



Mahidol University  
International College



# A Simplified Multi-Floor Classification-Based Indoor Positioning System Study

Optimizing Grid Size and Feature Selection for ML-Based IPS

---

Burin Intachuen<sup>1</sup>      Mhadhanagul Charoenphon<sup>1</sup>      Tanakorn Mankhetwit<sup>1</sup>  
Charnon Pattiyanon<sup>2</sup>

November 7, 2025

<sup>1</sup>Department of Computer Engineering, Mahidol University (International College), Thailand

<sup>2</sup>Department of AI and Computer Engineering, CMKL University, Thailand

# Table of contents

1. Problem & Contributions
2. Methodology
3. Results
4. Key Findings & Implications
5. Future Directions

# Problem, Approach & Contributions

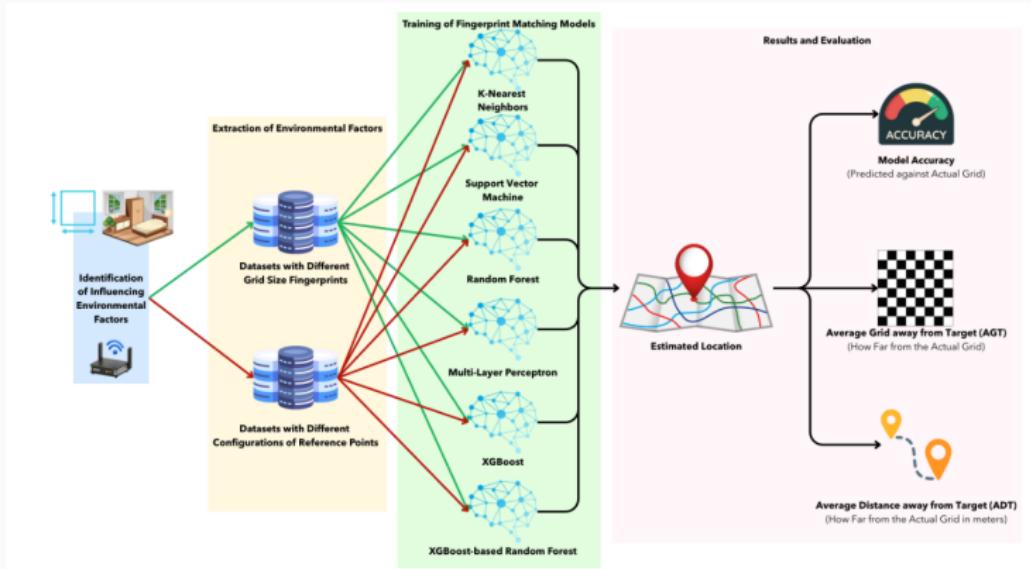
**Challenge:** GPS fails indoors, need ML-based IPS using Wi-Fi fingerprinting

**Research Gap:** Limited understanding of grid size and feature selection impact

## Our Contributions:

1. New evaluation metrics: AGT and ADT
2. Grid size vs. precision trade-off analysis ( $1 \times 1\text{m}$  to  $15 \times 15\text{m}$ )
3. Feature filtering impact ( $1,799 \rightarrow 378$  BSSIDs)
4. Verified local optimal grid size can be identified ( $7 \times 7\text{m}$ )

# Research Overview



**Figure 1:** Four-part research methodology

# Environmental Factors

**Key factors identified:**

## 1. Grid Size

- Larger grids → easier classification, lower precision
- Smaller grids → higher precision, more data collection effort

## 2. Low-Relevance BSSIDs

- Signals from distant access points
- Introduce noise and increase computational cost
- Need systematic filtering approach

# New Evaluation Metrics

**Average Grid from Target (AGT):**

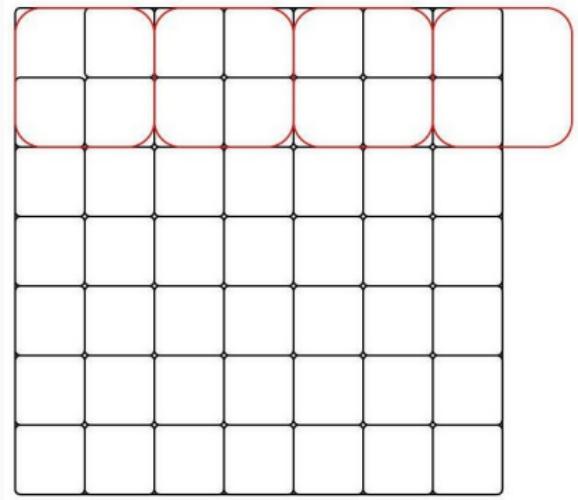
$$AGT = \sqrt{(x_{target} - x_{estimate})^2 + (y_{target} - y_{estimate})^2}$$

**Average Distance from Target (ADT):**

$$ADT = \sqrt{[w_g \times (x_{target} - x_{estimate})]^2 + [h_g \times (y_{target} - y_{estimate})]^2}$$

- AGT: Grid-based distance (adaptable to any grid size)
- ADT: Converts AGT to physical distance in meters using grid dimensions

# Grid Interpolation Approach



**Figure 2:** Grid aggregation visualization

## Benefits:

- Avoid re-taping entire area
- Generate multiple grid sizes from  $1 \times 1\text{m}$  base
- Reduce human measurement errors

## Method:

- Average RSSI from 5 sampling points
- Test 8 grid sizes:  $1 \times 1\text{m}$  to  $15 \times 15\text{m}$

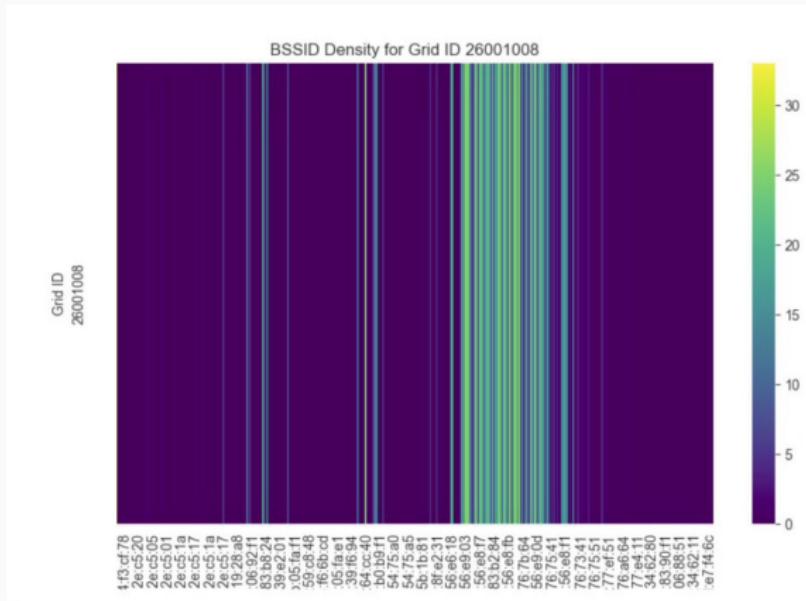
# Dataset Overview

**Table 1:** Dataset metadata by grid size

Floor	Grid Size	# Grids	# Data Points
1st	1×1m	309	8,086
1st	7×7m	17	445
1st	15×15m	5	130
2nd	1×1m	174	4,554
2nd	7×7m	12	314
2nd	15×15m	4	105

- Total: 12,640 RSSI samples
- Filtered: 378 BSSIDs (from 1,799)

# RSSI Heatmap Analysis



**Figure 3:** Heatmap showing RSSI values from different access points in a sample grid

- Darker colors = lower signal strength

# Machine Learning Models

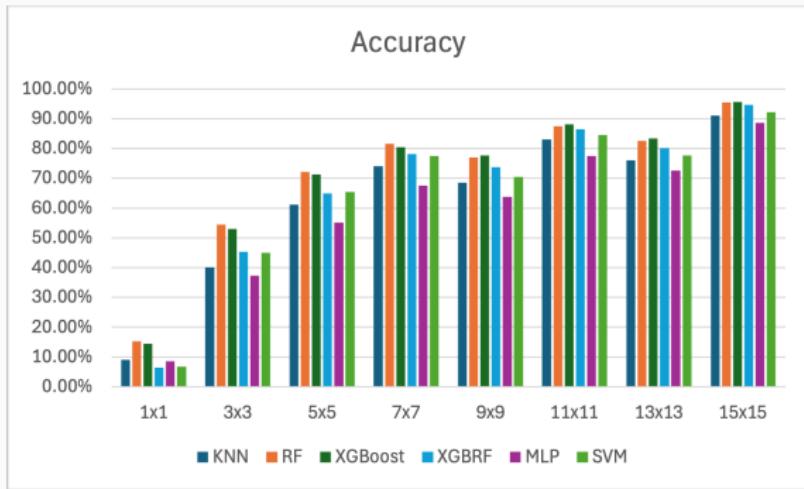
## Models evaluated:

- k-Nearest Neighbor (kNN)
- Random Forest (RF)
- Support Vector Machine (SVM)
- Multi-Layer Perceptron (MLP)
- XGBoost
- XGBoost + Random Forest hybrid

## Training approach:

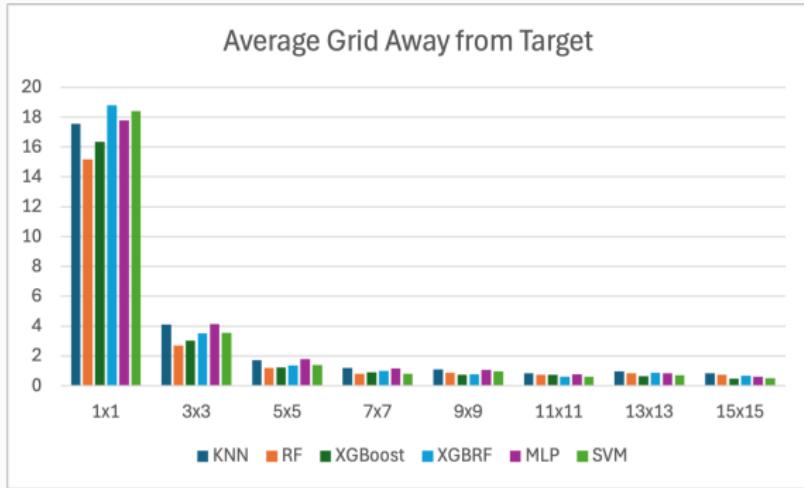
- Standard process with hyperparameter tuning
- Consistent dataset across all models
- Comparison of filtered vs. unfiltered BSSIDs

# Accuracy by Grid Size



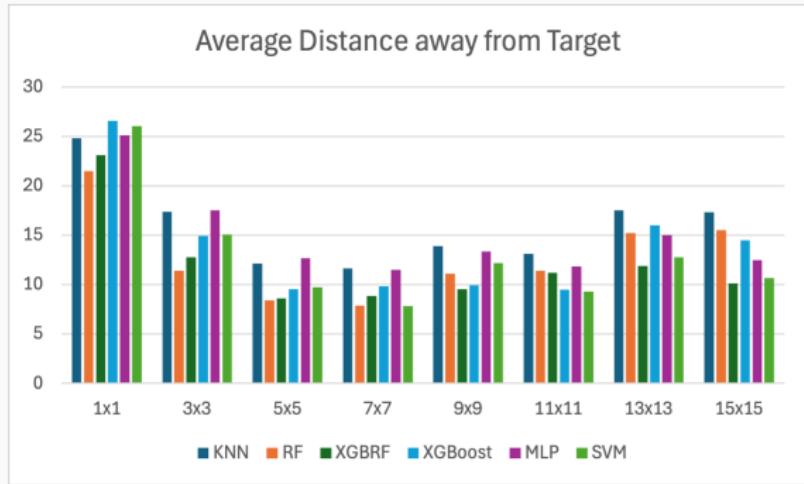
**Figure 4:** Model accuracy across different grid sizes (Filtered)

# AGT Results



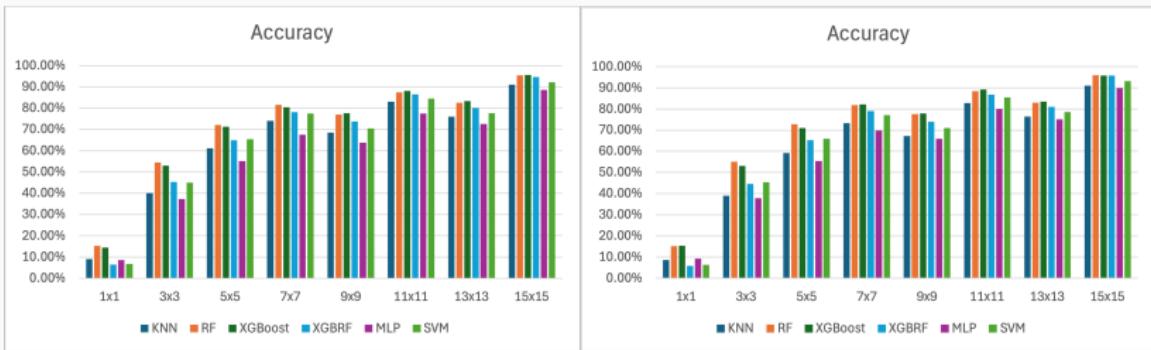
**Figure 5:** Average Grid from Target (Filtered)

# ADT Results



**Figure 6:** Average Distance from Target in meters (Filtered)

# Feature Filtering Impact



**Figure 7:** Filtered (378 BSSIDs)

**Figure 8:** Unfiltered (1,799 BSSIDs)

# Key Findings

## 1. Optimal Grid Size: $7 \times 7\text{m}$

- Best balance between accuracy and spatial resolution
- 80% accuracy with lowest AGT/ADT values

## 2. Feature Filtering Works

- 79% reduction in features ( $1,799 \rightarrow 378$ )
- Maintained performance, improved stability
- Significantly reduced training time

## 3. Best Models: RF and XGBoost

- Consistently outperformed other models
- Well-suited for fingerprint-matching tasks
- Better than complex deep learning in this setting

# Practical Implications

## For IPS Implementation:

- Don't assume bigger grids imply better accuracy - accuracy diminishes beyond optimal point
- Moderate grid sizes can be optimal
- Simple feature filtering is effective

## Limitations:

- Site-specific results (university campus)
- Intensive data collection required (can be automated but still daunting)
- May need recalibration for different environments
- Architecture and materials affect performance

## Future Directions

- Advanced deep learning models (attention-based, GNN-MLP)
- Real-time adaptive grid configurations
- Synthetic data augmentation and sensor fusion
- Transfer learning across buildings

**Questions?**

# Implementation Challenges

## Floor tiling variations:

- Different elevations and angles
- Addressed by reference-point-based grid construction

## Edge grid effects:

- Corner vs. center RSSI variations
- Minimal impact on overall performance
- Misclassifications typically involve adjacent grids

## Hardware constraints:

- RTX 3080 Ti struggled with unfiltered dataset
- Filtering made training feasible across all grid sizes