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 remains a pressing concern, as over 75% of EU citizens reside in cities and transportation accounts for roughly 24% of greenhouse gas emissions [1], [2]. Managing public transport effectively hinges on accurate passenger flow data, yet traditional counting approaches—manual surveys, ticketing systems, infrared sensors, and Automatic Passenger Counting (APC) devices—come with drawbacks such as high costs, incomplete coverage, and privacy concerns [3], [4]. Moreover, conventional ticketing often fails to capture complete passenger journeys when exit validation is absent [5].

The widespread adoption of Wi-Fi-enabled devices opens up new avenues for non-intrusive passenger sensing. Such devices generate Received Signal Strength Indicator (RSSI) signatures that allow movement pattern recognition while preserving anonymity [6], [7]. Although Channel State Information (CSI) approaches can achieve high accuracy, they demand specialized hardware and intensive computation [8]. RSSI-based methods, by contrast, offer a practical alternative using standard networking equipment [9].

This paper puts forward a new approach to passenger boarding and alighting classification using temporal RSSI sequences. By tracking signal evolution over ten-second observation windows, our methodology distinguishes four movement patterns: remaining inside the vehicle, remaining at the stop, boarding, and alighting—all without requiring precise localization.

The principal contributions of this work are: (1) an RSSI-based movement classification framework that exploits temporal signal evolution; (2) an experimental dataset comprising approximately 1,360 labelled samples across four classes; (3) a thorough evaluation of 38 machine learning classifiers; (4) feature importance analysis; and (5) a privacy-preserving sensing approach that operates without device identification.

The remainder of this paper reviews related work (section II), describes the experimental setup (section III), presents exploratory data analysis (section IV), presents results (section V) and discussion (section VI), and concludes with future directions (section VII).

II. Related Work

A. Automatic Passenger Counting Systems

Conventional APC systems rely on infrared sensors, pressure mats, or video-based detection [3]. Pronello and Garzón Ruiz [4] observed that claimed 98% accuracy frequently drops to 53–74% under real-world conditions. Deep learning approaches can reach up to 94% accuracy [10], though vision-based systems remain hampered by occlusion, lighting variability, and privacy concerns.

B. Wi-Fi-Based Passenger Sensing

The ubiquity of Wi-Fi-enabled devices has spurred interest in wireless signal analysis for mobility monitoring. Myrvoll et al. [11] pioneered probe request analysis for passenger counting, while Nitti et al. [6] reported 100% detection in static settings and 94% in dynamic ones. CSI-based systems yield richer information, with Guo et al. [8] reaching over 94% accuracy, though they require specialized hardware. RSSI remains practical with off-the-shelf equipment. Fabre et al. [12] found Light Gradient Boosting effective for Wi-Fi-based ridership estimation, and Simončič et al. [7] attained over 96% accuracy despite Media Access Control (MAC) randomization.

C. RSSI Fingerprinting and Movement Classification

RSSI fingerprinting is well-established for indoor localization [9]. Wang et al. [13] proposed treating RSSI as temporal sequences, aligning with our methodology. Servizi et al. [14] addressed boarding detection using Bluetooth sensing, while Cerqueira et al. [5] demonstrated the importance of complete journey patterns for Origin-Destination (OD) matrix inference.

D. Research Gap

Despite progress in passenger counting and wireless sensing, machine learning classification of movement patterns from RSSI time series has received comparatively little attention. Existing approaches tend to focus on aggregate counting rather than fine-grained movement classification. Our work addresses this gap by casting passenger movement detection as a supervised classification problem using temporal RSSI evolution, thereby enabling real-time trajectory classification with a single access point at the vehicle door and keeping infrastructure requirements to a minimum.

III. Experimental Setup

This section outlines the physical data collection environment and the machine learning experimental framework devised for passenger movement classification.

A. Physical Data Collection Setup

Controlled experiments were carried out in an indoor environment that emulates public transport interactions, allowing reproducible data collection under both isolated and noisy conditions.

1) Environmental Configuration: The environment was divided into two zones (Figure 1). Zone A (Vehicle Interior) consisted of a closed room simulating a bus interior, with a Wi-Fi access point positioned adjacent to the doorway. Zone B (Bus Stop) was the corridor outside, representing the boarding area. The wall and door between zones introduce signal attenuation, generating distinctive RSSI patterns during transitions.

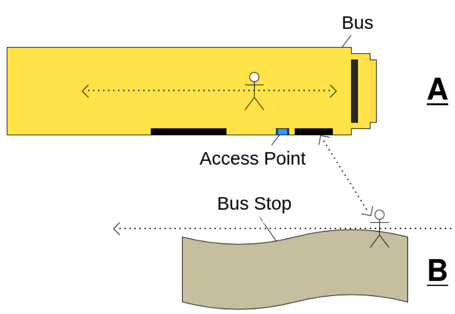

Figure 1. Experimental environment simulating a public transport scenario.

2) Data Acquisition: Data was collected using a Python script interfacing with the Acess Point via Ethernet. Each trial comprised a **10-second window** with **10 RSSI samples at 1 Hz**. Devices maintained periodic low-overhead traffic (ICMP) to ensure consistent RSSI reporting. Four mobile devices from three manufacturers were used to introduce hardware variability.

Four movement classes were defined: **A→A** (remaining inside), **B→B** (remaining at stop), **A→B** (alighting), and **B→A** (boarding). Collection occurred under two scenarios: *Isolated* (single device, 20 repetitions per class per device) and *Noisy* (four devices simultaneously performing paired movements).

3) Preprocessing and Dataset Structure: Raw data underwent the following preprocessing steps: (1) *Temporal Aggregation*, where 10 RSSI measurements per trial were aggregated into a feature vector $\mathbf{R} = [r_1, \ldots, r_{10}]$; (2) *Device Isolation* using MAC addresses; (3) *Labeling* with movement class and noise indicator; and (4) *Feature Filtering* to retain only RSSI values.

Prior to model training, *Feature Scaling* was applied using standardization (z-score normalization), which transforms each feature to zero mean and unit variance. This step is essential for classifiers sensitive to feature magnitudes, such as Support Vector Machines (SVMs), Gaussian Process (GP) classifiers, Multi-Layer Perceptrons (MLPs), and logistic regression, as these distance-based and gradient-based methods may otherwise suffer from degraded convergence speed and classification performance when operating on raw RSSI value ranges.

The resulting dataset contains 1,356 samples, each representing a 10-second trajectory with a corresponding movement class label, enabling analysis of both absolute signal strength and temporal evolution.

4) Experimental Scenarios: To comprehensively evaluate the impact of data collection conditions on classification performance, three distinct experimental scenarios were defined:

Combined Dataset: The complete dataset containing all 1,356 samples from both isolated and noisy collection conditions. This scenario represents the most realistic deployment setting, where the classifier must generalize across varying environmental conditions.

Isolated-Only Dataset: A subset containing exclusively samples collected under isolated conditions (single device, $n = 160$). This scenario provides an upper-bound estimate of classification performance under ideal conditions with minimal signal interference.

Noisy-Only Dataset: A subset containing exclusively samples collected with simultaneous multi-device activity ($n = 1,196$). This scenario evaluates classifier robustness under challenging conditions that closely approximate real-world public transport environments.

Comparing across these scenarios allows for a quantitative assessment of how environmental noise affects classification accuracy and sheds light on the operational boundaries of RSSI-based movement detection.

B. Machine Learning Experimental Framework

The following subsection details the classifier selection rationale, evaluation methodology, and performance metrics adopted in our experimental protocol.

1) Classifier Selection and Justification: A total of 38 classification algorithms spanning multiple paradigms were evaluated to ensure thorough benchmarking. The classifier families were chosen based on their established effectiveness in RSSI-based classification tasks [15], [16]:

Support Vector Machines (SVM): SVMs with Radial Basis Function (RBF) and linear kernels were included due to their demonstrated superiority in Wi-Fi fingerprinting tasks. Prior studies on indoor localization using RSSI have shown SVMs achieving accuracies exceeding 90% for location classification [9], [17]. The RBF kernel effectively captures non-linear relationships in signal strength patterns.

Ensemble Methods: Random Forest and Extra Trees were selected for their robustness to noise and ability to model complex decision boundaries without extensive hyperparameter tuning [15]. Gradient boosting variants (XGBoost, LightGBM, CatBoost) were included based on their state-of-the-art performance in tabular classification tasks, with CatBoost demonstrating particular effectiveness for categorical features [12].

Gaussian Process Classifier: GPs provide probabilistic predictions with uncertainty quantification, particularly valuable for RSSI data where signal variability is inherent. The RBF kernel enables automatic adaptation to the intrinsic dimensionality of temporal RSSI sequences.

Neural Networks: MLPs with varying architectures (small, medium, large) were evaluated to assess whether deeper representations improve classification over traditional methods for this feature space dimensionality.

Stacking and Voting Ensembles: Meta-learning approaches combining heterogeneous base learners were included to leverage complementary classifier strengths, a strategy shown to improve robustness in transportation sensing applications [10].

2) Data Partitioning Strategy: The dataset was partitioned using **stratified sampling** with an 80%/20% train-test split. Stratified sampling ensures that class distributions are preserved in both partitions, which is critical for multi-class classification problems where class imbalance could otherwise bias model evaluation. This approach maintains the original proportion of each movement class (AA, BB, AB, BA) in both training and testing sets.

3) Cross-Validation Protocol: Model training employed **5-fold stratified cross-validation**, a methodology widely recommended for robust classifier evaluation. Stratified K-fold cross-validation maintains class ratios across all folds, ensuring that minority classes receive adequate representation during training and validation. This technique reduces variance in performance estimates compared to simple hold-out validation.

To assess result stability, experiments were repeated with three random seeds (3, 5, and 42), and metrics were aggregated across runs. This multi-seed evaluation quantifies classifier sensitivity to random initialization and data shuffling.

4) Hyperparameter Optimization: To ensure fair comparison and optimal performance across classifier families, systematic hyperparameter optimization was conducted using Optuna [18], a state-of-the-art Bayesian optimization framework. Optuna employs the Tree-structured Parzen Estimator (TPE) algorithm to efficiently explore high-dimensional hyperparameter spaces, focusing search efforts on promising regions.

For each tunable classifier, extensive hyperparameter searches were performed with budgets ranging from 50 to over 1,300 trials depending on model complexity. The optimization objective was accuracy during 5-fold stratified cross-validation on the training set. This approach yields classifier configurations specifically adapted to the RSSI feature space characteristics.

The hyperparameter search spaces were defined based on recommended ranges from the literature and practical considerations. Ensemble methods (Random Forest, Extra Trees, Gradient Boosting variants) were tuned over tree depth, number of estimators, and regularization parameters. SVMs were optimized for the regularization parameter C and kernel coefficients. Neural network configurations explored layer architectures, learning rates, and regularization strengths. The complete hyperparameter search spaces and optimal configurations for each classifier are documented in the Appendix.

5) Performance Metrics: Four complementary metrics were employed to provide comprehensive performance characterization:

Accuracy: The proportion of correctly classified samples. While intuitive, accuracy can be misleading for imbalanced datasets.

Weighted F1-Score: The harmonic mean of precision and recall, weighted by class support. This metric balances false positives and false negatives while accounting for class distribution.

Balanced Accuracy: The arithmetic mean of per-class recall values, ensuring equal contribution from each class regardless of prevalence.

Matthews Correlation Coefficient (MCC): Selected as the primary evaluation criterion, as this metric offers a reliable and balanced evaluation of classification models, particularly in scenarios involving imbalanced datasets or when assessing performance across multiple classes [19], [20]. Matthews Correlation Coefficient (MCC) ranges from -1 to $+1$, producing high scores only when all confusion matrix categories achieve strong results.

IV. Exploratory Data Analysis

Ahead of classifier training, an exploratory data analysis was carried out to gauge the discriminative potential of temporal RSSI signatures and to characterize signal behaviour across different passenger movement classes.

A. Dataset Composition

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

B. Temporal RSSI Characteristics

The temporal evolution of RSSI values serves as the primary discriminative feature between movement classes. Figure 2 depicts the mean RSSI trajectory over the 10-second observation window for each class.

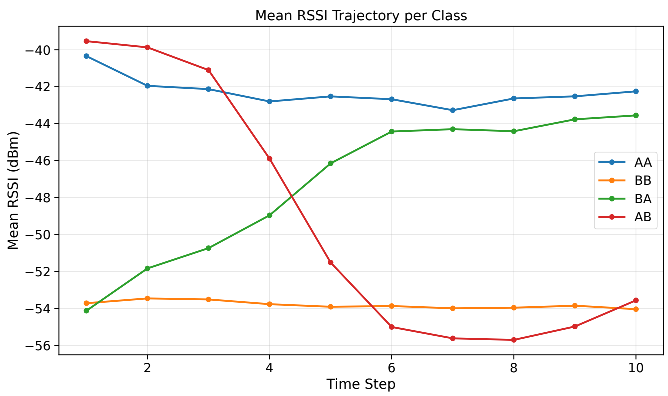

Figure 2. Temporal evolution of RSSI values over the 10-second observation window for each movement class.

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 contrasting temporal patterns offer a compelling rationale for leveraging RSSI sequences in movement classification.

V. Results

This section presents the experimental results from training and evaluating an extensive set of machine learning classifiers on the RSSI-based passenger movement dataset across three experimental scenarios: combined, isolated-only, and noisy-only conditions. The evaluation covers 38 classification algorithms assessed across multiple random seeds to ensure statistical robustness.

A. Comparative Analysis Across Experimental Scenarios

Table I presents the performance comparison of top classifiers across the three experimental scenarios, revealing the substantial impact of data collection conditions on classification performance.

The isolated-only scenario yielded the highest classification performance, with K-Nearest Neighbors (KNN) (k=5) reaching an MCC of 0.907—a 20% improvement over the combined dataset. Simpler classifiers such as KNN showed the most pronounced performance gains in the isolated scenario.

The noisy-only scenario exhibited performance levels comparable to the combined dataset, with CatBoost attaining the highest MCC of 0.770. The GP classifier maintained robust performance (MCC: 0.755) across both combined and noisy-only conditions, yet showed notable degradation in the isolated scenario (MCC: 0.414).

Table I
Performance Comparison Across Experimental Scenarios (MCC)

Classifier	Combined	Isolated-Only	Noisy-Only
KNN (k=5)	0.692 ± 0.029	**0.907 ± 0.061**	0.704 ± 0.025
KNN (k=3)	0.690 ± 0.046	0.882 ± 0.068	0.702 ± 0.043
LinearSVC	0.726 ± 0.039	0.867 ± 0.060	0.716 ± 0.031
LogisticRegression (L2)	0.743 ± 0.044	0.866 ± 0.028	0.731 ± 0.009
SVC (Linear)	0.750 ± 0.023	0.850 ± 0.088	0.731 ± 0.037
StackingEnsemble	0.749 ± 0.020	0.851 ± 0.122	0.768 ± 0.023
ExtraTrees	0.737 ± 0.013	0.836 ± 0.041	0.755 ± 0.021
GaussianProcess	**0.756 ± 0.033**	0.414 ± 0.052	0.755 ± 0.028
SVC (RBF)	0.755 ± 0.021	0.825 ± 0.101	0.754 ± 0.016
CatBoost	0.746 ± 0.017	0.782 ± 0.063	**0.770 ± 0.013**

B. Per-Class Analysis

Table II presents the per-class accuracy, recall, F1-score, and MCC for the best classifier (GP).

Table II
Per-Class Performance Metrics (GP)

Class	Accuracy	Recall	F1-Score	MCC
AA (Inside)	0.838	0.750	0.791	0.785
BB (Stop)	0.838	0.897	0.878	0.785
BA (Boarding)	0.838	0.824	0.855	0.785
AB (Alighting)	0.838	0.882	0.828	0.785
Weighted Avg	**0.838**	**0.838**	**0.838**	**0.785**

The static state at the bus stop (BB) exhibited the highest recall (89.7%) and F1-score (0.878). The boarding movement (BA) achieved recall of 82.4%, F1-score of 0.855, and MCC of 0.785. The alighting class (AB) demonstrated high recall (88.2%) but lower F1-score (0.828). The static state inside the vehicle (AA) presented the lowest recall (75.0%) and F1-score (0.791). The overall accuracy reached 83.8% with an MCC of 0.785.

C. Hyperparameter Configuration

While extensive hyperparameter optimization was conducted for nine classifier families using Optuna (detailed in section A), the GP classifier employed default configurations with a RBF kernel, which proved highly effective without explicit tuning. The GP configuration is presented in Table III.

Table III
GP Hyperparameters

Parameter	Value
Kernel	$1.0 \times \text{RBF}(1.0)$
Kernel Length Scale	Optimized during fitting
Optimizer	L-BFGS-B
Max Iterations	100
Multi-class Strategy	One-vs-Rest

The RBF kernel automatically learns the optimal length scale parameter during training, adapting to the intrinsic dimensionality of the RSSI temporal sequences. Notably, the GP classifier achieved top performance without requiring the extensive hyperparameter search applied to other classifiers, suggesting that its probabilistic formulation is particularly well-suited to the RSSI feature space.

D. Confusion Matrix Analysis

Figure 3 presents the normalized confusion matrix for the GP classifier, which achieved the best average performance across experimental seeds.

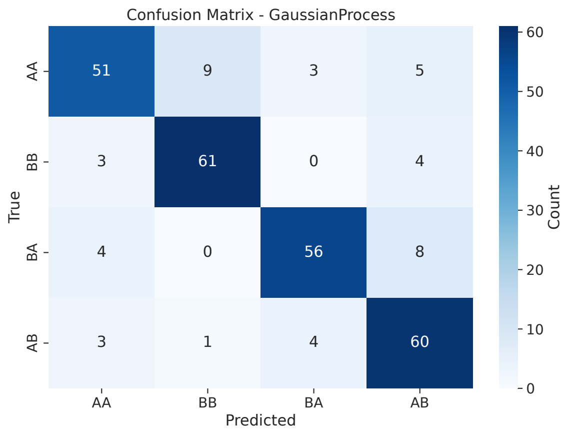

Figure 3. Confusion matrix for the GP classifier, demonstrating strong diagonal dominance with minimal inter-class confusion.

The confusion matrix reveals that the primary source of classification errors occurs between spatially adjacent classes. The AA class (remaining inside) is occasionally misclassified as BA (boarding), and minor confusion exists between AB (alighting) and BB (remaining at stop).

E. Model Stability Analysis

To evaluate the robustness of classifier rankings across different experimental conditions, Figure 4 illustrates the accuracy variability for the top classifiers across the three random seeds.

The analysis confirms that top-performing classifiers maintain consistent relative rankings across seeds, with kernel-based methods (GP, SVC) and ensemble approaches (Stacking, CatBoost) displaying the lowest variability.

F. Feature Importance

Analysis of feature importance across interpretable classifiers revealed that the initial RSSI measurements (features 1–3) contribute most significantly to classification decisions, as illustrated in Figure 5.

The first RSSI sample exhibits the highest importance (normalized score: 2.29), followed by samples at positions 6 and 2.

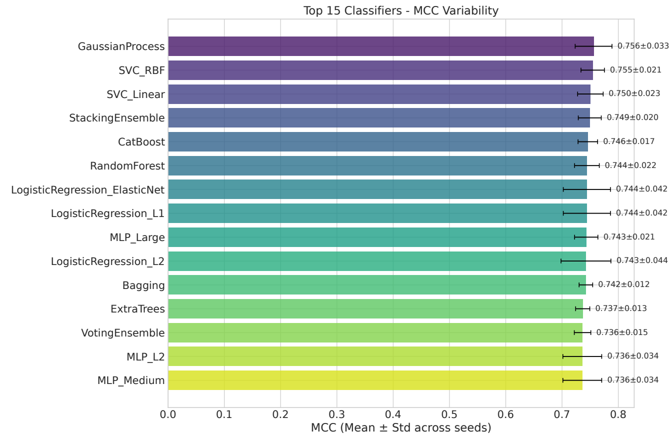

Figure 4. MCC variability for top classifiers, in combined dataset, demonstrating consistent ranking stability.

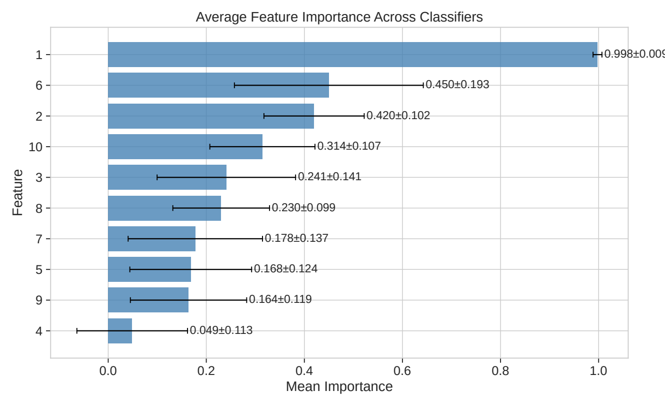

Figure 5. Mean feature importance across classifiers, with standard deviation, to the $[0, 1]$ range using min-max normalization

VI. Discussion

The aforementioned experimental results confirm the feasibility of using temporal RSSI sequences for non-intrusive passenger movement classification in public transport scenarios. This section analyses the key findings across the three experimental scenarios, examines how data collection conditions influence performance, and discusses the practical implications for real-world deployment.

A. Classifier Performance Analysis

The comparative evaluation across experimental scenarios paints a nuanced picture of classifier suitability for RSSI-based movement classification. On the combined dataset, the GP classifier attained the highest MCC (0.756), exhibiting strong discriminative capability for the temporal RSSI patterns. This performance stems from its probabilistic framework and the flexibility of the RBF kernel in modelling non-linear decision boundaries. Unlike parametric models that assume specific functional forms, GPs adapt their complexity to the underlying data distribution, an advantage for RSSI patterns that exhibit complex spatial dependencies.

The flexibility of the RBF kernel enables the GP to model complex decision boundaries without requiring extensive manual hyperparameter tuning, making it particularly suitable for RSSI-based classification where signal patterns exhibit non-linear spatial dependencies. SVMs with RBF and linear kernels achieved comparable performance on the combined dataset, confirming that kernel-based methods are well-suited for this classification task. The margin-maximization principle of SVMs provides robust generalization, particularly relevant given the overlap between movement classes observed in the t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization.

B. Impact of Data Collection Conditions

Perhaps the most striking finding emerges from the comparison between experimental scenarios. The isolated-only dataset yielded markedly superior classification performance, with KNN (k=5) reaching an MCC of 0.907—a 31% relative improvement over its performance on the combined dataset (MCC: 0.692). This pronounced increase can be traced to the absence of inter-device signal interference during data collection, resulting in cleaner RSSI patterns with more distinct class separations. Two primary factors account for this enhancement:

Signal Clarity: In isolated conditions, RSSI measurements are unaffected by inter-device interference, co-channel contention, or access point load variations. The resulting signal patterns exhibit clearer temporal trajectories with reduced variance within each movement class.

Class Separability: The absence of noise enables more distinct decision boundaries between movement classes. Simpler classifiers such as KNN, which rely on local neighbourhood structure, benefit disproportionately from this increased separability, suggesting that the underlying class boundaries are well-defined when noise is absent. This explains the pronounced performance gains observed for KNN in the isolated scenario.

Conversely, complex models such as the GP suffered marked performance degradation in the isolated scenario (MCC: 0.414), despite attaining the best results on the combined dataset. This apparent paradox is explained by the limited sample size ($n = 160$) of the isolated dataset, which proves insufficient for the GP to reliably estimate its kernel hyperparameters without overfitting.

The noisy-only scenario, representing the most realistic operational conditions, showed performance levels on par with the combined dataset. CatBoost attained the highest MCC (0.770) in this scenario, suggesting that gradient boosting methods are particularly robust to signal interference. The consistency between noisy-only and combined results indicates that the combined dataset's performance is largely driven by the noisy samples, which make up 88% of the total data.

C. Model Stability Considerations

The observed stability of classifier rankings across experimental seeds has notable implications for deployment. Kernel-based methods (GP, SVC) and ensemble approaches (Stacking, CatBoost) showed the lowest variability, a desirable property for deployment scenarios where model retraining may occur with different data partitions. This consistency lends confidence that selected classifiers will maintain their relative performance across varying operational conditions.

D. Classifier Selection Guidelines

The experimental results inform practical recommendations for classifier selection based on deployment context:

High-interference environments: For scenarios with multiple simultaneous devices, gradient boosting methods (CatBoost, XGBoost) and ensemble approaches (StackingEnsemble) offer the best balance of accuracy and robustness.

Controlled environments: In settings with minimal device density, simpler classifiers such as KNN or logistic regression can achieve superior performance while offering reduced computational overhead and improved interpretability.

General deployment: For systems that must operate across varying conditions, SVC with RBF kernel provides consistent performance with acceptable variance, making it a reliable default choice.

E. Per-Class Error Analysis

The confusion patterns reveal insights into the physical characteristics of each movement class. The static state inside the vehicle (AA) exhibited the lowest recall (75.0%), primarily due to misclassification as boarding (BA). This confusion is attributable to the spatial proximity of both classes to the access point, resulting in similar high-RSSI signatures that reflect the similarity in RSSI magnitude when devices are positioned near the access point. Although the temporal dynamics differ (AA maintains relatively stable values while BA shows an increasing trend), this distinction may be subtle in short observation windows.

Conversely, the bus stop class (BB) achieved the highest recall (89.7%), attributable to the consistent low RSSI values observed when devices remain outside the vehicle. The physical barrier separating Zone B from the access point provides natural signal attenuation that enables clear class discrimination. The transitional classes (AB and BA) benefited from their characteristic monotonic RSSI trends, with alighting (AB) achieving 88.2% recall. However, the lower F1-score (0.828) for the alighting class indicates some false positives from the AA class. The boarding movement (BA) achieved strong results (F1-score: 0.855), benefiting from the distinctive increasing RSSI pattern as devices approach the access point.

The overall accuracy of 83.8% and MCC of 0.785 confirm robust multi-class discrimination across all movement categories, demonstrating that RSSI-based temporal patterns provide sufficient information for reliable passenger movement classification.

F. Confusion Matrix Interpretation

The observed confusion between spatially adjacent classes provides important insights for system design. The AA–BA confusion reflects the similarity in RSSI magnitude when devices are positioned near the access point, suggesting that additional features or longer observation windows might improve discrimination between static and transitional states. Similarly, the AB–BB confusion arises because both classes share lower RSSI values characteristic of the exterior zone, though the transitional nature of AB provides some discriminative information through temporal patterns.

G. Feature Importance Insights

The feature importance analysis revealed that initial RSSI measurements (features 1–3) contribute most significantly to classification decisions, capturing the starting position and providing immediate context for distinguishing static states from transitional movements. This finding aligns with the temporal dynamics of passenger movements, where early signal readings capture the initial position before any state transition occurs.

The elevated importance of feature 6 (mid-trajectory) indicates that classifiers also rely on signal evolution to confirm movement direction, validating the choice of sequential RSSI measurements over aggregate statistics. The observed pattern suggests that classifiers leverage both the starting signal strength and mid-trajectory measurements to infer movement direction, while later features provide confirmatory information about the final position. This temporal dependency structure supports the use of sequence-based classification approaches for RSSI-based passenger detection.

The cross-classifier variability in feature importance, quantified by the standard deviation across normalized importance values, provides additional insight into classifier behaviour. Feature 1 exhibits the highest mean importance (0.998) with the lowest relative variability (standard deviation of 0.009), indicating universal agreement among classifiers regarding the discriminative value of the initial RSSI measurement. Conversely, mid-trajectory features (4, 5, 7, and 9) exhibit higher relative variability, with coefficient of variation values exceeding 0.7. This disparity reflects fundamental differences in how classifiers exploit temporal information: tree-based ensemble methods tend to favour early features that establish initial position, whereas kernel-based methods such as SVMs and GPs distribute importance more evenly across the temporal sequence to capture signal dynamics. Feature 6 occupies an intermediate position, with moderate importance (0.450) but elevated variability (standard deviation of 0.193), suggesting that its utility for classification is model-dependent and potentially influenced by the specific decision boundaries each classifier learns.

H. Limitations and Considerations

Several limitations should be acknowledged. First, the controlled experimental environment, while designed to simulate public transport conditions, may not capture all sources of variability present in operational settings, such as passenger density fluctuations, vehicle movement, and diverse access point placements.

Second, the 10-second observation window, while suitable for capturing typical boarding and alighting actions, may be insufficient for detecting slower movements or hesitant passengers. Adaptive window lengths could potentially improve classification accuracy in such cases.

Third, the current approach assumes consistent device behaviour; however, variations in device hardware, operating system power management, and user-initiated WiFi state changes may affect RSSI reporting in practice.

Fourth, the limited sample size of the isolated dataset ($n = 160$) constrains the reliability of performance estimates for complex classifiers in that scenario. Future work should expand isolated data collection to enable more robust comparisons.

I. Practical Implications

The experimental findings hold meaningful implications for real-world deployment. The notable performance improvement observed in isolated conditions (MCC up to 0.907) suggests that signal interference is the chief limiting factor for classification accuracy. Deployment strategies that mitigate interference—such as dedicated frequency channels, directional antennas, or temporal multiplexing—could markedly enhance system performance.

Nevertheless, the classification performance achieved under noisy conditions (MCC $>$ 0.77 with CatBoost) shows that RSSI-based movement classification remains viable as a complementary technology for passenger counting systems, even in challenging environments. The consistency of results across the combined and noisy-only scenarios lends confidence that models trained on realistic data will generalize well to operational settings.

VII. Conclusions

This paper has presented a new approach for classifying passenger movements in public transport using temporal sequences of Wi-Fi RSSI measurements. The proposed methodology enables non-intrusive distinction between four fundamental movement classes using only standard Wi-Fi access point infrastructure.

The GP classifier attained the highest performance, with an accuracy of 81.6%, recall of 81.6%, F1-score of 81.5%, and MCC of 0.756, validated across multiple random seeds. SVMs and regularized logistic regression variants yielded comparable results, confirming that both kernel-based and linear methods can effectively exploit the temporal structure of RSSI sequences.

Per-class analysis revealed that static states at the bus stop (BB) and transitional movements (AB, BA) are more readily distinguished owing to their characteristic signal patterns, whereas the static state inside the vehicle (AA) posed the greatest classification challenge because of its proximity-based similarity with the boarding class.

The results demonstrate that RSSI-based passenger movement classification offers a viable, cost-effective, and privacy-preserving complement to existing APC technologies, requiring no specialized hardware beyond standard Wi-Fi infrastructure.

A. Future Work

Several directions for future research emerge from this work. First, validation in operational public transport environments is essential to gauge the impact of real-world factors such as vehicle movement, passenger density variations, and diverse access point configurations. Second, the integration of complementary sensor modalities such as accelerometer data or Bluetooth Low Energy beacons could enhance classification accuracy and robustness. Third, the development of adaptive observation windows that adjust to movement speed could improve detection of hesitant or slower passengers. Fourth, investigation of federated learning approaches would enable model improvement across multiple vehicles while preserving data privacy. Finally, the extension of the methodology to estimate complete OD matrices through temporal aggregation of boarding and alighting events represents a natural progression toward comprehensive passenger flow analytics.

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] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] C. Wiboonsiriruk, E. Phaisangittisagul, C. Srisurangkul, and I. Kumazawa, "Efficient passenger counting in public transport based on machine learning," in *IEEE TENCON 2023*, 2023.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] N. Singh, S. Choe, and R. Punmiya, "Machine learning based indoor localization using wi-fi rssi fingerprints: An overview," *IEEE Access*, vol. 9, pp. 127 150–127 174, 2021.

[16] B. Zholamanov, A. Saymbetov, M. Nurgaliyev, A. Bolatbek, G. Dosymbetova, N. Kuttybay, S. Orynbassar, A. Kapparova, N. Koshkarbay, and Ö. F. Beyca, "Rssi fingerprint-based indoor localization solutions using machine learning algorithms: A comprehensive review," *Smart Cities*, vol. 8, no. 5, p. 153, 2025.

[17] C. Jain, G. V. S. Sashank, V. N, and S. Markkandan, "Low-cost ble based indoor localization using rssi fingerprinting and machine learning," in *IEEE WiSPNET 2021*, 2021.

[18] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2019, pp. 2623–2631.

[19] D. Chicco and G. Jurman, "The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, 2020.

[20] ——, "The matthews correlation coefficient (mcc) should replace the roc auc as the standard metric for assessing binary classification," *BioData Mining*, vol. 16, no. 1, p. 4, 2023.

Appendix

This appendix documents the hyperparameter optimization process conducted using Optuna [18], including the search spaces explored and the optimal configurations identified for each classifier family.

A. Optimization Methodology

The Tree-structured Parzen Estimator (TPE) sampler was employed with consistent random seeds to ensure reproducibility. Each classifier underwent extensive optimization with the number of trials shown in Table IV. The optimization objective was 5-fold stratified cross-validation accuracy on the training set.

Table IV
Optuna Hyperparameter Optimization Summary

Classifier	Trials	Best CV Accuracy
Random Forest	1,303	89.08%
Extra Trees	1,250	91.45%
Gradient Boosting	1,170	87.51%
XGBoost	1,150	87.51%
LightGBM	51	79.89%
SVC (RBF)	950	92.22%
KNN	950	85.88%
MLP	950	90.68%
Logistic Regression	900	85.97%

B. Optimal Hyperparameter Configurations

The following tables present the best hyperparameter configurations identified through the optimization process.

1) **Random Forest** (Table V)

Table V
Random Forest Best Hyperparameters

Parameter	Value
n_estimators	97
max_depth	13
min_samples_split	2
min_samples_leaf	1
max_features	sqrt
bootstrap	False

2) **Extra Trees** (Table VI)

Table VI
Extra Trees Best Hyperparameters

Parameter	Value
n_estimators	182
max_depth	9
min_samples_split	9
min_samples_leaf	1
max_features	sqrt

3) **Gradient Boosting** (Table VII)

Table VII
Gradient Boosting Best Hyperparameters

Parameter	Value
n_estimators	156
learning_rate	0.0616
max_depth	12
min_samples_split	4
min_samples_leaf	2
subsample	0.950

4) **XGBoost** (Table VIII)

Table VIII
XGBoost Best Hyperparameters

Parameter	Value
n_estimators	104
learning_rate	0.253
max_depth	13
min_child_weight	2
subsample	0.800
colsample_bytree	0.714
reg_alpha (L1)	2.66×10^{-6}
reg_lambda (L2)	1.368

5) **LightGBM** (Table IX)

Table IX
LightGBM Best Hyperparameters

Parameter	Value
n_estimators	83
learning_rate	0.0333
max_depth	8
num_leaves	25
min_child_samples	45
subsample	0.970
colsample_bytree	0.923
reg_alpha (L1)	4.15×10^{-5}
reg_lambda (L2)	1.02×10^{-7}

6) **SVC (RBF Kernel)** (Table X)

Table X
SVC (RBF Kernel) Best Hyperparameters

Parameter	Value
C (regularization)	0.634
gamma	scale
kernel	RBF

7) **K-Nearest Neighbors** (Table XI)

Table XI
KNN Best Hyperparameters

Parameter	Value
n_neighbors	4
weights	distance
metric	minkowski
p	5

8) **MLP Neural Network** (Table XII)

Table XII
MLP BEST HYPERPARAMETERS

Parameter	**Value**
hidden_layer_sizes	(249, 140, 95)
activation	tanh
alpha (L2 penalty)	0.000991
learning_rate	constant
learning_rate_init	0.00696
max_iter	500
early_stopping	True

9) **Logistic Regression** (Table XIII)

Table XIII
LOGISTIC REGRESSION BEST HYPERPARAMETERS

Parameter	**Value**
C (regularization)	36.81
l1_ratio	0.214
solver	saga
max_iter	1000

C. Search Space Definitions

Table XIV summarizes the hyperparameter search ranges explored during optimization.

Table XIV
HYPERPARAMETER SEARCH SPACE RANGES

Classifier	**Parameter**	**Range**	**Scale**
Tree Ensembles	n_estimators	[50, 300]	Linear
	max_depth	[3, 30]	Linear
	min_samples_split	[2, 20]	Linear
	min_samples_leaf	[1, 10]	Linear
	max_features	{sqrt, log2, None}	Categorical
Boosting Methods	learning_rate	[0.01, 0.3]	Log
	subsample	[0.6, 1.0]	Linear
	reg_alpha	[10^{-8}, 10]	Log
	reg_lambda	[10^{-8}, 10]	Log
SVC (RBF)	C	[0.01, 100]	Log
	gamma	{scale, auto}	Categorical
KNN	n_neighbors	[1, 20]	Linear
	weights	{uniform, distance}	Categorical
	metric	{euclidean, manhattan, minkowski}	Categorical
	p	[1, 5]	Linear
MLP	n_layers	[1, 3]	Linear
	n_units_per_layer	[32, 256]	Linear
	alpha	[10^{-5}, 0.1]	Log
	learning_rate_init	[10^{-4}, 0.1]	Log
	activation	{relu, tanh}	Categorical
Logistic Regression	C	[0.001, 100]	Log
	l1_ratio	[0, 1]	Linear