Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 14, No. 3, September 2026, pp. 861~873
ISSN: 2089-3272, DOI: 10.52549/ijeei.v14i3.8256

Accuracy and Stability Trade-off in RSSI Window Aggregation for MLP-Based Wi-Fi Indoor Positioning

Wahyu Arthanugraha[1], Abdul Halim Ismail[2], Shahrul Nizam Yaakob[3]
[1]Informatics Engineering Study Program, Universitas Muhammadiyah Parepare, Indonesia
[2]Department of Mechatronic, Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis, Malaysia
[3]Department of Computer, Faculty of Intelligence Computing, Universiti Malaysia Perlis, Malaysia

Article Info

Article history:

Received Jun 13, 2026
Revised Sep 20, 2026
Accepted Oct 6, 2026

Keywords:

Wi-Fi indoor positioning
RSSI window aggregation
Multilayer Perceptron
Fingerprinting
Accuracy and stability trade-off

ABSTRACT

Wi-Fi RSSI fingerprinting is practical for indoor positioning, but short-term signal fluctuations can reduce classification consistency. This study evaluates how fixed RSSI window size affects the accuracy, stability, and sample availability of an MLP-based indoor positioning system. RSSI from three Access Points was collected at 45 ordered corridor positions. Non-overlapping windows with $n = 1,3,5,7,9,11,15,21$ and 29 were represented by the mean and standard deviation of three RSSI channels. The MLP was assessed using stratified 5-fold cross-validation with supplementary order-blocked validation. Relative to the $n = 1$ baseline Accuracy of 68.24%, $n = 7$ achieved 90.60 ± 0.71% Accuracy, 95.08 ± 1.31% Top-2 Accuracy, 0.9047 ± 0.0060 Macro-F1, and 0.4319 ± 0.1364 Ordinal Error. Larger windows achieved higher mean Accuracy but sharply reduced window-level samples and increased fold-to-fold variability. Order-blocked validation produced more conservative estimates and shifted the strongest performance region toward $n = 9$ to $n = 11$. These findings show that fixed-window selection should consider predictive performance, evaluation stability, sample availability, and validation design rather than maximum Accuracy alone.

Copyright © 2026 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Wahyu Arthanugraha,
Informatics Engineering Study Program,
Universitas Muhammadiyah Parepare,
Building F, 1st Floor, Parepare City, South Sulawesi, Indonesia 91131.
Email: wahyusaja212@gmail.com

1. INTRODUCTION

Indoor positioning is a key component of location-based services in environments such as campus buildings, hospitals, shopping malls, laboratories, logistics warehouses, and industrial facilities. Unlike outdoor positioning, where Global Navigation Satellite System (GNSS) is widely available, indoor environments attenuate, reflect, and obstruct satellite signals. RSSI-based Wi-Fi fingerprinting remains widely used because it can exploit existing network infrastructure, is relatively low-cost, and requires no specialized positioning hardware [1–7]. However, RSSI measurements remain sensitive to multipath propagation, non-line-of-sight conditions, environmental noise, interference, device orientation, and human activity [8–13].

As machine learning and deep learning have advanced, data-driven methods have increasingly been used to model nonlinear relationships between RSSI fingerprints and user locations. Recent studies have reported improved localization performance using neural networks, convolutional and recurrent architectures, autoencoders, semi-supervised learning, data augmentation, few-shot learning, federated learning, and explainable deep learning [14–23]. However, greater model complexity does not necessarily guarantee robustness when the RSSI representation remains sensitive to transient signal variations. Recent work also shows that preprocessing, standardization, feature construction, fingerprint correction, data augmentation, and

feature representation can strongly affect model stability, particularly when only a limited number of Access Points is available or measurements are collected under noisy real-world conditions [24–27].

Temporal preprocessing of noisy signals can involve both segmentation and filtering. Fixed-length signal segmentation may use overlapping or non-overlapping windows. Overlapping windows increase the number of available segments and temporal coverage, but they also introduce greater redundancy and dependence between adjacent segments [28]. Recent indoor localization studies have likewise applied several forms of RSSI smoothing. Exponential and simple moving averages, as well as moving-median filtering, have been used to stabilize Wi-Fi RSS measurements before localization [29]. Kalman and Gaussian-based approaches have also been investigated for constructing more stable RSSI representations, particularly because indoor RSSI distributions may not always follow a simple Gaussian pattern [30]. Other recent studies have combined Kalman, Gaussian, and mean filtering [31], while comparative evaluations have examined Kalman, weighted moving average, and Gaussian filters and shown that their behavior depends on both signal conditions and the amount of temporal smoothing applied [32]. More generally, statistical descriptors such as the mean, variance, and standard deviation remain established time-domain features for summarizing local signal behavior within a temporal segment [33]. These studies confirm that temporal filtering and statistical aggregation are established preprocessing strategies. The unresolved issue addressed here is therefore not whether RSSI should be smoothed, but how the fixed aggregation length affects predictive performance, evaluation stability, and sample availability when the classifier and evaluation pipeline are kept unchanged.

Previous work on Wi-Fi RSS fingerprint database construction has shown that fingerprint representation quality can influence positioning performance [34]. A small window may retain short-term signal variation, whereas a large window provides stronger aggregation at the cost of fewer window-level samples and longer observation intervals. Window size should therefore be evaluated as a trade-off among predictive performance, evaluation stability, Ordinal Error, and sample availability.

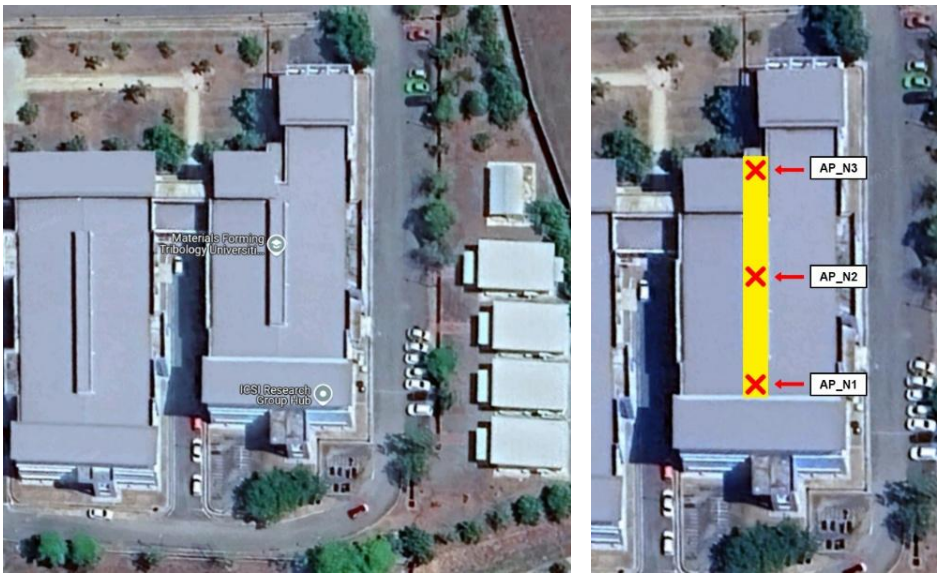
Based on these considerations, this study examines how fixed RSSI window aggregation affects the performance and stability of an MLP-based Wi-Fi indoor positioning system. Rather than introducing a new filtering algorithm or neural architecture, the experiments are designed to isolate the effect of window size by keeping the MLP architecture, feature construction procedure, preprocessing, and evaluation protocol unchanged across all tested configurations. The evaluated fixed window sizes were $n = 1,3,5,7,9,11,15,21$ and 29. Performance was assessed using Accuracy, Top-2 Accuracy, Macro-F1, Ordinal Error, fold-to-fold variability, and the number of available window-level samples. This controlled design allows the performance gain obtained from stronger RSSI aggregation to be examined together with its consequences for evaluation stability and sample availability. The main contribution is therefore a quantitative characterization of the accuracy and stability trade-off associated with fixed RSSI window aggregation and an empirical basis for selecting a practical window size under a limited-Access-Point indoor positioning setting.

2. RESEARCH METHOD

2.1. Research Design and Experimental Site

This study used a controlled experimental design to evaluate the effect of fixed RSSI window size on MLP-based indoor position classification. The model architecture, preprocessing procedure, and evaluation settings were kept unchanged across window sizes so that performance differences could be examined primarily in relation to RSSI aggregation.

The data were collected in a straight indoor corridor at the ICSI Research Hub, Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis. The testbed was approximately 45 m long and contained 45 ordered reference positions spaced at 1 m intervals. Three fixed Access Points were distributed longitudinally along the corridor to provide the RSSI channels AP_N1, AP_N2, and AP_N3. The Access Points were mounted at a height of 1.5 m, with AP_N2 located approximately midway between AP_N1 and AP_N3, giving an interval of about 22.5 m between consecutive Access Points. This arrangement forms a controlled one-dimensional positioning testbed while preserving naturally occurring RSSI variability in the indoor environment.



(a) (b)

Figure 1. Experimental site and Access Point placement in the indoor corridor testbed

Figure 1 illustrates the experimental environment and the physical deployment of the Access Points. Figure 1(a) shows the experimental site at the ICSI Research Hub, Universiti Malaysia Perlis, whereas Figure 1(b) shows the longitudinal placement of AP_N1, AP_N2, and AP_N3 along the corridor. The three-AP arrangement follows the same linear direction as the ordered reference positions, providing a deliberately simple and reproducible spatial configuration for evaluating RSSI window aggregation.

The evaluated testbed therefore represents a constrained one-dimensional indoor positioning scenario. It does not reproduce room-rich, corner-dominated, multi-floor, or predominantly non-line-of-sight layouts. Accordingly, the findings are interpreted only for the evaluated single-corridor environment with a fixed three-AP deployment and are not presented as evidence of generalization across different building geometries, Access Point layouts, or measurement sessions.

2.2. RSSI Data Acquisition and Dataset Preparation

RSSI measurements were recorded at 45 consecutive reference positions separated by 1 m and labeled from L1 to L45, resulting in a 45-class ordered position-classification problem. The three Access Points, AP_N1, AP_N2, and AP_N3, were TP-Link TD-W8961N units supporting IEEE 802.11b/g/n in the 2.400-2.4835 GHz band. The manufacturer specifies a transmit power below 20 dBm EIRP. The AP placement, height, and hardware configuration were kept unchanged during data acquisition. The exact configured transmit-power level, operating Wi-Fi channel, and channel bandwidth used during the measurement session were not preserved in the dataset metadata and are therefore not inferred in this study.

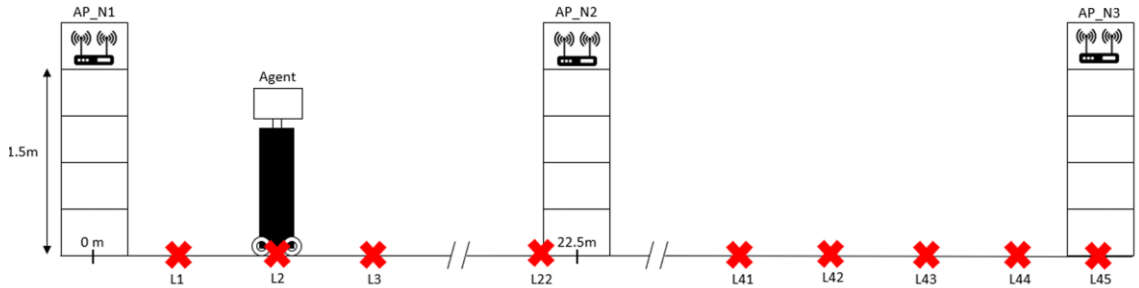


Figure 2. Sequential RSSI data acquisition layout across 45 ordered reference positions

Figure 2 illustrates the sequential RSSI data-acquisition layout across the 45 reference positions. The 1 m spacing gives the position labels a natural order, allowing classification errors to be interpreted in terms of label displacement along the corridor.

At each reference position, RSSI values from the three Access Points were recorded using Homedale on an MSI GL63 8RE laptop running Windows 10. The acquisition interval was approximately 2 s per measurement. The receiver orientation was not retained in the available acquisition metadata and is therefore reported as unavailable rather than inferred. Each raw sample contained three RSSI values and one position label. The resulting dataset consisted of the columns AP_N1, AP_N2, AP_N3, and label. Before model development, records with missing values in these fields were removed, labels were converted to numerical form, and the ordered class sequence from 1 to 45 was verified.

Before model evaluation, RSSI characteristics were examined separately for each Access Point and reference position. The distributional diagnostics included the mean, median, standard deviation, interquartile range, skewness, kurtosis, and the Jarque-Bera normality test at $\alpha = 0.05$. For descriptive comparison, the proportion of position groups with an absolute skewness greater than 0.5 was also calculated. Local RSSI dips were identified at interior reference positions by comparing the mean RSSI at each position with the average mean RSSI of its two adjacent positions. These diagnostics were used only to characterize the measured RSSI data and were not included as model input features.

2.3. RSSI Window Aggregation and Feature Construction

RSSI window aggregation was used to construct the model input features. For the single-sample baseline, one RSSI observation formed one input unit. For larger window sizes, consecutive RSSI observations collected at the same position were grouped before feature extraction.

The RSSI data were grouped into non-overlapping windows for each window size n, with window construction performed separately within each position label. Non-overlapping segmentation was retained so that each raw RSSI observation contributed to only one window, avoiding duplicated observations across adjacent window-level feature vectors and allowing the reduction in sample availability at larger window sizes to be measured explicitly. Because explicit timestamps or burst identifiers were not available in the curated dataset, consecutive samples within each position were arranged according to their original row order. When the number of samples at a position was not exactly divisible by the selected window size, the remaining observations at the end of the sequence were excluded and recorded as dropped samples. Each valid window was represented by the mean and standard deviation of the three RSSI channels, resulting in six window-level features.

$$\mathcal{X}_n = [\mu AP_N1, \mu AP_N2, \mu AP_N3, \sigma AP_N1, \sigma AP_N2, \sigma AP_N3] \quad (1)$$

where μ and σ denote the sample mean and sample standard deviation of RSSI values within each window, respectively. These statistics are used as compact empirical descriptors of local signal level and dispersion and do not imply that the RSSI observations follow a Gaussian distribution. Because indoor RSSI can exhibit asymmetric or non-Gaussian behavior, a median-and-IQR representation was also included in the supplementary representation benchmark to evaluate whether a more robust statistical description improves classification performance.

$$n \in \{1,3,5,7,9,11,15,21,29\} \quad (2)$$

Window size $n = 1$ served as the single-sample baseline, while larger values represented progressively stronger RSSI aggregation. The selected window sizes were used to examine whether increased aggregation consistently improved predictive performance or introduced trade-offs in sample availability, fold-to-fold variability, and evaluation reliability.

2.4. MLP Model and Training Configuration

A Multilayer Perceptron was used for the 45-class indoor position-classification task. The network received six window-level features and contained two hidden layers with 100 and 50 neurons. Training used Scaled Conjugate Gradient with cross-entropy loss. SCG determines the optimization step size adaptively, so no fixed learning rate was specified. Training stopped when either 400 epochs were reached or the performance gradient fell below 1×10^{-6}. Internal data division was disabled because model evaluation was performed externally.

Before training, window-level features were first transformed using robust scaling based on the median and interquartile range estimated from the training fold. The resulting inputs were then processed using the default MATLAB patternnet input-processing functions, removeconstantrows and mapminmax. The preprocessing settings were estimated from the training fold and subsequently applied to the corresponding test fold. Table 1 summarizes the experimental and model configuration.

Table 1. Experimental and model configuration of the RSSI window aggregation study

Component	Configuration
Testbed	Straight indoor corridor, approximately 45 m, 45 ordered reference positions at 1 m spacing
Access Point configuration	3 × TP-Link TD-W8961N, approximately 22.5 m between consecutive APs, AP height 1.5 m
Wireless configuration	IEEE 802.11b/g/n, 2.400-2.4835 GHz, rated transmit power <20 dBm EIRP
Data acquisition	Homedale on MSI GL63 8RE with Windows 10, approximately 2 s per measurement
Dataset	RSSI channels AP_N1, AP_N2, AP_N3, 45 ordered position classes
Window aggregation	Non-overlapping fixed windows, $n = 1,3,5,7,9,11,15,21,29$
Window-level features	Mean and standard deviation of three RSSI channels, 6 input features
MLP architecture	Two hidden layers with 100 and 50 neurons, 45-class output
Training	Scaled Conjugate Gradient, cross-entropy loss, no fixed learning-rate parameter
Stopping conditions	Maximum 400 epochs or minimum performance gradient 1×10^{-6}
Data handling	Training-fold robust scaling followed by removeconstantrows and mapminmax, internal data division disabled
Evaluation protocol	Primary stratified 5-fold CV with supplementary order-blocked 5-fold sensitivity analysis
Metrics	Accuracy, Macro-F1, Top-2 Accuracy, Ordinal Error (label-index MAE)

2.5. Evaluation Protocol and Accuracy-Stability Trade-off Analysis

The primary evaluation used stratified 5-fold cross-validation with a deterministic seed, in which window-level samples were randomly assigned to folds within each position class while preserving class representation. All preprocessing parameters were derived from the training fold only. This included the external robust-scaling parameters and the input-processing settings used by the MLP network, which were subsequently applied to the corresponding test fold. Because the window-level samples were constructed from sequential observations, an additional order-blocked validation was performed to assess sensitivity to possible dependence between nearby windows. For this analysis, windows within each position label were retained in their original order and divided into five contiguous blocks. In each fold, one block was held out for testing, while one adjacent window on each side of the test block was excluded from training as a guard interval. The same MLP architecture, preprocessing procedure, and evaluation metrics were retained. Order-blocked 5-fold validation was feasible for window sizes $n = 1$ to $n = 21$. The window size $n = 29$ was excluded because at least one position class contained only four window-level samples. Since explicit timestamps were unavailable, this order-blocked procedure is interpreted as a sensitivity analysis based on acquisition-order information rather than as a fully time-indexed validation protocol.

To examine whether the observed performance was specific to the selected mean-and-standard-deviation representation, a supplementary representation benchmark was conducted at $n = 7$ using the same order-blocked 5-fold protocol. Five configurations were compared: non-overlapping mean and standard deviation features, non-overlapping median and interquartile range (IQR) features, moving-median filtering followed by mean and standard deviation extraction, Gaussian filtering followed by mean and standard deviation extraction, and an overlapping sliding window with $n = 7$ and stride 3. The moving-median filter used a span of five samples, while the Gaussian filter used a five-sample kernel with $\sigma = 1$. For the sliding-window configuration, two adjacent windows on each side of the test block were excluded from training because consecutive windows shared raw observations. All configurations used the same MLP architecture, SCG training procedure, train-fold preprocessing pipeline, and evaluation metrics.

Model performance was evaluated using Accuracy, Macro-F1, Top-2 Accuracy, and Ordinal Error. Accuracy measures the proportion of correctly classified positions, while Macro-F1 summarizes class-wise performance with equal contribution from each position class. Top-2 Accuracy indicates whether the true position label is included among the two highest model scores. In this study, Ordinal Error is mathematically the mean absolute error (MAE) computed over the ordered position-label indices. The term Ordinal Error is retained only to emphasize that the class labels represent sequential reference positions rather than unrelated nominal categories. It is not introduced as a distinct error metric.

$$OE = MAE_{\text{label}} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (3)$$

where y_i and $\hat{y}_i$ denote the true and predicted position-label indices, respectively, and N is the number of evaluated samples. Since the 45 reference positions are ordered at 1 m intervals along the corridor, an absolute label difference of one corresponds to one reference-position interval in this testbed. Accordingly, this metric describes the average prediction displacement in ordered label units. It should not be interpreted as a general two-dimensional Euclidean localization error.

Performance was summarized using the mean and standard deviation across the five folds. Mean values represent predictive performance, while fold-to-fold standard deviation is used as an indicator of evaluation stability. Sample availability was considered jointly with these metrics because increasing the

window size reduces the number of available window-level samples. In this study, stability refers to fold-to-fold variability in the evaluation metrics and should not be interpreted as a direct measure of RSSI signal variance.

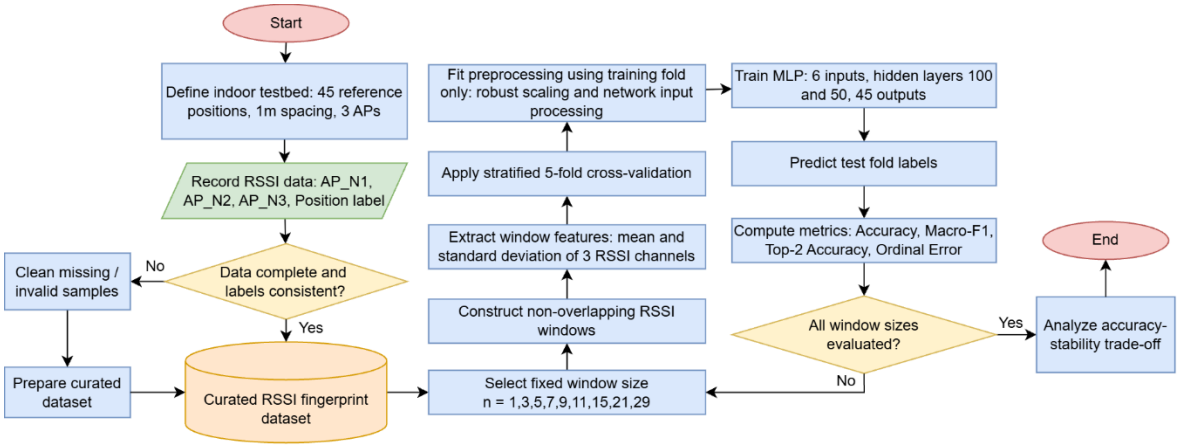


Figure 3. Proposed workflow for RSSI window aggregation and MLP-based Wi-Fi indoor positioning evaluation

Figure 3 summarizes the primary experimental workflow used for fixed-window evaluation. The procedure includes testbed definition, RSSI acquisition and validation, non-overlapping window construction, mean-and-standard-deviation feature extraction, stratified 5-fold cross-validation, train-fold preprocessing, MLP training, and metric computation. The procedure was repeated for all fixed window sizes. Supplementary order-blocked validation and alternative-representation benchmarking were subsequently conducted using the procedures described in Section 2.5.

3. RESULTS AND DISCUSSION

3.1. RSSI Characteristics across Reference Positions

Before model evaluation, RSSI characteristics were examined across the 45 reference positions to assess the spatial variation available for position discrimination. Figure 4 summarizes the mean RSSI profiles of AP_N1, AP_N2, and AP_N3 along the corridor.

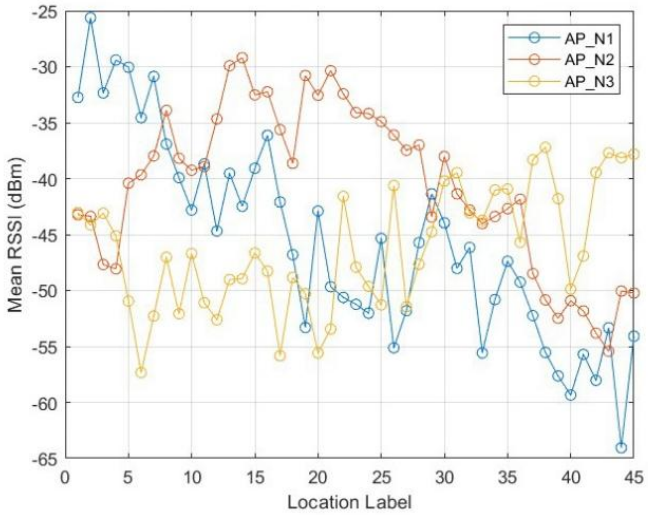


Figure 4. Mean RSSI variation across reference positions for three Access Points

Figure 4 shows that the mean RSSI does not vary monotonically across the 45 reference positions. A local-dip analysis was therefore conducted by comparing the mean RSSI at each position with the average of its two adjacent positions. The strongest local negative deviations were observed at L44 for AP_N1 with a dip of 10.35 dB, at L29 for AP_N2 with a dip of 5.92 dB, and at L27 for AP_N3 with a dip of 7.30 dB. Other pronounced deviations were also observed at L19 and L33 for AP_N1, L18 and L4 for AP_N2, and L17 and

L25 for AP_N3. These results confirm that the RSSI profile contains localized spatial variations rather than a smooth distance-dependent trend. Because synchronized information on obstruction, channel condition, and environmental activity was not recorded, the individual dips are not assigned to a specific propagation mechanism. They are interpreted conservatively as location-dependent RSSI variations that are consistent with the behavior of indoor wireless propagation.

Distributional diagnostics also showed that RSSI within individual position and Access Point groups was frequently non-Gaussian. The Jarque-Bera test rejected normality for 73.3% of AP_N1 position groups, 82.2% of AP_N2 groups, and 68.9% of AP_N3 groups. An absolute skewness greater than 0.5 occurred in 40.0%, 37.8%, and 31.1% of the corresponding groups. Nevertheless, the median absolute differences between the mean and median remained below 1 dB for all three channels, while the median position-wise IQR values were 5 dB for AP_N1 and 4 dB for AP_N2 and AP_N3. These findings indicate that non-Gaussian behavior is common but not uniformly associated with strong skewness. Accordingly, mean and standard deviation are treated as empirical window descriptors rather than Gaussian-distribution parameters. This interpretation is also supported by the representation benchmark in Table 4, where median-and-IQR features produced lower Accuracy than the mean-and-standard-deviation representation under the same order-blocked protocol.

3.2. Overall Performance across Fixed RSSI Window Sizes

Table 2 summarizes MLP performance across the evaluated fixed RSSI window sizes. The $n = 1$ configuration served as the single-sample baseline, while larger window sizes represent aggregation of consecutive RSSI observations. Performance was assessed using Accuracy, Top-2 Accuracy, Macro-F1, and Ordinal Error.

Table 2. Performance of MLP-based Wi-Fi indoor positioning across fixed RSSI window sizes

Window size (n)	Window samples	Dropped samples	Accuracy (%)	Top-2 Accuracy (%)	Macro-F1	Ordinal Error
1	6292	0	68.24 ± 2.29	83.50 ± 1.43	0.6774 ± 0.0230	1.8751 ± 0.1577
3	2083	43	83.50 ± 1.82	91.93 ± 1.13	0.8321 ± 0.0173	0.8126 ± 0.1210
5	1241	87	86.48 ± 2.13	93.76 ± 2.41	0.8607 ± 0.0248	0.6539 ± 0.1714
7	881	125	90.60 ± 0.71	95.08 ± 1.31	0.9047 ± 0.0060	0.4319 ± 0.1364
9	678	190	91.20 ± 2.60	96.33 ± 0.90	0.9084 ± 0.0252	0.3890 ± 0.2723
11	552	220	91.50 ± 2.65	96.13 ± 1.81	0.9081 ± 0.0274	0.3927 ± 0.1390
15	399	307	91.55 ± 2.83	96.81 ± 1.84	0.8982 ± 0.0261	0.4774 ± 0.3515
21	277	475	93.44 ± 4.53	96.45 ± 2.89	0.9166 ± 0.0593	0.2413 ± 0.1658
29	188	840	94.26 ± 4.30	97.35 ± 2.88	0.9267 ± 0.0542	0.4121 ± 0.3159

Table 2 shows a substantial performance improvement from the single-sample baseline to short-window aggregation, followed by smaller gains as the window size increases.

3.3. Effect of Window Aggregation on Accuracy, Macro-F1, and Ordinal Error

The greatest performance improvement occurred when the input representation changed from the single-sample baseline to short-window aggregation. Accuracy increased by 15.26 percentage points, from 68.24% at $n = 1$ to 83.50% at $n = 3$. This result indicates that the single-sample representation was less effective than short-window aggregation for the evaluated dataset.

At $n = 5$, Accuracy increased to 86.48% and reached 90.60% at $n = 7$. Under the primary stratified evaluation, $n = 7$ produced the lowest fold-to-fold Accuracy standard deviation at 0.71 percentage points while maintaining high predictive performance. This result identifies $n = 7$ as a stable operating point under the primary evaluation protocol, rather than as a universally optimal window size.

Macro-F1 follows a pattern consistent with Accuracy. At $n = 1$, Macro-F1 is 0.6774, then increases to 0.8321 at $n = 3$, 0.8607 at $n = 5$, and 0.9047 at $n = 7$. The increase in Macro-F1 indicates that window aggregation not only improves overall accuracy but also helps the model maintain more balanced performance across the 45 position classes.

Ordinal Error, interpreted here as MAE over the ordered position-label indices, also decreased after window aggregation was applied. It declined from 1.8751 at $n = 1$ to 0.8126 at $n = 3$, 0.6539 at $n = 5$, and 0.4319 at $n = 7$. Because the reference positions are sequentially ordered at 1 m intervals, the reduction indicates that incorrect predictions tend to occur across fewer position-label intervals. This interpretation is specific to the one-dimensional ordered testbed and should not be generalized as a Euclidean positioning error.

3.4. Accuracy and Sample Availability Trade-off

Although larger windows tended to improve Accuracy, they also reduced the number of available window-level samples. The total decreased from 6,292 at $n = 1$ to 2,083 at $n = 3$, 881 at $n = 7$, 277 at $n =$

21, and 188 at $n = 29$. As sample availability declines, the evaluation may become more sensitive to fold partitioning.

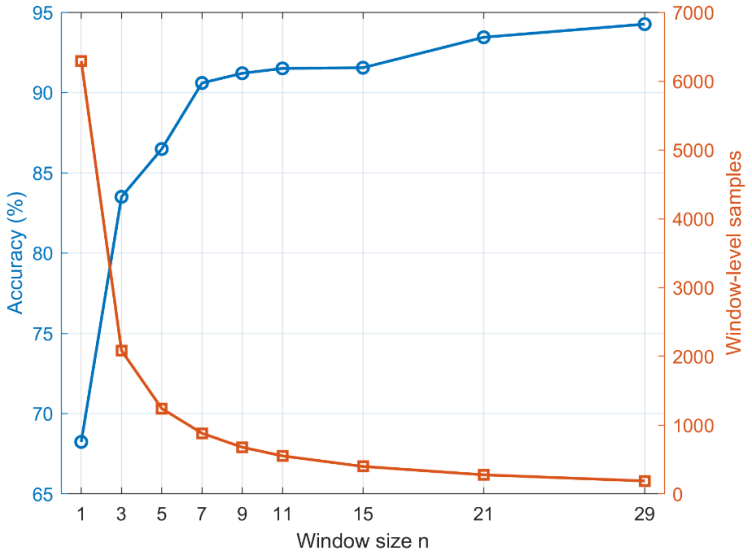


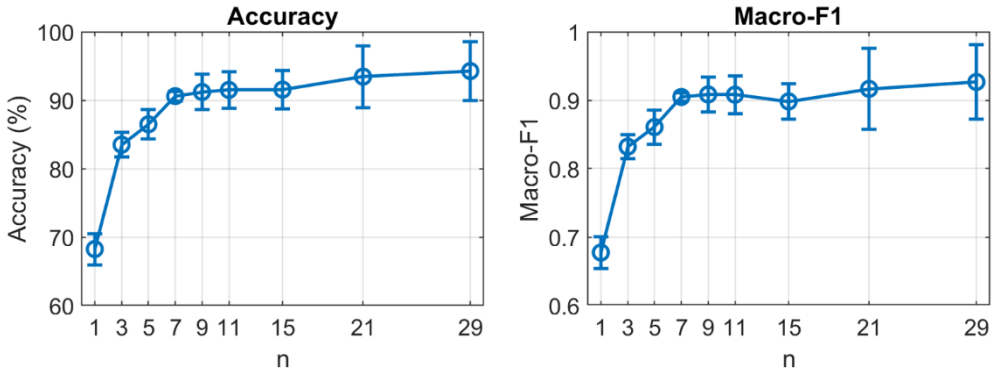
Figure 5. Accuracy and sample availability trade-off across fixed RSSI window sizes

Figure 5 shows that higher average Accuracy is accompanied by a sharp reduction in window-level sample availability. Consequently, window-size selection cannot be based on Accuracy alone. As the number of window-level samples decreases, evaluation becomes increasingly sensitive to data partitioning, as reflected by the larger fold-to-fold variability observed at the largest window sizes.

At $n = 29$, average Accuracy reached 94.26%, but only 188 window-level samples remained and the fold-to-fold standard deviation increased to 4.30 percentage points. In contrast, $n = 7$ achieved $90.60 \pm 0.71\%$ Accuracy with 881 window-level samples. Under the primary stratified evaluation, $n = 7$ therefore provides a favorable balance among predictive performance, evaluation stability, and sample availability.

3.5. Fixed-Window Sensitivity Analysis

Figures 6 and 7 summarize the sensitivity of MLP performance to fixed RSSI window size. Figure 6 presents Accuracy and Macro-F1, while Figure 7 shows Ordinal Error and the corresponding number of available window-level samples. Error bars in Figures 6(a), 6(b), and 7(a) represent the standard deviation across the five stratified folds. Figure 7(b) reports the resulting number of window-level samples and therefore does not include error bars.



(a) (b)

Figure 6. Predictive performance across fixed RSSI window sizes: (a) Accuracy and (b) Macro-F1

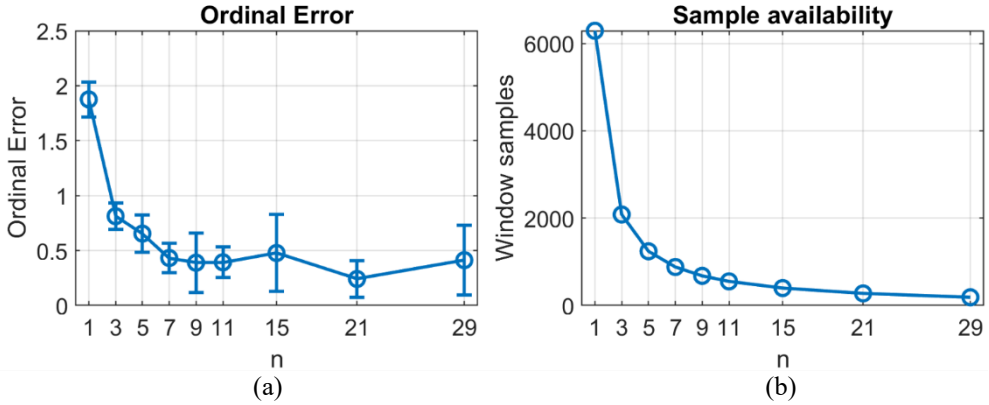


Figure 7. Prediction error and sample availability across fixed RSSI window sizes: (a) Ordinal Error and (b) window-level sample availability

Under the primary stratified evaluation, $n = 7$ provides a favorable balance among Accuracy, Macro-F1, Ordinal Error, fold-to-fold variability, and sample availability. Larger windows, particularly $n = 21$ and $n = 29$, achieve higher average Accuracy but retain substantially fewer window-level samples and exhibit greater variability across folds. This interpretation is further examined using the order-blocked validation reported in Table 3.

Table 3. Sensitivity of fixed-window performance to order-blocked validation

n	Window samples	Stratified Accuracy (%)	Order-blocked Accuracy (%)	Order-blocked Macro-F1	Order-blocked Ordinal Error
1	6292	68.24 ± 2.29	63.10 ± 5.40	0.6240 ± 0.0525	2.2318 ± 0.2788
3	2083	83.50 ± 1.82	80.56 ± 2.64	0.8005 ± 0.0242	0.9532 ± 0.1787
5	1241	86.48 ± 2.13	86.36 ± 2.73	0.8598 ± 0.0298	0.6547 ± 0.2134
7	881	90.60 ± 0.71	88.26 ± 1.99	0.8747 ± 0.0211	0.6331 ± 0.0809
9	678	91.20 ± 2.60	89.84 ± 1.43	0.8906 ± 0.0125	0.5091 ± 0.0951
11	552	91.50 ± 2.65	91.45 ± 2.94	0.9070 ± 0.0319	0.4466 ± 0.1693
15	399	91.55 ± 2.83	90.99 ± 3.70	0.8959 ± 0.0445	0.5950 ± 0.3171
21	277	93.44 ± 4.53	86.49 ± 3.27	0.8264 ± 0.0357	0.7008 ± 0.3311
29	188	94.26 ± 4.30	Not feasible	–	–

Table 3 shows that preserving the within-position ordering produced a more conservative assessment than the primary stratified cross-validation. The largest differences occurred at the smallest and largest window sizes, whereas several intermediate configurations remained comparatively stable. For $n = 7$, Accuracy decreased from 90.60 ± 0.71% to 88.26 ± 1.99%. The blocked evaluation also changed the relative behavior of the intermediate windows. Window $n = 9$ produced the smallest Accuracy standard deviation at 1.43%, while $n = 11$ achieved the highest blocked Accuracy of 91.45 ± 2.94% and the highest Macro-F1 of 0.9070 ± 0.0319. In contrast, $n = 21$ showed a substantial performance reduction, and $n = 29$ could not support five blocked folds because some position classes contained fewer than five windows.

These results indicate that the choice of a fixed RSSI window is sensitive not only to the amount of temporal aggregation but also to the validation protocol. The primary stratified evaluation identifies $n = 7$ as a strong balance between performance, fold-to-fold variability, and sample availability, but the order-blocked analysis does not support treating $n = 7$ as a universal optimum. Instead, the favorable region extends approximately from $n = 7$ to $n = 11$, with different operating points emphasizing sample availability, stability, or predictive performance.

Table 4. Comparison of alternative RSSI representations under order-blocked validation

Representation	Window samples	Accuracy (%)	Top-2 Accuracy (%)	Macro-F1	Ordinal Error
Non-overlap mean+SD, $n = 7$	881	88.26 ± 1.99	94.70 ± 1.60	0.8747 ± 0.0211	0.6331 ± 0.0809
Non-overlap median+IQR, $n = 7$	881	79.17 ± 3.04	89.05 ± 2.01	0.7812 ± 0.0319	1.1486 ± 0.1479
Median filter + mean+SD	881	78.86 ± 2.72	89.15 ± 1.89	0.7767 ± 0.0252	1.0462 ± 0.2304
Gaussian filter + mean+SD	881	81.10 ± 1.98	91.32 ± 1.92	0.8021 ± 0.0186	1.1120 ± 0.2011
Sliding $n = 7$, stride 3, mean+SD	2021	90.66 ± 2.28	95.65 ± 1.50	0.9028 ± 0.0224	0.4583 ± 0.1281

Table 4 compares the main $n = 7$ representation with alternative statistical, filtering, and windowing schemes under the same order-blocked evaluation. Among the non-overlapping configurations, the mean-and-standard-deviation representation achieved the highest Accuracy at 88.26 ± 1.99% and a Macro-F1 of 0.8747

± 0.0211. Replacing these features with median and IQR reduced Accuracy to 79.17 ± 3.04%. Similarly, additional moving-median and Gaussian preprocessing did not improve the non-overlapping baseline. Gaussian filtering reached 81.10 ± 1.98%, while moving-median filtering produced 78.86 ± 2.72%. These results indicate that additional smoothing did not provide a performance advantage for the evaluated non-overlapping representation.

The overlapping sliding-window configuration produced the strongest predictive results, reaching 90.66 ± 2.28% Accuracy, 95.65 ± 1.50% Top-2 Accuracy, a Macro-F1 of 0.9028 ± 0.0224, and an Ordinal Error of 0.4583 ± 0.1281. It also increased the number of window-level samples from 881 to 2021. This improvement, however, should be interpreted together with the additional redundancy introduced by overlapping windows. The guard interval prevented direct overlap across the train-test boundary, but adjacent windows within each partition remained statistically dependent. Thus, sliding windows provide an alternative operating point with greater temporal coverage and sample availability, rather than evidence that overlapping segmentation is universally superior.

3.6. Discussion

Across the evaluated configurations, RSSI window aggregation improved predictive performance relative to the single-sample baseline, while larger windows reduced sample availability and increased sensitivity to validation design.

The reduction in Ordinal Error observed after window aggregation should be interpreted as an empirical result rather than as a theoretically derived consequence of RSSI variance reduction. Although temporal aggregation can reduce short-term variability in the input representation, this study does not assume or derive a deterministic relationship between within-window RSSI variance and the MAE of the predicted position-label indices. These quantities describe different stages of the positioning pipeline: the former characterizes variability in the input signal, whereas the latter measures prediction displacement in the ordered output space. Their simultaneous improvement is therefore reported as an empirical association observed under the evaluated configurations.

Accuracy gains should be considered together with sample availability and fold-to-fold variability. Larger windows aggregate more consecutive observations but leave fewer window-level samples, making cross-validation results more sensitive to data partitioning. Because the MLP architecture and training configuration were held constant, the observed differences primarily reflect changes in RSSI representation and data availability.

The supplementary representation benchmark further shows that preprocessing choice can materially affect the resulting fingerprint representation. The unfiltered mean-and-standard-deviation representation outperformed the median-and-IQR representation and the configurations using additional median or Gaussian filtering under the same non-overlapping protocol. This suggests that, for the present dataset, the local variations retained by the original window representation contain useful position-discriminative information that may be weakened by additional smoothing. The sliding-window configuration achieved higher predictive performance and provided more window-level samples, but this advantage was accompanied by stronger dependence among neighboring windows. Similar sliding-window approaches have been used in RSSI-based positioning to increase temporal coverage and suppress interference [35], while median, Gaussian, and moving-average-based filtering have also been investigated as preprocessing alternatives [29,30,32].

The distributional analysis provides additional context for the representation benchmark. Although normality was rejected for most position-specific RSSI groups, the magnitude and direction of skewness varied across positions and Access Points. This indicates that no single distributional model adequately describes all RSSI observations in the evaluated testbed. Therefore, the stronger performance of mean-and-standard-deviation features relative to median-and-IQR features should not be interpreted as evidence that RSSI follows a Gaussian distribution. Rather, the results indicate that, for the present dataset, local average signal level and within-window dispersion retained more position-discriminative information for the MLP. This interpretation remains empirical and specific to the evaluated measurement environment.

More elaborate temporal representations are also possible. For example, wavelet-based RSSI feature extraction has been used to transform signal sequences into discriminative temporal features before classification [36]. Such an approach, however, changes both the feature space and the associated modeling choices, making it less suitable as a direct controlled comparison with the six-dimensional statistical representation evaluated here. The present benchmark therefore focuses on alternatives that can be incorporated without changing the MLP architecture or the fundamental dimensionality of the input.

The combined results show that no single fixed window should be regarded as universally optimal. Under the primary stratified evaluation, $n = 7$ provides a favorable balance because it retains 881 window-level samples while achieving 90.60 ± 0.71% Accuracy. However, the order-blocked sensitivity analysis shifts the strongest operating region toward $n = 9$ and $n = 11$. Window $n = 9$ provides the lowest fold-to-fold

Accuracy variability under blocked validation, whereas $n = 11$ provides the highest blocked Accuracy and Macro-F1. This difference indicates that window-size selection is partly dependent on how sequential dependence is handled during evaluation. Consequently, the practical interpretation is not that $n = 7$ is universally superior, but that moderate window sizes in the range of approximately $n = 7$ to $n = 11$ provide the most defensible balance for the evaluated dataset.

The interpretation of these results remains limited by the deliberately constrained measurement topology. The dataset was collected in a single straight 45 m corridor using three fixed Access Points under one hardware and spatial configuration. Therefore, the evaluated setting does not capture the propagation diversity associated with room-rich layouts, corners, multiple floors, or strongly non-line-of-sight conditions. The AP hardware, frequency band, nominal transmit-power specification, and sampling interval were documented. However, the exact operating channel, channel bandwidth, configured transmit-power level, and receiver orientation were not retained in the measurement metadata. These limitations restrict propagation-specific interpretation and direct extrapolation to other wireless configurations. Future evaluation should therefore include multi-session and multi-environment measurements with varied AP layouts and fully recorded radio parameters.

4. CONCLUSION

RSSI window aggregation substantially improved MLP-based indoor position classification relative to the single-sample baseline. Under the primary stratified evaluation, the baseline achieved 68.24 ± 2.29% Accuracy, while $n = 7$ achieved 90.60 ± 0.71% with 881 window-level samples. Larger windows produced higher average Accuracy but substantially reduced sample availability and increased fold-to-fold variability. Supplementary order-blocked validation provided a more conservative assessment and shifted the strongest operating region toward $n = 9$ to $n = 11$. These results show that fixed-window selection should consider predictive performance, evaluation stability, sample availability, and validation design rather than maximum Accuracy alone.

The additional representation benchmark showed that median and Gaussian prefiltering did not improve the non-overlapping mean-and-standard-deviation representation. An overlapping sliding window produced higher predictive performance and more window-level samples, but also introduced greater dependence between neighboring windows. Distributional analysis further showed that RSSI was frequently non-Gaussian, supporting the interpretation of mean and standard deviation as empirical descriptors rather than assumptions about the underlying distribution.

Future work should examine adaptive window aggregation and evaluate the approach across multiple sessions, environments, AP layouts, and fully documented radio configurations. Comparisons with alternative learning architectures may also clarify whether the observed aggregation effects remain consistent across different classifiers.

REFERENCES

[1] He S, Chan SHG. Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons. IEEE Communications Surveys and Tutorials 2016;18:466–90. https://doi.org/10.1109/COMST.2015.2464084.

[2] Shang S, Wang L. Overview of WiFi fingerprinting-based indoor positioning. IET Communications 2022;16:725–33. https://doi.org/10.1049/cmu2.12386.

[3] Singh N, Choe S, Punmiya R, Kaur N. XGBLoc: XGBoost-Based Indoor Localization in Multi-Building Multi-Floor Environments. Sensors 2022;22:6629. https://doi.org/10.3390/s22176629.

[4] Lin X, Gan J, Jiang C, Xue S, Liang Y. Wi-Fi-Based Indoor Localization and Navigation: A Robot-Aided Hybrid Deep Learning Approach. Sensors 2023;23:6320. https://doi.org/10.3390/s23146320.

[5] Martín-Frechina S, Dura E, Miralles I, Torres-Sospedra J. From Fingerprinting to Advanced Machine Learning: A Systematic Review of Wi-Fi and BLE-Based Indoor Positioning Systems. Sensors 2025;25:6946. https://doi.org/10.3390/s25226946.

[6] Neupane I, Shahrestani S, Ruan C. Wi-Fi RSS Fingerprinting-Based Indoor Localization in Large Multi-Floor Buildings. Electronics (Basel) 2025;15:183. https://doi.org/10.3390/electronics15010183.

[7] Zholamanov B, Saymbetov A, Nurgaliyev M, Bolatbek A, Dosymbetova G, Kuttybay N, et al. RSSI Fingerprint-Based Indoor Localization Solutions Using Machine Learning Algorithms: A Comprehensive Review. Smart Cities 2025;8:153. https://doi.org/10.3390/smartcities8050153.

[8] Yang L, Finnerty P, Ohta C. Indoor localization using router-to-router RSSI and transfer learning for dynamic environments. Ad Hoc Networks 2025;178:103938. https://doi.org/10.1016/j.adhoc.2025.103938.

[9] Silva I, Pendão C, Moreira A. Collection of a Continuous Long-Term Dataset for the Evaluation of Wi-Fi-Fingerprinting-Based Indoor Positioning Systems. Sensors 2022;22:8585. https://doi.org/10.3390/s22228585.

[10] Maduranga MWP, Tilwari V, Abeysekera R. Improved-RSSI-based indoor localization by using pseudo-linear solution with machine learning algorithms. Journal of Electrical Systems and Information Technology 2024;11:10. https://doi.org/10.1186/s43067-024-00138-0.

[11] Wan S, Gu C, Shu Y, Shi Z. Last-seen time is critical: Revisiting RSSI-based WiFi indoor localization. Signal Processing 2025;231:109756. https://doi.org/10.1016/j.sigpro.2024.109756.

[12] Yang T, Cabani A, Chafouk H. A Survey of Recent Indoor Localization Scenarios and Methodologies. Sensors 2021;21:8086. https://doi.org/10.3390/s21238086.
[13] Ma Z, Shi K. Few-Shot Learning for WiFi Fingerprinting Indoor Positioning. Sensors 2023;23:8458. https://doi.org/10.3390/s23208458.
[14] Cha J, Lim E. A hierarchical auxiliary deep neural network architecture for large-scale indoor localization based on Wi-Fi fingerprinting. Appl Soft Comput 2022;120:108624. https://doi.org/10.1016/j.asoc.2022.108624.
[15] Feng X, Nguyen KA, Luo Z. A survey of deep learning approaches for WiFi-based indoor positioning. Journal of Information and Telecommunication 2022;6:163–216. https://doi.org/10.1080/24751839.2021.1975425.
[16] Gao B, Yang F, Cui N, Xiong K, Lu Y, Wang Y. A Federated Learning Framework for Fingerprinting-Based Indoor Localization in Multibuilding and Multifloor Environments. IEEE Internet Things J 2023;10:2615–29. https://doi.org/10.1109/JIOT.2022.3214211.
[17] Bi J, Zhao M, Yao G, Cao H, Feng Y, Jiang H, et al. PSOSVRPos: WiFi indoor positioning using SVR optimized by PSO. Expert Syst Appl 2023;222:119778. https://doi.org/10.1016/j.eswa.2023.119778.
[18] Zhu H, Cheng L, Li X, Yuan H. Neural-Network-Based Localization Method for Wi-Fi Fingerprint Indoor Localization. Sensors 2023;23. https://doi.org/10.3390/s23156992.
[19] Turgut Z, Kakisim AG. An explainable hybrid deep learning architecture for WiFi-based indoor localization in Internet of Things environment. Future Generation Computer Systems 2024;151:196–213. https://doi.org/10.1016/j.future.2023.10.003.
[20] Arslantas H, Okdem S. Indoor Localization With an Autoencoder-Based Convolutional Neural Network. IEEE Access 2024;12:46059–66. https://doi.org/10.1109/ACCESS.2024.3382135.
[21] Yoo J. Wi-Fi Fingerprint Indoor Localization by Semi-Supervised Generative Adversarial Network. Sensors 2024;24:5698. https://doi.org/10.3390/s24175698.
[22] Kim D, Park J-H, Suh Y-J. A Wi-Fi Fingerprinting Indoor Localization Framework Using Feature-Level Augmentation via Variational Graph Auto-Encoder. Electronics (Basel) 2025;14:2807. https://doi.org/10.3390/electronics14142807.
[23] Li S, Kim KS, Tang Z, Smith JS. Hierarchical Stagewise Training of Linked Deep Neural Networks for Multibuilding and Multifloor Indoor Localization Based on Wi-Fi RSSI Fingerprinting. IEEE Sens J 2025;25:23341–51. https://doi.org/10.1109/JSEN.2024.3455554.
[24] Esmaeili Gorjan H, Gil Jiménez VP. Improving Indoor WiFi Localization by Using Machine Learning Techniques. Sensors 2024;24:6293. https://doi.org/10.3390/s24196293.
[25] Tang Z, Li S, Kim KS, Smith JS. Multi-Dimensional Wi-Fi Received Signal Strength Indicator Data Augmentation Based on Multi-Output Gaussian Process for Large-Scale Indoor Localization †. Sensors 2024;24. https://doi.org/10.3390/s24031026.
[26] Reyes JMR, Ho IW-H, Mak M-W. Wi-Fi CSI fingerprinting-based indoor positioning using deep learning and vector embedding for temporal stability. Expert Syst Appl 2025;264:125802. https://doi.org/10.1016/j.eswa.2024.125802.
[27] Shen Q, Cui Y. WIFI Fingerprint Correction for Indoor Localization Based on Distance Metric Learning. IEEE Sens J 2024;24:36167–77. https://doi.org/10.1109/JSEN.2024.3462759.
[28] Aviles M, Alvarez-Alvarado JM, Robles-Ocampo J-B, Sevilla-Camacho PY, Rodríguez-Reséndiz J. Optimizing RNNs for EMG Signal Classification: A Novel Strategy Using Grey Wolf Optimization. Bioengineering 2024;11:77. https://doi.org/10.3390/bioengineering11010077.
[29] Sun C, Zhou J, Jang K, Kim Y. Indoor Localization Based on Integration of Wi-Fi with Geomagnetic and Light Sensors on an Android Device Using a DFF Network. Electronics (Basel) 2023;12:5032. https://doi.org/10.3390/electronics12245032.
[30] Tseng CH, Tsaur W-J. FFK: Fourier-Transform Fuzzy-c-means Kalman-Filter Based RSSI Filtering Mechanism for Indoor Positioning. Sensors 2023;23:8274. https://doi.org/10.3390/s23198274.
[31] Wei Z, Zhou Z. A Combined Filtering Method for ZigBee Indoor Distance Measurement. Sensors 2024;24:3164. https://doi.org/10.3390/s24103164.
[32] Antonopoulos K, Skandamis T, Alogdianakis G, Faliagka E, Antonopoulos CP, Voros N. Indoor Localization and ADL Monitoring via RSSI-Driven ML with Feedback Process. Electronics (Basel) 2025;14:3759. https://doi.org/10.3390/electronics14193759.
[33] Pawan, Dhiman R. Machine learning techniques for electroencephalogram based brain-computer interface: A systematic literature review. Measurement: Sensors 2023;28:100823. https://doi.org/10.1016/j.measen.2023.100823.
[34] Ismail AH, Kitagawa H, Tasaki R, Terashima K. WiFi RSS fingerprint database construction for mobile robot indoor positioning system. 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE; 2016, p. 001561–6. https://doi.org/10.1109/SMC.2016.7844461.
[35] Gomes EL, Fonseca MSP, Lazzaretti AE, Munaretto A, Guerber CR. Sliding Window, Hierarchical Classification, Regression, and Genetic Algorithm for RFID Indoor Positioning Systems. Expert Syst Appl 2024;238:122298. https://doi.org/10.1016/j.eswa.2023.122298.
[36] Kamal AHM, Alam MdGR, Hassan MR, Apon TS, Hassan MM. Explainable indoor localization of BLE devices through RSSI using recursive continuous wavelet transformation and XGBoost classifier. Future Generation Computer Systems 2023;141:230–42. https://doi.org/10.1016/j.future.2022.11.001.

BIOGRAPHY OF AUTHORS



Wahyu Arthanugraha
Wahyu Arthanugraha is currently a lecturer at the Informatics Engineering Study Program, Faculty of Engineering, Universitas Muhammadiyah Parepare. He obtained his Bachelor's degree in Informatics Engineering from Universitas Muhammadiyah Parepare in 2020. He then earned his Master's degree in Computer Systems from Universitas Handayani Makassar in 2023. He is currently pursuing his Ph.D. in Computer Engineering at Universiti Malaysia Perlis (UniMAP), Malaysia.



Abdul Halim Ismail
Abdul Halim Ismail is currently an Associate Professor at Universiti Malaysia Perlis (UniMAP) serving at the Department of Mechatronics, Faculty of Electrical Engineering & Technology. He obtained his Bachelor of Engineering (Electrical and Electronics Engineering) and Master of Science (Control and Automation Engineering) from Universiti Putra Malaysia (UPM), respectively in 2007 and 2010. In 2017, he was awarded the Doctor of Engineering specializing in robotics engineering from Toyohashi University of Technology, Japan.



Shahrul Nizam Yaakob
Shahrul Nizam Yaakob obtained his B.Eng. (Hons.) in Electrical-Electronic Engineering from Universiti Teknologi Malaysia (UTM) and his M.Sc. by Research in Computer Engineering from Universiti Malaysia Perlis (UniMAP). He received his Ph.D. in Computer System Engineering from the University of South Australia (UniSA), Australia.