Journal of King Saud University – Computer and Information Sciences 35 (2023) 101696

HOSTED BY

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Robust and fast algorithm design for efficient Wi-Fi fingerprinting based indoor positioning systems

Asim Abdullah [a,*], Omar Abdul Aziz [a,*], Rozeha A. Rashid [a], Muhammad Haris [b,c], Mohd Adib Sarijari [a]

[a] Telecommunication Software and Systems Research Group, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, Johor, Malaysia
[b] Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru 81310, Johor, Malaysia
[c] Department of Computer Science & Bioinformatics, Khushal Khan Khattak University, Karak 27200, Khyber Pakhtunkhwa, Pakistan

ARTICLE INFO

Article history:
Received 1 May 2023
Revised 24 July 2023
Accepted 29 July 2023
Available online 6 August 2023

Keywords:
Indoor positioning
Wi-Fi
Fingerprinting
RSSI
Radio map construction
Data filtering
Area classification

ABSTRACT

Indoor positioning systems (IPS) based on Wi-Fi fingerprinting have gained significant attention due to their potential for providing location-based services. Large scale IPS deployments require implementation of robust, accurate, and fast algorithms. Data analytics assisted algorithms provide positioning accuracy improvements, however their integration within real-time and scalable solutions significantly depends on the computational complexity. Therefore, we propose robust and computationally efficient algorithms for performance enhancements in Wi-Fi fingerprinting-based IPS. A robust radio map algorithm based on enhanced statistical cluster initialization was designed for efficient indoor environment characterization. The proposed data filtering algorithm leveraged smart clustering to mitigate real-time data variations. The designed area classification algorithm was based on smart dual-band data aggregation. We evaluated the performance of proposed algorithms based on accuracy and computation time. The performance evaluations signified accuracy and computational enhancement, in comparison to related benchmark techniques. The data filtering and area classification algorithms required 40% less computation time. Simultaneously, 14.5% and 36% accuracy improvements were recorded for the area classification and radio map algorithms respectively. The proposed algorithms have the potential to significantly enhance IPS performance in a variety of real-time applications, including indoor navigation and asset tracking.

1. Introduction

The demanding requirements of contemporary and future indoor navigation applications, render the Global Positioning System (GPS) based localization, as an inefficient solution (Zafari et al., 2019; Brena et al., 2017; Din et al., 2018). The demand for efficient Indoor Positioning Systems (IPS) has become prominent, with Internet of Things (IoT) emergence (Zafari et al., 2019; Yang et al., 2021). Robust IPS based tracking systems (Frankó et al., 2020) are also paramount for the realization of smart factories (Osterrieder et al., 2020) and smart warehouses (Buntak et al., 2019). Therefore, innovative IPS dedicated research works are pertinent and prevalent (Lin et al., 2023). The ubiquity of Wi-Fi networks, promising localization performance, and wide range applications continue to promote research interest in Wi-Fi fingerprinting based IPS (Bellavista-Parent et al., 2021; Hayward et al., 2022).

As the application range of Wi-Fi fingerprinting based IPS is wide, enhanced robustness and system performance are imperative for future solutions (Shang and Wang, 2022). In order to facilitate large scale deployments, it is necessary to design real-time and scalable fingerprinting algorithms. Radio map construction and area classification algorithms designed for computational efficiency, degrade localization accuracy (Torres-Sospedra et al., 2022). The incurred inaccuracy in the constructed radio map, is mainly due to inadequate characterization of temporal variations in Received Signal Strength Identifier (RSSI) data. Robust area classification relies on the efficiency of area partitioning stage. The robustness however, is limited by the random cluster initialization of the area partitioning algorithm (Zhang et al., 2020). Data filtering algorithms provide accuracy enhancement by statistical conditioning of real-time samples. Although the conventional data filters are computationally efficient, their robustness to multipath varia-

* Corresponding authors.
E-mail addresses: abdullahasim@graduate.utm.my (A. Abdullah), omar@utm.my (O.A. Aziz).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2023.101696
1319-1578/

tions is inadequate for future applications (Koledoye et al., 2018). On the other hand, adaptive gaussian and clustering based filtering algorithms provide robust accuracy, at the cost of high computational complexity (Yang et al., 2020; Shi et al., 2020). Therefore, fingerprinting algorithm designs that incorporate robustness, accuracy, and computational cost are integral for implementation of efficient future systems.

In this paper, we address the aforementioned gap by designing robust, computationally efficient algorithms for real-time and scalable Wi-Fi fingerprinting based IPS. The algorithms are designed to enhance accuracy and computational efficiency, thereby providing robust solutions for real-time IPS implementation. Firstly, robust indoor environment characterization is ensured by generation of accurate radio maps. The characterization is leveraged by statistical clustering of Received Signal Strength Identifier (RSSI) data. Secondly, the filtering algorithm efficiently estimates appropriate RSSI from real-time data. Computational enhancement is ensured by statistics driven, smart execution of the computation intensive k-means clustering based data filtering. Lastly, the proposed area classification algorithm, performs rigorous offline analysis, to smartly partition the environment into appropriate sub-areas. In the online stage, classification is completed by intelligent dual-band RSSI data aggregation. Due to the modular design of the algorithms, IPS realizations or deployments can select subsets of proposed algorithms, as per application and system requirements. The important contributions of our presented work are summarized as:

- A set of smart, modular algorithms that enhance accuracy and computational efficiency of Wi-Fi fingerprinting based IPS.
- Development of a radio map construction algorithm–Statistical Clustered Radio Map (SCRM), to improve positioning accuracy.
- Introduction of the Statistically Enhanced RSSI Clustering (SERC) data filtering algorithm, for computational enhancement without accuracy compromise.
- Development of the Dual-Band Area Classification (DBAC) algorithm, that simultaneously enhances accuracy, and computational efficiency.

The remainder of this paper is organized as follows. In Section 2, related fingerprinting algorithms, designed for IPS performance enhancement are reviewed. Section 3 explains the design methodology, and Section 4 reports the performance evaluation of proposed algorithms. The main conclusions of this work are finally discussed in Section 5.

2. Related works

The demand of accurate and robust Location Based Services (LBS), continues to foster enhanced positioning IPS techniques (Xie et al., 2022a; Biswas et al., 2023; Zhang et al., 2023). Performance enhancement of fingerprinting based IPS requires robust algorithms including radio map construction, data filtering, and area classification algorithms (Roy and Chowdhury, 2021; Roy and Chowdhury, 2022). Radio map construction algorithms are required to create maps of radio signal characteristics within an indoor environment. The target of such algorithms is to generate a robust radio map that assists the positioning algorithm to estimate real-time location with high accuracy. Conventional algorithms store one sample per Reference Point (RP), using mean or median sampling (Jung et al., 2017). The traditional algorithms assist low computational positioning, however high errors are incurred in multipath conditions. A fingerprint clustering based algorithm was proposed as a robust solution with improved accuracy over conventional algorithms (Wang et al., 2019). The performance degradation in areas with weak or variable signal strengths however is not completely addressed. Clustering was utilized to generate smaller radio maps and reduce computation time of online positioning (Torres-Sospedra et al., 2022). Reducing the radio map size however has an adverse impact on accuracy. Gaussian Process Regression (GPR) based radio map construction algorithms were proposed to generate radio maps with minimal data acquisition (Zou et al., 2017; Huang et al., 2019). GPR based algorithms provide a solution for small datasets, however they are confronted with challenges of significant performance degradation in large environments. Dual-band Wi-Fi data were incorporated in the radio map to achieve higher accuracy than single band data (Yiu et al., 2017). Multiple samples per RP in the radio map, also increased localization accuracy (Huang et al., 2021). However, incorporation of multi-sample, dual-band data in radio map construction is not addressed. A precise Dual-Band Radio Map (DBRM) algorithm, stored multiple dual-band Wi-Fi samples MAC-wise, in the radio map (Ozdemir and Ceylan, 2020). Robustness of DBRM is although restrained due to inadequate characterization of temporal variations.

Data filtering algorithms estimate the appropriate RSSI value over multiple real-time samples, to provide clean data required for high accuracy localization, by the positioning algorithm. The performance of conventional, computationally efficient windowed filters was analyzed by (Koledoye et al., 2018). Evaluated filtering algorithms included moving average, moving exponential average, and moving median. The data filters showed performance improvement in comparison to using raw RSSI data. However, large window size requirement for improved accuracy and degraded performance in non line of sight scenario, restrict robustness of the algorithms in multipath conditions. The moving window approach is also a computationally efficient solution in tracking scenarios (Hu and Hu, 2023). The Savitzky-Golay filter was proposed as an efficient digital filter capable of smoothing non-linear, noisy input data (Krishnan and Seelamantula, 2013). The sensitivity of Savitzky-Golay filter to outliers in RSSI data however, is known to degrade accuracy of location estimations. The gaussian filtering algorithm was applied on real-time RSSI data to improve accuracy (Yang et al., 2020). The matrix convolution within the algorithm however, significantly increases computational complexity in high dimensional data scenarios usually incurred by large environments. The Isolation Forest algorithm, filtered RSSI data by anomaly detection (Guo et al., 2021). The filtering algorithm lacks adaptation to real-time variations, since it relies on offline training. The RSSI Clustering (RC) filter, proposed k-means clustering with smart cluster initialization, to remove outliers in successive RSSI samples (Shi et al., 2020). The RC filtering algorithm significantly outperforms conventional algorithms, in terms of accuracy. Due to the clustering process however, the algorithm is considerably computationally intensive.

Area classification algorithms are conveniently implemented in two stages. In the offline stage area partitioning of the indoor environment is performed. Area classification is completed in the online stage by estimation of sub-area as per real-time data. Finally, the positioning algorithm has only to search within the sub-area for final location estimate. Area partitioning was implemented by sub-area definition according to data similarity in terms of euclidean space (Wang et al., 2015). The proposed area classification algorithm was based on gradient descent search among possible sub-areas. The area partitioning according to data similarity does not effectively deal with overlapping areas. Also the iterative nature of gradient descent based area classification incurs added computational complexity. Area partitioning based on k-means clustering, and Bayes theorem for area classification were proposed by (Zhang et al., 2020). The performance of area classification however, is significantly dependent on the cluster initialization of the

unsupervised k-means clustering based area partitioning. The MAX algorithm partitioned the environment according to Access Points (APs) with highest RSSI. The MAX area classification was based on sorting real-time RSSIs from all detected APs (Xie et al., 2022b). Although the proposed algorithm is computationally efficient, yet area partitioning based on one AP per sub-area limits the positioning accuracy in multipath scenarios. Partitioning of rooms into two halves was proposed in (Biswas et al., 2023). The area classification was performed by selecting the half, that provided highest sum of RSSI probabilities. The partitioning scheme however lacks adaptability to smaller or larger environments where two segment partitioning could be an over-fit or under-fit solution.

In order to efficiently address the requirements of real-time and scalable, future IPS solutions, algorithms necessarily have to be designed for simultaneous accuracy and computational enhancements.

3. Materials and methods

In order to enhance the performance of indoor positioning systems, robust fingerprinting algorithms are proposed in this paper. Firstly, a radio map construction algorithm is devised to enhance accuracy. A data filtering algorithm that reduces computational complexity without accuracy compromise is also designed. Lastly, an area classification algorithm that jointly enhances accuracy and computational efficiency is proposed. All proposed algorithms are formulated and detailed in this section. The algorithm designs are also analyzed from the scalability perspective. Specifically, the capability to handle demanding computational requirements of large-scale deployment is discussed. Algorithm design and methodology details are preceded by introduction of the proposed system and dataset utilized for performance evaluations.

3.1. The proposed system

The proposed system integrates robust and efficient radio map construction, data filtering, and area classification algorithms for performance enhancement of fingerprinting based IPS. The modular design of proposed algorithms facilitates their incorporation within conventional fingerprinting solutions, in addition to forming the integrated system. Fig. 1 illustrates the proposed system architecture in the form of a flowchart.

In the offline phase, a radio map is constructed by leveraging k-means clustering with smart cluster initialization. The initialization is based on quantile estimation of sample windows containing equal number of RSSI samples. Using the radio map, further area partitioning is performed in the offline phase. The indoor environment is divided into labeled sub-areas defined by locations having common three strongest Access Points (APs). The online phase is initiated by calculating the standard deviation (σ) of v samples from the same Wi-Fi MAC. A low σ value indicates lower noise and mean filtering is performed in such scenario. A substantially

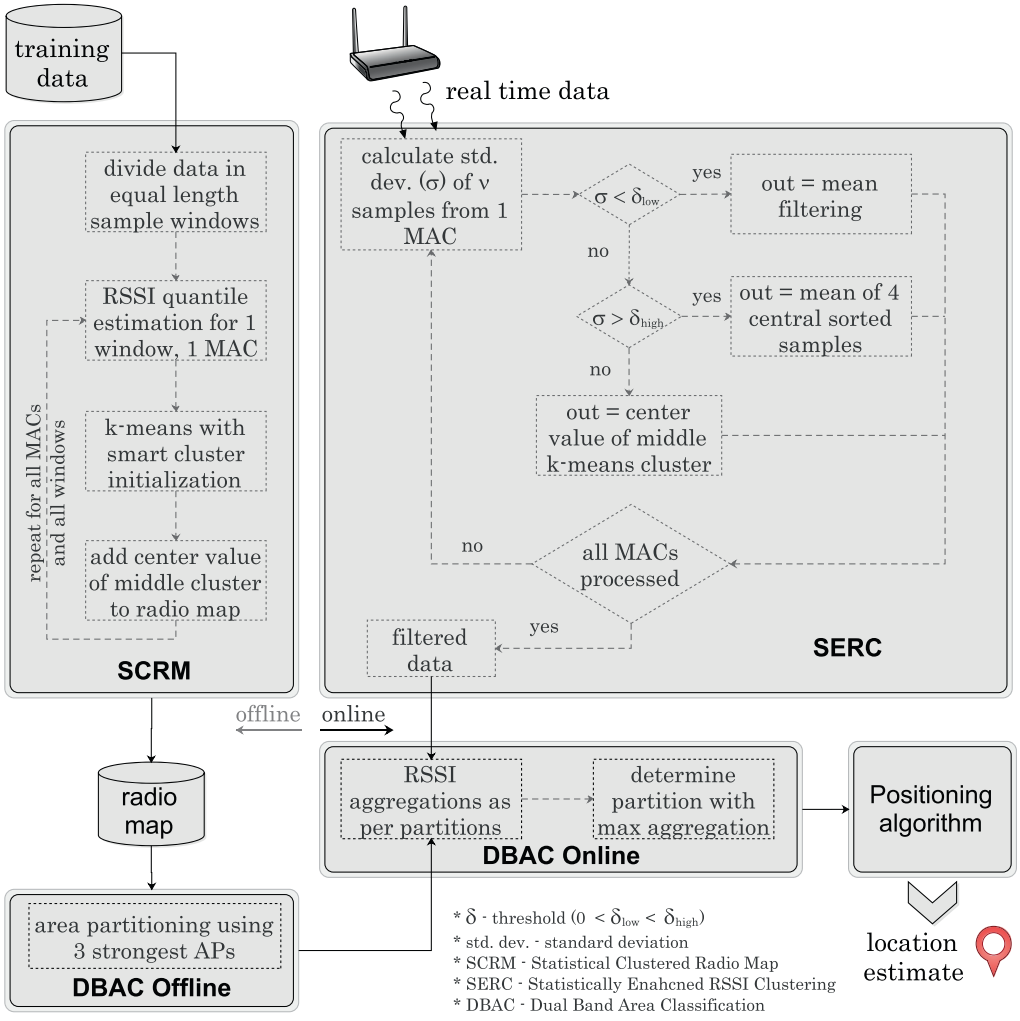

Fig. 1. Flowchart representation of proposed system.

high σ value represents high variations and outliers. In this case, mean of four central sorted samples is used as filtered output. If σ is between the threshold values, k-means clustering is executed using three clusters (Shi et al., 2020). The center of middle cluster is the filter output in this situation. Next, the filtered data are aggregated according to area partitions. The partition corresponding to highest aggregation is declared as classified area. Finally, the positioning algorithm estimates the location by searching within the identified area only. Prior to detailed explanation of algorithm design methodologies in Sections 3.4–3.6, details of the utilized dataset are described.

3.2. Dataset and indoor environment

Prior to description of the indoor environment and corresponding dataset utilized in this work, we detail the data requirements for implementation and evaluation of designed algorithms. As the proposed radio map construction algorithm is designed to statistically incorporate RSSI variations, extensive raw training data corresponding to several minutes of data per location, are required. Similarly the filtering algorithm requires extensive test data to evaluate the performance of RSSI variations mitigation. Furthermore, the area classification algorithm is designed for dual-band Wi-Fi environments. For robustness assessment, evaluations in multi-floor, multi-building scenarios are recommended (Roy and Chowdhury, 2022). In order to implement and evaluate the algorithms, this work utilized the UTMInDualSymFi dataset (Abdullah et al., 2022). The dataset is a comprehensive source of dual-band Wi-Fi data acquired at multi-building environments. Data corresponding to several consecutive minutes of acquisition per reference and test locations are provided in the dataset. The data is far more extensive than other multi-building datasets (Torres-Sospedra et al., 2014; Lohan et al., 2021), that provide few seconds of training data and only one sample per test location. Furthermore, the two mentioned datasets do not contain dual-band Wi-Fi data. Therefore, the utilization of these datasets, for implementation of the proposed radio map construction, data filtering, and area classification algorithms is not practicable. The UTMInDualSymFi dataset promotes high accuracy positioning due to availability of dual-band data collected at a high resolution of 1 m Reference Point (RP) spacing (Ozdemir and Ceylan, 2020; Obeidat et al., 2021). Furthermore, perfect network device homogeneity ensured wireless hotspots interference-free, Wi-Fi data within the dataset. The dataset also assists device heterogeneity evaluations by providing individual as well as merged data of two devices. Complete details of the dataset are available in the data descriptor (Abdullah et al., 2023). Other pertinent details of data provided in the dataset are summarized in Table 1.

3.3. Wi-Fi RSSI data formulation

Firstly, we define the notation convention adopted in this work. Matrices are denoted with uppercase bold symbols (e.g., $\mathbf{X}$). Lower case bold symbols represent vectors (e.g., $\mathbf{x}$). Scalars are represented by lower case italic symbols (e.g., x). We also adopt the typical terminologies of indoor fingerprinting positioning, in which, a Reference Point (RP), refers to a position with known location. Test Point (TP), is an unknown location that is to be estimated by the positioning algorithm using a constructed radio map.

The radio maps, training and test datasets contain two-dimensional Wi-Fi RSSI data, and therefore are represented as matrices. Hence $\mathbf{RM}, \mathbf{TR}$ and $\mathbf{TS}$ denote the radio map, training and test data respectively of a specific building. A single row within the matrices, constitutes one training/test/radio map sample. There are r RPs, t TPs, and m MACS within the data of considered building. The radio map contains s samples of each RP, and therefore $\mathbf{RM}$ is a, $rs \times m$ matrix. In the training data, at RP a, there are $n^{(a)}$ samples. The total samples within $\mathbf{TR}$ are $n^{(r)}$, such that

$$n^{(r)} = \sum_{i=1}^{r} n^{(i)} \tag{1}$$

Similarly, the total samples within $\mathbf{TS}$ are $n^{(t)}$, such that

$$n^{(t)} = \sum_{i=1}^{t} n^{(i)} \tag{2}$$

Eqs. (1) and (2) cater the scenario of unequal number of samples at RPs and TPs respectively. From Eqs. (1) and (2), $\mathbf{TR}$ is formulated as a, $n^{(r)} \times m$ matrix, and $\mathbf{TS}$ is a, $n^{(t)} \times m$ matrix. $\mathbf{TR}$ is defined as $\mathbf{TR} = [\mathbf{TR}_1 \mathbf{TR}_2 \dots \mathbf{TR}_r]^T$, where $\mathbf{TR}_x$ is the training data at RP-x, and is defined as $\mathbf{TR}_x = \left[\mathbf{tr}_{x,1}\ \mathbf{tr}_{x,2}\ \dots\ \mathbf{tr}_{x,n^{(x)}}\right]^T$; $1 \leqslant x \leqslant r$. The vectors within $\mathbf{TR}_x$, denote one complete sample and are defined as $\mathbf{tr}_{x,a} = \{\alpha_{x,1,a}\ \alpha_{x,2,a}\ \dots\ \alpha_{x,m,a}\}$; $1 \leqslant a \leqslant n^{(a)}$, where $\alpha_{x,p,q}$ is the RSSI (in dBm), at RP-x of MAC-p, in the q-th sample. By the same formulation $\mathbf{TS} = [\mathbf{TS}_1 \mathbf{TS}_2 \dots \mathbf{TS}_t]^T$; $\mathbf{TS}_y = \left[\mathbf{ts}_{y,1}\ \mathbf{ts}_{y,2}\ \dots\ \mathbf{ts}_{y,n^{(y)}}\right]^T$; $1 \leqslant y \leqslant t$; $\mathbf{ts}_{y,b} = \{\beta_{y,1,b}\ \beta_{y,2,b}\ \dots\ \beta_{y,m,b}\}$; $1 \leqslant b \leqslant n^{(b)}$, where $\beta_{y,p,q}$ is the RSSI (in dBm), at TP-y of MAC-p, in the q-th sample. Also, $\mathbf{RM} = [\mathbf{RM}_1 \mathbf{RM}_2 \dots \mathbf{RM}_r]^T$; $\mathbf{RM}_z = [\mathbf{rm}_{z,1}\ \mathbf{rm}_{z,2}\ \dots\ \mathbf{rm}_{z,s}]^T$; $1 \leqslant z \leqslant r$; $\mathbf{rm}_{z,c} = \{\gamma_{z,1,c}\ \gamma_{z,2,c}\ \dots\ \gamma_{z,m,c}\}$; $1 \leqslant c \leqslant s$, where $\gamma_{z,p,q}$ is the RSSI (in dBm), at RP-z of MAC-p, in the q-th sample of radio map. We also define MAC-wise vectors, that comprise data of a specific MAC. Formally, $\mathbf{trm}_{x,d} = \{\alpha_{x,d,1}\ \alpha_{x,d,2}\ \dots\ \alpha_{x,d,n^{(x)}}\}^T$; $\mathbf{tsm}_{y,e} = \{\beta_{y,e,1}\ \beta_{y,e,2}\ \dots\ \beta_{y,e,n^{(y)}}\}^T$; $\mathbf{rmm}_{z,f} = \{\gamma_{z,f,1}\ \gamma_{z,f,2}\ \dots\ \gamma_{z,f,s}\}^T$; $1 \leqslant d \leqslant m$; $1 \leqslant e \leqslant m$; $1 \leqslant f \leqslant m$. For practical implementations the TP labels are not used with test data, but are required for results verification.

Table 1
Summary of utilized dataset features.

Building	Floors	Wings	Total MACs	Training samples	Test samples
CX1	2	4	71	58,760	48,216
F04	2	6	64	54,240	41,333

3.4. Statistical Clustered Radio Map (SCRM) Algorithm

The proposed radio map construction algorithm–SCRM, statistically extracts RSSI data from recorded training samples, to achieve improved characterization of the wireless network. The algorithm is based on k-means clustering with enhanced statistical cluster initialization. Specifically, the cluster centers are initialized based on quantile estimation using a subset of RSSI samples, received from the same AP/MAC. The samples are clustered in three clusters (lower, middle, higher values), and the center of middle cluster is stored into the radio map. SCRM algorithm generates the radio map $\mathbf{RM}$, from training data $\mathbf{TR}$, of each building. The implementation schematic of SCRM within Wi-Fi fingerprinting based IPS is illustrated in Fig. 2.

The SCRM algorithm is initialized by arranging samples of each RP into s windows, each having $sw = \lfloor n^{(x)} \div s \rfloor$ samples. Essentially the data in $\mathbf{TR_x}$ are divided in sample windows. Depending on the implementation scheme, MAC-wise data available in $\mathbf{trm_{x,a}}$ can also be readily utilized. From each window one RSSI value, representative of the data is appended to the $\mathbf{rmw}$ vector, which eventually is incorporated in the final radio map $\mathbf{RM}$. The sw RSSI samples are sorted in ascending order and denoted as $\mathbf{sd}$. The

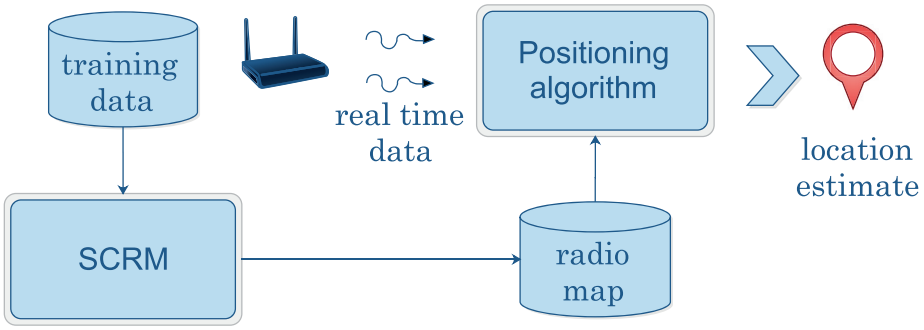

Fig. 2. SCRM implementation within Wi-Fi fingerprinting based IPS.

quantile of each sorted sample is estimated by forming **qts**, as in Eq. (3).

$$\mathbf{qts} = \left\{\frac{0.5}{sw}, \frac{1.5}{sw}, \frac{2.5}{sw}, \ldots \frac{(sw-0.5)}{sw}\right\} \tag{3}$$

To statistically cluster the data in three clusters, the 15th, 45th, and 85th quantile samples are used as initial cluster centers. The selected quantile values are utilized (a) to target consistent clustering performance by non-random initialization, (b) to provide reasonable distance between adjacent cluster centers and (c) define the middle cluster with widest range. Cluster initialization with other similar quantile values that meet the defined criteria, can also be readily used due to the iterative and convergence properties of clustering. To estimate the desired quantiles from **qts**, firstly **qt** is defined as $\mathbf{qt} = \{0.15, 0.45, 0.85\}$. The sample indices closest to the required quantiles are calculated using Eq. (4).

$$id_i = \arg\min_i \mathbf{qt_i} \quad \forall i = \{1,2,3\} \tag{4}$$

where $\mathbf{qt_i} = \mathbf{qts} - qt(i)\sum_{j=1}^{sw}\{1\};\ 1 \leqslant i \leqslant 3$. The initial cluster centers are subsequently calculated as in Eq. (5).

$$cint_i = \mathbf{sd}(id_i) \quad \forall i = \{1,2,3\} \tag{5}$$

Using the cluster centers, k-means clustering is applied to the RSSI data. The center of second (middle) cluster is added to **rmw**. The entire process is repeated for data of all windows, of all MACs, at all RPs.

The SCRM design facilitates scalable implementation, with the parameter *s*. Leveraging a lower value of *s* would construct a smaller radio map, thereby enhancing scalability prospect. Construction of a manageable radio map could however compromise accuracy. From the design perspective it is asserted that, SCRM supports scalability. The pseudocode of proposed SCRM algorithm is given in Algorithm 1.

Algorithm 1. SCRM algorithm pseudocode

Require: r, s, m, **TR**; **TR** $\neq \{\}$; $r > 0$, $s > 0$, $m > 0$ ▷(Section 3.3)
$k \leftarrow 3$
RM $\leftarrow \{\}$
for $x = 1$ *to* r **do** ▷(for each RP)
  $sw \leftarrow \lfloor n^{(x)} \div s \rfloor$ ▷# samples in a window, integer division
  RMP $\leftarrow \{\}$
  for $i = 1$ *to* m **do** ▷(for each MAC)
    $rmw \leftarrow \{\}$
    $dtm \leftarrow trm_{x,i}$ ▷(data at RP-x, MAC-i)
    for $w = 1$ *to* s **do** ▷(for each window)
      $fs \leftarrow nw \times (w-1) + 1$
      $data \leftarrow \{dtm[fs],\ dtm[fs+1], \ldots, dtm[fs+sw-1]\}$
      $sd \leftarrow \texttt{sort}(data);\ qts \leftarrow \{\frac{0.5}{sw}, \frac{1.5}{sw}, \ldots \frac{(sw-0.5)}{sw}\}$
      $id1 \leftarrow \texttt{arg min}(\texttt{qnts} - 0.15);\ \texttt{id2} \leftarrow \texttt{arg min}(\texttt{qnts} - 0.45)$
      $id3 \leftarrow \texttt{arg min}(\texttt{qnts} - 0.85);\ \texttt{cint} \leftarrow \{\texttt{sd[id1]}, \texttt{sd[id2]}, \texttt{sd[id3]}\}$
      $cfin \leftarrow \texttt{kmeans}(data, k, cint)$ ▷($cint$–initial cluster centers)
      $rmw \leftarrow \{rmw, cfin[2]\}^T$
    end for
    RMP $\leftarrow [\mathbf{RMP} \quad rmw]$
end for
  RM$_x \leftarrow$ **RMP**
  end for

3.5. Statistically Enhanced RSSI Clustering (SERC) Algorithm

The proposed data filtering algorithm–SERC, efficiently estimates the appropriate RSSI from multiple samples of real-time

varying data. The algorithm is an improved version of the RSSI Clustering (RC) data filter (Shi et al., 2020). Since the RC filtering algorithm is based on the computationally intensive k-means clustering, the proposed enhancements are specifically targeted to reduce computational cost with negligible accuracy compromise. The performance enhancements are manifested by smart execution of the clustering process. Fig. 3 shows the implementation scheme to incorporate the SERC algorithm within Wi-Fi fingerprinting based IPS.

We denote that the SERC algorithm filters out one RSSI value $\hat{\rho}_a$ (MAC-a) per MAC, from v consecutive received samples i.e, **TST**, which is defined in Eq. (6).

$$\mathbf{TST} = \begin{bmatrix} \rho_{1,1} & \rho_{1,2} & \cdots & \rho_{1,m} \\ \vdots & \vdots & \ldots & \vdots \\ \rho_{v,1} & \rho_{v,2} & \cdots & \rho_{v,m} \end{bmatrix} \tag{6}$$

where, $\rho_{a,b}$ is the RSSI of MAC-b in the a-th received test sample. The **TST** data are essentially any v consecutive samples within the building test data **TS**, excluding the TP labels. As suggested in (Shi et al., 2020), the value of v is considered in the range $10 \leqslant v \leqslant 60$. From Eq. (6), the data of any MAC-i form the vector $\mathbf{tstm_i} = \{\rho_{1,i}, \rho_{2,i}, \rho_{v,i}\}^T$. Firstly the standard deviation is calculated as per Eq. (7).

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^{v}(\rho_{j,i} - \mu_i)^2}{v}} \qquad 1 \leqslant i \leqslant m \tag{7}$$

where $\mu_i = 1/v\sum_{j=1}^{v}\rho_{j,i}$; $1 \leqslant i \leqslant m$. The samples of $\mathbf{tstm_i}$ are then sorted in ascending order, denoted as $\mathbf{sd}_i$. We define two parametric thresholds $\delta_{low}, \delta_{high}$ that control the execution of clustering process for filtering RSSI data. If σ_i is lower than δ_{low}, clustering is by-passed and μ_i is declared as filter output. Setting δ_{low} to a low value would signify that in case of low variations in RSSI, SERC simplifies to the mean filter. If σ_i is higher than δ_{high}, this suggests very high variations in RSSI. In this scenario, mean of four central samples in $\mathbf{sd}_i$ is declared the filter output. The clustering procedure is avoided in this case also, due to the fact that high variations and outliers would significantly degrade clustering performance. The clustering algorithm is executed when $\delta_{low} \leqslant \sigma_i \leqslant \delta_{high}$. Cluster initialization as adopted in (Shi et al., 2020) is followed. Three clusters with initial values of $\mathbf{cint} = \{\mathtt{min}(\mathbf{tstm_i}), \mu_i, \mathtt{max}(\mathbf{tstm_i})\}$ are used for clustering of data in $\mathbf{tstm_i}$. The output cluster centers are denoted as the vector **cf**, and $\mathbf{cf}(2)$ is taken as $\hat{\rho}_i$. The SERC output according to parametric conditions is also described in Eq. (8).

$$\hat{\rho}_i = \begin{cases} \mu_i & \sigma_i < \delta_{low} \\ \frac{1}{4}\sum_{j=c_s}^{c_s+3} sd_i(j) & \sigma_i > \delta_{high} \\ \mathbf{cf}(2) & \text{otherwise} \end{cases} \tag{8}$$

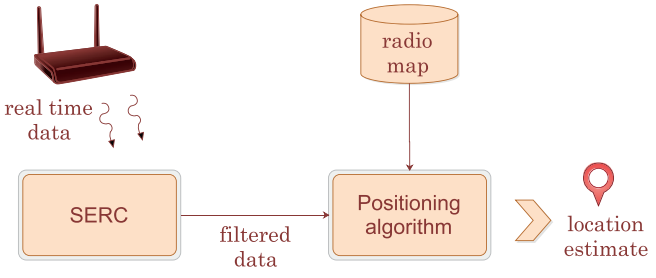

Fig. 3. SERC implementation within Wi-Fi fingerprinting based IPS.

where, $c_s = \lfloor 0.5(v-2) \rfloor$ and $\mathbf{cf}(2)$ is the center of middle cluster after application of k-means clustering with $k_m = 3$. It must be noted that generally the number of clusters in k-means algorithm is denoted by k. In this paper, we used k_m instead, in order to differentiate from the k parameter in k-Nearest Neighbors (kNN) algorithm, which was used as the benchmark positioning algorithm. In order to effectively implement SERC filtering in real-time tracking applications, the algorithm could be readily implemented by using moving windows of v latest RSSI samples. An initial (one time) delay, equivalent to acquisition time of v samples would be incurred to estimate the initial position. Subsequent filtered data would be available at each current sample. To ensure accurate tracking, the initial position is recommended to be estimated in static condition.

From the scalability perspective, utilizing the same v value would ensure consistent filtering performance and comparable computational requirement. In a large-scale environment, the total non-detected sources would significantly increase. Due to the smart clustering execution within SERC, the undetected sources would be efficiently catered with mean filtering. Essentially, in a scaled scenario the proportion of mean filter execution would be higher as compared to non-scaled conditions. Therefore, the scalability of SERC data filtering approach is duly asserted. The pseudocode of SERC is reported as Algorithm 2.

Algorithm 2. SERC algorithm pseudocode

Require
$m,\ \delta_{low},\ \delta_{high};\ m > 0,\ \delta_{low} > 0,\ \delta_{high} > 0$ ▷(Section 3.3)
TST $\leftarrow v$ consecutive test samples
▷(**TST** is a $v \times m$, *matrix*)
$k_m \leftarrow 3$
for $i = 1$ *to* m **do** ▷(for each MAC)
$\mathbf{tstm_i} = \{\rho_{1,i}, \rho_{2,i}, \ldots, \rho_{v,i}\}^T$
$sd_i \leftarrow \mathtt{sort}(\mathbf{tstm_i})$
$\sigma_i \leftarrow \mathtt{stdev}(\mathbf{tstm_i})$ ▷(Eq. (7))
if $\sigma_i < \delta_{low}$ **then**
$\hat{\rho}_i \leftarrow \mathtt{mean}(data)$
else
if $\sigma_i > \delta_{high}$ **then**
$c_s \leftarrow \lfloor 0.5(v-2) \rfloor$
$\hat{\rho}_i \leftarrow \frac{1}{4}\sum_{j=c_s}^{c_s+3} sd_i(j)$
else
$dc \leftarrow \mathtt{min(data)};\ \mathtt{nc} \leftarrow \mathtt{mean(data)};\ \mathtt{ic} \leftarrow \mathtt{max(data)}$
$cint \leftarrow \{dc, nc, ic\}$
$cf \leftarrow \mathtt{kmeans}(data, k, cint)$ ▷(*cint*–initial cluster centers)
$\hat{\rho}_i \leftarrow cf[2]$
end if
end if
end for

3.6. Dual Band Area Classification (DBAC) Algorithm

The proposed area classification algorithm–DBAC efficiently classifies real-time input samples to labeled sub-areas of the entire environment. The DBAC is implemented in offline and online stages. In the offline phase, area partitioning according to three strongest APs, is performed. In the online phase, RSSIs of dual-band APs are aggregated potential area partitions-wise. Among the partitions, the one with highest aggregated RSSI is classified as the final output. Consequently, the positioning algorithm has

to search only within the classified area partition, for final location estimation. The methodology to implement the proposed DBAC algorithm within Wi-Fi fingerprinting based IPS, is illustrated in Fig. 4.

In the offline stage, firstly the matrix **A** is formed. Using the radio map **RM**, the RSSIs within $rm_{x,i}$ are sorted in ascending order. The first three indices of the sorted data, (corresponding to AP/MAC number) are recorded in **A**, as given in Eq. (9) and (10).

$$\mathbf{idx}_i = \text{argsort}(\mathbf{rm}_{x,i}) \qquad 1 \leqslant x \leqslant r;\ 1 \leqslant i \leqslant s \tag{9}$$

$$\mathbf{A}_{x\times i} = \{\mathbf{idx}_i[1], \mathbf{idx}_i[2], \mathbf{idx}_i[3]\} \quad 1 \leqslant x \leqslant r;\ 1 \leqslant i \leqslant s \tag{10}$$

From **A**, all unique rows are found and recorded as **Ar**. The occurrence of each unique row is also calculated and stored in **cnt**. Using **Ar** the final matrix **AR** is extracted by applying the threshold δ_{ar} as shown in Eq. (11).

$$\mathbf{AR}_i = \begin{cases} \mathbf{Ar}_y & \mathbf{cnt}(y) > \delta_{ar} \\ \Phi \text{ (discard)} & \text{otherwise} \end{cases} \tag{11}$$

where Φ represents an empty vector, and signifies discarding of potential areas according to set threshold. Also, $1 \leqslant y \leqslant act$ and $1 \leqslant i \leqslant uct$. Essentially, **AR** groups locations at which the three strongest APs are the same, to a common labeled area. Leveraging three strongest APs has also been proposed in (Moreira et al., 2015). The DBAC area partitioning, additionally handles the dual-band Wi-Fi APs scenario. To incorporate the dual-band situation, APs are identified using MACs of both bands. Area partitioning using only the strongest AP (Xie et al., 2022b) suffers accuracy degradation in multipath conditions due to inefficient resolution of, locations at boundaries of multiple sub-areas. Using three highest RSSI APs rather than one, defines sub-areas with more distinct boundaries and reduces the possibility of nearby locations being grouped/classified to different areas (Le Dortz et al., 2012; Peng et al., 2020). Partitioning based on more than three APs, would further decrease classification of close proximity locations to different areas, at the cost of increased computations in the online phase (Obeidat et al., 2021).

In the online phase, the area classification of real-time input vector **ts**, is accomplished by aggregating the dual-band RSSIs according to **AR**. The aggregation is given in Eq. (12).

$$\mathbf{agg}_u = \sum_{z=1}^{3} \mathbf{ts}[\mathbf{AR}_u[z]] \qquad 1 \leqslant u \leqslant uct \tag{12}$$

The area corresponding to highest aggregated RSSI is declared as the classified area, as given in Eq. (13).

$$\phi_a = \arg\max_u \mathbf{agg}_u \quad \forall u = \{1, 2, 3, \ldots, uct\} \tag{13}$$

Since the three strongest APs based area partitioning, is aptly implementable under large-scale conditions, the formed area partition sizes would be comparable to the non-scaled scenario. Consequently, the required computations for the positioning algorithm would also be similar in both scenarios. Therefore, the scalability of DBAC algorithm is manifested. The pseudocode of DBAC is provided as Algorithm 3.

Algorithm 3. DBAC algorithm pseudocode

Require RM; δ_{ar}
  $\mathbf{A} \leftarrow \{\}$; $\mathbf{Ar} \leftarrow \{\}$; $\mathbf{AR} \leftarrow \{\}$; **agg** $\leftarrow \{\}$; **cnt** $\leftarrow \{\}$; $act \leftarrow 0$; $uct \leftarrow 0$
  procedure OFFLINE STAGE
    for $x = 1$ *to* r **do**
      for $i = 1$ *to* s **do**
        $\mathbf{idx}_i \leftarrow \texttt{argsort}(\mathbf{rm}_{x,i})$
        $\mathbf{A}_{x\times i} \leftarrow \{\mathbf{idx}_i[1], \mathbf{idx}_i[2], \mathbf{idx}_i[3]\}$
      end for
    end for
    for $p = 1$ *to* rs **do**
      for $q = 1$ *to* i **do**
        if $\mathbf{A}_p == \mathbf{A}_q$ **then**
          $cnt[q]++$; `break`
        end if
        if $p == q$ **then**
          $act++$; $\mathbf{Ar}_{act} \leftarrow \mathbf{A}_p$
        end if
      end for
    end for
    for $y = 1$ *to* act **do**
      if $\mathbf{cnt}[y] > \delta_{ar}$ **then**
        $uct++$; $\mathbf{AR}_{uct} \leftarrow \mathbf{Ar}_y$
      end if
    end for **end procedure**
  procedure ONLINE STAGE
    for $u = 1$ *to* uct **do**
      $\mathbf{agg}[u] \leftarrow \sum_{z=1}^{3} \mathbf{ts}[\mathbf{AR}_u[z]]$
      if $u > 1$ **then**
        if $\mathbf{agg}[u] > \mathbf{agg}[\phi_a]$ **then**
          $\phi_a \leftarrow u$
        end if
      else
        $\phi_a \leftarrow u$
      end if
    end for **end procedure**

4. Results and discussion

This section details the performance evaluations of proposed algorithms, in comparison to state-of-the-art benchmarks. To ensure comprehensive evaluations, performance was measured in terms of accuracy and computation time. Algorithms that are

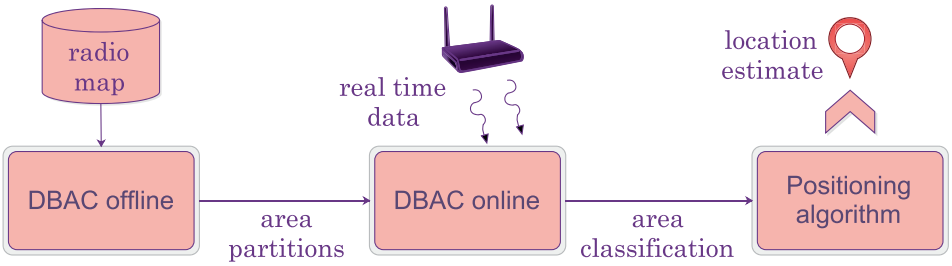

Fig. 4. DBAC implementation method within Wi-Fi fingerprinting based IPS.

implemented in offline phase of fingerprinting positioning were only evaluated by accuracy. The three metrics used for performance evaluations are described below.

i. Classification accuracy: The percentage of correct floor and wing classification by the positioning algorithm.
ii. Positioning error: The 75th percentile of the positioning errors in meters. The same metric was adopted in (Torres-Sospedra et al., 2018; Mendoza-Silva et al., 2018) and encouraged for fair comparisons.
iii. Computational time: The time in seconds taken for execution on the same computing platform.

The proposed algorithms were evaluated using data of two buildings CX1 and F04. Furthermore, device wise evaluations were also performed in order to fairly incorporate device heterogeneity scenario. Throughout this section, 'Device 1' and 'Device 2', refer to the scenario that only test samples and radio map of the corresponding device were used. 'Device 1&2' refers to the scenario that all test samples and merged radio map (i.e. device heterogeneity) was utilized. The performance evaluations were implemented in MATLAB® on a 3.6 GHz Intel® Core™ i7 work station with 16 GB RAM. The schematics depicted in Figs. 2–4, were implemented for the performance evaluations. The values of internal parameters used within implementation of each proposed algorithm, are reported in Table 2.

The parametric selections were determined by one of the following criteria/methods:

i. To ensure fair comparisons, parameter values as used in comparative benchmark algorithms were adopted. Therefore the values of v, k_m, and s were set to same values as in corresponding benchmark algorithms.
ii. Selection of values from a suitable range, that optimized performance in terms of accuracy and computational cost. Values of the thresholds $\delta_{ar}, \delta_{low}$, and δ_{high} were therefore determined by this method.

Table 2
Internal parameter values used for implementation and performance evaluation of proposed algorithms.

Algorithm	Parameter	Value
SERC	v	10
	k_m	3
	δ_{low}	1.5 dB
	δ_{high}	7.5 dB
SCRM	s	3
	k_m	3
DBAC	δ_{ar}	15

4.1. Benchmark positioning algorithm

The proposed fingerprinting algorithms were devised for performance optimization of indoor positioning systems, and had to be implemented along with a benchmark positioning algorithm. The algorithms were therefore implemented along with the k-Nearest Neighbors (kNN)-based positioning algorithm (Mendoza-Silva et al., 2018). It is widely used as a benchmark in fingerprinting IPS (Rojo et al., 2019; Retscher, 2020). Furthermore, the benchmark accuracy of UTMInDualSymFi dataset was also determined using the same algorithm. All internal parameters of kNN as adopted in (Abdullah et al., 2023), were replicated for the evaluations.

4.2. SCRM performance evaluation

The performance of SCRM algorithm was evaluated in comparison to the Dual-Band Radio Map (DBRM) construction technique proposed by (Ozdemir and Ceylan, 2020), and the performance baselines established of the dataset (Abdullah et al., 2023). Since radio map algorithms are implemented in the offline phase of fingerprinting IPS, the computational evaluations were not performed. The classification accuracy achieved with SCRM in CX1 and F04 buildings are respectively given in Tables 3 and 4.

The dataset baselines are reported for a conventional mean sampling algorithm. In building CX1 the baseline classification accuracy was 99% or higher. Despite the very high baseline, the performance of DBRM (Ozdemir and Ceylan, 2020) and SCRM generated radio maps were better than the baseline, across all values of k. Similarly in building F04, DBRM and SCRM recorded higher accuracy than the baseline, across all tested values of k for all device scenarios. Therefore, in statistical terms, classification performance of DBRM and SCRM was better than the established baseline. The recorded results also proved the marginal classification enhancement shown by SCRM in comparison to DBRM. Throughout the multi-scenario evaluations, corresponding to various buildings, devices, and k values, the classification improvement achieved by SCRM over DBRM was 0.1% to 1.7%. Evaluations from

Table 3
Classification accuracy using radio maps generated by SCRM algorithm in building CX1. kNN-based classification results are reported for various values of k.

Device	Classification accuracy (%)								
	$k = 3$			$k = 5$			$k = 7$		
	Mean	DBRM	SCRM	Mean	DBRM	SCRM	Mean	DBRM	SCRM
1	99.6	99.8	99.9	99.5	99.8	99.9	99.6	99.8	99.9
2	99.4	99.6	99.8	99.0	99.2	99.7	99.5	99.6	99.8
1 & 2	99.5	99.7	99.8	99.1	99.6	99.9	99.2	99.5	99.7

Table 4
Classification accuracy using radio maps generated by SCRM algorithm in building F04. kNN-based classification results are reported for various values of k.

Device	Classification accuracy (%)								
	$k = 3$			$k = 5$			$k = 7$		
	Mean	DBRM	SCRM	Mean	DBRM	SCRM	Mean	DBRM	SCRM
1	89.2	90.8	91.3	90.8	91.2	91.8	90.4	90.6	90.9
2	97.5	98.7	99.3	99.2	99.4	99.7	99.0	99.2	99.5
1 & 2	94.8	95.1	95.8	94.6	94.9	95.8	93.6	94.1	95.7

the perspective of positioning accuracy were also conducted. Comparative results of the 75th percentile of positioning errors in the two buildings are presented in Tables 5 and 6.

The reported results demonstrate the superiority of DBRM and SCRM constructed radio maps, in comparison to the baseline – mean sampling generated radio maps. DBRM and SCRM exhibited enhanced positioning accuracy as compared to the baseline, for all test scenarios including across buildings, devices, and k values. Combined with the classification results, the positioning accuracy findings reassert the statistical eminence of DBRM and SCRM over the baseline. In building CX1, SCRM generated radio maps showed 35% to 52% superior accuracy over DBRM. The average positioning improvement provided by SCRM as compared to DBRM in building F04 was 28%, across devices. Therefore, enhanced positioning shown by SCRM also signifies the algorithm's robustness in multi-building and device heterogeneity conditions. The positioning algorithm was run sample-by-sample to generate the results presented for the SCRM technique. Along with the construction of multi-sample (per location) radio maps with improved cluster initialization, dataset features like hotspot interference free dual-band data and high resolution RP spacing (1 m), also assisted achieving high accuracy performance. The performance of SCRM constructed radio maps is related to the deployed Wi-Fi network density. In less favorable scenarios as sparse network coverage conditions, SCRM constructed radio map could suffer from inadequate area coverage. The adverse impact would be catered by the enhanced network characterization, such that an adequate level of accuracy is maintained.

Overall, the SCRM algorithm contributed to significant positioning accuracy enhancement in comparison to benchmark techniques. It is therefore asserted that the introduction of enhanced statistical cluster initialization, within the proposed algorithm, adequately characterized RSSI variations in the resultant radio maps. The performance analyses of SCRM manifested robustness and efficiency, for implementation within multi-building, multi-device fingerprinting positioning applications.

Table 5
75th percentile of positioning errors using radio maps generated by SCRM algorithm in building CX1. kNN-based positioning results are reported for various values of k.

Device	75th percentile positioning error (m)								
	$k = 3$			$k = 5$			$k = 7$		
	Mean	DBRM	SCRM	Mean	DBRM	SCRM	Mean	DBRM	SCRM
1	2.23	2.05	1.24	2.43	2.21	1.58	2.65	2.58	1.61
2	4.51	3.43	1.78	4.64	3.55	1.69	4.65	3.71	1.82
1 & 2	2.25	2.52	1.76	2.32	2.16	1.67	2.71	2.60	1.85

Table 6
75th percentile of positioning errors using radio maps generated by SCRM algorithm in building F04. kNN-based positioning results are reported for various values of k.

Device	75th percentile positioning error (m)								
	$k = 3$			$k = 5$			$k = 7$		
	Mean	DBRM	SCRM	Mean	DBRM	SCRM	Mean	DBRM	SCRM
1	1.16	1.10	1.06	1.48	1.38	1.12	1.75	1.70	1.24
2	1.43	1.36	1.04	1.62	1.52	1.07	2.24	2.14	1.36
1 & 2	1.37	1.31	1.08	1.64	1.55	1.08	1.82	1.76	1.28

Table 7
Classification accuracy results of filtering algorithms in building CX1. Baseline accuracy corresponds to no filtering. kNN-based classification results are reported for various values of k.

Device	Classification accuracy (%)								
	$k = 3$			$k = 5$			$k = 7$		
	No filter	RC	SERC	No filter	RC	SERC	No filter	RC	SERC
1	99.1	99.5	99.3	99.1	99.5	99.4	99.2	99.7	99.5
2	99.0	99.5	99.3	99.1	99.7	99.4	99.2	99.6	99.4
1 & 2	99.2	99.9	99.8	98.9	99.9	99.8	99.5	99.9	99.7

Table 8
Classification accuracy results of filtering algorithms in building F04. Baseline accuracy corresponds to no filtering. kNN-based classification results are reported for various values of k.

Device	Classification accuracy (%)								
	$k = 3$			$k = 5$			$k = 7$		
	No filter	RC	SERC	No filter	RC	SERC	No filter	RC	SERC
1	93.3	98.3	97.9	93.7	98.5	98.2	93.2	97.9	97.2
2	98.3	98.7	98.4	98.2	98.5	98.4	98.0	98.4	98.2
1 & 2	97.5	99.9	99.3	97.4	99.9	99.5	97.5	99.9	99.6

4.3. SERC – Performance evaluation

This section reports the comparative performance of our proposed data filtering algorithm SERC, in terms of classification, positioning accuracy, and computational complexity; across multi-building and multi-device scenarios. Classification accuracy results for evaluated data filtering algorithms in buildings CX1 and F04 are reported in Tables 7 and 8 respectively.

Due to high baseline classification in building CX1, the filtering algorithms showed marginal improvements. In the case of building F04, the average accuracy recorded for SERC filtering was 2.5% higher than the baselines. As targeted, the performance of SERC was consistently comparable to the RC filtering algorithm (Shi et al., 2020). Results of the 75th percentile of positioning errors are presented for buildings CX1 and F04, in Tables 9 and 10 respectively.

Significant positioning accuracy enhancement by both data filtering algorithms is evident from the results of both buildings. The average accuracy enhancement provided by SERC in comparison to the baselines was around 37% and 38% respectively, in buildings CX1 and F04. Therefore, significant enhancement was achieved with SERC data filter in comparison to no filtering. Throughout the results, it is also evident that the accuracy of SERC and RC filter, was comparable. The computational time for execution of both filtering algorithms was also measured and are reported in Tables 11,12, for buildings CX1 and F04 respectively.

It is evident from the performance evaluation in both buildings that, the SERC on average took 42% less computation time for execution. Therefore, a significant enhancement in computational efficiency was achieved by SERC in comparison to the RC filtering algorithm (Shi et al., 2020). In order to comprehensively evaluate the performance of proposed SERC algorithm, the worst degradation in accuracy was also measured. Relative to the RC algorithm performance, the worst recorded degradation in 75th percentile errors were 9 cm and 8 cm respectively, for buildings CX1 and F04. From the comparative accuracy degradation values and computation time, it is clearly established that the significance of computational enhancement, outweighed the corresponding degraded accuracy. Therefore, SERC filtering provided high positioning accuracy, in much lesser execution time. All reported results for the SERC filter were determined by running the positioning algorithm on SERC filtered data from v consecutive samples. The availability of dual-band data at high resolution with network device homogeneity, in addition to smart clustering, assisted the achievement of performance improvements. The SERC filtering algorithm is designed to ensure efficient performance in beneficial and less favorable conditions. In unfavorable scenarios such as severe multipath fading or dynamic environment changes that cause large RSSI variations, SERC would provide typical performance. This is due to efficient outlier removal and utilization of suitable central samples from sorted RSSI data.

In summary, the proposed SERC algorithm showed pertinent performance enhancement across the multi-building and multi-device test scenarios. The proposed smart clustering within the SERC data filter, induced significant reduction in computational complexity, while maintaining accuracy, throughout the test cases. Therefore, viability of incorporating the proposed data filtering algorithm in real-time and scalable fingerprinting positioning solutions is manifested.

4.4. DBAC – Performance evaluation

In this section the performance evaluation of the proposed area classification algorithm DBAC is reported. In addition to compara-

Table 9
75th percentile of positioning errors for implemented filtering algorithms in building CX1. Baseline accuracy corresponds to no filtering. kNN-based classification results are reported for various values of k.

Device	75th percentile positioning error (m)								
	$k = 3$			$k = 5$			$k = 7$		
	No filter	RC	SERC	No filter	RC	SERC	No filter	RC	SERC
1	2.32	1.51	1.58	2.67	1.55	1.62	2.24	1.61	1.69
2	2.27	1.36	1.41	2.25	1.54	1.57	2.32	1.49	1.53
1 & 2	2.96	1.38	1.42	3.25	1.62	1.64	3.32	1.85	1.88

Table 10
75th percentile of positioning errors for implemented filtering algorithms in building F04. Baseline accuracy corresponds to no filtering. kNN-based classification results are reported for various values of k.

Device	75th percentile positioning error (m)								
	$k = 3$			$k = 5$			$k = 7$		
	No filter	RC	SERC	No filter	RC	SERC	No filter	RC	SERC
1	1.08	0.96	1.02	1.13	0.97	0.99	1.44	1.03	1.12
2	1.67	0.96	0.98	1.89	0.97	0.99	2.32	1.06	1.08
1 & 2	2.26	1.05	1.14	2.63	1.04	1.09	2.25	1.15	1.19

Table 11
Computation time measured for execution of data filtering algorithms in building CX1. Experiments were performed on the same work station and averaged over 100 repetitions.

Device	Total samples	Computation time (seconds)		SERC computational Improvement (%)	SERC accuracy degradation (m)
		RC	SERC		
1	24,194	178.2	100.6	43.5	0.08
2	24,082	176.1	101.7	42.2	0.05
1 & 2	48,216	356.6	207.6	41.7	0.04

Table 12
Computation time measured for execution of data filtering algorithms in building F04. Experiments were performed on the same work station and averaged over 100 repetitions.

Device	Total	Computation time (seconds)		SERC	SERC accuracy
	samples	RC	SERC	Improvement (%)	degradation (m)
1	20,610	151.6	87.3	42.4	0.09
2	20,723	150.7	85.4	43.3	0.04
1 & 2	41,333	305.3	179.8	41.1	0.09

Table 13
Classification accuracy using DBAC algorithm in building CX1. kNN-based classification results are reported for various values of k.

Device	Classification accuracy (%)					
	$k = 3$		$k = 5$		$k = 7$	
	MAX	DBAC	MAX	DBAC	MAX	DBAC
1	89.2	99.5	89.4	99.6	89.1	99.2
2	86.3	99.6	86.5	99.5	86.2	99.3
1 & 2	87.6	99.6	87.8	99.8	87.1	99.3

tive accuracy evaluations, computational analysis was also conducted for comprehensive assessments. The performance of DBAC algorithm was evaluated in comparison to the heap sort based algorithm, MAX (Xie et al., 2022b). The classification accuracy results with implementation of both algorithms are reported in Tables 13 and 14.

From the results of both buildings, the superior classification accuracy of DBAC is evident. The DBAC outperformed the MAX algorithm in all tested scenarios. In building CX1, DBAC provided a consistent 99% accuracy whereas the MAX algorithm consistently recorded below 90%. The DBAC performance was observed as, on average 13% superior than the MAX area classification. In building F04, the average classification enhancement was 16% in comparison to MAX performance. The vivid superiority of DBAC is also evident by the fact that throughout the test cases in building F04, the worst performance was still better than MAX algorithm performance. The area classification algorithms were also compared in terms of positioning accuracy. In Tables 15 and 16, comparisons of positioning accuracy results in both buildings are given.

Across different k values, both buildings and devices, DBAC showed improvement throughout in terms of positioning accuracy. More precisely, in building CX1 an average 6% enhancement was shown by DBAC in comparison to MAX algorithm. In building F04 the mean improvement was 7.5%. The mean of all 75th per-

Table 14
Classification accuracy using DBAC algorithm in building F04. kNN-based classification results are reported for various values of k.

Device	Classification accuracy (%)					
	$k = 3$		$k = 5$		$k = 7$	
	MAX	DBAC	MAX	DBAC	MAX	DBAC
1	86.9	96.1	86.5	96.4	90.4	91.5
2	77.4	99.2	77.8	99.4	77.3	99.1
1 & 2	82.2	94.5	82.6	94.8	82.1	94.1

Table 15
75th percentile of positioning errors for implemented area classification algorithms in building CX1. kNN-based classification results are reported for various values of k.

Device	75th percentile positioning error (m)					
	$k = 3$		$k = 5$		$k = 7$	
	MAX	DBAC	MAX	DBAC	MAX	DBAC
1	1.16	1.05	1.38	1.33	1.82	1.76
2	1.32	1.13	1.47	1.42	1.91	1.78
1 & 2	3.2	3.03	2.87	2.83	2.94	2.81

Table 16
75th percentile of positioning errors for implemented area classification algorithms in building F04. kNN-based classification results are reported for various values of k.

Device	75th percentile positioning error (m)					
	$k = 3$		$k = 5$		$k = 7$	
	MAX	DBAC	MAX	DBAC	MAX	DBAC
1	1.06	0.91	0.98	0.93	1.28	1.21
2	1.08	1.03	1.14	1.06	1.27	1.21
1 & 2	2.37	2.04	2.23	2.03	2.48	2.26

Table 17
Computation time measured for execution of area classification algorithms in building CX1. Experiments were performed on the same work station and averaged over 100 repetitions.

Device	Computation time, per sample (seconds)		DBAC
	MAX	DBAC	Improvement (%)
1	0.79	0.47	40.5
2	0.81	0.49	39.5
1 & 2	1.87	1.11	40.6

Table 18
Computation time measured for execution of area classification algorithms in building F04. Experiments were performed on the same work station and averaged over 100 repetitions.

Device	Computation time, per sample (seconds)		DBAC
	MAX	DBAC	Improvement (%)
1	0.58	0.35	39.6
2	0.56	0.34	39.2
1 & 2	1.35	0.79	41.4

centile errors in buildings CX1 and F04 was recorded as 1.85 meters and 1.35 meters respectively. For all test data scenarios, the positioning algorithm was run sample-by-sample to determine the findings reported for the DBAC technique. High accuracy results were achieved due to availability of dual-band data at high resolution, free of hotspot interference within the dataset, in addition to efficient partitioning and data aggregation. The computation times for execution of area classification algorithms were also measured and are reported in Tables 17,18.

The computational results across buildings and devices, clearly signify the superiority of proposed DBAC, as it consistently required 40% lesser computation time as compared to the MAX algorithm. It must be further noted, that higher recorded computation times for the multi-device scenario, was due to higher samples per location, than in the single device radio maps (Abdullah et al., 2022). Sparse Wi-Fi network coverage is a scenario, that would result in lower computational enhancement over conventional techniques, than as reported in achieved results. The efficient dual-band data aggregation would still maintain typical accuracy, under such an unfavorable condition.

Overall, the proposed area classification algorithm DBAC, showed performance enhancements in terms of classification, positioning, and computational efficiency. The targeted computationally efficient, high classification performance was achieved by DBAC as it showed significant improvement to state-of-the-art. The attained improvements were specifically reaped due to, incorporation of robust area partitioning and efficient dual-band RSSI aggregation, within the algorithm.

5. Conclusions

We proposed novel design approach for implementation of robust radio map construction, data filtering, and area classification algorithms in Wi-Fi fingerprinting based IPS. Efficient characterization of RSSI temporal variations was ensured by our radio map construction algorithm–SCRM. Apt estimation of real-time RSSI in multipath conditions, was efficiently accomplished by the data filtering algorithm–SERC. Reduced positioning algorithm search space, was aptly implemented in the proposed area classification algorithm–DBAC. To evaluate the performance, the proposed algorithms were implemented along with kNN-based positioning technique. The robustness of proposed algorithms was ensured by rigorous performance evaluations inclusive of heterogeneous positioning devices, dual-band Wi-Fi data, and multi-building test scenarios.

The conducted evaluations revealed the robustness and efficiency of SCRM, SERC, and DBAC algorithms, throughout the multi-building and multi-device investigations. According to the performance evaluations, the SERC and DBAC algorithms consistently showed a 40% reduction in computation time, across buildings. In addition to the computational enhancement, SERC provided positioning accuracy within 9 cm of benchmark data filtering performance. Along with computation improvement, DBAC algorithm also provided 14.5% superior accuracy, across buildings. The average positioning accuracy enhancement provided by SCRM was 36%. The performance evaluations indicate significant enhancements provided by proposed algorithms. The accuracy enhancement showed by SCRM generated radio maps, validates the algorithm's effective indoor environment characterization approach. The up to mark positioning accuracy provided by SERC, with significant execution speed enhancement, endorses the robust and efficient filtering approach in multipath scenario. Furthermore, the significant computational enhancement and encouraging accuracy improvement demonstrated by DBAC, manifests the efficacy of area classification approach. The reported results indicate that the proposed algorithms significantly improve accuracy and computational efficiency. Therefore, the algorithms are feasible for incorporation in high-precision real-time Wi-Fi fingerprinting based IPS. Furthermore, the modular design of the algorithms allows for flexibility in integrating certain subsets of the algorithms, depending on the specific IPS application and system requirements. Therefore, the fingerprinting algorithms offer a promising solution for enhancing the computational efficiency and accuracy of IPS, with potential applications in indoor navigation and asset tracking. The internal parameters used within the algorithms were selected based on statistical data analysis. Future work will include selection of algorithmic parameters using optimization algorithms, to enable seamless implementation across diverse datasets.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research study was supported by the Ministry of Higher Education (MOHE), Malaysia and Universiti Teknologi Malaysia (UTM), through the Collaborative Research Grant (CRG) under Grant R.J130000.7351.4B437.

References

Abdullah, A., Haris, M., Aziz, O.A., Rashid, R.A., Abdullah, A.S., 2022. UTMInDualSymFi: A dataset of dual-band Wi-Fi RSSI data in symmetric indoor environments [dataset]. https://zenodo.org/record/7260097.

Abdullah, A., Haris, M., Aziz, O.A., Rashid, R.A., Abdullah, A.S., 2023. UTMInDualSymFi: A dual-band Wi-Fi dataset for fingerprinting positioning in symmetric indoor environments. Data 8, 14. https://doi.org/10.3390/data8010014.

Bellavista-Parent, V., Torres-Sospedra, J., Perez-Navarro, A., 2021. New trends in indoor positioning based on wifi and machine learning: A systematic review. In: 2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN), IEEE. pp. 1–8.

Biswas, D., Barai, S., Sau, B., 2023. New rssi-fingerprinting-based smartphone localization system for indoor environments. Wirel. Netw. 29, 1281–1297. https://doi.org/10.1007/s11276-022-03188-2.

Brena, R.F., García-Vázquez, J.P., Galván-Tejada, C.E., Mu noz-Rodriguez, D., Vargas-Rosales, C., Fangmeyer, J., 2017. Evolution of indoor positioning technologies: A survey. J. Sens. 2017. https://doi.org/10.1155/2017/2630413.

Buntak, K., Kovačić, M., Mutavdžija, M., 2019. Internet of things and smart warehouses as the future of logistics. Teh. glas. 13, 248–253. https://doi.org/10.31803/tg-20190215200430.

Din, M.M., Jamil, N., Maniam, J., Mohamed, M.A., 2018. Review of indoor localization techniques. Int. J. Eng. Technol. (UAE) 7, 201–204. https://doi.org/10.14419/ijet.v7i3.12.16024.

Frankó, A., Vida, G., Varga, P., 2020. Reliable Identification Schemes for Asset and Production Tracking in Industry 4.0. Sensors 20, 3709. https://doi.org/10.3390/s20133709.

Guo, Y., Zheng, J., Zhu, W., Xiang, G., Di, S., 2021. iBeacon indoor positioning method combined with real-time anomaly rate to determine weight matrix. Sensors 21. https://doi.org/10.3390/s21010120.

Hayward, S., van Lopik, K., Hinde, C., West, A., 2022. A survey of indoor location technologies, techniques and applications in industry. Internet Things, 100608. https://doi.org/10.1016/j.iot.2022.100608.

Hu, J., Hu, C., 2023. A wifi indoor location tracking algorithm based on improved weighted k nearest neighbors and kalman filter. IEEE Access 11, 32907–32918. https://doi.org/10.1109/ACCESS.2023.3263583.

Huang, B., Xu, Z., Jia, B., Mao, G., 2019. An online radio map update scheme for WiFi fingerprint-based localization. IEEE Internet Things J. 6, 6909–6918. https://doi.org/10.1109/JIOT.2019.2912808.

Huang, B., Yang, R., Jia, B., Li, W., Mao, G., 2021. A theoretical analysis on sampling size in WiFi fingerprint-based localization. IEEE Trans. Veh. 70, 3599–3608. https://doi.org/10.1109/TVT.2021.3066380.

Jung, S.H., Moon, B.C., Han, D., 2017. Performance evaluation of radio map construction methods for Wi-Fi positioning systems. IEEE Trans. Intell. Transp. Syst. 18, 880–889. https://doi.org/10.1109/TITS.2016.2594479.

Koledoye, M.A., De Martini, D., Rigoni, S., Facchinetti, T., 2018. A comparison of RSSI filtering techniques for range-based localization. In: 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 761–767.

Krishnan, S.R., Seelamantula, C.S., 2013. On the selection of optimum Savitzky-Golay filters. IEEE Trans. Signal Process. 61, 380–391. https://doi.org/10.1109/TSP.2012.2225055.

Le Dortz, N., Gain, F., Zetterberg, P., 2012. Wifi fingerprint indoor positioning system using probability distribution comparison. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2301–2304. https://doi.org/10.1109/ICASSP.2012.6288374.

Lin, Q., Son, J., Shin, H., 2023. A self-learning mean optimization filter to improve bluetooth 5.1 AoA indoor positioning accuracy for ship environments. Inf. Sci. 35, 59–73. https://doi.org/10.1016/j.jksuci.2023.01.019.

Lohan, E.S., Torres-Sospedra, J., Gonzalez, A., 2021. WiFi RSS measurements in Tampere University multi- building campus, 2017 [Dataset]. Available online: https://zenodo.org/record/5174851.

Mendoza-Silva, G.M., Richter, P., Torres-Sospedra, J., et al., 2018. Long-term WiFi fingerprinting dataset for research on robust indoor positioning. Data 3, 3. https://doi.org/10.3390/data3010003.

Moreira, A., Nicolau, M.J., Meneses, F., Costa, A., 2015. Wi-fi fingerprinting in the real world - rtls@um at the evaal competition. In: 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–10. https://doi.org/10.1109/IPIN.2015.7346967.

Obeidat, H., Shuaieb, W., Obeidat, O., Abd-Alhameed, R., 2021. A review of indoor localization techniques and wireless technologies. Wirel. Pers. Commun. 119, 289–327. https://doi.org/10.1007/s11277-021-08209-5.

Osterrieder, P., Budde, L., Friedli, T., 2020. The smart factory as a key construct of industry 4.0: A systematic literature review. Int. J. Prod. Econ. 221, 107476. https://doi.org/10.1016/j.ijpe.2019.08.011.

Ozdemir, B.N., Ceylan, A., 2020. Constructing a precise radio map and application of indoor positioning with dual-frequency Wi-Fi fingerprinting method. Measurement 163, 107997. https://doi.org/10.1016/j.measurement.2020.107997.

Peng, X., Chen, R., Yu, K., Ye, F., Xue, W., 2020. An improved weighted k-nearest neighbor algorithm for indoor localization. Electronics 9. https://doi.org/10.3390/electronics9122117.

Retscher, G., 2020. Fundamental concepts and evolution of Wi-Fi user localization: An overview based on different case studies. Sensors 20, 5121. https://doi.org/10.3390/s20185121.

Rojo, J., Mendoza-Silva, G.M., Ristow Cidral, G., Laiapea, J., et al., 2019. Machine learning applied to Wi-Fi fingerprinting: The experiences of the ubiqum challenge. In: 2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–8.

Roy, P., Chowdhury, C., 2021. A survey of machine learning techniques for indoor localization and navigation systems. J. Intell. Robot. Syst. 101, 63. https://doi.org/10.1007/s10846-021-01327-z.

Roy, P., Chowdhury, C., 2022. A survey on ubiquitous WiFi-based indoor localization system for smartphone users from implementation perspectives. CCF Trans. Pervasive Comput. Interact. 4, 298–318.

Shang, S., Wang, L., 2022. Overview of wifi fingerprinting-based indoor positioning. IET Commun. 16, 725–733. https://doi.org/10.1049/cmu2.12386.

Shi, Y., Shi, W., Liu, X., Xiao, X., 2020. An RSSI classification and tracing algorithm to improve trilateration-based positioning. Sensors 20, 4244. https://doi.org/10.3390/s20154244.

Torres-Sospedra, J., Montoliu, R., Martínez-Usó, A., Arnau, T.J., Avariento, J.P., Benedito-Bordonau, M., Huerta, J., 2014. UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems [dataset]. Available online: https://archive.ics.uci.edu/ml/datasets/ujiindoorloc.

Torres-Sospedra, J., Jiménez, A., Moreira, A., et al., 2018. Off-line evaluation of mobile-centric indoor positioning systems: The experiences from the 2017 IPIN competition. Sensors 18, 487. https://doi.org/10.3390/s18020487.

Torres-Sospedra, J., Richter, P., Moreira, A., Mendoza-Silva, G.M., Lohan, E.S., Trilles, S., Matey-Sanz, M., Huerta, J., 2022. A comprehensive and reproducible comparison of clustering and optimization rules in wi-fi fingerprinting. IEEE Trans. Mob. Comput. 21, 769–782. https://doi.org/10.1109/TMC.2020.3017176.

Wang, B., Zhou, S., Liu, W., Mo, Y., 2015. Indoor localization based on curve fitting and location search using received signal strength. IEEE Trans. Ind. Electron. 62, 572–582. https://doi.org/10.1109/TIE.2014.2327595.

Wang, B., Liu, X., Yu, B., Jia, R., Gan, X., 2019. An improved WiFi positioning method based on fingerprint clustering and signal weighted euclidean distance. Sensors 19, 2300. https://doi.org/10.3390/s19102300.

Xie, Y., Wang, K., Huan, H., 2022a. Bpnn based indoor fingerprinting localization algorithm against environmental fluctuations. IEEE Sens. J. 22, 12002–12016. https://doi.org/10.1109/JSEN.2022.3172860.

Xie, Y., Wang, T., Xing, Z., Huan, H., Zhang, Y., Li, Y., 2022b. An improved indoor location algorithm based on back propagation neural network. Arab. J. Sci. Eng. 47, 13823–13835. https://doi.org/10.1007/s13369-021-06529-z.

Yang, B., Guo, L., Guo, R., Zhao, M., Zhao, T., 2020. A novel trilateration algorithm for RSSI-based indoor localization. IEEE Sens. J. 20, 8164–8172. https://doi.org/10.1109/JSEN.2020.2980966.

Yang, T., Cabani, A., Chafouk, H., 2021. A survey of recent indoor localization scenarios and methodologies. Sensors 21, 8086. https://doi.org/10.3390/s21238086.

Yiu, S., Dashti, M., Claussen, H., Perez-Cruz, F., 2017. Wireless rssi fingerprinting localization. Signal Process. 131, 235–244. https://doi.org/10.1016/j.sigpro.2016.07.005.

Zafari, F., Gkelias, A., Leung, K.K., 2019. A survey of indoor localization systems and technologies. IEEE Commun. Surv. 21, 2568–2599. https://doi.org/10.1109/COMST.2019.2911558.

Zhang, H., Liu, K., Jin, F., Feng, L., Lee, V., Ng, J., 2020. A scalable indoor localization algorithm based on distance fitting and fingerprint mapping in wi-fi environments. Neural. Comput. Appl. 32, 5131–5145. https://doi.org/10.1007/s00521-018-3961-8.

Zhang, L., Liu, K., Pan, Z., Pan, L., Gao, R., Zhang, Q., et al., 2023. An indoor unknown radio emitter positioning approach using improved rssd location fingerprinting. Int. J. Antennas Propag. 2023. https://doi.org/10.1155/2023/5462081.

Zou, H., Jin, M., Jiang, H., Xie, L., Spanos, C.J., 2017. WinIPS: WiFi-based non-intrusive indoor positioning system with online radio map construction and adaptation. IEEE Trans. Wirel. Commun. 16, 8118–8130. https://doi.org/10.1109/TWC.2017.2757472.