventions and guide investment decisions.

Literature Review

Before the widespread use of automated datasets, agencies relied primarily on rule-based frameworks to guide bus stop design and placement. Foundational guidance—most notably TCRP Report 19 (Fitzpatrick et al. 1996)—codified how placement (near-, far-, and mid-block), curb geometry (curbside vs. bay), and amenities (pads, shelters, lighting, sidewalks) shape both operational performance and rider experience, functioning as a de facto typology for stop design. Complementary manuals extended this checklist approach: the American Public Transportation Association (APTA) published security guidelines that emphasized visibility and lighting (APTA 2010); the *Transit Street Design Guide* (NACTO 2016) formalized context-sensitive street and stop treatments widely adopted in U.S. cities; regional standards such as Berkeley–Charleston–Dorchester Council of Governments (BCDCOG 2021) provided spacing thresholds tailored to local conditions; and federal guidance—such as the *Bus Stop Optimization Report* from the Federal Transit Administration—reinforced consistent stop-selection and placement practices (Wu et al. 2010). These tools remain valuable for quick, standardized decisions and baseline consistency, but because they are largely static rule sets, they can overlook the operational variability and dynamics of multimodal transfers later revealed by real-world, event-level data.

As automated transit datasets matured, agencies and researchers gained continuous system-wide visibility instead of relying on patchy field counts (Lu et al. 2020). Integrating APC, AVL, AFC, and GTFS feeds allows agencies to link ridership, operations, and spatial context at the stop-event level. Examples include AFC+AVL fusion for OD inference, APC for passenger load/dwell time and validation, and GTFS as the common schema tying together stops, trips, and blocks. This multi-source approach underpins today's stop-level diagnostics and planning workflows (Ge et al. 2021; Li et al. 2018; Liu et al. 2025). Using APC and AVL, agencies can now link headways, dwell times, and onboard passenger loads to diagnose delay patterns and reliability (Wong and Khani 2018). Spatial attributes such as route count and rail proximity help situate each stop within the multimodal network (Ge et al. 2021). AFC data add dimensions of transfer intensity, fare type, and temporal demand profiles, while integrated AFC + APC frameworks have been used to infer OD flows and transfer chains (Devillaine et al. 2012; Wu et al. 2019). Because automated feeds often contain systematic errors, best-practice guidance emphasizes data-quality screening and balancing before analysis (TCRP 2006; Olivo et al. 2019). Similarly, in the context of fare compliance and monitoring, other studies have integrated AFC and APC data to estimate fare evasion and the associated revenue loss (Egu and Bonnel 2020; Gonzales et al. 2024).

Recent studies increasingly employ unsupervised learning to uncover latent structures in stop- and station-level behavior—supporting tasks from typology building to detecting operational and demand patterns—without relying on predefined outcomes (Cats 2024). These methods classify stops based on shared operational, spatial, and behavioral characteristics, revealing functional roles that inform planning and design. For example, Lin et al. (2024) apply k-means clustering to APC data from King County Metro to profile stop-level ridership before and during COVID-19, linking clusters to socio-economic and land-use contexts. Extending this, Cheng et al. (2024) use a fuzzy clustering framework on high-resolution smart-card data for London rail stations, allowing partial membership and capturing nuanced intra-day and intra-week usage overlooked by hard assignments. In parallel, Montoya et al. (2024) integrate passenger activity, built environment, and accessibility via k-prototypes to derive context-specific functional stop types within urban networks. Beyond typology discovery, Degeler et al. (2021) show that clustering can extract recurrent operational phenomena—identifying "bunching swing" formations across The Hague's tram network and relating them to passenger load and delay propagation. Similarly, Abdalazeem and Oke (2023) apply hierarchical clustering to noisy mobile-ticketing traces from the Pioneer Valley Transit Authority to derive spatiotemporal trip typologies, revealing recurring passenger behaviors that aid demand estimation. Collectively, these studies point to interpretable, scalable, data-driven classifications that capture system-wide regularities while preserving local behavioral and spatial nuance.

Prior studies either rely on rule-based design guidance and checklists (Fitzpatrick et al. 1996; NACTO 2016; APTA 2010); focus on isolated operational metrics, such as dwell time or headway (Wong and Khani 2018; Daganzo and Pilachowski 2011); or limit attention to predicting/extracting ridership or fare-behavior patterns (Baek and Sohn 2016; Baghbani et al. 2023; Gonzales et al. 2024). Relatively few studies explicitly classify or cluster bus stops, and those that do often narrow the lens to a single data stream or build typologies from limited feature sets. They rarely articulate a clear policy pathway that links types to concrete operational or design actions (Montoya et al. 2024; Cats 2024; Cheng et al. 2024; Degeler et al. 2021). What remains missing is a reproducible, fully data-driven pipeline that integrates indicators across domains (passenger behavior, service performance, multimodal context), yields interpretable roles, and demonstrates out-of-sample separability, so the resulting types can directly inform design, operations, and safety interventions.

This paper addresses that gap by building a stop-level feature space that integrates operational, ridership, fare/behavioral, and multimodal context indicators into a single analysis framework. We derive an interpretable typology using Gaussian Mixture Models with principled model selection, explain type drivers via per-type Random Forest SHAP (SHapley Additive exPlanations) (Lundberg and Lee 2017), and verify recoverability with a supervised multiclass eXtreme Gradient Boosting (XGBoost) model. The outputs are not only statistically separable but also policy-ready. Each type is mapped to targeted operational controls, curb/stop design actions, spacing and assignment strategies, and safety treatments—providing a direct bridge from integrated measurement to implementation. To our knowledge, this is the first bus-stop typology to fuse AFC–APC–GIS data into indicators of ridership/fare behavior, service performance, and multimodal context, pair a Gaussian Mixture Model (GMM) discovery step with per-type SHAP interpretation and out-of-sample validation, and translate the resulting types into an actionable policy table.

Methods

We collected and standardized bus-stop-level indicators from multiple sources—operations and schedule feeds (headway, dwell time), APC summaries (boardings, alightings, onboard load), fare and transfer signals (transfer intensity, non-interaction rate), and a GIS layer for distance to the nearest rail station—into a single analysis file with consistent units and completeness screening. Using these indicators, we applied the GMM clustering to identify distinct stop types. Figure 1 summarizes the full workflow from data inputs through clustering, validation, and the policy table outputs.

Data Description and Aggregation

To interpret the drivers of each type, we trained four separate Random Forest classifiers, and used SHAP analyses to quantify feature contributions with respect to the probability of belonging to that type. For overall validation and separability testing, we additionally trained a single multiclass XGBoost classifier on the same features and cluster labels. This validation model provided performance diagnostics (confusion matrix, per-class metrics, ROC–AUC) and permutation-based feature importance, confirming the robustness and interpretability of the typology. Finally, we translated the results into actionable guidance by summarizing type-specific operational, design, and safety implications in a policy table.

This study uses data from the Massachusetts Bay Transportation Authority (MBTA) Research Database, which integrates records from the AFC system, APC system, and the inferred Origin–Destination–Transfer (ODX) table. These data sources jointly provide detailed information about passenger activity, fare transactions, and vehicle operations across the MBTA bus network. The combined dataset enables the analysis of stop-level operational performance, demand characteristics, and multimodal connectivity. Table 1 lists the stop-level indicators, sources, and units used to build the analysis file.

The AFC system records individual fare transactions, including card identifiers, transaction time, ticket type, and payment amount. The ODX data is a processed version of AFC data that includes validated fare transactions associated with passenger origins and destinations, along with the corresponding route, vehicle, and timestamp of the transaction. The APC system captures passenger boardings, alightings, and on-board load at each stop event using sensors installed on buses. A *stop event* is a single bus arrival and door-open episode at a specific stop, route, direction, and time, with

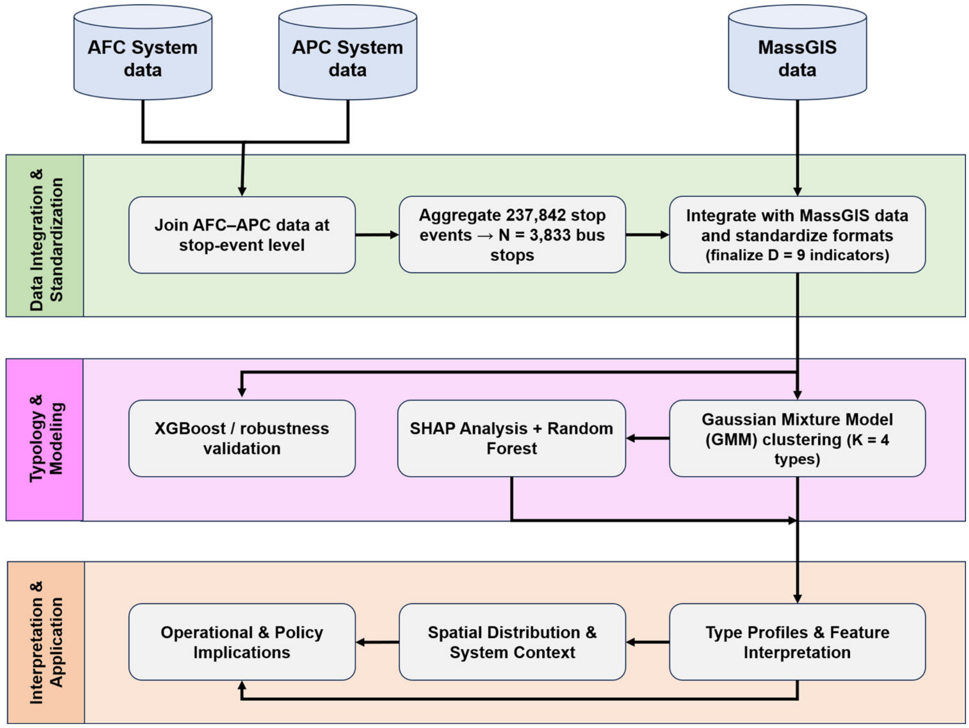


Fig. 1 Study workflow from data to policy

Table 1 Stop-level indicators, descriptions, and sources

Feature (unit)	Description	Data source
Headway (min)	Mean inter-arrival by stop–route–direction	AVL
Dwell time (min)	Mean door-open duration	AVL + APC
Passenger boardings (pax/event)	Mean boardings per stop event	APC
Passenger alightings (pax/event)	Mean alightings per stop event	APC
On-board load (pax)	Mean onboard passengers at arrival	APC
Route count (count)	Distinct routes per stop	GTFS
Transfers (transfers/event)	Mean number of transfer tickets (from rail to bus) inferred within the stopping window	AFC + ODX
Non-interaction rate (%)	Share of passenger boardings without a matched fare system record	AFC + ODX
Distance to nearest train station (m)	Straight-line distance from bus stop to nearest rail station	MassGIS MBTA

associated passenger counts from APC and any matched fare interactions from AFC/ODX. These data also contain the vehicle ID, route, direction, and timestamp of each stop, enabling the estimation of time-based operational measures, such as headway and dwell time. Figure 2 illustrates the temporal matching logic used to join AFC transactions with APC stop events, producing a unified stop-level dataset for analysis.

The joined working file is at the stop-event level. After standard cleaning, we aggregated 237,842 stop events into 3833 MBTA bus stops and computed nine stop-level indicators for modeling. Figure 3 illustrates the hourly profiles of stop events and total boardings across the four analyzed weekdays. Both measures show distinct AM and PM peaks (around 07:00–09:00 and 16:00–18:00, respectively), with consistent temporal patterns across weekdays, reflecting that service activity and passenger demand rise and fall together throughout the day.

From the aggregated stop-level dataset, we derived operational, ridership, and connectivity indicators. Operational measures include headway (average time between consecutive bus arrivals for the same route and direction) and dwell

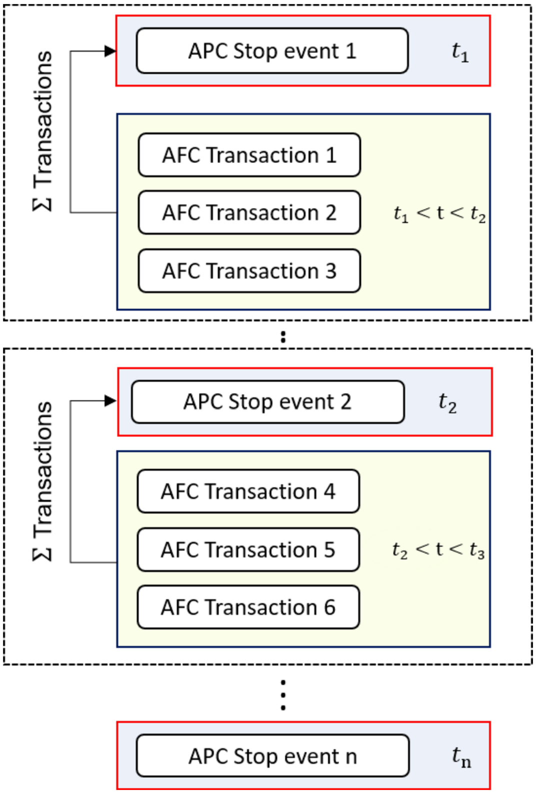


Fig. 2 AFC and APC data integration logic. Each AFC transaction occurring between consecutive APC stop events ($t_i \le t < t_{i+1}$) is counted and assigned to stop event at time t_i; n denotes the final observed stop event in the sequence (e.g., end of run or day)

time (average duration of door-open events). Passenger activity indicators include total boardings, alightings, and average on-board load. Transfer intensity and the non-interaction rate were calculated from matched AFC and APC records, where transfers represent fare-linked movements between routes, and the non-interaction rate measures the share of stop events without an associated fare transaction. Spatial context variables were added from the MBTA bus stop and rail station GIS layers, including the number of bus routes serving each stop and the distance to the nearest rail station.

Together, these features describe the role of each stop within the network. Figure 4 shows histograms of all nine indicators, with medians marked, to show the range and skewness we model against. Operational variables capture service supply and reliability, ridership measures reflect demand intensity, fare-based indicators reveal transfer and payment behavior, and spatial attributes describe network connectivity and multimodal proximity. Keeping all features at the stop level preserves variation across corridors and locations, allowing the subsequent clustering and classification analyses to uncover distinct functional types within the bus system.

Clustering Analysis

We clustered bus stops using a GMM (Dempster et al. 1977) applied to the standardized indicators described above. Let $\mathbf{x}_n \in \mathbb{R}^D$ denote the feature vector for stop n. The GMM assumes

$$p(\mathbf{x}_n \mid \Theta) = \sum_{k=1}^{K} \pi_k \, \mathcal{N}\big(\mathbf{x}_n \,\big|\, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\big),$$

where π_k are mixture weights ($\sum_{k=1}^{K} \pi_k = 1$), and $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the component mean vector and covariance matrix. Parameters $\Theta = \{\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^{K}$ are estimated using the expectation maximization (EM) algorithm.

In the E-step, responsibilities are

$$\gamma_{nk} = \frac{\pi_k \, \mathcal{N}\big(\mathbf{x}_n \,\big|\, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\big)}{\sum_{j=1}^{K} \pi_j \, \mathcal{N}\big(\mathbf{x}_n \,\big|\, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\big)}.$$

In the M-step, we update as follows:

$$\pi_k = \frac{N_k}{N}, \qquad \boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk} \, \mathbf{x}_n,$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk} \, (\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\top},$$

with $N_k = \sum_{n=1}^{N} \gamma_{nk}$.

Prior to fitting, all features were standardized. Model selection varied the number of components $K \in \{2, \ldots, 14\}$ and the covariance parameterization $\boldsymbol{\Sigma}_k \in \{\texttt{full}, \texttt{tied}, \texttt{diag}, \texttt{spherical}\}$. Each configuration used the library defaults for initialization (k-means) with five restarts ($n\_init = 5$) and fixed regularization (default reg_covar). The Bayesian Information Criterion (BIC):

$$\mathrm{BIC} \;=\; -2 \log \hat{L} + p \log N,$$

where $\hat{L}$ is the maximized likelihood and p is the number of free parameters, provided the objective for selection. We inspected BIC curves across K and covariance types and used the minimum BIC together with an elbow check (notably in the range $K = 3$–5) to choose the final solution. Figure 5 plots BIC across candidate K and covariance forms, with a clear elbow and minimum at the selected four-component model (diagonal covariance). The selected model yielded four clusters, which we interpret as distinct stop roles in the bus network.

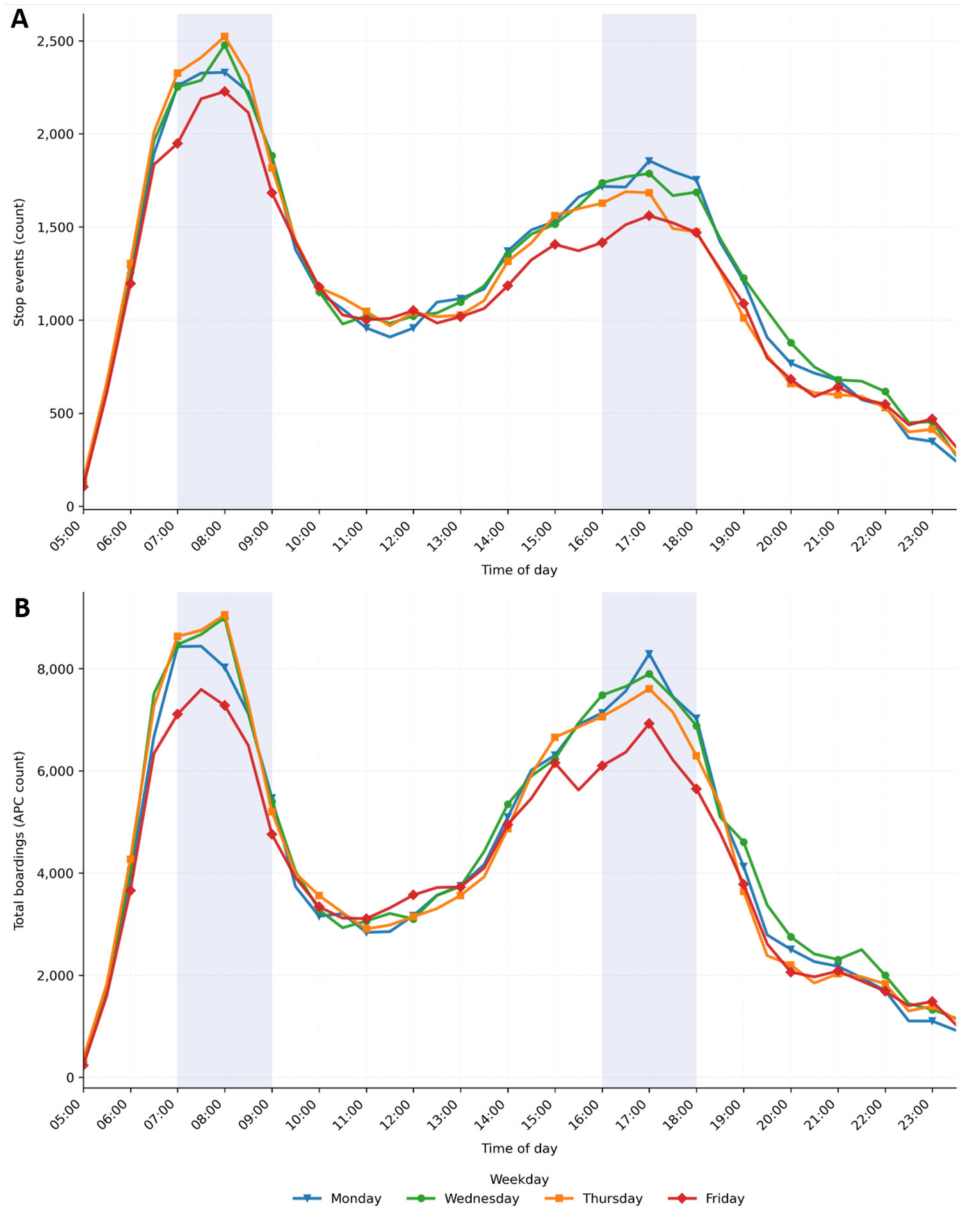


Fig. 3 Time-of-day profiles across 4 weekdays (Monday, Wednesday, Thursday, and Friday) aggregated in 30-min bins. **A** Number of stop events. **B** Total number of boardings (APC count). Shaded bands indicate the morning (07:00–09:00) and evening (16:00–18:00) peak periods

Type Interpretation and SHAP Analysis

After clustering, each bus stop was assigned a type label $\hat{k} \in \{1, \ldots, K\}$ by the GMM. To interpret how the original indicators contribute to each type, we trained four separate Random Forest classifiers (one per type) (Breiman 2001) and applied SHAP (Lundberg and Lee 2017) to quantify feature contributions.

For each type k, we constructed a binary target

$$y^{(k)} = \begin{cases} 1, & \text{if the stop belongs to type } k, \\ 0, & \text{otherwise,} \end{cases}$$

and fitted a Random Forest on the standardized indicators. Given an input vector

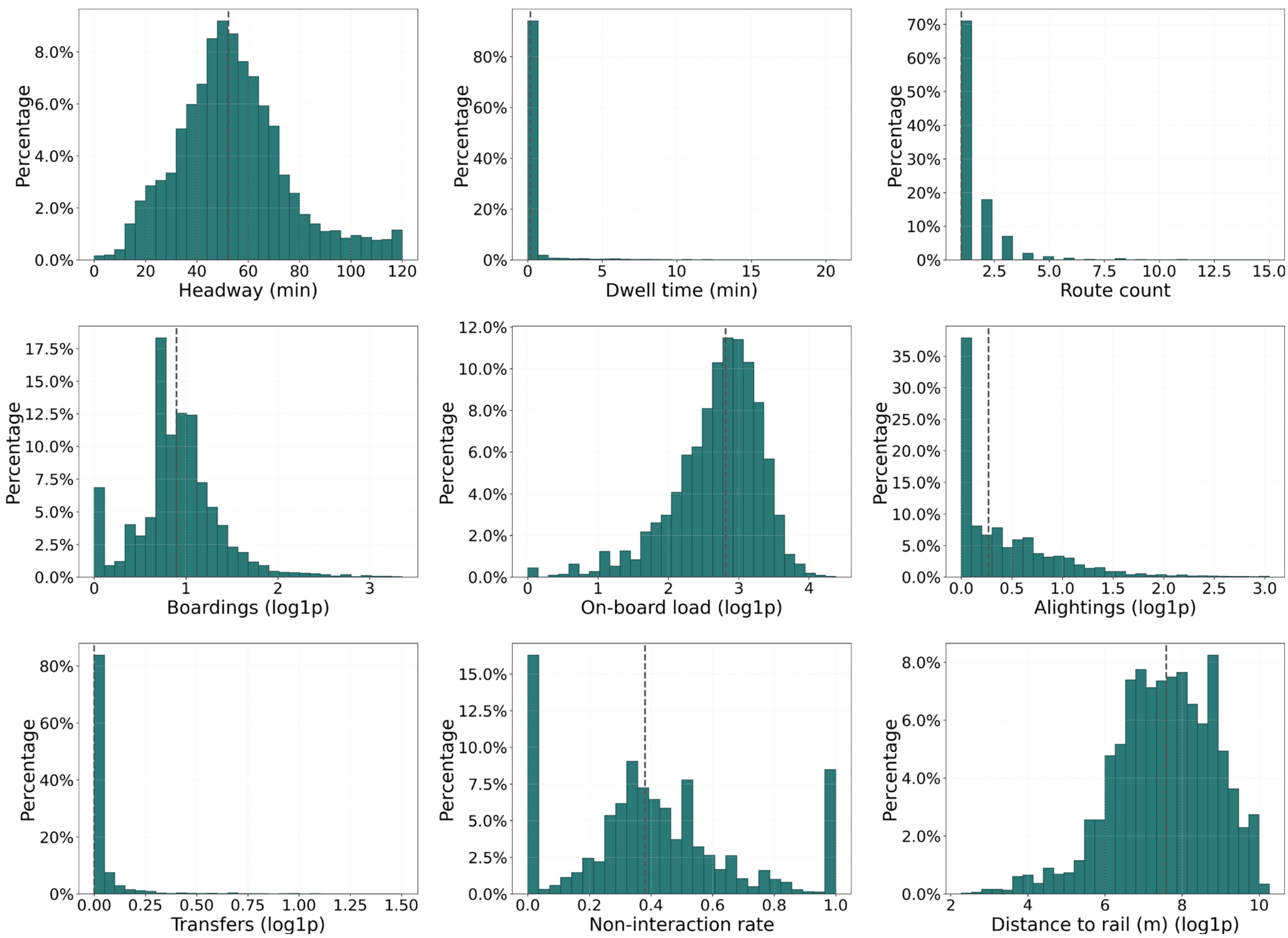
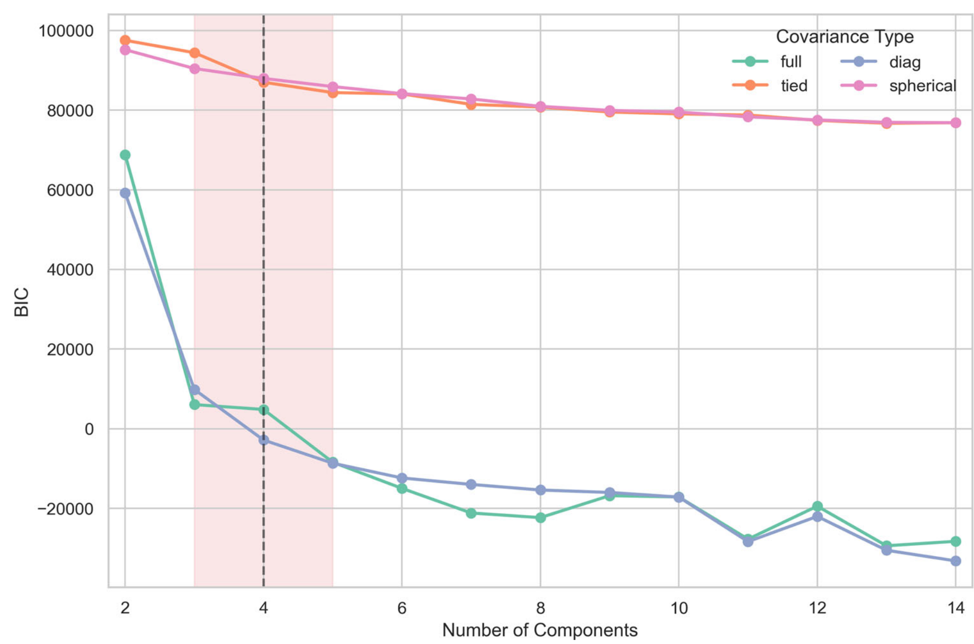


Fig. 4 Histograms of nine stop-level indicators used in the typology; dashed lines mark medians

Fig. 5 GMM model selection via BIC; the four-component model with diagonal covariance was chosen as optimal



$\mathbf{x} = [x_1, \ldots, x_D]^\top$

of D indicators, the forest aggregates, over T trees, class probabilities $p_t^{(k)}(1 \mid \mathbf{x})$ that a stop with features $\mathbf{x}$ belongs to type k. The ensemble estimate is

$$\hat{P}^{(k)}(1 \mid \mathbf{x}) = \frac{1}{T}\sum_{t=1}^{T} p_t^{(k)}(1 \mid \mathbf{x}),$$

which we interpret as the likelihood that $\mathbf{x}$ exhibits the operational and spatial characteristics of type k.

To explain the one-vs-rest classifiers, we computed SHAP values with respect to the probability output $\hat{P}^{(k)}(1 \mid \mathbf{x})$. For a given stop n with feature vector $\mathbf{x}_n$, the model output admits the standard SHAP decomposition,

$$f^{(k)}(\mathbf{x}_n) \; = \; \mathbb{E}_{\mathbf{X}}\Big[f^{(k)}(\mathbf{X})\Big] \; + \; \sum_{d=1}^{D} \phi_{nd}^{(k)},$$

where $\mathbb{E}_{\mathbf{X}}[f^{(k)}(\mathbf{X})]$ is the expected output over the background distribution of feature vectors, and $\phi_{nd}^{(k)}$ is the contribution of feature d to the probability that stop n belongs to type k. Positive SHAP values increase this probability, while negative values decrease it.

Formally, each SHAP value $\phi_{nd}^{(k)}$ is a Shapley value computed over all subsets of the remaining features:

$$\phi_{nd}^{(k)} = \sum_{S \subseteq \{1,\ldots,D\}\setminus\{d\}} \frac{|S|!\,\big(D - |S| - 1\big)!}{D!} \times \Big[f_n^{(k)}(S \cup \{d\}) - f_n^{(k)}(S)\Big],$$

where S ranges over all subsets of $\{1, \ldots, D\}$ that do not contain d, and $f_n^{(k)}(S)$ denotes the model prediction for stop n when only the features in subset S are included. For each of the four one-vs-rest classifiers, we generated per-type SHAP beeswarm plots to visualize the magnitude and direction of feature effects on $\hat{P}^{(k)}(1 \mid \mathbf{x})$ across all stops.

Assessing Typology Separability

To assess how well the identified stop types can be recovered from their original indicators, we developed a supervised classification framework. The aim was to test whether the four clusters derived from the unsupervised GMM analysis correspond to statistically separable and reproducible patterns in the underlying operational and ridership data. This framework also yields a practical model that can assign new stops to a type based on their indicators without rerunning the unsupervised pipeline.

We trained and tested multiple classifiers using the same nine standardized indicators and the GMM labels: XGBoost (Chen and Guestrin 2016), Random Forest (Breiman 2001), Artificial Neural Network (ANN) (Rosenblatt 1958), Logistic Regression (Cox 1958), and an RBF-kernel SVM (Cortes and Vapnik 1995). Each model was tuned via grid search with threefold cross-validation on the training split; the complete grids and sensitivity analyses are reported in Appendix B. Performance on the held-out test set was evaluated using log-loss, accuracy, precision, recall, macro-F1, and macro ROC–AUC.

Given the results in Table 2, we selected the multiclass eXtreme Gradient Boosting (XGBoost) classifier for final evaluation and visualization. We used the same nine standardized stop-level indicators as inputs and the GMM cluster label as the target.

Following the gradient-boosting formulation, at boosting iteration t, XGBoost minimizes a regularized objective:

$$\mathcal{L} = \sum_{n=1}^{N} \ell\big(y_n,\; \hat{y}_n^{(t-1)} + f_t(\mathbf{x}_n)\big) + \Omega\big(f_t\big), \tag{1}$$

where $\ell(\cdot)$ is the differentiable loss comparing the true label y_n with the current model prediction $\hat{y}_n^{(t-1)}$, f_t is the t-th regression tree added to the ensemble, and $\Omega(\cdot)$ penalizes tree complexity to control overfitting. Optimization proceeds via a second-order Taylor expansion around $\hat{y}_n^{(t-1)}$:

$$\mathcal{L}^{(t)} \approx \sum_{n=1}^{N} \Big[\, g_n\, f_t(\mathbf{x}_n) + \tfrac{1}{2}\, h_n\, f_t(\mathbf{x}_n)^2 \Big] \; + \; \Omega\big(f_t\big), \tag{2}$$

with $g_n = \partial\ell/\partial\hat{y}_n$ and $h_n = \partial^2\ell/\partial\hat{y}_n^2$ denoting the first and second derivatives (gradient and Hessian).

As summarized in Table 2, XGBoost achieved the best overall performance among the candidates, with test accuracy of 0.981 and macro-$F1$ of 0.968. All models performed strongly, confirming that the GMM-based labels are highly recoverable from the original indicators. We, therefore, used the XGBoost model for final evaluation and visualization, including one-vs-rest ROC curves, a confusion matrix, and permutation-based feature importance. Full hyperparameter grids and tuning outcomes are provided in Appendix B.

Testing the Model Sensitivity to Data Availability and Feature Uncertainty

Model Sensitivity to Data Availability To explore model performance under different data availability conditions, we conducted a reduced-input retraining analysis using the multiclass XGBoost validation model. The objective was to evaluate how well the four stop types can be recovered when the available indicator set is constrained, and to distinguish

Table 2 Evaluation metrics on the test set for the top classification models

Method	Log loss	Accuracy	Precision*	Recall*	$F1^*$	ROC–AUC
XGBoost	0.043	0.981	0.976	0.963	0.968	1.000
Random Forests	0.078	0.978	0.966	0.963	0.964	0.999
ANN (MLP)	0.049	0.976	0.962	0.961	0.961	0.999
Logistic Regression	0.120	0.956	0.945	0.928	0.936	0.997
SVM (RBF)	0.105	0.955	0.945	0.923	0.934	0.997

Columns marked with * are macro-averaged across classes

indicators that are most essential from those that are complementary.

We considered two retraining protocols. First, we retrained XGBoost using only the top-k indicators ranked by the baseline model's global SHAP importance (mean absolute SHAP value), varying k from 2 to 9. This feature-budget experiment quantifies how predictive performance changes as the available indicator set becomes more compact. Second, we retrained XGBoost after excluding indicators associated with one data source at a time (APC, AFC/ODX, AVL, or GTFS/GIS), using the indicator-to-source mapping in Table 3. This source-based retraining experiment represents scenarios, where only a subset of data feeds is available for computing the indicators.

All retrained models used the same train/test split and the same tuned XGBoost specification as the baseline model, and were evaluated on the held-out test set using accuracy (reported as test error = 1 − accuracy). These experiments provide a robustness assessment of the typology under reduced indicator and data-source conditions.

Indicator Uncertainty Sensitivity Analysis

To evaluate the stability of the stop typology under measurement uncertainty, we conducted a Monte Carlo simulation in which indicator values were perturbed using empirically derived noise. This analysis quantifies the sensitivity of typology recoverability to errors in individual indicators, while holding the remaining indicators unchanged (Saltelli et al. 2008).

For each indicator d (excluding the route count and distance to the nearest rail station), we constructed an empirical noise distribution by converting the observed values into standardized deviations. Let $\{x_{nd}\}_{n=1}^{N}$ denote the observed values of indicator d across stops, with median $m_d = \text{median}(\{x_{nd}\})$ and standard deviation $s_d = \text{sd}(\{x_{nd}\})$. We first computed robust deviations:

$$\tilde{z}_{nd} = \frac{x_{nd} - m_d}{s_d}, \qquad n = 1, \ldots, N \tag{3}$$

and then normalized them to ensure zero mean and unit variance:

$$z_{nd} = \frac{\tilde{z}_{nd} - \bar{\tilde{z}}_d}{\text{sd}(\{\tilde{z}_{nd}\})}. \tag{4}$$

where $\bar{\tilde{z}}_d$ denotes the sample mean of $\{\tilde{z}_{nd}\}$ and median centering reduces sensitivity to skewness and extreme observations. The resulting set $\mathcal{Z}_d = \{z_{nd}\}_{n=1}^{N}$ forms the empirical noise distribution for indicator d, from which terms were sampled with replacement. Sampling from an empirical distribution preserves the higher-order characteristics of the observed data without imposing parametric assumptions (Davison and Hinkley 1997; Saltelli et al. 2008).

For all considered indicators except the non-interaction rate, perturbations were applied multiplicatively. Let x denote the original indicator value and x^* the perturbed value. Then

$$x^* = x(1 + pz), \qquad z \sim \mathcal{Z}_d, \tag{5}$$

where p represents the noise magnitude ($p \in \{0.01, 0.05, 0.1\}$) and z denotes a draw from the empirical distribution $\mathcal{Z}_d$. For the non-interaction rate, perturbations were applied in logit space to preserve the bounded nature of the variable between 0 and 1, ensuring that perturbed values remained within the feasible range:

$$x^* = \text{logit}^{-1}(\text{logit}(x) + pz)\,. \tag{6}$$

For each indicator d and noise level p, we generated datasets in which only the values of indicator d were perturbed, while all remaining indicators were kept unchanged. Specifically, if $\mathbf{X}$ denotes the original feature matrix, the perturbed dataset $\mathbf{X}^*_{(d)}$ was obtained by replacing column d with the perturbed values x^* computed using the formulations above while leaving all other columns intact. This procedure isolates the effect of uncertainty in a single indicator on typology recoverability.

The model was then retrained on each perturbed dataset across $R = 200$ Monte Carlo replications. Model performance was compared against the baseline accuracy obtained from the model trained on the original data.

Results

Type Profiles

The typology analysis revealed four distinct bus stop types. These types differ in ridership levels, headway patterns, dwell times, and their role within the multimodal network. Some stops emerge as high-volume destinations or transfer points, while others function as feeders to rail or serve as local neighborhood access points. The results show that bus stops are not uniform elements of the system but perform a range of functions that influence passenger experience and overall service performance.

To interpret the types, we relied on a set of complementary tools. Figure 6 (CDF plots) illustrates differences in ridership and operational measures such as headways and dwell times across the four types. Comparison of these plots shows that the distribution of indicators differs by bus stop type. In many cases, one type stands apart from the others, revealing defining characteristics of the clusters. Figure 7 combines SHAP summary plots (left) with radar profiles (right): the SHAP plots rank indicators by their contribution to type probability (e.g., transfers, dwell time, and route count emerge as strong discriminators), while the radar profiles show each type's mean standardized levels across all indicators. Coordinated views provide additional context by comparing stops within the broader system, and SHAP values from the Random Forest models identify the features most influential in predicting type membership. Together, these visualizations produce a clear and interpretable typology.

We assigned descriptive labels to the four types of bus stops to reflect their defining roles: *Multi-route Destinations*, *Rail-Adjacent Distributors*, *High-Demand Hubs*, and *Low-Ridership Locals*. These labels emphasize both the observed activity patterns and the contribution of each type to system performance within the multimodal network.

Multi-route Destinations

These stops are marked by high passenger alightings, the convergence of multiple bus routes, and moderate dwell times, though they generally experience relatively low levels of transfer activity. Their defining function is to act as major destinations within the network, where buses deliver riders to employment centers, shopping districts, or key civic and service hubs. The combination of relatively high off-boarding volumes and multiple route connections indicates that they primarily serve as end-point destinations rather than transfer nodes, shaping travel flows toward specific corridors or activity clusters.

Rail-Adjacent Distributors

These stops are characterized by transfer activity from rail passengers and their close proximity to train stations. They are typically served by a relatively high number of bus routes, each operating at short headways, which allows them to accommodate steady flows of passengers transferring from rail. While overall passenger volumes are not as large as at major hubs, their function is specialized: they provide an important link between rail and bus services, redistributing riders arriving by train across multiple bus routes.

High-Demand Hubs

These stops are defined by very high passenger volumes and extended dwell times, reflecting intense turnover. They are served by many routes, and what distinguishes them most is the sheer concentration of ridership. Transfer activity is significant, often creating terminal-like conditions even without formal designation. Unlike *Rail-Adjacent Distributors*, these locations are not usually near rail stations; their demand is generated within the bus network along major corridors or activity centers. The combination of high ridership, overlapping services, and long dwell times underscores their role as *High-Demand Hubs*, where the operational focus is maintaining reliability, managing dwell times, and keeping passenger circulation efficient under concentrated demand.

Low-Ridership Locals

Bus stops of this type are characterized by low passenger activity, short dwell times, and limited transfers. They are generally served by fewer routes, and boarding/alighting volumes remain modest. Their defining role is neighborhood-level accessibility—ensuring coverage in residential or lower-density areas rather than acting as destinations or transfer points.

They connect local riders to the broader network, often serving as the first/last point of contact. Their contribution lies in extending the system's reach, making them essential for coverage-oriented planning even if their operational profile is less intensive.

Spatial Typology Distribution

Figure 8 provides two complementary views of spatial patterns: (A) stop-level type assignments across the network and (B) tract-level dominant types summarizing prevailing roles by area. Panel A shows each studied MBTA stop colored by its assigned type, illustrating how different functional roles are distributed throughout the network. Panel B aggregates these results at the census-tract level, highlighting the dominant cluster type in each area. The clearest patterns

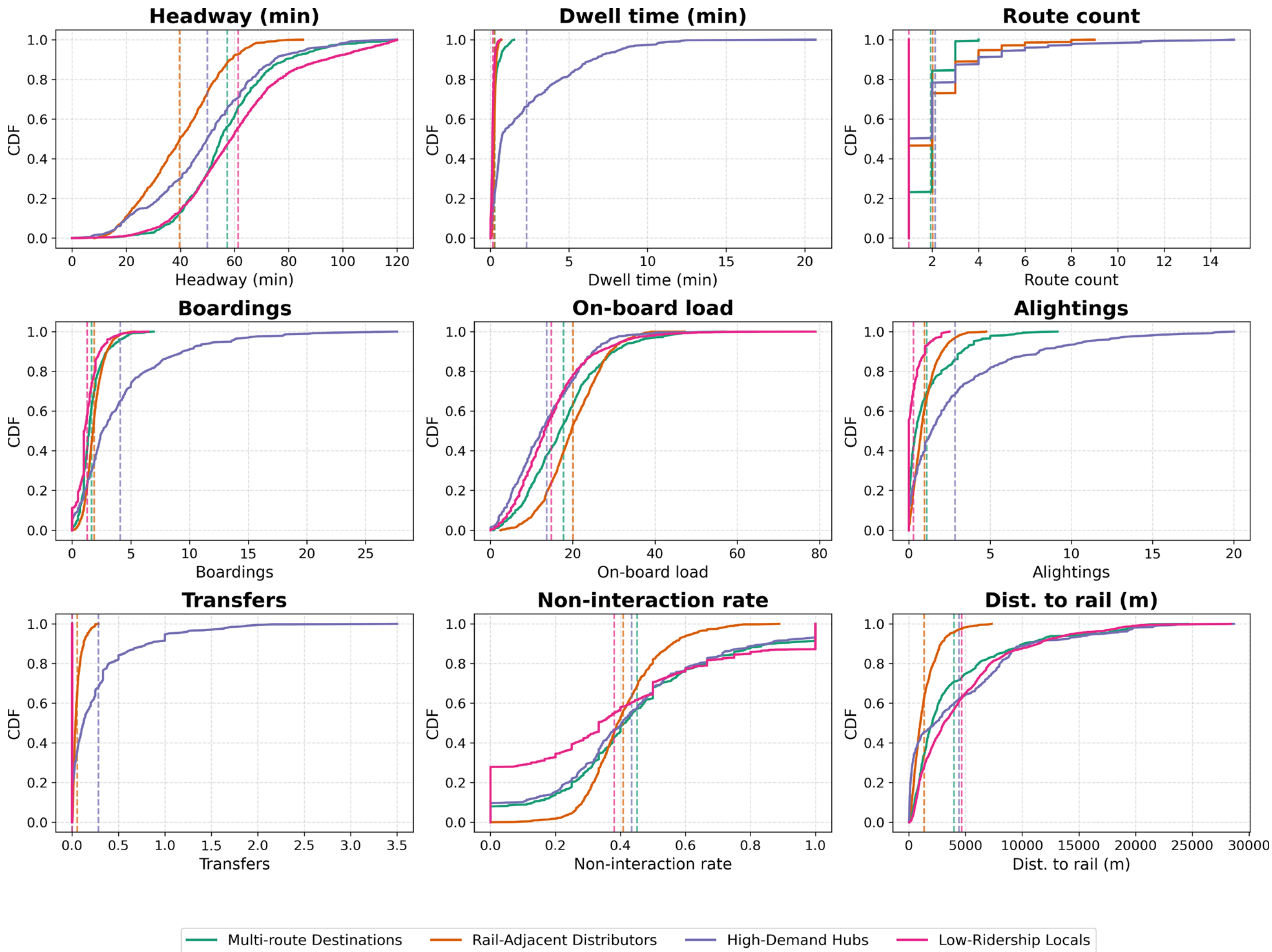


Fig. 6 Cumulative distribution (CDF) plots of the nine stop-level indicators by type. Solid curves show the distribution for each stop type, and dashed vertical lines mark the corresponding mean values

appear for *Rail-Adjacent Distributors*, which concentrate around the downtown core and along major rail corridors, and for *Low-Ridership Locals*, which spread widely across outer neighborhoods and suburban areas. In contrast,*High-Demand Hubs* and *Multi-route Destinations* form strong clusters along high-frequency corridors and key transfer locations, reflecting their intensive operational and passenger activity.

The overall geography of types shows both order and intermixing. Adjacent stops along the same corridor or near the same station can belong to different types, revealing functional variation within short distances. The tract-level summary reinforces this mosaic structure: some areas are dominated by a single role, while others host a mix of complementary types. This pattern suggests that the typology captures underlying service and demand dynamics rather than merely mapping fixed spatial zones—demonstrating how operational context and passenger behavior jointly shape the geography of bus stop functions.

Model Separability

We validated the recoverability of the four clusters using a gradient-boosted tree classifier (XGBoost). The model was trained on the original stop-level indicators using a stratified train/test split, with hyperparameters tuned via grid search. On the held-out test set, the tuned model achieved accuracy = 0.981, macro precision = 0.976, macro recall = 0.963, macro $F1 = 0.968$, macro ROC–AUC = 0.999, and log loss = 0.043.

Figure 9 summarizes the evaluation. Panel A shows the multiclass confusion matrix, which is sharply diagonal: nearly all stops are correctly classified into their true cluster, with only a small number of confusions between the two operationally similar mid-spectrum types (Classes 2 and 3). Classes 1 and 4 exhibit almost perfect separation, reflecting their distinct operational signatures.

Panel B presents one-vs-rest ROC curves, all of which closely hug the upper-left corner. Three clusters achieve an

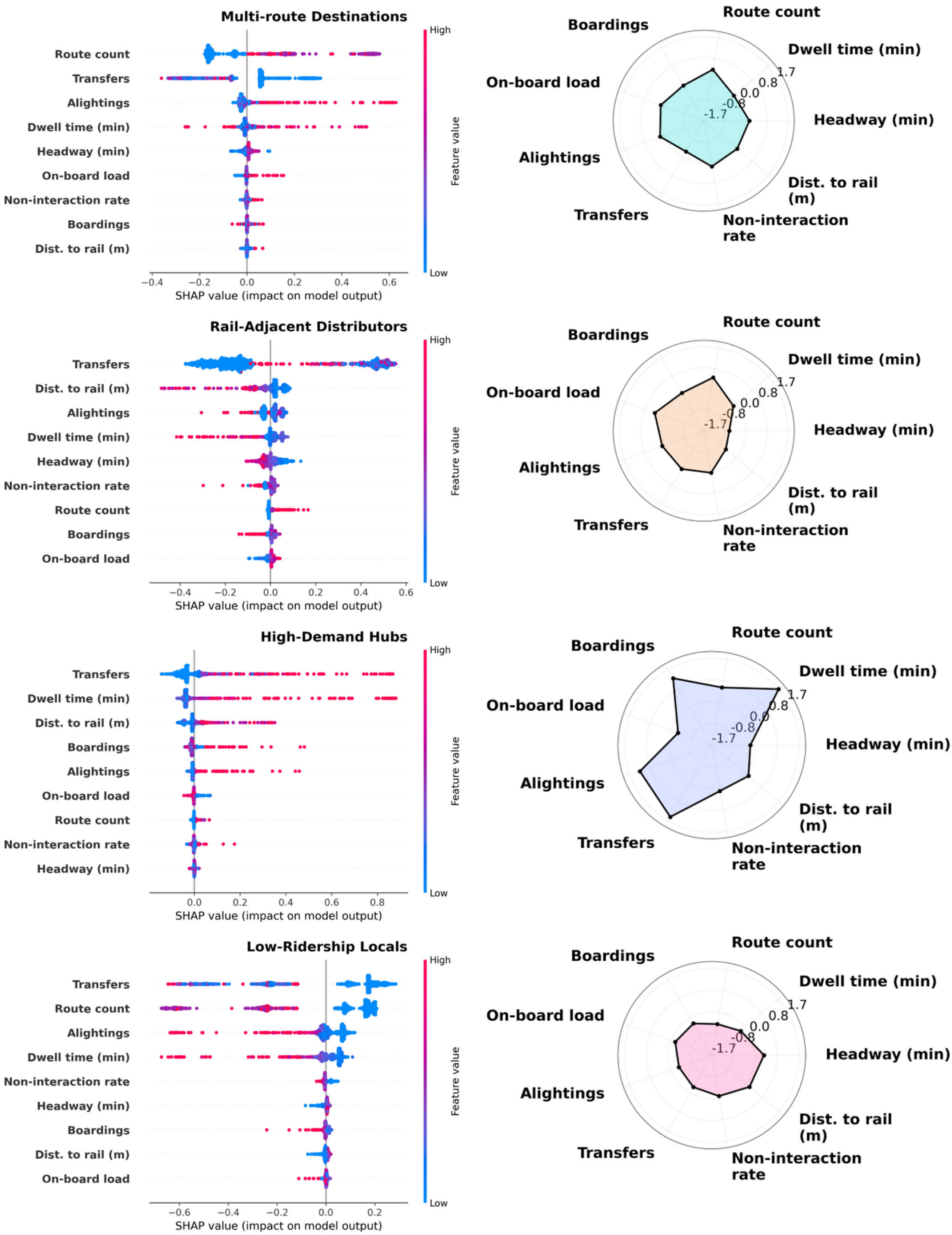


Fig. 7 SHAP plots (left) and radar profiles (right) for the four stop types. SHAP shows each indicator's impact on type probability, sorted by mean |SHAP|; radar plots show mean standardized indicator levels by type

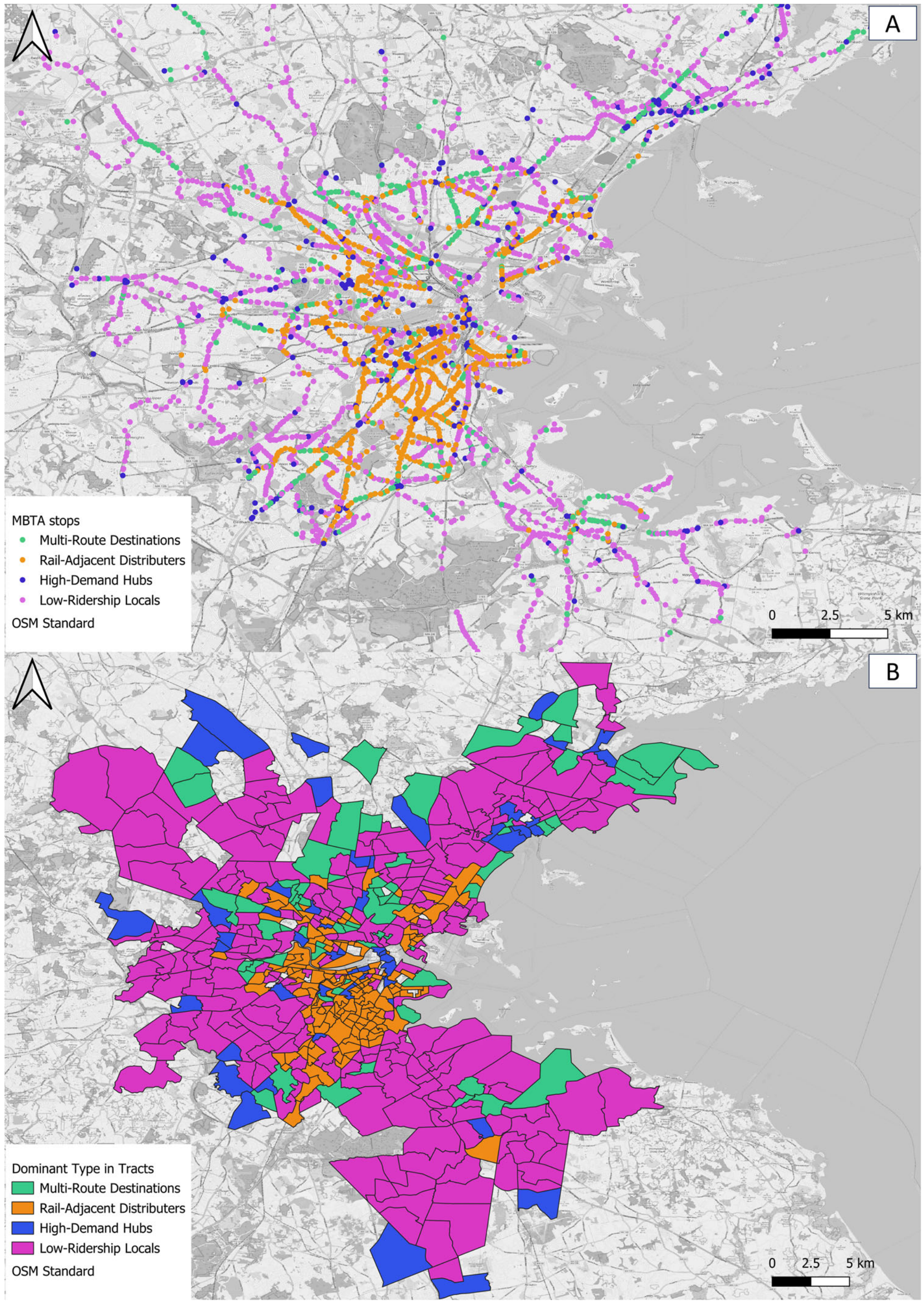


Fig. 8 Spatial distribution of bus stop types. **A** MBTA bus stops classified by typology: each point represents a studied stop colored by its assigned type. **B** Dominant cluster type by census tract, summarizing the prevailing stop role within each area

AUC of 1.000, and the remaining cluster reaches 0.999, confirming extremely strong separability across the four types.

Panel C displays per-cluster SHAP beeswarm plots, which explain the drivers of each type's predictions. The distinguishing features vary across clusters in interpretable ways. For Multi-route Destinations, transfers and route count dominate the decision boundary. Rail-Adjacent Distributors are characterized most strongly by distance to rail and dwell time, with non-interaction rate and headway reinforcing the separation. Low-Ridership Locals are shaped by consistently low levels of ridership indicators (on-board load, boardings, alightings). High-Demand Hubs show strong positive influence from ridership-related indicators—particularly dwell time, alightings, and load—along with elevated transfer volumes.

Panel D shows the overall feature importance based on global SHAP values aggregated across all clusters. Transfers, route connectivity, and dwell-related measures stand out as the strongest contributors to classification, followed by ridership indicators and spatial context variables. This summary

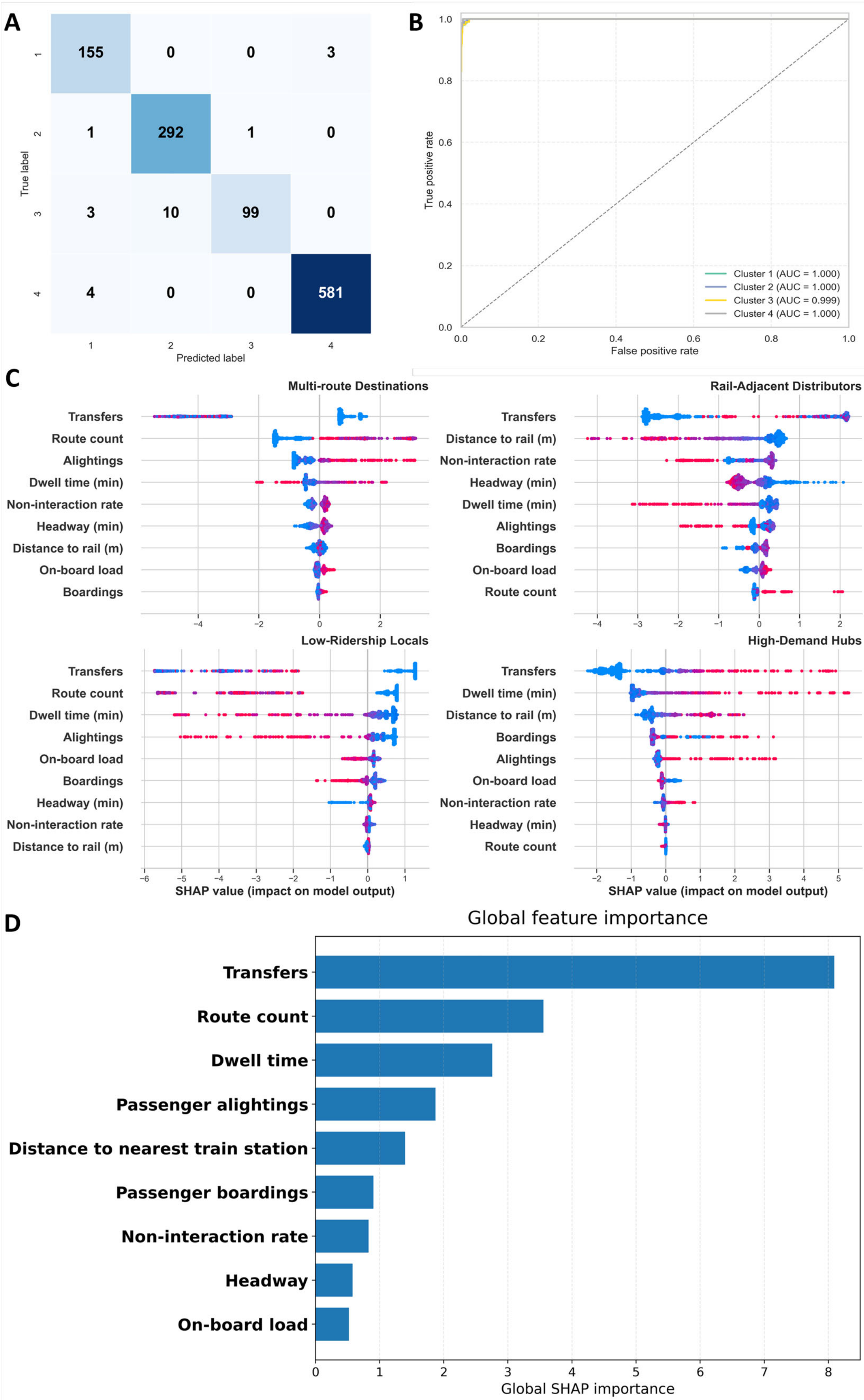


Fig. 9 Typology separability results of the multiclass XGBoost model. **A** Confusion matrix. **B** One-vs-rest ROC curves with AUC values. **C** SHAP summary plots illustrating key features driving each type's classification within the multiclass model. **D** Global SHAP importance ranking (mean absolute SHAP values) across all clusters

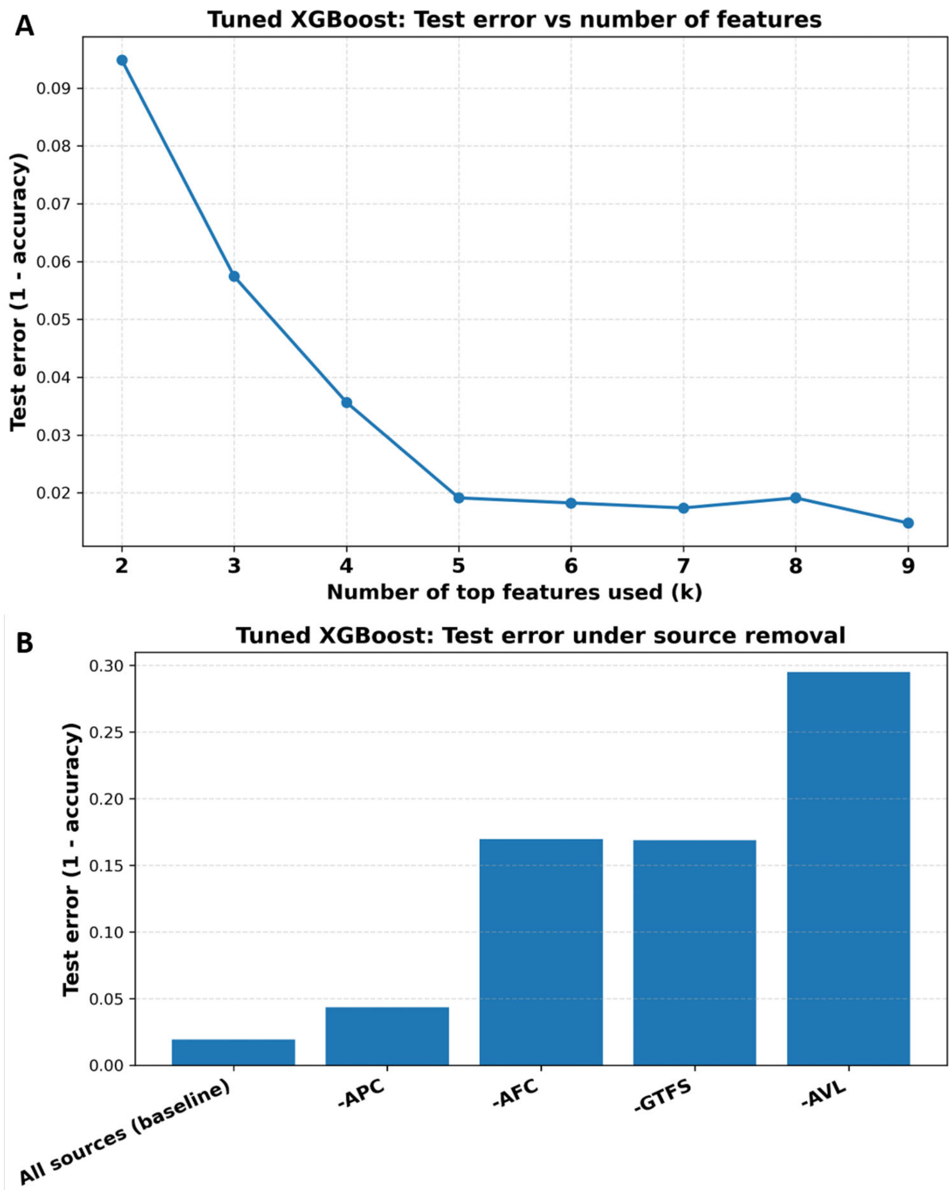


Fig. 10 Robustness of the tuned multiclass XGBoost model under reduced input conditions. **A** Test error as a function of the number of indicators retained (top-*k* by global SHAP importance). **B** Test error when indicators associated with one data source are excluded at a time, based on the mapping in Table 3

complements the cluster-specific patterns shown in Panel C by highlighting the features that matter most across the typology as a whole.

Overall, the boosted classifier confirms that the unsupervised clustering captures meaningful, operationally interpretable patterns rather than artifacts of the grouping process.

Robustness Under Constrained Data Availability

To further assess how the learned stop-type structure behaves under constrained data conditions, we conducted a reduced-input retraining analysis using the multiclass XGBoost validation model. Figure 10 summarizes two complementary experiments. Panel A shows test error as a function of the number of indicators retained (top-*k* by global SHAP importance from Fig. 9D), demonstrating that predictive performance improves rapidly with the inclusion of the most informative indicators and then stabilizes. Panel B shows performance when indicators associated with one data source are excluded at a time, based on the mapping in Table 3. Together, these experiments provide a practical way to evaluate how the typology performs when only partial data are available.

The results also help distinguish core indicators from those that are more supplementary for recovering the typology. Panel A suggests that a relatively small subset of the highest-ranked features—approximately the first five indicators identified through the global SHAP ranking—captures most of the predictive signal, with only modest improvements beyond that point. However, these influential features originate from multiple automated data streams. Consistent

Table 3 Bus stop typology features and data sources, ordered by global SHAP importance

Feature	GTFS/GIS	AVL	AFC	APC
Transfers		✓	✓	
Route count	✓			
Dwell time	✓	✓		
Passenger alightings		✓		✓
Distance to nearest train station	✓			
Passenger boardings		✓		✓
Non-interaction rate		✓	✓	✓
Headway	✓	✓		
On-board load		✓		✓

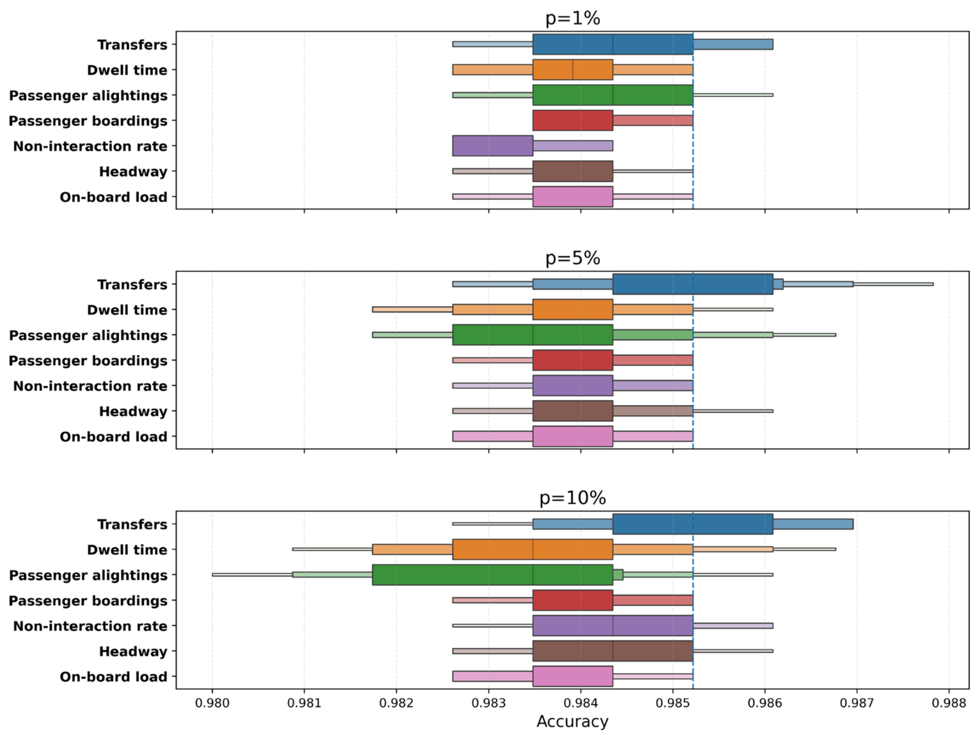


Fig. 11 Model accuracy under simulated indicator uncertainty. Boxen plots show the distribution of accuracy across 200 Monte Carlo replications for each indicator at noise magnitudes of 1%, 3%, and 5%. The dashed vertical line indicates baseline accuracy

with this, Panel B shows that excluding different sources leads to varying levels of performance degradation, indicating that full coverage of the most important predictors ultimately requires access to the major feeds. In practical terms, agencies with integrated AVL, AFC, APC, and schedule or GIS data are best positioned to compute the complete feature set and achieve near-baseline performance, while the reduced-input results provide guidance on expected robustness in environments with more limited data availability.

Robustness to Indicator Uncertainty

We further evaluated typology stability under simulated measurement uncertainty using the Monte Carlo noise-injection framework described in Sect. "Assessing Typology Separability". Figure 11 summarizes model accuracy across 200 replications for each indicator under representative noise magnitudes (1%, 3%, and 5%). Overall, the classifier performance remains highly stable across all indicators and noise levels considered. Median accuracy changes are on the order of 10^{-3}, indicating negligible degradation relative to the baseline model trained on the original data.

Some indicators exhibit slightly greater sensitivity than others, particularly ridership-related measures and dwell-related variables, but the magnitude of the effect remains small even under higher noise levels. Importantly, no indicator produces substantial deterioration in classification performance, confirming that the learned typology is robust to realistic levels of measurement uncertainty in the underlying operational data streams.

Table 4 Type-specific policies and rationale: operational control, curb/stop design, spacing/assignments, and safety actions aligned to each role

Type	Policy dimension	Brief explanation
Multi-route Destinations	Operations	Manage as small hubs—design multi-bay configurations and assign berths/bays by route family and throughput. Control bunching using headway-based holding (capped) and conditional TSP (Transit Signal Priority) to stabilize arrivals and departures (Bartholdi and Eisenstein 2012; Daganzo and Pilachowski 2011; Smith et al. 2005)
	Curbside design & safety	Keep the bus curb clear—relocate micromobility and PUDO to short side pockets; provide protected crossings and continuous ADA-compliant sidewalks sized for higher alightings and safe circulation, and integrate micromobility access, where it supports first- and last-mile connectivity (Mitman et al. 2020; Architectural and Transportation Barriers Compliance Board 2023; Tabassum et al. 2025; Ewing et al. 2024)
	Stop spacing & assignments	When multi-route stops are closely spaced, selectively consolidate or relocate stops and adjust route assignments to cut friction delays and balance passenger loads; evidence shows that clustered interlined stops increase dwell and imbalance, while consolidation improves performance (NACTO 2016; Stewart and El-Geneidy 2016; Devunuri et al. 2024)
Rail-Adjacent Distributors	Transfer design	Minimize rail-to-bus walking distance and ensure safe, direct, ADA-compliant, weather-protected paths. Provide intuitive wayfinding and real-time next-bus and rail headway information to improve transfer predictability (NACTO 2016; Architectural and Transportation Barriers Compliance Board 2023; Liu and Miller 2020)
	Capacity & dwell management	Provide platforms and shelters sized to absorb rail-arrival waves and maintain safe, comfortable waiting space; on high-transfer routes, off-board fare payment and all-door boarding help reduce average dwell times and stabilize operations during surges (Tirachini 2013; Fletcher and El-Geneidy 2013)
	Service coordination	Where feasible, coordinate bus departures with rail schedules (accounting for platform-to-stop walking time); otherwise regulate departures using holding or self-coordinating headway control to maintain regular service and suppress bunching (Bartholdi and Eisenstein 2012; Daganzo and Pilachowski 2011)
High-Demand Hubs	Speed & reliability	Implement off-board fare collection and all-door boarding to minimize dwell time. Add near-hub bus-only approaches and conditional TSP to avoid throat gridlock and stabilize travel times (Smith et al. 2005; NACTO 2016; Tirachini 2013)
	Crowd management	Widen platforms and size waiting areas for peak passenger loads; channel queues to reduce conflicts; provide clear line signage and integrated real-time information; staff high-density periods to maintain safe circulation, supported by passenger-density monitoring technologies where applicable (TCQSM 2013; NACTO 2016; Darsena et al. 2022)
	Service design & spacing	At multi-bay locations, use dispatch control and load-balanced bay assignments; consider banked departures where applicable; review upstream and downstream stop spacing to reduce friction delay. (TCQSM 2013; Tirachini 2014; NACTO 2016)
Low-Ridership Locals	Spacing & speed	Use selective stop consolidation where access impacts are minimal to improve run time and reliability while maintaining coverage (NACTO 2016; Devunuri et al. 2024)
	Service quality	Emphasize headway regularity and on-time performance over raw frequency; install real-time signage only where observed usage warrants and enhances predictability (TCQSM 2013)
	Safety & security	Improve safety at low-ridership stops by ensuring good visibility, adequate nighttime lighting, and clearly marked crossings; empirical assessments show that poor sight distance, weak lighting, and inadequate pedestrian facilities around bus stops are linked to higher crash risk (NACTO 2016; Cheranchery et al. 2019; Ulak et al. 2021)

TSP transit signal priority, *PUDO* pick-up/drop-off, *ADA* Americans with Disabilities Act

Discussion and Policy Implications

The distinct functional roles identified in our typology enable us to tailor type-specific policy and design strategies. Stops with high passenger activity, such as *High-Demand Hubs*, require measures that manage dwell time, boarding efficiency, and crowd circulation, while *Low-Ridership Locals* benefit more from coverage-oriented reliability and accessible design features. These contrasting needs highlight why a uniform, one-size-fits-all approach to stop management is ineffective and why differentiated interventions are essential for improving both efficiency and equity across the network.

Accordingly, our policy guidance (summarized in Table 4) organizes the recommended actions into broad categories—operations, curb and stop design, spacing and assignments, and safety and access. The implications are derived from established transit operations research and industry guidance, linking each stop's observed role to the most relevant evidence-based strategies for improving reliability, passenger experience, and overall system performance.

Looking at the bus stop type classification results and their spatial distribution in Fig. 8, dominant cluster types are evident across much of the city, revealing clear spatial patterns. Yet, there are also areas, where different stop roles sit very close together. In these locations, the mix of functions can produce uneven ridership and operational patterns across neighboring stops, contributing to long dwell times, bus bunching, and difficult transfers. Recognizing these differences of roles enables targeted adjustments to design, stop spacing, headway management, and route assignments to rebalance loads and improve reliability. Importantly, the framework also supports scenario testing of potential changes, helping agencies anticipate how stop roles may shift before implementation.

Conclusion

This paper introduces a data-driven, interpretable bus stop typology derived from features that reveal each stop's functional role in operations, ridership and transfers, and its position within the multimodal network. While some elements of these roles may be correlated with simpler contextual characteristics, such as land use, proximity to rail, or aggregate boarding levels, the proposed typology captures how stops function through the interaction of multiple operational and behavioral dimensions. This functional perspective enables distinctions among stops that would not be apparent from static or single-criterion classifications alone.

Using typology-informed insights, agencies can make better decisions with regard to bus stop operations (e.g., headway management, stop spacing), design (e.g., bay assignments, shelter and pad sizing, wayfinding), and safety (e.g., protected crossings, queue management), allowing investments to be prioritized according to the role each stop plays within the network rather than solely its location or demand level.

In practice, agencies can phase in improvements by first testing small operational changes—such as slightly tighter headways or modest stop-level adjustments—before committing to larger design or capital projects. Using the same indicators that underlie the typology, they can periodically reassess stop roles to evaluate whether these changes shift stops in the intended direction and confirm that observed benefits align with policy goals. The approach is intended to be transferable: other agencies may apply the framework using their own data and, where possible, augment it with additional features to support more targeted decisions. The reduced-input retraining analysis further suggests that stop types can be assigned for new observations within the same system using a limited subset of the highest-ranked indicators—approximately the first five features identified in the global importance analysis. Because these influential indicators draw on multiple automated data streams, agencies with access to the data required to compute them are typically also able to derive the full feature set and implement the typology.

Appendix A Glossary of Acronyms

Table 5 defines the acronyms used throughout the paper.

Table 5 Glossary of acronyms used in the paper

Acronym	Definition
AFC	Automated Fare Collection
APC	Automatic Passenger Counter
AVL	Automatic Vehicle Location
GTFS	General Transit Feed Specification
MBTA	Massachusetts Bay Transportation Authority
ODX	Origin–Destination–Transfer table
OD	Origin–Destination
GMM	Gaussian Mixture Model
RF	Random Forest
ANN	Artificial Neural Network
SVM	Support Vector Machine
RBF	Radial Basis Function
SHAP	Shapley Additive exPlanations
ROC–AUC	Area Under the Receiver Operating Characteristics Curve
TSP	Transit Signal Priority
PUDO	Pick-up/Drop-off
ADA	Americans with Disabilities Act

Table 6 Hyperparameter grid for the GMM clustering sweep

Parameter	Values explored	Notes
`n_components`	{2,3,...,14}	BIC evaluated for each k
`covariance_type`	{full, tied, diag, spherical}	Standard covariance families
`n_init`	5	Multiple initializations for stability
`random_state`	42	Ensures reproducibility

Table 7 Grid for the one-vs-rest Random Forest models used in the SHAP analysis

Parameter	Values	Criterion
`n_estimators`	{100, 200}	Highest mean CV F1
`max_depth`	{`None`, 10, 20}	Highest mean CV F1
`min_samples_split`	{2, 5}	Highest mean CV F1
`min_samples_leaf`	{1, 2}	Highest mean CV F1
`class_weight`	{`None`, `balanced`}	Highest mean CV F1

Appendix B Model Hyperparameter Grids

B.1 GMM Clustering Model

To select the number of clusters and covariance structure, we performed a grid sweep over all combinations of covariance types and component counts. For each configuration, the model was initialized with standardized indicators and fitted using `n_init = 5`. Bayesian Information Criterion (BIC) values were computed for all models, and an elbow region between $k = 3$ and $k = 5$ guided selection.

Table 6 summarizes the grid used in the clustering analysis.

The BIC elbow plot corresponding to this grid search is included in Fig. 5.

B.2 One-vs-Rest Random Forest Models

For type interpretation, we trained four separate one-vs-rest Random Forest classifiers. The grid in Table 7 was applied to each binary problem, and the best configuration was chosen based on mean validation $F1$ score.

B.3 Multiclass Classification Models

To test separability of the four clusters, we trained multiple multiclass classifiers. Each model was tuned with threefold stratified cross-validation. Table 8 lists all hyperparameter grids.

Table 8 Hyperparameter grids for multiclass classification models

Model	Parameter	Values
Random Forest	`n_estimators`	{300, 600}
	`max_depth`	{`None`, 12, 20}
	`min_samples_split`	{2, 5}
	`min_samples_leaf`	{1, 2}
Logistic Regression	`C`	{0.1, 1.0, 5.0}
	`penalty`	{`l1`, `l2`}
ANN (MLP)	`hidden_layer_sizes`	{(64,), (128,), (128, 64)}
	`alpha`	{1e−4, 1e−3}
	`learning_rate_init`	{1e−3, 5e−4}
SVM (RBF)	`C`	{0.5, 1.0, 5.0}
	`gamma`	{`scale`, 0.1, 0.01}
XGBoost	`max_depth`	{3, 5}
	`learning_rate`	{0.1, 0.05}
	`subsample`	{0.8, 1.0}
	`colsample_bytree`	{0.8, 1.0}
LightGBM	`learning_rate`	{0.1, 0.05}
	`max_depth`	{− 1, 10}
	`num_leaves`	{31, 63}
	`subsample`	{0.8, 1.0}

Acknowledgements This work was funded in part by the Massachusetts Department of Transportation Office of Transportation Planning (OTP). We thank the staff of MassDOT and MBTA for providing access to relevant data.

Author Contributions M.A. designed the study, processed and integrated the MBTA AFC/APC/GIS data, developed the modeling framework, carried out all analyses, and wrote the initial manuscript draft. J.O. contributed to the methodological design, advised on the clustering and machine learning approach, assisted in interpreting the results, and revised the manuscript. E.J.G. contributed to the study design, provided domain expertise on transit operations and planning, helped interpret the findings and policy implications, and revised the manuscript. All authors discussed the results and implications and approved the final manuscript.

Data Availability The data used in this study were obtained from the Massachusetts Bay Transportation Authority (MBTA) under a confidentiality agreement and cannot be shared publicly. However, the full analysis workflow and Python code are available at: https://github.com/Mahdi-Azhdari/bus-stop-typology.

Declarations

Conflict of interest The authors declare no competing interest.

References

Abdalazeem M, Oke J (2023) Extracting spatiotemporal bus passenger trip typologies from noisy mobile ticketing boarding data. Data Sci Transp 5(3):20

APTA (2010) Bus stop design and placement: security considerations. Technical Report APTA SS-SIS-RP-008-10, American Public Transportation Association (APTA) Transit Infrastructure Security Working Group, Washington, D.C. Approved June 26, 2010. https://www.apta.com/wp-content/uploads/Standards_Documents/APTA-SS-SIS-RP-008-10.pdf

Architectural and Transportation Barriers Compliance Board (2023) Accessibility Guidelines for Pedestrian Facilities in the Public Right-of-Way. Federal Register, 88(151):53604–53664. Final Rule, 36 CFR Part 1190. https://www.access-board.gov/prowag/

Baek J, Sohn K (2016) Deep-learning architectures to forecast bus ridership at the stop and stop-to-stop levels for dense and crowded bus networks. Appl Artif Intell 30(9):861–885

Baghbani A, Bouguila N, Patterson Z (2023) Short-term passenger flow prediction using a bus network graph convolutional long short-term memory neural network model. Transp Res Rec 2677(2):1331–1340

Bartholdi JJ III, Eisenstein DD (2012) A self-coördinating bus route to resist bus bunching. Transp Res Part B Methodol 46(4):481–491

BCDCOG (2021) Transit and bus stop design guidelines. Technical Report, Berkeley–Charleston–Dorchester Council of Governments (BCDCOG), Charleston, SC. Prepared for CARTA and TriCounty Link. Accessed 07 Nov 2025. https://www.bcdcog.com/bus_stop_design_guidelines/

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Cats O (2024) Identifying human mobility patterns using smart card data. Transp Rev 44(1):213–243

Cheng Y, Hatzichristos T, Kostellou A, Fujiyama T, Argyropoulou K, Spyropoulou I (2024) Understanding the intra-day and intra-week ridership patterns of urban rail transit stations in London using a fuzzy clustering approach. J Public Transp 26:100099

Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 785–794

Cheranchery MF, Bhattacharyya K, Salih M, Maitra B (2019) A proactive approach to assess safety level of urban bus stops. Int J Inj Contr Saf Promot 26(3):260–270

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297

Cox DR (1958) The regression analysis of binary sequences. J R Stat Soc Ser B Stat Methodol 20(2):215–232

Daganzo CF, Pilachowski J (2011) Reducing bunching with bus-to-bus cooperation. Transp Res Part B Methodol 45(1):267–277

Darsena D, Gelli G, Iudice I, Verde F (2022) Sensing technologies for crowd management, adaptation, and information dissemination in public transportation systems: a review. IEEE Sens J 23(1):68–87

Davison AC, Hinkley DV (1997) Bootstrap methods and their application, vol 1. Cambridge University Press, Cambridge, UK

Degeler V, Heydenrijk-Ottens L, Luo D, Van Oort N, Van Lint H (2021) Unsupervised approach towards analysing the public transport bunching swings formation phenomenon. Public Transport 13(3):533–555

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B (Methodol) 39(1):1–22

Devillaine F, Munizaga M, Trépanier M (2012) Detection of activities of public transport users by analyzing smart card data. Transp Res Rec 2276(1):48–55

Devunuri S, Lehe LJ, Qiam S, Pandey A, Monzer D (2024) Bus stop spacing statistics: theory and evidence. J Public Transp 26:100083

Egu O, Bonnel P (2020) Can we estimate accurately fare evasion without a survey? Results from a data comparison approach in Lyon using fare collection data, fare inspection data and counting data. Public Transport 12(1):1–26

Ewing R, Kalantari HA, Yang W et al (2024) Increasing public transportation ridership by integrated transportation planning with micromobility services in the state of Utah. Technical Report, Utah. Dept. of Transportation. Research Division

Fitzpatrick K, Hall K, Perkinson D, Nowlin L, Koppa R (1996) Guidelines for the location and design of bus stops. Technical Report TCRP Report 19, Transit Cooperative Research Program, Transportation Research Board, Washington, DC. https://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_19-a.pdf

Fletcher G, El-Geneidy A (2013) Effects of fare payment types and crowding on dwell time: fine-grained analysis. Transp Res Rec 2351(1):124–132

Ge L, Sarhani M, Voß S, Xie L (2021) Review of transit data sources: potentials, challenges and complementarity. Sustainability 13(20):11450

Gonzales EJ, Gao S, Azhdari M (2024) Measuring fare payment compliance on MBTA buses. Technical Report 24-065, Massachusetts Department of Transportation. https://rosap.ntl.bts.gov/view/dot/82466

Li T, Sun D, Jing P, Yang K (2018) Smart card data mining of public transport destination: a literature review. Information 9(1):18

Lin JY-J, Chen C, Angah O (2024) Socio-economic and spatial disparity of bus ridership impacts in King County, Washington, during COVID-19. Transp Res Rec 2678(12):295–308

Liu L, Miller HJ (2020) Does real-time transit information reduce waiting time? An empirical analysis. Transp Res Part A Policy Pract 141:167–179

Liu D, Guo J, Gu Y, King M, Han LD, Brakewood C (2025) Analyzing transit systems using general transit feed specification (GTFS) by generating spatiotemporal transit networks. Information 16(1):24

Lu K, Liu J, Zhou X, Han B (2020) A review of big data applications in urban transit systems. IEEE Trans Intell Transp Syst 22(5):2535–2552

Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 30:4765–4775

Mitman MF, Davis S, Ballús Armet I, Knopf E (2020) Curbside management practitioners guide. Technical Report, Institute of Transportation Engineers, Washington, DC. Produced by the ITE Complete Streets Council. https://www.ite.org/pub/C75A6B8B-E210-5EB3-F4A6-A2FDDA8AE4AA

Montoya JA, Escobar DA, Moncada CA (2024) Public transportation through the clustering analysis of bus stops. Case study Manizales, Colombia. Geo J Tour Geosites 56(4):1448–1453

NACTO (2016) Transit street design guide. Island Press, Washington

Olivo A, Maternini G, Barabino B et al (2019) Empirical study on the accuracy and precision of automatic passenger counting in European bus services. Open Transp J 13(1):250–260

Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 65(6):386

Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2008) Global sensitivity analysis: the primer. Wiley, West Sussex, UK

Smith HR, Hemily B, Ivanovic M (2005) Transit signal priority (TSP): a planning and implementation handbook. Handbook, U.S. Department of Transportation, Federal Transit Administration, ITS Joint Program Office, Washington, D.C. https://nacto.org/wp-content/uploads/2016/02/transit_signal_priority_handbook_smith.pdf

Stewart C, El-Geneidy A (2016) Don't stop just yet! A simple, effective, and socially responsible approach to bus-stop consolidation. Public Transport 8(1):1–23

Tabassum N, Kalantari HA, Kaniewska J, Ameli SH, Ewing R, Yang W, Promy NS (2025) Ways of increasing transit ridership-lessons learned from successful transit agencies. Case Stud Transport Policy 19:101362. https://doi.org/10.1016/j.cstp.2024.101362

TCQSM (2013) Transit capacity and quality of service manual, 3rd edn. The National Academies Press, Washington, DC. https://doi.org/10.17226/24766 . https://nap.nationalacademies.org/catalog/24766/transit-capacity-and-quality-of-service-manual-third-edition

TCRP (2006) Using archived AVL–APC data to improve transit performance and reliability. Technical Report Report 113, Transit Cooperative Research Program (TCRP), Washington, D.C. Prepared by Cambridge Systematics, Inc., Kittelson & Associates, Inc., and Texas A&M Transportation Institute. https://nap.nationalacademies.org/catalog/13932/using-archived-avl-apc-data-to-improve-transit-performance-and-reliability

Tirachini A (2013) Bus dwell time: the effect of different fare collection systems, bus floor level and age of passengers. Transp A Transport Sci 9(1):28–49

Tirachini A (2014) The economics and engineering of bus stops: spacing, design and congestion. Transp Res Part A Policy Pract 59:37–57

Ulak MB, Kocatepe A, Yazici A, Ozguven EE, Kumar A (2021) A stop safety index to address pedestrian safety around bus stops. Saf Sci 133:105017

Wang Z, Pel AJ, Verma T, Krishnakumari P, Brakel P, Van Oort N (2022) Effectiveness of trip planner data in predicting short-term bus ridership. Transp Res Part C Emerg Technol 142:103790

Welch TF, Widita A (2019) Big data in public transportation: a review of sources and methods. Transp Rev 39(6):795–818

Wong E, Khani A (2018) Transit delay estimation using stop-level automated passenger count data. J Transp Eng Part A Syst 144(3):04018005

Wu L, Kang JE, Chung Y, Nikolaev A (2019) Monitoring multimodal travel environment using automated fare collection data: data processing and reliability analysis. J Big Data Anal Transp 1(2):123–146

Wu W, Gan A, Cevallos F, Shen LD (2010) Optimization models for prioritizing bus stop facility investments for riders with disabilities. Final Report FTA-FL-04-7104-2010.03, Federal Transit Administration, U.S. Department of Transportation, Washington, D.C. https://www.transit.dot.gov/research-innovation/optimization-models-prioritizing-bus-stop-facility-investments-riders-disabilities