Evaluating Wi-Fi Fingerprinting for Enhanced Indoor Positioning in Campus Environments*

*A Case Study at University West, Sweden

Rasmus Andersson
Department of Engineering Science
University West
Trollhättan 461 32, Sweden
rasmus.andersson.5@student.hv.se

William Tagesson
Department of Engineering Science
University West
Trollhättan 461 32, Sweden
william.tagesson@student.hv.se

Rashid Ali
Department of Engineering Science
University West
Trollhättan 461 32, Sweden
rashid.ali@hv.se

***Abstract*—Wi-Fi fingerprinting indoor positioning systems (FP-IPS) based on received signal strength indicator (RSSI) are crucial for delivering location-based services in indoor environments where GPS is impractical. This study evaluates the performance of four KNN algorithms (Traditional, Regions-based, Weighted Average, and Median Filtering) for RSSI-based Wi-Fi FP-IPS at University West's indoor campus. Key metrics assessed include accuracy, precision, and computational cost. The Regions-based algorithm achieved the lowest average error of 5.2 meters and the shortest prediction time of 0.01 seconds for $k = 5$, highlighting its computational efficiency and superior accuracy. In contrast, the Traditional algorithm showed higher errors (average 18.4 meters) and similar computational efficiency (0.01 seconds), while the Weighted Average and Median Filtering algorithms balanced accuracy and computational cost. These findings provide valuable insights for engineers implementing Wi-Fi FP-IPS, demonstrating the region-based algorithm's effectiveness in real-world applications.**

***Index Terms*—wi-fi fingerprinting, indoor positioning system, indoor localization, wi-fi**

I. Introduction

GPS is widely used for positioning and navigation, offering accuracy within approximately 5 meters [1]. However, its performance is impaired in environments obstructed by non-line of sight (NLOS) barriers like buildings, limiting its effectiveness indoors. To overcome this, indoor positioning systems (IPS) have been developed, utilizing technologies such as magnetic, infrared, ultrasonic, ultrawide band, Bluetooth, and Wi-Fi, with Wi-Fi emerging as a popular choice due to its existing infrastructure [2]. Received Signal Strength Indicator (RSSI) fingerprinting is widely used in Wi-Fi-based IPS. In this approach, users capture signal data emitted by access points (APs), which an algorithm then compares against an RSSI-based fingerprint database to estimate the user's position [3].

Various algorithms can be used to develop an RSSI-based Wi-Fi IPS, with a notable example being the K-nearest neighbor (KNN) algorithm. KNN operates by selecting the K-nearest reference points (RPs) and calculating the Euclidean distance to determine the closest RP [4]. KNN-based algorithms are popular due to their simplicity and scalability, making them easy to implement with existing RP datasets through numerical analysis. This study focuses on evaluating four distinct variations of KNN algorithms to assess IPS performance metrics, including accuracy, precision, and computational cost in a campus environment. The primary research questions are:

- How do different KNN algorithm variations perform on campus?
- Which Wi-Fi fingerprinting algorithm shows optimal performance in this setting?

II. Related Research

Wi-Fi Fingerprinting uses captured RSSI values to locate users by comparing data within a fingerprint database. Basri *et al.* [5] outline IPS construction, evaluating Wi-Fi and Bluetooth technologies and emphasizing Wi-Fi fingerprinting's essential stages. Hu *et al.* [6] introduce Self-Adjusted Weight KNN (SAWKNN), demonstrating its superior performance. Hoang *et al.* [7] propose Soft Range Limited KNN (SRL-KNN), improving accuracy with user movement constraints. Lee *et al.* [8] utilize a random forest algorithm, employing smart watch data for indoor localization. Turabieh and Sheta [9] introduce Layered-Recurrent Neural Network (L-RNN), achieving superior accuracy with Non-Linear Regression (NLR). Quezada-Gaibor *et al.* [10] propose a data cleansing algorithm for Wi-Fi fingerprinting datasets, reducing positioning errors efficiently. These studies contribute to advancing Wi-Fi fingerprinting and indoor localization, enhancing accuracy and efficiency in real-world environments.

III. System Design

A. Campus Environment

The study collected fingerprinting data within the City Campus of University West (Högskolan Väst, HV), Sweden, focusing on Level 1 of Block I and Block J. This area

This work was part of an undergraduate final thesis project within the Wireless @ West Networking Research Group at the Department of Computer Science, University West, Sweden. We thank OpenAI for ChatGPT access, improving our writing clarity and accuracy.

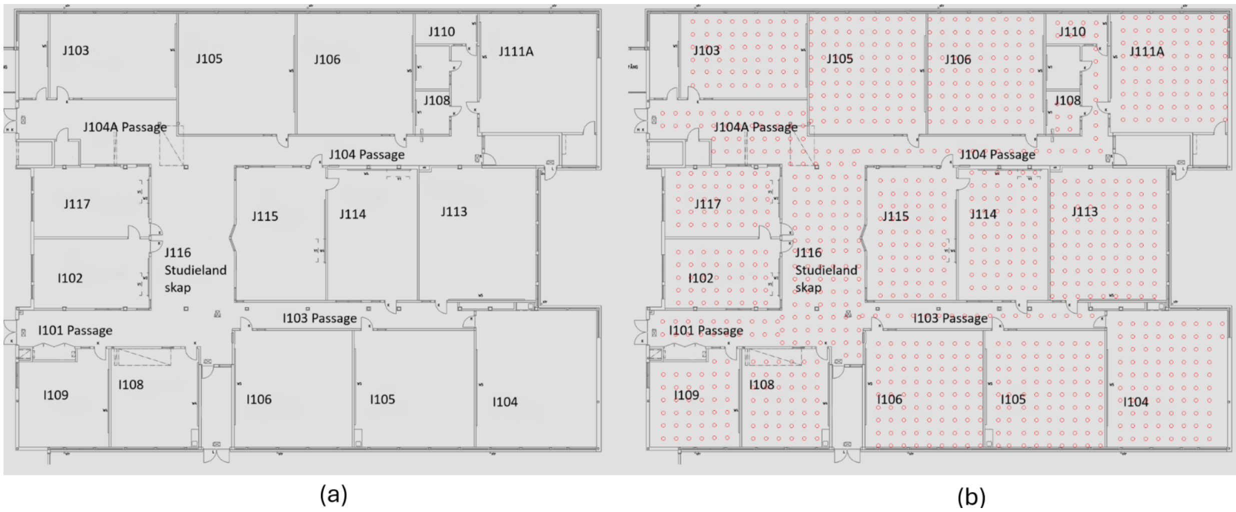

Fig. 1. Campus environment and fingerprinting reference points. (a) Map depicting the campus environment encompassing Level 1 of Block I and Block J at the City Campus of University West, Sweden. (b) The same map highlighting fingerprinting Reference Points (RPs) collected within this environment, denoted by red points.

represents a typical indoor setting on the campus, enabling systematic Wi-Fi signal data gathering to construct the fingerprint database. The campus map (Fig. 1(a)) highlights the specific locations where fingerprinting RPs were recorded (shown as red points in Fig. 1(b)), providing a foundational dataset for subsequent IPS analysis and experimentation.

B. Fingerprint Dataset

The initial phase involves gathering fingerprint data to establish a comprehensive database. Here, a fingerprint refers to the distinctive set of characteristics or attributes representing RSSI values or sensors at precise locations within a building. Multiple RPs constitute the fingerprint database, utilized by algorithms to determine a device's location. Essential data captured include RSSI and the unique identifier (BSSID). RSSI-based Wi-Fi fingerprinting typically comprises offline and online phases. **Offline Phase:** During the offline phase, fingerprints were systematically collected to construct the database. This involves collecting RPs in classrooms and corridors, spaced one meter apart, to build a fingerprinting database. A Galaxy A52 5G smartphone with the GetSensorData 2.0 application facilitated data recording [11]. Data collection occurred for about 20 seconds at each RP, using a systematic sampling approach. RPs were gathered on the first floor of Blocks I and J of the HV campus, as illustrated in Fig. 1(a), with corresponding real-world positions recorded. The dataset and maps are available on our GitHub repository [12]. After preprocessing the raw data, fingerprints from each RP were used to build the database. RPs, representing real-world positions, were mapped onto pixels corresponding to the layout of the first floor of Blocks I and J of the HV campus. The map dimensions are 1920x1356 pixels, providing approximately 35.7 pixels per meter for precise spatial alignment (Fig. 1(a)). With this conversion, RPs can accurately be depicted on the map, as shown in Fig. 1(b).

Online Phase: In the online fingerprinting phase, test data points are collected for testing purposes, compared against the fingerprint database. Four distinct KNN algorithms (refer to Section III-C) are used for positional estimation of these test points (TPs). We collected TPs using the same application used for RP collection. Unlike the systematic sampling for RPs, TP data is collected at random locations, ensuring that TPs do not overlap with RPs.

C. Indoor Positioning Algorithms (K-Nearest Neighbor)

The KNN algorithm computes device proximity to predefined points by selecting the k-nearest neighbors and calculating distances. In Wi-Fi fingerprint IPS, it identifies the nearest RPs to TPs, providing localization. Four KNN variants were evaluated: traditional, weighted average, median filtering, and region-based, the latter offering an enhanced solution for campus environments.

1) Traditional KNN Algorithm: Modern KNN IPS algorithms employ diverse learning methods to predict the optimal RP, requiring a sizable and homogeneous dataset. Traditional KNN algorithms rely on fixed calculations, resulting in consistent outcomes without learning. They take online data for a specific TP and the entire fingerprint database as inputs, iterating through RP positions and BSSIDs. By comparing BSSIDs and calculating RSSI differences, they determine multiple candidate positions based on RSSI errors, selecting the candidate with the lowest overall error.

2) Weighted Average KNN Algorithm: The weighted average variant prioritizes earlier BSSID findings based on K, utilizing descending RSSI-ordered datasets. It employs a weighted approach inversely proportional to error, preventing division by zero and assigning higher weights to smaller errors. The algorithm calculates the sum of weighted error RSSI values and divides by the sum of reciprocals of the weights, utilizing a "data overlapping" technique.

3) Median Filtering KNN Algorithm: This method employs a sliding window algorithm for filtering and optimizing outcomes. Consecutive BSSIDs overlap, enhancing position estimation and error averaging. The window comprises five loops, with every fifth loop incorporating overlapping data from the previous loop, while others add current index data. Utilizing the data overlapping generates substantial data. In median filtering, RSSI values are averaged using a window of RSSI values, from which the median is calculated (filtered values). Each error value within the window is compared to its corresponding filtered value to ensure proximity to measured actual values. The algorithm retains actual values and filters incorrect ones.

4) Regions-based KNN Algorithm: Partitioning the experimental area into distinct regions enhances accuracy by limiting the number of accessible RPs to a TP. Previously, TPs could select from all RPs, but with this modification, computational time is expected to decrease and large-scale errors should diminish. Regions are delineated based on the BSSID broadcasted from surrounding APs (in our case, six) of all RPs. Each RP is categorized into a region based on the BSSID with the highest RSSI value, assuming the corresponding AP belongs to that region. Subsequently, all RPs are categorized into one of the six regions, with data separated accordingly. Before predicting the RP to which a TP belongs, the TP is assigned a region using the same method as the RPs. The algorithm then executes the default K-NN algorithm, as mentioned in Section III-C1, with the TP selecting from RPs within the same region.

IV. Performance Evaluation

For performance evaluation, accuracy, precision, and computational cost were selected as metrics. Accuracy measures the proximity of positioning results to the actual location of a TP, averaged across error measurements for each TP. Precision evaluates the consistency of distinct variations of KNN algorithms across all tests, regardless of individual test accuracy. Cumulative Distribution Function (CDF) and quartile values visualize error probability and distribution. Computational cost measures the time needed to compute results for one TP, represented as the average time from start to finish for all TPs in milliseconds, varying based on the four KNN algorithms and k value.

In Fig. 2, the box plot illustrates the error distribution in meters for the four KNN algorithms at two different values of k (5 and 15). For $k = 5$, the Traditional algorithm shows the highest error variability and median error. The Regions-based algorithm demonstrates lower errors and variability compared to the Traditional method. The Weighted Average algorithm exhibits slightly higher errors than the Regions-based method but lower than the Traditional approach. The Median Filtering algorithm shows moderate accuracy but with a wider spread. For $k = 15$, the Traditional algorithm's errors decrease compared to $k = 5$ but still shows considerable variability. The Regions-based algorithm maintains the lowest error distance among the four methods. The Weighted Average algorithm has similar performance to the Regions-based method but with slightly higher errors. The Median Filtering algorithm displays increased errors and variability compared to the other methods. Outliers are present in all algorithms, indicating occasional significant deviations from the typical error range. The Regions-based algorithm consistently performs best in terms of lower average errors and variance.

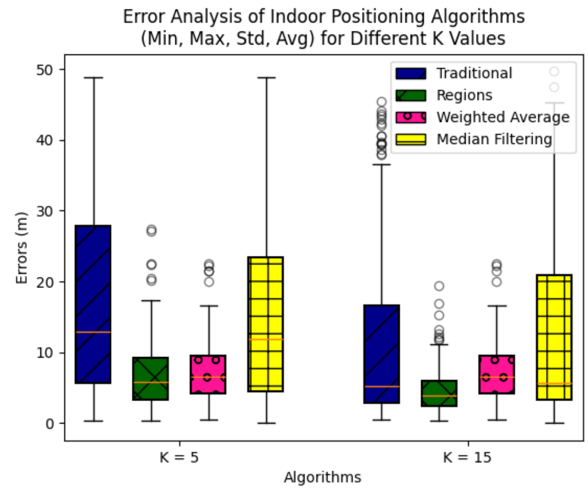

Fig. 2. Box plot showing the accuracy errors (m) of four KNN algorithms (Traditional, Regions-based, Weighted Average, and Median Filtering) for $k = 5$ and $k = 15$. The plot includes average and variance values with outliers.

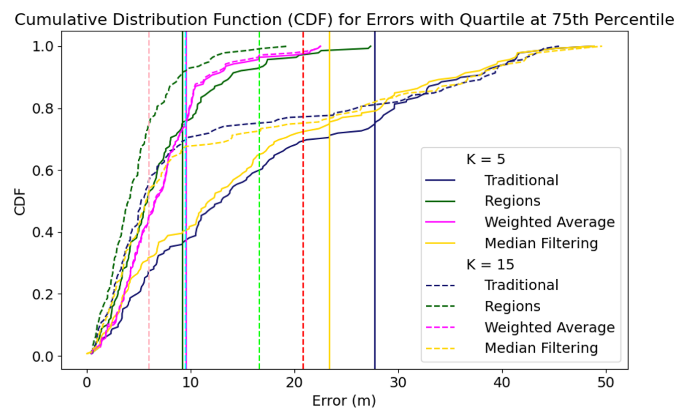

Fig. 3. Cumulative Distribution Function including 75^{th} percentile (Q3) of accuracy errors for KNN algorithms with $k = 5$ and $k = 15$.

The CDF (Fig. 3) plot shows the cumulative probability distribution of accuracy errors to demonstrate the precision of four KNN algorithms at $k = 5$ and $k = 15$. For $k = 5$, the Traditional algorithm (solid blue line) exhibits the slowest increase in the CDF, indicating higher errors. The Regions-based algorithm (solid green line) shows the steepest curve, indicating lower errors and better performance. The Weighted Average algorithm (solid pink line) also performs well, closely following the Regions-based algorithm. The Median Filtering algorithm (solid yellow line) has a moderate performance, with errors higher than the Regions-based and Weighted Average but lower than the Traditional algorithm. For $k = 15$, the Traditional algorithm (dashed blue line) shows improved performance compared to $k = 5$ but still has higher errors than the other methods. The Regions-based algorithm (dashed green line) continues to show the best performance with the lowest errors. The Weighted Average algorithm (dashed pink line) maintains good performance, similar to the Regions-based method. The Median Filtering algorithm (dashed yellow line) shows slightly decreased errors compared to $k = 5$. The vertical dashed lines represent the 75^{th} percentile (Q3) for each algorithm, indicating the error value below which 75% of the data points fall. The Regions-based algorithm consistently shows the lowest Q3 values, highlighting its superior accuracy in reducing errors compared to the other methods. The bar chart in Fig. 4 shows the average prediction time (in seconds) for the algorithms. The Traditional algorithm has very low prediction times for both $k = 5$ and $k = 15$, indicating high computational efficiency. Howeve, the Regions-based algorithm demonstrates the lowest prediction times, lower than the Traditional algorithm. The Weighted Average algorithm has moderate prediction times, higher than the Traditional and Regions-based algorithms but still relatively efficient. The Median Filtering algorithm has the highest prediction times for both $k = 5$ and $k = 15$, with a significant increase as k rises, indicating a higher computational cost. Overall, the Regions-based approach is as efficient as the Traditional method and outperforms the Weighted Average and Median Filtering methods in terms of prediction time, although the Median Filtering algorithm incurs a much higher computational cost. These results are crucial for drawing conclusions on algorithm performance for RSSI-based Wi-Fi fingerprinting IPSs.

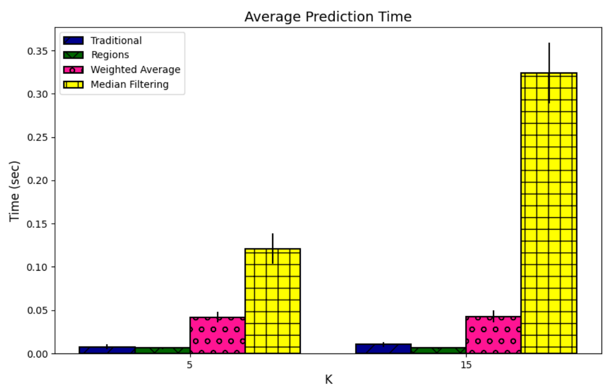

Fig. 4. Average prediction time of each algorithm with different k (5 and 15) values, highlighting the computational efficiency of the region-based approach.

Computational efficiency and prediction accuracy are vital considerations for real-world applications. The findings indicate that the region-based algorithm offers a compelling balance of low prediction time and high accuracy, making it a promising choice for practical implementation in indoor positioning systems in campus environments.

V. Conclusion

This study emphasizes the importance of tuning and evaluating Wi-Fi fingerprinting IPS algorithms, especially in campus environments where accurate localization is critical. Through quantitative experiments, we assessed various KNN algorithms within University West's indoor campus. Our findings show that the region-based algorithm consistently outperforms traditional, weighted average, and median filtering algorithms in minimizing accuracy errors and prediction time. These results improve indoor positioning accuracy and offer valuable insights for engineers implementing Wi-Fi FP-IPS, highlighting the region-based algorithm's suitability for real-world applications.

The limitations of this research include the susceptibility of Wi-Fi signals to reflection and obstruction (effecting RSSI values), the potential inadequacy of median filtering, the limited number of regions studied (only six), and data collection using a single phone, which may limit generalizability. Additionally, as our future work, we aim to use multiple end devices with different types (smartphone, laptop, tablet etc) to construct a more comprehensive dataset, considering the individual device differences.

References

[1] GPS.gov, "Gps accuracy," accessed: Mar. 2, 2024. [Online]. Available: https://www.gps.gov/systems/gps/performance/accuracy/#how-accurate

[2] L. Qi, Y. Liu, Y. Yu, L. Chen, and R. Chen, "Current status and future trends of meter-level indoor positioning technology: A review," *Remote Sensing*, vol. 16, no. 2, 2024. [Online]. Available: https://www.mdpi.com/2072-4292/16/2/398

[3] C. S. Álvarez Merino, E. J. Khatib, H. Q. Luo-Chen, A. T. Muñoz, and R. B. Moreno, "Evaluation and comparison of 5g, wifi, and fusion with incomplete maps for indoor localization," *IEEE Access*, vol. 12, pp. 51 893–51 903, 2024.

[4] T. Srivastava, "A Complete Guide to K-Nearest Neighbors (Updated 2024) — analyticsvidhya.com," https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/, [Accessed 25-04-2024].

[5] C. BASRI and A. El Khadimi, "Survey on indoor localization system and recent advances of wifi fingerprinting technique," in *2016 5th International Conference on Multimedia Computing and Systems (ICMCS)*, 2016, pp. 253–259.

[6] J. Hu, D. Liu, Z. Yan, and H. Liu, "Experimental analysis on weight K -nearest neighbor indoor fingerprint positioning," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 891–897, 2019.

[7] M. T. Hoang, Y. Zhu, B. Yuen, T. Reese, X. Dong, T. Lu, R. Westendorp, and M. Xie, "A soft range limited k-nearest neighbors algorithm for indoor localization enhancement," *IEEE Sensors Journal*, vol. 18, no. 24, pp. 10 208–10 216, 2018.

[8] S. Lee, J. Kim, and N. Moon, "Random forest and wifi fingerprint-based indoor location recognition system using smart watch," *Human-centric Computing and Information Sciences*, vol. 9, no. 1, Feb. 2019. [Online]. Available: http://dx.doi.org/10.1186/s13673-019-0168-7

[9] H. Turabieh and A. Sheta, "Cascaded layered recurrent neural network for indoor localization in wireless sensor networks," in *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)*, 2019, pp. 1–6.

[10] D. Quezada-Gaibor, L. Klus, J. Torres-Sospedra, E. S. Lohan, J. Nurmi, C. Granell, and J. Huerta, "Data cleansing for indoor positioning wi-fi fingerprinting datasets," in *2022 23rd IEEE International Conference on Mobile Data Management (MDM)*, 2022, pp. 349–354.

[11] A. R. Jiménez, F. Seco, and J. Torres-Sospedra, "Tools for smartphone multi-sensor data registration and gt mapping for positioning applications," in *2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2019, pp. 1–8.

[12] R. Andersson, W. Tagesson, and R. Ali, "Dataset: Evaluating Wi-Fi Fingerprinting for Enhanced Indoor Positioning in Campus Environments," Apr. 2024. [Online]. Available: https://github.com/wirelessATwest/PEWFIPS-HV