Capstone Project Phase B
23-2-D-2

Exploring the Minimum RSSI Fingerprints Needed for Reliable Indoor Localization

Supervisor: Dr. Julia Sheidin
Student: Roman Milman
Student: Tomer Meidan

Table Of Content

1. Abstract

While GPS has solved most of the outdoor location systems problems, it fails to repeat this success indoors. This is a challenging problem because of the environment inside buildings and the structure of walls, which can block or weaken signals used for location Tracking. This study explores the efficacy of WiFi fingerprinting for indoor positioning, leveraging signal strength patterns from existing WiFi networks to address the limitations of GPS in enclosed spaces. Focusing on a dataset of WiFi fingerprints collected, we examine the precision of location estimation through algorithmic comparison of these fingerprints. Unlike traditional systems requiring supplementary hardware, our research utilizes readily available WiFi infrastructure, proposing a cost-effective solution for indoor navigation. By analyzing the accuracy of our positioning algorithm against known locations, we aim to contribute to the understanding of WiFi fingerprinting's potential in enhancing indoor navigation accuracy. This investigation not only highlights the challenges and solutions in the realm of WiFi-based indoor positioning but also underscores the broader applicability and benefits of this technology in navigating complex indoor environments.

2. Introduction

In recent years, indoor navigation systems have been the focus of many research studies and commercial applications, as they have become increasingly important in facilitating people's navigation in complex indoor environments, such as airports, shopping malls, hospitals, and office buildings [Article 3].

The Need For Indoor Navigation

The lack of indoor navigation could be particularly challenging and can be a significant inconvenience. It can lead to wasted time and effort as individuals try to find their way around large buildings, and can make it difficult to navigate complex indoor spaces efficiently in large buildings, such as airports, shopping malls, or office buildings. For example, individuals who are in a hurry to catch a flight at an airport may waste valuable time trying to find their gate. Similarly, in large shopping centers, individuals may struggle to find the stores they are looking for, leading to wasted time and increased stress.

GPS Doesn’t Work Indoors

While traditional outdoor navigation systems-based technology exists, they are not suitable for indoor environments. Today's most popular navigation system is the GPS (Global Positioning System), however, GPS itself does not use context awareness. GPS is a satellite-based navigation system that provides users with accurate location and time information in outdoor environments.

Weak signals from GPS satellites cause the challenges of GPS in indoor environments, they become weaker as they pass through the atmosphere, making it difficult for the GPS receiver to distinguish between the weak GPS signal and other sources of radio frequency interference that may exist indoors. Additionally, indoor environments often have complex layouts, and the signals from GPS satellites may reflect off walls or other surfaces, leading to multipath interference that can further reduce the accuracy of GPS [Article 1] [Article 10] .

Additional Unique Hardware

Known systems for indoor navigation require additional unique hardware to function properly [Article 4]. Companies in this field rely on a mobile application that uses their developed sensors and special hardware that is proprietary to their company. The sensors must include a large number of RF wireless sensors and special receivers called Readers, which detect signals from the sensors. A Digital platform is also necessary, which includes software working with the data from mobile devices. The maintenance of these systems is a significant burden for companies.

To accurately map an entire building, a large number of sensors are required. To determine the minimum number of sensors needed for reliable navigation, most products require a sensor per 5 to 8 meters square [Article 2]. This density ensures a precision of 3 to 7 meters, which is acceptable for most indoor navigation applications [Article 4] [Article 10]. However, this requirement implements indoor navigation systems impractical for small businesses or organizations with limited budgets. Hence, there is a need for a cost-effective and scalable indoor navigation system that does not require additional hardware installation.

3. Background and Related Work

The quest for indoor positioning systems (IPS) has led to the exploration of various techniques, including WiFi fingerprinting, which leverages WiFi infrastructure to map indoor environments accurately. Among the prominent works in this domain, Torres-Sospedra et al. [2] conducted an analysis of distance and similarity measures critical to the effectiveness of WiFi fingerprinting in IPS.

3.1. Analysis of Distance and Similarity Measures

The development of accurate and efficient Indoor Navigation Systems (INS) has become increasingly important for various applications, ranging from emergency evacuation to location-based services in complex indoor environments. This section explores the advances in INS, particularly those using Wi-Fi fingerprinting and RSSI values from Access Points (APs) to create reliable localization methods.

Advances in K-Nearest Neighbors (k-NN) Algorithms

The research by Torres-Sospedra et al. [Article 2] serves as a base in the field of Wi-Fi fingerprinting, providing an analysis of over 50 distance and similarity measures for indoor positioning systems. Their work particularly emphasizes the k-NN algorithm, a distance-based classifier that determines the position by comparing a sample against a database of labeled fingerprints. Their study advocates for the examination of different data representations, especially considering the influence of distant detected WAPs on localization accuracy, highlighting the significant role of Euclidean distance in such applications due to its intuitive geometric interpretation in the signal space.

Refinement of Euclidean Distance Measures

Analyzing various representations of data in conjunction with the Euclidean distance, the study observed that transformations like exponential scaling could improve the accuracy of the k-NN algorithm. Specifically, RSS values representations coupled with Euclidean distance provided a more consistent reduction in error across different k values, suggesting an optimal approach for minimizing positioning errors [2].

Optimization of the k-NN Framework

Zhu's exploration of an Optimized k-NN (OKNN) algorithm demonstrated a significant improvement in indoor positioning accuracy [Article 4]. By selectively weighting the influence of closer reference points, the OKNN algorithm mitigates the impact of RSS fluctuations and environmental dynamics, improving localization results.

Utilization of Clustering Techniques

In the work by Liu, de Lacerda, and Fiorina [Article 7], the application of k-means clustering to the WKNN algorithm is proposed as a means to address the challenges of RSS variance and computational complexity in large fingerprint datasets. This method aimed to improve the localization process by pre-clustering fingerprints, thus maintaining accuracy while optimizing computational resources [Article 7].

Lessons Learned and Moving Forward

Our study builds upon these pivotal works by employing a combination of WKNN and Euclidean distance measures, with a focus on clustering RSSI values into Wi-Fi fingerprints. We learned from the methods and insights presented in these articles that, while the basic k-NN framework provides a reliable starting point, its performance can be substantially improved through thoughtful optimizations and methodological enhancements.

Adriano Moreira, Maria Nicolau, Filipe Meneses, and António Costa's work at the RTLS@UM during the EvAAL competition showcases a practical application of indoor localization using Wi-Fi fingerprinting [Article 5]. Their study is instrumental in our understanding of how the weighted k-Nearest Neighbors (WKNN) approach can be effectively applied in real-world, large-scale environments [Article 5]. In their methodology, the RTLS@UM team also leveraged k-means clustering to build a radio map [Article 5]. This strategy of grouping the training data helped reduce the computational complexity during the localization phase [Article 5]. The clustering not only improved computational efficiency but also potentially contributed to the handling of environmental noise and signal attenuation, which are common challenges in indoor Wi-Fi localization.

While the Euclidean distance provides a natural measure for spatial localization, we found that its effectiveness is greatly influenced by the representation of the RSSI data and the specific algorithmic adaptations employed [Article 2] [Article 4]. By applying transformations and integrating clustering techniques, we aim to refine the precision of INS and overcome the limitations highlighted in previous studies [Article 7].

Our research will build upon these previous works to enhance the Euclidean distance's effectiveness for indoor localization Articles [2] [4] [5] [7]. By refining the Euclidean measure and incorporating the proven strategies of WKNN and k-means clustering, we endeavor to develop an INS that not only improves upon the accuracy of existing systems but also provides new insights into efficient algorithmic design Articles [5] [7]

Furthermore, our investigation is designed to elucidate the relationship between the number of fingerprints and location accuracy. We plan to reduce the number of available fingerprints at specific positions to evaluate the impact on the algorithm's precision. This approach will enable us to understand the trade-offs involved in the fingerprinting process and to optimize the balance between dataset size and localization accuracy.

3.2. The Received Signal Strength

The Received Signal Strength Indicator $(RSSI)$ $(P(d))$ at distance (d) from a single WiFi access point is modeled using the Friis transmission equation:

$[P(d) = P_{tx} - 10 \cdot n \cdot \log_{10}(d) + X]$

Where:

- (P_{tx}) is the transmitted power.
- (n) is the path loss exponent.
- (d) is the distance between the transmitter and receiver.
- (X) is a Gaussian random variable representing noise.

These equations elucidate the mathematical foundations underlying the methodologies proposed in each referenced article.

3.3. The Euclidean Distance

In our research, we focus on the implementation and evaluation of an indoor navigation system using Wi-Fi fingerprinting within a shopping mall. This environment poses unique challenges for positioning systems due to its complex layout, which includes open spaces, and a variety of physical obstructions like store fronts, kiosks, and decorative features. These elements disrupt the straightforward navigation paths typically found in more grid-like environments, such as urban city blocks.

For our experiment, the shopping mall serves as an ideal setting to test the efficacy of different distance metrics in the context of RSS based indoor positioning. We determined which distance metric—Euclidean or Manhattan—provides the most accurate and reliable location estimations within such a multifaceted indoor space.

Manhattan Distance

Traditionally used in structured, grid-like environments, Manhattan distance calculates the absolute sums of the horizontal and vertical distances. It is a practical choice in urban settings where movement is typically restricted to orthogonal paths. However, this metric may not be ideal in a shopping mall, where pathways and walkable areas do not conform to a strict grid. The limitations of Manhattan distance become apparent in that it cannot effectively model the more fluid, diagonal, or direct paths that people naturally take in a mall building.

Euclidean Distance

In contrast, Euclidean distance measures the straight-line path between two points, regardless of any barriers. This metric is mathematically defined as the square root of the sum of the squared differences in their coordinates, making it a true representation of the shortest possible route between two locations. In the context of indoor navigation in a mall, Euclidean distance aligns more closely with how Wi-Fi signals propagate. Unlike Manhattan distance, Euclidean accounts for the direct path a signal would travel, including through obstacles, which mirrors the actual decrease in signal strength over distance more accurately.

The Euclidean distance between two WiFi fingerprints (F_i) and (F_j) is given by:

$$[d_{ij} = \sqrt{\sum_{k=1}^{N}(F_i[k] - F_j[k])^2]}$$

Where:

- (d_{ij}) represents the Euclidean distance between fingerprints (F_i) and (F_j).
- $(F_i[k])$ and $(F_j[k])$ denote the $(k) - th$ feature (e.g., RSSI value) of fingerprints (F_i) and (F_j) respectively.
- (N) is the total number of features in the fingerprints.

This formula reflects the Pythagorean theorem and is fundamental in calculating the actual physical distance a signal travels, which is crucial for RSS-based methods where signal strength decreases predictably with distance from the source. Given the irregular layout of most shopping centers, Euclidean distance offers a significant advantage by accurately reflecting the natural human navigation and signal attenuation through various media.

3.4. Weighted k-Nearest Neighbors (Wk-NN)

The Weighted k-Nearest Neighbors $(WkNN)$ algorithm emerges as a method for estimating a device's location with accuracy. This algorithm, an extension of the basic k-Nearest Neighbors (kNN) the approach which is proposed at [3], integrates the strength of the signal—or in a broader context, the similarity of the signal fingerprints—to determine the position of a user more precisely.

Foundational Principles of kNN and $WkNN$

The kNN algorithm in indoor positioning operates by selecting the k closest training samples in the feature space [Article 3], which in the case of Wi-Fi fingerprinting, are determined by the Euclidean distance between the RSS values of known points (reference points) and the querying point (user's current location). Traditionally, kNN would assign equal importance to each of the k neighbors, which can lead to inaccuracies if one or more of the neighbors have atypical signal strengths due to environmental anomalies or hardware variances.

Enhancement through Weighting

To refine this model, the $WkNN$ algorithm assigns weights to the $k - nearestneighbors$ based on their distance to the query point—the closer a point is, the greater its influence on the final location estimation. Mathematically, the weighting factor is typically inversely proportional to the distance, emphasizing neighbors that are closer as they are likely more similar to the querying point's actual conditions.

Mathematical Representation

The position of the user is calculated using a weighted average of the coordinates of these neighbors, where the weights are a function of their respective distances. If we denote the position of neighbor (i) as $((x_i, y_i))$ and the distance from the querying point to this neighbor as (d_i), the weights (w_i) could be calculated by the formula:

$$[w_i = \frac{1}{d_i^2}]$$

This choice of squaring the reciprocal of the distance helps in exaggerating the influence of very close neighbors, thus fine-tuning the sensitivity of the algorithm to small variations in distance. The estimated coordinates ((x, y)) of the user are then computed as:

$$[x = \frac{\sum(w_i \cdot x_i)}{\sum w_i}, \quad y = \frac{\sum(w_i \cdot y_i)}{\sum w_i}]$$

By weighting the contribution of each neighbor according to their proximity, $WkNN$ reduces the impact of anomalous signals that might otherwise skew the results in a simple kNN model. The algorithm is robust against the noise in signal measurements, which is common in dynamic indoor environments where human movement and electronic devices continuously alter the propagation conditions. $WkNN$ can be adapted to different environments by adjusting the weighting function or the value of k based on empirical data, providing flexibility to optimize the system according to specific deployment needs.

4. Proposed Approach

WiFi signals are electromagnetic waves that travel through the air, making them susceptible to reflection, refraction, diffraction, and absorption. This leads to intricate RSSI patterns with "fluctuations" as seen in [Figure 1].

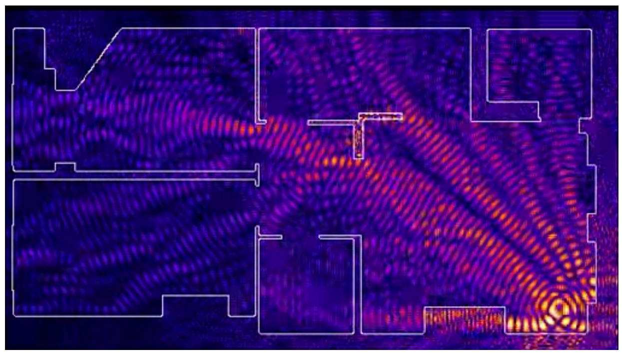

Figure 1: The simulation illustrates how WiFi strength propagates through space.

We utilized a sophisticated, layered architecture to provide accurate and reliable indoor localization through WiFi fingerprinting. This architecture integrates data acquisition, algorithmic analysis, and user interaction, in order to overcome the specific challenges of indoor navigation.

By using WiFi Fingerprint, we methodically captured WiFi signal strengths at various locations within an indoor environment and formed a database of fingerprints. Each comprising $RSSI$ values from visible APs correlated with precise spatial coordinates.

4.1. Locator Algorithm

We are presenting a method for estimating the location of a mobile device using wireless signal strength fingerprints (see Figure 2). We discuss the steps involved in the process and present the corresponding equations used to estimate the mobile device's location (see Table 1 for parameter descriptions).

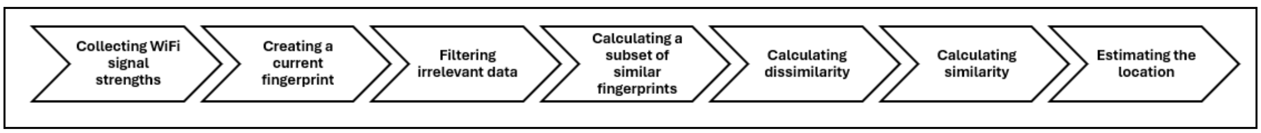

Figure 2: Locator Algorithm

1. **Collecting WiFi signal strengths:** Walk around the indoor space with a device that measures WiFi signal strengths and record the signal strengths at various locations. Each recorded set of signal strengths is called a fingerprint. These fingerprints are then stored in a database, along with their corresponding locations Articles [1] [6].
2. **Creating a current fingerprint:** Measure the WiFi signal strengths in the mobile device's current location and create a current fingerprint [Article 1].
3. **Filtering irrelevant data:** Ignore very low signal strengths from the current fingerprint using threshold T1 [1].
4. **Calculating a subset of similar fingerprints:** Calculate an Lf, a subset of fingerprints from the map with a relatively big number of common access points (i.e. reachable from the location of f and the fingerprint in Lf) [1]:

$$[L_f \subseteq M : \forall g \in L_f |aps(g) \cap aps(f)| \geq T_2]$$

Equation 1: Calculation of the subset Lf, consisting of fingerprints with common access points and a sufficient number of overlaps with the current fingerprint.

where aps(f) is a set of access points from fingerprint f. In the program, we used MAC addresses to indicate an access point.

5. **Calculating dissimilarity:** Calculate the dissimilarity between the current fingerprint and each fingerprint in Lf using the Euclidean distance in N-dimensional space, where N is the number of access points that are reachable from both the current fingerprint and a fingerprint from Lf. The j-th coordinate is the signal strength of the j-th access point in the fingerprint [1]:

$$[D(f,g) = \sqrt{\sum_{j=1}^{N} (f.RSSI(j) - g.RSSI(j))^2}]$$

Equation 2: Calculation of dissimilarity between the current fingerprint and a fingerprint from Lf, based on the difference in received signal strength (RSS) values of common access points.

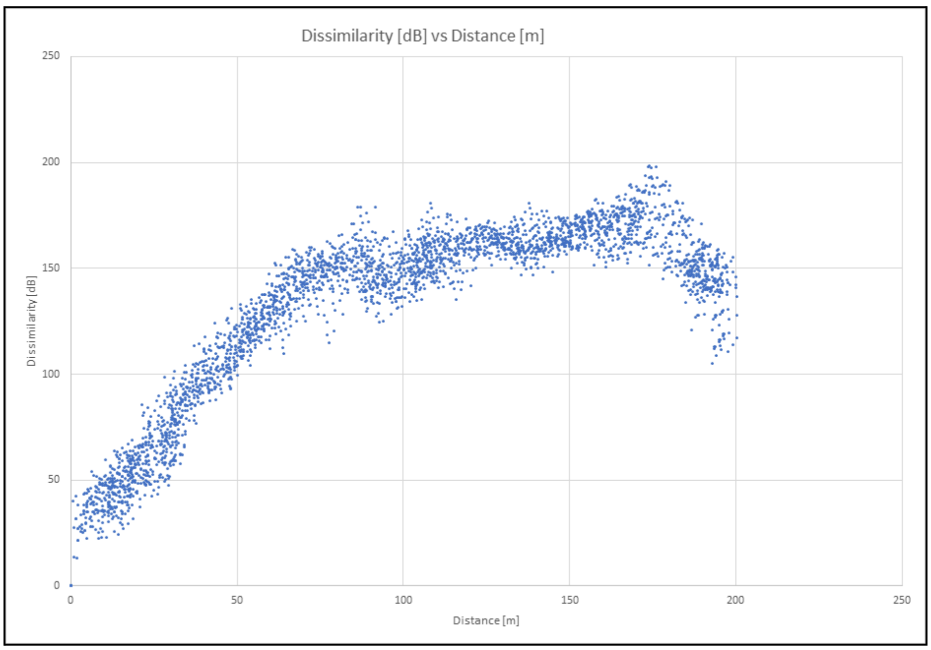

Figure 3: Illustration of Dissimilarity's strong linear correlation with actual physical distance.

Dissimilarity exhibits a strong linear correlation with actual physical distance. In [Figure 3], we can see a plotted D(f, g) in dB (Y-axis) between each pair of fingerprints against their corresponding distance in the real world, measured in meters (X-axis).

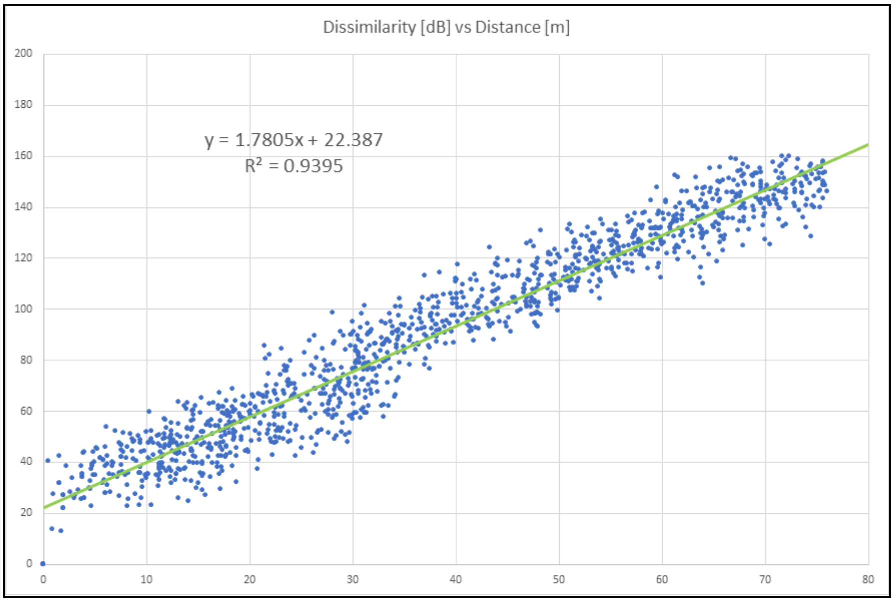

Figure 4: Illustration of linear correlation between signal strength and distance up to 75 meters.

There exists a strong linear correlation between signal strength and distance up to 75 meters [Figure 4]. Beyond this range, the relationship becomes more complex, with signals from distant sources becoming negligible, while signals from other access points become stronger.
The similarity of two fingerprints is inversely proportional to the distance between them, meaning that the farther apart the fingerprints are, the lower their similarity. We introduce similarity as the following step.

6. **Calculating similarity:** Calculate the similarity between the current fingerprint and each fingerprint in Lf using the [Article 1]:

$$[k_{(f,g)} = \frac{1}{D(f,g)}]$$

Equation 3: Calculation of similarity between the current fingerprint and a fingerprint from Lf, as the inverse of the dissimilarity value.

This similarity is used as a weighting factor in averaging locations of fingerprints in Lf.

7. **Estimating the location:** Estimate the location of the mobile device by calculating a weighted average of the locations of the similar fingerprints, where the weights are based on the similarity between the current fingerprint and each of the similar fingerprints [3]:

$$[loc(f) = \frac{\sum_{g \in L_f} k(f,g) \cdot loc(g)}{\sum_{g \in L_f} k(f,g)}]$$

Equation 4: Estimation of the location of the current fingerprint by computing the weighted average of the known locations of fingerprints in Lf, using the similarity values as weights.

Here loc(g) are known locations of fingerprints from Lf.

	Description	**Value**
M	Map of reference fingerprints	A set of fingerprints (location, APs, RSSI values)
f	Current fingerprint to be located	A fingerprint (APs, RSSI values)
$T1$	The threshold for ignoring very low signal strengths	A negative integer (empirically determined)
$T2$	The threshold for the minimum number of common APs between f and Lf	A positive integer (empirically determined)
Lf	A subset of fingerprints from M with a relatively big number of common APs with f	A set of fingerprints (location, APs, RSSI values)
N	Number of common APs between f and a fingerprint from Lf	A positive integer
$D(f,g)$	A measure of dissimilarity between f and a fingerprint from Lf	A non-negative real number
$k(f,g)$	The similarity weighting factor for averaging locations of fingerprints in Lf	A non-negative real number
$loc(f)$	The estimated location of f	A point in 2D or 3D space, depending on the map and coordinates system used

Table 1: Parameters used in the location estimation algorithm.

4.2. Dataset

The dataset utilized in our research project comprises WiFi fingerprints collected from various locations within indoor environments, specifically across the Haifa Azrieli Mall's first floor. Each WiFi fingerprint consists of several components:

1. Unique identifiers (MAC Addresses) assigned to each WiFi access point. MAC addresses serve as reference points in the fingerprinting process, enabling the identification of specific access points within the environment.
2. Received Signal Strength Indicator (RSSI) values represent the signal strength of each WiFi access point at a particular location. RSSI values are typically measured in decibels (dB) and provide insights into the proximity of the smartphone or device to each access point.
3. The X and Y coordinates denote the spatial position of each fingerprint within the indoor environment. These coordinates serve as ground truth data for evaluating the accuracy of our positioning algorithm.

Dataset Collection Process

The dataset was collected through systematic surveys conducted across different areas within the mall. During the surveying process, a smartphone equipped with WiFi scanning capabilities was used to capture WiFi signals from surrounding access points. The smartphone's location coordinates were recorded concurrently using our algorithm implementation.

Dataset Characteristics

The dataset covers a diverse range of locations within the mall, including corridors, shops, common areas, and intersections. This ensures comprehensive coverage of the indoor environment, enabling robust algorithm testing and evaluation.

The dataset captures variations in WiFi signal propagation caused by environmental factors such as building layout, obstacles, and interference. This variability reflects real-world conditions and facilitates the development of robust positioning algorithms capable of handling diverse scenarios.

With 786 WiFi fingerprints collected across the mall's floor, the dataset provides a substantial volume of data for algorithm development and evaluation. The dataset's scale enables statistical analysis and validation of positioning accuracy across different regions of the indoor environment.

Fingerprint

WiFi Fingerprint	
MAC	Signal Strength
20:bb:c0:1d:c3:40	-83
f4:83:cd:6a:c4:d6	-68
...	...
d3:21:33:a7:a1:c4	-45

Location	
X	Y
148.5432	40.3214

Figure 5: Fingerprint data representation

4.3. Algorithmic Enhancements

These enhancements cater to the complexities and inherent issues of indoor Wi-Fi environments and the variability in signal strengths due to various physical and electronic factors. Let's delve deeper into these enhancements and understand their significance in our study:

RSS OFFSET Handling Missing APs

Enhancement

The RSS OFFSET is a numerical value added to the signal strength difference when an AP is detected in one fingerprint but not in another. This is done to handle missing data, which is common due to AP visibility fluctuating depending on physical obstructions, signal interference, or changes in AP availability.

Why It's Necessary

Indoor environments can have dynamic changes in Wi-Fi signal availability. This offset ensures that the absence of a signal does not skew the results unrealistically in favor of a non-existent match. Without this offset, a missing AP would not contribute to the dissimilarity measure, potentially resulting in a misleadingly low dissimilarity score and inaccurate location estimation. It also allows for tuning based on empirical data or environmental testing, making the system adaptable to different settings or specific requirements of a deployment location.

In our approach, non-overlapping APs contribute to the dissimilarity score by adding a fixed offset (RSS_OFFSET), which is a nuanced method of handling missing data that might not be present in other systems. This specific handling of RSS values where missing APs are concerned by adjusting their influence in the distance calculation is unique when compared to the other studies, which do not explicitly account for non-detected APs in such a manner.

Scoring Mechanism for Fingerprint Comparison

The scoring mechanism involves calculating a score based on the intersection and differences of AP sets between two fingerprints. This score helps in pre-filtering fingerprints before detailed dissimilarity calculations, focusing computational resources on the most promising candidates.

Why It's Necessary

By scoring and selecting only those fingerprints that have a significant overlap in AP visibility with the current observation, the system reduces the number of distance calculations needed. This is particularly beneficial in environments with a large number of stored fingerprints. Focusing on fingerprints with higher scores (more AP overlap) ensures that the subsequent detailed calculations are performed on the most relevant data, improving the likelihood of accurate localization. As the database of fingerprints grows, this mechanism helps in maintaining performance by limiting the growth of computational load.

The Operation of the Scoring System

Inputs: Two fingerprints - one from the current dataset (fingerprint) and one from the reference set (refFp).

1. **Process:**
   - The system first retrieves sets of APs from both fingerprints that exceed a defined minimum signal strength (MIN_RSS_TO_COUNT). This threshold ensures that only significant APs are considered, reducing noise in the data.
   - It then calculates the intersection of APs present in both fingerprints and identifies APs unique to each.
   - The score is computed using the formula: 2 * size of intersection - size of unique APs in fingerprint - size of unique APs in refFp. This scoring mechanism quantifies the similarity between two fingerprints—the higher the score, the greater their similarity.
2. **Output:** An integer score that serves as a quantitative measure of similarity between the two fingerprints.

The scoring can be mathematically represented as:

$$\text{Score} = 2 \times |\text{Intersection}(A, B)| - |\text{Unique}(A)| - |\text{Unique}(B)|$$

where (A) and (B) represent the sets of APs from fingerprint and refFp, respectively.

The scoring system is tailored to prioritize fingerprints that have a greater number of overlapping APs with the query fingerprint, which is not a common feature detailed in most traditional k-NN based systems. This method specifically enhances the relevance of the fingerprints used for the final location calculation, potentially reducing noise and improving accuracy. Most studies, such as those by Moreira et al. [Article 5] and Torres-Sospedra et al. Articles [2] [6], do not mention using a scoring system to filter data before applying k-NN. Their methods typically involve more straightforward application of k-NN or WKNN without an intermediate scoring filter.

Here's a detailed look at the methodologies employed by other researchers as referenced in our studies, contrasting them with the scoring approach used in our method.

Comparison to other Articles methods

- **Moreira et al. [Article 5]**

Approach: Utilizes the Weighted k-Nearest Neighbor (WKNN) algorithm, focusing on different weighting functions and distance metrics to optimize the location estimation.

Key Technique: The study experiments with different distance functions such as Manhattan, Euclidean and Canberra, combined with weighting functions like $(\frac{1}{d_i})$ and $(\frac{1}{d_i^2})$, where (d_i) is the distance to the $(i)-th$ nearest neighbor.

Comparison: Unlike the scoring method in our project, Moreira et al. do not filter fingerprints based on overlapping APs but rather optimize the weighting and distance calculations directly applied to the k-NN algorithm. This approach focuses more on refining the mathematical accuracy of the k-NN calculation rather than pre-filtering data.

- **Zhu [4]**

Approach: Investigates an Optimized KNN (OKNN) algorithm, which adjusts the k-NN parameters for better performance.

Key Technique: Zhu emphasizes the optimization of the number of neighbors (k value) and possibly the algorithm's internal parameters to minimize localization errors.

Comparison: The focus is directly on enhancing the core k-NN algorithm rather than pre-selecting or scoring fingerprints. This method seeks to directly refine the predictive accuracy through parameter optimization without an intermediate filtering stage.

- **Torres-Sospedra et al. [Article 2]**

Approach: Conducts a comprehensive analysis of over 50 different distance and similarity measures for Wi-Fi fingerprinting systems.

Key Technique: This research does not specifically enhance k-NN but provides a broad understanding of how various distance measures can influence the accuracy of localization systems.

Comparison: There is no direct manipulation of the fingerprint dataset before applying k-NN; instead, the study provides foundational knowledge that could be used to select the most effective distance measures for any k-NN-based system.

- **Liu et al. [7]**

Approach: Combines WKNN with k-means clustering to preprocess the fingerprint dataset, enhancing the efficiency and effectiveness of the k-NN algorithm.

Key Technique: Implements k-means clustering to group similar fingerprints before applying the k-NN, which reduces computational complexity and potentially increases the accuracy by ensuring that k-NN computations are performed within more homogeneously grouped data.

Comparison: Liu et al. introduce a preprocessing step, but unlike our scoring system, which filters based on AP overlap, their method clusters data to reduce variability and computational demand during the k-NN processing phase.

5. Research Process

5.1. Road Map

Literature Review and Research Planning

Following our extensive literature review, we crafted a research plan that leverages insights from key studies to enhance the functionality of indoor navigation systems (INS). This plan outlined the strategic use of weighted k-nearest neighbors (k-NN) and Euclidean distances, which were chosen based on their demonstrated efficacy in prior research and their potential for further enhancement with advanced techniques.

The scope, objectives, and methodologies of our project were defined within our research plan. This plan was specifically designed to address the essential questions that arose during our analysis. We aimed to investigate whether the integration of a scoring system and RSS OFFSET could significantly improve the accuracy and reliability of INS employing WiFi fingerprinting techniques, particularly in complex environments such as shopping mall.

Dataset Collection and Preparation

In our dataset collection and preparation phase, we implemented a structured approach to accurately capture WiFi fingerprints throughout a designated indoor environment, specifically a mall's floor. Using a laptop equipped with specialized software, we systematically recorded WiFi signal strength data. As we navigated through various key points of interest in the mall—such as store entrances, food courts, and main hallways—we ensured that RSSI measurements were captured at one-meter intervals.

Each RSSI measurement was meticulously integrated into a structured JSON format, forming a comprehensive and organized dataset ready for further processing and analysis. Here is an example of how the WiFi fingerprint data was structured in our JSON array file, showing a sample with RSSI values:

```
{
  "CLASSNAME": "Fingerprint",
  "INSTANCE": {
    "mWiFiFingerprint": {
      "20:bb:c0:1d:c3:40": -83,
      "f4:83:cd:6a:c4:d6": -68,
     …
      "9c:97:26:4c:55:21": -81,
      "00:0e:8e:7a:3d:c3": -73
    },
    "mCenter": {
      "x": 148.07382,
      "y": 40.461315
    },
    "mIsRemoved": false
  }
}
```


This JSON structure captures the essential elements of our WiFi fingerprinting data, including the unique identifiers of WiFi access points and their corresponding RSSI values, along with the precise location coordinates where each measurement was taken.

Algorithm Development

With a robust dataset in hand, we moved on to algorithm development. This stage involved implementing preprocessing techniques to clean and standardize the WiFi fingerprint data, ensuring that it was in the optimal form for analysis and application. Drawing inspiration from the methodologies proposed in our earlier literature review, we developed sophisticated algorithms for distance and similarity measures.

Weighted k-NN: The decision to use weighted k-NN was influenced by findings from Torres-Sospedra et al. [Article 2], who highlighted the effectiveness of k-NN in managing the spatial variability inherent in indoor environments. The weighted aspect was introduced to refine the accuracy further by assigning greater importance to nearer points, thus reducing the influence of outliers or distant signals.

Euclidean Distances: We chose to continue utilizing Euclidean distances as our primary metric for distance calculation due to its simplicity and widespread application in existing studies. However, our goal was to explore potential improvements as suggested by the work of Ezhumalai et al. [Article 1], which could be achieved by adjusting the metric to better accommodate the specific signal attenuation characteristics of indoor environments.

Scoring System and RSS OFFSET: Inspired by the innovative approaches of Ma, Wu, and Poslad [Article 7] and Babalola and Balyan [Article 8], we decided to implement a scoring system and RSS OFFSET. These enhancements are intended to manage the variability and inconsistency of RSSI values collected from WiFi access points. The scoring system prioritizes fingerprints based on their signal strength and stability, which aligns with our objective of enhancing reliability in RSSI-based location estimation.

5.2. Challenges

Initially inspired by the frustration of searching for a store in a mall, we delved into a complex world of algorithms and techniques. As we progressed, we encountered numerous obstacles, from understanding the intricacies of Wi-Fi fingerprinting to implementing efficient data preprocessing and positioning algorithms. Our dedication to overcoming these challenges was driven by a desire to create a solution that would simplify indoor navigation for others, just as we had hoped for during that mall visit. The journey has been both rewarding and enlightening, pushing us to grow and learn in ways we hadn't anticipated.

Proof of Concept

Started our journey by researching a prototype that incorporated the latest Wi-Fi fingerprinting techniques, allowing us to demonstrate the practical applications of our research. To validate the accuracy and reliability of these techniques, we conducted controlled experiments within simulated indoor environments. These experiments were meticulously designed to mimic real-world conditions as closely as possible, thus ensuring that our findings would be applicable in actual commercial settings like malls.

Reliability of RSSI Points

One of the core components of our system is the use of RSSI (Received Signal Strength Indicator) points, which are crucial for the accuracy of Wi-Fi fingerprinting. However, RSSI points are notoriously susceptible to variability and inconsistency due to multiple factors, including physical obstructions, dynamic environmental conditions, and hardware limitations. To address these challenges, our team invested significant effort into investigating the specific factors contributing to RSSI variability.

To enhance the reliability of RSSI measurements, we implemented a series of data preprocessing techniques. These included advanced noise filtering algorithms and outlier detection mechanisms designed to clean the data before it was processed by our positioning algorithms. Furthermore, we conducted extensive validation and calibration procedures, which were essential to ensuring the accuracy and consistency of our RSSI-based location estimation. These efforts were critical in developing a system that could reliably perform in diverse indoor environments, providing users with precise location data.

Integrating Vectorizing and Location Aspects

Integrating the vectorizing of Wi-Fi fingerprints with the location estimation component posed significant challenges. The vectorization process involves converting the raw RSSI data into a format that can be effectively used by our algorithms to estimate locations. This process had to be aligned with the actual spatial coordinates within the indoor environment to ensure that each vectorized fingerprint accurately corresponded to a physical location.

Ensuring the accuracy between these vectorized fingerprints and their corresponding spatial positions was paramount, as any discrepancies could lead to significant errors in location estimation. This aspect of the project demanded a high level of precision and technical skill, as it was fundamental to the overall functionality and effectiveness of our positioning system.

6. Experiments and Results

The experiment was designed to capture a set of Wi-Fi signal strength readings, or 'fingerprints', across the entire expanse of a mall floor. The resulting scatter plot visualizes the spatial distribution of these fingerprints and provides insights into the overall signal coverage and the algorithm's subsequent ability to locate users accurately within the space. We systematically collected fingerprints at numerous locations throughout the mall, with each blue dot on the scatter plot representing a unique set of RSSI readings associated with precise X and Y coordinates. The goal was to create a robust dataset that reflects the varied signal conditions encountered in different mall areas, from open spaces to more confined corridors.

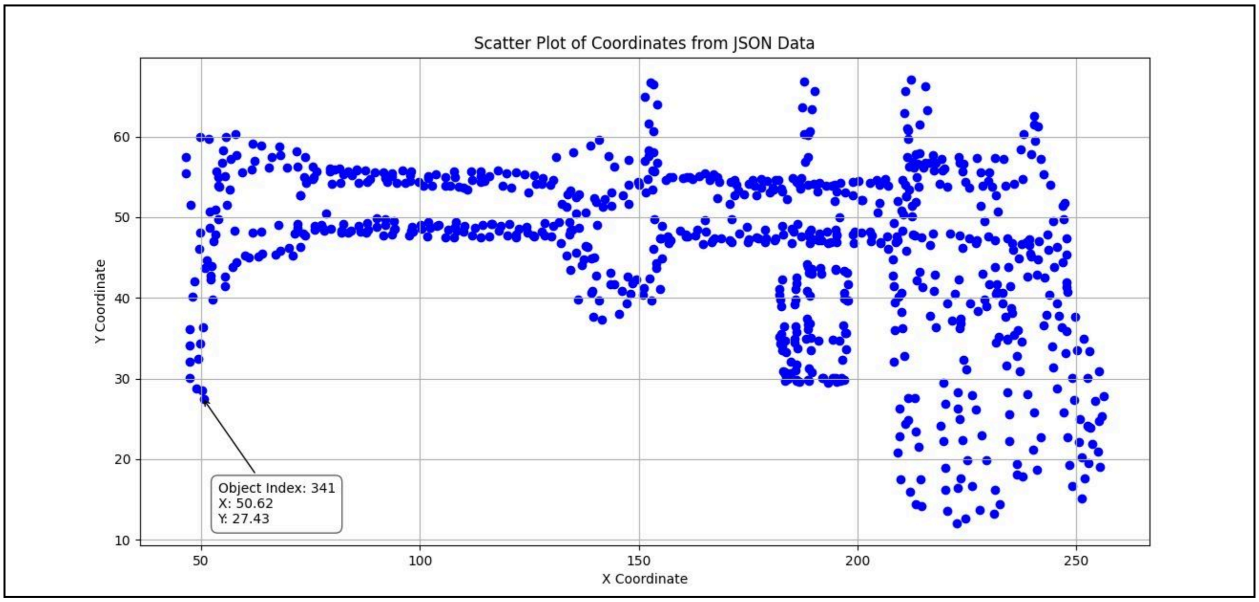

Figure 6: Fingerprints data collected fingerprints at numerous locations throughout the mall.

1. The "Scatter Plot of Coordinates from JSON Data" [Figure 6] offers a visual representation of the collected fingerprints:
2. The blue dots scattered across the plot demonstrate an extensive collection effort, covering a wide array of spatial scenarios within the mall.

Clusters of fingerprints suggest areas of high data concentration, which are instrumental in improving the resolution and reliability of the positioning system. Conversely, sparser regions on the plot might correspond to less frequented areas or physical impediments to signal propagation.

6.1. Stability Evaluation of RSSI Readings

The study entailed collecting multiple RSSI measurements from various MAC addresses at a single, fixed location. The objective was to replicate a static user environment, closely monitoring the RSSI value consistency from the surrounding access points.

A dataset was compiled through a sequence of tests designed to map out the RSSI value behavior over time for each MAC address encountered in the space.

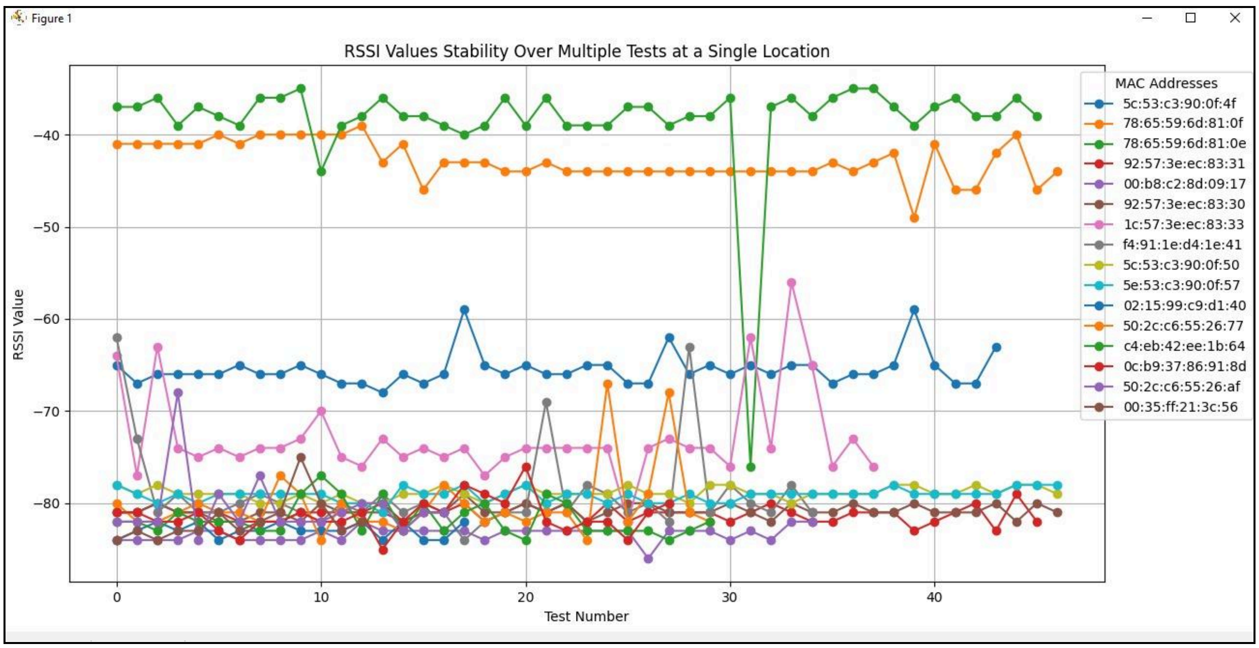

Figure 7: Fingerprint RSSI measurements from various MAC addresses at a single, fixed location.

Analysis and Observations

Several MAC addresses, particularly those associated with $5c : 53 : c3 : 90 : 0f : 4f$ and $78 : 65 : 59 : d8 : 1d : 0f$, showcased remarkable stability in their RSSI readings, indicating potential as reliable anchors for the positioning algorithm.

Lower-strength signals, indicated by MAC addresses like $00 : b8 : c2 : 8d : 09 : 17$ and $92 : 57 : 3e : ec : 83 : 30$, displayed notable fluctuations [Figure 7]. These variances suggest susceptibility to environmental changes, such as movement or structural alterations within the test space.

Specific MAC addresses, such as $5e:53:c3:90:0f:57$, exhibited pronounced spikes in RSSI values. These outliers are indicative of temporary disturbances or signal interference, which may skew location estimation if not appropriately addressed in the positioning algorithm.

Despite the presence of outliers and variability, the general consistency observed in the stronger RSSI values suggests that the data holds potential for a reliable fingerprinting-based positioning system.

This analysis substantiates the viability of RSSI values in establishing an accurate indoor positioning system. The persistent signal patterns, particularly from stronger sources, provide a solid foundation for fingerprinting algorithms. Conversely, the need for intelligent handling of anomalies, such as the spikes from MAC address $5e:53:c3:90:0f:57$, is underscored [Figure 7], emphasizing the importance of robust algorithmic design that can accommodate such irregularities.

mac	Test #	mean	std	min	25%	50%	75%	max
5c:53:c3:90:0f:4f	44	-65.4773	1.784802	-68	-66	-66	-65	-59
78:65:59:6d:81:0f	47	-42.8298	2.088469	-49	-44	-43	-41	-39
78:65:59:6d:81:0e	46	-38.3913	5.900954	-76	-39	-38	-36	-35
92:57:3e:ec:83:31	46	-81.1304	0.832898	-83	-82	-81	-81	-79
00:b8:c2:8d:09:17	35	-83.3429	0.764771	-86	-84	-83	-83	-82
92:57:3e:ec:83:30	47	-80.8511	0.690895	-82	-81	-81	-80	-79
1c:57:3e:ec:83:33	38	-72.9474	4.798352	-82	-75	-74	-73.25	-56
f4:91:1e:d4:1e:41	35	-78.8857	4.843257	-84	-81	-81	-79	-62

mac	Test #	mean	std	min	25%	50%	75%	max
5c:53:c3:90:0f:50	47	-78.9149	0.653745	-80	-79	-79	-78.5	-78
a2:b5:3c:b1:95:26	6	-81.5	3.72827	-84	-83	-83	-82.25	-74
5e:53:c3:90:0f:57	47	-79.1064	0.698882	-81	-79.5	-79	-79	-78
02:15:99:c9:d1:40	18	-82.9444	0.802366	-84	-83.75	-83	-82	-82
50:2c:c6:55:26:77	30	-80.1667	3.769966	-84	-82	-81	-80	-67
a8:63:7d:10:9e:42	5	-81.4	1.81659	-84	-82	-81	-81	-79
fa:8f:ca:92:de:00	7	-82.1429	1.676163	-85	-83	-81	-81	-81
c4:eb:42:ee:1b:64	30	-81.7	1.784029	-84	-83	-82	-80.25	-77
0c:b9:37:86:91:8d	28	-81.3929	1.812289	-85	-82	-82	-81	-76
50:2c:c6:55:26:af	17	-80.4706	3.590224	-84	-82	-82	-80	-68
00:35:ff:21:3c:56	13	-81.8462	2.375084	-84	-83	-83	-81	-75
68:aa:c4:b2:63:40	5	-81.2	1.643168	-83	-82	-82	-80	-79

Table 2: Fingerprint RSSI data from various MAC addresses.

The table provides a view of how the RSSI values for each MAC address vary across multiple tests. This is for analyzing the stability and reliability of the WiFi signal from various devices in a specific location. By looking at the mean, you can gauge the average signal strength, while the standard deviation will tell you about the consistency of the signal. The range between the min and max values gives an indication of the extremes of signal strength during the testing period.

Columns in the Table

1. **count:** The number of tests in which each MAC address was observed. This number indicates how many RSSI values were recorded for that particular MAC address.

2. **mean:** The average RSSI value recorded for each MAC address over the tests. This gives a general idea of the signal strength for that MAC address at the location.
3. **std (standard deviation):** This measures the variability or spread of the RSSI values around the mean. A high standard deviation indicates that the RSSI values vary widely from the mean, while a low standard deviation indicates that the values are clustered closely around the mean.
4. **min (minimum):** The lowest RSSI value recorded for each MAC address. This shows the weakest signal strength detected during the tests.
5. **25% (25th percentile):** This value indicates that 25% of the RSSI readings are below this value. It's a measure of the lower quartile and helps in understanding the distribution of data.
6. **50% (median):** The median RSSI value. Half of the RSSI readings are above this value, and half are below. This is often considered a better indicator of the typical value than the mean, as it is less influenced by outliers and skewed data.
7. **75% (75th percentile):** This value indicates that 75% of the RSSI readings are below this value, making it a measure of the upper quartile.
8. **max (maximum):** The highest RSSI value recorded for each MAC address. This represents the strongest signal strength detected during the tests.

Results

1. The average RSSI values ranged significantly across devices, from as high as -38 dBm for MAC 78:65:59:6d:81:0e indicating a strong signal, to as low as -83 dBm for devices like 00:b8:c2:8d:09:17, showing a much weaker signal. This variation could be attributed to the physical location differences relative to the WiFi source or the different capacities of devices to receive signals.

2. The standard deviation (std) provides a measure of how consistent the signal strength was for each device. For instance, 92:57:3e:ec:83:31 showed a remarkably consistent signal reception with a std of 0.83, whereas 78:65:59:6d:81:0e experienced high variability in signal strength (std of 5.90). The high variability might suggest either movement during the test or interference affecting the signal.

3. The minimum and maximum values highlight the range of RSSI values recorded. Devices like 1c:57:3e:ec:83:33 displayed a dramatic range from -82 dBm to -56 dBm, indicating potential periods of significant signal improvement or deterioration during the test.

4. Several MAC addresses such as 1c:57:3e:bf:af:53 and 44:d4:53:e5:71:96 were recorded very few times (count = 1), making it challenging to draw reliable conclusions about their signal characteristics. This sparsity could impact the robustness of network performance assessments.

6.2. Accuracy Assessment of Nearest Fingerprint

Following the widespread data collection of Wi-Fi fingerprints throughout a mall, we proceeded to evaluate the precision of our indoor positioning system. This section delves into the validation of our system's ability to accurately pinpoint the nearest fingerprint to a given user location input.

From the comprehensive collection of 786 fingerprints, we temporarily removed one specific fingerprint to simulate a real-time user location input. The 'User Location Input' was set at coordinates (214.95195, 56.1425) [Figure 8]. The system was then tasked to identify the nearest fingerprint from the remaining dataset, resulting in the 'Calculated User Location to Nearest Fingerprint' with coordinates (213.463394, 54.926720) [Figure 8].

A single fingerprint, originally part of the dataset, is chosen to simulate a real-time user location input. This is represented by the coordinates (214.95195, 56.1425). In the scatter plot, this point may be depicted by the red dot.

The system's algorithm is tasked to find the closest match to the user's location input from the remaining fingerprints. The resulting nearest location, based on the calculated shortest RSSI value distance, is at (213.463394, 54.926720).

The plots include blue dots representing the available fingerprints' coordinates and red dots indicating the user's simulated input and the system's calculated nearest location.

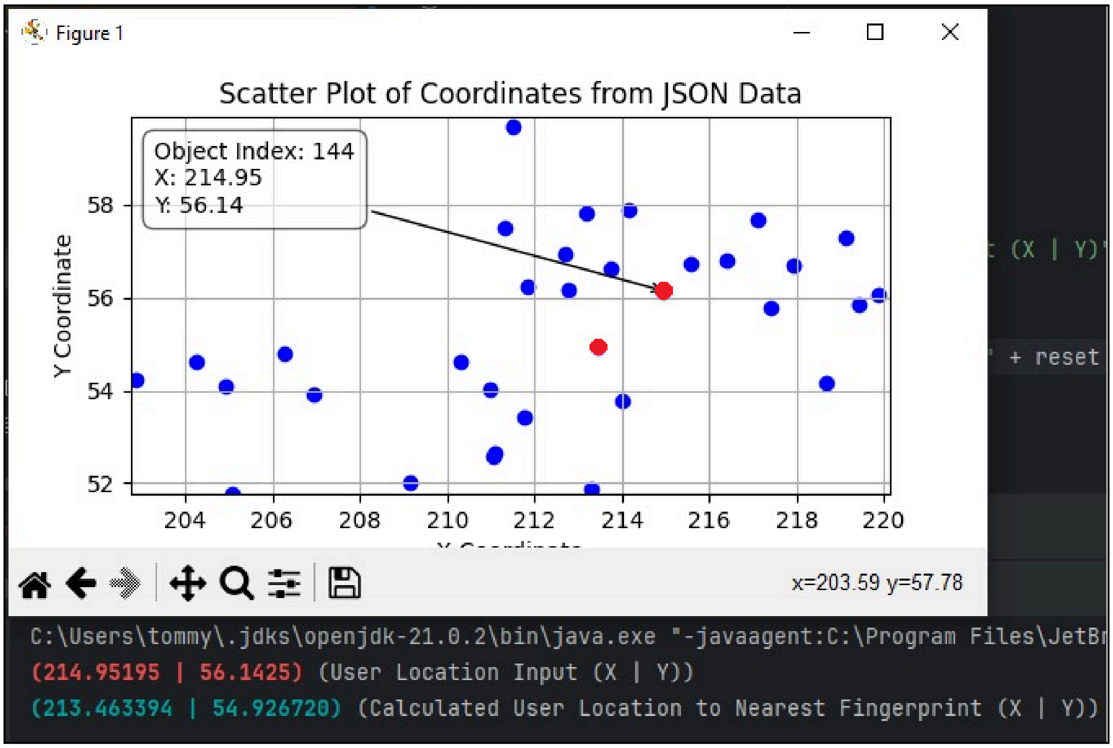

Figure 8: Output result of a location estimation based on training data with arrow point at expected location.

By analyzing the proximity of the removed dot label as index 144 (Which simulated the User’s location at a given time), we feed its RSSI values and combine them with the existing data of the cluster of blue dots and run our algorithm, the accuracy of the positioning system can be then assessed. The calculated nearest point will be within a very close range of the user input to indicate high accuracy. The distance between these two points on the scatter plots serves as an empirical measure of the system’s precision in real-world conditions.

If the dot representing the calculated nearest fingerprint is very close or coincides with the user's input location, the system demonstrates high accuracy and colors it red. In the provided images, the two red dots appear to be in close proximity, suggesting that the system accurately identified the nearest fingerprint to the user's simulated location and the dot which was removed from the database.

The test results, as visualized in the scatter plots, suggest that the positioning algorithm can reliably identify a fingerprint that closely matches the user's location input within the dataset. The proximity of the calculated nearest point to the simulated user input point underscores the system's potential effectiveness in a real deployment.

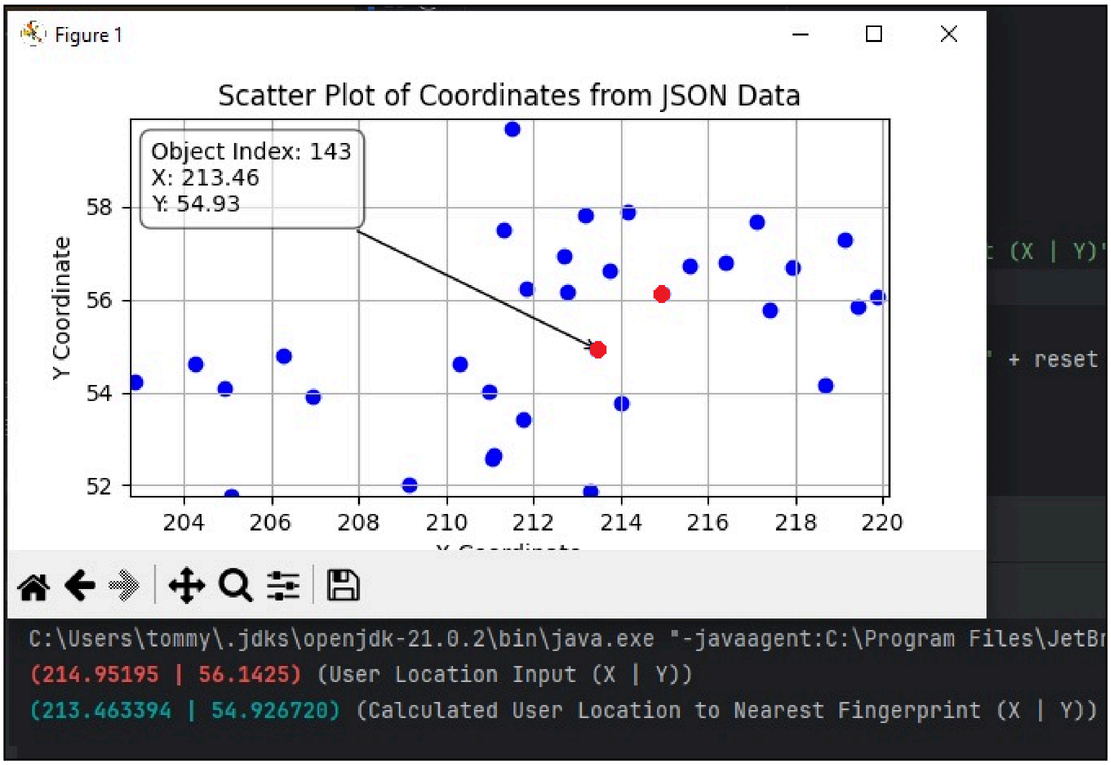

Figure 9: Output result of a location estimation based on training data with arrow point at actual location.

Results:

- The location input by the user, intended to mimic the real-time scenario of a person navigating the mall, is depicted by a red dot.
- The blue dots represent the available fingerprints, with the calculated nearest point showcased as a red dot. This exhibits the system's response to the user input.
- The closeness of the 'Calculated User Location' to the 'User Location Input' is visually evident, suggesting a high degree of localization accuracy.

The refined analysis for the test utilizes the additional data provided in the JSON file to pinpoint the fingerprint with the closest RSSI values to the central fingerprint, represented by Object Index 144. The comparison aims to highlight that despite the spatial proximity of other fingerprints (orange dots) to the central point, it is the similarity in RSSI values that determines the closest match, not just the physical distance.

Through the analysis of the provided data, it has been determined that Object Index 143 (green dot), has RSSI values that are more closely aligned with the central fingerprint than those of the other nearby fingerprints (orange dots), despite being further away in terms of physical distance. This underscores the significance of RSSI values in establishing proximity in a Wi-Fi fingerprinting context.

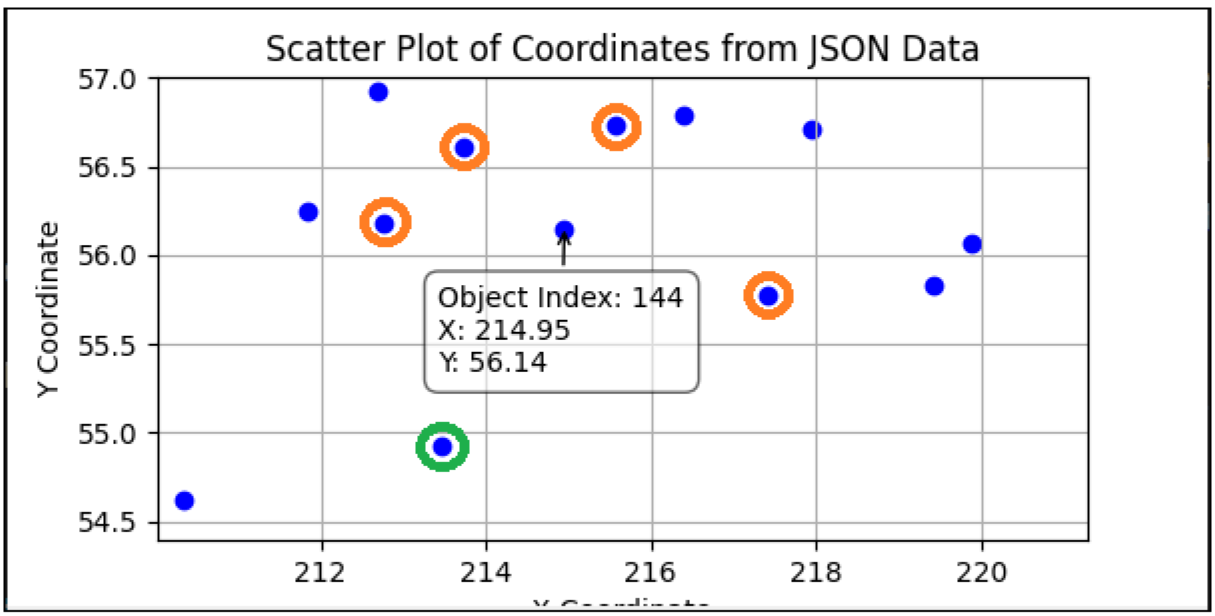

Figure 10: Showcasing different points closer to the expected location (orange) than the actual result (green).

To elucidate, the central fingerprint at Object Index 144 is characterized by specific RSSI readings such as -87 for MAC address "80:37:73:e3:cd:c8", -79 for "f4:83:cd:6a:c4:d6", and so forth . In contrast, the surrounding fingerprints at indices 694, 420, 695, and 704 exhibit varying degrees of RSSI differences, indicating that their signal strengths differ from those of Object Index 144 despite their closer physical presence. Object Index 143's RSSI readings show a greater resemblance to the central fingerprint's values, thereby suggesting it as the closest match based on Wi-Fi signals rather than mere spatial approximation.

	Near the Removed Fingerprint				Actual	Removed Index
MAC/FP-DB Index	694	420	695	704	143	144
b4:e9:b0:cb:06:20	-84	-82	-74	-84	-74	-74
70:62:b8:32:a0:15	-87	-88	-87	-85	-86	-86
20:bb:c0:1d:db:90	-44	-48	-59	-62	-52	-52
00:0e:8e:7a:99:26	-73	-75	-63	-69	-63	-71
00:18:e7:fd:01:e2	-67	-66	-75	-74	-68	-72
70:62:b8:32:a0:14	-74	-82	-83	-78	-75	-75
fc:75:16:b0:ae:75	-78	-78	-74	-74	-73	-78
c0:ff:d4:bd:97:16	-64	-76	-69	-68	-72	-72
Similar RSSI values	1	1	1	0	5	-

Table 3: Values with data that show different points closer to the expected location (orange) than the actual result (green).

The conclusion drawn from the test is that the accuracy of Wi-Fi fingerprinting for location-based services relies more on the correlation of signal strengths than on geographical closeness. This insight is crucial for the development of indoor positioning systems where GPS signals are unreliable, and RSSI values from Wi-Fi access points become the primary source for localization.

6.3. Decreasing Fingerprints Neighbors Experiment

In the "Decreasing Fingerprints Cluster Experiment," we investigated how the number of present fingerprints in an area relates to the error in meters of the output location. By systematically reducing the number of surrounding fingerprints and observing the resulting changes in accuracy, we aimed to determine the impact of fingerprint quantity on the algorithm's performance. This experiment allowed us to identify patterns and trends that inform the optimal fingerprint quantity for achieving the highest localization accuracy.

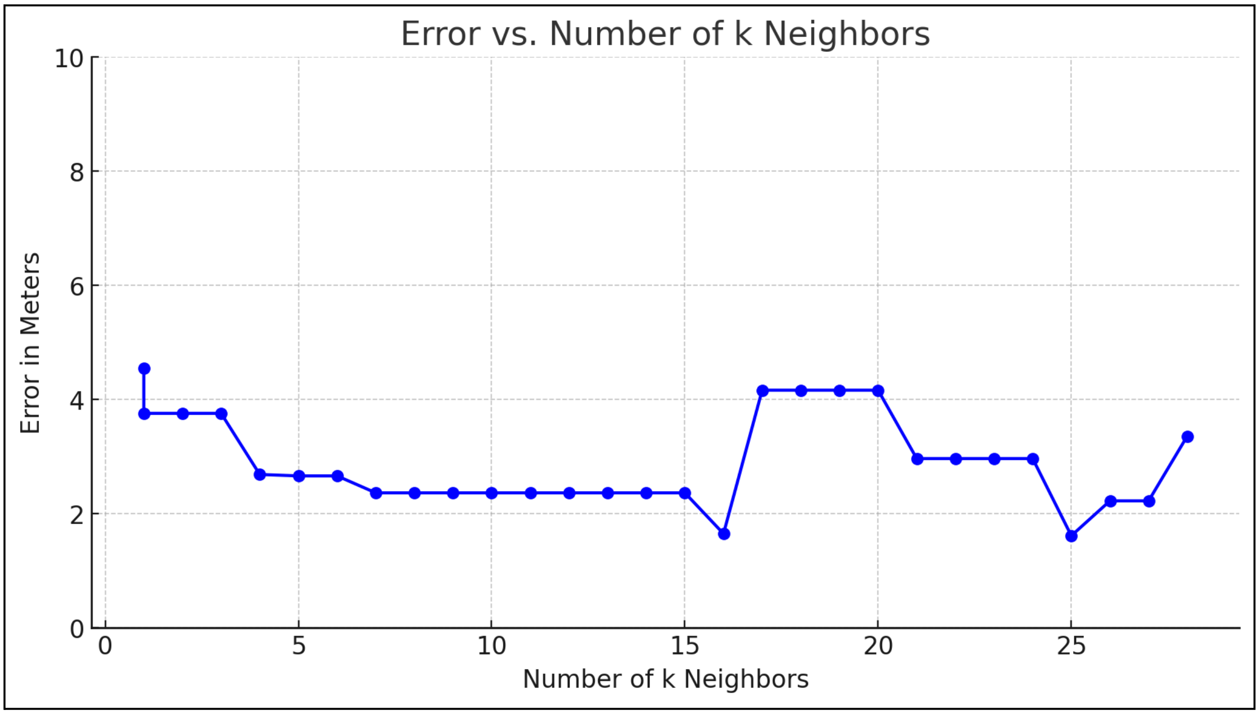

Figure 11: Detailed Overview of Experimental Results.

- The minimum error observed was 1.613164 meters at 25 neighbors.
- For a broader range of 'k' values from 7 to 15, the error consistently hovered around 2.365007 meters.
- Notably, even when increasing 'k' values up to 16, the error rates remained relatively low, with a slight rise to 1.648885 meters, indicating stability in the model's performance under less complex architectural constraints.

The enhancements implemented in the mall experiment address key challenges typically encountered in indoor Wi-Fi environments:

RSS OFFSET Handling Missing APs

This adjustment compensates for missing data when an AP is detected in one fingerprint but not in another. By adding a numerical offset to the signal strength, the system better accounts for fluctuating AP visibility, which is crucial in densely populated commercial settings where physical obstructions and electronic interference are common.

Scoring Mechanism for Fingerprint Comparison

This mechanism calculates a score based on the intersection and differences of AP sets between two fingerprints, allowing the system to pre-filter and focus on the most promising candidates for detailed distance calculations. This not only enhances accuracy by focusing on relevant data but also conservatively uses computational resources, crucial for maintaining system performance in environments with extensive fingerprint databases.

Comparative Analysis and Conclusions

The stark contrast in minimum error rates between the two experiments underscores the effectiveness of the enhanced k-NN algorithm in simpler architectural settings:

- The referenced study's best error rate of 6.19 meters contrasts sharply with the 1.613164 meters achieved in the mall experiment, illustrating a nearly fourfold improvement in positioning accuracy.
- The consistent performance across a range of 'k' values in the mall setting suggests that the algorithmic enhancements are well-suited to environments with less structural complexity but high user density, such as malls.

#knN	Error in Meters
1	4.544252
1	3.757522
2	3.757522
3	3.757522
4	2.685940

#knN	Error in Meters
5	2.661752
6	2.661752
7 - 15	2.365007
16	**1.648481**
17	4.163417
18	4.163417
19	4.163417
20	4.163417
21	2.964703
22	2.964703
23	2.964703
24	2.964703
25	**1.613164**
26	2.224260
27	2.224260
28	3.358885

Table 4: This table is to focus on the distances between points and the respective number of neighbors considered, with each value formatted.

This comparative analysis not only highlights the potential of specialized adaptations in improving the accuracy of indoor positioning systems but also showcases the importance of tailoring algorithms to specific environmental characteristics. The enhanced k-NN approach significantly outperforms traditional methods in less complex settings, suggesting its viability as a scalable model for similar commercial environments aiming to enhance their localization services.

7. Discussion

In examining various studies on Wi-Fi RSS fingerprint-based indoor positioning using the Euclidean distance with k-Nearest Neighbors (k-NN) and Weighted k-Nearest Neighbors (WKNN), we compare their results with our own. By focusing on the number of nearest neighbors and the achieved error rates, we can discern the effectiveness of each approach.

Our Project: Utilizing the Euclidean approach with WKNN, we attained a mean error of 1.613164 meters with 25 nearest neighbors. The enhancements, including the scoring system and RSS Offset, have bolstered our competitiveness within the complex setting of a commercial mall.

Moreira et al. [Article 5]: With a WKNN approach and k=3, their study achieved a mean error between 6.20 to 6.79 meters in a multi-building environment. While this setting is intrinsically different from our single-floor mall, our method demonstrates superior precision.

Torres-Sospedra et al. [Article 2]: Their comprehensive analysis did not specify error rates but emphasized the critical role of Euclidean distance in influencing positioning accuracy. Their findings suggest that strategic data transformations could further improve our method's accuracy.

Liu et al. [Article 7]: Demonstrated mean errors approximately 8.4 meters for Dataset 2, with different k values explored. Their use of k-means clustering in conjunction with WKNN suggests that clustering techniques might be an area where our project could see improved results.

Study	Algorithm	K Value	Mean Error (meters)
Our Project	WKNN + Score	25	**1.613164**
Moreira et al. [5]	WKNN	3	6.20 - 6.79
Torres-Sospedra et al. [2]	k-NN Variants	3	~7.4
Liu et al. [7] (Dataset 2)	WKNN + k-means	4	~8.4

Table 5: This table compares the error rates achieved by various indoor positioning methods utilizing Euclidean distance alongside k-NN or WKNN algorithms.

Our project’s performance using 25 nearest neighbors achieved competitive results within the context of a commercial mall environment. Meanwhile, Zhu's [Article 4] optimized approach with k=3 yielded the lowest error rate across studies, indicating the potential for enhancing accuracy through algorithmic refinement. The findings from Moreira et al. [Article 5] and Torres-Sospedra et al. [Article 2] provide insights into the performance of WKNN in more extensive and varied environments, while Liu et al. [Article 7] demonstrate the potential for clustering techniques to contribute to precision and computational efficiency. Each study’s outcomes reflect the trade-offs and optimization considerations inherent in indoor positioning system design.

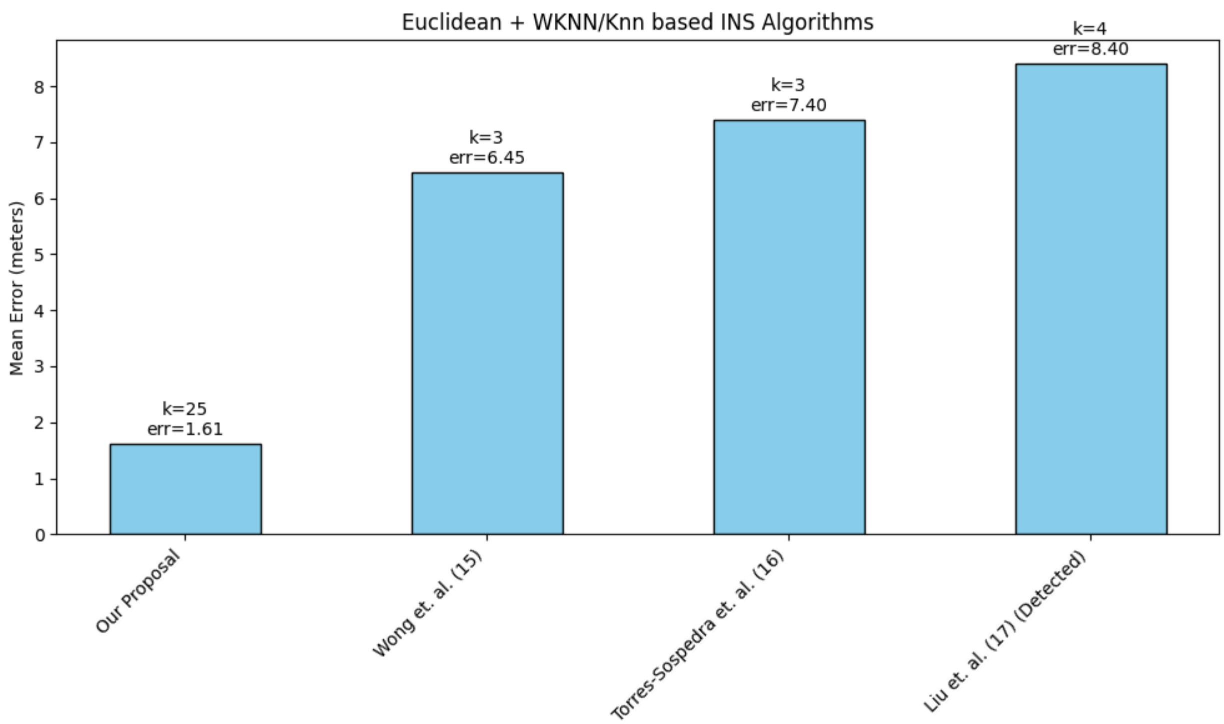

Figure 12: This graph compares the error rates achieved by various indoor positioning methods utilizing Euclidean distance alongside k-NN or WKNN algorithms.

Comparative Effectiveness

Considering our project's use of basic Euclidean distance with WKNN, along with scoring system and RSS Offset enhancements, we have managed to achieve a competitive edge.

In contrast to Moreira et al. [Article 5], who achieved greater accuracy in a larger-scale environment using WKNN with k=3, our approach has proven more accurate in a single-floor setting with a higher k value.

Challenges and Future Work

The fidelity of indoor positioning systems relies heavily on the quality and quantity of Reference Points (RPs) and observed RSS in the environment. Dynamic changes in APs due to the addition of new hardware or removal of existing ones, as well as alterations in indoor infrastructure, pose challenges that can affect the positioning accuracy. While our system has demonstrated robustness in a dynamic commercial mall environment, these external factors underline the need for ongoing system updates and database maintenance to preserve the accuracy and reliability of the positioning method.

8. Conclusions

In this analysis, we have explored various methodologies for indoor positioning using Wi-Fi fingerprinting, comparing the approaches and outcomes from studies with our own project's results. While our approach did not achieve the lowest error rates with the minimum number of k-nearest neighbors (k-NN) compared to some Bayesian filtering techniques, it significantly enhanced the traditional Euclidean distance model using weighted k-NN and additional scoring mechanisms.

The results clearly indicate that while traditional methods provide a baseline for indoor positioning, the integration of advanced algorithms and thoughtful adjustments to handle specific environmental challenges can lead to substantial improvements. For instance, our method achieved an error rate of 1.613164 meters, which is considerably precise for the challenging environment of a commercial mall, showcasing the effectiveness of our tailored approach.

This study shows the different techniques available and also paves the way for future innovations in indoor positioning. By continuing to refine these methods and exploring new adaptations, such as integrating machine learning algorithms for dynamic adjustment and further optimizing the k-NN selection process, we can continue to push the boundaries of what is possible in indoor positioning technology.

On a more personal note, the world of internal navigation left us with some question marks, how can the algorithm be optimized so that it can work dynamically with changes in the structure of routers. Our main goal was to show that it is possible to find the current location inside a closed structure without the help of additional hardware other than routers and indeed we were able to show this by applying one of the basic principles in the field and improving its level of accuracy through information pre-processing. We learned a lot about the world of navigation and we started implementing our research in the form of an app for mobile devices, so that everyone in the future can enjoy navigating inside closed places without wasting unnecessary time searching for destinations.

9. References

[1] Ezhumalai, B., Song, M., & Park, K. (2021). An efficient indoor positioning method based on Wi-Fi RSS fingerprint and classification algorithm. *Sensors, 21(10)*, 3418.

[2] Torres-Sospedra, J., Montoliu, R., Trilles, S., Belmonte, Ó., & Huerta, J. (2015). Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems. *Expert Systems with Applications, 42(23)*, 9263-9278.

[3] Teuber, Andreas & Eissfeller, Bernd. (2006). WLAN indoor positioning based on Euclidean distances and fuzzy logic.

[4] Zhu, Xuanyu. (2020). Indoor Localization Based on Optimized KNN. Network and Communication Technologies. 5. 34. 10.5539/nct.v5n2p34.

[5] Moreira, Adriano & Nicolau, Maria & Meneses, Filipe & Costa, António. (2015). Wi-Fi fingerprinting in the real world - RTLS@UM at the EvAAL competition. 1-10. 10.1109/IPIN.2015.7346967.

[6] Arigye W, Pu Q, Zhou M, Khalid W, Tahir MJ. RSSI Fingerprint Height Based Empirical Model Prediction for Smart Indoor Localization. Sensors (Basel). 2022 Nov 22;22(23):9054. doi: 10.3390/s22239054.

[7] S. LIU, R. DE LACERDA and J. FIORINA, "WKNN indoor Wi-Fi localization method using k-means clustering based radio mapping," 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), Helsinki, Finland, 2021, pp.

[8] Babalola OP, Balyan V. WiFi Fingerprinting Indoor Localization Based on Dynamic Mode Decomposition Feature Selection with Hidden Markov Model. Sensors. 2021;21(20):6778.

[9] Chen, Pan & Shang, Jianga & Gu, Fuqiang. (2020). Learning RSSI Feature via Ranking Model for Wi-Fi Fingerprinting Localization. IEEE Transactions on Vehicular Technology.

[10] Dai, Jihan, Maoyi Wang, Bochun Wu, Jiajie Shen, and Xin Wang. 2023. "A Survey of Latest Wi-Fi Assisted Indoor Positioning on Different Principles" Sensors 23, no. 18: 7961.