Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

A dynamic K-nearest neighbor method based on strong access point credibility for indoor positioning

Yuting YANG[1], Tao ZHANG[1], Wu HUANG[‡2]

[1]*Research institute, Chengdu Techman Software Co.,Ltd, Chengdu 644000, China*

[2]*Department of Software Engineering, Sichuan University, Chengdu 610044, China*

E-mail: yytxwzj@163.com; zhangtao@alu.uestc.edu.cn; huangwu@scu.edu.cn

Received May 7, 2024; Revision accepted Jan. 19, 2025; Crosschecked

Unedited

Abstract: High-precision indoor positioning offers valuable information support for various services such as patient monitoring, equipment scheduling management, and laboratory safety. An traditional indoor positioning technology, fingerprint indoor positioning, often employs the K-nearest neighbors (KNN) algorithm to identify the closest K reference points (RPs) via received signal strength (RSS) for location prediction. However, RSS is susceptible to environmental interference, leading to the selection of RPs that are not physically the closest to the user. Moreover, using a fixed K value is not the optimal strategy. In this work, we propose a novel approach, the dynamic K-nearest neighbor method based on strong access point (AP) credibility (SAPC-DKNN), for indoor positioning. In SAPC-DKNN, we leverage prior knowledge of RSS path loss and employ the RSS fluctuation area to quantify the significance of different APs. We integrate the similarity of AP sets within the range of strong APs and formulate a weighted distance metric for RSS based on the credibility of strong APs. Additionally, we introduce a dynamic K value algorithm (ND-DKA) based on neighbor density for the automatic optimization of the K value for each test point. Experimental evaluations conducted on three datasets demonstrate that our method significantly reduces the average positioning error by 18.67%-66.48% compared to the state-of-the-art KNN methods.

Key words: RSS path loss; Indoor fingerprint positioning; Dynamic K-nearest neighbor
https://doi.org/10.1631/FITEE.2400366 **CLC number:**

1 Introduction

The rapid progress of wireless communication technology has led to increased attention concerning location-based services that rely on positioning technology (Gu et al., 2009; Liu et al., 2007; Song et al., 2019). These services are ubiquitous in daily life; are found in places like roads, playgrounds, underground parking lots, shopping malls, laboratories, and classrooms; and play a vital role in supporting travel navigation, equipment scheduling, and ensuring laboratory and personnel safety. Although the Global Navigation Satellite System (GNSS) excels in outdoor environments (Hegarty and Chatre, 2008), its signals are influenced by factors such as geomagnetism, weather, and building obstructions in indoor settings, which cause signal distortion and hinder the provision of accurate and reliable positioning information (Nguyen et al., 2024; Zhao et al., 2023). In the current landscape, effective indoor positioning is attainable through technologies like Wi-Fi (Yang and Shao, 2015; Cha and Lim, 2022), Bluetooth (Liu et al., 2013; Pu and You, 2018), ZigBee (Dong et al., 2019), radio frequency identification (RFID) (Chon et al., 2004; Ni et al., 2003), and ultra-wideband (UWB) (Xianjia et al., 2021). Wi-Fi, in particular, stands out due to its low power consumption, high real-time performance, low deployment cost, and robust environmental adaptability, making it popular for wide application in indoor positioning and the

‡ Corresponding author
© Zhejiang University Press 2025

Internet of Things (Salamah et al., 2016; Ma et al., 2008; Dong et al., 2022). Due to the widespread deployment of Wi-Fi infrastructure in indoor settings, indoor positioning methods based on Wi-Fi are commonly categorized into two types: ranging-based and fingerprint-based (Rusli et al., 2016; Nguyen et al., 2023). In ranging-based methods, the received signal strength (RSS) value is typically employed to compute the distance between the test point (TP) and the access point (AP) using the RSS propagation loss model (Ciurana et al., 2007; Dag and Arsan, 2018). However, these methods are susceptible to environmental variations and lead to increased estimation errors due to changing obstacles. In contrast, the fingerprint positioning algorithm does not determine the target location by estimating the distance between the TP and the AP. Instead, it estimates the location by assessing the RSS distance between the TP and the offline-collected database (Mingzhe et al., 2015; Alitaleshi et al., 2023). During the offline phase, the fingerprint database is created and stores the RSSI and corresponding coordinates of all APs collected at each reference point (RP) using mobile devices. In the online phase of position estimation, the mobile device gathers the user's current location information, and the RP information in the database is used to estimate the location of the target point with the highest probability. The appeal of the fingerprint-based algorithm lies in its capacity to enhance positioning accuracy without incurring additional hardware costs, making it a subject of interest for many scholars (He and Chan, 2015; Ren et al., 2023; Ayinla et al., 2024).

Numerous fingerprint-based indoor positioning methods, including the K-nearest neighbor (KNN) algorithm (Sadowski et al., 2020; Chen et al., 2022), deep learning (Nabati and Ghorashi, 2023; Arslantas and Okdem, 2024), Bayesian probability method (Madigan et al., 2005), support vector machine (Brunato and Battiti, 2005), and clustering (Le et al., 2021; Lin et al., 2023) have been proposed. The KNN method, compared with deep learning and SVM methods, demonstrates acceptable positioning accuracy with a simple structure and low computational complexity. As a result, our research focuses on methods based on KNN. Many improved KNN approaches have been developed to enhance positioning accuracy (Lin et al., 2023). These enhancements involve changes to the RSS distance measurement method and the conversion of the traditional Euclidean distance (ED) to alternative metrics, such as cosine distance and Spearman distance (Xie et al., 2016; Peng et al., 2016). However, these improved methods do not consider the positioning errors caused by features. Subsequently, Zou et al. (2017) proposed the use of the signal trend index as a new weighting scheme for WKNN, which improved positioning accuracy. Nevertheless, an exponential relationship exists between the RSS and physical distance, and relying solely on the RSS similarity cannot accurately reflect the physical distance between the TP and RP (Zhang et al., 2022). Addressing this, Hu et al. (2015) suggested using the Jaccard coefficient to calculate AP set similarity as a weight, effectively mitigating the physical estimation errors caused by RSS instability. However, they overlooked the sequence of AP strength and the relationship between the RSS and physical distance.

In the fingerprint positioning method grounded in KNN, many researchers adopt a uniform K value across all test samples to derive prediction outcomes directly. For instance, Li et al. (2015) employed a K value of 2, whereas Hu et al. (2015) opted for 3. Discrepancies like these primarily stem from varying network environments. However, such static KNN approaches might overlook opportunities to enhance accuracy through flexible adjustment of the K value. Thus, several scholars have delved into determining the optimal K through extensive experimentation. For instance, Ma et al. (2015) identified the optimal K as 4, whereas Torres-Sospedra et al. (2015) found it to be 13. Nevertheless, this approach demands substantial human effort, and further enhancements in accuracy are warranted. Studies have revealed that customizing the K value for individual test samples can lead to enhanced positioning performance (Umair et al., 2014). Consequently, certain scholars establish a threshold criterion, incorporating only new neighbors if the distance between the user's RSS vector and the neighbor point's fingerprint falls below the threshold (Hu et al., 2018a,b). Moreover, constraints based on physical distance are imposed (Costa et al., 2006; Lee et al., 2016). However, determining these thresholds also necessitates extensive experimentation.

Existing enhanced KNN methods (Zhang and Mao, 2022; Hoang et al., 2018) encounter certain challenges:

Unedited

- The conventional weighted KNN method overlooks the prior knowledge of RSS path loss and the significance of features at the AP level. It relies solely on RSS similarity to measure physical distance, which is suboptimal. Because RSS weakens over time and distance, the physical distance between the RP and AP determines the noise level of the currently received signal. Signal fluctuation varies for RPs at different locations, and the degree of fluctuation is inconsistent. Stronger signals received by the AP result in less noise. Hence, emphasis should be placed on the order and similarity of strong APs. Consequently, relying solely on RSS similarity during the position estimation stage leads to larger errors.

- The performance of KNN methods is influenced by the choice of the K value (Oh and Kim, 2018). When the K value is small, the positioning result is susceptible to interference from noise points, leading to increased errors. Conversely, when the K value is large, adjacent points may encompass numerous points of different categories, also contributing to error escalation. Hence, conventional static KNN methods may overlook opportunities for enhancing accuracy by adjusting the K value. Empirical assignment of K values fails to adequately address actual requirements across diverse network environments. Manually determining the optimal K value for individual samples proves challenging and resource-intensive. Moreover, many dynamic KNN approaches entail additional threshold settings, necessitating extensive experimentation for acquisition.

To overcome the challenges mentioned above, we propose a dynamic K-nearest neighbor (DKNN) method based on strong AP credibility for indoor positioning (SAPC-DKNN). This method leverages prior knowledge of RSS path loss, uses the relationship between the physical distance between the RP and AP and the RSS, quantifies the credibility of different RSS values, and integrates the similarity of strong AP levels. Notably, SAPC-DKNN eschews fixed K values or the design of thresholds for dynamic K adjustment; instead, it adaptively selects the RP based on improved RSS distance measurement results and the concept of neighbor density. In contrast to prior studies (Zhang et al., 2022; Hu et al., 2015), SAPC-DKNN directly incorporates prior knowledge of RSS attenuation, theoretically quantifies the credibility of different RSS values, and considers two key aspects of the strong AP level: order and similarity. This method exhibits strong robustness and does not require the collection of challenging-to-obtain information such as the user's walking speed, direction, and time. Achieving favorable positioning results across various environments is feasible using only the collected RSS data.

2 Related work

The nearest neighbor (NN) algorithm is widely applied in indoor positioning. Bahl and Padmanabhan (2000) introduced the RADAR system, utilizing wireless LAN for indoor positioning and employing the NN algorithm to determine the coordinates of the target point. Experiments substantiate that this method achieves higher positioning accuracy than the model propagation method under similar conditions. Following this work, Wu et al. (2009) investigated two memoryless positioning techniques—KNN and naive Bayes and compared them with the trilateration method across various scenarios. Experimental results indicate that KNN is the simplest and most accurate positioning method. However, these methods may not consistently yield optimal results. Thus, some enhanced KNN algorithms were proposed to improve indoor positioning effectiveness (Chen et al., 2022).

Furthermore, some studies have investigated improved methods for RSS distance measurement. Xie et al. (2016) introduced the use of Spearman distance to refine the KNN algorithm, demonstrating improved positioning performance through simulation experiments. Peng et al. (2016) proposed the Sim-KNN and IW-KNN methods, with the Sim-KNN method incorporating both ED and cosine similarity to gauge fingerprint similarity. There are also weighting methods applied at the RSS value level; for instance, Zou et al. (2017) used the RSS trend index as a new weighting scheme for weighted KNN, resulting in enhanced positioning accuracy. Subsequently, some scholars integrated historical information into KNN, leading to a notable improvement in positioning accuracy (Zhang et al., 2022; Hoang et al., 2018). Although experimental results reveal

Unedited

that these methods surpass traditional KNN in terms of positioning accuracy, they fall short of optimal performance because they neglect prior knowledge of RSS attenuation and AP-level information.

Liu et al. (2022) observed that adjusting the parameter K can markedly enhance positioning performance. In comparison to the NN method, varying K for different TPs can yield a positioning performance improvement of more than 45%. Subsequently, scholars explored algorithmic research on K value selection in KNN. For instance, Oh and Kim (2018) introduced an adaptive KNN algorithm, adjusting the K value based on the ED between the target position and the fingerprint database. By considering only RPs within a specific ED range, positioning accuracy improved by more than 30%. Umair et al. (2014) proposed a dynamic KNN method that sorts based on the difference in ED and sets a threshold to derive a dynamic K value, which proved more accurate than the traditional method. Costa et al. (2006) presented an adaptive approach for establishing thresholds in neighbor selection within sensor networks. However, their consideration was limited to the distance between sensors, and they found that a two-stage selection algorithm for neighboring sensors was inadequate. Consequently, Lee et al. (2016) introduced the Dk-NN method, which determines the optimal K value by accounting for changes in topology and distance among NNs.

Although the aforementioned dynamic KNN methods significantly reduce positioning errors, their approach requires the introduction of new manually adjusted parameters, which increases labor costs and optimization difficulties. In this work, we propose a novel improved dynamic KNN indoor positioning method, integrating prior knowledge of RSS attenuation and AP-level information to achieve a more accurate prediction goal.

3 Our method

Our method is divided into two parts: (1) strong AP weighted distance measurement method and (2) density-based dynamic K value algorithm.

3.1 Strong AP weighted distance measurement method

The path loss model for estimating RSS is characterized using a logarithmic attenuation model (Latif and Parasuraman, 2022), that is:

$$RSS = RSS_0 - 10nlog_{10}(d) \tag{1}$$

where RSS_0 is the RSS at 1 m from the AP, n is the path attenuation coefficient, ranging from 2 (free space) to 6 (complex indoor environment) (Latif and Parasuraman, 2022), and d is the distance from the RP to the AP.

When there are two RPs at the same location but at different times, the RSS from the same AP is denoted as RSS_1 and RSS_2. The RSS change, ΔRSS, is then represented as

$$\Delta RSS = RSS_1 - RSS_2 = 10log_{10}(d)(n_1 - n_2) \tag{2}$$

where $(n_1 - n_2)$ consistently maintains a specific distribution over time, whereas ΔRSS and d exhibit a positive correlation (Jin et al., 2015). Consequently, when a particular RP is distant from a specific AP, a smaller RSS corresponds to larger d and ΔRSS. These parameters indicate heightened signal fluctuations and increased noise at that location, diminishing the credibility of the RSS signal under the corresponding AP. Additionally, scholars have highlighted that RSS similarity inadequately reflects the proximity of locations (Zhang et al., 2022). Hence, relying solely on RSS measurements for distance assessment may not accurately depict the physical closeness of two points. This underscores the importance of prioritizing attention on the AP with robust signals.

We introduced a method that employs the RSS fluctuation area as AP credibility to account for the significance of the RSS corresponding to a single AP. If the strongest signal RSS_{stro} at the current RP is identified, the maximum and minimum distances (d_{stro}^{max} and d_{stro}^{min}) of the RP from the corresponding AP can be determined by Eq. (1):

$$d_{stro}^{min} = 10^{\frac{RSS_0 - RSS_{stro}}{10n_{max}}} \tag{3}$$

$$d_{stro}^{max} = 10^{\frac{RSS_0 - RSS_{stro}}{10n_{min}}} \tag{4}$$

where RSS_0 is a constant value, and we designate the maximum RSS in the database as RSS_0. $n_{min} = 2$, $n_{max} = 6$. By using d_{stro}^{min} and d_{stro}^{max} as radii, the fluctuation space S_{stro} of RSS_{stro} can be derived (represented by the area of the yellow circle in Fig. 1):

$$S_{stro} = \pi({d_{stro}^{max}}^2 - {d_{stro}^{min}}^2) \tag{5}$$

Unedited

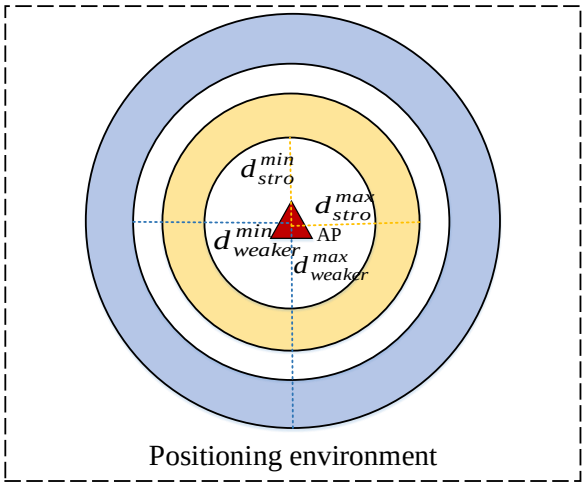


Fig. 1 RSS fluctuation range under strong and weak AP. d_{stro}^{min} and d_{stro}^{max} (yellow dotted lines) are the radii of the inner and outer circles of the yellow ring, representing the minimum and maximum distances from the RP of the strong signal (RSS_{stro}) to the AP. d_{weak}^{min} and d_{weak}^{max} (blue dotted lines) are the radii of the inner and outer circles of the blue ring, representing the minimum and maximum distances from the RP of the weak signal (RSS_{weak}) to the AP. AP, access point; RP, reference point; RSS, received signal strength.

As illustrated in Fig. 1 and according to Eqs (3) and (4), the distance from RSS_{weak}, weaker than RSS_{stro}, to the AP can be derived. The area Sweak (in blue) of the ring bounded by it can be determined from Eq. (5). Comparing the areas of the blue and yellow rings ($S_{weak} > S_{stro}$), it is evident that as n remains constant, larger RSS values correspond to smaller RSS fluctuation spaces for the respective APs, indicating higher credibility of the AP.

The traditional KNN method frequently uses the ED of the RSS to ascertain the physical proximity between two points. Nevertheless, it is acknowledged that RSS similarity alone may not completely convey physical distance (Zhang et al., 2022). In response to this limitation, certain researchers have suggested employing the Jaccard coefficient to assess the difference in AP sets for estimating physical location Hu et al. (2015). Motivated by this insight, we assembled a set comprising corresponding APs with the strongest signals and subsequently performed set similarity calculations between the points to be measured and RPs. Diverging from previous work Hu et al. (2015), which exclusively considered APs with signals, our approach incorporates the order of RSS signal strength. Given the RSSs for a TP and RP, along with their AP IDs, we sort the RSSs. This sorting process arranges the RSSs from strong to weak, accompanied by their corresponding AP (IDs) sequences denoted as $\boldsymbol{L}_{tp}$ and $\boldsymbol{L}_{rp}$:

$$L_{tp} = \begin{bmatrix} RSS_1^{tp} & RSS_2^{tp} & ... & RSS_M^{tp} \\ AP_1^{tp} & AP_2^{tp} & ... & AP_M^{tp} \end{bmatrix} \tag{6}$$

$$L_{rp} == \begin{bmatrix} RSS_1^{rp} & RSS_2^{rp} & ... & RSS_M^{rp} \\ AP_1^{rp} & AP_2^{rp} & ... & AP_M^{rp} \end{bmatrix} \tag{7}$$

where M is the number of APs. We take the first N_t APs in L_{tp} to form a set, perform similarity matching with the first N_r strong APs in L_{rp}, and obtain the AP set similarity vector $\boldsymbol{\alpha}$:

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \alpha_2 & ... & \alpha_N \end{bmatrix} \tag{8}$$

$$\alpha_i = \begin{cases} 1 & \{AP_i^{rp}\} \in \{AP_{start}^{tp}, ..., AP_{end}^{tp}\} \\ 0.0001 & \{AP_i^{rp}\} \notin \{AP_{start}^{tp}, ..., AP_{end}^{tp}\} \end{cases} \tag{9}$$

$$start = \begin{cases} i - \lfloor (N_t - 1)/2 \rfloor, & i \neq 1 \\ 1, & i = 1 \end{cases} \tag{10}$$

$$end = i + \lfloor (N_t - 1)/2 \rfloor \tag{11}$$

where $i \in \{1, 2, ..., N_r\}$, N_t and N_r denote the size of the RSS fault tolerance space and the scope of focusing on strong APs, respectively. N_r is acquired from the database and $N_t = 3$ is a fixed coefficient, which will be further discussed in subsequent sections.

At this time, the RSS fluctuation space S_i corresponding to the first N_r strong APs can be obtained using Eq. (5), and after normalization, the RSS credibility vector is formed:

$$\boldsymbol{\lambda} = [\lambda_1, \lambda_2, ..., \lambda_{N_r}] \tag{12}$$

$$\lambda_i = \frac{exp(-S_i)}{\sum_{i=1}^{N_r} exp(-S_i)} \tag{13}$$

We update the weighted ED (Zhang et al., 2022) of the RSS between the current TP $\boldsymbol{L}_{tp}$ and RP $\boldsymbol{L}_{rp}$ to be measured as:

$$D_{SAPC} = \frac{1}{\boldsymbol{\lambda}\boldsymbol{\alpha}^T} \sqrt{\sum_{m=1}^{M} \omega_m (RSS_m^{tp} - RSS_m^{rp})^2} \tag{14}$$

Unedited

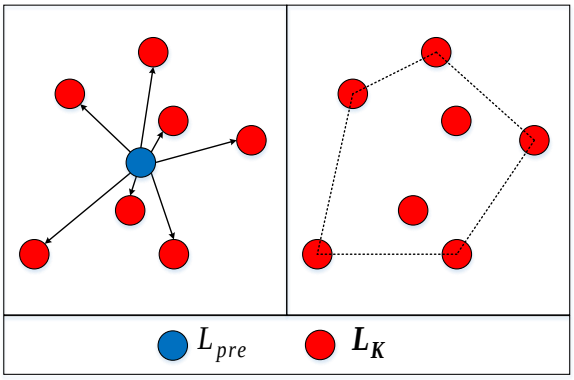


Fig. 2 Example diagram of ND-DKA calculation approaches.

$$\omega_m = \frac{W_m^{rp} W_m^{tp}}{\sum_{m=1}^{M} W_m^{rp} W_m^{tp}} \tag{15}$$

$$W_m^{rp} = (1/RSS_m^{rp})^2 \tag{16}$$

$$W_m^{tp} = (1/RSS_m^{tp})^2 \tag{17}$$

where M is the number of APs. Ultimately, the preliminary positioning result L_{pre} is obtained by identifying the weighted coordinates L_k corresponding to the K minimum EDs D_{SAPC}^k:

$$L_{inlP}^K = \frac{\sum_{k=1}^{K} \frac{1}{D_{SAPC}^k} * L_k}{\sum_{k=1}^{K} \frac{1}{D_{SAPC}^k}} \tag{18}$$

3.2 Dynamic K-value algorithm based on neighbor density

In the original KNN positioning method, determining the optimal k value emerges as a pivotal challenge. Ni et al. (2003) addressed this by selecting the k value through experimentation. They explored various k values, specifically $k = 2, 3, 4, 5$, to assess position estimation errors. The results indicate that, on average, the optimal effect is observed at $k = 4$. However, obtaining this empirical value demands considerable effort and may lack generalizability to other datasets. Additionally, prior studies have demonstrated that adjusting parameter K for different TPs holds potential for improving localization performance (Umair et al., 2014; Lee et al., 2016; Oh and Kim, 2018). Therefore, for each individual TP across diverse datasets, it is essential and effective to determine the optimal K value for predicting each sample. Lee et al. (2016) proposed adjusting the optimal K value based on the topology and distance changes of its NNs. Drawing inspiration from this approach, we introduce a density-based K value selection algorithm, named ND-DKA (A dynamic K value Algorithm based on Neighbor Density). In contrast to the method by Lee et al. (2016), ND-DKA does not require experiential threshold setting. Instead, we prioritize the proximity of neighbors to the TP and the density of neighbors during NN selection. ND-DKA employs two criteria for evaluating K values: (1) If the selected K RPs are near the predicted position, and (2) the K RPs are close to each other, then K is deemed optimal.

A. The proximity of the neighbor to TP: quantify the proximity of K RPs to their weighted geometric centers.

If the K RPs $\boldsymbol{L}_K = \{L_1(x_1, y_1), L_2(x_2, y_2), ..., L_k(x_k, y_k)\}$ are known, the preliminary predicted position $L_{inlP}^K(x_{pre}^K, y_{pre}^K)$ is obtained. As illustrated in Fig. 2 (left), to quantify the proximity between $\boldsymbol{L}_K$ and L_{pre}, the initial step involves calculating the physical distance between L_{pre} and $\boldsymbol{L}_K$.

$$dis_k = \sqrt{(x_{pre} - x_i)^2 + (y_{pre} - y_i)^2} \tag{19}$$

From this, we can get the closeness of K RPs to the predicted position, that is

$$\beta_K = \frac{1}{K} \sum_{k=1}^{K} (dis_k - \frac{1}{K} \sum_{k=1}^{K} dis_k)^2 \tag{20}$$

B. The density of neighbors: quantify the proximity of K RPs to each other.

As depicted in Fig. 2 (right), K RPs can encompass a minimal area, known as the convex hull area. This area reflects the density of RP distribution. An increase in the K value may result in the expansion of the convex hull area of RPs and potentially fail to meet the evaluation standards. Therefore, we use the unit convex hull area as an indicator of whether RPs are close or far apart. Initially, the Graham scan algorithm is employed to calculate the convex hull of the a1 point set, followed by the determination of the convex hull area h_K. The algorithm flow is outlined in Algorithm 1.

Unedited

The ultimate criterion for evaluating the K value is the fusion of both methods, forming an assessment sequence U across various K values:

$$U = \{u_{K_{min}}, u_K, ..., u_{K_{max}}\} \tag{21}$$

$$u_K = \frac{\beta_K}{\sum_{K=K_{min}}^{K_{max}} \beta_K} + \frac{h_K}{\sum_{K=K_{min}}^{K_{max}} h_K} \tag{22}$$

where, K_{min}=3, K_{max}=30 are empirical values.

Algorithm 1 Convex hull area algorithm

Require: The initial selection of the first K RP sets $\boldsymbol{L}_K = \{L_1(x_1, y_1), L_2(x_2, y_2), ..., L_k(x_k, y_k)\}$ determined by (18).

Ensure: Convex hull area h_K

1: Select the starting point: select the coordinate with the smallest y coordinate (if there are multiple points with the smallest y coordinate, select the leftmost one) as P_1.

2: Polar angle sorting: calculate the polar angle A_i of each point P_i relative to P_1 except P_1, and sort the remaining points in ascending order of polar angle:

$$A_i = atan2(y_i - y_1, x_i - x_1)$$

3: Initialize the stack: put P_1 and P_2 into the empty stack S in sequence.

4: Scan and build convex hull:

5: **FOR**{j=3: length($\boldsymbol{L}_K$)}

6: Calculate the cross product $B_j = (x_{P_2} - x_{P_1})(y_{P_j} - y_{P_1}) - (y_{P_2} - y_{P_1})(x_{P_j} - x_{P_1})$

7: **While** $B_j < 0$ and the number of elements in stack S is > 1

8: Pop the top element of the stack

9: **End While**

10: Push point P_j onto stack S

11: **End FOR**

12: Enter $S = \{(x_1, y_1), ..., (x_n, y_n)\}$ to calculate the convex hull area:

$$h_K = \frac{1}{2}|\sum_{j=1}^{n}(x_j y_{j+1} - x_{j+1} y_j)|/K$$

13: **return** h_K

Ultimately, the K value corresponding to the minimum value in U is chosen as the optimal K value, and its predicted position serves as the final positioning outcome:

$$L_{finP} = L_{inlP}^K(x_{pre}^K, y_{pre}^K)_{min(U)} \tag{23}$$

4 Experiment

4.1 Dataset description

We chose four benchmark Wi-Fi fingerprint datasets, namely Tampere (Lohan et al., 2017), UJIIndoorLoc (Torres-Sospedra et al., 2014), GeotecDatabase (Torres-Sospedra et al., 2016), and SODIndoorLoc (Bi et al., 2022), to assess the efficacy of our method. Prior to conducting the experiments, the dataset underwent preprocessing, retaining only the necessary data for the experiment. The retained data include signal strength and real-world coordinates (longitude and latitude) of the APs at all RPs.

Tampere. The dataset was gathered from a four-story university building situated in Tampere, Finland, spanning the period from January to August 2017. It encompasses 687 training fingerprints and 3951 test or estimated fingerprints. The collection was conducted in a crowd-sourced manner, involving 21 distinct devices and users. Coordinates are meticulously recorded in three dimensions, with each sample comprising 992 AP signals.

UJIIndoorLoc. The dataset was released by Jaume I University in Spain and encompasses multiple adjacent buildings and floors. Table 1 presents the relevant dataset information. Each record in this dataset includes RSS data from 520 APs (WAP001–WAP520), along with details such as building ID (BUILDINGID), collection device ID (PHONEID), and longitude and latitude (LONGITUDE and LATITUDE) indicating its location.

GeotecDatabase. The dataset was gathered and released by the GEOTEC Laboratory. As depicted in Table 2, each data record includes RSS data from 97 APs (wap1–wap97), along with details such as longitude and latitude.

SODIndoorLoc. The dataset was collected in the vicinity of the south and east sides of floors within corridors, offices, and conference rooms. Considered as a complement to UJIIndoorLoc, they share a consistent format. The stored data are labeled by building numbers, namely CETC331, HCXY, and SYL, with the respective APs containing 53, 347, and 363 RSSs.

Unedited

Table 1 UJIIndoorLoc data set (some examples)

WAP001	WAP002	...	WAP519	WAP520	LONGITUDE	LATITUDE	BUILDINGID	PHONEID
100	100	...	100	100	-7541.2643	4864920.778	1	23
100	100	...	100	100	-7536.6212	4864934.225	1	23
...	...	...	...	...	...	...	...	...
100	100	...	100	100	-7536.1658	4864897.859	1	10

Table 2 GeotecDatabase dataset (some examples)

wap1	wap2	...	wap96	wap97	lon	lat	TIMESTAMP
100	100	...	-57	-53	-8220.54659	4865048.064	1450452282
100	100	...	-55	-56	-8220.54659	4865048.064	1450452285
...	...	...	...	...	...	...	...
100	100	...	-72	-58	-8226.447664	4865062.628	1450786990

4.2 Performance of SAPC-DKNN

To evaluate the efficacy of our method in a real-world localization setting, we partitioned the experiment into two distinct phases.

A. Explore the positioning effect of SAPC-DKNN in a 3D environment

We selected the Tampere dataset for 3D indoor positioning prediction and juxtaposed it against the latest methods. The experimental outcomes are delineated in Table 3. To underscore impartiality, the Baseline method results in Table 3 are exclusively sourced from the original text. Notably, amid the majority of deep learning-based approaches, ours exhibits superior performance. Notably, compared to the optimal deep learning-based method GrowNetLoc, SAPC-DKNN yields a 47% increase in positioning accuracy. This method integrates two distinct networks: a gradient boosting neural network (GrowNet) and a long short-term memory (LSTM) network, effectively navigating complex environments for 3D positioning. Nonetheless, under limited data circumstances, it may not yield anticipated outcomes. Moreover, contrasting with the KNN-based method, RWKNN achieves optimal positioning accuracy at 6.53m, surpassing deep learning-based alternatives. Incorporating indoor movement constraints, it employs search rectangles and trajectory constraints to mitigate spatial ambiguity. In contrast, our proposed approach still outperforms it by 37%, underscoring the efficacy of AP credibility-based methods in mitigating positioning errors arising from RSS instability. In addition, among fingerprint-based positioning methods, the KNN method can achieve acceptable positioning accuracy with a simple structure and low computational complexity.

Table 3 Performance of SAPC-DKNN on Tampere for 3D positioning

Year	Method	Loss (m)
2024	CNN (Arslantas and Okdem, 2024)	8.13
2024	SALLoc (Ayinla et al., 2024)	9.52
2024	eAaT+ (Nguyen et al., 2024)	8.14
2023	GrowNetLoc (Zhao et al., 2023)	7.82
2023	EA-CNN (Alitaleshi et al., 2023)	9.52
2023	bAaT (Nguyen et al., 2023)	8.53
2023	GMM + WKNN (Lin et al., 2023)	11.17
2022	Vanilla LSTM (Dong et al., 2022)	8.66
2022	HADNN (Cha and Lim, 2022)	9.05
2022	RWKNN (Chen et al., 2022)	6.53
2019	CNNLoc (Song et al., 2019)	10.88
	SAPC-DKNN	**4.14**

B. Explore the 2D positioning effect of SAPC-DKNN under the walking trajectory of a single person

The previous section confirmed that, compared with deep learning-based methods, the KNN method can achieve better positioning accuracy with a simple structure and lower complexity. We compared the results of SAPC-DKNN on three data sets with KNN (Sadowski et al., 2020), WKNN (Zhang and Mao, 2022), STI-WKNN (Zou et al., 2017), Spearman-KNN (Xie et al., 2016), SIM-KNN (Peng et al., 2016), and SRL-KNN (Hoang et al., 2018). To verify the performance of our method in a 2D environment, we used actual positioning application habits and divided them according to different users on three public data sets to verify the effect of online indoor positioning in a simulated real environment. The experimental results are shown in Table

Unedited

Table 4 Average positioning error for different methods. The bold values are our methods.

Data	SAPC-DKNN (K∈[3,30])	KNN (K=5)	WKNN (K=5)	STI-WKNN (K=5)	Spearman-KNN (K=5)	SIM-KNN (K=5)	SRL-KNN (K=3)
B0-P0	**5.66±3.61**	9.46±6.23	9.52±6.26	8.70±5.88	16.05±10.95	9.52±6.26	8.26±5.11
B0-P4	**4.18±3.92**	6.87±5.11	6.84±5.08	6.73±4.59	10.97±9.10	6.84±5.08	5.76±4.45
B0-P21	**6.20±5.89**	13.02±11.12	13.00±11.11	13.62±11.10	13.46±9.22	13.00±11.11	15.29±11.55
B1-P4	**6.90±4.27**	10.67±7.22	11.33±7.95	10.72±8.82	16.65±10.75	11.33±7.95	10.15±6.84
B1-P12	**8.01±6.58**	12.38±9.15	12.37±9.12	12.16±9.60	12.83±9.48	12.37±9.12	16.49±20.20
B1-P13	**7.70±7.22**	11.11±8.82	11.04±8.88	10.51±9.37	12.55±10.36	11.04±8.88	11.34±10.27
Val-1	**4.56±2.20**	5.87±2.69	5.87±2.72	5.87±2.87	5.39±2.32	5.87±2.73	5.57±2.64
Val-2	**4.68±1.97**	5.73±2.70	5.73±2.70	5.76±2.80	5.87±3.07	5.72±2.75	5.74±2.77
Val-3	**4.52±1.90**	6.43±3.02	6.45±3.02	6.27±2.96	7.31±3.44	6.44±3.03	6.14±2.89
Val-All	**4.59±2.08**	6.02±2.94	6.03±2.93	6.04±2.91	6.47±3.22	6.04±2.94	5.92±2.79
Testing-CETC331	**2.96±2.63**	4.75±4.31	4.75±4.31	4.66±4.03	4.87±4.05	4.79±4.33	4.72±4.07
Testing-HCXYAll	**2.20±1.93**	3.15±2.99	3.01±2.79	2.84±2.58	4.35±4.17	3.00±2.80	2.98±2.82
Testing-SYL-All	**3.68±5.01**	5.19±5.27	5.18±5.27	4.78±5.25	5.18±5.30	5.08±5.30	5.02±5.30

4. To ensure fair results, each method uses the optimal K value. Specifically, for UJIIndoorLoc, we categorized walking trajectories based on building names and user device IDs. Three random trajectories per building (denoted as B for BUILDINGID and P for PHONEID, with B0–P0 representing the trajectory where BUILDINGID and PHONEID are both 0) were selected for prediction. This allowed us to assess positioning performance in both single-user and multi-user contexts. Across these six test tracks, SAPC-DKNN demonstrated superior performance compared to the other six advanced methods, including a significant performance gap. On B0–P4, we achieved the highest accuracy positioning error of 4.18 m, marking a 61.9% improvement over the least accurate baseline method and a 27.43% enhancement compared to the best-performing SRL-KNN method. Subsequently, B0–P0 exhibited a positioning error of 5.66 m. Notably, B0–P4 and B0–P0 represent short and long trajectories, respectively, underscoring the consistent stability of SAPC-DKNN across trajectory lengths. Across the B0 and B1 trajectories, the positioning accuracy gap consistently ranged between 1 and 2 m, affirming SAPC-DKNN's robust performance in both single-user and multi-user scenarios. Furthermore, our observations highlight Spearman-KNN as the least effective method across all six trajectories. This discrepancy arises, in part, because Spearman-KNN neglects the ranking effect of APs when calculating fingerprint similarity. Instead, it solely considers overall data, potentially resulting in identical fingerprints being associated with different locations. It is crucial to note that our method successfully addresses this challenge and maintains the highest accuracy by leveraging prior knowledge of RSS attenuation performance and exploiting AP sequence information.

The GeotecDatabase test set exhibits temporal continuity, with a few instances featuring significant time intervals, notably 4238 s and 17290 s. We attribute these intervals to inconsistencies in the data collection process, and use them as time nodes to segment the test set into three distinct walking trajectories: Val-1, Val-2, and Val-3. These trajectories, along with the entire test set, are independently subjected to predictions to showcase the real-time positioning efficacy in scenarios involving both continuous walking and intermittent user positioning in authentic environments. Similar to the observed performance on UJIIndoorLoc, SAPC-DKNN demonstrates positioning differences of <0.1 m on Val-1, Val-2, and Val-3, all aligning closely with the results of Val-All. This substantiates the capacity of SAPC-DKNN to excel in both continuous walking and discontinuous positioning settings.

On SODIndoorLoc, we conducted experimental verification using all unprocessed test sets (Testing-CETC331, Testing-HCXY-All, and Testing-SYL-All) across three environments. SRL-KNN, incorporating principles of human walking, ranks second only to SAPC-DKNN in Testing-SYL-All. However, it neglects the impact of RSS fluctuations, leading to biased positioning outcomes. KNN, representing traditional fingerprint positioning methods, exhibits a positioning accuracy 27.68% lower than SAPC-DKNN on Val-3 and 37.68% lower on Testing-CETC331. Although WKNN, SIM-KNN, and STI-WKNN aim to enhance KNN by consid-

Unedited

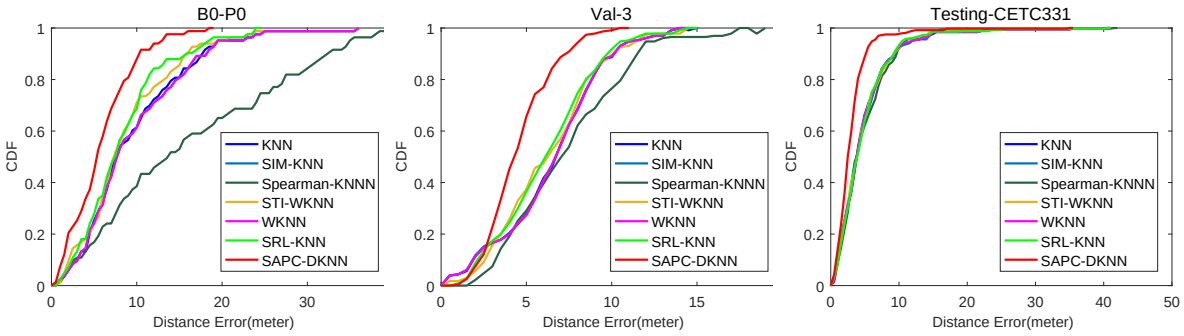


Fig. 3 Comparison of the CDF among seven methods of indoor positioning. CDF, cumulative distribution function, KNN, K-nearest neighbor.

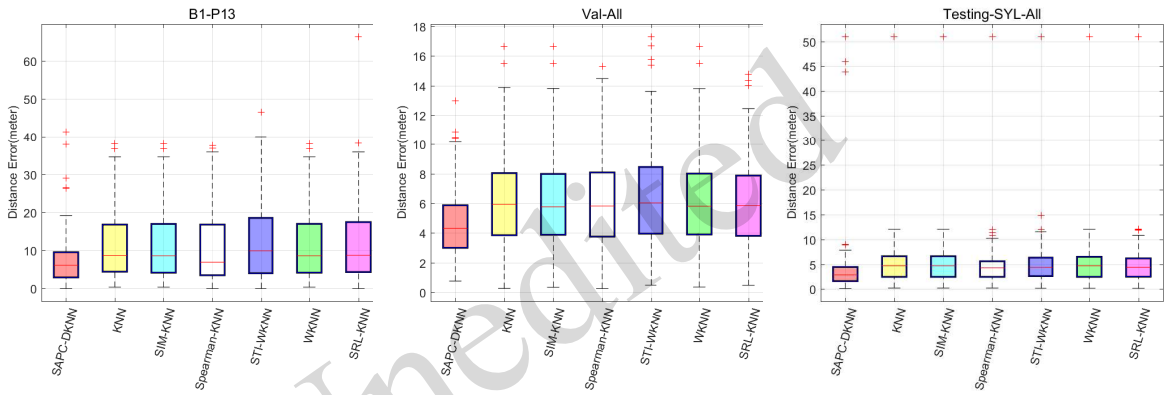


Fig. 4 Comparison of distance errors among seven methods of indoor positioning. KNN, K-nearest neighbor.

Unedited

ering weights and fingerprint distances, their performance improvements across the three datasets are relatively modest, and their generalization performance is average. This suggests that employing different strategies such as assigning weights at distinct RP locations and adopting varied fingerprint calculation methods may not be highly effective for datasets with varying trajectory lengths and environmental backgrounds. Figs. 3 and 4 present comparisons of distance error cumulative distribution function curves and error box plots for seven positioning methods across three datasets, respectively. In Fig. 3, it is evident that SAPC-DKNN consistently exhibits leftward-shifted error curves on all three datasets, with maximum errors consistently smaller than the other six methods. Specifically, on B0–P0, SAPC-DKNN accounts for 80% within the 10-m positioning error range, surpassing SRL-KNN at approximately 60%, and outperforming the remaining methods. Notably, on Val-3 and Testing-CETC331, the SAPC-DKNN curve stands distinct from the six comparison methods, underscoring its superior performance. Fig. 4 illustrates that SAPC-DKNN achieves the lowest positioning errors at the median, maximum, minimum, and upper quartile. In summary, SAPC-DKNN leverages prior knowledge of the RSS decay model, quantifies AP importance, and mitigates RSS uncertainty, resulting in consistently superior accuracy across multiple datasets. These findings affirm the model's good generalization performance and stability.

4.3 Research on parameters

Following the sorting of APs for both the point to be measured and the RP, SAPC-DKNN chooses the Received Signal Strength (RSS) corresponding to N_t and N_r APs for distance measurement. The selection of N_t and N_r significantly influences the ultimate positioning accuracy. Specifically, the value of N_t determines the extent of the RSS fault tolerance space, while N_r dictates the scope of focus on strong APs. Optimal values for N_t and N_r contribute to the optimization of positioning accuracy.

We made an assumption that the value of N_r

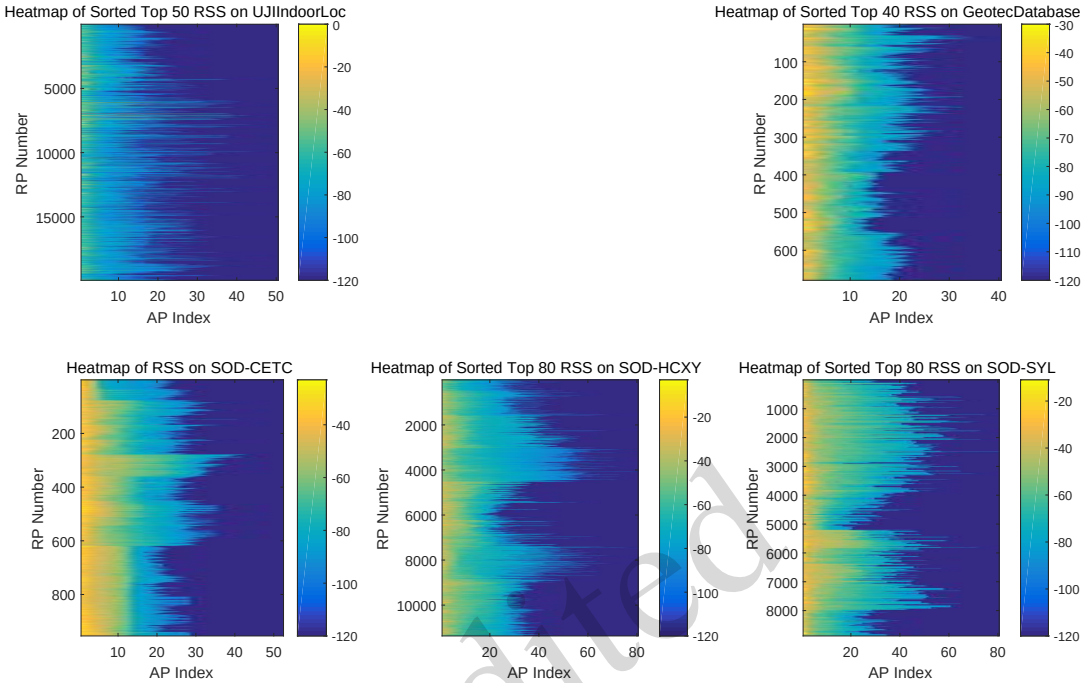


Fig. 5 RSS distribution heat map on three datasets. RP, reference point; RSS, received signal strength.

Unedited

might be associated with the distribution of RSS sizes within the fingerprint database. The chosen N_r should fulfill the requirement of having a sufficient number of strong APs to cover all RPs in the fingerprint database. This selection serves to prevent excessive weak APs extracted from RPs, thereby averting inaccurate matching results during the process of matching RPs with points to be measured. To identify the suitable N_r for various datasets, we visualized the RSS distribution within the fingerprint database across three datasets. By arranging the RSS values corresponding to each RP in ascending order, we generated an RSS heat map 5. As depicted in Fig. 5, distinct N_r values can be selected based on the RSS distribution within the fingerprint database across various datasets. For instance, on the UJIIndoorLoc dataset, the total count of strong APs is approximately 10 ($N_r \approx 10$), whereas on the GeotecDatabase, it amounts to approximately 12 ($N_r \approx 12$). Across different environments within the SODIndoorLoc dataset, the distribution of strong AP counts remains relatively consistent, at approximately 15 ($N_r \approx 15$).

To corroborate our assumptions regarding the value of N_r and to determine the appropriate selection of N_t, we randomly selected two test sets for experiments across three types of datasets. The results, depicted in Fig. 6 and 7, reveal insights. In Fig. 6, with N_t fixed at three, the curves across the six test sets exhibit varying trends. Different data sets exhibit distinct peak values. Notably, on Val-3 and Val-All, there is a conspicuous minimum error, corresponding to N_r values of 15 and 14, respectively. When $N_r \approx 12$, a minor error persists, consistent with our hypothesis. Intriguingly, on both the UJIIndoorLoc and SODIndoorLoc datasets, the error trend tends to plateau. However, when $N_r \approx 10$ and $N_r \approx 15$, respectively, improved values are obtained. When N_r reaches its maximum value, it represents the total number of APs. At this point, both strong and weak APs are included in the algorithm. As shown in Fig. 6, as N_r approaches the total number of APs, the prediction error initially increases, reaches a peak, and then gradually converges. This indicates that incorporating more weak RSS signals when calculating the weights reduces positioning performance. The reason is that the noise in weak RSS signals is significantly larger than the effective signal, causing the useful information to be overwhelmed by noise. Moreover, noise removal is

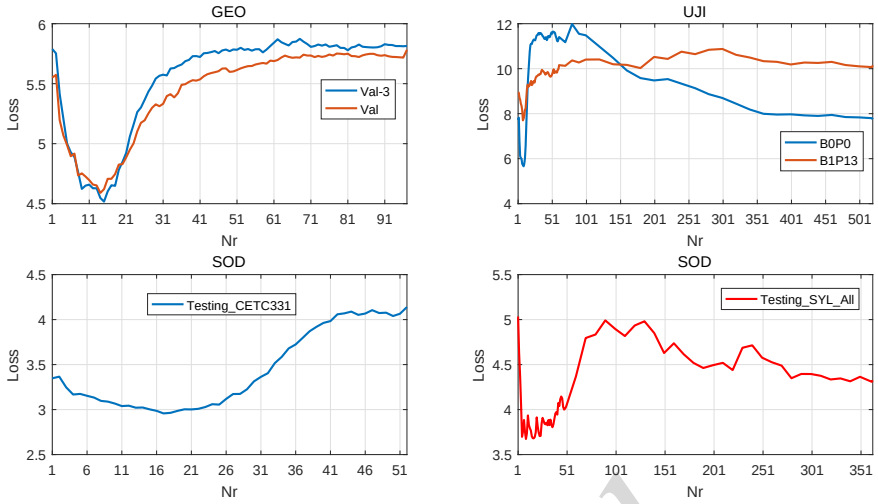


Fig. 6 Distance errors against N_r.

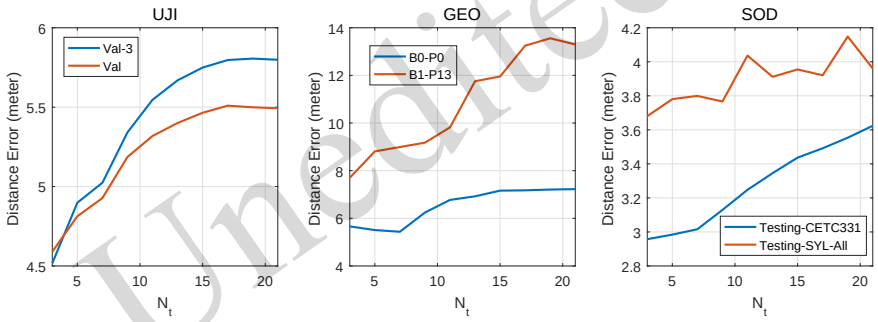
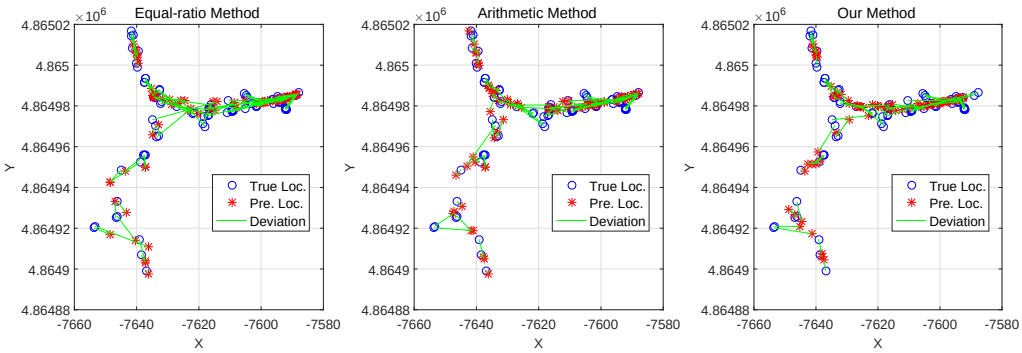


Fig. 7 Distance errors against N_t.

Unedited

challenging due to its complexity. Consequently, selecting too many weak RSS signals is not an optimal strategy. Hence, our assumption regarding N_r is affirmed. By analyzing the RSS distribution within the fingerprint database, we can determine the number of RPs covered by the majority of strong APs, representing the appropriate N_r. The results in Fig. 7 are based on experimental outcomes of the Fig. 6. Following the determination of N_r, errors exhibit an upward trend across the six test sets. When N_t is set to 3, minimal errors are observed for Val-3, Val-All, B1-P13, and Testing-CETC331. Although this trend is not consistent for B0-P0 and Testing-SYL-All, the error variation is not substantial compared to when N_t is set to 3. Consequently, based on our comprehensive experiment, we adopt $N_t = 3$ as the overall model design.

In general, selecting a small N_t limits the space for AP ranking changes induced by RSS fluctuations and overly relies on the magnitude of the RSS value. Conversely, a large N_t increases the RSS fault tolerance space, potentially screening out APs with greater physical distance. This is not conducive to enhancing positioning accuracy. Similar considerations apply to the value of N_r. A narrow focus on strong APs limits the stability of the judgment basis, whereas an overly broad focus introduces weaker APs and interferes with RSS credibility calculations and diminishing positioning accuracy. In real-world application scenarios, the value of N_r may vary across different environments, but it can be determined from the RSS distribution within the fingerprint database. Regarding the value of N_t, experimental findings across three distinct datasets suggest that $N_t = 3$ can be directly applied to real-world scenarios. This confirmation underscores the continued effectiveness of SAPC-DKNN in practical applications.



(a) (b) (c)

Fig. 8 Performance of different quantitative AP credibility methods on B0–P0. Fig. (a), (b) and (c) show the results of using the Arithmetic Method, the Equal Ratio Method, and our method, respectively. AP, access point.

Unedited

4.4 RSS credibility discussion

In our method, AP credibility is depicted by the RSS fluctuation area, which contributes to optimal positioning accuracy. The credibility of each strong AP hinges on its respective RSS value and serves as a key aspect of our proposed approach. However, considering the potential for enhanced accuracy, could we derive improved results by overlooking specific RSS values and instead employing straightforward arithmetic or geometric relationships to assess the strength of the first N_r APs? To explore this, we introduced two alternative methods for quantifying AP credibility:

1. Arithmetic Method:

The arithmetic method establishes the credibility of various strong and weak APs as a vector linked to the AP order and N_r. In other words, it transforms Eq. (13) as follows:

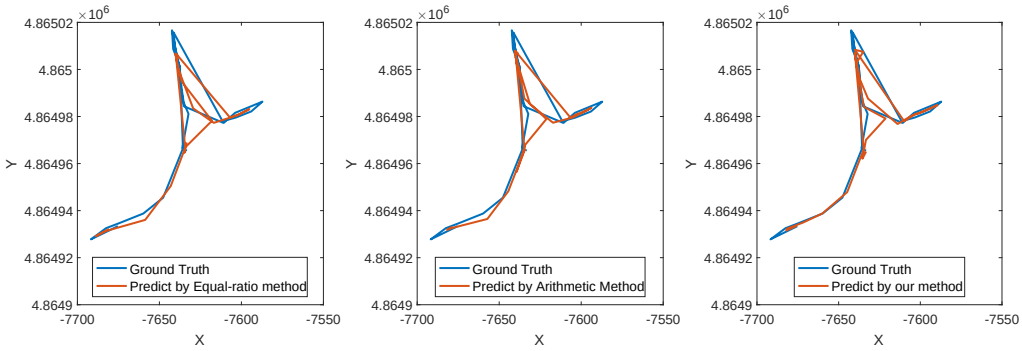
$$\lambda_i = (N_r - i + 1)/N_r \tag{24}$$

2. Equal Ratio Method:

Similar to the arithmetic method, the equal ratio method converts the credibility between strong and weak APs into an equal ratio relationship, transforming Eq. (13) as follows:

$$\lambda_i = 1/(N_r)^i \tag{25}$$

We conducted experimental verification by randomly selecting two test sets from the three datasets, and the results are presented in Table 4. Across the six test sets, the positioning accuracy of the arithmetic method is inferior to our proposed method, and the equal ratio method exhibits lower accuracy than the arithmetic method. This discrepancy arises because the equal difference or proportional relationship is independent of the RSS value corresponding to the AP but is solely linked to the strength order and N_r. Compared with our method, the arithmetic method diminishes the credibility gap between the strongest N_r APs. If the i-th AP, following the strength ranking, significantly differs in strength from the i+1-th AP, the arithmetic method fails to reflect such a gap. The equal ratio method encounters a similar issue, because it cannot dynamically quantify differences between APs with distinct strengths. In contrast, SAPC-DKNN dynamically quantifies AP credibility based on the strength of its corresponding RSS, which is closely tied to the strength of APs. We visualized the deviation between the actual position and the predicted position using the three quantization methods on the B0-P0 test set. Illustrated in Fig. 8, the green line represents the physical distance between the actual and predicted positions and serves as a metric for prediction accuracy. In Fig. 8, from right to left, the area covered by the green lines widens and deepens, affirming the superior positioning accuracy of our method. Furthermore, Fig. 9 depicts the actual and predicted trajectories of the B0-P4 test set. When the blue and red lines overlap, optimal prediction accuracy is achieved. Notably, Fig. 9a reveals numerous non-overlapping trajectories in the upper part, indicating a suboptimal prediction effect. In contrast, Fig. 9b exhibits fewer non-overlapping tra-



(a) (b) (c)

Fig. 9 Performance of different quantitative AP credibility methods on B0-P4. Fig. (a), (b) and (c) show the results of using the Arithmetic Method, the Equal Ratio Method, and our method, respectively. AP, access point.

Unedited

jectories, confirming that the equal ratio method is inferior to the equal difference method. In the lower part of Fig. 9c, the predicted trajectory almost perfectly aligns with the actual trajectory, a feature absent in both the arithmetic and equal ratio methods.

4.5 Effectiveness of AP set similarity

AP set similarity delineates the likeness between the strong APs of the TP and RP, serving as an effective measure of their overall similarity. To assess the effectiveness of our proposed AP set similarity method, we excluded this method on three datasets: this transformation simplifies the AP set similarity to:

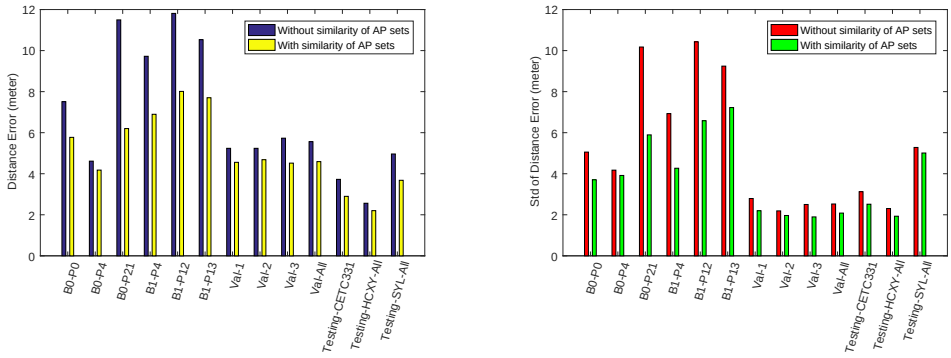
$$\boldsymbol{\alpha} = [1\ 1\ \ldots\ 1]_{N_r} \tag{26}$$

Fig. 10 compares the results between using and not using AP set similarity. As depicted in Fig. 10a, across three distinct datasets, the average positioning error of the AP set similarity method consistently outperforms the method without AP set similarity. Across diverse datasets, our proposed AP set similarity method consistently attains superior positioning accuracy, affirming the method's effectiveness and robustness. Additionally, in the standard deviation results presented in Fig. 10b, the method that does not use AP set similarity consistently exhibits a higher standard deviation compared to the method incorporating it. This suggests that the AP set similarity method can maintain a specific positioning prediction error range, preventing extremely biased prediction results. In contrast, the method without AP set similarity reaches a prediction standard deviation of 10 m, indicating its ineffectiveness in handling samples that are challenging to predict accurately. This underscores the effectiveness of our proposed AP set similarity, which processes AP sorted results based on RSS size rather than directly handling RSS values, thereby eliminating issues arising from small fluctuations in RSS and errors in similarity calculation.

4.6 Ablation experiment

The rationale derived from the RSS decay model demonstrates that RSS similarity inadequately reflects the proximity of locations, with the stability of strong APs surpassing that of weak APs. Building on this insight, we leveraged prior knowledge of the RSS decay law to quantify the credibility of strong APs, leading to the development of a novel method for measuring RSS distance that incorporates the weighting of strong APs. Simultaneously, achieving the optimal target for KNN prediction with a static K value is challenging. To address this, we introduced a a dynamic K value algorithm based on Neighbor Density (ND-DKA).

To assess the efficacy of the combined strong AP weighted RSS distance measurement method with ND-DKA, we conducted experiments by randomly selecting two test sets across three types of datasets. The results are detailed in Table 5, where "no-Weight" denotes the RSS ED without strong AP weighting, that is, the transformation of Eq. (14) into Eq. (27), and "no-Kdynamic" signifies the re-



(a) (b)

Fig. 10 Comparison of distance errors with and without using similarity of AP sets. Fig. (a) and (b) show the average positioning error results and standard deviation results respectively. AP, access point.

Table 5 Comparison of different methods for quantifying AP credibility. The bold values are our methods.

Method	B0-P21	B1-P12	Val-1	Val-All	Testing-CETC331	Testing-SYL-All
Equal Ratio Method	8.79 ± 8.54	9.51 ± 8.06	5.41 ± 2.46	5.39 ± 2.33	3.48 ± 3.22	4.59 ± 5.18
Arithmetic Method	7.37 ± 6.34	8.77 ± 5.62	4.72 ± 1.99	4.78 ± 1.81	3.11 ± 2.64	4.01 ± 5.08
Ours	**6.20 ± 5.89**	**8.01 ± 6.58**	**4.56 ± 2.20**	**4.59 ± 2.08**	**2.96 ± 2.63**	**3.68 ± 5.01**

Unedited

sult without the ND-DKA method.

$$D = \sqrt{\sum_{m=1}^{M} \omega_m (RSS_m^{tp} - RSS_m^{rp})^2} \tag{27}$$

Table 6 reveals that "Weight + Kdynamic" consistently performs optimally across the six test sets. Specifically, on B0–P0, "Weight + Kdynamic" outperforms "no-Weight + Kdynamic" by 22.76%. Similarly, without the ND-DKA method, "Weight + no-Kdynamic" surpasses "no-Weight + no-Kdynamic" by 21.88%, demonstrating an effective enhancement in positioning accuracy. This affirms the effectiveness of the strong AP weighting method in KNN. When applying the strong AP weighting method, "Weight + Kdynamic" exhibits a 6.67% improvement on B1–P13 and a 24.12% improvement on Testing-SYL-All compared to "Weight + no-Kdynamic." Notably, when the strong AP weighting method is not employed, the results consistently favor the ND-DKA method, with "no-Weight + Kdynamic" consistently outperforming "no-Weight + no-Kdynamic." This underscores the superiority of the method using ND-DKA. Experimental results across various datasets indicate the relative robustness of our ND-DKA method, showcasing its capacity to automatically optimize and enhance positioning prediction accuracy. Overall, "Weight + Kdynamic" demonstrates improvements of 24.67%, 20.03%, and 27.32%, respectively, on Val-3, Val-All, and Testing-CETC331 compared to "no-Weight + no-Kdynamic." This verification indicates that our method consistently achieves stable and accurate positioning results under both continuous walking and non-continuous environments. It further affirms the efficacy of the strong AP weighted RSS distance measurement method combined with ND-DKA in significantly improving the prediction effectiveness of indoor positioning.

Additionally, Fig. 11 illustrates the sample error trend on GeotecDatabase and SODIndoorLoc when employing and not employing AP weighting. The prediction error curve using the AP weighting method is consistently below the red line, and on the SODIndoorLoc test set, prediction errors are evident in the red line, which is attributed to RSS instability. In contrast, the use of the AP weighting method results in sample errors confined within a specific range, exhibiting minimal fluctuations. This underscores the capability of the weighting method to mitigate inaccuracies in similarity calculation stemming from abnormal RSS fluctuations, thereby maintaining a consistently robust prediction effect.

Table 6 Ablation results of SAPC-DKNN. The bold values are our methods.

Method	B0-P0	B1-P13	Val-3	Val	Testing-CETC331	Testing-SYL-All
no-Weight + no-Kdynamic	8.09 ± 5.60	10.83 ± 9.40	6.00 ± 2.96	5.74 ± 2.88	3.99 ± 3.58	5.00 ± 5.19
no-Weight + Kdynamic	7.47 ± 5.03	10.60 ± 9.23	5.68 ± 2.32	5.53 ± 2.56	3.75 ± 3.12	4.92 ± 5.23
Weight + no-Kdynamic	6.32 ± 4.20	8.25 ± 7.44	4.83 ± 2.60	4.85 ± 2.63	3.18 ± 3.04	4.85 ± 5.18
Weight + Kdynamic	**5.66±3.61**	**7.70±7.22**	**4.52±1.90**	**4.59±2.08**	**2.96±2.63**	**3.68±5.01**

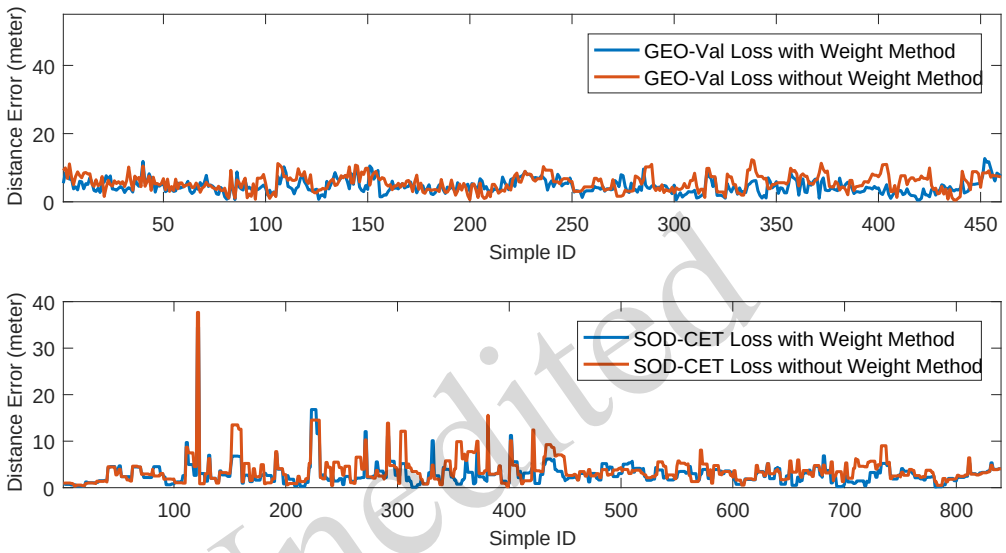


Fig. 11 Comparison of distance errors with and without AP weighting on GeotecDatabase (top) and SODIndoorLoc (bottom). AP, access point.

Unedited

To further investigate the effectiveness of ND-DKA, we conducted experiments with three selected test sets. Fig. 12 illustrates the prediction results under different K values when ND-DKA is not applied. Fig. 13 illustrates the distribution of automatically selected k-values when using ND-DKA. The k-value range in ND-DKA is set between 3 and 30. A minimum value of k=3 satisfies the geometric constraint that three points must be coplanar to ensure the basic computational functionality of Algorithm 1 for the convex hull area. The maximum value of k=30 is chosen as a reasonable upper limit based on experimental observations. The frequency of larger k-values decreases significantly, indicating their limited contribution to classification performance. The selection of $K = 30$ balances algorithmic efficiency with the requirements of diverse datasets.

In Fig. 12, the prediction errors across the three datasets exhibit a peak at a specific K value. Beyond this point, the errors increase as K continues to grow. However, compared with the method employing ND-DKA, the errors remain higher. Additionally, the optimal K value that achieves the best overall prediction performance varies across datasets. The optimal solution for the static K-value is a global optimum calculated based on the average error across the entire dataset. However, it may not reflect the most suitable number of neighbors for an individual sample. Due to the heterogeneity in feature distributions across samples, some samples may rely heavily on a small number of highly correlated RPs, whereas others require more RPs to synthesize information effectively. Therefore, although the static K-value method is appropriate at a global level, it may not suit every individual sample. The core advantage of ND-DKA lies in its ability to adaptively adjust the K-value according to the characteristics of each sample, achieving a local optimum. As illustrated in Fig. 13, on Testing_CETC331, the K-value distribution for ND-DKA method reveals that high-frequency K-values are concentrated between 3 and 6. This indicates that most samples only require a small num-

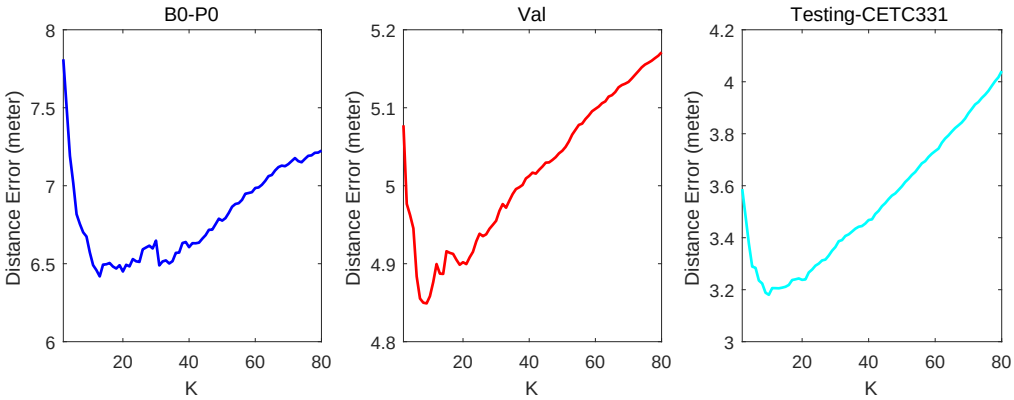


Fig. 12 Comparison of distance errors against K values without using ND-DKA on three datasets.

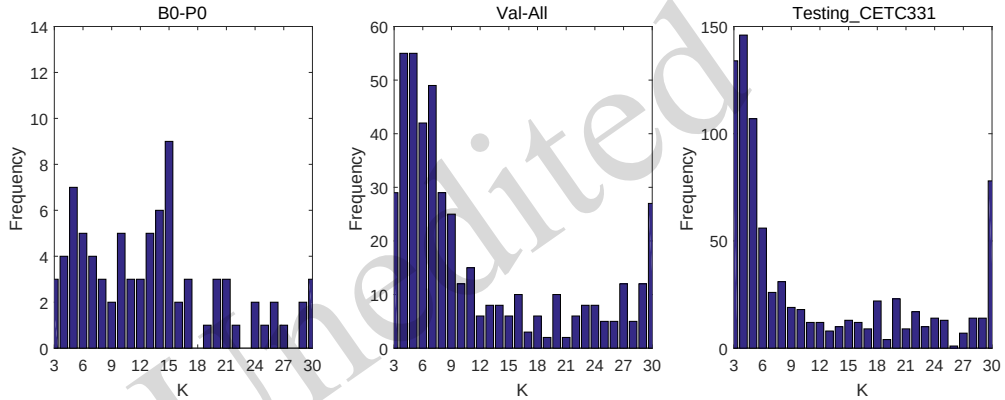


Fig. 13 Frequency distribution histogram of K values selected using ND-DKA on three datasets.

Unedited

ber of highly correlated reference points (RPs) to achieve better prediction results. Simultaneously, the dynamic distribution includes some larger K-values (e.g., above 10) to account for samples requiring additional RPs. These additional RPs may have lower Euclidean Distance (ED) correlations but can provide significant potential contributions. Furthermore, on different datasets, the selected K-value distributions vary, demonstrating that ND-DKA optimizes at the individual sample level through its adaptive mechanism. This approach avoids completely excluding RPs with low ED correlations and comprehensively considers the contributions of all RPs across different scenarios. Consequently, ND-DKA outperforms the static method as a whole. In summary, our approach demonstrates robustness and effectiveness.

5 Conclusions

In conclusion, we introduced a Dk-NN method based on strong AP credibility for indoor positioning (SAPC-DKNN). Leveraging the progression of the RSS path loss model, this method quantifies AP credibility and devises an RSS distance measurement technique incorporating strong AP weighting. In comparison with alternative weighting methods, SAPC-DKNN uses both the order and strength information of the AP level to mitigate similarity errors arising from RSS fluctuations, thereby significantly enhancing positioning accuracy. Furthermore, through integration with a dynamic K value algorithm based on NN density, SAPC-DKNN achieves heightened precision and stability in real-time positioning performance. Our results demonstrate that SAPC-DKNN adeptly addresses key challenges inherent in KNN, such as the mismatch between RSS similarity and physical distance, RSSI instability, and uncertainty in optimal K-value so-

lutions. Experimental findings reveal a substantial reduction in average positioning error, ranging from 18.67% to 66.48%, surpassing the performance of the traditional KNN method. In future research endeavors, we aim to extend the application of the attenuation model's prior knowledge to more intricate prediction models, striving for even more precise indoor positioning solutions.

Contributors

Yuting YANG designed the research, processed the data, and drafted the paper. Tao ZHANG and Wu HUANG revised and finalized the paper.

Conflict of interest

All the authors declare that they have no conflict of interest.

References

Alitaleshi A, Jazayeriy H, Kazemitabar J, 2023. Ea-cnn: A smart indoor 3d positioning scheme based on wi-fi fingerprinting and deep learning. *Engineering Applications of Artificial Intelligence*, 117:105509. https://doi.org/10.1016/j.engappai.2022.105509

Arslantas H, Okdem S, 2024. Indoor localization with an autoencoder based convolutional neural network. *IEEE Access*, (12):46059-46066. https://doi.org/10.1109/ACCESS.2024.3382135

Ayinla SL, Abd Aziz A, Drieberg M, 2024. Salloc: An accurate target localization in wifi-enabled indoor environments via sae-alstm. *IEEE Access*, 12:19694-19710.

Bahl P, Padmanabhan VN, 2000. Radar: An in-building rf-based user location and tracking system. Proceedings IEEE INFOCOM 2000 Conference on computer communications Nineteenth annual joint conference of the IEEE computer and communications societies (Cat No 00CH37064), 2:775-784. https://doi.org/10.1109/INFCOM.2000.832252

Bi J, Wang Y, Yu B, et al., 2022. Supplementary open dataset for wifi indoor localization based on received signal strength. *Satellite Navigation*, 3(1):25. https://doi.org/10.1186/s43020-022-00086-y

Brunato M, Battiti R, 2005. Statistical learning theory for location fingerprinting in wireless lans. *Computer Networks*, 47(6):825-845. https://doi.org/10.1016/j.comnet.2004.09.004

Cha J, Lim E, 2022. A hierarchical auxiliary deep neural network architecture for large-scale indoor localization based on wi-fi fingerprinting. *Applied Soft Computing*, 120:108624. https://doi.org/10.1016/j.asoc.2022.108624

Chen G, Guo X, Liu K, et al., 2022. Rwknn: A modified wknn algorithm specific for the indoor localization problem. *IEEE Sensors Journal*, 22(7):7258-7266. https://doi.org/10.1109/JSEN.2022.3155902

Chon HD, Jun S, Jung H, et al., 2004. Using rfid for accurate positioning. *Positioning*, 1(8):0-0.

Ciurana M, Cugno S, Barcelo-Arroyo F, 2007. Wlan indoor positioning based on toa with two reference points. 2007 4th Workshop on Positioning, Navigation and Communication, p.23-28. https://doi.org/10.1109/WPNC.2007.353607

Costa JA, Patwari N, Hero III AO, 2006. Distributed weighted-multidimensional scaling for node localization in sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, 2(1):39-64. https://doi.org/10.1145/1138127.1138129

Dag T, Arsan T, 2018. Received signal strength based least squares lateration algorithm for indoor localization. *Computers & Electrical Engineering*, 66:114-126.

Dong Y, Arslan T, Yang Y, 2022. An encoded lstm network model for wifi-based indoor positioning. 2022 IEEE 12th International Conference on Indoor Positioning and Indoor Navigation (IPIN), p.1-6. https://doi.org/10.1109/IPIN54987.2022.9918116

Dong ZY, Xu WM, Zhuang H, 2019. Research on zigbee indoor technology positioning based on rssi. *Procedia Computer Science*, 154:424-429. https://doi.org/10.1016/j.procs.2019.06.060

Gu Y, Lo A, Niemegeers I, 2009. A survey of indoor positioning systems for wireless personal networks. *IEEE Communications surveys & tutorials*, 11(1):13-32. https://doi.org/10.1109/SURV.2009.090103

He S, Chan SHG, 2015. Wi-fi fingerprint-based indoor positioning: Recent advances and comparisons. *IEEE Communications Surveys & Tutorials*, 18(1):466-490. https://doi.org/10.1109/COMST.2015.2464084

Hegarty CJ, Chatre E, 2008. Evolution of the global navigation satellitesystem (gnss). *Proceedings of the IEEE*, 96(12):1902-1917. https://doi.org/10.1109/JPROC.2008.2006090

Hoang MT, Zhu Y, Yuen B, et al., 2018. A soft range limited k-nearest neighbors algorithm for indoor localization enhancement. *IEEE Sensors Journal*, 18(24):10208-10216. https://doi.org/10.1109/JSEN.2018.2874453

Hu J, Liu D, Yan Z, et al., 2018a. Experimental analysis on weight K-nearest neighbor indoor fingerprint positioning. *IEEE Internet of Things Journal*, 6(1):891-897. https://doi.org/10.1109/JIOT.2018.2864607

Hu J, Liu H, Liu D, 2018b. Toward a dynamic k in k-nearest neighbor fingerprint indoor positioning. 2018 IEEE International Conference on Information Reuse and Integration (IRI), p.308-314. https://doi.org/10.1109/IRI.2018.00054

Hu X, Shang J, Gu F, et al., 2015. Improving wi-fi indoor positioning via ap sets similarity and semi-supervised affinity propagation clustering. *International Journal of Distributed Sensor Networks*, 11(1):109642. https://doi.org/10.1155/2015/109642

Jin R, Xu H, Che Z, et al., 2015. Experimental evaluation of reducing ranging-error based on receive signal strength indication in wireless sensor networks. *IET Wireless Sensor Systems*, 5(5):228-234. https://doi.org/10.1049/iet-wss.2013.0139

Latif E, Parasuraman R, 2022. Online indoor localization using doa of wireless signals. *arXiv preprint arXiv:220105105*, . https://arxiv.org/abs/2201.05105

Unedited

Le Y, Zhang H, Shi W, et al., 2021. Received signal strength based indoor positioning algorithm using advanced clustering and kernel ridge regression. *Frontiers of Information Technology & Electronic Engineering*, 22(6):827-838.
https://doi.org/10.1631/FITEE.2000093

Lee I, Kwak M, Han D, 2016. A dynamic k-nearest neighbor method for wlan-based positioning systems. *Journal of Computer Information Systems*, 56(4):295-300.
https://doi.org/10.1080/08874417.2016.1164000

Lin H, Purmehdi H, Fei X, et al., 2023. Two-stage clustering for improve indoor positioning accuracy. *Automation in Construction*, 154:104981.
https://doi.org/10.1016/j.autcon.2023.104981

Liu H, Darabi H, Banerjee P, et al., 2007. Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(6):1067-1080.
https://doi.org/10.1109/TSMCC.2007.905750

Liu S, Jiang Y, Striegel A, 2013. Face-to-face proximity estimationusing bluetooth on smartphones. *IEEE Transactions on Mobile Computing*, 13(4):811-823.
https://doi.org/10.1109/TMC.2013.44

Liu S, De Lacerda R, Fiorina J, 2022. Performance analysis of adaptive k for weighted k-nearest neighbor based indoor positioning. 2022 IEEE 95th Vehicular Technology Conference:(VTC2022-Spring), p.1-5.

Lohan E, Torres-Sospedra J, Richter P, et al., 2017. Crowdsourced wifi-fingerprinting database and benchmark software for indoor positioning. *Zenodo Repos*, .
https://doi.org/10.5281/zenodo.889798

Ma J, Li X, Tao X, et al., 2008. Cluster filtered knn: A wlan-based indoor positioning scheme. 2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks, p.1-8.
https://doi.org/10.1109/WOWMOM.2008.4594840

Ma R, Guo Q, Hu C, et al., 2015. An improved wifi indoor positioning algorithm by weighted fusion. *Sensors*, 15(9):21824-21843.
https://doi.org/10.3390/s150921824

Madigan D, Einahrawy E, Martin RP, et al., 2005. Bayesian indoor positioning systems. Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies, 2:1217-1227.
https://doi.org/10.1109/INFCOM.2005.1498348

Mingzhe X, Jiabin C, Chunlei S, et al., 2015. The indoor positioning algorithm research based on improved location fingerprinting. The 27th Chinese Control and Decision Conference (2015 CCDC), p.5736-5739.
https://doi.org/10.1109/CCDC.2015.7161827

Nabati M, Ghorashi SA, 2023. A real-time fingerprint-based indoor positioning using deep learning and preceding states. *Expert Systems with Applications*, 213:118889.
https://doi.org/10.1016/j.eswa.2022.118889

Nguyen SM, Le DV, Havinga PJ, 2023. Learning the world from its words: Anchor-agnostic transformers for fingerprint-based indoor localization. 2023 IEEE International Conference on Pervasive Computing and Communications (PerCom), p.150-159.
https://doi.org/10.1109/PERCOM56429.2023.10099376

Nguyen SM, Le DV, Havinga PJ, 2024. Seeing the world from its words: All-embracing transformers for fingerprint-based indoor localization. *Pervasive and Mobile Computing*, 100:101912.
https://doi.org/10.1016/j.pmcj.2024.101912

Ni LM, Liu Y, Lau YC, et al., 2003. Landmarc: Indoor location sensing using active rfid. Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003(PerCom 2003), p.407-415.
https://doi.org/10.1109/PERCOM.2003.1192765

Oh J, Kim J, 2018. Adaptivek-nearest neighbour algorithm for wifi fingerprint positioning. *Ict Express*, 4(2):91-94.
https://doi.org/10.1016/j.icte.2018.04.004

Peng Y, Fan W, Dong X, et al., 2016. An iterative weighted knn (iw-knn) based indoor localization method in bluetooth low energy (ble) environment. 2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), p.794-800.
https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0127

Pu YC, You PC, 2018. Indoor positioning system based on ble location fingerprinting with classification approach. *Applied Mathematical Modelling*, 62:654-663.
https://doi.org/10.1016/j.apm.2018.06.031

Ren Q, Wang Y, Liu S, et al., 2023. Fstnet: Learning spatial–temporal correlations from fingerprints for indoor positioning. *Ad Hoc Networks*, 149:103244.
https://doi.org/10.1016/j.adhoc.2023.103244

Rusli ME, Ali M, Jamil N, et al., 2016. An improved indoor positioning algorithm based on rssi-trilateration technique for internet of things (iot). 2016 International Conference on Computer and Communication Engineering (ICCCE), p.72-77.
https://doi.org/10.1109/ICCCE.2016.28

Sadowski S, Spachos P, Plataniotis KN, 2020. Memoryless techniques and wireless technologies for indoor localization with the internet of things. *IEEE Internet of Things Journal*, 7(11):10996-11005.
https://doi.org/10.1109/JIOT.2020.2992651

Salamah AH, Tamazin M, Sharkas MA, et al., 2016. An enhanced wifi indoor localization system based on machine learning. 2016 International conference on indoor positioning and indoor navigation (IPIN), p.1-8.
https://doi.org/10.1109/IPIN.2016.7743586

Song X, Fan X, Xiang C, et al., 2019. A novel convolutional neural network based indoor localization framework with wifi fingerprinting. *IEEE access*, 7:110698-110709.
https://doi.org/10.1109/ACCESS.2019.2933921

Torres-Sospedra J, Montoliu R, Martínez-Usó A, et al., 2014. Ujiindoorloc: A new multi-building and multi-floor database for wlan fingerprint-based indoor localization problems. 2014 international conference on indoor positioning and indoor navigation (IPIN), p.261-270.
https://doi.org/10.1109/IPIN.2014.7275492

Torres-Sospedra J, Montoliu R, Trilles S, et al., 2015. Comprehensive analysis of distance and similarity measures for wi-fi fingerprinting indoor positioning systems. *Expert Systems with Applications*, 42(23):9263-9278.
https://doi.org/10.1016/j.eswa.2015.08.013

Unedited

Torres-Sospedra J, Montoliu R, Mendoza-Silva GM, et al., 2016. Providing databases for different indoor positioning technologies: Pros and cons of magnetic field and wi-fi based positioning. *Mobile Information Systems*, 2016(1):6092618.
https://doi.org/10.1155/2016/6092618

Umair MY, Ramana KV, Dongkai Y, 2014. An enhanced k-nearest neighbor algorithm for indoor positioning systems in a wlan. 2014 IEEE Computers, Communications and IT Applications Conference, p.19-23.

Wu D, Xu Y, Ma L, 2009. Research on rss based indoor location method. 2009 Pacific-Asia Conference on Knowledge Engineering and Software Engineering, p.205-208.
https://doi.org/10.1109/KESE.2009.67

Xianjia Y, Qingqing L, Queralta JP, et al., 2021. Applications of uwb networks and positioning to autonomous robots and industrial systems. 2021 10th Mediterranean Conference on Embedded Computing (MECO), p.1-6.
https://doi.org/10.1109/MECO52532.2021.9460266

Xie Y, Wang Y, Nallanathan A, et al., 2016. An improved k-nearest-neighbor indoor localization method based on spearman distance. *IEEE signal processing letters*, 23(3):351-355.
https://doi.org/10.1109/LSP.2016.2519607

Yang C, Shao HR, 2015. Wifi-based indoor positioning. *IEEE Communications Magazine*, 53(3):150-157.
https://doi.org/10.1109/MCOM.2015.7060497

Zhang H, Wang Z, Xia W, et al., 2022. Weighted adaptive knn algorithm with historical information fusion for fingerprint positioning. *IEEE Wireless Communications Letters*, 11(5):1002-1006.
https://doi.org/10.1109/LWC.2022.3152610

Zhang J, Mao H, 2022. Wknn indoor positioning method based on spatial feature partition and basketball motion capture. *Alexandria engineering journal*, 61(1):125-134.
https://doi.org/10.1016/j.aej.2021.04.078

Zhao Y, Gong W, Li L, et al., 2023. An efficient and robust fingerprint based localization method for multi floor indoor environment. *IEEE Internet of Things Journal*, .
https://doi.org/10.1109/JIOT.2023.3298603

Zou H, Jin M, Jiang H, et al., 2017. Winips: Wifi-based non-intrusive indoor positioning system with online radio map construction and adaptation. *IEEE Transactions on Wireless Communications*, 16(12):8118-8130.
https://doi.org/10.1109/TWC.2017.2757472

Unedited