Received October 14, 2021, accepted November 21, 2021, date of publication November 23, 2021,
date of current version December 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3130429

A Novel Approach for User Equipment Indoor/Outdoor Classification in Mobile Networks

PEDRO ALVES[1], THAÍNA SARAIVA[2], MARÍLIA BARANDAS[1,3], DAVID DUARTE[2,5], DINIS MOREIRA[1], RICARDO SANTOS[1,3], RICARDO LEONARDO[1,3], HUGO GAMBOA[1,3], (Senior Member, IEEE), AND PEDRO VIEIRA[4,5]

[1]Associação Fraunhofer Portugal Research, 4200-135 Porto, Portugal
[2]Consultoria em Telecomunicações, Celfinet, 1495-764 Cruz Quebrada, Portugal
[3]Laboratório de Instrumentação, Engenharia Biomédica e Física da Radiação (LIBPhys-UNL), Departamento de Física, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCT-NOVA), Monte da Caparica, 2829-516 Caparica, Portugal
[4]Instituto Superior de Engenharia de Lisboa (ISEL), 1959-007 Lisboa, Portugal
[5]Instituto de Telecomunicações (IT), Instituto Superior Técnico, 1049-001 Lisboa, Portugal

Corresponding author: Pedro Alves (pedro.alves@fraunhofer.pt)

This work was supported in part by the project Geolocation Non-Assisted by GPS for Mobile Networks in Indoor and Outdoor Environment (GARMIO), co-funded by Portugal 2020, framed under the COMPETE 2020 (Operational Programme Competitiveness and Internationalization); and in part by the European Regional Development Fund (ERDF) from European Union (EU) under Grant POCI-01-0247-FEDER-033479.

ABSTRACT The ability to locate users and estimate traffic in mobile networks is still one of the major challenges when it comes to planning and optimizing the networks. Since indoor location is not always possible or precise, having the ability to distinguish indoor from outdoor traffic can be a valuable alternative and/or improvement. In this paper, two different machine learning algorithms are presented to classify a user's environment, whether indoor or outdoor, using only data from a Long Term Evolution (LTE) network. To test both algorithms, two different measurement campaigns were done. Both campaigns used a smartphone to gather data from the user's side. The first measurement campaign was done across 6 different cities, ranging from small rural areas to large urban environments, while the second was only done on a large urban city. On the second campaign, Network Traces (NT) data was also collected from the network side. The first algorithm consists on a Random Forest (RF) and the second relies on a Long Short Term Memory (LSTM), thus covering both more traditional machine learning and deep learning approaches. The results varied from 0.75 to 0.91 on the F1-Score, depending on the validation strategy, showing promising results.

INDEX TERMS Indoor outdoor detection, machine learning algorithms, long term evolution, measurement campaigns, smartphone, network traces.

I. INTRODUCTION

According to [1], 4th Generation (4G) technology will remain the dominant mobile access network for a while longer. During the first quarter of 2021, 4G subscriptions increased around 100 million, reaching 58% of all mobile subscriptions. It is projected to peak during the year at 4.8 billion subscriptions, before dropping to around 3.9 billion subscriptions by the end of 2026, when the subscribers will migrate to 5th Generation (5G). Also, the global 4G population coverage was over 80% at the end of 2020 and is forecast to reach around 95% in 2026. 4G networks are evolving to deliver increased network capacity and faster data throughputs, and the Communication Service Providers (CSPs) are continuously adapting their service packaging towards consumers. Improving network planning by estimating indoor traffic demand will contribute to more efficient network deployments. Predicting an accurate indoor traffic ratio is especially useful to operators rolling out high frequency coverage. Traditionally, it has been assumed that almost 80% of mobile data traffic is generated indoors. Now, methods are being developed to accurately estimate the proportion of traffic in outdoor base stations that is due to indoor usage.

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

In urban environments, the main mobile traffic is generally indoor, which is hard to serve from outdoor macro base stations due to radio signal attenuation. Considering the high indoor traffic demand, the mobile operators may make in-building solutions more economic. Still, to properly serve the outdoor traffic, macro site densification may be needed. Having a realistic classification of the indoor traffic ratio provides a solid indicator for network investment decisions at mid-term. A specific research shows that in a dense urban scenario, the average indoor traffic from outdoor cells was about 37%, increasing for 64% in urban environments. These results suggest that the operator could consider deploying in-building solutions where possible and then augmenting the number of outdoor small cells. By quantifying the traffic demand and coverage from both inside and outside, the additional resources that would be needed can be determined so that the least amount of radio power is sacrificed to penetration losses [1].

The extensive usage of smartphones in the modern society makes these devices a crucial platform that serves the communication, entertainment and work needs of many people. Also, the ability to access the internet anywhere represent a strong motivation for producing mobile applications based on Location-based Systemss (LBSs). The LBS field plays an important role in many domains, including tracking, navigation, safety-related services, advertising, tourism, healthcare monitoring, intelligent transportation, among other services. The commercial location services can be used for value-added location-aware purposes, such as local area advertisement and targeted marketing. However, all mobile applications based on LBS have the current user positioning as common requirement. Since mobile users can be in many places such as open sky outdoors, crowded avenues, indoor environments, the next generation of positioning systems has to perform well both indoors and outdoors [2]. At the engineering level, the location services available internally inside the Mobile Network Operators (MNOs) can use the end-user location to assist in planning and optimizing their networks, such as optimizing the handover parameters in terms of user mobility, detecting overshooting, coverage holes, pilot pollution and interference situations, detecting cells with capacity problems, among others. For instance, a very dense and high traffic service area can be optimized by adding more capacity to the base stations, increasing the Quality of Service (QoS) and Quality of Experience (QoE) of each user. To perform this, an Indoor/Outdoor traffic classification can be the way to adapt the network to enhance the user services, in the environment type they are in. Understanding if the users are inside or outside the buildings in a certain area can leverage the operators to optimize the network in terms of antenna tilting, power transmission, configured frequency bands or configuration parameters, among others, in order to compensate (or not) the additional penetration losses. This can even lead to a specific radio planning in areas high-rise or highly occupied buildings, such as using outdoor small cells or indoor distributed antenna systems.

The environment where the connection occurs has a strong influence on user expectations. Thus, estimating the context of a specific session is a key to evaluating user experience in mobile network management. Examples of contextual factors are the device types, user age and gender, previous experiences, time of day and mainly the user location. Specifically, the user location has a fundamental role, as it determines the demanded services, the link conditions or the expected service performance. To sum up, a precise estimation of the user environment is critical for evaluating the QoE [3] [4].

Geo-positioning in mobile networks continues to be demanding for future applications and use cases like the Internet of Things (IoT), emergency services and vehicle-to-everything (V2X) [5], [6]. Recently, this type of technique is being designed to address requirements and needs of several 5G verticals, in the business, public and entertainment spheres. By using radio measurements collected by mobile devices on wireless networks, it is possible to estimate their own positions, assuming that the considered location is not shared by the devices through signaling.

Another important aspect within the MNO business is the drive testing. These measuring campaigns are often used in radio planning and optimization, namely for propagation model tunning, network troubleshooting and benchmarking. Conventional drive testing is a manual, inefficient, and expensive process of collecting radio network information by conducting measurement campaigns. Typically, a vehicle equipped with measurement devices and a Global Positioning System (GPS) receiver to obtain geographical location, is used to collect and analyse the radio conditions. For indoor environments, engineers perform ''walk tests'', using measurement devices that can be carried by hand [11]. Thus, Minimization of Drive Tests (MDT) is the concept that allows the operators to use own users' devices to collect radio measurements and associated location information, in order to assess network performance while reducing the Operational Expenditure (OPEX) [12]. Unfortunately, MDT is rarely enabled for all users and is not continuous over time, which means that anonymous call traces provided by network equipment often lack detailed location information. Thus, the network re-planning and optimization has to be done based on network traces geolocated by prediction algorithms, with location errors of hundreds of meters, which is excessive to estimate the user connection context [3].

Considering the presented motivation and with the detected lack of efficient classifiers for indoor/outdoor traffic classification in mobile networks, this research aims to bring added value in this area. Two different approaches will be explored, the first one based on traditional machine learning techniques, and the second one considering deep learning, using LSTM algorithms. Also, two distinct data sources from a live cellular network are used. Firstly, data from Mobile Terminals (MTs) were collected by common test smartphones. This data was obtained from an in-house application developed by Fraunhofer Institute that uses the telephony manager to access and record the cellular data.

Secondly, NT are also considered. Concretely, signaling information and radio measurements exchanged between the mobile users and the network, in dedicated mode are gathered and used for training the indoor/outdoor classifier. It was considered a live LTE network based on macro Evolved NodeBs (eNBs) and with mainly three frequency bands. Because of the operator deployment strategic, micro eNBs were not used. Since the goal of this work is not the user precise positioning inside the buildings, but instead the environment classification (indoor/outdoor), it was not considered neither indoor cells nor other kind of Wi-Fi private networks, as well as GPS signalling.

In the following, the major contributions of this paper are to:

- Perform a equivalence demonstration between data collected from a smartphone (MT) and from the network side (NT);
- Explore an extensive dataset containing several cities, from densely urban to rural towns, with several daily activities;
- Present the ability to classify the end user's environment, indoor or outdoor, with two different approaches, while resorting only to the common data fields between the smartphone (MT) and the network (NT);
- Share the versatility of both solutions regarding the different real world scenarios, such as dense urban environments and also rural.

This paper is structured as follows. Section II presents the most recent work related with this topic, where some performance metrics are compared. Section III shares the adopted methodology for the identification of indoor (inside buildings) and outdoor environments. Section IV shows the model construction details for two different approaches on the mission to distinguish indoor from outdoor environments. The first approach uses a traditional machine learning algorithm and the second a deep learning algorithm. Section V presents the obtained results and analysis, being followed by the conclusions in Section VI.

II. RELATED WORK

Several methodologies have been proposed that attempt to distinguish the environment in which a mobile phone is, whether it is indoor or outdoor. These methodologies also resort to different sources of data, namely sensors available inside the mobile phone and also cell data traces available on the network side, after internal network recordings.

The most favorable results usually come from systems that use GPS, where it is very common to surpass 90% accuracy. In [15] and [16], the authors use GPS data to classify the phone's environment. In [15], a manually constructed decision tree is explored, *i.e.* the authors devise the different parameters, such as signal strength, signal to noise ratio, gain, among others, into different intervals in order to separate indoor from outdoor samples. The evaluation was performed on several testing datasets achieving an accuracy of 89 to 98%, depending on the specific dataset. The authors in [16] use a Hidden Markov Model with stacking ensemble, which was also tested with several testing datasets and achieved an accuracy between 92 and 99%. The authors tested on continuous data, and when the system encountered an environment transition, the maximum observed classification delay was of 4 seconds. The problem with the GPS usage is that it consumes a lot of energy and is only available in the smartphone. Thus, it cannot be freely used from the network side.

In [17], the authors used the magnetometer present in all modern smartphones to solve the classification problem. The base assumption is that the construction materials used in today's buildings, as well as other electrical equipment present in indoor environments will be sufficient to distinguish both environments. The chosen algorithm was a Naive Bayes and the achieved accuracy was 83%.

A very different and novel approach was presented in [2], where the authors used the accelerometer and gyroscope, which are part of the Inertial Measurement Unit (IMU) sensors present in all modern smartphones, to not only classify six activities: staying still (no activity), skip, jog, walk, going up the stairs, and going down the stairs, but also if the activity was performed indoors or outdoors. One of the findings was that, not only each individual has a unique way of doing each activity, but they also had slightly different behaviors depending if it was indoor or outdoor. The constructed dataset was collected during a five-year period, with several smartphones in different positions, and is publicly available. The chosen algorithm was the Adaboost and the authors achieved an accuracy of 99%. Similarly to ''using GPS'', the authors in [17] and [2] resort to sensors that exist solely in the smartphone, and are not available from the network side.

Regarding the use of cell data to do this classification, several works have been published.

In [18], the authors use network cell data from the smartphone, along with the light sensor, to construct a map, called CIMAP, where the different global cell Identifiers are labeled as either being an indoor cell, outdoor, indoor edge, outdoor edge or hybrid. An edge cell is a cell that mostly covers an area, either indoor or outdoor, but it still affects a small area of the opposite environment. An hybrid cell is a cell that covers both environments. The map was constructed using a crowdsourcing approach where the light sensor is the main environment indicator, and a higher intensity of light means that the smartphone is most likely outside. This information was cross-referenced between all smartphones to find matches for the same location. For example, a cell ID that is always observed when the light intensity is low, will be labeled as an indoor cell. After the map is constructed, any new Base Station (BS) connection is classified based on the light sensor if available, otherwise the cellular data is used to perform the classification. The reported result is an accuracy above 98% but no information regarding the test dataset is given, only that a real test was conducted.

The authors in [19] started by dividing the indoor/outdoor classification into four different labels, deep indoor, light indoor (close to windows), semi outdoor (close to tall buildings), and open outdoor. Then, using a smartphone, the authors collected exclusively Global System for Mobile Communications (GSM) (2G) data every 0.5 seconds and calculated statistical features, such as mean and standard deviation, to train a classifier. The training dataset was collected by volunteers in 4 specific environments on a university campus, where each volunteer walked around each site for 10 minutes. The testing dataset was a specified path on the same campus, making sure that all 4 different environments were present, but without being at the same locations that were used on the training set. The chosen classifier was a k-Nearest Neighbors (k-NN), with a 97% accuracy was achieved when making sure that at least 4 BSs were visible on each scan. When this constraint was dropped, the authors used 8-second windows and achieved a 95% accuracy using a RF. The high accuracy might be a result of the low coverage area of this study, and the proximity of the locations in which the training and testing dataset were collected.

The solution proposed in [20] uses a single smartphone with the TEMS application [21], resulting in 24912 instances of data, where each instance has: Time stamp, Timing Advance, Latitude and Longitude, Evolved Universal Mobile Telecommunications System Terrestrial Radio Access (E-UTRAN) cell identifier, Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ). From these data, the authors calculated 20 different features and created a RF of CART trees. Then, the Out-of-Bag (OOB) method is explored to train and test the RF classifier. This means that each tree was trained and tested with a different dataset. The authors also performed feature selection and varied the number of trees to find the fastest solution without compromising the results. The final model had 4 features and 15 trees and achieving a 99% accuracy. Additionally, the obtained OOB error converges to the cross validation error. However, for binary classification problems, it has been shown that OOB can overestimate the true prediction error depending on several factors, such the choices of RF parameters, the dataset size and in balanced settings [22].

The authors in [3] use two sources of information to estimate the probability of a mobile device being indoor, while establishing a connection to a LTE eNB. The first source was a map divided into small tiles, 10×10m, where each tile has a classification of land use, such as paths, offices, services, residential, open space. By knowing the location of the LTE BSs, a circular area, called Cell, is established based on the TA values, and is then divided into rings of a fixed thickness. By measuring the amount of connections that occur in each ring of each cell, and combining the percentage of area occupied by each land use type, the authors calculated the probability of an incoming connection being from an indoor device. The authors established that this probability is approximated by the indoor connection ration, *i.e.* the ratio of indoor connections in a given cell. This map based information is, according to the authors, publicly available through local municipalities or crowdsourcing. The second source of information are Cell Traces, measurements done at the BS where information regarding the connection can be obtained, such as: connection time, throughput and average RSRP. These Cell Traces are used to train a logistic regression model to estimate the same probabilities that the connection is from an indoor device. The training dataset consisted of a land area of 125 km^2, with 400 LTE BSs of which 320 were used as a training set. For the evaluation results, the authors used the correlation between the output of the map-based estimate and the model prediction, which led to R^2 values between 0.79 and 0.99.

A very extensive dataset is used in [23], collected during a 9 month period, across France, and using different smartphone models. 40% of this data was labeled. Using Drive Tests (DTs) performed by a top North American Operator in New York as a guide, the authors selected a specific portion of the dataset from Paris in order to simulate DTs. This portion only had clearly defined indoor and outdoor samples. Regarding the indoor samples, these did not contain unclear indoor locations such as balconies or open buildings. Regarding the outdoor samples, these were only pedestrian or vehicular, with a limited speed, above ground and in urban environments. From these DT, several models were trained and the results were impressive, 99% F1 weighted score, which was expectable due to the very clear nature of the data, either an open space or indoor spaces without windows and possible signal leaks. When using the original crowdsourcing data, the results were lower, 83% by choosing a Support Vector Machine (SVM) algorithm. To improve the results, the authors developed a two step classifier. The first part consisted on a Bayesian Gaussian Mixture (BGM), which is an unsupervised learning model that used the 60% of unlabeled data, as training data. The second step used a supervised model, both a SVM and a Feed Forward Neural Network (FFNN) which received the classification of the BGM as a feature along with the data. The system achieved an F1 weighted score of 89% with the SVM and 94% with the Deep Learning (DL) model.

In Table 1, a comparison between the previously mentioned works regarding data type, used algorithm, reported results and some additional considerations is presented. The first five references depend on the user equipment to gather specific data which will be used to classify the environment. Given the context presented in Section I, that the ability to classify a user environment as either indoor or outdoor is important to the planning of mobile networks, this dependence is a major deterrent to the use of these proposed solutions. In this work, we intend to use cell data as a viable alternative, which solves this problem.

Although the authors in, [3], [19], [20], [23], use only cell data, there are some technical aspects that this work attempts to improve. The first work from Wang *et al.* [19] considers a dataset that includes only a small geographic area, a university campus, and uses 2G technology, which is phasing-out and has been replaced by both 3G and 4G. Regarding, the work of Zhang *et al.* [20], the dataset is more

TABLE 1. Comparison of existing solutions for indoor and outdoor environment classification.

Work	Data	Algorithm	Results	Study considerations
[15]	GPS	Decision Tree	89-98 % Accuracy	User equipment dependency
[16]	GPS	Hidden Markov Model	92-99 % Accuracy	User equipment dependency
[17]	Magnetometer	Naive Bayes	83 % Accuracy	User equipment dependency
[2]	IMU	Adaboost	99 % Accuracy	User equipment dependency
[18]	Cell data, Light Sensor	State Machine	98 % Accuracy	User equipment dependency
[19]	Cell data	k-NN	95-97 % Accuracy	Small geographic area (university campus) and 2G technology
[20]	Cell data	RF	99% Accuracy	Medium geographic area (one urban environment)
[3]	Cell data	Logistic Regression	0.79-0.99 R^2	Large geographic area (one city metropolitan area) and map of land use requirement
[23]	Cell data	BGM + FFNN	94 % F1 weighted	Large geographic area including different types of environment

extensive than the previous one, however, it only considers an urban environment of a single city. Additionally, the train and test approach used by the authors, OOB, is debatable and can lead to over estimations of true predictions, as previously mentioned. The novelty presented in [3] depends on the accessibility to an updated map of the city and the ratio's estimation between indoor and outdoor area. A potential problem could be an outdoor event, which could skew the typical incoming traffic. The need to constantly update the map is also a major drawback.

Finally, Saffar *et al.* [23] present the most complete solution found at the time of writing, where the authors solve all of the previously mentioned problems. However, although the authors have a large dataset that includes different types of environments, the details of the data distribution by the type of environment and the ability of their algorithm to be used in a new and unseen location is not shown.

This work will focus on showing the impact on the algorithm's performance when using data from a new and unseen location from different environment types. Moreover, this work will also attempt to achieve equal or better results than the previous works using real-time data collected in collaboration with a MNO.

III. METHODOLOGY

The adopted methodology for the classification of indoor and outdoor environments consists of mainly five steps: (1) data collection; (2) data pre-processing; (3) feature extraction; (4) training and testing; (5) performance evaluation. Figure 1 shows the general schema including the five steps and its correspondence with data source, i.e., mobile terminal data or network traces. Due to the growing availability of smartphones and their easy data collection process, the primary purpose of the proposed solution was to train a model using cellular network data collected from common smartphones and then perform a validation using NT data. Therefore, the data acquisition step includes two different measurement

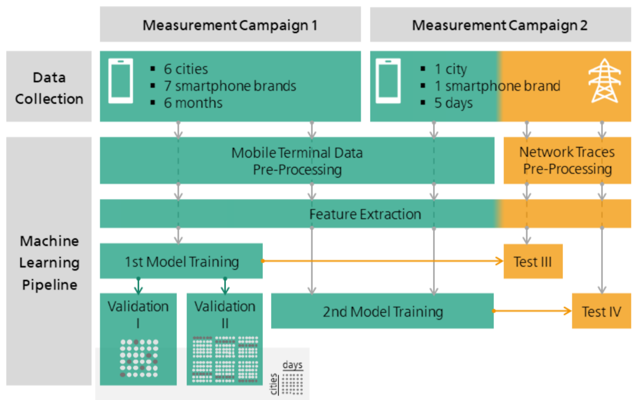


FIGURE 1. Overview of the proposed methodology for the classification of indoor and outdoor environments.

campaigns. Firstly, a large scale measurement campaign (1) using common smartphones was performed through 6 cities with different characteristics and over the course of 6 months. This measurement campaign main purpose was to collect a large and diverse dataset to train machine learning models. The collected data from this campaign includes only data acquired from the MT side. On the other hand, a second measurement campaign with the aim of testing the trained models with NT was planned with a MNO to map the signaling for each UE profile in the network traces. Thus, this campaign includes both MT and NT data. The data pre-processing step consists of parsing raw data, filtering and data segmentation tasks. The feature extraction step includes the extraction of features that can describe the environment, namely variations of serving Cell ID and RSRP. For the model training, a traditional machine learning approach and a deep learning approach were tested. Finally, four validation strategies were used to evaluate the performance of the trained models. In the following sections, a detailed description of all steps is presented.

A. DATA COLLECTION

1) DATA DESCRIPTION

When connected to a BS, an end-user is constantly exchanging data with the network, in order to inform about its radio conditions. All these data can be recorded and logged recursively by the mobile operator, by setting a command on the Operations Support System (OSS). These data are called Network Traces, and they are recordings of protocol events resulting from all the exchanged signaling between the mobile users and the network, making them a powerful source of analyze, monitor and optimize the network performance. Regarding the LTE network, the Cell Trace is a file in the original format (.bin) collected directly from the mobile operator OSS, *i.e.*, from the live network. It contains all the communications exchanged between the end-users and the network on the interface Uu in 15-minute Recording Periods (ROPs). All Cell Trace events contain common types of information, like the event name and identification, the timestamp of its recording, user identification in the specific area, as well as serving cell identification. However, they all also contain specific information that characterizes their function. But, the user geography location is unknown in these data, turning this work useful for operators. Thus, using these always available data, the mobile operator can classify the user environment, without its real position given by GPS and any additional modifications. An useful Cell Trace event that should be considered is the Radio Resource Control (RRC) MEASUREMENT REPORT. This event and its processing allows the retrieval of parameters like RSRP and RSRQ, key measures of signal level and quality for modern LTE networks, Cell ID and Physical Cell Identity (PCI).

On the Mobile Terminal side, data was obtained using an Android in-house application that uses the Telephony Manager to access and record the cellular data. The available fields change from different manufacturers given that each one has a slightly different Android version, which in turn reports and allows access to different cell data fields. All manufacturers allow the access to important parameters, such as RSRP, RSRQ and PCI. The Cell ID was available to the registered BS but not for the neighbouring BSs.

In summary, the cellular network datasets consist in the following fields:

- **Time**: timestamp of the measurement set.
- **RSRP**: the average power of Resource Elements (REs) that carry cell specific Reference Signals over the entire bandwidth. Reporting range: -140 dBm to -44 dBm [24].
- **RSRQ**: the ratio between RSRP and Received Signal Strength Indicator (RSSI) measured over the same bandwidth. It indicates the quality of the received reference signal. Reporting range: -19.5 dB to -3 dB [24].
- **PCI**: the identifier of a cell in the physical layer of the LTE network. The number of PCIs are limited to 504. Range: 0 to 503.
- **Cell ID**: a generally unique number used to identify each LTE cell inside the operator network.

2) MEASUREMENT CAMPAIGN 1

The first dataset was collected using seven smartphones from different manufacturers (Samsung, Google, Huawei, Oneplus, Asus, Motorola), with different Android versions (Android 7-10), from stock to custom Read Only Memorys (ROMs) in order to assure a more diverse and representative dataset. An Android ROM is the system image which can be installed into a smartphone. Depending on the manufacturer and Android version, new sensor data is received every 2 to 10 seconds. Due to restrictions imposed by Apple, it is not possible to use a smartphone with iOS to collect cellular data, since the necessary Application Programming Interfaces (APIs) are not publicly available.

Each batch of data (aggregation of several sensor data) was collected over a minimum period of 5 minutes where the user could be stationary, walking, running, riding a bike or driving. In the indoor scenarios, the user was restricted to the building itself, and if the acquired environment was labeled light indoor, the user had to remain close to the window. If the label was just indoor, the user should not be too close to windows. The buildings ranged from apartments to office buildings, assuring that there were no indoor BSs present. As for the outdoor samples, most are a representation of everyday life scenarios such as walking through a city/park, riding a bicycle, walking a dog, and driving. Some samples also covered the scenario where the user was outside, near tall buildings. A small number of samples were collected with a transition between environments. These had a minimum of 3 minutes, either indoor or outdoor, followed by a minimum of 3 minutes on the opposite environment.

The dataset has a total of 31 hours, 17 minutes and 43 seconds, over 6 different cities of different types, urban, sub-urban and rural, and was collected over the course of 6 months. From this, approximately 18 hours are indoor, of which 2.65 hours are next to windows. The remaining 15.5 hours are outdoor, of which 4.85 hours are on high speed roads, such as highways.

In Figure 2, the representation of the entire dataset distribution divided by each type of environment is shown. Highways category represent data that was collected in highways or high speed roads, covering several different locations and cities. Indoor represents the samples collected indoor far from windows and the Light Indoor environment represent samples collected near windows or balconies. Outdoor represents all samples collected outdoor, from open spaces to samples collected near taller buildings.

Figure 3 shows the representation of the dataset divided by cities and then by each type of environment. In this figure, the samples of high speed roads are not represented since they cover multiple cities.

3) MEASUREMENT CAMPAIGN 2

The second measurement campaign resulted in two additional datasets, one from the MT and the other from the NT.

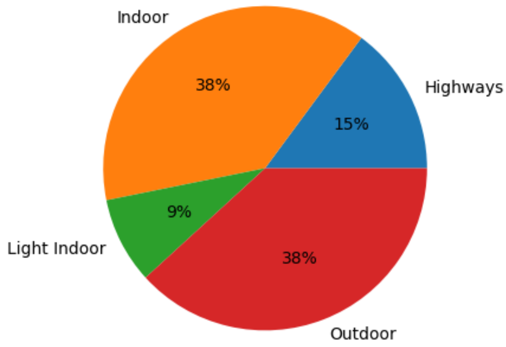


FIGURE 2. **Environment type distribution of Mobile Terminal dataset from measurement campaign 1.**

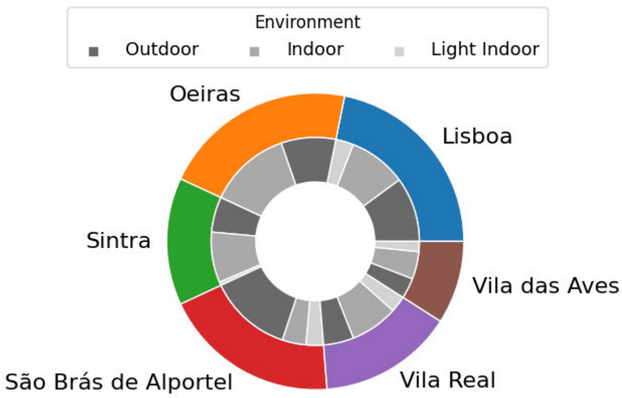


FIGURE 3. **Smartphone dataset distribution. The outer chart represents the distribution over the main six cities of the dataset and the inner chart represents the environment distribution (indoor, light indoor and outdoor) for each independent city.**

FIGURE 4. **Urban environment of measurement campaign 2 characterized by an uniform terrain morphology and city blocks.**

FIGURE 5. **Urban environment of measurement campaign 2 characterized by an irregular terrain morphology and city blocks poorly organized.**

The MT dataset was obtained using a Samsung smartphone with Android 9. This device collected data in five different days in the Lisboa city, for outdoor environments.

The measuring campaign included four different tests, where the user could be stationary, walking, driving in a urban and highway environments. For each mobility profile, several routes were covered in the measuring campaign. These different routes were chosen with the goal of reaching as many radio frequency scenarios as possible, such as Line-of-Sight (LoS) Non-Line-of-Sight (NLoS) with high buildings and narrow roads, street canyons, open areas environments like green parks and gardens, and a university campus, with a mix of building heights and construction types, and gardens. One of the urban areas could be characterized by a terrain morphology quite uniform and city blocks well defined, presented in Figure 4, instead the other, with an irregular terrain morphology and city blocks poorly organized, shown in Figure 5.

This measurement campaign was aligned with the MNO in order to record the Network Traces of a specific BSs cluster, corresponding to the same area of the MT measuring campaign. This process allowed to map the signaling obtained from the measurement campaign for each User Equipment (UE) profile in the network traces, obtaining the NT dataset.

Additionally, a set of different mobile services was periodically run for 20 seconds in each profile/environment. These mobile services included voice, web-browsing, file transfer, and some applications involving data packets exchange. This measurement campaign resulted in the acquisition of approximately 16 hours and 30 minutes of network signalling information from the MT and 1 hour and 33 minutes from NT.

B. DATA PRE-PROCESSING

1) NETWORK TRACES

From the network perspective, as mentioned before, a Cell Trace is an LTE raw file with a period of 15 minutes containing all the signalling exchanged between the mobile users and the network, including the specific mobile terminal used on measuring campaigns.

In order to identify the MTs used in the measurement campaigns, the Traces data is filtered by the Temporary Mobile Subscriber Identity (TMSI) of the measuring terminal, ensuring consistency in both data sources. The TMSI is used instead of International Mobile Subscriber Identity (IMSI), which is a unique user identifier on the network, to protect subscriber from being identified and also ensure more security against radio interface hackers. The TMSI can be intercepted right at the first RRC CONNECTION REQUEST message and is used to link the next protocol messages, creating the concept of user session. Then, the RRC MEASUREMENT REPORT message used by the proposed approach, can be filtered from Network Traces for the same measuring terminal.

2) MOBILE TERMINAL

Regarding the MT, the cellular data is received in call batches. Each batch has the registered BS information, followed by, if existent, neighbour BS information, as well as the batch's timestamp. From time to time, a lost batch can also be received, which comes out of order. Since there is no way to know the correct timestamp of these lost batches, they are filtered out from the collected samples.

In a single batch, multiple BS technologies, LTE (4G), Wideband Code-Division Multiple Access (WCDMA) (3G) and GSM (2G) can be present. All samples were collected using the default option of choosing the strongest available signal, which translated to being connected to LTE BSs most of the time. Additionally, and since the Traces data was limited to LTE technology, the cellular data on the smartphone was also filtered to ignore WCDMA and GSM data, including samples where a connection to a LTE BS was not possible. This last step was possible since the only situations in which no LTE connection was available occurred while in deep indoor, usually in below ground environments.

C. FEATURE EXTRACTION

For the feature extraction process, a set of features that can describe the character of the environment was extracted. These features include variations of the serving Cell ID and RSRP for each time window. A description of each feature can be seen in Table 2.

TABLE 2. List of features extracted from serving cell ID and RSRP for each time window.

USC	Unique Serving Cells
CC	Cell Changes
ACRT	Average Cell Residence Time
MAX	RSRP Maximum value
MIN	RSRP Minimum value
RANGE	Difference between the maximum and minimum RSRP values
AVG	RSRP Average value
VAR	RSRP Variance value
STD	RSRP Standard Deviation
MED	RSRP Median value
RBrS	RSRP Ratio Beyond r Sigma *(r=1)*
ABS-E	RSRP Absolute Energy given by the sum over the squared values
CID	RSRP Complexity-Invariant Distance
SKEW	RSRP Skewness calculated with the adjusted Fisher-Pearson standardised moment coefficient G1
KURT	RSRP Kurtosis calculated with the adjusted Fisher-Pearson standardised moment coefficient G1
LSBM	RSRP Longest Strike Below Mean
LSAM	RSRP Longest Strike Above Mean
PRVV	RSRP Percentage of Reoccurring Values to all Values
PEAKS	Number of RSRP peaks of at least support n in the window *(n=5)*

D. TRAINING AND TESTING

Due to the nature of NT, it is not always easy to gather data, more specifically labeled data. It requires planning and access to specific equipment and software which might not always be available. To circumvent these limitations, some samples of cellular data were collected with a smartphone and compared to the available NT data of the same time and location. It was verified that the available information from the smartphone was a subset of the NT. Thus, by using the smartphone, it is possible to gather a dataset which can be used to train a classifier that can later be applied to NT data as well. Due to the easy process of data collection using smartphones and the limitations of acquiring NT data, this strategy was employed for models training and testing. Thus, the training set is composed by MT data acquired over 6 cities in Portugal, ranging from urban to rural environments. As regards testing set, a set from MT data and an independent set composed by NT from measurement campaign 2 acquired on two different locations of the same city were used.

Regarding the model selection, a traditional machine learning approach and a deep learning approach were selected to classify an environment as indoor or outdoor. For the traditional machine learning approach, a range of algorithms (Naive Bayes, Decision Trees, RF, Adaboost, k-NN, SVM) were investigated and due to the superior performance of RF classifier in this particular classification task, RF was selected. For the deep learning approach, a LSTM was selected due to be capable of extracting the temporal dependencies of the network data patterns and learning to discriminate labels. A detailed description of models construction is presented in section IV.

E. PERFORMANCE EVALUATION

As in Figure 1, four different validation strategies were used for models' performance evaluation. The validation strategy (I) and (II) used only MT data and strategy (III) and (IV) used the independent dataset of NT for performance validation. The first validation strategy (I) was performed with a mix of all cities in both train and test set. For the second validation strategy (II) and given that 6 cities were available, it was decided to use the leave on out approach, in this case with cities, to train and validate the classifier. The training was done using the remaining 5 cities and the testing with an entirely new city. This proves a more robust and faithful approach than most of the literature mentioned in Section II, since it is common to train and test with data from a single location/city. Our approach can also show potential difficulties on passing from an urban environment to a rural one, where the data collected indoor in an urban location can be similar to the data collected outdoor in a rural location. For example, the RSRP and RSRQ might be lower, in a rural environment, since there are fewer BSs and thus its easier to be further away from one. The amount of handovers, i.e. when the smartphone changes the registered BS, can also decrease and thus potentially raise the overlap between urban indoor and rural outdoor.

Regarding validation strategies (III) and (IV), their test set was composed by NT data of measurement campaign (2) collected in two urban areas with different characteristics (see Section III-A for more details). However, due to the above mentioned limitations of acquiring NT data, plus the imposed governmental restrictions due to the covid-19 pandemic status at the time of data collections, the available data is only in an outdoor environment. For the validation strategy (III), the training set was composed by all data from the measurement campaign (1). The main purpose of this validation strategy was to evaluate the performance of the models using an independent test set with a different data source, i.e. network traces. Then, the aim of validation strategy (IV) was to re-train the models using a training set with additional data from the same locations of the NT test set, i.e., all MT data from both measurement campaigns was used for the training set. Although the added data was from the same locations, it was guaranteed that data from both sets set was from different days.

IV. MODELS CONSTRUCTION

In this work, two different approaches to distinguish indoor from outdoor environment were tested: 1) Traditional machine learning approach using a Random Forest classifier; and 2) Deep Learning approach using a LSTM model.

A. TRADITIONAL MACHINE LEARNING APPROACH

A time based sliding-window segmentation method was used to identify the signal changes over a short period of time. Different overlapping window sizes were used to capture different data granularity, namely 15, 30, 60, 90 and 120 seconds. Higher values usually lead to better differentiation between the different activities due to the more available information. However, longer windows increase the estimation latency, and extra-large windows can worsen the performance as they may span different scenarios within the same window [15]. Hence, all features were calculated within several overlapping sliding windows with a fixed size. Although a maximum of 120 seconds window size is used, due to the overlapping strategy, this process results in a classification every 15 seconds, except on the first 120 seconds of data collection.

The reason for the minimum duration of the windows, 15 seconds, is that in a worst case scenario, the smartphone only updates the cell data every 10 seconds, thus it is not possible to use short time windows as was used in the literature as in [19].

Although most of the acquisitions were performed in an indoor or outdoor environment, there were few acquisitions where a transition between environments occurred during the data collection. In this situation, if a transition occurs within a 15 second window, the label will correspond to the longest duration within the window, i.e if 10 seconds were indoor, then the label will be indoor. In the case that it is exactly 7.5 to 7.5 seconds, the label will be of the last type of environment.

For the feature extraction, the set of features described in Section III-C was extracted, which are the variations of the serving Cell ID and RSRP for each time window.

In the training phase, a range of traditional machine learning algorithms were investigated to distinguish between indoor and outdoor environments. Due to the superior performance of RF classifier in this particular classification task, RF was employed using a bootstrap approach.

As the feature extraction process results in a total of number of 95 features (5 window sizes times 19 features), a feature selection algorithm was used to improve the accuracy and computational performance of the algorithm. The Sequential Forward Feature Selection [25] was employed using a weighted F1-Score as performance metric and a group k-fold cross-validator as validation method. Depending of the used approach, the groups can represent a specific day, or an entire city. Moreover, the RF parameters were optimized using a cross-validated grid-search.

B. DEEP LEARNING APPROACH

This approach uses the data as a temporal series for classification. The LSTM model was chosen given its ability to model and deal with the long-term temporal dependencies of the input sequences. Unlike traditional Recurrent Neural Networks (RNNs), LSTM were designed to better deal with long-short term memory and to overcome the vanishing/exploding gradient problem, where the model may stop learning early in the process if the error's gradient value during backpropagation is too small or to big [26].

LSTM's, instead of using nodes, are composed of special ''memory blocks'' containing three key gate units: forget gate, input gate and output gate. These gates essentially control the amount of information that the network should keep and forget from the original input sequence. Concretely, the forget gate decides which information should be thrown away from the cell state; the input gate decides which new information is added to the cell state; the output gate decides how much the internal state should be passed to the next step. By using this gating mechanism, LSTMs can explicitly model long-term dependencies, making them attractive for a variety of time-related problems [27].

Similarly to the previous approach, due to the different sampling frequency of cellular network data between

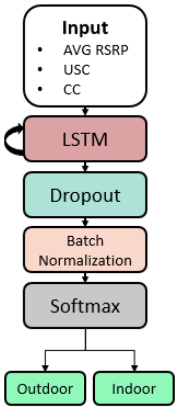


FIGURE 6. LSTM network architecture for the classification of indoor and outdoor environment.

smartphones, features were calculated within 15 seconds windows. Within the features, listed in Table 2, the input features of LSTM were empirically chosen. The selected features were RSRP average value, unique serving cells and cell changes, being the total number of features (N) equals to 3.

Each time series was divided and grouped with a fixed window length (W) of 8, which represents a total of 120 seconds. This W value is the number of timesteps used as input in the LSTM architecture. A total number of 5297 sequences were obtained for entire dataset using a fixed window (W) and one timestep shift (15 seconds).

Thus, the LSTM input array is three-dimensional array given by $B \times W \times N$ (batch size, timesteps, features).

The neural network architecture is shown in Figure 6. In the first layer, LSTM receives a three-dimensional array input. In each recurrent units, sigmoid activation function is used. LSTM layer returns hidden state of the last time step, additionally in order to reduce overfitting, the weight constraint is considered to force the weights to have a magnitude equal to or less than a certain limit, in this case the maximum norm type constraint was considered. The next layer is a dropout layer. This layer is based on random neurons deactivation during the training process of a neural network, which aim to minimize the overfitting of the model on the training data [28]. The third layer is a batch normalization layer which standardizes layer's inputs, to have zero mean and unit variance, for each mini-batch, allowing the stabilization of the learning process and reducing the required number of training epochs. Finally, the output layer is the softmax activation function, which generates the probability distribution for each output class.

The hyperparameters were tuned using a 5-fold cross-validated random grid-search. This method provides multiple hyperparameters for the neural network, in which their combinations are chosen randomly and applied in cross validation, in order to find the best model. Table 3 shows the final range of hyperparameters.

To analyze the impact of different hyperparameters on the LSTM performance, an illustration of F1-Score versus the number of epochs for each hyperparameter is shown in Figure 7. Regarding batch size, three candidates from the geometric progression of 2 were tested, namely 128, 256 and 512. Figure 7a shows the behavior with different batch sizes, where the performance improves with the decrease of batch sizes. In the case of dropout rate (Figure 7b) and weight constraint (Figure 7e), all tested values show a similar performance with the increase of epochs. Figure 7c shows the influence of different learning rates on the performance, where it is possible to see a significant increase in the performance with the increase of learning rate. Finally, regarding the number of neurons, the more complex the network is, the higher the obtained performance (see Figure 7d).

TABLE 3. Possible values for each neural network hyperparameter.

Hyperparameters	Range Values
batch_size	[128,256,512]
dropout_rate	[0.2,0.4,0.5]
epochs	[5,10,20,30,50,100,150,200,250,300]
learning_rate	[0.0001,0.001,0.01]
neurons	[10,30,50]
weight_constraint	[1,3,4,5]

V. RESULTS AND DISCUSSION

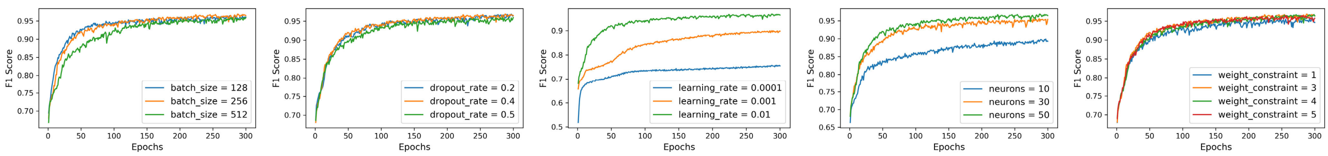
In this Section a detailed description of the experimental results and their discussion is presented.

A. PRELIMINARY STUDY

Before the models' performance evaluation, a preliminary study comprising an analysis of (1) cell data differences between different devices, (2) cell data differences between data acquired from the mobile terminal and network traces, and (3) cellular network data statistics was conducted.

In order to test potential differences between smartphones, two collections were conducted with 5 different smartphones, namely a Google Pixel XL, a OnePlus 6T, a Samsung A9, a Samsung S10e and a Huawei Mate 20 Pro. The first data collection was indoor, with all smartphones on the same location and under the same conditions, and the second was outdoor, in a similar manner as the first. Regarding the signal properties, the maximum obtained difference between smartphones was of 2 dB, which is small considering potential signal fluctuations that can occur. Regarding the number of BS seen, the obtained results were consistent on 4 out of 5 smartphones. While the Samsung S10e was able to detect, on average 2 to 3 BS, both indoor and outdoor, the other 4 smartphones had an average of 10 BS while indoor and 11 when outdoor.

The second study was conducted at a shopping center, in Lisbon, to validate the measurements of the developed mobile application and compare them to the data from the NT. The measurement campaign lasted several hours, and it ranged across the entire building with multiple floors. The study aimed at matching the exact timestamps on both data sources and confirming that, for the same timestamp, both sources agree on the registered BS, Cell ID, RSRP, RSRQ



(a) Batch size. (b) Dropout rate. (c) Learning rate. (d) Number of neurons. (e) Weight constraint.

FIGURE 7. LSTM performance variation versus number of epochs with different hyperparameters.

TABLE 4. Statistical analysis of network information for each analysed city.

City	Type	Average RSRP / dBm		Average # detected BSs		Average # Handovers / min.		Macro BSs Density	Average LTE Bands
		Indoor	Outdoor	Indoor	Outdoor	Indoor	Outdoor	# / km^2	per BS
Lisbon	Urban	-95.01	-85.41	7.19	10.76	0.05	0.82	4.95	2.50
Oeiras	Urban	-95.79	-94.14	4.83	6.26	0.78	0.75	2.71	1.79
Sintra	Sub-urban	-88.78	-86.22	3.16	5.71	0.10	2.15	1.05	1.18
Vila Real	Sub-urban/Rural	-100.44	-91.70	11.04	11.46	0.38	0.70	1.50	2.22
São Brás Alportel	Sub-urban/Rural	-107.64	-85.20	2.49	4.29	0.01	0.2	0.63	2.00
Vila das Aves	Rural	-111.59	-104.94	2.77	2.97	0.20	0.61	0.28	2.00

and PCI. The results of this study were successful and it was possible to verify that the data was the same across both data sources.

Finally, in Table 4, it is enumerated the different characteristics observed for each city, the average RSRP, the average number of Base Stations seen, the average number of handovers per minute, the macro BSs density and average number of LTE bands per BS. These statistics are divided into indoor and outdoor environment for each city. A handover is the switching of the BS to which the smartphone is currently registered, i.e., the smartphone changes connections from BS 1 to BS 2. In this table cities are sorted according to the type of environment. Lisboa and Oeiras are categorized as urban, Sintra as sub-urban, Vila Real and São Brás de Alportel as sub-urban/rural and Vila das Aves as rural. Note that Vila Real and São Brás de Alportel are categorized as a mixed of sub-urban and rural environment since data was collected in both the city center (sub-urban environment) and in more isolated locations, similar to a rural environment. It is possible to see a slight decrease regarding the average RSRP as we go to more rural locations, especially when indoor. The average number of BS also decreases, with Vila Real being the exception. Regarding the number of handovers, it appears that there is no direct connection with the type of environment. The abnormally high number in Sintra is explained with the fact that a more significant amount of outdoor data was collected while driving, which naturally increases the number of handovers since the user was covering greater distances in a smaller amount of time.

B. EXPERIMENTAL RESULTS

To evaluate the models' performance in distinguishing indoor and outdoor environments, different combinations of train and test sets were used, namely:

TABLE 5. Size of train and test set for each validation strategy.

Approach		Train Time / h	Train # Instances	Test Time / h	Test # Instances
I		16.42	3941	9.02	2163
	Lisbon	20.52	4922	4.90	1179
	Oeiras	20.63	4951	4.78	1150
II	Sintra	21.68	5203	3.73	898
	Vila Real	20.98	5036	4.43	1065
	São Brás Alportel	20.28	4951	5.15	1236
	Vila das Aves	23.03	5528	2.38	573
III		25.43	6104	1.93	332
IV		32.32	7446	1.93	332

I. Train and test set from the first measurement campaign using a random group k-fold approach, where each group represents a given acquisition day;
II. Train and test set from the first measurement campaign using a leave one out approach, where each city is left out of the training phase and used exclusively for testing;
III. Train set composed by all data from the first measurement campaign and test set composed of the Traces dataset;
IV. Train set composed by all MT data from both measurement campaigns and a test set composed by the Traces dataset;

The number of instances of both train and test set after pre-processing for each validation strategy can be consulted in Table 5.

In order to compare the implemented algorithms with the literature, the first approach consists of a training and testing

TABLE 6. Approach II performance measures with Random Forest and LSTM classifiers. Each city name represents the validation set using a leave one city out strategy. The F1-Score was calculated using a weighted averaging to account for label imbalance.

City	Type	Accuracy		Precision		Recall		F1-Score	
		RF	LSTM	RF	LSTM	RF	LSTM	RF	LSTM
Lisbon	Urban	0.94	0.86	0.97	0.91	0.95	0.86	0.94	0.86
Oeiras	Urban	0.92	0.76	0.90	0.92	0.98	0.76	0.92	0.77
Sintra	Sub-urban	0.92	0.80	0.91	0.82	0.96	0.80	0.92	0.80
Vila Real	Sub-urban/Rural	0.84	0.70	0.83	0.67	0.95	0.70	0.85	0.66
São Brás Alportel	Sub-urban/Rural	0.96	0.92	0.99	0.82	0.92	0.92	0.96	0.92
Vila das Aves	Rural	0.76	0.73	0.81	0.72	0.84	0.73	0.76	0.63
Average		0.90	0.80	0.90	0.81	0.94	0.80	0.90	0.80

sets with a mix of all cities was performed. To avoid having sequential data collections separated, *i.e* one in the training dataset and the other on the testing dataset which could lead to overfitting, it was ensured that a day of data collections was, in its entirety, either in the training dataset or the testing dataset. All cities had several days of data collections with both types of labels ensuring a diverse training and testing dataset. After applying the group k-fold split, where each group represents an acquisition day, the training dataset had 66% of the total instances and the testing dataset had the remaining 34%. The obtained weighted F1-Score using the RF algorithm was 0.95, which is aligned with the best results in the literature, even considering the higher diversity in our dataset. Regarding the LSTM approach, the obtained results were lower than RF, achieving a value of 0.75.

In the second approach, a leave one city out approach was employed to show the differences of using an entirely new location for the testing set. For this, one city was left out of the training phase and used exclusively for testing phase. The process is repeated for all the remaining cities and the results are shown separately for each city. Due to the imbalance of data within each city, the accuracy, precision, recall and weighted F1-Score were chosen as evaluation metrics. The obtained results for the RF and LSTM can be seen in Table 6.

From the Table 6, it is possible to observe that for the urban and sub-urban environments, the results from both models are overall higher, ranging from a F1-Score of 0.92 to 0.96 using the RF model and ranging from 0.92 to 0.77 when using the LSTM model. This is somewhat expected, firstly since four out of the six cities are either urban or sub-urban. Secondly, in an urban environment, there is a higher density of BSs, which means that not only is the smartphone able to switch from BS to BS more frequently, but also the received signal will be stronger, on average, while outdoor, since the BSs usually have a shorter serving range.

Regarding the rural cities, the results were lower, which was also expected since a rural environment is inevitably different from an urban one. The exception being São Brás Alportel. A possible reason is the differences in the terrain morphology where data was recorded. Vila das Aves and Vila Real are located in the north of Portugal, both these cities have an irregular morphology terrain, which, associated with buildings can increase the signals' attenuation. Regarding the city center of São Brás Alportel, where the majority of outdoor data was collected, the terrain morphology is flatter than both Vila Real and Vila das Aves, which mitigates this problem. Nonetheless, for the RF only in one city, Vila das Aves, the results (F1-Score of 0.76) were significantly lower which could potentially be mitigated by collecting more data. Regarding the lower results of the LSTM, observed in rural and sub-urban cities, it can be justified by having a strong temporal dependence, and in these types of environments for having less variability in the signal level, the model cannot effectively distinguish the type of environment.

The average results of all cities resulted in a weighted F1-Score of 0.90 and 0.80 for RF and LSTM, respectively. RF model obtained a lower performance in approach II when compared with the approach I. These results prove that using a test set acquired in the same city/location of the training set can produce an overfitted result. The performance decrease from approach I to II can be explained by the ability of the RF to learn some characteristics that are unique from each city and/or type of environment. Regarding LSTM the behavior between approaches is the opposite of the RF. The average weighted F1-Score of approach II was 0.80 compared with the 0.75 of approach I. One possible explanation for these results can be the differences between the training size of both approaches. In approach I, the training size was composed by approximately 16.5 hours, and in approach II the training size ranged from 20 hours to 23 hours depending on the city being evaluated.

In the approaches III and IV, a different data source, NT, was used for the test set. As previously explained in Section III-D, only data from the outdoor environment was available. For the approach III, the training set contained the first measurement campaign data, the same which was used for both scenarios I and II. Regarding the approach IV, a second measurement campaign was made in the city of Lisboa, part of which in the same location as the test dataset, containing the NT, but in different days.

TABLE 7. Results summary of all tested approaches with Random Forest and LSTM. The F1-Score was calculated using a weighted averaging to account for label imbalance.

		# Measurement Campaign (Data Source)		F1-Score	
	Splitting Criteria	Train Data	Test Data	RF	LSTM
I	groups k-fold	1 (MT)	1 (MT)	0.95	0.75
II	leave one city out	1 (MT)	1 (MT)	0.90	0.80
III	independent	1 (MT)	2 (NT)	0.75	0.80
IV	independent	1, 2 (MT)	2 (NT)	0.79	0.91

Table 7 summarizes the results of all approaches and both models (RF and LSTM) using as metric the weighted F1-Score.

Regarding the RF, the obtained F1-Score of approaches III, IV were lower than expected when compared with approaches I and II. After investigating the reason of this performance decrease, it was concluded that most of the mislabeled windows happened in a single location, between tall buildings, of the measurement campaign represented in Figure 5. This specific scenario represents an environment that can be categorized as semi-outdoor, which is an already difficult situation to classify. In the case of the LSTM, this challenging situation seems to not affect the performance of the model, achieving even higher results than in the previous approaches. This performance increase can also be explained by the increase in the training set. In these approaches the training size was 25.5 and 32.3 hours for approach III and IV, respectively.

Note the increase in the results from scenario III to IV, which was expected, and consistent with the conclusion that by gathering more data from a new location, it will improve the results of the classifiers in that location. This increase was more significant in the LSTM approach, obtaining a weighted F1-Score of 0.91.

These results could greatly benefit from more data collection, specially indoor, but, as previously mentioned in Section III-D, that was not possible. With more data, both for the training and testing datasets, a better representation of the models reliability can be achieved.

Regarding the comparison of the obtained results with similar studies in the literature, the first validation strategy used to train and test the algorithms is the most comparable to the validation strategies of existing works. The RF achieved similar results (F1-Score of 0.95) to the presented works (see Table 1), while the LSTM underperformed (F1-Score of 0.75). These results were achieved without the limitations previously presented in section I with the exception of the solution proposed in [23] which is the most similar to the one presented in this paper.

Additionally, the proposed solution also goes beyond the work presented in [23] by being tested and validated under stricter conditions, such as being tested with a completely new location. The RF performance was slightly lower with a 5% decrease on the average results with the leave one city out approach and the LSTM with a 0.8 F1-Score. The two additional validation strategies are also a novelty given that the testing dataset was obtained in real time by an MNO. The RF results were lower (0.75 and 0.79) while the LSTM obtained good results, especially in the IV strategy (F1-Score of 0.91). These two last strategies are a closer representation to the deployment performance.

C. VISUALIZATION

Due to the large geographic area of both measurement campaigns, it is not possible to visualize the model's performance using the entire dataset. Therefore, for visualization purposes, a dedicated acquisition in an area of approximately 0.496 km^2 was performed. This acquisition includes indoor data in seven different buildings (cafes, pharmacy, private apartments, offices and university buildings) and outdoor data acquired while the user was stationary, walking, running and driving a car. For indoor environment, acquisitions close to windows were also considered and for outdoor environment, both open areas and scenarios near tall buildings were covered.

A representation of both models' predictions overlaid with the map is shown in Figure 8. The indoor and outdoor predictions are represented with orange and blue circles, respectively. Regarding the ground truth, buildings were colored blue and the outdoor environment orange. For the geolocalization of data points, GPS coordinates and cell data were recorded simultaneously using the in-house Android application previously mentioned in section III-A. Due to measurement errors caused by the limitations of GPS positioning, the recorded coordinates had to be corrected to correctly describe the user's positions in the map. Different techniques for map matching using GPS can be employed to improve the visualization process, such as the work of H. Cheng et al [29]. However, for this visualization process, the recorded coordinates inside buildings and a few outdoor coordinates with low positioning precision were manually corrected using checkpoints manually annotated during data acquisition. This visualization limitation will be addressed in future research to improve the visualization capabilities of the proposed work.

Regarding the analysis of Figure 8 two areas are highlighted to show the misclassifications of both models. The areas identified by number 1 and 2, represent an office building and a university campus with three different buildings, respectively. In the office building it is possible to see that RF classifies five indoor labels as outdoor. These misclassifications correspond to the transition between environments (E) and to the office's open space, which have several windows surrounding (W). In the case of LSTM model, only one misclassification in the transition between environments occurred (E), where an outdoor label was classified as indoor. In the area 2, both models misclassified the location C, which is a connection made of glass between the two buildings. Moreover, there is some confusion in the transition between environments (E) for both models. The RF also misclassified

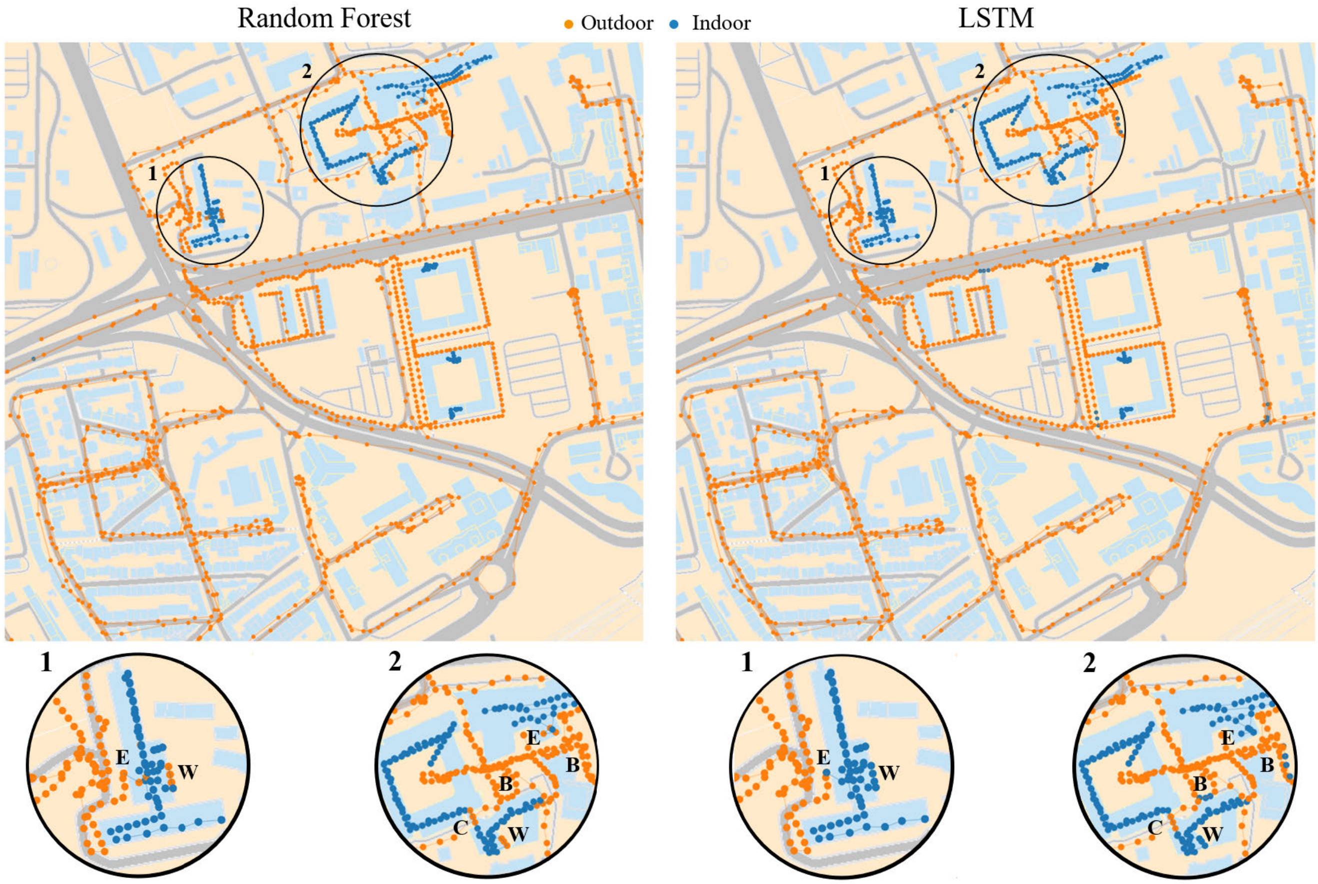


FIGURE 8. Representation of geo-positioned Random Forest and LSTM predictions. Indoor and outdoor predictions are colored in blue and orange, respectively. Buildings and outdoor environment in the map are also colored in blue and orange, respectively. Area 1 represents an office building and area 2 an university campus. E: Entrance, W: Window, C: Connection between buildings, B: Outdoor near building.

location W (close to a big window) and location E (building entrance), and the LSTM misclassified locations B (near tall buildings). Besides that, there are only a few outdoor labels wrongly classified as indoor. It is also worth mentioning that there is no misclassification in the map's lower left corner. This area represents an urban environment characterized by an irregular terrain morphology and tight roads surrounded by tall buildings.

VI. CONCLUSION

In this paper, a solution for solving the classification problem of whether a user's smartphone is being used indoor or outdoor is proposed, relying solely on cellular network data. In order to tackle the existent difficulty in gathering labeled data, a mixed approach was developed where a low cost mobile app can be used to gather data around a city, which was proved to be a subset of the available data on the network side. From this data collection, a dataset was created with measurements from six different cities, differing in both size and density. A second dataset was also created using network traces to represent the network data and test the scalability of smartphone data. An important difference between the first dataset and the ones analyzed from the literature, is the diversity of data that ranged from urban to rural environments, which can be a better representation of a real world scenario.

In the sequence, two different algorithms were trained, a RF and a LSTM. The first algorithm, representing a traditional machine learning approach, obtained very good results, using the mobile terminal as a data source, having a weighted F1-Score, 0.92-0.96 for urban environments, but showing a slight difficulty in adapting to a more rural location, dropping the weighted F1-Score to 76%. When testing with network traces, the F1-Score decreased to a maximum of 0.79, due to the high number of samples in a semi-outdoor environment. This could potentially be mitigated by doing a small data collection with smartphones to further improve the robustness of the algorithm.

Regarding the DL approach, the training set size seems to have a huge impact on the performance of the algorithm. Comparing with the RF, the results from different cities ranged from 0.77 to 0.92 in urban cities, and 0.63 to 0.66 in rural cities. However, when the network traces were used as testing set, with a training size of approximately 32.3 hours, the algorithm's performance increased to a value of 0.91.

The obtained results evidence the impact on models' performance when using different validation strategies. In this work, it is shown that a common validation strategy can overestimate the models' performance, and the importance of using a test set from a new and unseen location from different environment types.

For future work, additional data collections would be engaged, not only to increase the robustness of the algorithms but also to give a better view of the feasibility of deploying such a system. Additionally, the improvement of visualization capabilities using map matching techniques will be addressed in future research.

REFERENCES

[1] Ericsson. (Jun. 2021). *Ericsson Mobily Report*. [Online]. Available: https://www.ericsson.com/en/mobility-report

[2] A. E. Kelishomi, A. H. S. Garmabaki, M. Bahaghighat, and J. Dong, "Mobile user indoor-outdoor detection through physical daily activities," *Sensors*, vol. 19, no. 3, p. 511, Jan. 2019.

[3] J. L. Bejarano-Luque, M. Toril, M. Fernandez-Navarro, R. Acedo-Hernandez, and S. Luna-Ramirez, "A data-driven algorithm for indoor/outdoor detection based on connection traces in a LTE network," *IEEE Access*, vol. 7, pp. 65877–65888, 2019.

[4] H. Huang, Q. Zeng, R. Chen, Q. Meng, J. Wang, and S. Zeng, "Seamless navigation methodology optimized for indoor/outdoor detection based on WIFI," in *Proc. Ubiquitous Positioning, Indoor Navigat. Location-Based Services (UPINLBS)*, 2018, pp. 1–7.

[5] H. Ullah, N. Gopalakrishnan Nair, A. Moore, C. Nugent, P. Muschamp, and M. Cuevas, "5G communication: An overview of vehicle-to-everything, drones, and healthcare use-cases," *IEEE Access*, vol. 7, pp. 37251–37268, 2019.

[6] J. Kaur, M. A. Khan, M. Iftikhar, M. Imran, and Q. E. Haq, "Machine learning techniques for 5G and beyond," *IEEE Access*, vol. 9, pp. 23472–23488, 2021.

[7] L. Jarvis, J. McEachen, and H. Loomis, "Geolocation of LTE subscriber stations based on the timing advance ranging parameter," in *Proc. Mil. Commun. Conf.*, 2011, pp. 180–187.

[8] X. Ye, X. Yin, X. Cai, A. Perez Yuste, and H. Xu, "Neural-network-assisted UE localization using radio-channel fingerprints in LTE networks," *IEEE Access*, vol. 5, pp. 12071–12087, 2017.

[9] J. A. Del Peral-Rosado, J. A. Lãpez-Salcedo, F. Zanier, and G. Seco-Granados, "Position accuracy of joint time-delay and channel estimators in LTE networks," *IEEE Access*, vol. 6, pp. 25185–25199, 2018.

[10] R. Borralho, D. Duarte, A. Quddus, and P. Vieira, "Developing a LTE localization framework using real network data towards RAN optimization through context knowledge," in *Proc. 23rd Int. Symp. Wireless Pers. Multimedia Commun. (WPMC)*, Oct. 2020, pp. 1–6.

[11] V. Guimaraes, L. Castro, S. Carneiro, M. Monteiro, T. Rocha, M. Barandas, J. Machado, M. Vasconcelos, H. Gamboa, and D. Elias, "A motion tracking solution for indoor localization using smartphones," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Madrid, Spain, Oct. 2016, pp. 1–8, doi: 10.1109/ipin.2016.7743680.

[12] W. A. Hapsari, A. Umesh, M. Iwamura, M. Tomala, B. Gyula, and B. Sebire, "Minimization of drive tests solution in 3GPP," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 28–36, Jun. 2012.

[13] R. Keating, M. Säily, J. Hulkkonen, and J. Karjalainen, "Overview of positioning in 5G new radio," in *Proc. 16th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Oulu, Finland, 2019, pp. 320–324.

[14] S. Dwivedi, R. Shreevastav, F. Munier, J. Nygren, I. Siomina, Y. Lyazidi, D. Shrestha, G. Lindmark, P. Ernström, E. Stare, S. M. Razavi, S. Muruganathan, G. Masini, Å. Busin, and F. Gunnarsson, "Positioning in 5G networks," 2021, *arXiv:2102.03361*.

[15] R. S. Pissardini and E. S. Fonseca Junior, "Automatic detection of indoor and outdoor scenarios using NMEA message data from GPS receivers," 2018, *arXiv:1804.05907*.

[16] Y. Zhu, H. Luo, Q. Wang, F. Zhao, B. Ning, Q. Ke, and C. Zhang, "A fast indoor/outdoor transition detection algorithm based on machine learning," *Sensors*, vol. 19, no. 4, p. 786, 2019.

[17] I. Ashraf, S. Hur, and Y. Park, "MagIO: Magnetic field strength based Indoor- outdoor detection with a commercial smartphone," *Micromachines*, vol. 9, no. 10, p. 534, Oct. 2018.

[18] Z. Liu, H. Park, Z. Chen, and H. Cho, "An energy-efficient and robust indoor-outdoor detection method based on cell identity map," *Proc. Comput. Sci.*, vol. 56, pp. 189–195, Jan. 2015.

[19] W. Wang, Q. Chang, Q. Li, Z. Shi, and W. Chen, "Indoor-outdoor detection using a smart phone sensor," *Sensors*, vol. 16, no. 10, p. 1563, Sep. 2016.

[20] L. Zhang, Q. Ni, M. Zhai, J. Moreno, and C. Briso, "An ensemble learning scheme for indoor-outdoor classification based on KPIs of LTE network," *IEEE Access*, vol. 7, pp. 63057–63065, 2019.

[21] *TEMS*. Accessed: Aug. 2, 2021. [Online]. Available: https://www.infovista.com/tems

[22] S. Janitza and R. Hornung, "On the overestimation of random forest's out-of-bag error," *PLoS ONE*, vol. 13, no. 8, Aug. 2018, Art. no. e0201904.

[23] I. Saffar, M. L. A. Morel, K. D. Singh, and C. Viho, "Machine learning with partially labeled data for indoor outdoor detection," in *Proc. 16th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Las Vegas, NV, USA, Jan. 2019, pp. 1–8.

[24] *LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer; Measurements*, document ETSI TS 136 214, 3GPP, 2017.

[25] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, 1997.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, pp. 1735–1780, Dec. 1997.

[27] A. Murad and J.-Y. Pyun, "Deep recurrent neural networks for human activity recognition," *Sensors*, vol. 17, no. 11, p. 2556, Nov. 2017.

[28] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdi, "DropOut: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Nov. 2014.

[29] H. Cheng, S. Ma, H. Lee, and M. Cho, "Millimeter wave path loss modeling for 5G communications using deep learning with dilated convolution and attention," *IEEE Access*, vol. 9, pp. 62867–62879, 2021.

PEDRO ALVES received the M.S. degree in electrical engineering and computer science from the Faculdade de Engenharia, Universidade do Porto, in 2018.

In 2018, he started collaborating with the Fraunhofer Center for Assistive Information and Communication Solutions (AICOS), on indoor positioning with a major emphasis on mobile network data. Since then, he has also worked with computer vision in the medical field and embedded testing.

THAÍNA SARAIVA received the bachelor's degree in electrical engineering from the Universidade Federal do Tocantins (UFT), in 2016. She is currently pursuing the M.Sc. degree in electrical and computer engineering with the Instituto Superior de Engenharia de Lisboa (ISEL). Since May 2018 she has been working at CELFINET, in the Research Department, where she has been developing research work in the area of energy saving, artificial intelligence and machine learning.

MARÍLIA BARANDAS received the M.S. degree in biomedical engineering from the Faculdade de Ciências e Tecnologia of NOVA University of Lisbon (FCT NOVA), in 2013, where she is currently pursuing the Ph.D. degree in biomedical engineering doctoral program. She has been working as a Scientist at the Fraunhofer Center for Assistive Information and Communication Solutions (AICOS), since 2015. Prior to joining AICOS, she was an Assistant Lecturer at the Department of Electrical and Computer Engineering, FCT NOVA, and a Research Engineer at the Centre of Technology and Systems, Computational Intelligence Research Group. Her main research interests include knowledge extraction, probability theory, machine learning, and explainable artificial intelligence.

DAVID DUARTE received the M.S. degree in electronic, telecommunications and computers engineering from the Instituto Superior de Engenharia de Lisboa (ISEL), in 2014. In 2014, he started as a Researcher at the Instituto de Telecomunicações (IT), focusing in the areas of planning and optimization of mobile networks. Since 2016, he has been working as a Research Lead at Celfinet company, specifically in energy saving, network performance and fault management, and smart operations researching areas. He has also the responsibility of M.S. theses coordination of new students, ensuring the collaboration between the company and the universities.

DINIS MOREIRA received the M.S. degree in bioengineering from the Faculty of Engineering, University of Porto, in 2015.

In 2016, he started collaborating with the Fraunhofer Center for Assistive Information and Communication Solutions (AICOS). Since then, he has also been working in several projects related to pattern recognition, data mining, and machine learning. His research interests include fall prevention, human movement characterization and activity recognition, based on inertial sensor data.

RICARDO SANTOS received the B.S. and M.S. degrees in biomedical engineering from the Faculdade de Ciências e Tecnologia of NOVA University of Lisbon (FCT NOVA), Portugal, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree in biomedical engineering.

Since 2018, he has been working as a Junior Researcher at the Fraunhofer Center for Assistive Information and Communication Solutions (AICOS), Portugal, until 2020, and currently as a Scientist. His work focuses on the indoor location field, and the development of intelligent systems with AI methods for different areas, such as medical diagnosis and prognosis.

RICARDO LEONARDO received the M.Sc. degree in biomedical engineering from the Faculdade de Ciências e Tecnologia of NOVA University of Lisbon (FCT NOVA), in 2018, where he is Currently pursuing the Ph.D. degree in biomedical engineering.

In 2017, he started collaborating with the Fraunhofer Center for Assistive Information and Communication Solutions (AICOS), focusing his research on indoor positioning solutions based on pervasive data sources, namely inertial data, sound, RF signals, and geomagnetism. He has also worked on applying computer vision and deep learning to retinal images for explainable computer aided diagnosis and image quality, enhancement, and synthesis.

HUGO GAMBOA (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the Instituto Superior Técnico of University of Lisbon (IST UL), in 2007.

He co-founded and is the President of PLUX, a company that develops bio-signals monitoring wearable technology. He is currently a Researcher at the Laboratory for Instrumentation, Biomedical Engineering and Radiation Physics (LIBPhys), Faculdade de Ciências e Tecnologia of NOVA University of Lisbon (FCT NOVA), where he is also an Associate Professor at the Physics Department. Since 2014, he has been a Senior Researcher at the Fraunhofer Center for Assistive Information and Communication Solutions (AICOS). His research interests include bio-signals processing and instrumentation.

PEDRO VIEIRA received the B.Eng., M.S., and Ph.D. degrees in electrical and computer engineering from the Instituto Superior Técnico (IST), Technical University of Lisbon, Portugal, in 1997, 2003, and 2008, respectively. Since 1997, he has been with the Department of Electronics, Telecommunications and Computer Engineering, Lisbon Polytechnic Institute (ISEL). He is currently an Adjunct Professor at ISEL and a Researcher at the Instituto de Telecomunicações (IT), where he is researching aspects of wireless communications, including radio propagation, radio network planning and optimization, and SON systems. He is also a Senior Engineer registered at the Portuguese Engineering Order (OE). Since 2008, he has been an OE Telecommunications Specialist. He is leading the research activities at Celfinet, a Portuguese technological consultancy company, where he is engaging in applied research to create new methodologies, tools and algorithms for the mobile communications business.