RSSI-Based Passenger Movement Classification for Non-Intrusive Public Transport Monitoring

Author One*, Author Two†, Author Three‡ and Author Four§
Department of Whatever, Whichever University
Wherever
Email: *author.one@add.on.net, †author.two@add.on.net, ‡author.three@add.on.net, §author.four@add.on.net

***Abstract*—Accurate monitoring of passenger flows in public transport is essential for service optimization and network planning, yet traditional counting methods face limitations in coverage, cost, and privacy preservation. This paper presents a novel approach for classifying passenger movements using temporal sequences of WiFi Received Signal Strength Indicator (RSSI) measurements. We introduce a dataset collected in a controlled experimental environment that simulates public transport scenarios, capturing the distinctive signal patterns associated with four fundamental movement classes: boarding the vehicle, alighting from the vehicle, remaining inside, and remaining at the bus stop. By analyzing the temporal evolution of RSSI values over ten-second observation windows, our approach enables non-intrusive distinction between static and transitional states without requiring specialized hardware or compromising passenger anonymity. Experimental evaluation using multiple machine learning classifiers demonstrates the feasibility of RSSI-based movement classification, providing a cost-effective complement to existing automatic passenger counting systems for intelligent transportation applications.**

***Index Terms*—Passenger counting, RSSI fingerprinting, WiFi sensing, public transport, machine learning, intelligent transportation systems, urban mobility**

I. Introduction

Urban mobility represents one of the most pressing challenges facing modern cities, as more than 75% of European Union citizens currently reside in urban areas [1]. Inefficient transportation systems contribute approximately 24% of greenhouse gas emissions, underscoring the urgent need for intelligent and sustainable mobility solutions [2]. Public transport networks play a pivotal role in addressing these challenges; however, their effective management requires accurate and continuous data regarding passenger flows across the network. Real-time passenger occupancy information supports dynamic scheduling and resource allocation, enabling transport operators to enhance service quality while contributing to sustainable urban development [3].

Traditional approaches to passenger counting and origin-destination (OD) matrix estimation rely on manual surveys, ticketing systems, or dedicated hardware such as infrared sensors and automated passenger counting (APC) devices [4]. While these methods have proven useful, they present significant limitations including high deployment costs, incomplete spatial coverage, privacy concerns associated with video-based systems, and the inability to provide continuous, non-intrusive monitoring of passenger movements [5]. Furthermore, conventional ticketing systems often fail to capture the complete passenger journey, as exit validation is frequently absent in many public transport networks [6]. Recent comprehensive reviews of passenger counting technologies have identified wireless sensing as a particularly promising direction, combining cost-effectiveness with privacy preservation [7].

The proliferation of personal wireless devices, particularly smartphones, has opened new avenues for non-intrusive passenger detection and tracking. WiFi-enabled devices continuously emit probe requests and maintain connections with nearby access points, generating Received Signal Strength Indicator (RSSI) signatures that can be leveraged for localization and movement pattern recognition [8]. Unlike camera-based systems, WiFi sensing approaches inherently preserve passenger anonymity while providing valuable insights into mobility patterns [9]. This characteristic is particularly relevant in the context of increasing privacy regulations, as wireless signal analysis enables presence monitoring without collecting personally identifiable visual data.

Recent advances in Channel State Information (CSI) extraction have further enhanced the capabilities of WiFi-based passenger counting systems, achieving remarkable accuracy rates exceeding 94% in controlled environments through cooperative sensing with multiple receivers [10]. However, CSI-based approaches require specialized hardware for channel estimation and are computationally intensive, limiting their practical deployment in resource-constrained settings. In contrast, RSSI-based methods offer a more accessible alternative, requiring only standard networking equipment while still providing meaningful signal fingerprints for classification tasks [11]. The inherent trade-off between sensing granularity and deployment practicality motivates the exploration of RSSI-based solutions for large-scale transit applications.

This paper presents a novel approach to passenger boarding and alighting classification using temporal sequences of RSSI measurements. The proposed methodology enables the distinction between four fundamental movement patterns: remaining inside the vehicle, remaining at the bus stop, boarding the vehicle, and alighting from the vehicle. By analyzing the temporal evolution of signal strength over a ten-second observation window comprising ten sequential measurements, our approach captures the distinctive signatures associated with each movement pattern without requiring precise localization or continuous tracking of individual devices.

The main contributions of this work are threefold. First, we introduce a publicly available dataset specifically designed for RSSI-based passenger movement classification, collected in a controlled experimental environment that simulates real-world public transport scenarios with both isolated and noisy conditions. Second, we present a comprehensive analysis of temporal RSSI patterns across different movement classes, demonstrating the feasibility of distinguishing between static and transitional states using standard WiFi access point equipment. Third, we evaluate multiple machine learning classifiers for this task, providing insights into the most effective approaches for real-time passenger flow estimation in practical deployments.

The remainder of this paper is organized as follows. Section II reviews related work in passenger counting, WiFi-based sensing, and RSSI fingerprinting. Section III describes the experimental setup and data collection methodology. Section V presents the classification results, followed by a discussion in Section VI. Finally, Section VIII concludes the paper and outlines directions for future work.

II. Related Work

The challenge of accurately monitoring passenger flows in public transport has attracted considerable research attention, resulting in diverse technological approaches ranging from dedicated sensing hardware to opportunistic wireless signal analysis. This section reviews the most relevant contributions across three interconnected domains.

A. Automatic Passenger Counting Systems

Traditional Automatic Passenger Counting (APC) systems primarily rely on infrared sensors, pressure mats, or video-based detection installed at vehicle entrances [4]. Pronello and Garzón Ruiz [5] evaluated commercial video-based APC systems under real-world conditions, revealing that claimed accuracy rates of 98% often deteriorate to 53–74% in practice. Computer vision approaches have advanced with deep learning, with Wiboonsiriruk et al. [12] achieving 94% accuracy using object detection and tracking. However, vision-based systems remain constrained by occlusion, lighting variations, and privacy concerns regarding visual data collection.

B. WiFi-Based Sensing for Passenger Detection

The ubiquity of WiFi-enabled devices has motivated research into wireless signal analysis for mobility monitoring. Myrvoll et al. [13] pioneered WiFi signatures for public transport passenger counting through probe request analysis. Nitti et al. [8] developed iABACUS, achieving 100% detection accuracy in static scenarios and approximately 94% in dynamic conditions.

Channel State Information (CSI) offers richer information than RSSI alone. Guo et al. [10] proposed an RSSI-assisted CSI-based counting system achieving accuracy exceeding 94% through adaptive feature fusion. However, CSI requires specialized hardware and is computationally demanding, whereas RSSI is readily available from standard equipment. Fabre et al. [14] compared machine learning algorithms for WiFi-based ridership estimation, finding Light Gradient Boosting Machine provided accurate boarding and alighting predictions. Simončič et al. [9] developed a non-intrusive WiFi detection system achieving over 96% accuracy despite MAC address randomization.

C. RSSI Fingerprinting and Movement Classification

RSSI fingerprinting has been extensively studied for indoor localization [11]. Recent work extends RSSI analysis to trajectory and movement classification. Wang et al. [15] proposed treating continuously measured RSSI values as temporal sequences, aligning with our methodology of leveraging signal strength evolution over time. Servizi et al. [16] addressed bus boarding and alighting detection using smartphone-based Bluetooth sensing, highlighting the complexity of distinguishing transitional states. Cerqueira et al. [6] demonstrated the importance of understanding complete passenger journey patterns for origin-destination matrix inference.

D. Research Gap and Contribution

While significant progress has been made in both passenger counting and wireless signal-based sensing, the specific application of machine learning to classify passenger movement patterns from RSSI time series remains underexplored. Existing WiFi-based approaches have primarily focused on aggregate counting or statistical inference, rather than developing classifiers that can distinguish fine-grained movement patterns such as boarding versus alighting.

Our work addresses this gap by framing passenger movement detection as a supervised classification problem, where the temporal evolution of RSSI values over a short observation window serves as the input feature vector. This approach enables real-time classification of individual device trajectories, which can subsequently be aggregated to estimate passenger flows and contribute to origin-destination matrix construction.

Furthermore, unlike studies that rely on multiple access points or complex sensor fusion, our approach uses a single WiFi access point positioned at the vehicle door, minimizing infrastructure requirements while still achieving discriminative power through the temporal dynamics of signal strength.

III. Experimental Setup

To evaluate the feasibility of RSSI-based passenger movement classification, a series of controlled experiments were conducted in an indoor environment designed to emulate real-world public transport interactions. The experimental setup aimed to reproduce the spatial constraints of a bus and an adjacent bus stop, while enabling reproducible data collection under both isolated and noisy conditions. This section describes the environmental configuration, the data acquisition methodology, and the preprocessing pipeline applied to the collected data.

A. Environmental Configuration

The experimental environment was divided into two distinct zones, as illustrated in Fig. 1, representing the interior of a public transport vehicle and the corresponding boarding area and bus stop. Zone A, designated as the Vehicle Interior, consisted of a closed room used to simulate the interior of a bus. The room provides moderate isolation from external interference, allowing controlled observation of RSSI variations. A WiFi access point (AP) was installed inside the room, positioned adjacent to the doorway, replicating a realistic placement of onboard communication equipment near the vehicle entrance. Zone B, representing the Bus Stop, was established in the corridor immediately outside the room and was designated as the boarding area. This zone represents the external environment where passengers wait before boarding or after alighting from the vehicle.

The physical separation imposed by the wall and door between the two zones introduces signal attenuation, generating distinctive RSSI patterns when a device transitions between Zone A and Zone B. These characteristics are essential for replicating real-world conditions and for enabling the discrimination between static passenger states and transitional movements.

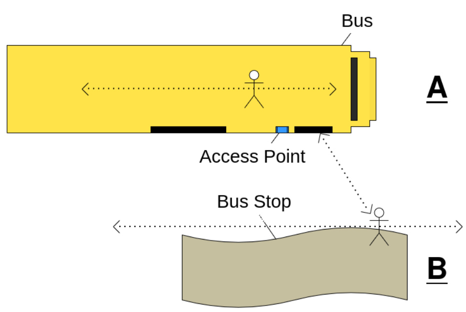

Figure 1. Experimental environment used to simulate a public transport scenario.

B. Data Acquisition Methodology

Data collection was performed in real time using a custom Python-based extraction script. The script interfaced directly with the WiFi access point via an Ethernet connection, enabling continuous retrieval of network metadata associated with connected devices.

Each experimental trial was conducted over a **10-second observation window**, during which **10 consecutive RSSI samples** were recorded at a sampling rate of **1 Hz**, such that each sample was acquired one second apart. This temporal resolution was selected to capture the dynamic evolution of signal strength as a subject moves within or between zones, over a time interval representative of typical boarding or alighting actions in public transport, while remaining compatible with real-time processing constraints.

To ensure consistent RSSI reporting during data collection, the mobile devices generated periodic low-overhead network traffic (ICMP echo requests) toward the access point. This approach maintained active communication with the AP while introducing negligible additional interference.

For each detected device, three primary attributes were extracted: the **MAC Address** served as a unique device identifier, used solely for separating different devices during preprocessing, the **RSSI** (dBm), representing the Received Signal Strength Indicator, served as the primary feature for movement classification and **Traffic Metadata**, consisting of transmitted and received byte counters, was collected by the AP but not retained for model training.

To account for hardware heterogeneity, four different mobile devices, from three separate manufacturers and different generations, were used throughout the data collection process, introducing variability in antenna characteristics and transmission power. This diversity improves the robustness of the resulting dataset and reduces device-specific bias.

C. Movement Classes and Experimental Scenarios

Four fundamental movement classes were defined to cover all possible passenger state transitions relative to the vehicle. The first class, Remaining Inside **(A $\rightarrow$ A)**, corresponds to static presence or localized movement within the vehicle interior. The second class, Remaining at Stop **(B $\rightarrow$ B)**, represents static presence or localized movement within the bus stop area. The third class, Alighting **(A $\rightarrow$ B)**, captures the transition from inside the vehicle to the bus stop through the front door. Finally, the fourth class, Boarding **(B $\rightarrow$ A)**, describes the transition from the bus stop into the vehicle through the front door.

To evaluate system robustness under varying interference conditions, data collection was performed under two distinct scenarios. The first scenario, Isolated Collection, involved trials conducted with a single active device at a time, minimizing channel contention and external interference. Two of the four devices were used in this scenario, with 20 repetitions per device for each movement class, providing clean baseline RSSI signatures. The second scenario, Noisy or Simultaneous Collection, involved trials conducted with all four devices operating simultaneously, performing two paired different movement classes in parallel. This setup introduces signal interference and collisions, approximating realistic passenger density conditions in public transport environments.

D. Data Preprocessing and Structuring

The raw data captured by the access point consisted of nested `Python` dictionaries, where each one-second capture contained network metadata for all connected devices. For each 10-second trial, ten such snapshots were recorded.

The preprocessing pipeline transformed this raw data into a machine-learning-ready dataset through a sequence of four steps. First, Temporal Aggregation was performed, where for each device and trial, the 10 sequential RSSI measurements were aggregated into a single feature vector $\mathbf{R} = [r_1, r_2, \ldots, r_{10}]$, allowing classifiers to exploit temporal trends, slopes, and variance rather than relying solely on instantaneous signal strength. Second, Device Isolation was applied, where data corresponding to each device was isolated using its MAC address, enabling independent trajectory reconstruction even during simultaneous collection scenarios. Third, Labeling was performed by assigning each RSSI sequence the corresponding movement class (**AA**, **BB**, **AB**, or **BA**) along with a boolean noise label indicating whether the trial was collected in isolation or under simultaneous device activity. Fourth, Feature Filtering was applied, where non-essential attributes, such as transmitted/received byte counts and connection duration, were discarded to reduce dimensionality and prevent overfitting.

E. *Final Dataset Structure*

The resulting dataset was exported in a CSV format, where each row represents a complete 10-second trajectory for a single device. Each instance includes the device MAC, 10 RSSI features, the movement class label, and the noise indicator.

This structure enables the analysis of both absolute signal strength and its temporal evolution, which is critical for distinguishing between static presence and transitional passenger movements, forming a solid foundation for supervised learning experiments.

IV. Exploratory Data Analysis

Prior to classifier training, an exploratory data analysis was conducted to assess the discriminative potential of temporal RSSI signatures and to characterize signal behavior across different passenger movement classes.

A. *Dataset Composition*

The final dataset is approximately balanced, comprising around 340 samples per movement class. For each class, 40 samples were collected under isolated conditions, while the remaining samples were obtained during simultaneous device activity, introducing controlled signal interference. This balance ensures that the exploratory analysis and the results reflect both ideal and realistic operating conditions.

B. *Temporal RSSI Characteristics*

The temporal evolution of RSSI values constitutes the primary discriminative feature between movement classes. Fig. 2 illustrates the mean RSSI trajectory over the 10-second observation window for each class.

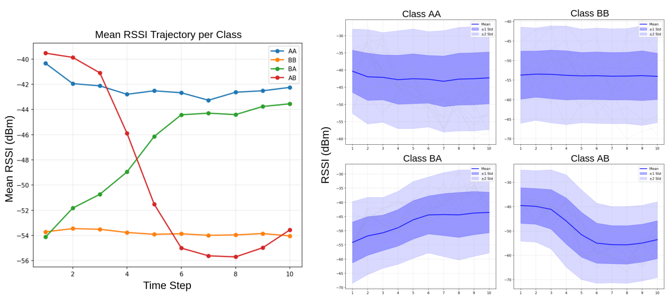

Figure 2. Temporal evolution of RSSI values over the 10-second observation window for each movement class.

Static states (**AA and BB**) exhibit relatively stable signal levels over time, albeit with distinct average magnitudes due to their spatial separation from the access point. In contrast, transitional movements display clear monotonic trends. The boarding class (**B** $\rightarrow$ **A**) shows a consistent increase in RSSI as the devices move toward the access point, while the alighting class (**A** $\rightarrow$ **B**) presents a pronounced decrease as physical obstructions attenuate the signal.

These opposing temporal patterns provide strong intuition for leveraging RSSI sequences in movement classification.

C. *Feature Space Separability*

To further inspect the structure of the 10-dimensional RSSI feature vectors, a t-Distributed Stochastic Neighbor Embedding (t-SNE) projection was applied for visualization purposes. The resulting two-dimensional embedding, shown in Fig. 3, reveals the formation of four predominantly distinct clusters corresponding to the defined movement classes.

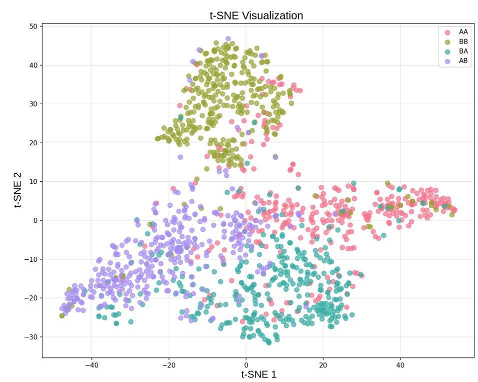

Figure 3. t-SNE projection for visualization of each class structure

While partial overlap is observed primarily associated with samples collected under noisy conditions the overall clustering suggests that temporal RSSI patterns retain sufficient class-dependent structure to support supervised learning approaches. This analysis is intended as a qualitative inspection of feature separability rather than a quantitative performance evaluation.

V. Results

VI. Discussion

VII. Contributions

VIII. Conclusions

A. *Future Work*

References

[1] Eurostat, "Urban and rural living in the eu," https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20230117-1, 2023, accessed: 2025-01-20.

[2] European Environment Agency, "Transport and environment report 2023," https://www.eea.europa.eu/publications/transport-and-environment-report-2023, 2023, accessed: 2025-01-20.

[3] L. Mikkelsen, R. Buchakchiev, T. Madsen, and H. P. Schwefel, "Public transport occupancy estimation using wlan probing," in *IEEE 8th International Workshop on Resilient Networks Design and Modeling (RNDM)*, 2016, pp. 302–308.

[4] B. Barabino, M. Di Francesco, and S. Mozzoni, “An offline framework for handling automatic passenger counting raw data,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 6, pp. 2443–2456, 2014.
[5] C. Pronello and X. R. Garzón Ruiz, “Evaluating the performance of video-based automated passenger counting systems in real-world conditions: A comparative study,” *Sensors*, vol. 23, no. 18, p. 7719, 2023.
[6] S. Cerqueira, E. Arsenio, and R. Henriques, “Inference of dynamic origin-destination matrices with trip and transfer status from individual smart card data,” *European Transport Research Review*, vol. 14, p. 49, 2022.
[7] A. Radovan, L. Mršić, G. Đambić, and B. Mihaljević, “A review of passenger counting in public transport concepts with solution proposal based on image processing and machine learning,” *Eng*, vol. 5, no. 4, pp. 3441–3467, 2024.
[8] M. Nitti, F. Pinna, L. Pintor, V. Pilloni, and B. Barabino, “iabacus: A wi-fi-based automatic bus passenger counting system,” *Energies*, vol. 13, no. 6, p. 1446, 2020.
[9] A. Simoni, M. Mohori, and A. Hrovat, “Non-intrusive privacy-preserving approach for presence monitoring based on wifi probe requests,” *Sensors*, vol. 23, no. 5, p. 2588, 2023.
[10] J. Guo, W. Zhuang, Y. Mao, and I. Ho, “Rssi-assisted csi-based passenger counting with multiple wi-fi receivers,” in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2025.
[11] C. Agualimpia-Arriaga, S. Govindasamy, B. Soni, C. Páez-Rueda, and A. Fajardo, “Rssi-based indoor localization using machine learning for wireless sensor networks: A recent review,” in *IEEE ANDESCON 2024*, 2024.
[12] C. Wiboonsiriruk, E. Phaisangittisagul, C. Srisurangkul, and I. Kumazawa, “Efficient passenger counting in public transport based on machine learning,” in *IEEE TENCON 2023*, 2023.
[13] T. A. Myrvoll, J. Håkegård, T. Matsui, and F. Septier, “Counting public transport passenger using wifi signatures of mobile devices,” in *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2017.
[14] L. Fabre, C. Bayart, Y. Kone, O. Manout, and P. Bonnel, “A machine learning approach to estimate public transport ridership using wi-fi data,” *IEEE Transactions on Intelligent Transportation Systems*, 2025.
[15] Y. Wang, H. Cheng, and M. Q.-H. Meng, “A learning-based sequence-to-sequence wifi fingerprinting framework for accurate pedestrian indoor localization using unconstrained rssi,” *IEEE Internet of Things Journal*, 2025.
[16] V. Servizi, D. R. Persson, F. Pereira, H. Villadsen, P. Bækgaard, I. Peled, and O. A. Nielsen, “"is not the truth the truth?": Analyzing the impact of user validations for bus in/out detection in smartphone-based surveys,” *IEEE Transactions on Intelligent Transportation Systems*, 2023.