

Received 26 February 2024, accepted 28 March 2024, date of publication 4 April 2024, date of current version 16 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3384625

RESEARCH ARTICLE

Evaluation and Comparison of 5G, WiFi, and Fusion With Incomplete Maps for Indoor Localization

CARLOS SIMÓN ÁLVAREZ-MERINO¹, EMIL J. KHATIB¹, (Member, IEEE),
HAO QIANG LUO-CHEN¹, ANTONIO TARRÍAS MUÑOZ¹,
AND RAQUEL BARCO MORENO¹

Telecommunication Research Institute (TELMA), ETSI de Telecomunicación, Universidad de Málaga, 29010 Málaga, Spain

Corresponding author: Emil J. Khatib (emil@uma.es)

This work was supported in part by the Framework of the Maori Project funded by the European Union (EU)-NextGenerationEU under Grant TSI-063000-2021-53, and in part by the University of Malaga for Open Access Charge: Universidad de Málaga/ Consorcio de Bibliotecas Universitarias de Andalucía (CBUA).

ABSTRACT Precise positioning will play a key role in future 5G/6G services. The upcoming location-based services drive the necessity of high-precision positioning to indoors. In fingerprinting, which is the most commonly used location algorithm indoors, comprehensive radio maps are essential for a precise localization service and highly influence on the result of the final position of the user. A Machine Learning (ML) algorithm that supports missing reference points information when maps are incomplete that are used during the training phase may improve the robustness and reliability of the localization service. In this work, we compare the performance of the classical fingerprinting technique and different Decision Tree Regressor (DTR)-based algorithms that are Decision Tree Adaboost (DTA), Linear Tree Adaboost (LTA) and Random Forest (RF). The experiments were carried out with real 5G and WiFi data in an indoor scenario to test the performance of the techniques. Additionally, we demonstrate the benefits of fusion of technologies when positioning with radio maps. Finally, an evaluation of the robustness from the different methods was carried out when missing information in radio maps during the training phase.

INDEX TERMS Fingerprinting, adaboost, random forest, 5G, WiFi, fusion, indoor localization.

I. INTRODUCTION

As an increasing number of customer services rely on location to satisfy the needs of both users and network operators, Localization-as-a-Service (LaaS) is becoming increasingly vital for 5G and 6G networks [1]. LaaS is critical in enabling new location-based services such as autonomous robots and vehicles [2], smart education [3] or e-Health [4]. The 3GPP has set a target of achieving high localization accuracy for 5G networks, aiming for submeter accuracy in certain cases such as autonomous driving, where location accuracy below 10 cm is envisioned [5], and an

The associate editor coordinating the review of this manuscript and approving it for publication was Yeon-Ho Chung¹.

accuracy of below 3 meters in most cases (both indoors and outdoors) [6]. Combining context-aware data from the Internet of Things (IoT) collected through WiFi networks with 5G information can enhance the accuracy, reliability, and scalability of localization services [7], [8].

Accurate location estimation has become increasingly important in recent years, and the use of Global Navigation Satellite Systems (GNSS) is a common approach for achieving high accuracy in outdoor environments. However, issues like signal blocking, attenuation, and multipath effects make GNSS ineffective indoors, where many applications are being developed. To address this, supplementary technologies like 5G, WiFi, or Ultra Wide Band (UWB) are often used to determine location [9], [10]. In situations where energy

constraints on user devices require network-based location to conserve battery and optimize computational efficiency, the network estimates the User Equipment (UE) location based on data collected in the network infrastructure in a non-cooperative manner [11]. Some applications, such as beam management or automatic configuration of network parameters, may also require terminals to transmit their location using specific protocols [12], which can be complex and energy-intensive. An alternative solution to determining location is through network-based location [13]. This method involves the network utilizing data collected from the network infrastructure in a non-cooperative manner to estimate the location of a terminal.

Cellular networks like Long Term Evolution (LTE) are commonly used to locate users when GNSS is unavailable [14]. The most common approaches are location by proximity, ranging-based methods, Angle of Arrival (AoA) and fingerprinting. Location by proximity is the easiest method to determine the location of the UE because it assumes the location of the gNodeB (gNB) is the location of the UE and is used when high accuracy is not required [15]. Ranging-based methods, such as trilateration, involves using ranges obtained through methods such as Received Signal Strength Indicator (RSSI) or Time of Flight (ToF) [9] and can be very accurate if ranges are precise. The determination of the location involves estimating the interception of 4 spheres (or 3 in 2D location). Nevertheless, range estimations are not normally accurate, occasionally resulting in the non-convergence of circles or hyperbolas utilized in the trilateration process. To solve the uncertainty, techniques such as Least Squares (LS) or Weighted Least Squares (WLS) are used [16]. AoA measures the angle at which the signal reaches the UE from the gNB. Multiple Input Multiple Output (MIMO) systems are capable of transmitting with beamforming that can be used to implement the AoA approach [17]. Indoor environments pose reliability challenges for both range-based models and AoA due to the susceptibility of the models to signal blocking and reflections. While the received power might not follow a predetermined propagation model, in cases where the environmental conditions remain relatively stable, it is observed to remain constant over time. For instance, if we consider a location in close proximity to a WiFi AP and the measured power is unusually diminished due to an obstacle such as a wall, this power level will remain unaltered over time as long as the obstruction remains stationary. As a result, each point in space is associated with a set of paired values comprising reference point (RP) identifiers and unvarying received power levels. This principle underlies the concept of fingerprinting, these paired values conform a distinctive signature, commonly referred to as a *fingerprint*, which serves to uniquely identify each point in space [18].

Fingerprinting exhibits several primary drawbacks. It notably demonstrates high sensitivity to disparities between training and testing conditions arising from dynamic

propagation attributes such as temperature, humidity, and obstacles [19], [20], [21]. Additionally, it mandates an extensive preliminary map construction phase, which necessitates thoroughness [22]. This is imperative because unrecorded data points remain unusable for positioning during the operational phase. Lastly, the integrity of the radio map is compromised due to device heterogeneity stemming from variations in orientation and chip sensitivity [23].

Fingerprinting has some disadvantages, which include the requirement for a long map-building phase in advance. This process must also be comprehensive because unmeasured points cannot be used for location in the exploitation phase. An alternative method for constructing maps is through crowdsourcing data from various UEs or sensors. This method may sacrifice precision but it is cost-effective. The system can use the measurements provided by these sources to create or update the radio map or models for localization systems [24], [25]. Other studies have explored the reconstruction of maps when maps are incomplete. In [26], they addressed this issue by leveraging the linear nature of signal propagation. Their objective was to create new data by considering the context of the existing map, allowing for the application of techniques like fingerprinting. It is important to note that this approach is constrained by the granularity of the radio map division, which directly impacts the final accuracy of the system. When the division is finer, precision increases, but it also necessitates a larger number of minimum required data points. Another avenue explored is the utilization of Deep Learning techniques with incomplete maps for the recovery of missing data points. However, it is observed that this approach has limitations and can only recover up to 50% of missing data in incomplete maps [27].

Supplementary techniques such as combining ranging with AoA can enhance the final location estimation of a UE, resulting in a higher degree of accuracy [9], [28]. The fusion of multiple technologies helps to increase the density of RPs in the scenario, providing more information for the final estimation stage. This reduces the cost of infrastructure or expands the coverage area [29].

The contributions of this paper are listed as follows:

- Implementation of fingerprinting and model-based algorithms utilizing real 5G and WiFi data.
- Evaluation of the performance of positioning systems when fusing different technologies employing map- and model-based methodologies.
- Examination of the the behavior of the algorithms when varying different percentages of missing reference points during the training phase.
- Proposition of model-based techniques when maps are incomplete with a percentage over 50% of missing reference points while minimizing the degradation on the localization performance.

The rest of the paper is organized as follows. Section II explains both the fingerprinting algorithm and various DTR-based techniques. Section III provides an overview of benefits of fusion of technologies. In Section IV, the

TABLE 1. Overview of acronyms.

Acronym	Definition
Adaboost	Adapting Boosting
AoA	Angle of Arrival
AP	Access Point
CDF	Cumulative Distribution Function
DTA	Decision Tree Adaboost
DTR	Decision Tree Regressor
IoT	Internet of Things
gNB	gNodeB
GNSS	Global Navigation Satellite Systems
LaaS	Localization-as-a-Service
LTA	Linear Tree Adaboost
LTE	Long Term Evolution
MIMO	Multiple Input Multiple Output
ML	Machine Learning
RF	Random Forest
RMSE	Root Mean Square Error
RP	Reference Point
RSSI	Received Signal Strength Indicator
ToF	Time of Flight
UE	User Equipment
UWB	Ultra Wide Band
WL	Weak Learner

experimental setup and the scenario are described. Section V analyzes the results of the outcomes that were obtained from the data collection campaign and the implementation of various location methods with different experiments. Finally, Section VI presents the conclusions of this work.

The acronyms in this paper are listed in the Table 1 as follows:

II. LOCATION TECHNIQUES

In indoor environments, techniques like trilateration can be challenging due to the possibility of signal blocking and reflections, which can result in significant errors. ToF based ranging estimation can reduce these errors, but it can be expensive due to the hardware requirements [9]. In indoor environments, however, there are typically multiple radio signals that can be measured and reported without hardware modifications. In cellular networks, UEs are required to measure all visible base stations and report this information to the serving base station to determine the best cell [30]. In stable indoor environments, the received power tends to remain constant, making radio map techniques particularly useful. This section will provide an overview of various techniques suitable for these types of scenarios.

A. FINGERPRINTING

Classical fingerprinting is a localization technique that involves generating a unique *fingerprint* of wireless signal strength and other characteristics of a particular location. This *fingerprint* can be later used to identify the location of a device. The process of creating a *fingerprint* involves measuring wireless signal characteristics at various points within an area, such as a building or campus. In static

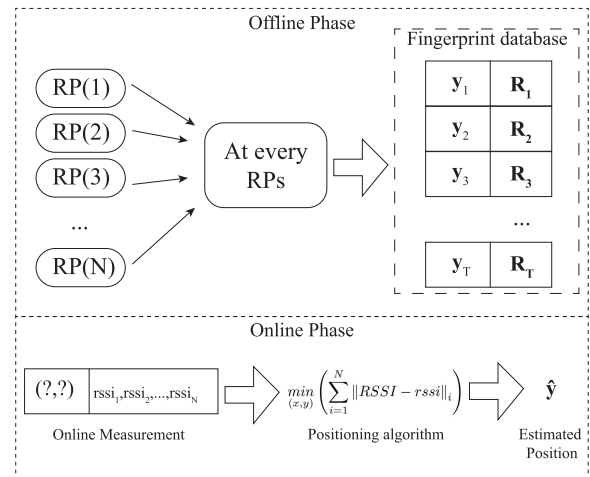


FIGURE 1. Fingerprinting method.

environments where changes are minimal, the received power at a specific point in space remains relatively constant over time. As a result, each point \mathbf{y}_i can be represented in a database of T entries using a vector that contains the RSSI measurements of the N APs received by UE with known positions in WiFi and gNBs in 5G networks, $\mathbf{R} = (RSSI_1, RSSI_2, \dots, RSSI_N)$, as shown in Figure 1. Consequently, each point has a distinctive RSSI vector that constitutes its *fingerprint* allowing for accurate location tracking.

Fingerprinting involves two distinct phases, as illustrated in Figure 1. The first phase, known as the offline or training phase, involves creating a radio map by assigning a unique *fingerprint* to each point on a regular grid. In the second phase, called the online or exploitation phase, the terminal measures the surrounding gNBs and generates a new vector $(rssi_1, rssi_2, \dots, rssi_N)$. This vector is then compared to the different points on the map to find the most similar *fingerprint*. To determine the most probable position \hat{y} on the grid, the algorithm seeks the point that minimizes the Euclidean distance between the new vector of RSSI and the *fingerprint* database, as described in the positioning algorithm equation shown in Figure 1.

The level of accuracy in the fingerprinting method depends on several factors such as the size of the grid used during the training phase, the size of the input vector, the variance of the measured power for each component, and the accuracy of the UE's measurements. If the input vector has fewer than N entries, the final estimation will be degraded. In the case of multiple final estimations, the system may randomly provide one final location. In addition, there is a tradeoff between complexity and accuracy; this is crucial because using a smaller grid results in a shorter training phase but with lower accuracy, while a finer grid requires a longer training phase but with higher accuracy.

WiFi and cellular networks are commonly associated with fingerprinting due to the high density of stations in office

and residential areas [18], [31]. Fingerprinting can provide high accuracy with a reduced infrastructure investment, but it requires the creation of a radio map with a complex training phase. To maintain location precision, the maps need to be updated when there are changes in the environment. Furthermore, the maps must be comprehensive, meaning that all points on the grid must be systematically measured in order to properly locate users at any point.

B. DTR-BASED LOCATION

Fingerprinting has a major drawback in that it requires complete information maps about the environment. To address this issue, a commonly used approach is to employ ML algorithms to create an environment model [32], [33], [34], which can then be utilized for estimating the position during the exploitation phase. ML algorithms generate a comprehensive model of the scenario through the information provided in the training phase with certain RPs. Consequently, even without conducting measurements across the entirety of the scenario during the training phase, the ML model enables a localization service encompassing the entire designated area [35], [36]. In the context of a grid scenario when maps are incomplete, this study utilized DTRs, which are recognized for their simplicity and computational efficiency [37], to estimate the position. In addition, DTR-based methods were chosen over other ML algorithms because they offer the best precision for indoor localization [38].

By creating a set of hierarchical comparison rules that are applied sequentially, DTRs model the behavior of the localization system. The resulting path over a tree is determined by the outcome of each rule (branch), leading to a final node (leaf) that decides the output of the regressor as illustrated in Figure 2.

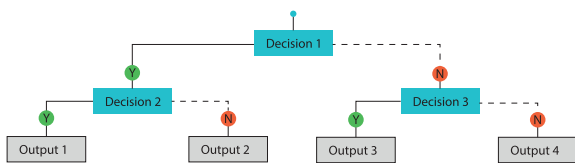


FIGURE 2. DTR functionality.

The DTR learning process comprises two phases: training and testing. In the training phase, the 80% of the available samples are randomly selected to form the training dataset D_{train} , while the remaining 20% of the dataset D_{test} is used for testing. The objective of the training phase is to create a tree that minimizes the regression error on the training set.

In this work, we study different DTR-based algorithms that were chosen due to its high accuracy and low complexity: Random Forest (RF) and two Adaboost-based training algorithms that are Decision Tree Adaboost (DTA) and Linear Tree Adaboost (LTA). DTA, in its final prediction of positions, combines outputs from different WLS by using a decision rule and taking their average [39]. On the other hand,

LTA creates an interpolation function by taking into account the different outputs of a set of decision rules [40], [41].

1) RANDOM FORESTS

RFs are a ML technique that employs a collection of individual models (known as *base models*) to generate a final prediction. This ensemble method is versatile and can be applied to various ML tasks such as classification, regression, or localization. RFs are especially useful for localization tasks because they can effectively aggregate the predictions of multiple decision trees to determine the location of a device [42], [43].

RFs employ the bootstrapping method to generate decision trees, which involves a random subset of the training data that is selected to create a single decision tree. This process is repeated several times, leading to a vast number of decision trees trained on various subsets of the data. To generate the final prediction of the localization process, the predictions of all the decision trees in the forest are averaged as depicted in Figure 3.

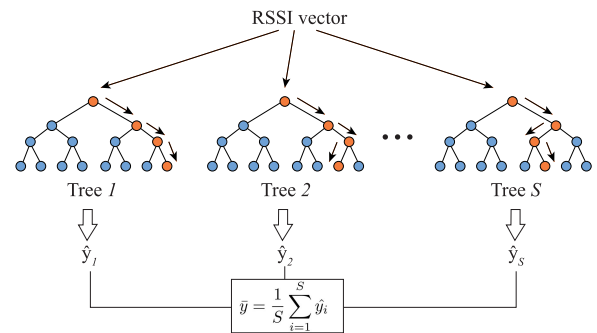


FIGURE 3. RF schema.

The implementation of RFs is relatively straightforward, as they utilize decision trees, which makes them computationally efficient. Additionally, RFs are resistant to data noise since average of all the location outputs mitigates the impact of any individual decision tree that might produce an inaccurate estimation.

Algorithm 1 explains with pseudocode the structure and formulation of the RF algorithm [44]. Given T samples and N features (in this case, APs) labeled $([R(1), \mathbf{y}_1], \dots, [R(T), \mathbf{y}_T])$ where the input vector is a tuple compound by a RSSI vector formed by $RSSI_n(t)$ from the n th AP and t th sample and the localization output \mathbf{y}_t that is the final position of the user in a 2D plane. For training the RF, we generate S trees as an arbitrary number of the size of the forest. On each iteration $i = 1, \dots, S$, we select a random subset of features ($S_{features}$) and samples ($S_{samples}$). Then, a tree is generated using $S_{samples}$ with the chosen features $S_{features}$.

During the testing phase of the RF algorithm, the input vector is $[rssi_1, rssi_2, \dots, rssi_N]$ which contains the measurements of different APs. Each tree produces an estimated location $\hat{\mathbf{y}}_i$ and the output of the RF algorithm is

Algorithm 1 RF Algorithm

- To model the forest (training):
 - **Input:** $([\mathbf{R}(1), \mathbf{y}_1], \dots, [\mathbf{R}(T), \mathbf{y}_T])$
 - foreach** $i = 1, \dots, S$ **do**
 - Select a random subset of features ($S_{features}$) from the input data
 - Extract a random subset of the samples ($S_{samples}$)
 - Build a tree using data of $S_{samples}$ with the selected features $S_{features}$
 - end foreach**
- To predict the final location:
 - Input the new vector of measurement $[rss_{i1}, \dots, rss_{iN}]$
 - foreach** $i = 1, \dots, S$ **do**
 - Estimates the location $\hat{\mathbf{y}}_i$ with the new input vector on the i -th tree
 - end foreach**
 - Output the final position $\bar{\mathbf{y}} = \frac{1}{S} \sum_{i=1}^S \hat{\mathbf{y}}_i$

$\bar{\mathbf{y}}$ as the average of all estimated locations. In the testing phase, we know the real location where the measurements are collected. Thus, the quality of the estimation by the RF can be analysed. The proposed approach has a significant advantage in that it is computationally efficient that can be used for real-time applications.

2) ADAPTING BOOSTING (ADABOOST)

Adaboost leverages the predictions of multiple individual models, known as *Weak Learners* (WLs), to arrive at a final prediction [45]. The WLs are generated through a process called boosting, which involves iteratively training the model on new subsets of the data, with each round emphasizing the data points that were incorrectly classified in the previous iteration. Figure 4 illustrates the method of combining the predictions of all the WLs in the ensemble to make the final prediction. Two Adaboost-based training algorithms are studied in this work: DTA and LTA. In the DTA method, the positions from various WLs associated with a decision rule are averaged in the final prediction [46], while in the LTA method, an interpolation function is developed between the different outputs within a set of decision rules [40], [41].

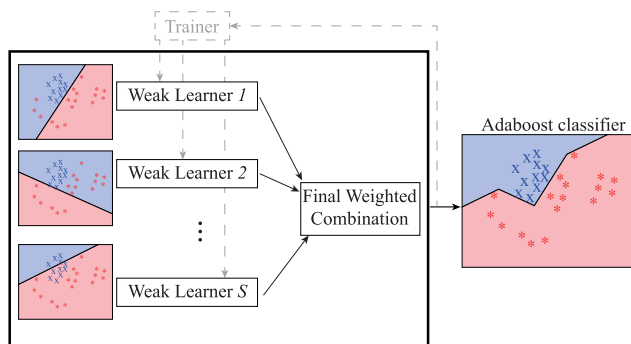


FIGURE 4. Adaboost process.

Adaboost is capable of adapting and learning from changes in data over time, making it crucial in dynamic environments where wireless characteristics are prone to variation. Although it achieves high accuracy, especially in LTA, a significant disadvantage of Adaboost is its reliance on extensive computational processing for the final estimation.

The structure and formulation of the Adaboost regressor training, as described in [45], is explained through the pseudocode in Algorithm 2. As in RF algorithm, given T samples and N APs labeled $([\mathbf{R}(1), \mathbf{y}_1], \dots, [\mathbf{R}(T), \mathbf{y}_T])$ where the input vector is a tuple compound by a RSSI vector formed by $RSSI_n(t)$ from the n th AP and t th sample and the output \mathbf{y}_t is the final position of the user in a 2D plane. For the Adaboost, there is a number S that determines the number of WLs (in RF determines the number of trees). On each iteration $i = 1, \dots, S$, the regression generates a WL function g_i with an associated weight w_i that adjust the regressor to minimise error. Initially, w_i is set equal to $1/S$. The weighted error e_i defines the error of the regression at the i iteration and sets λ_i . The parameter λ_i defines two characteristics of the regressor. First, λ_i defines the step size of the adapting boosting. Second, λ_i sets the new weight w_{i+1} in the WL. Then, the set of weights w_i are normalized for the next iteration. The final regression model $F(x)$ computes the combination of the weighted WLs that is the weighted average in DTA and the weighted linear regression in LTA.

The estimator uses $F(x)$ for the testing or exploitation phase. In these phases, the input of the regressor is a vector $[rss_{i1}, rss_{i2}, \dots, rss_{iN}]$ with the measurements of the APs extracted from the UE. The output of the algorithm is the estimated location $\bar{\mathbf{y}}$ of the UE.

Algorithm 2 Adaboost Algorithm

- **Input:** $([\mathbf{R}(1), \mathbf{y}_1], \dots, [\mathbf{R}(T), \mathbf{y}_T])$
- Initialize weights $w_i = \frac{1}{S}$ for every i
- Start with the null classifier $f_0(\vec{x}) = g_0(\vec{x}) = 0$
- $[\mathbf{R}(1), \mathbf{y}_1], \dots, [RSSI_N(T), \mathbf{y}_T]$
- foreach** $i = 1, \dots, S$ **do**
 - Fit some weak learner g_i
 - Calculate $e_i = \frac{\sum_{j=1}^S (e_j * w_j)}{\sum_{j=1}^S w_j}$
 - Set $\lambda_i = \frac{1}{2} \ln \left(\frac{1 - e_i}{e_i} \right)$
 - Update weights: $w_{i+1} = w_i e^{-\lambda_i y_i g_i}$
 - Normalize w_i to sum to one
 - The new model is $f_i = f_{i-1} + \lambda_i g_i$
- end foreach**
- Output the final model $\bar{\mathbf{y}} = F(x) = \sum_{i=1}^S \lambda_i g_i$

III. FUSION OF Wi-Fi/5G TECHNOLOGIES

As the demand for connectivity increases, the number of radio technologies available at a specific point in space has increased over time. This is especially true for indoors

environments, where the demand for broadband is higher. Thus, it is common that technologies such as WiFi and cellular networks are present in most indoor scenarios.

The fusion of 5G and WiFi can also enhance the user experience by providing seamless connectivity [47]. This is particularly important in indoor environments where users frequently move between different rooms and areas, each with varying signal strengths and qualities. With the integration of 5G and WiFi, the system can dynamically switch between the two technologies depending on the location and signal strength, ensuring a consistent and reliable connection.

Moreover, the fusion of these two technologies can also improve network efficiency and reduce costs [48]. With the increasing demand for high-speed connectivity, network operators are under pressure to provide faster and more reliable services. By utilizing both 5G and WiFi technologies, operators can optimize the use of available resources, thereby reducing network congestion and improving overall network performance. This can result in lower costs for both the network operator and the end-user [49].

In terms of localization, 5G and WiFi are two technologies that can be utilized to increase the coverage area, enhance the accuracy of the final location estimate through fusion in trilateration [29], or create denser areas for radio map creation. Furthermore, since both services are managed independently, they can act as backup options for each other in case one fails. Additionally, both technologies can offer unique services, such as wide spectrum service in case of 5G [50] or precise timestamp in trilateration for WiFi [51].

In this work, the fusion of 5G and WiFi, for the different localization algorithms, is direct as the system integrates the data of both technologies as input data. Fusion enables the system to expand the number of APs available for the radio map creation. Having a higher number of APs in the radio map allows the method to compare the context of the UE more thoroughly for the final location estimation. Moreover, a denser radio map reduce the impact of losing a single gNB or APs.

IV. EXPERIMENTAL SETUP

This section presents the configuration for obtaining real 5G data and WiFi from the University of Malaga. The 5G network belongs to the University of Malaga, and contains three indoors base stations which have been configured to reduce the interferences with commercial networks. The base stations are located at two different heights (2.5m and 3.5m) and a map of the scenario is shown in Figure 5. The three WiFi APs are Google WiFi mesh routers placed on shelves at a height of 2 meters in order to ensure the visibility of all APs in the entire map. Measurements were taken at *ground truth* points represented by orange dots and green dots. The scenario includes three laboratories and one hall with metallic elements that can cause signal blocking, attenuation, and multipath effects. The 5G gNBs are placed in the ceiling to provide good visibility and transmit at a power of 20 dBm

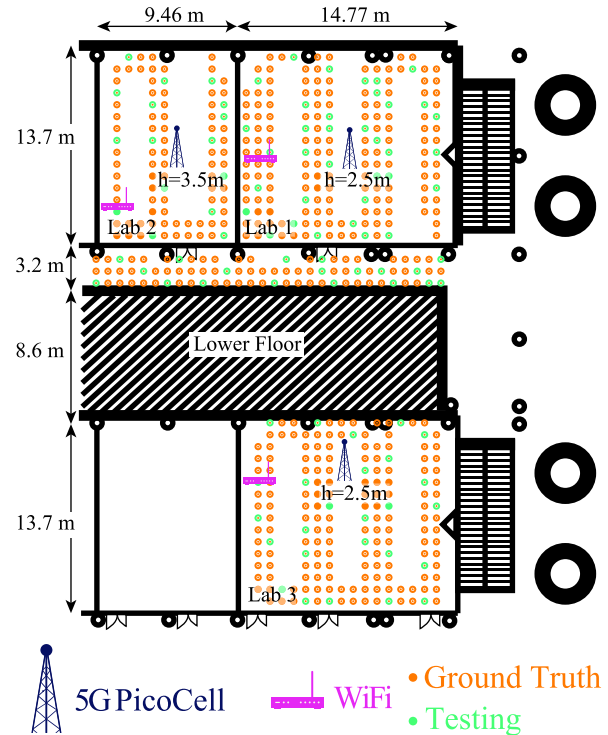


FIGURE 5. Map of the scenario.

at a frequency of 3774.990 MHz. Measurements were taken systematically over a grid of points marked on the floor as illustrated in Figure 5. Samples were taken 0.8 meters apart to cover the entire accessible area of the scenario.

The location target UE is a Motorola Edge 20 which runs Android 11. An application has been programmed to capture the RSSI of the serving and neighbor cells and WiFi APs information. The captured data is sent to a server over 5G, where the measurement samples are saved in a MySQL database to be further processed. The programmed application also allows to indicate the ground truth and send it along the taken measurements.

V. RESULTS

In this section, we present the localization results obtained from three different experiments. All experiments used a dataset of over 500 samples. The data was randomly split into a training set and a testing set (represented in Figure 5 as orange and green dots, respectively) with 20% of the measuring points allocated for testing. This process was repeated a thousand times, on each iteration the training and testing points are randomly chosen, using the Monte Carlo method, in which Figure 5 represents one example of this process, to produce accurate statistical results.

A. EVALUATION OF DIFFERENT METHODS

This experiment evaluated the performance of four localization techniques - fingerprinting, DTA, LTA, and RF. The DTA and LTA methods were trained with 50 WLs

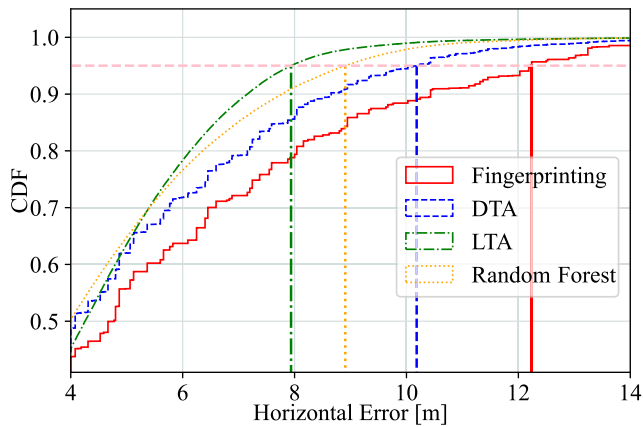


FIGURE 6. Cumulative Distribution Function of the error of different methods with 5G data.

as suggested in [52], and the number of trees in RF was set at 50 for fair comparison with Adaboost. In this experiment, the performance of the different methods is being evaluated solely using 5G technology. For fingerprinting, the training data was used to construct a radio map and for the rest of the methods, the training data was utilized to construct the trees or the WLS. The testing data was used to measure the precision of the different localization methods, with the results represented by the Cumulative Distribution Function (CDF) of the horizontal error in Figure 6. The 95th percentile (horizontal pink line) has been selected as the basis for location accuracy standard [53].

Fingerprinting (red) estimates the position of the UE on the radio map by identifying the closest point. The radio map is divided into a lattice, and fingerprinting determines the location of the UE within this lattice. DTA (blue) calculates the average the output of the different WLS. In case of averaging a regular radio map, the final result is always a lattice of the radio map. Notably, all measurements are acquired at the center of these lattices. Thus, both fingerprinting and DTA provide a lattice-based location that is translated into discrete error and a staggered step of the CDF. RF (yellow) performs better than DTA as it provides good performance on large and complex datasets and averages the linear estimation of the different trees. In contrast, DTA averages the final estimation based on WLS. LTA (green) is the top performer because it generates an interpolation function of the different WL outputs, significantly improving final localisation accuracy, even though, it reduces the computational efficiency.

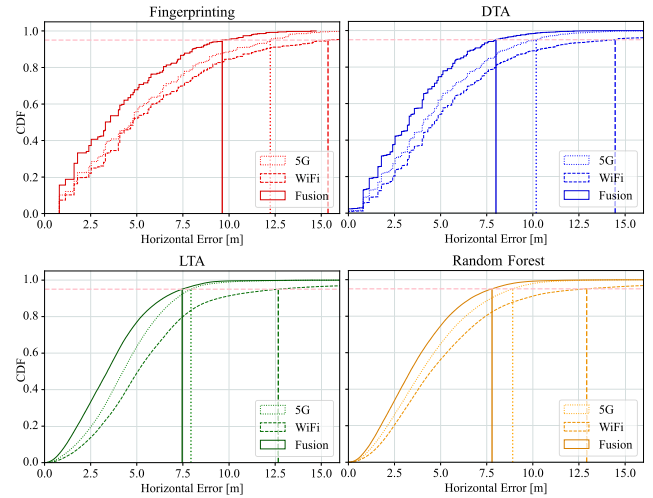


FIGURE 7. Cumulative Distribution Function of the error of 5G, WiFi and fusion for different methods.

Figure 6 clearly shows that ML methods significantly reduce errors compared to the fingerprinting technique. The accuracy of fingerprinting heavily relies on the radio map as it compares directly with the received signals from the UE. RF and LTA enhances the final location estimation compared to DTA because the final position is derived from a linear function. While LTA provides higher precision in positioning, it is not feasible for real-time applications due to its highly time-consuming nature. RF offers a trade-off between accuracy and computational efficiency, making it suitable for real-time applications.

B. EVALUATION OF FUSING TECHNOLOGIES

In this experiment, the behavior of the different algorithms are evaluated with different cases: with 5G and WiFi in isolation and the fusion of both. The goal was to determine if fusion enhances the precision of the localization system when different technologies appear in fingerprinting and DTR-based methods.

Although 5G promises high-precision positioning through the multi-RTT protocol, to the best of the authors’ knowledge, this protocol has not been implemented in any commercial device yet. On the other hand, WiFi has already created the 802.11mc protocol, which offers accurate ranging estimation through the RTT protocol and can achieve meter-level accuracy [9], [51]. This protocol is widely implemented in plenty of smartphones, but only a limited number of APs

TABLE 2. Performance comparison of different methods with 5G, WiFi and fusion.

	5G					WiFi					Fusion				
	μ	Mdn	RMSE	σ	95%	μ	Mdn	RMSE	σ	95%	μ	Mdn	RMSE	σ	95%
Fingerprinting	5.21	4.77	4.45	3.52	12.23	6.15	4.8	5.56	4.90	15.36	4.06	3.29	3.52	2.88	9.63
DTA	4.6	4.07	7.17	2.97	10.18	5.71	4.66	7.38	4.47	14.46	3.61	3.29	6.90	2.40	8.0
LTA	4.39	4.25	3.45	2.12	7.93	5.77	5.06	4.92	3.87	12.67	3.63	3.40	2.96	2.06	7.46
RF	4.33	3.98	3.69	2.46	8.9	5.42	4.49	4.85	4.95	12.91	3.70	3.36	3.24	2.19	7.78

TABLE 3. Comparison of different methods.

Method	Advantages	Disadvantages
Fingerprinting	Efficient and simple Suitable for real-time applications	It requires frequent updates on the map Not precise
Random Forest	It can handle large amount of data Robust to overfitting and can handle incomplete maps Relatively simple	Does not adapt to changes Significant computational resources for training
DTA	Adapts to changes in the environment Robust to overfitting and can handle incomplete maps	Not as effective as other methods Computationally intensive
LTA	Adapts to changes in the environment High accuracy Reduced frequency of retrainings Allows for larger radio maps	Computationally intensive Not suitable for real-time applications

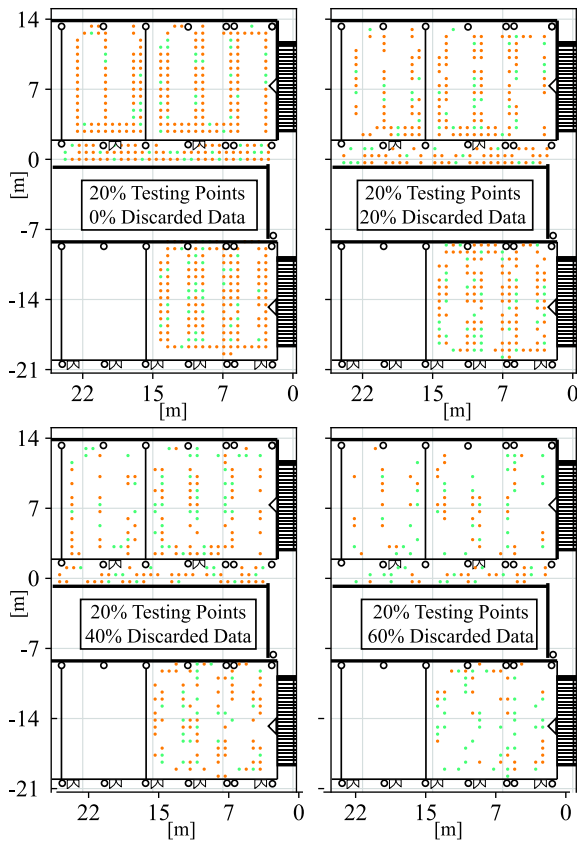


FIGURE 8. Experiments with different percentage of discarded data.

have adopted it [54]. Until RTT protocol gets integrated into 5G, using ML techniques and fingerprinting along with RSSI measurements can be highly beneficial. Additionally, capturing RSSI measurements does not increment the energy consumption of the terminal or demand special hardware [9].

Figure 7 represent the different cases of fingerprinting, DTA, LTA and RF with only 5G NR (solid line), only WiFi (dashed line) and fusion of 5G and WiFi (dotted line). As it can be observed, fusion improves the performance of the system in all cases. Combining different technologies increases the number of APs in the scenario, which improves the final estimation due to the availability of more information

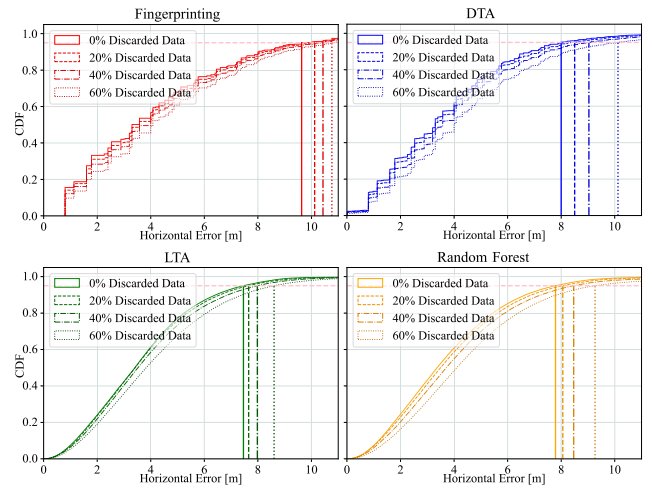


FIGURE 9. Cumulative density distribution of the error of different techniques with different percentage of discarded data.

of the environment. So, the more complete radio map, the better localization resolution will be. Among the techniques, fingerprinting yields the greatest improvement as it is the most radio map dependent. Despite of LTA having a slightly better overall performance, RF still provides a high level of accuracy that is comparable to LTA and it allows real-time location-based services.

Table 2 presents the performance of the different algorithms for 5G, WiFi and fusion data, characterized by metrics including the mean (μ), median (Mdn), Root Mean Square Error (RMSE), standard deviation (σ) and the 95% percentile of cumulative density error. All measurements are presented in meters for reference.

C. ROBUSTNESS OF THE METHODS WITH DIFFERENT LEVELS OF INCOMPLETE MAPS

In this experiment, the robustness of the different methods are evaluated with varying degrees of missing RPs in incomplete maps. In this case, the fusion of 5G and WiFi is used as the input data because it has demonstrated to always improve the performance of the localization results. The goal was to determine the robustness of the techniques by examining how well it performed as the percentage of missing data

in the radio map increased. This experiment consists on reducing the number of training points. To do this, the percentage of testing points was kept constant at 20% while the percentage of discarded data varied from 0% to 60% as shown in Figure 8.

The results of the experiment were represented by the CDFs of the horizontal error when different percentages of discarded data (0%, 20%, 40% and 60%) are used to evaluate the performance of the system. As it can be observed in Figure 9, except of the case of DTA with 60% of discarded data, the rest of the cases of DTR-based algorithms outperform fingerprinting without discarded data in terms of user localization accuracy. Therefore, using DTR-based algorithms, allows to achieve higher levels of accuracy even with maps that contain fewer data points.

Table 3 provides a brief summary of the different methods.

VI. CONCLUSION

Indoor positioning has become an increasingly important technology in recent years as it enables a wide range of applications, such as indoor navigation, asset tracking, and location-based services. However, the traditional method of using radio maps for indoor positioning has several significant drawbacks. One of the most significant issues with radio map techniques is the complex training process required. The process of creating a radio map involves collecting and analyzing a large amount of data from a given indoor environment. This data is used to build a map of the radio signal strength for each location in the area. However, this process can be time-consuming and expensive, which limits its applicability in scenarios where a large area must be covered. Moreover, another major issue with radio map techniques is that the fingerprints of the indoor environment can change over time due to changes in the scenario. These changes can affect the accuracy of the radio map, which requires frequent updates to maintain the effectiveness of the technique. This retraining can be very costly, both in terms of time and resources.

In this work, we have performed and compared fingerprinting, DTA, LTA and RF techniques with real 5G and WiFi data. First, DTR-based methods noticeably improves the localization performance compared with the regular fingerprinting. It is remarkable that LTA and RF have performed better than DTA and fingerprinting because the final location based on the interpolation between points.

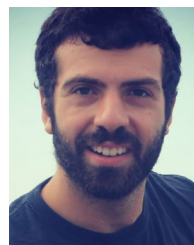
On the other hand, fusion of technologies have proven to provide better performance of the system. By combining 5G and WiFi, the number of APs in the scenario increases. This implies that during both the training phase and the operational phase, the different localization algorithms are provided with more information about the environment. This, in turn, results in improved final estimation by providing more environmental information. Furthermore, fusion has the potential to enhance connectivity, extend coverage, optimize resources for location-based services and minimize the expenses associated with deployment and infrastructure.

Related to the robustness of the different methods, LTA and RF maintains the error stable even when the percentage of missing RPs in incomplete maps becomes significant, up to 40% of missing data. Its robustness allows to cover larger areas, minimize the need for frequent retraining, or decrease the number of data points required on each map. Depending on the service being offered, DTR-based models, specifically LTA and RF, can be highly valuable tools for indoor positioning. In these experiment, LTA yields better results in both experiments than RF but it is not suitable for real-time applications. Nonetheless, RF provides a balance between accuracy and computational efficiency, making it ideal for real-time services. Although RF cannot adapt to environmental changes, DTA readjusts to them but with decreased localization accuracy. As a result, DTR-based models have imposed its applicability over fingerprinting. However, there is no single method that is universally best for location-based services, as it varies depending on the specific application and scenario.

REFERENCES

- [1] W. Jiang and H. D. Schotten, "The KICK-OFF of 6G research worldwide: An overview," in *Proc. 7th Int. Conf. Comput. Commun. (ICCC)*, Dec. 2021, pp. 2274–2279.
- [2] Q. Li, J. P. Queralta, T. N. Gia, Z. Zou, and T. Westerlund, "Multi-sensor fusion for navigation and mapping in autonomous vehicles: Accurate localization in urban environments," *Unmanned Syst.*, vol. 8, no. 3, pp. 229–237, Jul. 2020.
- [3] S. Fortes, J. Santoyo-Ramón, D. Palacios, E. Baena, R. Mora-García, M. Medina, P. Mora, and R. Barco, "The campus as a smart city: University of Málaga environmental, learning, and research approaches," *Sensors*, vol. 19, no. 6, p. 1349, Mar. 2019.
- [4] G. Paoletti, D. Masotti, F. Antoniazzi, T. S. Cinotti, and A. Costanzo, "Fall detection and 3-D indoor localization by a custom RFID reader embedded in a smart e-health platform," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 12, pp. 5329–5339, Dec. 2019.
- [5] H. Bagheri, M. Noor-A-Rahim, Z. Liu, H. Lee, D. Pesch, K. Moessner, and P. Xiao, "5G NR-V2X: Toward connected and cooperative autonomous driving," *IEEE Commun. Standards Mag.*, vol. 5, no. 1, pp. 48–54, Mar. 2021.
- [6] *5G:NG-RAN:NR Positioning Protocol A (NRPPa)*, document 38.455, version 15.0.0, 3GPP, 2018. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138400_138499/138455/15.00.00_60/ts_138455v15000p.pdf
- [7] S. Bartoletti, A. Conti, D. Dardari, and A. Giorgetti, "5G localization and context-awareness," Univ. Bologna, Bologna, Italy, Univ. Ferrara, Ferrara, Italy, Tech. Rep., 2018.
- [8] G. Naik, J.-M. Park, J. Ashdown, and W. Lehr, "Next generation Wi-Fi and 5G NR-U in the 6 GHz bands: Opportunities and challenges," *IEEE Access*, vol. 8, pp. 153027–153056, 2020.
- [9] C. S. Álvarez-Merino, H. Q. Luo-Chen, E. J. Khatib, and R. Barco, "WiFi FTM, UWB and cellular-based radio fusion for indoor positioning," *Sensors*, vol. 21, no. 21, p. 7020, Oct. 2021.
- [10] F. Morselli, S. Bartoletti, M. Z. Win, and A. Conti, "Localization in 5G ecosystem with Wi-Fi," in *Proc. IEEE 22nd Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Sep. 2021, pp. 441–445.
- [11] A. Sciarone, C. Fiandrino, I. Bisio, F. Lavagetto, D. Kliazovich, and P. Bouvry, "Smart probabilistic fingerprinting for indoor localization over fog computing platforms," in *Proc. 5th IEEE Int. Conf. Cloud Netw. (Cloudnet)*, Oct. 2016, pp. 39–44.
- [12] ETSI, *Study on NR Positioning Support*, document TS 38.855, Version 16.0.0, Release 16, 3GPP, Mar. 2019.
- [13] X. Song, X. Fan, C. Xiang, Q. Ye, L. Liu, Z. Wang, X. He, N. Yang, and G. Fang, "A novel convolutional neural network based indoor localization framework with WiFi fingerprinting," *IEEE Access*, vol. 7, pp. 110698–110709, 2019.

- [14] X. Ye, X. Yin, X. Cai, A. P. Yuste, and H. Xu, "Neural-network-assisted UE localization using radio-channel fingerprints in LTE networks," *IEEE Access*, vol. 5, pp. 12071–12087, 2017.
- [15] A. Satan and Z. Toth, "Development of Bluetooth based indoor positioning application," in *Proc. IEEE Int. Conf. Future IoT Technol. (Future IoT)*, Jan. 2018, pp. 1–6.
- [16] P. Teunissen and O. Montenbruck, *Springer Handbook of Global Navigation Satellite Systems*. Berlin, Germany: Springer, 2017.
- [17] D. Zhu, J. Choi, and R. W. Heath, "Auxiliary beam pair enabled AoD and AoA estimation in closed-loop large-scale millimeter-wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4770–4785, Jul. 2017.
- [18] J. H. Lee, B. Shin, D. Shin, J. Kim, J. Park, and T. Lee, "Precise indoor localization: Rapidly-converging 2D surface correlation-based fingerprinting technology using LTE signal," *IEEE Access*, vol. 8, pp. 172829–172838, 2020.
- [19] K. Maneerat and K. Kaemarungsi, "RoC: Robust and low-complexity wireless indoor positioning systems for multifloor buildings using location fingerprinting techniques," *Mobile Inf. Syst.*, vol. 2019, pp. 1–22, Feb. 2019.
- [20] Y.-X. Ye, A.-N. Lu, M.-Y. You, K. Huang, and B. Jiang, "Wireless localization based on deep learning: State of art and challenges," *Math. Problems Eng.*, vol. 2020, pp. 1–8, Oct. 2020.
- [21] S. J. Hayward, K. van Lopik, C. Hinde, and A. A. West, "A survey of indoor location technologies, techniques and applications in industry," *Internet Things*, vol. 20, Nov. 2022, Art. no. 100608.
- [22] R. Kawecki, S. Hausman, and P. Korbil, "Performance of fingerprinting-based indoor positioning with measured and simulated RSSI reference maps," *Remote Sens.*, vol. 14, no. 9, p. 1992, Apr. 2022.
- [23] C. Lamoureux and R. Chelouah, "Fusion particle and fingerprint recognition for indoor positioning system on mobile," *Eng. Appl. Artif. Intell.*, vol. 98, Feb. 2021, Art. no. 104082.
- [24] B. Lashkari, J. Rezazadeh, R. Farahbakhsh, and K. Sandrasegaran, "Crowdsourcing and sensing for indoor localization in IoT: A review," *IEEE Sensors J.*, vol. 19, no. 7, pp. 2408–2434, Apr. 2019.
- [25] L. Yang, N. Wu, B. Li, W. Yuan, and L. Hanzo, "Indoor localization based on factor graphs: A unified framework," *IEEE Internet Things J.*, vol. 10, no. 5, pp. 4353–4366, Mar. 2023.
- [26] Z. Gu, Z. Chen, Y. Zhang, Y. Zhu, M. Lu, and A. Chen, "Reducing fingerprint collection for indoor localization," *Comput. Commun.*, vol. 83, pp. 56–63, Jun. 2016.
- [27] J. Purohit, X. Wang, S. Mao, X. Sun, and C. Yang, "Fingerprinting-based indoor and outdoor localization with LoRa and deep learning," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.
- [28] X. Li, E. Leitinger, and F. Tufvesson, "RSS-based localization of low-power IoT devices exploiting AoA and range information," in *Proc. 54th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2020, pp. 651–656.
- [29] C. S. Alvarez-Merino, H. Q. Luo-Chen, E. J. Khatib, and R. Barco, "Opportunistic fusion of ranges from different sources for indoor positioning," *IEEE Commun. Lett.*, vol. 25, no. 7, pp. 2260–2264, Jul. 2021.
- [30] C. Wang, J. Yang, H. He, R. Zhou, S. Chen, and X. Jiang, "Neighbor cell list optimization in handover management using cascading bandits algorithm," *IEEE Access*, vol. 8, pp. 134137–134150, 2020.
- [31] S. Shang and L. Wang, "Overview of WiFi fingerprinting-based indoor positioning," *IET Commun.*, vol. 16, no. 7, pp. 725–733, 2022.
- [32] M. W. P. Maduranga and R. Abeysekera, "TreeLoc: An ensemble learning-based approach for range based indoor localization," *Int. J. Wireless Microw. Technol.*, vol. 11, no. 5, pp. 18–25, Oct. 2021.
- [33] M. W. P. Maduranga and R. Abeysekera, "Supervised machine learning for RSSI based indoor localization in IoT applications," *Int. J. Comput. Appl.*, vol. 183, no. 3, pp. 26–32, May 2021.
- [34] T. Janssen, R. Berkvens, and M. Weyn, "Comparing machine learning algorithms for RSS-based localization in LPWAN," in *Proc. Int. Conf. P2P, Parallel, Grid, Cloud Internet Comput.*, Antwerp, Belgium, Cham, Switzerland: Springer, 2019, pp. 726–735.
- [35] R. M. M. R. Rathnayake, M. W. P. Maduranga, V. Tilwari, and M. B. Dissanayake, "RSSI and machine learning-based indoor localization systems for smart cities," *Eng.*, vol. 4, no. 2, pp. 1468–1494, May 2023.
- [36] M. I. AlHajri, N. T. Ali, and R. M. Shubair, "Indoor localization for IoT using adaptive feature selection: A cascaded machine learning approach," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, pp. 2306–2310, 2019.
- [37] Y. Zhang, D. Li, and Y. Wang, "An indoor passive positioning method using CSI fingerprint based on AdaBoost," *IEEE Sensors J.*, vol. 19, no. 14, pp. 5792–5800, Jul. 2019.
- [38] A. Ç. Seçkin and A. Coskun, "Hierarchical fusion of machine learning algorithms in indoor positioning and localization," *Appl. Sci.*, vol. 9, no. 18, p. 3665, Sep. 2019.
- [39] C. Wang, S. Xu, and J. Yang, "AdaBoost algorithm in artificial intelligence for optimizing the IRI prediction accuracy of asphalt concrete pavement," *Sensors*, vol. 21, no. 17, p. 5682, Aug. 2021.
- [40] H. Zhang, D. Nettleton, and Z. Zhu, "Regression-enhanced random forests," 2019, *arXiv:1904.10416*.
- [41] I. Ilic, B. Görgülü, M. Cevik, and M. G. Baydogan, "Explainable boosted linear regression for time series forecasting," *Pattern Recognit.*, vol. 120, Dec. 2021, Art. no. 108144.
- [42] M. Malmström, I. Skog, S. M. Razavi, Y. Zhao, and F. Gunnarsson, "5G positioning—A machine learning approach," in *Proc. 16th Workshop Positioning, Navigat. Commun. (WPNC)*, Oct. 2019, pp. 1–6.
- [43] X. Guo, N. Ansari, L. Li, and H. Li, "Indoor localization by fusing a group of fingerprints based on random forests," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4686–4698, Dec. 2018.
- [44] L. Mentch and S. Zhou, "Randomization as regularization: A degrees of freedom explanation for random forest success," *J. Mach. Learn. Res.*, vol. 21, no. 1, pp. 6918–6953, 2020.
- [45] J. Li, Z. Chen, Z. Ran, Y. Xu, X. Zhu, and X. Yuan, "The GNSS spoofing detection method based on AdaBoost," in *Proc. 6th Int. Symp. Auto. Syst. (ISAS)*, 2023, pp. 1–6.
- [46] Y. Zhang, C. Wu, and Y. Chen, "A low-overhead indoor positioning system using CSI fingerprint based on transfer learning," *IEEE Sensors J.*, vol. 21, no. 16, pp. 18156–18165, Aug. 2021.
- [47] K. Arunachalam, S. S. Gantha, J. M. Ppallan, S. Jaiswal, S. Song, and A. Nigam, "Mobility aware Socket Layer (MaSL) for seamless connectivity in mobile networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2021, pp. 1–7.
- [48] C. N. Tadros, M. R. M. Rizk, and B. M. Mokhtar, "Software defined network-based management for enhanced 5G network services," *IEEE Access*, vol. 8, pp. 53997–54008, 2020.
- [49] W. Lehr, F. Queder, and J. Haucap, "5G: A new future for mobile network operators, or not?" *Telecommun. Policy*, vol. 45, no. 3, Apr. 2021, Art. no. 102086.
- [50] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo, and G. Seco-Granados, "Survey of cellular mobile radio localization methods: From 1G to 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 1124–1148, 2nd Quart., 2018.
- [51] C. Gentner, M. Ulmschneider, I. Kuehner, and A. Dammann, "WiFi-RTT indoor positioning," in *Proc. IEEE/ION Position, Location Navigat. Symp. (PLANS)*, Apr. 2020, pp. 1029–1035.
- [52] Scikit Learn. (2022). *AdaBoostRegressor*. Accessed: Feb. 1, 2023. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostRegressor.html>
- [53] N. Jakowski and M. M. Hoque, "Estimation of spatial gradients and temporal variations of the total electron content using ground-based GNSS measurements," *Space Weather*, vol. 17, no. 2, pp. 339–356, Feb. 2019.
- [54] A. Developer. (2021). *Wi-Fi Location: Ranging With RTT*. Accessed: Feb. 2, 2022. [Online]. Available: <https://developer.android.com/guide/topics/connectivity/wifi-rtt>



CARLOS SIMÓN ÁLVAREZ-MERINO received the B.Sc. degree in telecommunication technologies and the M.Sc. degree in telecommunication engineering and in telematics and networks from the University of Málaga, Spain, in 2017 and 2019, respectively. He is currently pursuing the Ph.D. degree. He is also a Research Assistant with the Communications Engineering Department.



EMIL J. KHATIB (Member, IEEE) received the Ph.D. degree, in 2017, with a focus on machine learning, big data analytics, and knowledge acquisition applied to the troubleshooting in cellular networks. He is currently a Postdoctoral Researcher with the University of Málaga. He has participated in several national and international projects related to 5G management and Industry 4.0 projects. Currently, he is working on the topic of security and localization in B5G/6G networks.



ANTONIO TARRÍAS MUÑOZ received the M.S. degree in telematics and telecommunication networks from the University of Malaga, Spain, in 2020. He is currently pursuing the Ph.D. degree in cellular communications. In 2019, he was a Research Assistant and an Assistant Professor with the Department of Communications Engineering.



HAO QIANG LUO-CHEN received the B.Sc. degree in sound and image engineering and the M.Sc. degree in telematics and telecommunication networks from the University of Málaga, Spain, in 2018 and 2021, respectively.



RAQUEL BARCO MORENO received the M.Sc. and Ph.D. degrees in telecommunication engineering. She is currently a Full Professor with the University of Malaga. She was with Telefónica and European Space Agency. She participated in the Mobile Communication Systems Competence Center, jointly created by Nokia and the University of Malaga. She has published more than 100 scientific articles, filed several patents, and led projects with major companies.

...