

Deep Attention-based Sequential Ensemble Learning for BLE-Based Indoor Localization in Care Facilities

Minh Triet Pham

*School of Computing and
Information Systems*

*The University of Melbourne
Melbourne, Australia*

minhtrietp@student.unimelb.edu.au

Quynh Chi Dang

*School of Computing and
Information Systems*

*The University of Melbourne
Melbourne, Australia*

quynhchid@student.unimelb.edu.au

Le Nhat Tan

Faculty of Applied Science

*Ho Chi Minh City University of Technology
Vietnam National University - Ho Chi Minh City
Ho Chi Minh City, Vietnam*

lenhattan@hcmut.edu.vn

Abstract—Indoor localization systems in care facilities enable optimization of staff allocation, workload management, and quality of care delivery. Traditional machine learning approaches to Bluetooth Low Energy (BLE)-based localization treat each temporal measurement as an independent observation, fundamentally limiting their performance. To address this limitation, this paper, developed as part of the ABC 2026 Activity and Location Recognition Challenge, introduces Deep Attention-based Sequential Ensemble Learning (DASEL), a novel framework that reconceptualizes indoor localization as a sequential learning problem. The framework integrates frequency-based feature engineering, bidirectional GRU networks with attention mechanisms, multi-directional sliding windows, and confidence-weighted temporal smoothing to capture human movement trajectories. Evaluated on real-world data from a care facility using 4-fold temporal cross-validation, DASEL achieves a macro F1 score of 0.4438, representing a 53.1% improvement over the best traditional baseline (0.2898). This breakthrough demonstrates that modeling temporal dependencies in movement patterns is essential for accurate indoor localization in complex real-world environments.

Index Terms—Indoor localization, BLE, sequential learning, attention mechanism, care facilities, GRU networks

I. INTRODUCTION

Indoor localization systems in care facilities enable optimization of staff allocation, workload management, and quality of care delivery [1, 2]. Accurate tracking of caregiver movements provides insights into care routines, enhances hand hygiene monitoring, and supports health interventions for elderly residents with conditions such as Alzheimer’s and dementia [3, 4]. Automated location recording eliminates manual logging burdens and provides objective data for facility management and quality improvement initiatives [5].

Bluetooth Low Energy (BLE) technology has emerged as a prominent indoor localization solution due to its low cost, minimal power consumption, and ease of deployment [6]. Indoor positioning methods have evolved from simple RSSI measurements to more advanced approaches such as CSI, RTT, and AoA, increasingly combined with Machine Learning techniques [7]. Traditional approaches predominantly rely on RSSI

fingerprinting with classical classification algorithms including K-Nearest Neighbors, Support Vector Machines, and Random Forest classifiers [8, 9]. However, these approaches share a fundamental limitation: they treat each temporal measurement as an independent observation, extracting statistical features from beacon signals and classifying each moment in isolation. The growing complexity of indoor environments requires solutions that can handle sensor noise, multipath fading effects, and temporal dependencies that traditional independent-window classification methods cannot adequately address [10].

BLE-based localization in real-world care facilities confronts significant data quality challenges. BLE signals experience substantial fluctuations in RSSI values caused by multipath propagation and environmental interference [11, 12]. Care facilities present particularly complex deployment scenarios characterized by beacon placement constraints and architectural complexity [8]. These environments exhibit several documented challenges: signal instability arising from multipath effects and device heterogeneity [13], spatial sparsity with limited beacon coverage in certain areas [11], temporal irregularity with variable detection rates [14], and severe class imbalance where common areas receive significantly more visits than individual patient rooms [15]. Traditional machine learning methods applied to such real-world datasets demonstrate limited performance despite various optimization efforts [9, 10]. This performance plateau reveals a critical limitation: the independence assumption inherent in traditional approaches discards temporal dependencies in human movement patterns [16]. Caregivers follow continuous trajectories through physical space, where location at time t strongly predicts location at time $t + 1$, and transitions between rooms produce gradual shifts in beacon signal patterns [17]. Traditional methods treat each temporal window as an isolated classification problem, thereby discarding contextual information about movement trajectories, dwell times, and transition dynamics.

This work addresses the location recognition component of the ABC 2026 Activity and Location Recognition Challenge,

which evaluates methods using macro F1 score as the primary metric for location classification. Our research objectives are threefold: (1) *Methodological goal*: achieve breakthrough performance improvement by leveraging temporal dependencies in human movement trajectories that traditional independent-window classification methods fundamentally discard; (2) *Technical improvements*: address RSSI instability through robust feature representations, handle the inference challenge where sequence boundaries are unknown during real-time prediction, and effectively manage severe class imbalance across 12-18 room locations; (3) *Evaluation targets*: maximize location classification macro F1 score while maintaining balanced performance across all room classes regardless of visit frequency and achieving substantial performance improvements suitable for practical deployment in real-world care facilities.

The key contributions of this work demonstrate measurable improvements addressing these objectives: (1) a novel sequential framework applying deep bidirectional recurrent networks with attention mechanisms to BLE-based care facility localization, explicitly modeling temporal movement trajectories rather than treating measurements as independent observations, (2) frequency-based representation addressing RSSI instability challenges inherent in real-world deployments, providing device-agnostic features robust to signal fluctuations, (3) two-level hierarchical ensemble combining multi-seed variance reduction with multi-directional positional robustness to handle unknown sequence boundaries during inference, (4) comprehensive baseline evaluation demonstrating that traditional paradigm optimization yields diminishing returns with performance plateauing at 0.2898 macro F1, and (5) DASEL framework achieving 0.4438 macro F1 score (53.1% improvement), with balanced performance across all room classes and robustness to device heterogeneity, suitable for real-world deployment.

II. METHODOLOGY

We propose and compare two distinct approaches (Section II-C1 and Section II-C2) representing fundamentally different paradigms in handling temporal sensor data. The traditional machine learning approach treats each temporal window independently, extracting statistical features from RSSI values and applying gradient boosting classification. In contrast, the Deep Attention-based Sequential Ensemble Learning (DASEL) framework reconceptualizes indoor localization as a sequential learning problem, leveraging the inherent temporal continuity of human movement patterns through deep learning architectures combining frequency-based feature engineering, bidirectional recurrent networks with attention mechanisms, multi-directional sliding windows, and multi-model ensemble learning.

A. Dataset

The dataset was provided by Kyushu Institute of Technology in conjunction with the ABC 2026 Activity and Location Recognition Challenge [18, 26]. The data collection involves

25 BLE beacons strategically installed throughout the 5th floor of a care facility. Each beacon continuously transmits signals that are detected by a mobile phone carried by a caregiver (User ID 90) who moves around the floor during normal work activities. The mobile phone records RSSI values from all detectable beacons with second-level temporal precision. Concurrently, an observer (User ID 97) manually tracks and annotates the caregiver’s location during specific time periods, providing ground truth labels for supervised learning.

The complete dataset spans four consecutive days (April 10-13, 2023) and covers 13-18 distinct rooms depending on the day, reflecting realistic deployment conditions where different days involve visits to different subsets of locations. After preprocessing and temporal alignment, the final labeled dataset contains approximately 1.1 million timestamped BLE sensor readings, each associated with a room label. The beacons return RSSI measurements in dBm, where values closer to zero represent stronger signals. Fig. 1 shows the spatial layout of the 5th floor with beacon placement, while Fig. 2 illustrates the overall class distribution across all four days.

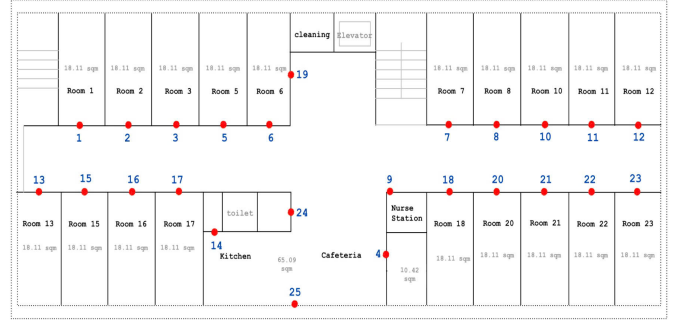


Fig. 1. 5th floor map with beacon positions and room layout [20, 26].

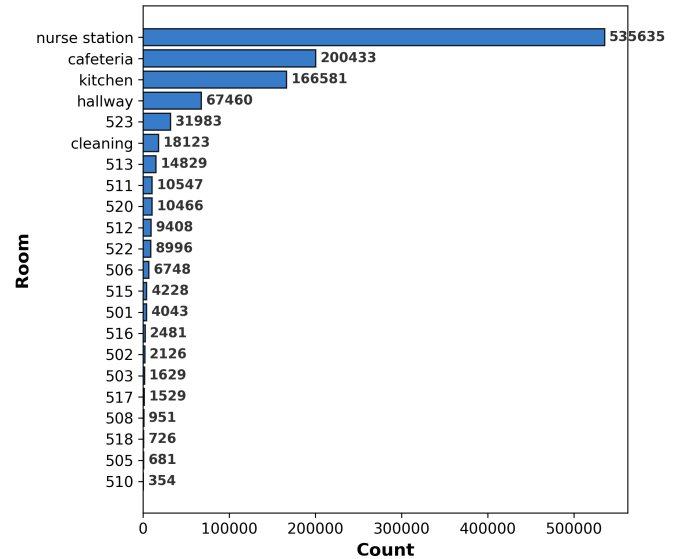


Fig. 2. Overall class distribution across all four days.

The dataset exhibits severe class imbalance, with common areas such as nurse station and cafeteria visited far more frequently than individual patient rooms. This imbalance poses significant challenges for classification models and justifies the use of macro F1 score as the evaluation metric, ensuring balanced performance across all locations regardless of visit frequency.

B. Preprocessing

The preprocessing pipeline transforms raw BLE sensor data and manual location annotations into a clean, labeled dataset suitable for machine learning. This step is essential to address data quality issues, ensure temporal alignment between sensor readings and labels, and create a supervised learning dataset with reliable ground truth.

The BLE sensor data from User ID 90 spanning the investigation period (April 10-13, 2023) was extracted from individual CSV files. Beacon signals were filtered to retain only transmissions from the 25 primary beacons installed on the 5th floor, excluding signals originating from beacons on other floors. MAC addresses were systematically mapped to beacon identifiers (1-25) for standardized reference. The location label data was processed to retain only valid annotations from the designated observer (User ID 97), ensuring complete temporal intervals and removing incomplete or deleted records.

Temporal alignment was performed to establish the supervised learning dataset. Each BLE sensor reading was matched with its corresponding room label based on temporal overlap between the sensor timestamp and the label's annotated time interval. All timestamps were standardized to second-level precision, consistent with established practices in indoor positioning systems where second-level temporal resolution is adequate [6, 19]. This temporal matching procedure successfully labeled approximately 1.1 million BLE records, representing 66% of the cleaned sensor data. Records without corresponding labels, comprising 34% of cleaned data, were excluded from the final dataset to maintain annotation quality, as these represent periods where active location tracking was not conducted or during unlabeled transitional intervals.

The final preprocessed dataset contains approximately 1.1 million labeled samples with the structure shown in Table I. Each record includes a timestamp (second precision), beacon ID (1-25), RSSI value, and the corresponding room label as the prediction target.

TABLE I
FINAL PREPROCESSED DATASET STRUCTURE

timestamp	mac_address	RSSI	room
2023-04-10 14:21:46+09:00	6	-93	kitchen
2023-04-10 14:21:46+09:00	4	-90	kitchen
2023-04-10 14:21:46+09:00	4	-97	kitchen
2023-04-10 14:21:46+09:00	6	-93	kitchen
2023-04-10 14:21:46+09:00	4	-90	kitchen
...

C. Model Training

1) *Traditional Machine Learning Baseline*: The traditional baseline approach treats each temporal window as an independent classification problem. Fig. 3 illustrates the complete workflow. Starting from individual BLE detection records, we construct a 25-dimensional beacon vector for each timestamp where position i contains the RSSI value if beacon i was detected, and 0 otherwise. Raw readings are grouped by timestamp into 1-second windows, and within each window we compute statistical features (mean, standard deviation, count) for each beacon, resulting in 75 features per window (25 beacons \times 3 statistics). These aggregated features are fed into an XGBoost classifier with balanced sample weighting to address class imbalance.

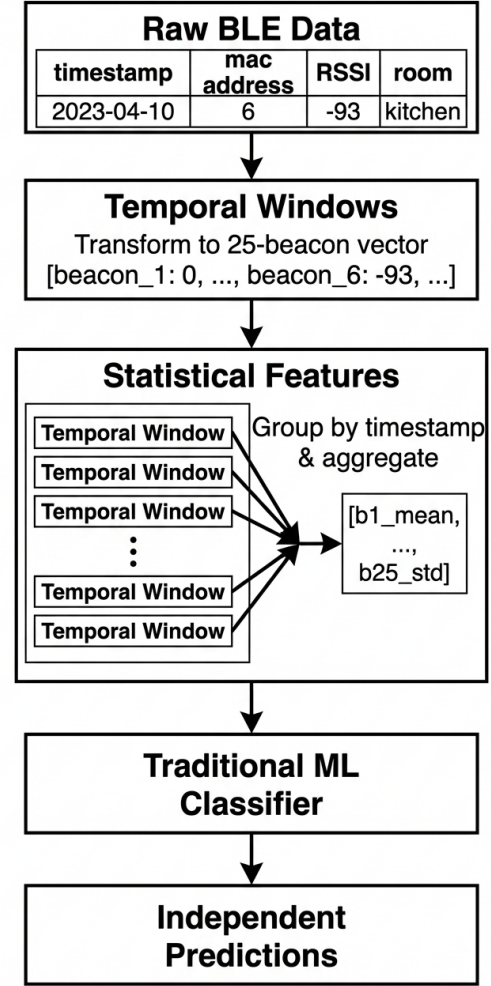


Fig. 3. Traditional machine learning workflow.

To establish a comprehensive baseline, we explored optimizations from different angles. Variation 1 employed dominant beacon features by augmenting the base statistical features with three additional categorical features representing the top three most frequently detected beacons within each window, aiming to capture additional spatial context through feature engineering. Variation 2 applied signal pattern-based

relabeling following Garcia and Inoue [20, 21], using KL divergence to identify majority class rooms whose signal patterns closely match minority class rooms, then relabeling matched samples to address class imbalance through data augmentation. Despite these optimization strategies targeting both feature engineering and data handling, all traditional methods remained within a narrow performance range.

2) *Deep Attention-based Sequential Ensemble Learning (DASEL)*: DASEL reconceptualizes indoor localization as a sequential learning problem through four integrated phases: (1) frequency-based feature engineering, (2) sequential model training using bidirectional recurrent networks with attention mechanisms, (3) multi-level ensemble inference combining multiple temporal perspectives and model initializations, and (4) temporal smoothing for spatial consistency. Fig. 4 provides an overview of the complete DASEL workflow.

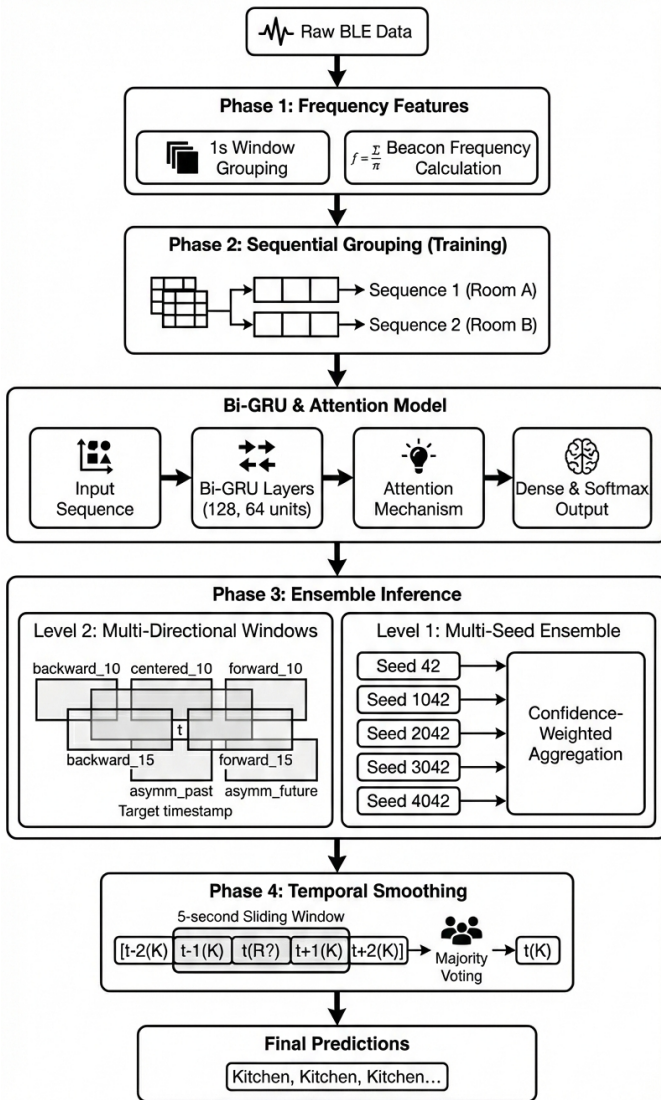


Fig. 4. DASEL complete workflow.

a) *Phase 1: Frequency-Based Feature Engineering*: Starting from preprocessed data (Section 2.2), we transform

raw RSSI measurements into frequency-based representations. For each BLE detection record, we construct a 25-dimensional beacon vector where position i contains the RSSI value if beacon i was detected, and 0 otherwise. BLE readings sharing the same timestamp are grouped into 1-second windows [6, 19]. Within each window, we calculate beacon appearance frequency as:

$$\text{frequency}_{i,t} = \frac{\text{count}_{i,t}}{\text{total_detections}_t} \quad (1)$$

where $\text{count}_{i,t}$ is the number of times beacon i was detected in window t , and $\text{total_detections}_t$ is the total number of all beacon detections in that window. For undetected beacons, frequency is 0. We use 23 beacons (beacons 1-23) as beacons 24-25 were never detected. Each window is represented by a 23-dimensional frequency vector with values normalized between 0 and 1.

b) *Phase 2: Sequential Model Training*: During training, we segment data into sequences using ground truth room labels. We identify consecutive timestamps where the room label remains constant:

$$\text{room_group_id} = \text{cumulative_sum}(\text{room_label} \neq \text{previous_room_label}) \quad (2)$$

Each contiguous block of the same room becomes one training sequence. Sequences shorter than 3 timesteps are discarded; sequences longer than 50 timesteps are truncated by taking the last 50 timesteps.

The model architecture (Fig. 5) consists of: masking layer (handles variable-length sequences, padding to 50 timesteps); first Bidirectional GRU layer (128 units); dropout (0.3); second Bidirectional GRU layer (64 units); dropout (0.3); attention mechanism with:

$$\text{attention_scores} = \tanh(W \cdot \text{sequence} + b) \quad (3)$$

$$\text{attention_weights} = \text{softmax}(\text{attention_scores}) \quad (4)$$

$$\text{context_vector} = \sum (\text{sequence} \times \text{attention_weights}) \quad (5)$$

dense layer (32 units, ReLU, 0.2 dropout); and output layer (softmax activation). The model is trained using sparse categorical cross-entropy loss with balanced class weights.

c) *Phase 3: Multi-Level Ensemble Inference*: Phase 3 implements a two-level ensemble strategy combining multi-seed model training with multi-directional sliding windows.

Level 1: Multi-Seed Model Ensemble. We train 5 independent models with different random initialization seeds [42, 1042, 2042, 3042, 4042]. For each of the 7 directional windows, we apply probability averaging across the 5 models (Algorithm 1), producing 7 probability distributions.

Level 2: Multi-Directional Window Combination. For each timestamp t requiring prediction, we create 7 temporal windows:

- backward_10: $[t - 9, \dots, t]$
- centered_10: $[t - 4, \dots, t + 5]$
- forward_10: $[t, \dots, t + 9]$
- backward_15: $[t - 14, \dots, t]$

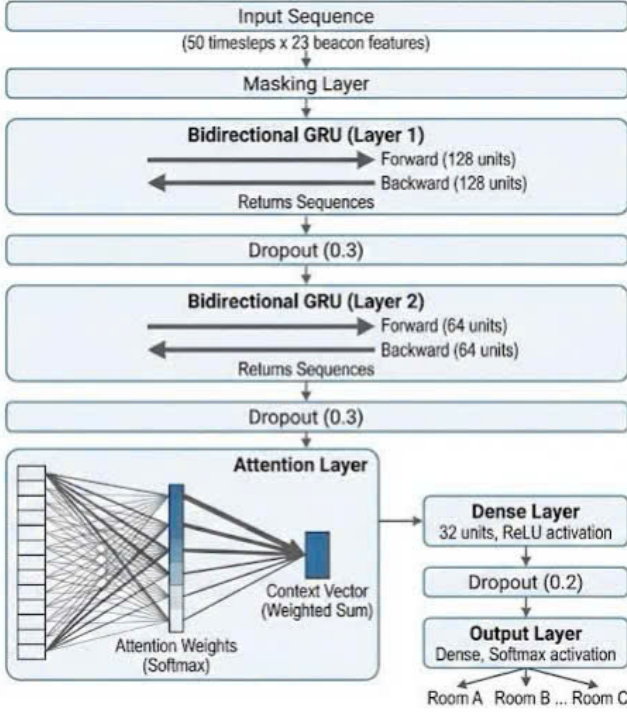


Fig. 5. DASEL model architecture.

Algorithm 1 Multi-Seed Model Ensemble

```

1: for each direction  $d \in \{\text{backward}_{10}, \text{centered}_{10}, \text{forward}_{10}, \text{backward}_{15}, \text{forward}_{15}, \text{asymm\_past}, \text{asymm\_future}\}$  do
2:   Extract sequences from direction  $d$ 
3:   for each model  $i$  ( $i = 1$  to  $5$ ) do
4:      $\text{probability}_i = \text{model}_i.\text{predict}(\text{sequences}_d)$ 
5:   end for
6:    $\text{averaged\_probability}_d = \text{mean}(\text{prob}_1, \dots, \text{prob}_5)$ 
7: end for

```

- forward_15: $[t, \dots, t + 14]$
- asymm_past: $[t - 11, \dots, t + 3]$
- asymm_future: $[t - 3, \dots, t + 11]$

Each directional window produces a probability distribution with confidence score (maximum probability). We aggregate using confidence-weighted voting (Algorithm 2).

d) *Phase 4: Temporal Smoothing*: For each prediction at timestamp t , we examine a 5-second temporal window $[t - 2, t - 1, t, t + 1, t + 2]$ and apply confidence-weighted voting (Algorithm 3, Fig. 6).

Algorithm 2 Confidence-Weighted Aggregation

```

1: for timestamp  $t$  with 7 directional probability distributions do
2:   for each direction  $d$  ( $d = 1$  to  $7$ ) do
3:      $\text{confidence}_d = \max(\text{probability\_distribution}_d)$ 
4:      $\text{weighted\_vote}_d = \text{probability\_distribution}_d \times \text{confidence}_d$ 
5:   end for
6:    $\text{final\_probability}_t = \sum(\text{weighted\_vote}_d) / \sum(\text{confidence}_d)$ 
7:    $\text{final\_prediction}_t = \arg \max(\text{final\_probability}_t)$ 
8: end for

```

Algorithm 3 Temporal Smoothing

```

1: for each timestamp  $j \in [t - 2, t - 1, t, t + 1, t + 2]$  do
2:    $\text{confidence}_j = \max(\text{probability\_distribution}_j)$ 
3:    $\text{weighted\_vote} += \text{probability\_distribution}_j \times \text{confidence}_j$ 
4: end for
5:  $\text{smoothed\_prediction}_t = \arg \max(\sum \text{weighted\_vote})$ 

```

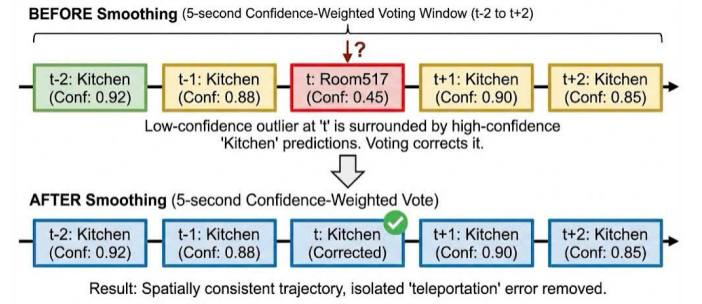


Fig. 6. 5-second temporal smoothing visualization.

D. Evaluation Protocol

We evaluate model performance using macro F1-score, the official metric specified by the ABC 2026 challenge organizers. Macro F1-score computes the F1-score independently for each room class and averages these scores, treating all rooms equally regardless of frequency. This metric is essential for imbalanced indoor localization where balanced performance across all locations is critical for comprehensive care facility monitoring.

We employ 4-fold cross-validation with temporal splitting rather than random partitioning. The dataset spanning four days (April 10-13, 2023) is split such that each fold uses one complete day as the test set and the remaining three days as training. Temporal splitting ensures test data is genuinely unseen and prevents data leakage from highly autocorrelated consecutive BLE readings, simulating realistic deployment where models trained on historical data must generalize to new time periods.

The four folds exhibit substantial variation in data size and class distribution, reflecting natural imbalance in real-world

data collection. Table II presents the detailed characteristics of each fold.

TABLE II
CROSS-VALIDATION FOLD CHARACTERISTICS

Fold	Test Day	Train Frames	Test Frames	Ratio	Train Classes	Test Classes
1	Day 4	962,294	30,619	31.4×	13	13
2	Day 3	951,141	143,401	6.6×	18	18
3	Day 2	747,816	333,507	2.2×	15	15
4	Day 1	465,004	590,447	0.79×	12	12

Fold 1 tests on the smallest dataset (Day 4 with 30,619 frames), providing the largest training set and a highly favorable train/test ratio of 31.4×. Conversely, Fold 4 presents the most challenging scenario: testing on the largest dataset (Day 1 with 590,447 frames) while training on the smallest subset (465,004 frames), resulting in a train/test ratio below 1 (0.79×). This tests the model’s ability to generalize when training data is scarce relative to deployment scale. The number of room classes also varies across folds, ranging from 12 to 18 unique locations, reflecting realistic deployment conditions where different days may involve visits to different subsets of rooms.

III. RESULTS AND ANALYSIS

Table III presents the comprehensive evaluation results comparing traditional machine learning approaches and the proposed DASEL framework across 4-fold cross-validation. All results are reported as macro F1 scores.

TABLE III
MACRO F1 SCORES ACROSS 4-FOLD CROSS-VALIDATION (F1-F4: FOLD 1-4; VAR.: VARIATION)

Approach	F1	F2	F3	F4	Mean±Std
Baseline	0.2819	0.2493	0.2665	0.3242	0.2805±0.0278
Var. 1	0.3009	0.2621	0.2634	0.3327	0.2898±0.0293
Var. 2	0.2830	0.2486	0.2633	0.3455	0.2851±0.0369
DASEL	0.5114	0.4207	0.4340	0.4082	0.4438±0.0295

The traditional machine learning approaches demonstrate limited and remarkably consistent performance. The baseline method achieved a mean macro F1 score of 0.2805 ± 0.0278 . Variation 1 with dominant beacon features achieved the highest traditional ML performance at 0.2898 ± 0.0293 , while Variation 2 with signal pattern-based relabeling yielded 0.2851 ± 0.0369 . Despite targeting different optimization strategies - feature engineering in Variation 1 versus data augmentation in Variation 2 - all traditional methods remained within a narrow 0.2805 to 0.2898 range. This consistent plateau across fundamentally different techniques suggests these methods have reached a fundamental ceiling imposed by the independence assumption paradigm.

The proposed DASEL framework achieves a mean macro F1 score of 0.4438 ± 0.0295 , with individual fold performances ranging from 0.4082 (Fold 4) to 0.5114 (Fold 1). The results demonstrate consistent performance improvements over traditional methods across all folds. The relatively stable standard deviation (0.0295) indicates robust performance across different temporal splits and varying data availability conditions, including the challenging Fold 4 scenario where training data is scarce relative to test data (train/test ratio 0.79×).

DASEL achieves a 53.1% relative improvement over the best traditional baseline (Variation 1: $0.2898 \rightarrow 0.4438$) and a 58.2% improvement over the basic traditional approach (Baseline: $0.2805 \rightarrow 0.4438$). The improvement is consistent across all folds: Fold 1 shows a 70.0% improvement ($0.3009 \rightarrow 0.5114$), Fold 2 shows 60.5% ($0.2621 \rightarrow 0.4207$), Fold 3 shows 64.8% ($0.2634 \rightarrow 0.4340$), and Fold 4 shows 22.7% ($0.3327 \rightarrow 0.4082$). Even in the most challenging scenario (Fold 4 with limited training data), DASEL maintains substantial performance gains, demonstrating that the sequential learning paradigm with multi-directional ensemble inference effectively captures temporal movement patterns that traditional independent-window approaches fundamentally cannot model.

IV. DISCUSSION

The dramatic performance difference between traditional methods and DASEL reveals that the fundamental limitation lies in treating each temporal window as an independent classification problem. Traditional machine learning methods extract statistical features from beacon signals and make predictions independently for each moment. Despite optimization strategies targeting feature engineering (Variation 1) and class imbalance handling (Variation 2), all methods plateaued within 0.2805 to 0.2898 macro F1 score, demonstrating that incremental improvements cannot overcome this independence assumption. Indoor localization is fundamentally a sequential problem - caregivers follow continuous trajectories where location at time t strongly predicts location at time $t + 1$, and room transitions produce gradual shifts in beacon patterns. Traditional methods discard this temporal continuity entirely. DASEL’s 53.1% improvement validates that sequential modeling is essential for breakthrough performance. By reconceptualizing indoor localization as a sequential learning problem and treating entire room visits as temporal sequences, DASEL captures movement trajectories, transition dynamics, and temporal patterns that independent-window classification fundamentally cannot model.

The bidirectional GRU architecture with attention mechanism directly addresses the fundamental characteristic of indoor localization data: temporal sequences where human movements follow continuous trajectories rather than independent observations. The core breakthrough comes from the bidirectional GRU layers [22, 24], which capture both past context (where the person came from) and future context (where they are going) - this sequential modeling alone accounts for the majority of performance improvement over traditional

methods. However, not all timesteps within a sequence are equally informative: during stable room occupancy, beacon patterns provide clear discriminative signals, while doorway passages and transitions produce multi-room ambiguities with overlapping signal characteristics. The attention mechanism [25, 27] addresses this data characteristic by learning to identify and emphasize key moments within sequences while downweighting noisy transitional periods, effectively simplifying complex sequential patterns into their most informative components. This architectural design - matching the model structure to the inherent temporal dependencies and variable signal quality in real-world movement data - directly translates to the observed 53.1% performance improvement, demonstrating that the sequential nature of indoor localization requires sequential modeling approaches rather than independent-window classification.

Frequency-based features provide critical practical advantages for real-world deployment. Raw RSSI measurements suffer from severe instability due to multipath signal propagation, human body absorption, interference from other wireless devices, and device-specific hardware characteristics. For example, Fig. 7 demonstrates this instability through box plot comparisons of RSSI distributions for three beacons appearing in both Kitchen and Cafeteria, revealing two critical limitations: high intra-room variance (within a single location, signal strength varies substantially), and significant inter-room overlap (mean RSSI values for the same beacon in different rooms differ by less than 2 dBm, smaller than the variance within each room).

In contrast, beacon appearance frequency captures which beacons are detected rather than how strongly, providing a more stable spatial signature. Fig. 8 illustrates the frequency distribution for three representative rooms, demonstrating that each room exhibits a distinct beacon appearance pattern. Each room shows characteristic detection patterns based on proximity to installed transmitters - nearby beacons appear frequently while distant beacons appear rarely or not at all. These presence patterns remain consistent despite environmental factors that destabilize individual RSSI values.

The frequency representation offers natural normalization (values between 0 and 1) without device-specific calibration, eliminating a major practical barrier to deployment across heterogeneous mobile devices.

The two-level hierarchical ensemble addresses distinct challenges that enhance prediction quality beyond the core sequential modeling framework. Multi-seed training with five different random initializations reduces variance by averaging predictions across models that learned slightly different patterns from unique initialization trajectories. This ensemble strategy provides robustness against unlucky initializations that might converge to poor local minima, ensuring evaluation metrics reflect the architecture’s true capability rather than initialization luck. The multi-directional sliding window strategy addresses a fundamental inference challenge: during training, the model learns from sequences with known room boundaries, but during deployment these boundaries are unknown. For

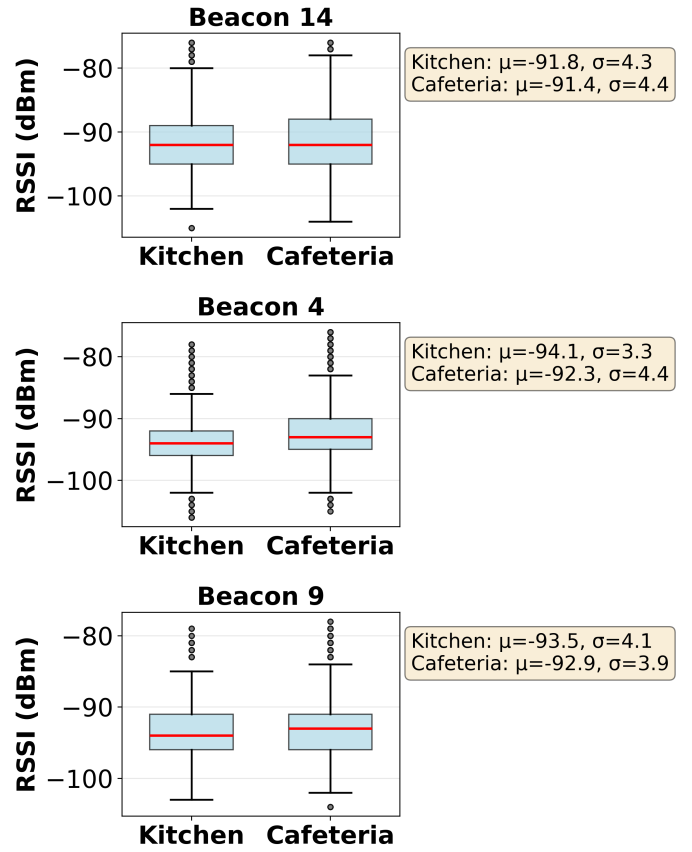


Fig. 7. RSSI distribution comparison between Kitchen and Cafeteria showing high variance and overlap.

any timestamp requiring prediction, we don’t know where it falls within its actual room visit - at the beginning, middle, or end. No single window configuration optimally captures all positions. Backward-looking windows fail when predicting near sequence start, forward-looking windows fail near sequence end, and centered windows struggle at boundaries. By employing seven complementary temporal perspectives, at least some windows will be well-aligned with clean signals regardless of true positional context.

The choice of window sizes (10 and 15 seconds) is empirically grounded in the training data’s sequence length distribution. Fig. 9 shows the distribution of training sequence lengths, revealing that the majority of room visits are concentrated between 10 and 200 seconds. The 10-second window captures the lower end of typical visit durations, ensuring even brief passages receive adequate temporal context without excessive contamination. The 15-second extended windows offer additional context for longer visits while remaining well below the average sequence length, minimizing risk of spanning multiple room transitions.

Confidence-weighted aggregation allows well-aligned windows with high-confidence predictions to dominate while poorly-aligned windows contaminated by transitions contribute less, creating a self-regulating ensemble.

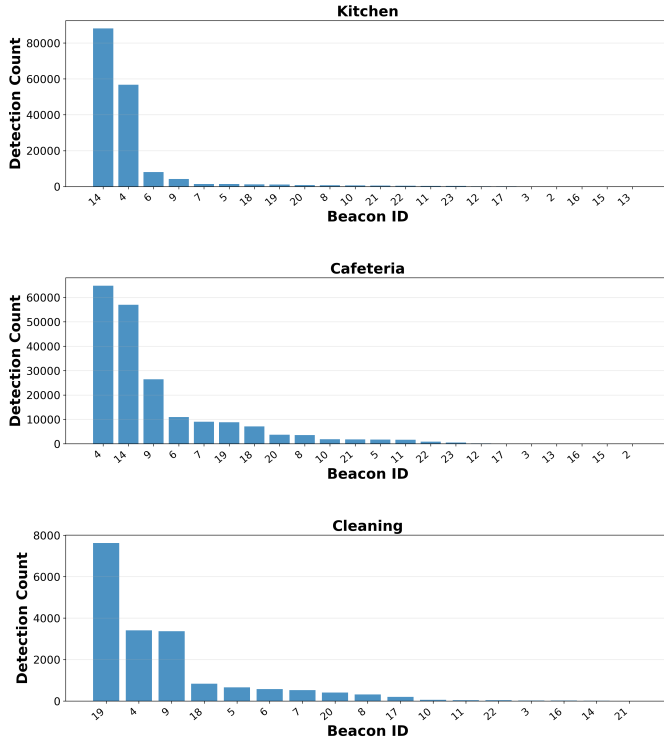


Fig. 8. Beacon frequency distribution for Kitchen, Cafeteria, and Cleaning showing distinct patterns.

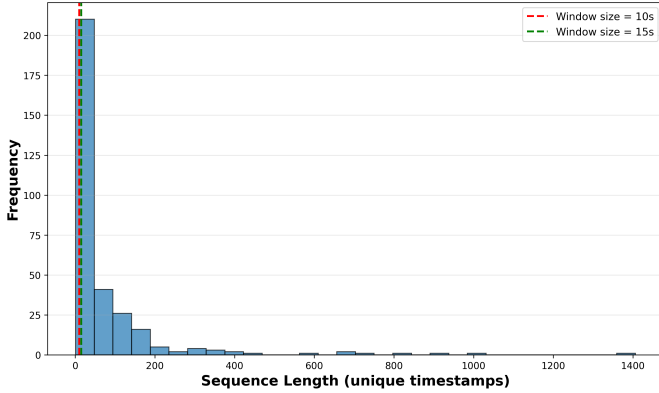


Fig. 9. Training sequence length distribution showing concentration between 10-200 seconds.

The 5-second temporal smoothing serves as final post-processing that enforces spatial and temporal consistency. Even after multi-directional ensemble aggregation, isolated prediction errors can occur where a single timestamp is incorrectly classified as a distant room despite surrounding predictions consistently indicating a nearby location. These isolated errors are often physically implausible - people don't teleport instantaneously across buildings. The smoothing window examines predictions within a local temporal neighborhood and applies confidence-weighted majority voting, effectively eliminating teleportation errors while preserving legitimate room transitions that show sustained directional change. The 5-

second window balances catching isolated errors (people typically remain in rooms longer than 5 seconds during meaningful visits) while preserving genuine transitions between adjacent rooms (which can occur within 5-10 seconds).

DASEL's balanced macro F1 score indicates reliable performance across all room classes including rarely-visited locations, which is essential for comprehensive care facility monitoring where accurate localization in individual patient rooms is as critical as tracking in common areas. The frequency-based representation combined with temporal smoothing produces spatially consistent predictions matching realistic human movement patterns. These characteristics - balanced performance, device robustness, and spatial consistency - make DASEL deployable in real-world settings without extensive system tuning. However, the two-level ensemble requires multiple forward passes per timestamp, which may challenge real-time deployment on resource-constrained mobile devices. Future work could explore model compression techniques such as knowledge distillation to maintain accuracy while reducing computational demands, adaptive window sizing based on detected movement patterns, and incorporating facility layout information to reduce physically implausible errors and improve transition modeling between adjacent rooms.

V. CONCLUSION

This research demonstrates that traditional machine learning approaches to BLE-based indoor localization face fundamental performance limitations rooted in their independence assumption rather than implementation details. Our proposed DASEL framework achieves breakthrough performance by reconceptualizing indoor localization as a sequential learning problem, explicitly modeling entire room visits as temporal sequences through bidirectional recurrent networks with attention mechanisms. The integration of frequency-based features for RSSI robustness, two-level hierarchical ensemble for variance reduction and positional uncertainty handling, and confidence-weighted temporal smoothing for spatial consistency creates a comprehensive system that captures movement trajectories, entry/exit patterns, and dwell-time dynamics that traditional methods discard. The resulting 53.1% improvement demonstrates that the sequential nature of indoor localization is not a secondary consideration but the fundamental property that must be modeled to achieve substantial performance gains. DASEL's balanced performance across all room classes and robustness to device heterogeneity make it practically deployable in real-world care facilities, establishing that breakthrough results in BLE-based indoor localization require treating human movement as the continuous temporal process it fundamentally is.

ACKNOWLEDGMENT

The authors would like to thank Kyushu Institute of Technology for providing the dataset in conjunction with ABC 2026.

REFERENCES

- [1] L. Bai, F. Ciravegna, R. Bond, and M. Mulvenna, "A Low Cost Indoor Positioning System Using Bluetooth Low Energy," *IEEE Access*, vol. 8, pp. 136858–136871, 2020.
- [2] M. M. Islam, A. Rahaman, and Md. R. Islam, "Development of Smart Healthcare Monitoring System in IoT Environment," *SN Comput. Sci.*, vol. 1, no. 185, 2020.
- [3] Y. Fisher and S. Efrati, "Real-time locating systems (RTLS) in healthcare: a condensed primer," *Int. J. Health Geogr.*, vol. 11, no. 25, 2012.
- [4] C.-W. Huang, Y.-T. Chan, C.-Y. Chen, M.-H. Chung, M.-Y. Lee, and Y.-T. Cheng, "Application of Indoor Positioning Systems in Nursing Homes: Enhancing Resident Safety and Staff Efficiency," *Sensors*, vol. 24, no. 18, p. 6099, 2024.
- [5] E. De Cock, B. Lemic, and J. Famaey, "Performance analysis of multiple Indoor Positioning Systems in a healthcare environment," *Int. J. Health Geogr.*, vol. 15, no. 7, 2016.
- [6] Y. Zhuang, J. Yang, Y. Li, L. Qi, and N. El-Sheimy, "Smartphone-Based Indoor Localization with Bluetooth Low Energy Beacons," *Sensors*, vol. 16, no. 5, p. 596, 2016.
- [7] S. Sadowski and P. Spachos, "RSSI-Based Indoor Localization With the Internet of Things," *IEEE Access*, vol. 6, pp. 30149–30161, 2018.
- [8] P. Davidson and R. Piché, "A Survey of Selected Indoor Positioning Methods for Smartphones," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1347–1370, 2017.
- [9] L. Li, X. Hu, K. Chen, and K. Tang, "WiFi Indoor Localization with CSI Fingerprinting-Based Random Forest," *Sensors*, vol. 18, no. 9, p. 2869, 2018.
- [10] A. Poullose and D. S. Han, "Indoor Localization: BLE, Machine Learning, and Kalman Filtering for Resource-Constrained Devices," *Preprints*, 2024, 2024021049. doi: 10.20944/preprints202402.1049.v1
- [11] F. Zafari, A. Gkelias, and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2568–2599, 2019.
- [12] I. Martín-Gorostiza, E. García-Garrido, M. Pizarro, D. Salido-Monzú, P. Mareca, and L. Lázaro-Galilea, "A Bluetooth Low Energy Indoor Positioning System with Channel Diversity, Weighted Trilateration and Kalman Filtering," *Sensors*, vol. 17, no. 12, p. 2927, 2017.
- [13] A. Basiri et al., "Indoor Location Based Services Challenges, Requirements and Usability of Current Solutions," *Comput. Sci. Rev.*, vol. 24, pp. 1–12, 2017.
- [14] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of Wireless Indoor Positioning Techniques and Systems," *IEEE Trans. Syst. Man Cybern. C*, vol. 37, no. 6, pp. 1067–1080, 2007.
- [15] D. Zhang, Y. Ma, Y. Zheng, X. Xie, and W. Y. Ma, "Robust Indoor Localization with Long-Term Stability," *IEEE Trans. Mobile Comput.*, vol. 18, no. 9, pp. 2156–2168, 2019.
- [16] X. Wang, Z. Wang, and B. O'Dea, "A TOA-Based Location Algorithm Reducing the Errors Due to Non-Line-of-Sight (NLOS) Propagation," *IEEE Trans. Veh. Technol.*, vol. 52, no. 1, pp. 112–116, 2003.
- [17] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-Based Fingerprinting for Indoor Localization: A Deep Learning Approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 763–776, 2017.
- [18] ABC Challenge 2026: Activity and Location Recognition Challenge in Care Facility, Kyushu Institute of Technology, 2026. [Online]. Available: <https://sites.google.com/view/activity-location-challenge>
- [19] J. Torres-Sospedra et al., "UJIIndoorLoc: A New Multi-building and Multi-floor Database for WLAN Fingerprint-based Indoor Localization Problems," in *Proc. Int. Conf. Indoor Positioning Indoor Navigation*, Busan, South Korea, 2014, pp. 261–270.
- [20] C. A. Garcia, Q. N. P. Vu, H. Kaneko, and S. Inoue, "A relabeling approach to signal patterns for beacon-based indoor localization in nursing care facility," *International Journal of Activity and Behavior Computing*, vol. 2024, no. 1, pp. 1–19, 2024. doi: 10.60401/ijabc.12
- [21] C. Garcia and S. Inoue, "Relabeling for indoor localization using stationary beacons in nursing care facilities," *Sensors*, vol. 24, no. 2, p. 319, 2024. doi: 10.3390/s24020319
- [22] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," in *Proc. Conf. Empirical Methods Natural Language Process.*, Doha, Qatar, 2014, pp. 1724–1734.
- [23] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," *arXiv preprint arXiv:1412.3555*, 2014.
- [24] S. Lee, B. Lim, and J. Lee, "Effective Indoor Robot Localization by Stacked Bidirectional LSTM Using Beacon-Based Range Measurements," *Sensors*, vol. 19, no. 9, p. 1997, 2019.
- [25] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," in *Proc. Int. Conf. Learning Representations*, San Diego, CA, USA, 2015.
- [26] C. Garcia, U. Dobhal, L. Zhao, X. Min, N. Nahid, and S. Inoue, "Decoding the Invisible: A Summary of the Location Recognition Challenge in Care Facilities," *International Journal of Activity and Behavior Computing*, vol. 2026, no. 2, 2026.
- [27] Z. Niu, G. Zhong, and H. Yu, "A Review on the Attention Mechanism of Deep Learning," *Neurocomputing*, vol. 452, pp. 48–62, 2021.