APPLIED RESEARCH

A Generative Adversarial Network (GAN) Fingerprint Approach Over LTE

LUIGI SERRELI[1], **(Member, IEEE), MAURO FADDA**[2], **(Senior Member, IEEE),**
ROBERTO GIRAU[3], **(Member, IEEE), PIETRO RUIU**[2],
DANIELE D. GIUSTO[1], **(Senior Member, IEEE),**
AND MATTEO ANEDDA[1], **(Senior Member, IEEE)**

[1]Department of Electrical and Electronic Engineering (UdR CNIT of Cagliari), University of Cagliari, 09123 Cagliari, Italy
[2]Department of Biomedical Science, University of Sassari, 07100 Sassari, Italy
[3]Department of Computer Science and Engineering (DISI), University of Bologna, 40126 Bologna, Italy

Corresponding author: Matteo Anedda (matteo.anedda@unica.it)

ABSTRACT Recent advancements in communication technologies have significantly enhanced localization techniques, improving both accuracy and operating modes. Initially, localization methods relied on global navigation satellite systems, offering high accuracy but proving inefficient in Non-Line-of-Sight scenarios. Furthermore, the absence of a passive mode, where the user can be localized without explicitly requesting it, renders these methods unsuitable for applications like passive tracking systems. Fingerprinting methods, a pattern matching techniques based on signal power estimation from target devices and distance estimation from reference points, can be seen as a valid and promising alternative. However, these methods face limitations due to extensive measurement campaigns needed to establish accurate sampling systems within specific areas and the substantial amount of data required for machine learning algorithms to achieve optimal performance. This study introduces a novel fingerprinting method capable of passive operation, involving all smartphones within a designated area, suitable for both indoor and outdoor scenarios. The proposed solution leverages Generative Adversarial Networks (GANs) to augment fingerprinting datasets, enhancing machine learning models' capabilities. Additionally, the offline phase's cost-effectiveness is improved by integrating a Bayesian system as a secondary machine learning component.

INDEX TERMS Fingerprinting, generative adversarial network, LTE, localization.

I. INTRODUCTION

In the last years, the development of new technologies has improved the accuracy for locating people, animals, or objects. Classic localization methods are based on wireless signals and were introduced by the Global Navigation Satellite System (GNSS) [1]. Despite this technology achieves high accuracy, it suffers in indoor environments (i.e., Non-Line-of-Sight scenarios - NLOS) for which this system is almost unusable. Another issue stems from the system being designated as active, necessitating users to

The associate editor coordinating the review of this manuscript and approving it for publication was Haipeng Yao.

indicate their intent for localization explicitly. This constraint renders the system impractical for various purposes, such as scenarios involving vehicular tracking (i.e., a passive tracking system). The ubiquity of smartphones has made instant communication with the surrounding world commonplace while also providing a myriad of functionalities that enhance daily life. Moreover, the vast majority of the population now has access to cutting-edge cellular networks, including 4G and even more advanced technologies. This advanced connectivity not only allows for faster internet browsing but also opens doors to a world of opportunities in terms of communication, work, and access to information at any time and place.

The fingerprinting technique is a type of radio frequency localization combined with signal characteristics such as Receive Signal Strength Indicator (RSSI) or Time of Arrival (ToA). RSSI and ToA are strictly linked to the type of communication between receiver and transmitter, which can be in line-of-sight (LOS) or non-line-of-sight (NLOS) conditions. In the NLOS case, the signal is affected by multipath and, therefore, is received as multiple paths leading to a reduction of power due to reflection, diffraction, and diffusion, if compared to the LOS case. On the contrary, in the case of ToA, the multipath is noticeable due to the different arrival times of the signal, which could be linked to a longer path that the signal took due to the lack of visibility between the transmitter and receiver.

Fingerprinting can be based on a range of communication technologies such as WiFi, GSM, and Bluetooth systems. Localization relies on pattern-matching techniques, where the core idea is to capture the power of signals received from target devices and compute the distance from strategic or reference points. The signal parameter or characteristic used as a reference is pre-measured, enabling subsequent comparisons when the system is operational and ready to locate terminals.

The fingerprinting technique consists of two phases: (i) the offline phase is commonly defined as the calibration phase [2], involves gathering reference characteristics to construct a radio map; (ii) The online phase is responsible for comparing the characteristics of a new signal that is not yet present on the map and for determining the receiver's position in space through suitable calculations or artificial intelligence methods.

The primary challenge associated with fingerprinting lies in the offline phase, a stage that, depending on the scenario, demands significant cost and time investment due to the lengthy and extensive measurement campaigns required [3]. These campaigns would necessitate a highly accurate system to systematically sample, step by step, the entire reference area. Alternative approaches to manually gathering the samples entail leveraging machine learning (ML) techniques [4], [5]. ML models are trained using real data to predict signal characteristics even in parts of the environment not covered during the data collection process. However, ML requires substantial data to achieve optimal performance for their models.

In order to increase the number of samples used for the training, data augmentation techniques can be used, which generate synthetic data from existing datasets [6], [7]. Among these techniques, Generative Adversarial Networks (GANs), are prominent methods used to augment fingerprinting datasets [8], [6], [9].

GANs was introduced by Goodfellow et al. in 2014 [10] representing a specific family of generative networks where two systems are trained simultaneously in a competitive manner:, a generator G and a discriminator D, respectively. G is designed to learn the probability distribution of the training data, while D aims to determine whether a sample comes from the training data or has been generated by G. The two networks compete to the extent that the training procedure for G maximizes the probability of D making a mistake.

Based on these assumptions, this work aims to develop a localization technique using fingerprinting of LTE signals and capable of operating in indoor and outdoor environments. A system leveraging GANs was employed as a modification of the method presented in [6] to augment real data collected during measurement campaigns. The use of GANs was mainly driven by the challenge of data sparsity. By generating realistic synthetic data, GANs significantly enhance our dataset's density, eliminating the need for extensive manual sampling, which is often impractical and time-consuming. Additionally, GANs produce synthetic data that reflects a wider range of signal variations caused by dynamic changes in real-world environments. Therefore the goal was to build a real-time system able to convert RSSI samples of users interacting with a base station (BS), also recognizing the user in a domestic or urban setting, where keeping track of user locations is essential for various purposes such as safety, security, and efficient urban management [11]. In pursuit of this objective, LTE technology was chosen owing to its maturity and widespread accessibility. LTE networks enjoy a well-established global presence, providing expansive coverage crucial for ensuring dependable localization. This significance is particularly pronounced in indoor settings, where LTE signals exhibit superior penetration through building materials compared to 5G. Additionally, the prevalence of LTE-compatible devices, as opposed to those supporting 5G, ensures greater compatibility and usability [12]. To accurately simulate an LTE environment, we employed the srsLTE software simulator.

The achieved spatial resolution is deemed highly precise, particularly considering that the primary aim of this study is to discern the presence of a mobile device with greater accuracy compared to traditional smartphone GPS positioning. It is noteworthy that this study does not aspire to attain even greater precision and accuracy, which would necessitate the concurrent utilization of cross-cutting technologies such as LTE, 5G, and WiFi [13]. Two reference scenarios, i.e., an indoor scenario consisting of a room in the University of Cagliari and an outdoor scenario relating to a dataset of RSSI fingerprints found online, were implemented to test the performance of the proposed system.

The rest of the paper is structured as follows. Section II outlines the technical background and related works on fingerprinting methods. Hardware and software configurations of the proposed system are presented in section III, while the methodology is detailed in section IV. Test and results performed to evaluate the proposed solution both in indoor and outdoor scenarios are discussed in section V. Finally, the conclusions are drawn in section VI.

II. BACKGROUND AND RELATED WORKS

The technique of fingerprinting LTE signals for localization purposes has been extensively explored in the literature.

An evaluation of fingerprinting using a simplified statistical matching algorithm with two similarity metrics is proposed in [14]. The automatic collection of radio frequency (RF) information is facilitated through the Minimization of Drive Testing (MDT) in LTE networks addressing non-ideal cell detection scenarios. Results indicate a substantial reduction of 76%, 34%, and 70% in computation time for rural, urban, and Hetnet cases, respectively.

Yan et al., proposes a CHAN-IPDR-ILS system for accurate indoor localization, featuring a step length estimation method that enhances accuracy by considering information from the fusion CHAN and the improved pedestrian dead reckoning (PDR) indoor localization system (CHAN-IPDR-ILS) [15]. Pedestrian localization is achieved through a motion model that combines acoustic estimation and dynamic improved PDR estimation. The fusion localization approach sets thresholds and confidence levels to mitigate accidental and cumulative errors.

Other studies, such as in [16], suggest strategies for Narrowband Internet of Things (NB-IoT) positioning using fingerprinting, incorporating coverage and radio data from multiple cells. The proposed strategies consistently achieve a minimum average positioning error of approximately 20 meters across various network scenarios, contrasting with the current state-of-the-art's average error of about 70 meters.

In [17] Pecoraro et al., propose approaches based on Channel State Information (CSI) instead of RSSI, demonstrating as methods based on RSSI shows inadequate performance in LTE networks. Additionally, the system, relying on CSI shape and statistical descriptors rather than direct CSI vectors, showcases more stable performance, particularly when there are changes in room configuration, such as moving furniture.

Zhang et al., present a fingerprint-based localization technique based on deep learning framework to collect real time channel state information knowledge from LTE eNodeB and to extract intrinsic features [18]. A time domain fusion approach is proposed to assemble multiple positioning estimations improving localization accuracy and robustness (i.e., Mean Distance Error of 0.47 meters for indoor and of 19.9 meters for outdoor scenarios) compared to previous methods in literature.

Compared to the existing literature, this study has focused on making the offline phase more cost-effective by integrating two machine learning systems into the system: a GAN approach and a Bayesian system. In addition, the proposed method was validated through practical experimentation by instrumentation, generating an LTE cell and testing various mobile devices with a view to both indoor and outdoor localization.

Concerning the data augmentation with GANs, various approaches can be found in the literature, primarily leveraging WiFi or Bluetooth signals. Below, the most relevant
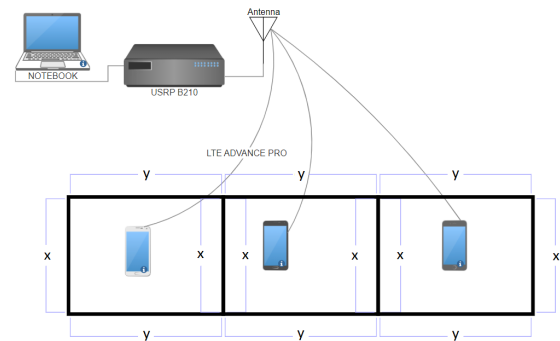

FIGURE 1. Hardware set-up.

literature approaches, highlighting the differences compared to the proposed solution, are presented.

In [8], a deep-learning-based framework for outdoor localization using a rich features set in LTE networks, called DeepFeat, has been proposed. To optimize computational efficiency and reduce complexity, a feature selection module is incorporated into the deep learning model, resulting in a notable 20.6% reduction in computation and complexity. Chi-squared algorithms to strategically narrow down the feature set to 12 inputs, irrespective of the area size, has been used. Additionally, to bolster the accuracy of the DeepFeat system, a One-to-Many augmenter is introduced, expanding the dataset and contributing to an overall improvement in system performance. The main differences with our study concern the generation of the augmented dataset. The technique called *One-to-Many* employed in [8] involves duplicating samples in k geographic directions. In contrast, in the proposed approach, a GAN is employed to learn the probability distribution from a real dataset and generate additional data.

In [6], Nabati et al. propose a method to reduce the data collection costs for human-centric localization systems. They leverage fingerprint-based localization and apply data augmentation through deep learning, utilizing received signal strength (RSS) or channel state information (CSI) in wireless sensor networks for user localization in various environments. GANs are employed to learn the distribution from a limited dataset, generating synthetic data to enhance positioning accuracy alongside real collected data. Experimental results on a benchmark dataset demonstrate that the proposed method, using a mix of 10% real data and 90% synthetic data, achieves positioning accuracy comparable to utilizing the entire dataset of collected data. The method employed in the current work is derived from the data generation approach used in [6]. The main difference lies in the type of signal utilized: in our case, an LTE signal is employed, whereas the paper uses WiFi signals.

Emerging research highlights the potential of GANs in addressing data scarcity issues. Specifically, studies have demonstrated the efficacy of GANs in generating data points to compensate for gaps caused by inaccessible data collection sites or signal degradation. In [9], the authors introduce a

data collection system leveraging a mobile robot. However, the robot is limited to reaching areas without obstacles. To overcome this limitation, GANs are employed to generate data in regions inaccessible to the robot.

Moreover in [19], GANs are utilized to generate data for Human Behavior Analysis. This is particularly beneficial in scenarios where there may be short periods of person tracking measurement losses, leading to gaps in tracking information.

III. EXPERIMENTAL SET-UP

This section provides a comprehensive overview of the hardware and software configuration implemented for the experimental setup. Additionally, it includes a description of the two distinct datasets utilized for outdoor and indoor scenarios. Following the methodology adopted in previous studies, we opted to utilize a single BS for LTE fingerprinting. This decision was influenced by the substantial cost associated with deploying multiple base stations [20]. Nonetheless, even with a single BS, exceptional localization outcomes can be attained by harnessing RSSI data from multiple reference points within the BS's coverage area [21], [22]. The hardware setup, depicted in Fig. 1, comprises a USRP Ettus B210 board, a MSI GE63 8RE laptop running Ubuntu 18.10, an LTE antenna, and several smartphones equipped with USIMs. The technical specifications of the laptop and LTE antenna are provided in Table 1, respectively.

The device employed for transmitting and receiving at the LTE working frequencies in the implementation of the real system is the Universal Software Radio Peripheral (USRP) [23]. The USRP is designed as a hardware platform that interfaces with a software processing system. Essentially, this device enables communication with User Equipments (UEs) as long as it can operate within the frequency bands of LTE technology.

The USRP Ettus B210 is an integrated board that guarantees a frequency range from 70MHz to 6 GHz and has been designed to be a low-cost board that combines an AD9361 transceiver, provides a bandwidth of 56 MHz, a programmable Field Programmable Gate Array (FPGA) and USB 3.0 connectivity. Several smartphones, including models from Samsung, Huawei, Xiaomi, and Apple, were utilized to simulate the UE. These smartphones were equipped with Universal Subscriber Identity Module (USIM), and their Mobile Country Code (MCC) and Mobile Network Code (MNC) had to align with the values declared in the srsLTE software.

To replicate a realistic LTE environment, an LTE network software simulator was employed. The srsLTE is an open-source software that facilitates the establishment of an end-to-end mobile network operating in 4G access mode. This software is comprised of three main components:

- **srsUE**, a comprehensive software for utilizing a device such as a *UE*;
- **srsENB**, a software that plays the role of eNodeB, providing access for UEs and sending data to the *Core Newtwork*;

TABLE 1. Technical specifications of the devices used for the experiments.

| Device	HW type	Specifications
Laptop	CPU	8th Gen. Intel®CoreT Mi7
	Graphics	GeForce®GTX 1060 with 6GB GDDR5
	RAM	16GB
	Storage	1x M.2 SSD 256GB
		1x M.2 SSD 128GB
		1x HDD 1TB
	I/O ports	1x Type-C USB3.1 Gen2
		1x Type-A USB3.1 Gen2
		1x RJ45
		1x SD (XC/HC)
		1x (4K @ 60Hz) HDMI
		1x Mini-DisplayPort
		2x Type-A USB3.1 Gen1
	Operating System	Ubuntu 18.04.5 LTS
USRP	Size	235x120x30mm
	Frequency	700MHz-2700MHz
	Standing Wave Ratio	≤ 2.5
	Input impedance	50Ω
	Cable length	2m
	Connector type	SMA
Antenna	Radiation Type	Vertical and Horizontal
	Model	LF-ANT4G01
	Frequency Range	800 MHz to 2700 MHz
	Polarization	Vertical, Horizontal
	Maximum input Power	200W
	Gain	88dBi
	Input Impedance	50Ω
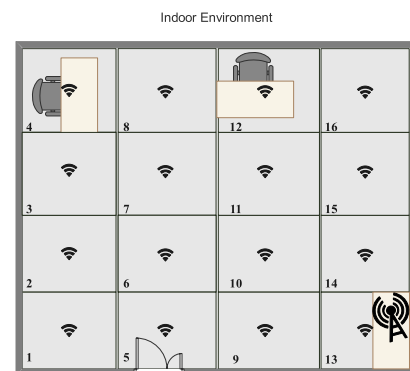
FIGURE 2. Indoor scenario.

- **srsEPC**, a lightweight implementation of the Core Network of an LTE system, including MME, HSS and Serving Gateway (SGW)/ Packet Data Network Gateway (PGW).

The srsLTE software runs under Linux and is compatible with various USRPs, including the Ettus B210.
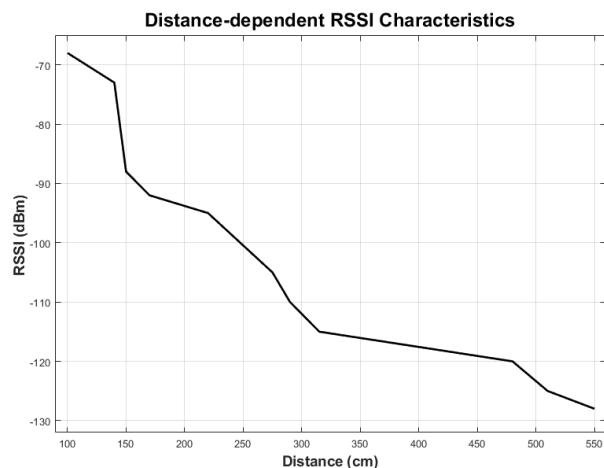
FIGURE 3. Pathloss calculated considering distance variation in indoor environment.

For the purpose of this work, two different datasets were utilized, characterizing two different scenarios. In the outdoor case, a dataset derived from a large area of a parking lot divided into 15-meter square cells was used. It's worth noting that the size of the cell determines the resolution of the accuracy. In the outdoor scenario, various time-varying variables can impact the system's performance. Typically, to address these challenges, multiple antennas are employed to enhance the coupling between RSSI and the reference cell. In the indoor scenario, a single antenna was chosen to be used to have a term of comparison with the outdoor case. Obviously, the accuracy of the two configurations is affected by the number of antennas considered and the number of samples per cell. Details on the configuration will be provided in section V.

IV. METHODOLOGY

The initial tests were conducted in an anechoic chamber to ensure a controlled environment. Using the srsLTE software, we simulated an LTE network to establish secure communication and collect fingerprint patterns along with the IMSI (International Mobile Subscriber Identity) of the user. This allowed us to receive a user identifier and store the encrypted value associated with the fingerprints.

An omnidirectional LTE antenna was positioned in the corner of the anechoic chamber, and the system automatically recorded data for each cell, considering the option to relocate the smartphone after a certain number of RSSI measurements. Within 30 minutes, the system recorded a sufficient number of samples to cover 10 selected cells. Modifications were made to the srsLTE code to enable the matching of RSSI with IMSI. This involved utilizing Wireshark for Python (Pyshark) to capture communication packets between the E-nodeB and the EPC software, where the IMSI is transmitted in plain text.

The experiment begins with the offline phase, obtaining the fingerprint dataset using the srsLTE software. Once srsLTE is started in both eNodeB and EPC components, the smartphones begin to perform the procedure for connecting to the network, releasing their identifier and thus starting the phase of measurement. As shown in Fig. 2, in the indoor scenario a grid of 16 cells is utilized, and within each 30 RSSI samples are collected, resulting in a total of 480 samples. Basically, it took 15 minutes as the goal is to have a rapid and low-cost measurement campaign. In Fig. 3, the pathloss based on the measurements conducted in an indoor environment is illustrated, as the variation of the distance from the antenna.

A. GENERATIVE AND BAYESIAN NETWORKS

The GAN used for augmenting the dataset is based on the work in [6] which built a network suitable for generating RSSI samples trained on a dataset of 250 samples for each reference cell. A schematic diagram of the GAN network used for this work is depicted in Fig. 4.

The network is initially trained with 25 samples per class, representing 10% of the training data. Subsequently, it is retrained using the remaining 90% of the data to generate synthetic data. This process is repeated 100 times using different random seeds to prevent and mitigate random effects.

With respect to the network proposed by Nabati et al., various adjustments were made to adapt it to our input data. Specifically, changes were implemented in the training parameters, including adjustments to the batch size, the number of GAN epochs, and the iteration count, to find the optimal configuration for our dataset.

After some preliminary experiments, we observed that the accuracy with 16-class discrimination was very low. For this reason, numerous Naïve-Bayes networks in parallel with different discrimination criteria and classes have been considered.

Naïve-Bayes network with multiple layers working in parallel can be used to extrapolate the probability that a certain RSSI belongs to a particular class $\mathcal{N} = \{n_1, \ldots, n_i, \ldots, n_m\}$, where n_i is the generic cell.

The initial layer of the network attempts to position the input data within two classes, denoted as $\mathcal{C}_1$ and $\mathcal{C}_2$, each comprising 8 cells on the grid. $\mathcal{C}1$ contains cells n_1 through $nm/2$, while $\mathcal{C}2$ contains cells $nm/2 + 1$ through n_m. As the network progresses through its inner layers, the number of classes increases, culminating in the final stage with 16 classes. Each layer will therefore use a different configuration, and each class $\mathcal{C}_i$ will be associated with a different set of cells, defining different *patterns*. Fig. 5 shows an example of a 4-level network with two pattern types, 2-class and 4-class, respectively.

A *pattern* is a configuration which splits the set of m starting classes into k classes with $k \leq m$. Seven different patterns $\mathcal{P}_{j,k}$ have been analyzed, where $\mathcal{P}$ indicates a pattern that brings j classes into k classes:

- $\mathcal{P}_{16,2}$ represents two classes containing two rows each, as illustrated in Fig. 6a;
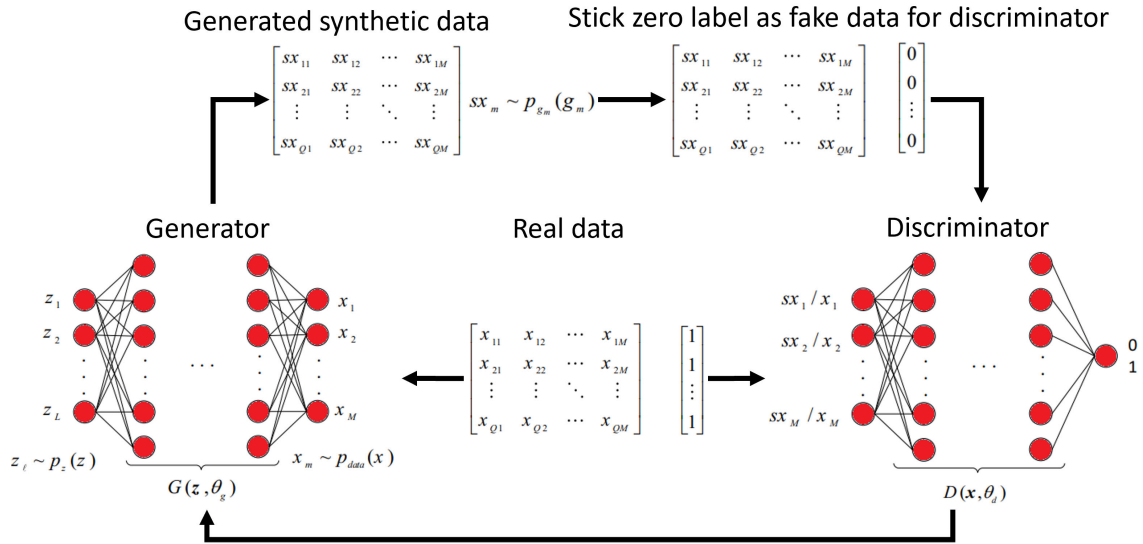
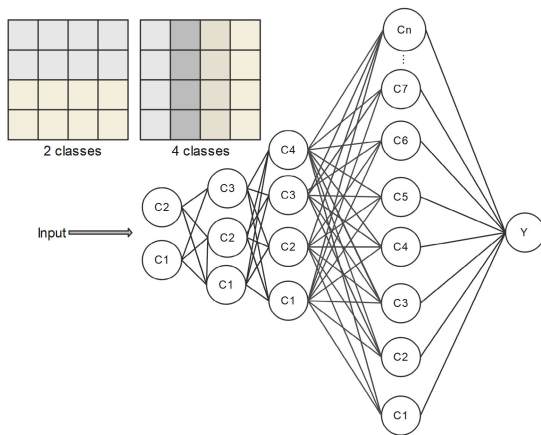**FIGURE 4.** GAN network architecture used to generate RSSI samples.

FIGURE 5. Bayesian neural network structure and examples of cell's patterns.

- $\mathcal{P}_{16,2}$: two classes containing two column each, as shown in Fig. 6b;
- $\mathcal{P}_{16,2}$: two classes containing two column each with a different configuration, as shown in Fig. 6c;
- $\mathcal{P}_{16,4}$: four quadrants as depicted in Fig. 6d;
- $\mathcal{P}_{16,4}$: four classes one for each row of the grid, Fig. 6e;
- $\mathcal{P}_{16,4}$: four classes one for each column of the grid, Fig. 6f;
- $\mathcal{P}_{16,16}$: the grid is split in 16 classes, each corresponding to one cell of the grid Fig. 6g;

B. POSITION CALCULATION AND ERROR EVALUATION
Each i-th layer of the Bayesian network has a probability that the RSSI r_j entering the system belongs to a certain class k. Therefore, the output from each layer is defined as $p_{i,k}$ where i is the i-th layer of the network, while k is the class to which

it belongs. The goal is to determine a score that identifies the highest probability where r_j precisely corresponds to the i-th cell. Therefore, P_k is defined as the probability that adheres to the relationship in Eq. 1

$$P_k = \sum_{i=1}^{n} P_{i,k} \forall k \in [1, m] \tag{1}$$

where n is equal to the number of layers in the network and m is the number of classes k for layer i.

The localization error is defined as the average Euclidean distance between the center of the real cell where the device is actually located and the center of the cell where the device has been localized. Starting from the assumption that it can be found at any point inside the cell, it is assumed that all the points inside the cell are still considered to be at the center. Therefore, since each cell is a square with side L, the minimum error that can be made is equal to $\frac{L}{2}$ if the user is in the cell border.

V. TEST AND RESULTS
This section introduces two reference scenarios: an indoor scenario comprising a room at the University of Cagliari and an outdoor scenario based on a dataset of RSSI fingerprints obtained from online sources.

A. INDOOR SCENARIO
The room was divided during the offline phase into 16 adjacent squares forming a 4×4 grid with the size of each side of a square equal to 1 m. As the room is a relatively stable environment, a fixed point was selected for the antenna placement, simulating an eNodeB of the LTE architecture. This setup allows for the identification of cells that consistently remain within direct visibility of the antenna. The sensitivity of the system is defined as half the side
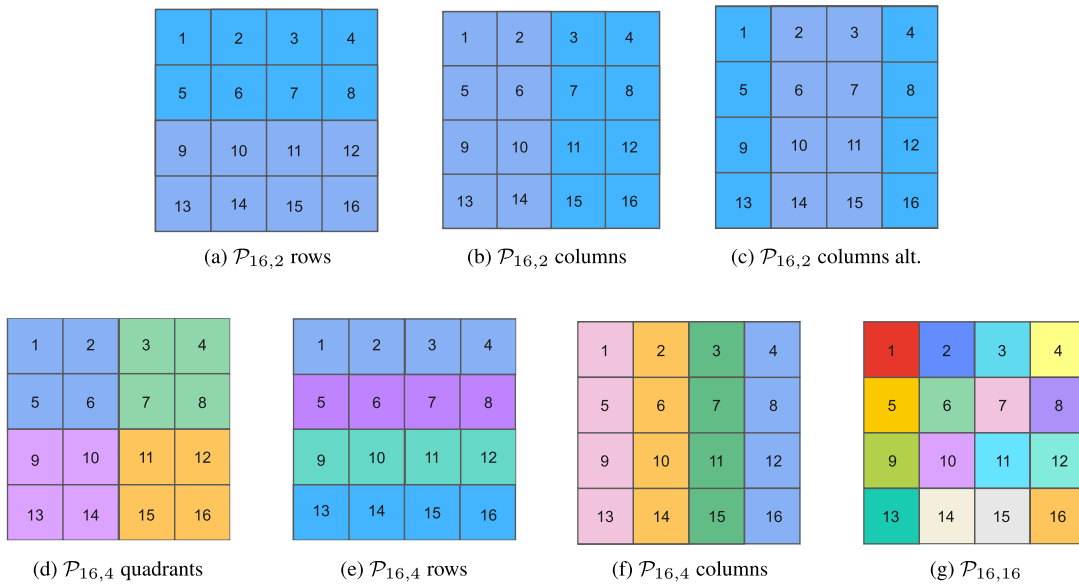
(a) $\mathcal{P}_{16,2}$ rows (b) $\mathcal{P}_{16,2}$ columns (c) $\mathcal{P}_{16,2}$ columns alt.

(d) $\mathcal{P}_{16,4}$ quadrants (e) $\mathcal{P}_{16,4}$ rows (f) $\mathcal{P}_{16,4}$ columns (g) $\mathcal{P}_{16,16}$

FIGURE 6. Different configuration in classes.

TABLE 2. Accuracy of the various layers of the Bayesian network.

	Pattern Type	Train Accuracy	Test Accuracy
Rows	$\mathcal{P}_{16,2}$	92.2 %	77%
Columns	$\mathcal{P}_{16,2}$	80.1%	55.2%
Alternate columns	$\mathcal{P}_{16,2}$	86.2%	77.1%
Quarter	$\mathcal{P}_{16,4}$	81.2%	59.3%
Rows	$\mathcal{P}_{16,4}$	80.1%	55.2%
Columns	$\mathcal{P}_{16,4}$	83.3%	60.4%
16 Classes	$\mathcal{P}_{16,16}$	79.3%	36.5%

of a single cell (i.e. $\pm 0.5m$), as there is no possibility of further refining the measurement. a total of 30 samples were collected during this phase. Fig. 2 shows the grid of the room.

The first goal is to evaluate the accuracy of the prediction of the Bayesian network using the RSSI dataset acquired during the offline phase. Each layer of the neural network was trained with 80% of the dataset and tested with the remaining 20%. Table 2 shows the accuracy level of the training and test part for each layer, as well as the pattern.

The test conducted with 16 classes revealed inadequate accuracy. Implementing only this level would lead to an imprecise system. Conversely, dividing the grid into two rows provides excellent accuracy, but the only available information would be whether the localized user is before or beyond the center of the room. The confusion matrices for the training and testing of the first layer of the network, respectively, are shown in Fig. 7a and in Fig. 7b. The accuracy is calculated as reported in Eq. 2, where *TP* stands for *True Positive*, *TN* for *True Negative*, *FP* for *False Positive* and *FN* for *False Negative*. The labels 1 and 2 indicate just how the initial set of cells was divided into two classes, class 1, and class 2.

2.

$$Accuracy = \frac{TP + TN}{TN + FP + FN + TP} \quad (2)$$

Applying the position calculation rule, which involves integrating information from all layers of the network, yields an accuracy of 73.75% with a distance error of approximately 66 cm. However, it was observed that the alternating columns pattern does not contribute to improving accuracy. Conversely, excluding it from the network allows for slightly higher accuracy, reaching 76.25% with an error of 64 cm.

1) RESULTS WITH AUGMENTED DATA

Several network configurations, including different epochs and percentages of real data, were experimented with for training the GAN network. The resulting training and testing accuracy varying these parameters are shown in Fig. 8. The highest accuracy is attained when the network is trained with 100% real data. However, it's notable that when using only 40% real data, the model doesn't stabilize even after 50 epochs, yet it steadily improves, reaching a respectable accuracy of 75%. This implies that increasing the number of epochs, even with a smaller portion of real data, can lead to substantial improvements in accuracy.

After training the GAN network with 100% of the dataset, it was employed to generate three additional datasets, consisting of 200, 400, and 700 samples for each class. In the case of the largest dataset, comprising 700 samples per class, a dataset totaling $700 \times 16 = 11,200$ samples was obtained. This dataset was then merged with the original dataset to proceed with the Bayesian approach.

Three tests were conducted with the Bayesian network, using 200, 400, and 700 samples, respectively. In the first
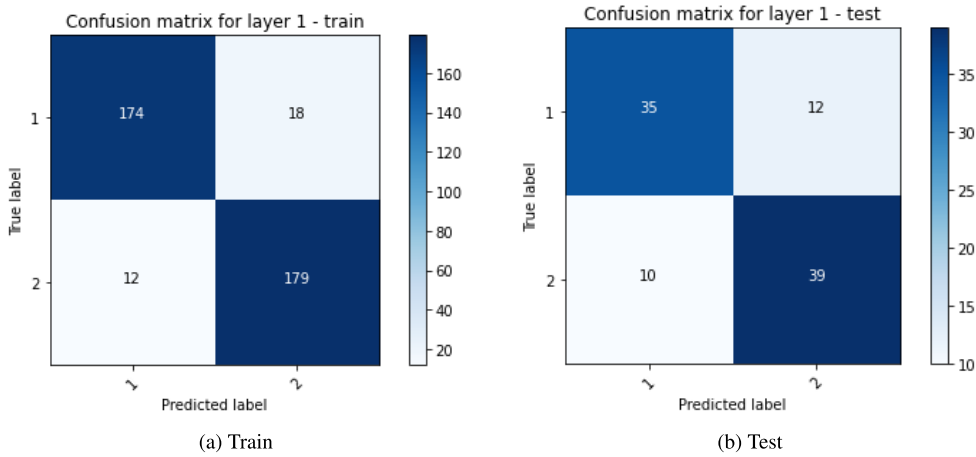
(a) Train
(b) Test

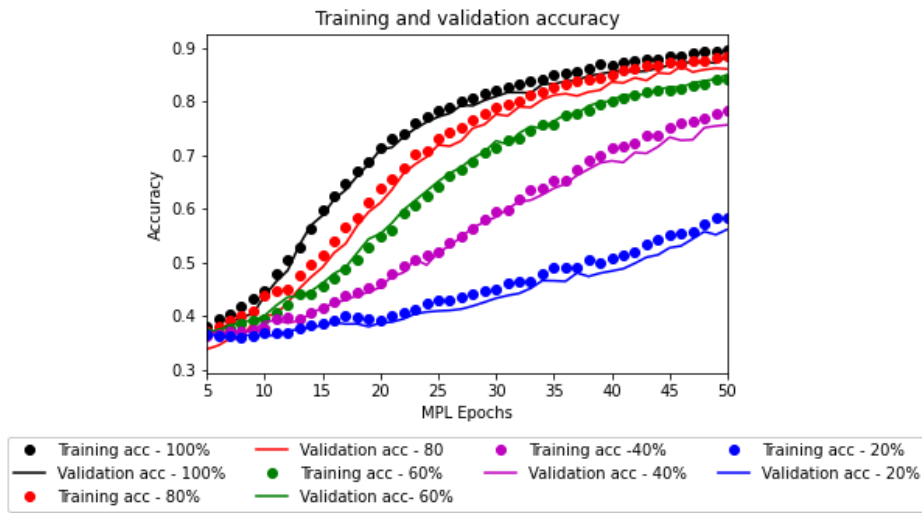**FIGURE 7.** Confusion matrix training the first layer of the network.

FIGURE 8. Accuracy of the GAN network training with different amounts of input data.

test, an accuracy of 77.5% with a precision of 68 cm was achieved. The second test showed an improved accuracy of 92.5% with a precision of 55 cm. Finally, in the third test, the accuracy remained high at 90.6%, with a precision of 58 cm. Comparing the results with the test that does not utilize the dataset generated by the GAN network, as shown in Table 3, an improvement in performance is evident with 400 and 700 samples. Although the first test achieved a higher success rate, the accuracy dropped slightly. For the other two cases, notable improvements in both accuracy and precision in terms of distance are noted. Between the two, the 700 samples appear too heavy for the Bayesian network which is affected by noise. The success rates for the three tests are shown in Fig. 9, respectively 200 (Fig. 9a), 400 (Fig. 9b), and 700 samples (Fig. 9c), calculated on a dataset of 160 samples (10 samples per cell) taken in the offline phase. In the heat-maps of Fig. 9, a lighter color indicates a more accurate prediction. The last two cases are much more precise and

TABLE 3. Performance with different number of input data.

Input	Accuracy [%]	Error [m]
Bayes	76.25	0.64
GAN-200 + Bayes	77.5	0.68
GAN-400 + Bayes	92.5	0.55
GAN-700 + Bayes	90.6	0.58

only cell 8 is problematic. This could be due to the presence of a table that produced scattering effects during the offline phase.

B. OUTDOOR SCENARIO
The Bayesian network was initially tested in an outdoor scenario. Subsequently, the same tests were repeated after augmenting the dataset with samples generated by the GAN network. As in the previous scenario, our expectation was to improve localization precision by integrating GAN data
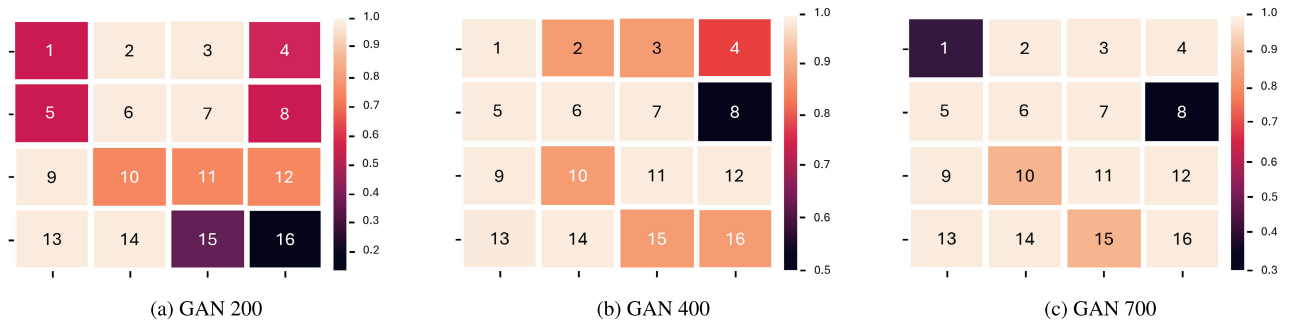
(a) GAN 200   (b) GAN 400   (c) GAN 700

FIGURE 9. Indoor scenario: heat maps of the success rates relating to the three Bayes+GAN tests.

as input for the Bayesian network. However, it's important to note that the enhancement in precision may not be as substantial as in the indoor case. The outdoor environment poses greater challenges due to its larger and more chaotic nature, characterized by dynamic obstacles such as vehicles and buses. The obstacles introduce phenomena like slow fading and NLOS conditions, resulting in multiple signal paths. Consequently, the performance of the system is compromised compared to the indoor scenario, where there is consistent visibility between the transmitter and receiver.

1) DATASET

The dataset used in this scenario was obtained online [24]. It involves a rather large environment involving a parking lot. The location was divided into square cells with sides $L = 15m$. This assumes that the precision of the system cannot be less than L as it is defined as the sensitivity of the localization system [25].

For the purpose of this work only one antenna was considered to compare the outdoor case with the indoor case. This dataset is composed of 25 cells (i.e., 5×5) with 500 samples per cell. To adapt the network already built for the indoor case, 16 adjacent cells (i.e., 4×4) were chosen, as in the previous case. Furthermore, only a small part of the dataset was used, namely 80 samples per cell. The rest of the dataset is used for testing purposes.

2) AUTONOMOUS NAÏVE-BAYES NETWORK

Due to the abundance of samples per cell in the outdoor dataset, we standardized the sample distribution to match that of the indoor case. Initially, we selected and tested 30 samples per cell using the Bayesian network without incorporating the GAN network. An accuracy of 69% was achieved with 30 samples, yielding a precision of 9.84 meters. This result indicates lower precision compared to the indoor setting. Consequently, the sample size was increased to 80. With this adjustment, the Bayesian network achieved a slightly higher accuracy of 73.4% with a precision of 9.53 meters. It is noteworthy that the precision deteriorated compared to the indoor scenario due to the larger cell size in this dataset. However, with this configuration and sample size,

TABLE 4. Accuracy of layers of the Bayesian network for the outdoor scenario.

	Pattern Type	Train Accuracy	Test Accuracy
Rows	$\mathcal{P}_{16,2}$	84.1 %	79.6%
Columns	$\mathcal{P}_{16,2}$	77.5%	72.3%
Alternate columns	$\mathcal{P}_{16,2}$	71.1%	69%
Quarters	$\mathcal{P}_{16,4}$	73%	56.2%
Rows	$\mathcal{P}_{16,4}$	70.9%	60.1%
Columns	$\mathcal{P}_{16,4}$	72.6%	60.5%
16 Classes	$\mathcal{P}_{16,16}$	77.7%	77.1%

comparable accuracy levels to those in the indoor scenario were achieved.

Also in this case, 80% of the dataset was chosen to train each layer of the network while 20% was used for the test.

By comparing the results shown in Table 4 with those in Table 2 which describe the results of the training and testing of each layer of the Bayesian network, outdoor and indoor respectively, we notice a substantial decrease in performance for the outdoor case. This is due to the particularly problematic scenario in terms of propagation.

3) RESULTS WITH AUGMENTED DATA

The GAN network was trained with part of the dataset found online with 80 samples per cell. Three datasets were subsequently generated, with 200, 300, 1000, and 2000 samples per each of the 16 cells (i.e., 4×4), respectively. Thus, for the largest case (i.e., 2000 samples per cell) a total of 32,000 samples were obtained. The methods and parameters used for the GAN network are the same as for the indoor case.

For the Bayesian network, the results were obtained using the same patterns as the indoor case. Table 5 presents the accuracy and precision, measured in meters, of the results categorized by the number of samples generated by the GAN network while maintaining a constant 80 samples from the original dataset.

The observed data indicates an increase in both accuracy and precision up to 500 samples. Beyond this point, further increasing the number of samples generated by the GAN
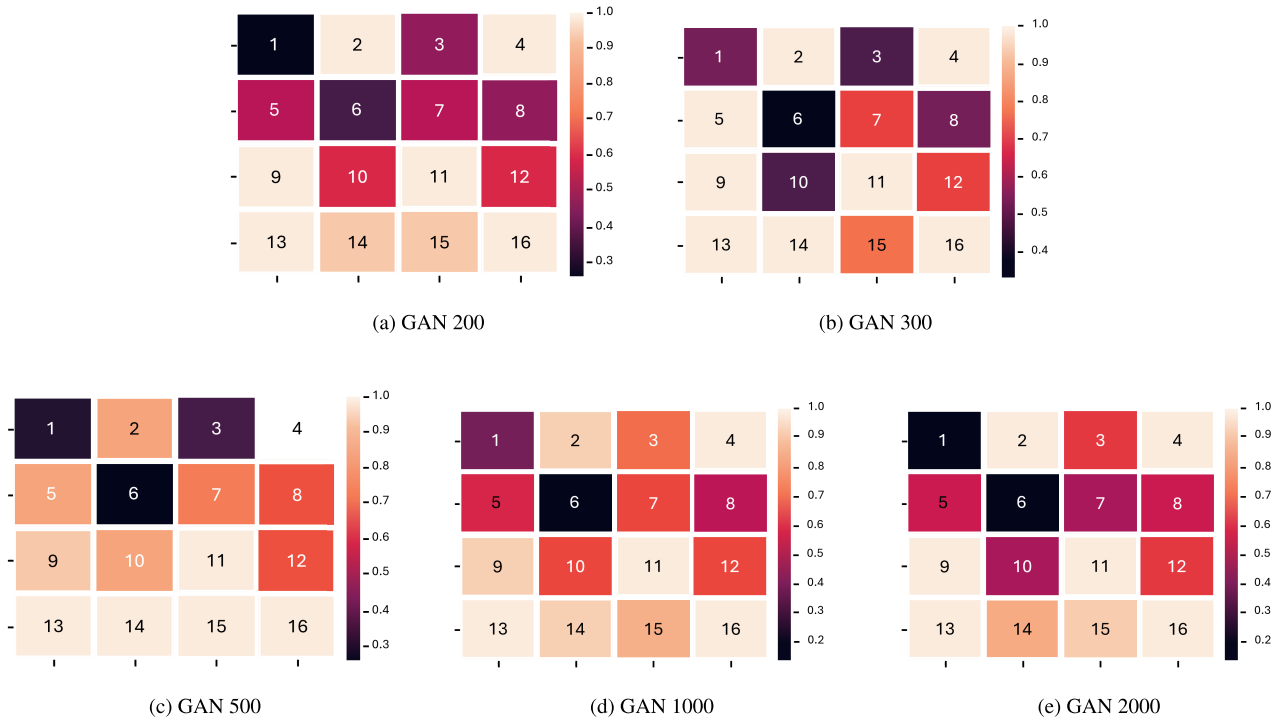
(a) GAN 200

(b) GAN 300

(c) GAN 500

(d) GAN 1000

(e) GAN 2000

FIGURE 10. Outdoor scenario: heat maps of the success rates relating to the five Bayes+GAN tests.

TABLE 5. Bayes + GAN test results in outdoor scenario.

GAN Samples	Accuracy [%]	Error [m]
200	75.83	9.4
300	78.75	9.1
500	80	9
1000	75.83	9.34
2000	74.16	9.53

network results in a decline in performance. This suggests that the network may begin to interpret the additional generated data as noise.

Figures 10a, 10b, 10c, 10d, and 10e depict heatmaps illustrating the success rate of individual cells. Cells closer to the antenna, located near cell 13, exhibit higher success rates. Conversely, cell number 6 appears completely black in cases 300, 500, and 1000, indicating a very low success rate, approaching 20% of identified positions. Additionally, the network encounters several issues, particularly on line number 2, which are partially addressed with the inclusion of 300-500 samples. This line corresponds to the subdivision made by the first pattern, which demonstrates higher accuracy, leading to clear differences in intersection areas. However, pronounced slow-fading and fast-fading events during the offline phase when obtaining the initial dataset may have influenced the results.

VI. CONCLUSION

This study introduces an autonomous LTE-based system designed for user positioning within its coverage area using fingerprinting techniques. A novel algorithm leveraging

IMSI recognition was developed to determine if a user has previously been located within a specific time-frame. Results demonstrate the reasonable capability of the autonomous network, employing the Bayesian approach, to differentiate between various cells during testing. Significant enhancements were observed by incorporating the GAN network into the proposed approach, leading to improved success rates and distance estimation precision, both indoors and outdoors. In the outdoor scenario, accuracy was compromised due to the chaotic nature of the environment, characterized by factors such as varying obstacles, moving vehicles, and changing conditions. Conversely, the indoor environment provided a more controlled setting, since NLOS communications was not needed The application of a GAN trained on a dataset consisting of 400 samples, in conjunction with a Bayesian network, resulted in an accuracy of 92.5%. Furthermore, a precision of 55 centimeters was achieved specifically within indoor scenarios.

Given the outdoor scenario, the application of a GAN trained on 500 samples, combined with a Bayesian network, resulted in an accuracy of 80%. It's important to emphasize that the primary focus of the study does not lie in achieving high accuracy for outdoor localization, as GPS technology is readily available and typically sufficient for outdoor scenarios.

REFERENCES

[1] A. Raskaliyev, S. H. Patel, T. M. Sobh, and A. Ibrayev, "GNSS-based attitude determination techniques—A comprehensive literature survey," *IEEE Access*, vol. 8, pp. 24873–24886, 2020.

[2] C. Xiang, S. Zhang, S. Xu, and G. C. Alexandropoulos, "Self-calibrating indoor localization with crowdsourcing fingerprints and transfer learning," 2021, *arXiv:2101.10527*.

[3] L. Serreli, R. Nonnis, G. Bingöl, M. Anedda, M. Fadda, and D. D. Giusto, "Fingerprint-based positioning method over LTE advanced pro signals with GAN training contribute," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Aug. 2021, pp. 1–5.

[4] X. Wang, X. Wang, and S. Mao, "CiFi: Deep convolutional neural networks for indoor localization with 5 GHz Wi-Fi," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[5] N. Singh, S. Choe, and R. Punmiya, "Machine learning based indoor localization using Wi-Fi RSSI fingerprints: An overview," *IEEE Access*, vol. 9, pp. 127150–127174, 2021.

[6] M. Nabati, H. Navidan, R. Shahbazian, S. A. Ghorashi, and D. Windridge, "Using synthetic data to enhance the accuracy of fingerprint-based localization: A deep learning approach," *IEEE Sensors Lett.*, vol. 4, no. 4, pp. 1–4, Apr. 2020.

[7] Hnavidan, "LocalizationGAN: Implementation of the paper using synthetic data to enhance the accuracy of fingerprint-based localization: A deep learning approach," *IEEE Sensors Lett.*, vol. 4, no. 4, pp. 1–4, Apr. 2020, Art. no. 6000204. [Online]. Available: https://github.com/hnavidan/LocalizationGAN

[8] A. Mohamed, M. Tharwat, M. Magdy, T. Abubakr, O. Nasr, and M. Youssef, "DeepFeat: Robust large-scale multi-features outdoor localization in LTE networks using deep learning," *IEEE Access*, vol. 10, pp. 3400–3414, 2022.

[9] H. Zou, C.-L. Chen, M. Li, J. Yang, Y. Zhou, L. Xie, and C. J. Spanos, "Adversarial learning-enabled automatic WiFi indoor radio map construction and adaptation with mobile robot," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 6946–6954, Aug. 2020.

[10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–12.

[11] M. Spanu, M. Bertolusso, G. Bingöl, L. Serreli, C. G. Castangia, M. Anedda, M. Fadda, M. Farina, and D. D. Giusto, "Smart cities mobility monitoring through automatic license plate recognition and vehicle discrimination," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Aug. 2021, pp. 1–6.

[12] T. Abubakr and O. A. Nasr, "Novel LSTM-based approaches for enhancing outdoor localization accuracy in 4G networks," *IEEE Access*, vol. 11, pp. 140103–140115, 2023.

[13] O. M. Gul, M. Kulhandjian, B. Kantarci, A. Touazi, C. Ellement, and C. D'amours, "Secure industrial IoT systems via RF fingerprinting under impaired channels with interference and noise," *IEEE Access*, vol. 11, pp. 26289–26307, 2023.

[14] Y. Han, H. Ma, L. Zhang, and L. L. Chen, "Performance evaluation of simplified matching algorithms for RF fingerprinting in LTE network," in *Proc. IEEE 9th Int. Conf. Anti-counterfeiting, Secur., Identificat. (ASID)*, Sep. 2015, pp. 24–28.

[15] J. Yan, Z. Huang, and X. Wu, "Smartphone based indoor localization using machine learning and multi-source information fusion," *IEEE Trans. Aerosp. Electron. Syst.*, early access, pp. 1–14, 2024.

[16] L. De Nardis, G. Caso, Ö. Alay, U. Ali, M. Neri, A. Brunstrom, and M.-G. Di Benedetto, "Positioning by fingerprinting with multiple cells in NB-IoT networks," in *Proc. Int. Conf. Localization GNSS (ICL-GNSS)*, Jun. 2022, pp. 01–07.

[17] G. Pecoraro, E. Cianca, S. Di Domenico, and M. De Sanctis, "LTE signal fingerprinting device-free passive localization robust to environment changes," in *Proc. Global Wireless Summit (GWS)*, Nov. 2018, pp. 114–118.

[18] H. Zhang, Z. Zhang, S. Zhang, S. Xu, and S. Cao, "Fingerprint-based localization using commercial LTE signals: A field-trial study," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2019, pp. 1–5.

[19] A. Belmonte-Hernández, G. Hernández-Peñaloza, D. M. Gutiérrez, and F. Álvarez, "Recurrent model for wireless indoor tracking and positioning recovering using generative networks," *IEEE Sensors J.*, vol. 20, no. 6, pp. 3356–3365, Mar. 2020.

[20] S. Abbas, Q. Nasir, D. Nouichi, M. Abdelsalam, M. Abu Talib, O. Abu Waraga, and A. U. R. Khan, "Improving security of the Internet of Things via RF fingerprinting based device identification system," *Neural Comput. Appl.*, vol. 33, no. 21, pp. 14753–14769, Nov. 2021.

[21] J. He and H. C. So, "A hybrid TDOA-fingerprinting-based localization system for LTE network," *IEEE Sensors J.*, vol. 20, no. 22, pp. 13653–13665, Nov. 2020.

[22] Y. Wang, S. Han, Y. Tian, C. Xiu, and D. Yang, "Is centimeter accuracy achievable for LTE-CSI fingerprint-based indoor positioning?" *IEEE Access*, vol. 8, pp. 75249–75255, 2020.

[23] E R Brand A National Instruments. *USRP Software Defined Radio (SDR) Online Catalog*. Accessed: May 16, 2024. [Online]. Available: https://www.ettus.com/products/

[24] (2021). *Rodrig1Alon*. [Online]. Available: https://github.com/rodrig1Alon/FingerprintingTLEDS

[25] X.-M. Yu, H.-Q. Wang, and J.-Q. Wu, "A method of fingerprint indoor localization based on received signal strength difference by using compressive sensing," *EURASIP J. Wireless Commun. Netw.*, vol. 2020, no. 1, pp. 1–13, Dec. 2020.
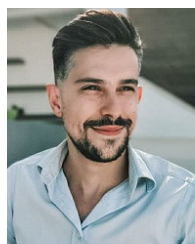
LUIGI SERRELI (Member, IEEE) received the bachelor's degree in electronic and computer engineering from the University of Cagliari, in 2019, the master's degree (Hons.) in internet engineering, in 2021, and the Ph.D. degree in electronic and computer engineering, with a research has focused on the Internet of Things and natural language processing, with an emphasis on generative large language models and deep learning technologies. His master's thesis "A Generative Adversarial Network (GAN) Fingerprint Approach Over LTE." He is currently engaged in projects related to digital twin technologies for IoT and service discovery.

MAURO FADDA (Senior Member, IEEE) received the Ph.D. degree in electronic and computer engineering from the University of Cagliari, in 2013. In 2020, he was a Visiting Professor with the Department of Electronics and Computers, Transilvania University of Brasov, Romania. He is currently an Assistant Professor with the University of Sassari, Italy. In March 2020, he was the Elected Chair of Italian Chapter of the Broadcast Technology Society of the Institute of Electrical and Electronic Engineering (IEEE). He has served as the chair for various international conferences and workshops. He is an Associate Editor of IEEE ACCESS and a Topic Editor of *Sensors*.

ROBERTO GIRAU (Member, IEEE) received the M.S. degree in telecommunication engineering and the Ph.D. degree in electronic engineering and computer science from the University of Cagliari, Cagliari, Italy, in 2012 and 2017, respectively. From 2012 to 2020, he was a Researcher with the Department of Electrical and Electronic Engineering, University of Cagliari, developing an experimental platform for the social the Internet of Things. He has been a Research Fellow with the Department of Computer Science and Engineering, University of Bologna, Bologna, Italy, since 2021. His main research interests include the IoT with particular emphasis on its integration with social networks, software engineering, smart cities, and cloud computing.

PIETRO RUIU received the master's degree in telecommunication engineering and the Ph.D. degree in electric, electronic, and communication engineering from the Polytechnic University of Turin, in 2006 and 2018, respectively. From 2013 to 2018, he held the position of the Head of the Infrastructure and Systems for Advanced Computing (IS4AC) Research Unit, Istituto Superiore Mario Boella (ISMB). He is currently an Assistant Professor with the University of Sassari, Italy. His research interests include computer vision, face recognition, machine learning, computing infrastructures, data center networks and architecture, and automatic resource provisioning. He has actively contributed to the academic community by serving as a technical program committee (TPC) member for international conferences and as a reviewer for esteemed international journals.

DANIELE D. GIUSTO (Senior Member, IEEE) received the Laurea (M.S.) degree in electronic engineering and the Dottorato di Ricerca (Ph.D.) degree in telecommunications from the University of Genoa, Genoa, Italy, in 1986 and 1990, respectively. He has been a Full Professor of telecommunications with the University of Cagliari, Cagliari, Italy, since 2002, where he has been a permanent Faculty Member with the Department of Electrical and Electronic Engineering, since 1994. His research interests include smart cities, sensor networks and the IoT, mobile and professional networks, and digital media. He was a member of the IEEE Standard Committee, from 2007 to 2010. He was a recipient of the IEEE Chester Sall Paper Award, in 1998, and the AEI Ottavio Bonazzi Best Paper Award, in 1993. He was Italian Head of Delegation in the ISO-JPEG Committee, from 1999 to 2018. He has been Italian Head of Delegation in the ISO-Smart Cities Committee, since its foundation in 2016.

MATTEO ANEDDA (Senior Member, IEEE) received the M.Sc. degree (summa cum laude) in telecommunication engineering and the Ph.D. degree in electronic and computer engineering from the University of Cagliari, in 2012 and 2017, respectively. He was a Visiting Erasmus Student with the University of the Basque Country, Bilbao, Spain, in 2010, for eight months, where he carried out the M.Sc. thesis under the supervision of Prof. P. Angueira. He was a Visiting Researcher with Dublin City University, Ireland, under the supervision of Prof. G. Muntean, in 2015, and Universidad de Montevideo, Uruguay, under the supervision of Prof. R. Sotelo, in 2016, for seven months each. He has been a Research Fellow with the Department of Electrical and Electronic Engineering, University of Cagliari, since 2017. In 2020, he was a Short Visiting Professor with the Department of Electronics and Computers, Transilvania University of Brasov, Romania. His research interests include real-time applications, 5G networks and network selection, the IoT and smart cities, adaptive multimedia streaming, and the heterogeneous radio access environment. He is a senior member of the IEEE Broadcast Technology, IEEE Communications, and IEEE Vehicular Technology Societies.

● ● ●