Computer Communications 157 (2020) 92–101

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

# A scalable fingerprint-based angle-of-arrival machine learning approach for cellular mobile radio localization

Robson D.A. Timoteo, Daniel C. Cunha *

*Centro de Informática (CIn) - Universidade Federal de Pernambuco (UFPE), Av. Jornalista Anibal Fernandes, s/n - 50740-560, Recife-PE, Brazil*

ARTICLE INFO

*Keywords:*
Wireless communications
Fingerprint-based localization
Angle-of-arrival
Machine learning

ABSTRACT

It is well-known that the performance of radio frequency fingerprinting is strongly dependent on the variation of the received strength signal indicator (RSSI), that can be caused by devices which came from different vendors (or even the same) and by changes of environment (from outdoor to indoor, for example). Given that, we propose an innovative and scalable fingerprint-based localization technique using the mobile user horizontal angle to estimate its position. Performance of traditional fingerprinting based on RSSI and of our proposal are compared under different scenarios (number of base stations), environments (outdoor, indoor, and indoor–outdoor), and presence/absence of noise. Numerical results show that the proposed method presents smaller error predictions for most of environments and scenarios investigated. In addition, our proposal is less sensitive to environment changing (from outdoor to indoor) and more stable when applied in indoor or outdoor environments, considering scenarios with fewer base stations (cellular networks in suburban or rural regions) and less data for training (less costly drive-tests performed by telecom operators).

## 1. Introduction

In recent years, advances in communication systems have contributed to powerful growth of using mobile phones to access a high variety of Internet services [1]. The demand for applications focusing on mobile devices has been increased across the globe. This kind of service has been applied to a huge diversity of mass-market applications, such as credit card transactions security [2], mobile location-based advertising [3], payment method based on location information [4], roadside assistance [5], and disease tracking [6]. In this context, the information about users localization has become essential for increasing the quality of such applications.

Concerning localization techniques, they can be classified according to the target environment, which can be indoor or outdoor. Each one has different challenges and constraints to reach its main objective that is to find the mobile user (MU). For example, the global positioning system (GPS) is one of the most widely used localization systems in the world [7,8], but it has limitations relative to indoor and outdoor environments. For indoor environments, GPS is inefficient due to attenuation caused by buildings. In case of outdoor positioning, since an unobstructed line-of-sight is necessary to GPS performs efficiently, it has its accuracy compromised by highly dense urban areas as well as by overcast weather conditions. Aside from that, GPS hardware is a power-hungry module in every mobile device, which leads to a high consumption of energy [9,10].

Due to the limitations of the GPS-based techniques, other alternatives, such as fingerprint-based location methods, have drawn considerable attention over the past few decades with the development of the cellular mobile radio localization methods [11]. Radio frequency (RF) fingerprinting is a well-established localization technique, based on received strength signal indicator (RSSI), that has got attention from research community [12–15] and has many advantages when compared to GPS-based methods. First, the low-power sensors incorporated in current mobile devices consume much lower power, even if operating regularly. Second, no additional hardware is needed in most of RF fingerprint-based techniques. Finally, one may wish to trade-off accuracy for energy efficiency according to the application prerequisites.

In spite of the benefits obtained by using radiolocalization, this kind of location technique has some challenges to be overcome. For example, the high variation of RSSI is a relevant limitation of RF fingerprint-based localization methods. Firstly, many mobile devices perform differently in respect to the mean signal strength, even those which came from the same vendor [16]. Next, considering a single mobile device, the RSSI can change meaningly when users move through different environments, for example, from outdoor to indoor [17,18]. This effect of the RSSI can have a huge impact on the localization accuracy due to the attenuation caused by the building walls [19]. Since about 80% of people spend most of their time indoors, this can be a serious drawback for any outdoor localization system [20]. Thus, the objective of this work is to propose an innovative and scalable

* Corresponding author.
*E-mail address:* dcunha@cin.ufpe.br (D.C. Cunha).

https://doi.org/10.1016/j.comcom.2020.04.014
Received 15 October 2019; Received in revised form 18 February 2020; Accepted 8 April 2020
Available online 13 April 2020

fingerprinting-based technique using horizontal angles obtained from the RF signals to estimate the MU position. Instead of using the absolute values of RSSI, our proposal employs the RSSI and propagation delay (PD) differences to obtain the MU horizontal angle. The main idea behind this proposal is that RSSI differences from base stations geographically positioned in the same place are similarly affected by the building attenuation or are not dependent on the device/vendor.

In order to acquire the MU horizontal angle, the proposed localization technique will employ a regressor. Since machine learning (ML) algorithms have been widely used to address positioning prediction in wireless networks [21–26], we decided to adopt a supervised ML algorithm to implement the regression. Considering that $K$-nearest neighbors ($K$-NN) algorithm has been receiving meaningful attention due to its performance regarding mobile localization systems [27–29], we adopt it as the ML algorithm in our experiments.

All things considered, this paper aims attention at a novel approach for RSSI-fingerprint-based localization and its comparison with the traditional fingerprinting technique based on RSSI. Numerical results show that the proposed method presents smaller error predictions for most of environments and scenarios investigated. In addition, our proposal is less sensitive to environment changing (from outdoor to indoor) and more stable when applied in indoor or outdoor environments, considering scenarios with fewer base stations (cellular networks in suburban or rural regions) and less data for training (less costly drive-tests performed by telecom operators). Finally, we compare the performance of our proposal to the Federal Communications Commission (FCC) benchmarks for network-based localization systems.

The remainder of this paper is organized as follows. Section 2 presents brief concepts about fingerprinting localization and the $K$-NN algorithm. In Section 3, we propose a novel fingerprint-based ML approach based on angle-of-arrival for mobile location. Numerical results and statistical analyses are accomplished in Section 4, while conclusions are drawn in Section 5.

## 2. Background

In this section, we provide sufficient background on fingerprinting localization techniques and machine learning focusing on $K$-NN algorithms.

### 2.1. Fingerprint-based positioning techniques

Fingerprinting (FP) can be seen as a group of positioning methods that explore detailed geographic maps of radio properties. These radio properties are used to create the correlation database (CDB), a data repository to store RF fingerprints. Each RF fingerprint, like a human fingerprint, must identify a unique geographical position [30]. In other words, radio properties are the primary keys of the CDB, being each one linked to a geographical location. To acquire this unique correspondence, the number of RF parameters should be highly enough [30]. Thus, an RF fingerprint can be composed by a broad variety of radio parameters, such as RSSI, time-of-arrival (ToA), angle-of-arrival (AoA), and time difference-of-arrival (TDoA).

There is an extensive range of fingerprinting methods and they generally have two phases named off-line and on-line [31]. In the first phase, the radio parameters mentioned above are measured. These measurements are about the mobile user (MU) location and are used to build the CDB. In this context, it is not possible to get measurements at all geographical positions. For example, in some cases, it is not possible to measure the RF parameters inside private areas, such as houses, hospitals, and schools. Therefore, a propagation model, a prediction algorithm, or even both, can be used to fulfill the creation of CDB regarding the unmeasured positions.

In the second phase, the CDB is used to infer the position of a searched MU. To do that, the ML algorithm tries to match the target fingerprint (from the sought MU) to a reference fingerprint that is stored in the CDB. The goal is to find, based on MU network parameters, the reference fingerprint which has the highest similarity to the target fingerprint. Finally, the estimated MU location will be the geographical position associated to the reference fingerprint at the CDB.

### 2.2. K-NN algorithm

ML can be seen as a data-driven approach, which can make accurate predictions by leveraging from past observations. Learning from data can be used in problems where we do not have an analytic solution, but using the data it is possible to construct an empirical model [32]. In this premise, learning from data, there are a huge number of algorithms which can be used. The choice will depend on the task at hand.

$K$-NN is a classifier that belongs to the family of instance-based learning algorithms [33]. In this family, training instances are stored and a new sample is predicted using the $K$-closest samples from the training set. This strategy uses a different approach when compared with other methods, such as artificial neural networks, which build a general description of the target function based on the training instances. So, in instance-based learning, generalization is only performed when a new instance is predicted.

$K$-NN algorithm can be utilized in regression or classification problems and it works as follows [34]: given a test instance $\mathbf{X}_i$, the first step is to find the $K$ closest training instances to $\mathbf{X}_i$; these are the neighbors of $\mathbf{X}_i$. More specifically, given that each instance is described by a $m$-dimensional feature vector $\mathbf{X} = [X_1, X_2, \ldots, X_m]$, the distance between two instances $\mathbf{X}_i$ and $\mathbf{X}_j$ is defined as $d(\mathbf{X}_i, \mathbf{X}_j)$, where:

$$d(\mathbf{X}_i, \mathbf{X}_j) = \sqrt{\sum_{r=1}^{m} (X_{ir} - X_{jr})^2} \tag{1}$$

It is worth to stress that different metrics can be used to calculate the distance between samples. In (1), the Euclidean distance, the most commonly used metric, is employed.

After calculating the $K$ neighbors of $\mathbf{X}_i$ using (1), the class of $\mathbf{X}_i$ is assigned as the most common class among its $K$ nearest neighbors for classification problems. However, for regression problems, the target value for $\mathbf{X}_i$ is the average of the values of its $K$ nearest neighbors, and is given by

$$\hat{f}(\mathbf{X}_i) \leftarrow \frac{\sum_{i=1}^{K} f(\mathbf{X}_i)}{K} \tag{2}$$

where $\mathbf{X}_i$ is a training set instance and $f(\mathbf{X}_i)$ is the target for $\mathbf{X}_i$.

An advantage of the $k$-NN algorithm is that there is no cost regarding the learning process and it is able to learn complex concepts by local approximation using a simple strategy.

## 3. Proposed method

In this section, we propose an FP-based AoA ML approach to infer the MU position. Before describing our proposal, it is necessary to define the concept of group of base stations (BSs), characterized as a set of $N_e$ BSs located at the same geographical coordinates (latitude and longitude). In this work, a group of BSs is denoted as $G$. Thus, $G_i, i = 1, 2, \ldots, N$ represents the $i$th group of BSs considered in our proposal.

The proposed method can be implemented in eight steps, which are described in Algorithm 1. These steps can be organized in two phases, named off-line and on-line. The off-line phase contains the steps 1 to 4, while the on-line phase includes the steps 5 to 8. In the off-line phase, there is no collection of RSSI and PD measurements regarding the sought MU.

The first step of the proposed technique is to define the localization area. We assume a regular grid model based on the boundaries established by the coverage region. Therefore, we obtain the final grid map splitting the coverage region into smaller square cells, whose side (in meters) is named *resolution*. Fig. 1 illustrates a localization area

**Algorithm 1:** Description of the proposed FP-based AoA ML positioning method.

1: Define the localization area (regular grid);
2: Collect the scanner measurements (RSSIs and PDs);
3: Train the ML algorithms to obtain the hypothesis functions $f_i(\cdot)$ for each group $G_i$ of BSs;
4: Build the CDB with horizontal angles and PD estimated for every cell in the regular grid;
5: Collect RSSI and PD from the sought MU to all BSs;
6: Given RSSI and PD, predict the horizontal angle $\widehat{\theta}_i$ for each group $G_i$ using $f_i(\cdot)$;
7: Apply CDB filtering by PD to create the reduced search space $S_r$;
8: Find the nearest point on the set $S_r$.

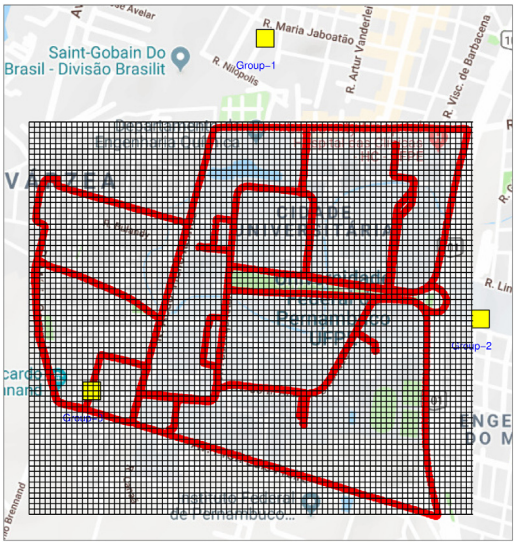


**Fig. 1.** Localization area represented by a 20 m-resolution regular grid ($q = 4000$ square cells) with the drive test path for sample collection (red points). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

represented by a regular grid with $q$ square cells, where each cell is associated with its center position (latitude and longitude). In our case, we consider a 20 m-resolution regular grid. It is worth to emphasize that the grid resolution can be a limiter of the technique's accuracy, since we assume that the MU is in the center of the square cell.

The second step consists of collecting RSSI and PD measurements in the localization area defined in the previous step. The PD can be seen as a round trip delay measurement and its value is an integer number from 0 to 56. These integer numbers represent a range of distances from MU to BS in steps of 234 m. Thus, the value 0 is mapped to the range $[0, 234[$ m, while the value 1 represents the interval $[234, 468[$ m, and so on. The maximum value of PD is 56, which represents distances greater than 13.1 km.

Assuming that $L$ field samples are collected in the localization area, each one consists of RSSI and PD measurements that are obtained for each BS and group of BSs, respectively. We denote as $\mathbf{P}_\ell$, the vector that represents the set of RSSI values (in dBm) for the $\ell$th sample, such that

$$\mathbf{P}_\ell = \{P_{\ell,i}^{(u)}\}, i = 1, \ldots, N; u = 1, \ldots, N_e, \quad (3)$$

where $P_{\ell,i}^{u}$ is the RSSI value for $u$th BS belonging to the $i$th group. In addition, we define $\mathbf{\Gamma}_\ell$ as the set of PD measurements such that

$$\mathbf{\Gamma}_\ell = \{\gamma_{\ell,i}\}, \quad i = 1, \ldots, N, \quad (4)$$

where $\gamma_{\ell,i}$ is the PD measurement for the $i$th group of BSs. For example, let us assume that there are $N = 2$ groups, each one with $N_e = 3$ BSs. In this scenario, each sample collected in a drive-test is composed of its geographical coordinates (latitude and longitude), a vector of RSSI measurements $\mathbf{P}_\ell$ given by

$$\mathbf{P}_\ell = \{P_{\ell,1}^{(1)}; P_{\ell,1}^{(2)}; P_{\ell,1}^{(3)}; P_{\ell,2}^{(1)}; P_{\ell,2}^{(2)}; P_{\ell,2}^{(3)}\}$$

and a set of PD measurements $\mathbf{\Gamma}_\ell = \{\gamma_{\ell,1}; \gamma_{\ell,2}\}$.

The next step is to train the ML algorithms, one for each group $G_i$, to predict the MU horizontal angle (target), denoted as $\widehat{\theta}_i$. It is worth to highlight that, due to the wide variation in AoA, a transformation function is required to smooth the target variable. A decreasing exponential function can be used for this task [35]. Thus, for AoA smoothing, the transformation function adopted, denoted as $t(x)$, is given by

$$t(x) = \frac{1}{1 + e^x}. \quad (5)$$

Fig. 2 shows the diagram of the process of training set building considering $N$ groups of BSs. Any regression algorithm can be trained, but the $k$-NN technique was exploited due to its good performance in localization problems [36,37]. The training set is created from data measurements (RSSI, PD, latitude, and longitude) collected in the field and fixed data (geographical coordinates of each group of BSs). To generate this set, a parser program was developed, using Python language, to extract and process the data.

Also in Fig. 2, we can see that the features of the training set are composed by RSSI and PD differences. Considering the $\ell$th sample collected in the field, the RSSI difference vector $\mathbf{\Delta}_\ell = \{\Delta_{\ell,i}^{(uj)}\}$ is calculated, being its elements given by

$$\Delta_{\ell,i}^{(uj)} = P_{\ell,i}^{(u)} - P_{\ell,i}^{(j)}, u, j = 1, 2, \ldots, N_e, u \neq j, u < j \quad (6)$$

where $P_{\ell,i}^{(u)}$ and $P_{\ell,i}^{(j)}$ are the RSSIs of the $u$th and the $j$th BSs, respectively, which belong to the group $G_i, i = 1, \ldots, N$. Thus, RSSI differences are calculated from BSs of the same group. In addition, we have the PD difference vector $\mathbf{T}_\ell$, that is also part of training data, and whose elements $\tau_\ell^{(mn)}$ are defined as

$$\tau_\ell^{(mn)} = \gamma_\ell^{(m)} - \gamma_\ell^{(n)}, m, n = 1, 2, \ldots, N; m \neq n, m < n, \quad (7)$$

where $\gamma_\ell^{(m)}$ and $\gamma_\ell^{(n)}$ are the smallest PDs related to the groups $G_m$ and $G_n$, respectively. It is noteworthy to emphasize that PD differences refer only to distinct groups of BSs. The smallest PD is considered since it is very likely that the BS with the smallest delay has the best line-of-sight in relation to the measuring point.

Consider again the previous example, where we have $N = 2$ groups of $N_e = 3$ BSs. In this scenario, we have the vector of RSSI differences

$$\mathbf{\Delta}_\ell = \{\Delta_{\ell,1}^{(12)}; \Delta_{\ell,1}^{(13)}; \Delta_{\ell,1}^{(23)}; \Delta_{\ell,2}^{(12)}; \Delta_{\ell,2}^{(13)}; \Delta_{\ell,2}^{(23)}\},$$

i.e., we have the number of combinations of size 2 from $N_e$ distinct elements for each group of BSs. In a similar way, we have a set of PD differences

$$\mathbf{T}_\ell = \{\tau_\ell^{(12)}; \tau_\ell^{(13)}; \tau_\ell^{(23)}\}$$

with different combinations of size 2 from $N$ elements.

At the end of the training set construction, the training data consist of $L$ RSSI differences vectors $\{\mathbf{\Delta}_\ell\}_{\ell=1}^{L}$ as well as $L$ PD differences ones $\{\mathbf{T}_\ell\}_{\ell=1}^{L}$. It is important to mention that the RSSI differences are used to allow indoor and outdoor localizations. The basic idea of using the RSSI difference, and not the absolute value directly, is that the attenuation suffered by the signal due to environment physical barriers will be approximately equal, since the RF signals have the same origin point and roughly the same frequencies [38].

Another relevant aspect is that the absolute value of the RSSI levels can vary depending on the handset vendor, and this variation can limit the accuracy of FP-based positioning systems [18]. However, the influence of this signal level variation decreases when the RSSI

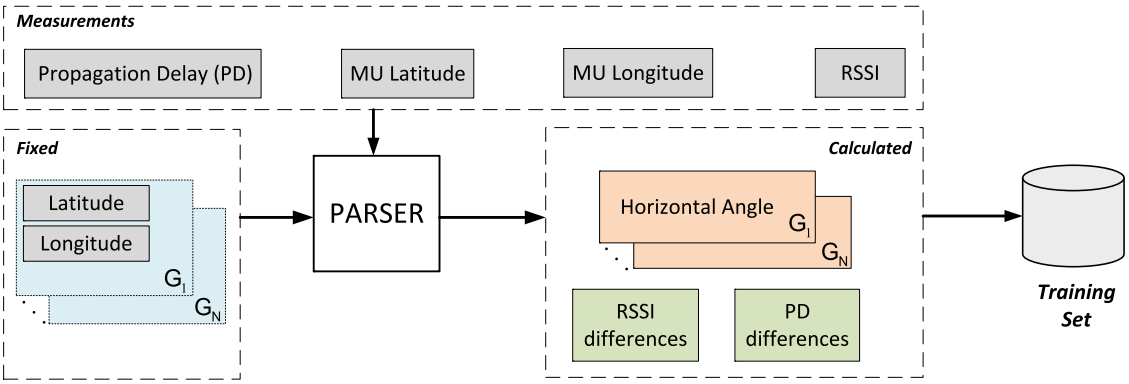


**Fig. 2.** Diagram of the training set building.

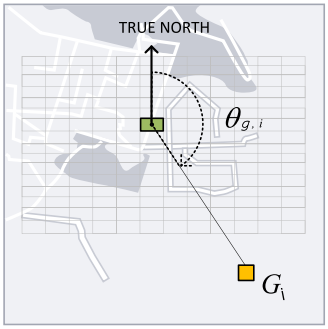


**Fig. 3.** CDB generation: azimuth $\theta_{g,i}$ of the $g$th square cell of the regular grid regarding the group $G_i$.

differences are assumed, since they remain almost constant in spite of the signal absolute level variation [39].

The fourth step of the proposed positioning method is the generation of the CDB associated to the localization area. The CDB is defined as the general search space given by the set $S$ of 3-tuples, such that

$$S = \{(\mathbf{p}_g, \Theta_g, \mathbf{t}_g) \in \mathbb{R}^2 \times \mathbb{R}^N \times \mathbb{R}^W,\ g = 1, 2, \ldots, q\}, \tag{8}$$

where the vector $\mathbf{p}_g = [p_g^{(a)}, p_g^{(b)}]$ represents the geographical coordinates of the $g$th square cell in the localization grid, being $p_g^{(a)}$ the longitude, and $p_g^{(b)}$ the latitude of the cell's center position. Besides that, the vector

$$\Theta_g = \{\theta_{g,i}\},\ i = 1, 2, \ldots, N,$$

represents the horizontal angles of the $g$th square cell regarding each of the $N$ groups. Fig. 3 shows how the horizontal angle $\theta_{g,i}$ is measured in a clockwise direction from the true North. Finally, the vector

$$\mathbf{T}_g = \{\tau_g^{(mn)}\},\ m, n = 1, 2, \ldots, N; m \neq n, m < n$$

represents the PD differences for each pair of groups of BSs, where the dimension of $\mathbf{T}_g$, denoted as $W$ in (8), is the combination of size 2 from $N$ distinct elements (groups of BSs).

The on-line phase starts with the fifth step, where the RSSI and PD measurements from the sought MU, denoted as

$$\mathbf{P}_{MU} = \{P_{MU,i}^u\},\ i = 1, \ldots, N; u = 1, \ldots, N_e$$

and

$$\mathbf{\Gamma}_{MU} = \{\gamma_{MU,i}\},\ i = 1, \ldots, N,$$

respectively, are acquired. The RSSI $P_{MU,i}^u$ is the estimated measure of power level that the MU is receiving from the $u$th BS belonging to the group $G_i$, while $\gamma_{MU,i}$ is the PD measurement for the $i$th group of BSs. Both measurements are obtained through the access network, as detailed in [30].

**Algorithm 2:** CDB Filtering to generate the reduced search space $S_r$

**Input:**
$\mathbf{T}_{MU}$, vector of MU PD differences
$S \leftarrow$ location area with $q$ squares
**Output:**
$S_r$, Reduced search space

```
1  begin
2      n_max ← maximum number of matches (all squares)
3      m_g ← NrMatches(T_MU, T_g),   g = 1, 2, ..., q
4      if n_max > 0 then
5          S_r ← squares where m_g = n_max
6          else
7              S_r ← S
8          end
9      end
10 end
11 return S_r
```

Given the MU measurements, the RSSI difference vector $\mathbf{\Delta}_{MU} = \{\Delta_{MU,i}^{(uj)}\}$ is obtained, whose elements are given by

$$\Delta_{MU,i}^{(uj)} = P_{MU,i}^{(u)} - P_{MU,i}^{(j)},\ u, j = 1, 2, \ldots, N_e, u \neq j, u < j, \tag{9}$$

while the PD difference vector $\mathbf{T}_{MU}$ has its elements $\tau_{MU}^{(mn)}$ given by

$$\tau_{MU}^{(mn)} = \gamma_{MU}^{(m)} - \gamma_{MU}^{(n)},\ m, n = 1, 2, \ldots, N, m \neq n, m < n. \tag{10}$$

After the calculation of $\mathbf{\Delta}_{MU}$ and $\mathbf{T}_{MU}$, these vectors are used as inputs of the hypothesis functions $f_i(\cdot)$ to predict $N$ MU horizontal angles, one for each group $G_i$, denoted as

$$\hat{\Theta}_{MU} = \{\hat{\theta}_{MU,i}\},\ i = 1, 2, \ldots, N.$$

After predicting the MU horizontal angles, the next step is to apply the Algorithm 2 for reducing the search area $S$ using CDB filtering [40]. The central idea of the Algorithm 2 is to maximize the number of matches between the expected and measured PDs. With this in mind and considering all square cells of the regular grid, the algorithm starts obtaining the maximum number of matches $n_{max}$ such that $\tau_{MU}^{(mn)} = \tau_g^{(mn)}$, where $\tau_{MU}^{(mn)}$ is the difference of measured PDs between the groups $G_m$ and $G_n$ and $\tau_g^{(mn)}$ is the expected PD difference between $G_m$ and $G_n$ for the $g$th square cell. After this, all square cells where the number of matches between the vectors $\mathbf{T}_{MU}$ and $\mathbf{T}_g$ equals to $n_{max}$ are included in the subset $S_r$ (line 5). If $n_{max} = 0$, there is no reduced space and the complete space $S$ is assigned to $S_r$ (line 7). In this case, the original localization area $S$ is returned.

The penultimate step of the Algorithm 1 is to estimate the MU position using the Euclidean distance as a similarity measure. The Euclidean

distance $d_g$ between the vectors $\hat{\Theta}_{MU}$ and $\Theta_g$ can be expressed as

$$d_g = \sqrt{\sum_{i=1}^{N} \left(\hat{\theta}_{MU,i} - \theta_{g,i}\right)^2}\ ,\ g = 1, 2, \ldots, q\ , \tag{11}$$

where $\hat{\theta}_{MU,i}$ is the MU horizontal angle predicted for the group $G_i$ and $\theta_{g,i}$ is the azimuth of $g$th square cell of the regular grid, regarding the group $G_i$. Finally, in the last step of the Algorithm 1, the best estimated position is the square cell which has the smallest Euclidean distance $d_g$, considering only the cells in the reduced space $S_r$.

## 4. Numerical results

In this work, the performance of our proposal is evaluated through computer simulations using the Python language, with emphasis on scikit-learn package [41].

Concerning the measurement setup, we assume mobile radio wave propagation measurements at 1.8 GHz third generation (3G) mobile network using W-CDMA air interface. By using an RF scanner, a drive test was performed to get the downlink pilot signal strength in an urban environment of approximately 1.6 km$^2$ located in the city of Recife-PE, Brazil. Red points in Fig. 1 indicate the positions where the measurements were captured by the RF scanner.

In order to compare the accuracy of the proposed algorithm with the RSS-based ones, two methods are implemented. The former is our proposal, which is described in Section 3 and is named as FP-AoA. The latter is the classical fingerprinting approach shown in [30], which is named as FP-RSSI. Regarding both methods, it is worth to mention two details about the implementation: (*i*) the $K$-NN algorithm is used to build the CDB; (*ii*) the procedure depicted in Algorithm 2 is employed to reduce the search space.

The two methods previously mentioned are deployed in three different scenarios related to the number of BSs in the urban area shown in Fig. 1. These scenarios are named as SCE-$N_b$, where $N_b$ represents the total number of BSs considered in the localization area. Fig. 4 illustrates all scenarios considered in this work. For example, Fig. 4(a) indicates that four BSs are used (two from the group $G_1$ and two from the group $G_2$). The gray cones represent the antennas, one per BS, with a horizontal angle of 63°. Besides that, for all groups, the azimuths of the antennas are 0°, 120°, and 240°. The azimuth 0° is indicated on Fig. 4(b). The other azimuths (120° and 240°) are measured in a clockwise direction from the azimuth 0°. Scenarios SCE-4 and SCE-9 follow the same azimuth orientation as scenario SCE-6.

Also in Fig. 4, we can see that various combinations of BSs for scenarios SCE-4 and SCE-6 are feasible. Let us assume the notation $G_i(A_u, A_j)$, where $A_u$ and $A_j$ indicate, respectively, the azimuths of the $u$th and $j$th BSs from the group $G_i$. For instance, in SCE-4, a possible combination of BSs is $[G_1(0°, 120°); G_3(0°, 120°)]$. However, the combination of BSs selected for SCE-4 in our experiment is $[G_1(120°, 240°); G_2(0°, 240°)]$. For SCE-6, the best combination considers all BSs from groups $G_1$ and $G_2$, while for SCE-9, all BSs from all groups represents the best configuration. The reason for choosing the settings previously mentioned (and shown in Fig. 4) is the fact that FP-RSSI method had the best tuning with those configurations. Due to the time-consuming tuning, the selection of BSs was done using only 50% of the outdoor samples, which were randomly chosen.

In order to train the ML models, 6007 external (outdoor) measurements were experimentally taken and divided into two subsets: a training dataset with 90% of the measurements and a test dataset with the remaining 10%. In addition, 3672 indoor measurements were taken to verify the accuracy of the positioning methods for users in an indoor environment. It is worth to emphasize that the indoor measurements are only used for testing stage.

To compare the localization methods FP-AoA and FP-RSSI, we define the distance prediction error $\eta$ as the difference (in meters) between the real and the predicted points. The validation technique denominated *cross-validation* is applied to check the generalization capacity of the localization methods. A widely used type of this technique is the $k$-fold cross-validation, a re-sampling technique in which the samples are randomly split into $k$ sets of approximately equal size [32]. These subsets are named folds and are divided into two groups: the test set with only one fold and the training set with $(k-1)$ folds. Algorithm 3 shows the pseudo-code of $k$-fold cross-validation. In this paper, we consider $k = 10$ and the average localization error $\bar{\eta}$ defined as

$$\bar{\eta} = \sqrt{\frac{1}{k}\sum_{j=1}^{k} \eta_j}\ , \tag{12}$$

in which $\eta_j$ is the localization error calculated for the $j$th test set (fold).

Regarding the ML model tuning, the 10-*cross-validation* technique also was applied to find the best value for the hyper-parameter $K$ of the $K$-NN regressors employed in the third step of the Algorithm 1. In this case, the parameter $K$ ($K$ is integer) was searched in the interval $[1; 11]$. The same interval was used for the regressors of the FP-RSSI technique. It is worth to highlight that, in case of FP-RSSI technique, each BS has an associated regressor, whereas, in case of FP-AoA method, each group of BSs has a regressor. As an example, Table 1 shows the best values for the hyper-parameter $K$ of the regressors obtained for the tenth fold, considering both localization techniques and all scenarios specified in this work. Due to space limitation, the best values for the hyper-parameter $K$ for the other nine folds have been omitted.

**Table 1**
Best values for the hyper-parameter $K$ of the regressors obtained for the tenth fold, considering both localization techniques and all scenarios specified in this work.

| Pos. Tech. | Scenario | Regressor ID | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| FP-RSSI | SCE-4 | 6 | 6 | 6 | 5 | – | – | – | – | – |
| | SCE-6 | 6 | 6 | 6 | 6 | 5 | 6 | – | – | – |
| | SCE-9 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 5 |
| FP-AoA | SCE-4 | 6 | 6 | – | – | – | – | – | – | – |
| | SCE-6 | 5 | 1 | – | – | – | – | – | – | – |
| | SCE-9 | 4 | 6 | 1 | – | – | – | – | – | – |

**Algorithm 3:** Description of $k$-fold cross-validation.

```
Partition data into k disjoint sets Z_1 ... Z_k
for j = 1 to k do
    Use Z_j for validation and the remaining for training.
    Calculate error on validation set Z_j.
end
Return average error on validation sets.
```

Besides the generalization capacity, another important aspect is the behavior of the localization method in the presence of outliers. These outliers can be generated by the shadow fading effects due to environment changes, such as modifications in vegetation, new buildings, weather conditions and/or vehicles of different sizes that move around the environment. In this article, we call this shadowing effect as noise and assume it follows a Rayleigh distribution with average of 12 dB and standard deviation equal to 7 dB, to simulate different levels of noise in the measurements from the sought MU [42]. In other words, the *noisy* environment mentioned in this paper refers to the Rayleigh-modeled noise added to the drive-test data. On the other hand, the *noise-free* environment considers only the drive-test data collected for our experiments.

Table 2 shows the average of distance prediction error (in meters), defined in (12), for FP-AoA and FP-RSSI localization methods regarding the 10-fold *cross-validation*. It is important to highlight that the smaller the prediction error, the better the accuracy of the localization method. We assume three MU noisy and noise-free environments where the

**Table 2**
Average distance prediction error $\bar{\eta}$ (in meters) for FP-AoA and FP-RSSI localization methods in three mobile user noisy and noise-free environments (outdoor, indoor, and indoor–outdoor) as well as three base station scenarios, regarding the 10-fold cross-validation.

| MU environment | Pos. Tech. | SCE-4 | | SCE-6 | | SCE-9 | |
|---|---|---|---|---|---|---|---|
| | | (noise-free) | (noisy) | (noise-free) | (noisy) | (noise-free) | (noisy) |
| Indoor–Outdoor | FP-AoA | 107.02 | 127.51 | 102.51 | 140.22 | 71.91 | 71.91 |
| | FP-RSSI | 127.94 | 150.22 | 129.58 | 150.41 | 84.74 | 84.74 |
| Outdoor | FP-AoA | 85.17 | 115.78 | 82.94 | 132.75 | 57.78 | 57.78 |
| | FP-RSSI | 96.45 | 162.41 | 91.45 | 165.14 | 47.65 | 47.65 |
| Indoor | FP-AoA | 117.12 | 126.08 | 111.55 | 143.68 | 78.45 | 78.45 |
| | FP-RSSI | 142.50 | 144.58 | 147.08 | 143.60 | 101.88 | 101.88 |

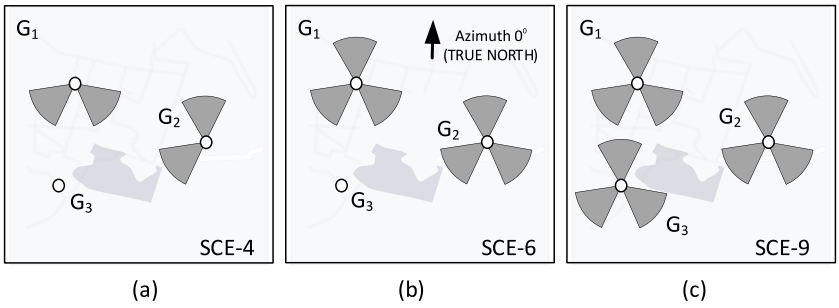


**Fig. 4.** Base stations considered for each scenario: (a) SCE-4, (b) SCE-6, and (c) SCE-9.

searched MU is located: (*i*) *outdoor*, when the MU is only outdoors; (*ii*) *indoor*, when the MU is only indoors; and (*iii*) *indoor–outdoor*, when the MU can be outdoors or indoors.

Firstly, we see that, in most of cases, FP-AoA overcomes FP-RSSI, except for indoor SCE-6 and outdoor SCE-9. Next, we can notice that the accuracy of FP-AoA and FP-RSSI methods becomes better with the increasing of the number of BSs. For example, considering only the noise-free outdoor environment, FP-AoA has a better performance in scenarios SCE-4 and SCE-6. For SCE-4 and SCE-6, FP-AoA provides a prediction error of 85.17 m and 82.94 m, respectively, while FP-RSSI reaches $\bar{\eta} = 96.45$ m for SCE-4 and $\bar{\eta} = 91.95$ m for SCE-6. In this way, we can see that the change of scenario (from SCE-4 to SCE-6) yields at most a prediction error difference of 4.5 m when FP-RSSI method is applied. On the other hand, when nine BSs (SCE-9) are considered, FP-RSSI method overcomes FP-AoA one, with average localization errors of $\bar{\eta} = 47.65$ m for the former, and $\bar{\eta} = 57.78$ m for the latter.

Also in Table 2, we notice that both localization methods have their accuracy compromised in both scenarios (SCE-4 and SCE-6) when noise is assumed. For FP-AoA, the prediction error increases 35.9% (85.17 m to 115.78 m) for SCE-4, and 60% (82.94 m to 132.75 m) for SCE-6. On the other hand, the increased prediction error is even higher for FP-RSSI, where we have an increment of 68.4% (96.45 m to 162.41 m) and of 80.6% (91.45 m to 165.14 m) for SCE-4 and SCE-6, respectively. Hence, in outdoor environments for SCE-4 and SCE-6, the FP-RSSI localization method is more sensitive to noise than the FP-AoA one. Nonetheless, when SCE-9 is assumed, both methods are not noise-sensitive. This characteristic can be explained by analyzing the behavior of Algorithm 2. This algorithm is in charge of reducing the search area using PD differences. Thus, a high number of BSs implies a more restrictive filter and, consequently, a smaller reduced area.

Considering only the indoor environment, we see that, in most of cases, the prediction error of the FP-AoA is smaller than the FP-RSSI in a range of 12.8% to 24.6%, according to Table 2. In this case, it is worth to emphasize that the FP-AoA is an outdoor localization method which can be used even if the MU gets into an indoor environment. In other words, in spite of FP-AoA being originally designed for outdoor localization, we can apply it to predict the MU position (not exactly) inside an indoor environment.

As we expected, the performance of both localization techniques get worse when the searched MU moves from outdoors to indoors. In view of this, Fig. 5 shows the increasing of prediction error (in percentage) for FP-AoA and FP-RSSI localization methods in all noisy and noise-free environments, when the MU gets into indoors. We can see that the FP-RSSI prediction error has higher increments than the FP-AoA one for all scenarios, where the highest one occurs in noise-free SCE-6. In this case, the FP-RSSI method has a increase of 60.83% (from 91.45 m to 147.08 m). Also in Fig. 5, we notice that error increments are lower in noisy environments, except for SCE-9. It does not mean that localization methods work better in a noisy environment, but that the search reduction becomes more relevant for both FP-AoA and FP-RSSI methods. Hence, the Algorithm 2 plays an essential role in noisy

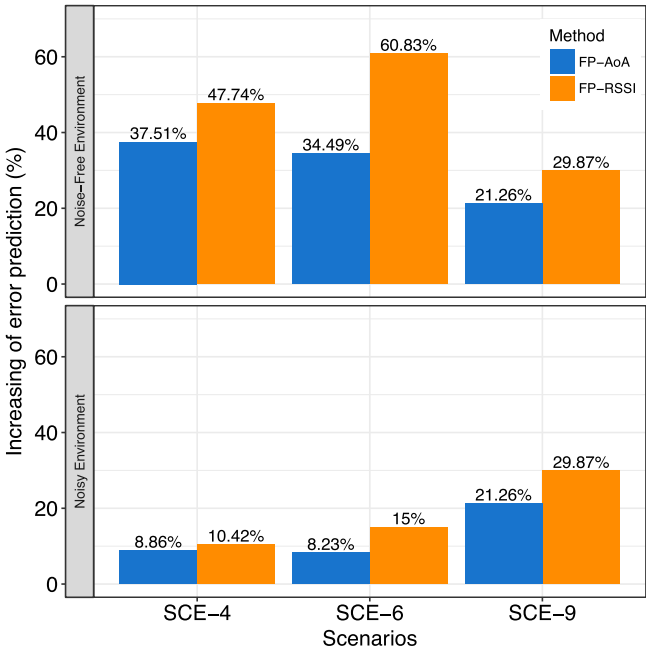


**Fig. 5.** Increasing of average distance prediction error (in percentage) for FP-AoA and FP-RSSI methods in all noisy and noise-free environments, when the searched MU moves from outdoors to indoors.

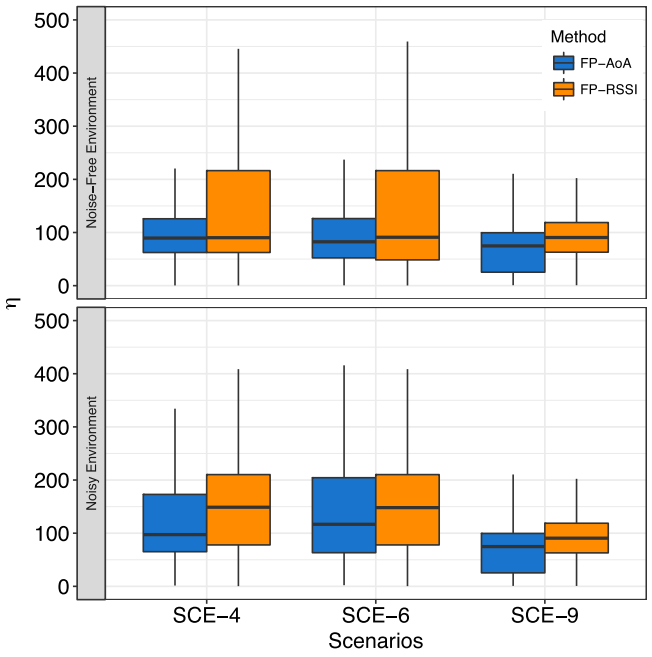


**Fig. 6.** Distribution of the distance error prediction $\eta$ (in meters) for FP-AoA and FP-RSSI methods in noisy and noise-free indoor–outdoor environments for all scenarios.

environments, and the type of environment changing has a smaller effect on it, because the Algorithm 2 relies on group PDs.

Concerning prediction algorithms (localization methods included), it is very important to know the full range of the error variation. Fig. 6 shows the distribution of the distance error prediction $\eta$ for FP-AoA and FP-RSSI methods in noisy and noise-free indoor–outdoor environments for all scenarios using box plots. Due to the 10-cross-validation technique, all instances (indoor and outdoor samples) are used as a test instance in one of the ten folds. In the box plot, an important aspect is the distribution range (box height) of the prediction errors. The narrower the range the better the stability of the localization technique. In Fig. 6, we see, for both localization methods, a narrower range for scenario SCE-9. Hence, the predictions for both algorithms are more stable when we compare SCE-9 to the other scenarios. In addition, we observe that FP-RSSI predictions are more steady than the FP-AoA ones. On the other hand, for scenarios SCE-4 and SCE-6, the ranges of FP-AoA predictions are tighter than the FP-RSSI ones in noise-free environments. Concerning the noisy environments for SCE-4 and SCE-6, we can assume that the localization methods are equivalent in terms of stability.

Another important aspect to analyze is the generalization capacity of the localization methods. In this paper, as previously explained, the 10-cross-validation technique is used to check the generalization capacity of FP-AoA and FP-RSSI methods. Thus, we can verify the performance of each localization approach when facing different datasets. Fig. 7 illustrates the distance prediction error $\eta_j$ for both localization methods, considering all noisy and noise-free environments (outdoor, indoor, and indoor–outdoor) and scenarios (SCE-4, SCE-6, and SCE-9) per each fold. FP-AoA is represented by the blue lines, while FP-RSSI by the orange ones. Regarding the presence/absence of noise, dashed lines represent noisy environments, whereas solid lines illustrate noise-free ones. Given a scenario, an environment, and a specific localization method, the noise sensitivity represents the variation of accuracy due to the presence of noise. Graphically, the noise sensitivity can be seen as the distance between solid and dashed lines corresponding to the same method. The smaller the distance between curves, the more noise-insensitive the localization method. Thus, both localization methods are noise-insensitive in all environments (indoor, outdoor, and indoor–outdoor), as we can observe at Fig. 7 (last row in the chart grid), since solid and dashed lines are overlapping.

For an outdoor environment in the scenario SCE-9, we see that the prediction error is smaller for FP-RSSI method in noisy and noise-free conditions. When we move from outdoor SCE-9 to the other scenarios, we notice that the FP-AoA method presents a lower prediction error, specially in scenario SCE-4, where FP-AoA has a higher accuracy for noisy and noise-free environments. In its turn, when the user changes its environment (from outdoor to indoor or to indoor–outdoor), in general, the FP-AoA exhibits a lower prediction error than the FP-RSSI for noisy and noise-free circumstances.

According to the FCC, for network-based localization systems, any MU that makes emergency 911 (E-911) calls requires a location accuracy to within 100 and 300 m for 67 and 90% of the cases, respectively [43]. Given that, Fig. 8 represents Cumulative distribution function of the distance prediction error for both localization methods, considering all noisy and noise-free indoor–outdoor environments for scenarios SCE-4 (left side), SCE-6 (center), and SCE-9 (right side). The two horizontal dashed lines represent the percentiles 67% and 90%, while the gray region in each graph represents the area outside of E-911 location accuracy requirements. We can see that only the noisy FP-AoA method (blue line in the right side graph) surpass the E911 accuracy specifications. In spite of that, we believe that other ML algorithms can be used to improve the performance of the both localization methods.

Another important aspect of any localization system is its scalability, specially with the advancement of Big Data technology in next-generation cellular networks [44]. Given that, we use the processing times spent in the model training and prediction stages to analyze the scalability of the methods. Fig. 9 shows the average normalized training runtime for FP-AoA and FP-RSSI methods in all noise-free indoor–outdoor scenarios using single-core and multi-core processing environments. Single and multi-core frameworks are examined to verify the parallelism ability of each localization technique. We assume the FP-RSSI method for scenario SCE-4 as a benchmark. As the prediction runtimes were almost the same for both localization methods (a variation less than 2%), we concentrate our analysis only on the training runtimes.

For scenario SCE-4, we see that the FP-AoA method has a training runtime 11.83% smaller than the benchmark for a single-core framework. When a multi-core framework is assumed, the reduction of training runtime reaches 39.25%. This fact combined with the performance improvement (smaller error prediction) give relevant advantages to the FP-AoA method when applied in less dense cellular environments, such as rural and suburban cellular networks.

Concerning the single-core framework, we can see that the training runtime for the FP-AoA increases exponentially when we move from SCE-4 to SCE-9. This behavior is explained by the increase in the number of features affecting the $K$-NN runtime. At the same time, when the scenario SCE-9 is assumed, a new group of BSs is included, which causes the FP-AoA method to calculate an additional horizontal angle for the localization grid. Regarding the FP-RSSI method, the increase in the training runtime is less intensive than the FP-AoA one.

In the multi-core framework (bottom of Fig. 9) and considering the FP-AoA method, we see that the training runtime growth rate due to the addition of BSs is smoother when compared to the single-core case. The main reason for this behavior is that the localization grid, shown in Fig. 1, can be built independently of the regressors models for the FP-AoA method. Accordingly, it is possible to parallelize both the ML model training tasks and the localization grid formulation, which favors the scalability for next-generation networks strongly dependent on massive amount of data. In addition, the parallelism can be a huge advantage for the application of automated machine learning (Auto-ML) [45]. Auto-ML frameworks can search for the best ML models automatically without re-build the localization grid for each tested model.

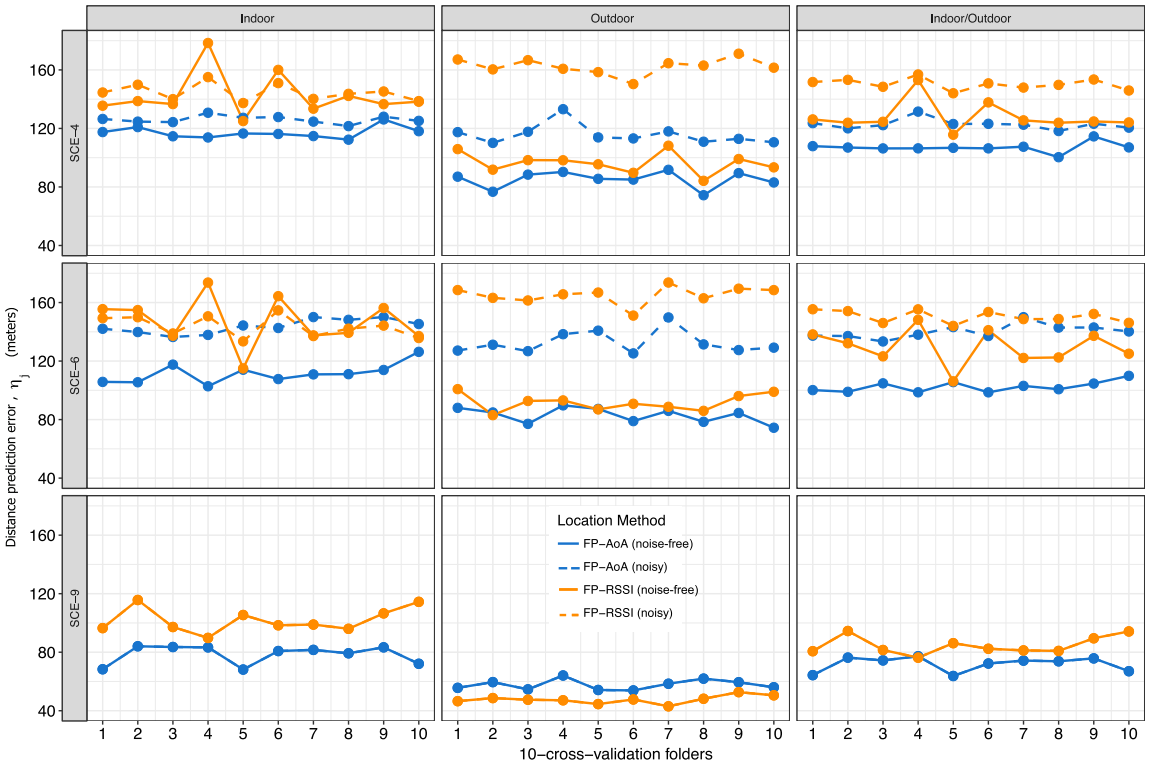


**Fig. 7.** Distance prediction error (in meters) for both localization methods, considering all noisy and noise-free environments (outdoor, indoor, and indoor–outdoor) and scenarios (SCE-4, SCE-6, and SCE-9) per each fold. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

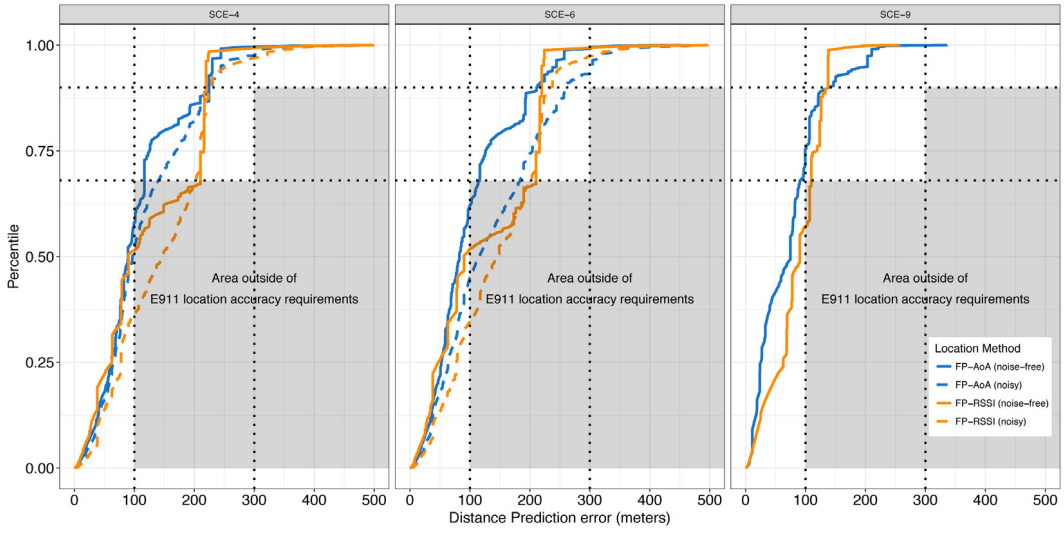


**Fig. 8.** Cumulative distribution function of the distance prediction error for both localization methods, considering all noisy and noise-free indoor–outdoor environments for all scenarios (SCE-4, SCE-6, and SCE-9). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Finally, to verify if the differences between the accuracy of the localization methods are statistically relevant, the Kruskal–Wallis non-parametric rank test is applied, which extends the Wilcoxon test for more than two samples [46]. The null hypothesis $H_0$ is that where all methods are equivalent and the alternative hypothesis $H_1$ is that where there are some difference between the methods. For the Kruskal–Wallis test performed over all dataset with a confidence level $\alpha = 0.05$, the $p$-value obtained was $2.2 \cdot 10^{-16}$. Thus, as $p \ll \alpha$, the hypothesis $H_0$ can be rejected and at least two approaches differ.

From the output of the Kruskal–Wallis test, we know that there are differences between the localization methods, but it is not possible to know which pairs of the methods are different. In this case, we can use the Wilcoxon rank sum test with corrections for multiple testing with $\alpha = 0.05$ to make pairwise comparisons [47]. It was obtained an adjusted $p$-value (using [47]) smaller than $\alpha$ for all pairs, except for the pair FP-RSSI SCE-4 and FP-RSSI SCE-6, with an adjusted $p$-value equals to 0.535. This $p$-value means that the FP-RSSI methods previously mentioned are statistically equivalent.

## 5. Conclusions

In this work, an innovative fingerprint-based localization technique using horizontal angles obtained from the RF signals was proposed to estimate the mobile user position. The main motivation behind our

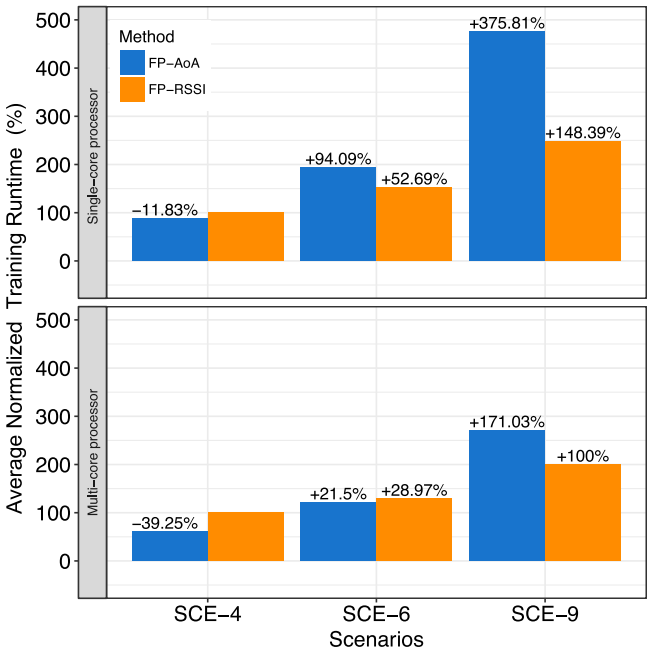


**Fig. 9.** Average normalized training runtime for FP-AoA and FP-RSSI localization techniques in noise-free indoor–outdoor environments for all scenarios. The FP-RSSI method for the scenario SCE-4 is assumed as a benchmark.

technique was to introduce a localization method designed for outdoor environments, but that can be applied to predict the user position inside an indoor environment. Different from the traditional fingerprint approach, that is based on the absolute values of received strength signal indicator (RSSI), our proposal employed RSSI and propagation delay differences to obtain the mobile user horizontal angle. The traditional fingerprint approach was also implemented for comparison. In both methods, mobile radio wave propagation measurements at a carrier frequency of 1.8 GHz W-CDMA were obtained in an urban environment in the city of Recife-PE, Brazil. Numerical results showed that our proposal presented a smaller average distance error prediction for all environments and scenarios investigated, except in noisy and noise-free outdoor environments for a scenario with nine base stations. In addition, the proposed method was less sensitive to environment changing (from outdoor to indoor) and more stable when applied to indoor or indoor–outdoor environments, considering scenarios with fewer base stations (cellular networks in suburban or rural regions) and less data for training (less costly drive-tests performed by telecom operators). Finally, the fingerprint-based angle-of-arrival method proposed in this work allowed the parallelization of both the machine learning model training and the localization grid formulation, which can be interesting to promote scalability in ultra-dense networks, such as 5G cellular systems.

## 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.

## CRediT authorship contribution statement

**Robson D.A. Timoteo:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft. **Daniel C. Cunha:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Supervision.

## References

[1] P. Deville, et al., Dynamic population mapping using mobile phone data, Proc. Nat. Acad. Sci. 111 (45) (2014) 15888–15893.
[2] A. Hammad, P. Faith, Location based authentication, U.S. Patent 10163100, 2017. Available at http://www.freepatentsonline.com/10163100.html.
[3] A. Gutierrez, S. O'Leary, N.P. Rana, Y.K. Dwivedi, T. Calle, Using privacy calculus theory to explore entrepreneurial directions in mobile location-based advertising: Identifying intrusiveness as the critical risk factor, Comput. Hum. Behav. 95 (2019) 295–306.
[4] A. Salomon, K.P. Mahaffey, Mobile communications device payment method utilizing location information, U.S. Patent 20190130382, 2019. Available at http://www.freepatentsonline.com/y2019/0130382.html.
[5] J. Trogh, D. Plets, E. Surewaard, et al., Outdoor location tracking of mobile devices in cellular networks, EURASIP J. Wirel. Commun. Netw. 2019 115 (2019) http://dx.doi.org/10.1186/s13638-019-1459-4.
[6] M. Nyhan, et al., Quantifying population exposure to air pollution using individual mobility patterns inferred from mobile phone data, J Expo. Sci. Environ. Epidemiol. 29 (2) (2019) 238–247.
[7] J. Yu, et al., Global navigation satellite system-based positioning technology for structural health monitoring: A review, Struct. Control Health Monit. 27 (1) (2020) 1–27.
[8] M.S. Grewal, A.P. Andrews, C.G. Bartone, Global Navigation Satellite Systems, Inertial Navigation, and Integration, fourth ed., John Wiley & Sons Inc., NJ, USA, 2020.
[9] L.A. Tawalbeh, A. Basalamah, R. Mehmood, H. Tawalbeh, Greener and smarter phones for future cities: Characterizing the impact of GPS signal strength on power consumption, IEEE Access 4 (2016) 858–868.
[10] K. Chen, et al., Modeling and improving the energy performance of GPS receivers for location services, IEEE Sens. J. (2019) http://dx.doi.org/10.1109/JSEN.2019.2962613.
[11] J.A. del Peral-Rosado, R. Raulefs, J.A. Lopez-Salcedo, G. Seco-Granados, Survey of cellular mobile radio localization methods: From 1G to 5G, IEEE Commun. Surv. Tutor. 20 (2) (2018) 1124–1148, Second Quarter 2018.
[12] D. Nouichi, M. Abdelsalam, Q. Nasir, S. Abbas, IoT devices security using RF fingerprinting, in: Proc. of the Int. Conf. on Advances in Science and Engineering Technology (ASET 2019), Dubai-UAE, 2019, pp. 1–7.
[13] M. Kose, S. Tascioglu, Z. Telatar, RF fingerprinting of IoT Devices based on transient energy spectrum, IEEE Access 7 (2019) 18715–18726.
[14] Q. Wu, et al., Deep learning based RF fingerprinting for device identification and wireless security, Electron. Lett. 54 (24) (2018) 1405–1407.
[15] H. Othman, N. At, C. Topal, Effectiveness of online RF fingerprinting for indoor localization, in: Proc. of the 26th Signal Processing and Communications Apps Conf. Izmir-Turkey, 2018, pp. 1–4.
[16] H. Huang, S. Lin, WiDet: Wi-Fi based device-free passive person detection with deep convolutional neural networks, Comput. Commun. 150 (2020) 357–366.
[17] E. Goldoni, et al., Experimental data set analysis of RSSI-based indoor and outdoor localization in LoRa networks, Internet Technol. Lett. 2 (1) (2019) 1–6.
[18] G. Lui, T. Gallagher, B. Li, A.G. Dempster, C. Rizos, Differences in RSSI readings made by different Wi-Fi chipsets: A limitation of WLAN localization, in: Proc. of the Int. Conf. on Localization and GNSS (ICL-GNSS 2011), Tampere-Finland, 2011, pp. 53–59.
[19] J. Bi, et al., Fast radio map construction by using adaptive path loss model interpolation in large-scale building, Sensors 19 (3) (2019) 1–19.
[20] R. Chen, L. Chen, Indoor positioning with smartphones: The state-of-the-art and the challenges, Acta Geod. Cartogr. Sin. 46 (2017) 1316–1326.
[21] G.B. Tarekegn, H.-P. Lin, A.B. Adege, Y.Y. Munaye, S.-S. Jeng, Applying long short-term memory (LSTM) mechanisms for fingerprinting outdoor positioning in hybrid networks, in: Proc. of the 2019 IEEE 90th Vehicular Tech. Conf. (VTC2019-Fall), Honolulu-USA, 2019, pp. 1–5.
[22] J. Gante, G. Falcão, L. Sousa, Enhancing beamformed fingerprint outdoor positioning with hierarchical convolutional neural networks, In Proc. of the ICASSP 2019-2019 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2019), Brighton-UK, 2019, pp. 1–5.
[23] M. Petric, A. Neskovic, N. Neskovic, et al., Indoor localization using multi-operator public land mobile networks and support vector machine learning algorithms, Wirel. Pers Commun. 104 (2019) 1573–1597, http://dx.doi.org/10.1007/s11277-018-6099-1.
[24] K.S. Anusha, R. Ramanathan, M. Jayakumar, Link distance-support vector regression (LD-SVR) based device free localization in indoor environment, Eng. Sci. Technol. Int. J. (2019) http://dx.doi.org/10.1016/j.jestech.2019.09.004.
[25] P. Dai, Y. Yang, M. Wang, R. Yan, Combination of DNN and improved KNN for indoor location fingerprinting, Wirel. Commun. Mob. Comput. (2019) 4283857, http://dx.doi.org/10.1155/2019/4283857, 9 pages.
[26] Y. Wang et. al, Robust and accurate Wi-Fi fingerprint location recognition method based on deep neural network, Appl. Sci. 10 (1) (2020) 321, http://dx.doi.org/10.3390/app10010321.
[27] D. Zou, W. Meng, S. Chen, D. An, A high robustness positioning algorithm for fingerprint localization system, in: Proc. of the Wireless Communications and Mobile Computing Conference (IWCMC 2016), Cyprus-Paphos, 2016, pp. 730–734.

[28] X. Ge, Z. Qu, Optimization Wi-Fi indoor positioning kNN algorithm location-based fingerprint, in: Proc. of the 7th IEEE Int. Conf. on Software Engineering and Service Science (ICSESS 2016), Beijing-China, 2016, pp. 135–137.
[29] J. Oh, J. Kim, Adaptive K-nearest neighbour algorithm for WiFi fingerprint positioning, ICT Express 4 (2) (2018) 91–94.
[30] R.S. Campos, L. Lovisolo, RF fingerprinting location techniques, in: HandBook of Position Location: Theory, Practice, and Advances, Wiley-IEEE Press, 2019, pp. 497–530.
[31] Q.D. Vo, P. De, A survey of fingerprint-based outdoor localization, IEEE Commun. Surv. Tuts. 18 (1) (2016) 491–506.
[32] Y.S. Abu-Mostafa, M. Magdon-Ismail, H.-T. Lin, Learning from Data, AMI Book New York, NY, USA, 2012.
[33] D.W. Aha, D. Kibler, M.K. Albert, Instance-based learning algorithms, Mach. Learn. 6 (1) (1991) 37–66.
[34] T. Mitchell, Machine Learning, McGraw-Hill, NY, USA, 1997.
[35] E.S. Gardner, Exponential smoothing: The state of the art, J. Forecast. 4 (1) (1985) 1–28.
[36] Q. Wang, et al., IWKNN: An effective bluetooth positioning method based on isomap and WKNN, in: Mobile Information Systems, 2016, 8765874, 1–11.
[37] X. Guo, L. Li, N. Ansari, B. Liao, Accurate WiFi localization by fusing a group of fingerprints via a global fusion profile, IEEE Trans. Veh. Technol. 67 (2018) 7314–7325.
[38] J.B. Andersen, S. Yoshida, T.S. Rappaport, Propagation measurements and models for wireless communications channels, IEEE Commun. Mag. 33 (1) (1995) 42–49.
[39] A.M. Hossain, Y. Jin, W.-S. Soh, H.N. Van, SSD: A robust RF location fingerprint addressing mobile devices heterogeneity, IEEE Trans. Mob. Comput. 12 (1) (2013) 65–77.
[40] R.S. Campos, L. Lovisolo, A fast database correlation algorithm for localization of wireless network mobile nodes using coverage prediction and round trip delay, in: Proc. of the IEEE 69th Vehicular Technology Conf. (VTC Spring 2009), Barcelona - Spain, 2009, pp. 1–5.
[41] F. Pedregosa, et al., Scikit-learn: Machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[42] H. Suzuki, A statistical model for urban radio propagation, IEEE Trans. Commun. 25 (7) (1977) 673–680.
[43] Second Report and Order on Wireless E911 Location Accuracy Requirements, Federal Communications Commission, Washington-US, 2010, Rep. 10-176.
[44] K. Sultan, H. Ali, Z. Zhang, Big data perspective and challenges in next generation networks, Future Internet 10 (56) (2018) 1–20.
[45] Q. Yao, et al., Taking the human out of learning applications: A survey on automated machine learning, 2019, Available at https://arxiv.org/pdf/1810.13306.pdf.
[46] J. Verzani, Using R for Introductory Statistics, Chapman and Hall/CRC, NY, USA, 2018.
[47] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B Stat. Methodol. 57 (1) (1995) 289–300.