IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Hydrogeological Risk Management in Smart Cities: a new Approach to Rainfall Classification based on LTE Cell Selection Parameters

ROBERTA AVANZATO AND FRANCESCO BERITELLI
Department of Electrical, Electronic and Computer Engineering, University of Catania, Italy
Corresponding author: Francesco Beritelli (e-mail: francesco.beritelli@unict.it).

ABSTRACT The sudden climate change, that has taken place in recent years, has generated calamitous phenomena linked to hydrogeological instability in many areas of the world. An accurate estimate of rainfall levels is fundamental in smart city application scenarios: it becomes essential to be able to warn of the imminent occurrence of a calamitous event and reduce the risk to human beings. Unfortunately, to date, traditional techniques for rainfall level estimation present numerous critical issues. This paper proposes a new approach to rainfall classification based on the LTE radio channel parameters adopted for the cell selection mechanism. In particular, this study highlights the correlation between the set of radio channel quality monitoring parameters and the relative rainfall intensity levels. Through a pattern recognition approach based on neural networks with Multi-Layer Perceptron (MLP), the proposed algorithm identifies five classes of rainfall levels with an average accuracy of 96 % and a F1 score of 93.6 %.

INDEX TERMS Rainfall Classification, Smart Sensors, 4G/LTE Technologies, Radio Signal Quality, Neural Networks

I. INTRODUCTION

Recently, several smart environment applications have been introduced, including smart transportation [1], smart healthcare [2], smart homes [3] and smart cities [4], due to the rapid growth of urban populations. Currently, urban performance depends, not only upon the physical infrastructure, but, also, on the availability and quality of knowledge, communication and social infrastructure [1]. The key enabler of these smart city applications is possibly the IoT (Internet of Things), which connects everyday objects and devices to network technologies.

In fact, today the advanced IoT (Internet of Things) sensing technologies cut across many areas of modern research, industry and daily life [5 – 7]. They facilitate detection, transmission and measurement of various environmental indicators.

Smart cities play a key role in transforming different areas of human life, touching upon such sectors as transportation, health, energy, and education. For example, the data regarding weather information are significantly increasing at a rapid pace. Identifying and obtaining valuable information from large amounts of weather data can be extremely beneficial in terms of agricultural development. Moreover, analytics of the weather data can help inform people in advance or alert them about possibly hazardous weather conditions (e.g. floods, extreme heat, droughts, and so on) [8].

For this reason, an accurate estimate of rainfall levels is fundamental in smart city application scenarios.

The massive amount of data collected by low-cost sensors plus the recent data analysis technologies help us greatly improve the modern rainfall classification process.

Big data analytics in cloud computing systems move from IoT to real-time control for smart cities.

The main existing rainfall level measurement methods employ tilt rain gauges, weather radars and satellites.

These traditional estimation techniques present a wide range of problems, for example:

- tilt rain gauges tend to underestimate the amount of rainfall, particularly in snowfall and heavy rainfall events and they are also sensitive to the inclination of the receiver and different types of dirt that may clog the water collection point. Moreover, rain gauges only record local information, measuring the level of precipitation in the specific geographic location where the gauge is installed. Information requests for any other

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011375, IEEE Access

point must be obtained by interpolating the available data provided by nearby rain checks, with the consequence that this information may be influenced by a higher error;
- weather radars have the advantage of being able to monitor a larger area, compared to the rain gauge, and to determine the real distribution of rainfall [9][10], but they are very expensive;
- the satellite ensures greater spatial and temporal resolution, but the estimate itself is less accurate [11][12].

For all these reasons, these systems are not very fast and accurate and are expensive to implement in smart cities. So, they cannot be used to estimate the intensity of rain in smart cities.

Aiming to implement rainfall estimate systems in smart cities, our idea is to use the already existing 4G/LTE network infrastructure. The innovative idea, proposed for the first time in this study, concerns the possibility to determine rainfall intensity based on the impact it has on the LTE radio channel parameters adopted for the cell selection mechanism. In particular, the study highlights the correlation between some parameters of the LTE system that measure the quality of the radio channel, i.e. the handover mechanism that selects the best base radio station for that particular mobile terminal. Compared to a previous study [13] [14], focused on the measurement of a single parameter related to the signal strength, in this study, the authors propose a new, larger set of parameters and use a wider class set of rain levels, including the "No rain" class and adding the "Shower" and "Cloudburst" classes.

Therefore, the first part of the study involves analysing the impact of different rain level statistics on the main parameters adopted by the LTE radio-mobile system for the cell selection mechanism to hook to.

Finally, keeping in mind the recent spread of artificial intelligence and machine learning techniques applied in many contexts, the second part of the study is devoted to defining a pattern recognition technique based on the average and variance of parameters that characterize the quality of the LTE radio channel and an MLP (Multi-Layer Perceptron) neural network, leaving out the use of the latest and most advanced machine learning techniques for future work.

The paper is organized as follows: Section II briefly summarizes the main studies regarding the rainfall classification based on a radio link, followed by Section III which illustrates the cell selection criteria adopted in LTE cellular networks; Section IV proposes a new approach to rainfall classification and Section V presents the testbed scenario; Section VI provides an overview of the neural networks adopted for this study; Section VII depicts the data analysis; Section VIII shows the results obtained using the MLP neural network; Section IX is devoted to discussion and comparison between the present study and other studies. Finally, Section X is devoted to conclusions.

II. RELATED WORKS

The urbanization phenomenon has caused the insurgence of a significant number of risks, concerns, and problems, which lead solicitous administrations to seek optimal solutions. According to researchers, such solutions can only be found in "smartness", where "smart" can be sustainable, liveable, secure, it can be green or connected. Indeed, "smart city" can be defined as the aim to reach all of these objectives through ICT (Information and Communication Technologies). Generally speaking, using ICT to smartify an object stands for adding two features to the normal functioning of the object: sensing and automation [15].

The smart city paradigm focuses on six characteristics:
- Smart Economy;
- Smart Governance;
- Smart Living;
- Smart Mobility;
- Smart Healthcare;
- Smart Environment.

This study is particularly concerned with the latter characteristic. In fact, in this paper we propose a smart rain gauge that exploits the impact of rain on radio signals, in particular that between the existing radio base stations using 4G/LTE technology (at frequencies between 1.8 GHz and 2.4 GHz) and the numerous mobile terminals in circulation.

Several past studies on this subject have only considered electromagnetic waves with frequencies greater than 10 GHz [16 – 18] since the impact of precipitation on the attenuation of electromagnetic waves, and therefore on the strength of the signal receiver, is best visible at high frequencies. A very interesting scenario, however, is represented by mobile radio systems that are widespread in the territory. Hence, in recent years, similar studies have been conducted considering frequencies used by cellular networks, therefore, less than 3 GHz, trying to analyze the effect of the impact of rain on the parameters that characterize the quality of the radio-mobile channel. In particular, in [19], the effect of rain on the Received Signal Strength Indicator (RSSI), i.e. the intensity of the signal received by the user, was studied.

Network analysis on 2G links was performed in two different geographies for nine non-consecutive days. Measurements of the signal quality of each telephone terminal were transferred to a computer at the end of the day, along with measurements obtained by a rain gauge. The effect of rain on the intensity of the cellular signal was studied by analyzing variations in RSSI values measured by the smartphone. Tests showed an RSSI drop during rainfall in 8 out of 9 cases. In general, however, the decline in RSSI did not lead to a clear and unambiguous distinction between the various levels of rainfall, since the decrease in power was insignificant. In [20], the measurements were taken for one year, using a transmitter/receiver system consisting of conventional antennas at the operating frequency of 2 GHz. The results showed reliable and accurate measurements for amounts of rain less than 1 mm for periods of 5 minutes. In our previous paper [21][22], we proposed a study on the estimation of the rainfall level, based on the intensity of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011375, IEEE Access

received signal in LTE systems, evaluating the parameters of mean, variance and instant value of the RSSI calculated in a sliding time window. These parameters were subsequently inserted into a Probabilistic Neural Network, which resulted in a satisfactory classification performance. The study deals with a first approach to using radio signal parameters for precipitation estimation, using frequencies from LTE technology (1.8 MHz/2.4 GHz). It is the first approach, as the distinction between rainfall levels (weak, moderate and strong) was not clearly evident and easily defined by the RSSI values. Previous studies suggest that it is particularly difficult to classify rainfall levels only by taking into consideration signal strength received at frequencies below 3Ghz. For this reason, in this paper we focus on the study and analysis of additional radio parameters, other than the strength of the signal received, i.e. RSSI, defined by the LTE technology, which are able to provide the most accurate estimate of the rainfall level. In addition, we assess the effectiveness, in terms of classification, of all radio channel statistical parameters provided by the LTE technology, i.e. mean and standard deviation. Finally, the obtained data are fed as input to a Multilayer Layer Perceptron (MLP) neural network [23] which differentiates between various levels of rainfall. The new rain gauge system, studied in this paper, offers great spatial resolution inasmuch as it is based on the impact of rain on radio routes between the mobile terminal and the base radio station. Base stations, in fact, have a fairly even and wide distribution in the territory and in cities. In smart cities of the future, characterized by intense use of the 5G radio route, this factor will be amplified as 5G base radio stations will have lower coverage radius and, therefore, much wider distribution in the territory than ever before.

This distribution allows to potentially convert each base radio station into a radio rain gauge. This technique comes with numerous advantages in terms of greater accuracy, speed and geographic accuracy, i.e. it will be possible to estimate the intensity of rain with high spatial precision.

At a functional level, even with the introduction of 5G technology, the proposed method will remain valid, as it will still be possible to use all the radio parameters to estimate the level of precipitation.

In perspective it is true that 5G uses higher frequencies offering a greater correlation between rainfall intensity levels and the RSSI parameter, but it is also true that 5G provides operating modes even at frequencies below 10 GHz. For this reason, for a more correct and robust classification of rainfall levels, all radio channel monitoring parameters proposed in this study are required to be used also in 5G systems.

III. CELL SELECTION CRITERIA IN LTE TECHNOLOGY

The smart city paradigm is a vision for future cities centred around the concept of connectivity. Indeed, connectivity is the core requirement for smart cities to exist, enabling tight integration among citizens, devices and service providers. However, it is also a means for interoperable access and interconnection among different services.

Studies have been conducted on the use of the LTE infrastructure to implement different methodologies and scenarios that form smart cities [24 – 26].

This section describes the main features of the LTE technology with particular reference to the parameters that characterize the quality of the radio channel between the mobile terminal and the base station, as well as the handover mechanism.

A. DESCRIPTION OF NETWORK PARAMETERS

In an LTE cellular network, when a mobile terminal moves between cells or can no longer have certain signal strength requirements from the cell it is hooked up with, it must perform the selection/re-selection operation of a base station. For such an operation it is necessary to measure the strength and signal quality of the neighbouring cells. In LTE, the E-UTRAN Node B, also known as Evolved Node B (abbreviated as eNodeB or eNB), is the element in E-UTRA of LTE that is the evolution of the element Node B in UTRA of UMTS. It is a hardware connected to the mobile phone network that communicates directly wirelessly with mobile handsets (UEs), like a base transceiver station (BTS) in GSM networks. In LTE, a UE (User Equipment) measures the following two parameters of the reference signal, signalling them to the electronic node:

- Reference Signal Received Power (RSRP);
- Reference Signal Received Quality (RSRQ).

From these two indexes, the eNB returns the Received Signal Strength Indicator (RSSI) parameter, which is the reference signal intensity indicator. The RSRP, which typically ranges between - 44 and - 140 dBm, is a good measure of the power of a specific sector, excluding noise and interference from other sectors. When the UE is near an LTE station, the average RSRP values are around -75 dBm, and around -120 dBm when the UE is near the edge of the cellular coverage area [27].

B. CELL ACQUISITION

After being turned on, the mobile device performs a low-level capture procedure to identify nearby LTE cells and find out how they are configured. The acquisition process is summarized in the following stages: The UE receives synchronization signals from all nearby cells. From the Primary Synchronization Signal (PSS), the mobile terminal reads the symbol timing and gathers information about the identity of the physical cell; from the Secondary Synchronization signal (SSS), the UE derives the frame timing, Physical Cell Identity (PCI), transmission mode (FDD or TDD), and duration of the cyclical prefix (normal or extended). At this point, the UE initiates the reception of cell-specific reference signals. These provide a reference of amplitude and phase for the channel estimation process, so they are essential for the following steps. The UE is then assigned the physical transmission channel and, subsequently, the UE reads the main block of information. The UE, after the first and second synchronization signals, receives the control format indicators.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011375, IEEE Access

Finally, the UE initiates the reception of the Physical Downlink Control Channel (PDCCH). This allows the UE to read the remaining blocks of system information (SIB), which are sent on the Physical Shared Downlink Channel (PDSCH). It captures all the remaining cell configuration data, such as the identities of the networks to which it belongs. The UE then initiates the reception of reference signals in the downlink channel [28]. Such signals are useful for the mobile terminal to:

- provide an amplitude and a phase reference to be used in the estimation of channels;
- measure the strength of the received signal according to the frequency;
- calculate the channel quality indicators.

These procedures are carried out while the terminal is in the "IDLE" state, i.e. there is no active phone call or data transfer.

C. CELL SELECTION

At this stage, the mobile terminal begins by performing the procedure of selecting the network and cell, which involves two main steps. Firstly, the UE selects a Public Land Mobile Network (PLMN) it will register with; secondly, it selects a cell that belongs to the selected network. Cell selection can be done in two ways. Usually, the UE has access to information stored on the potential frequencies and cells of the LTE service operator, starting from the last turning on or network selection procedure described above. If this information is not available, the device scans all supported LTE carrier frequencies and identifies the most powerful cell on each carrier on the selected network.

The selected cell is the one that meets several criteria, from release 9 onwards, standardized by 3GPP [29]. The most important criterion is show in (1):

$$S_{qual} > 0 \tag{1}$$

During the initial network selection, the UE calculates S_{qual} as follows:

$$S_{qual} = Q_{qualmeas} - Q_{qualmin} \tag{2}$$

In this equation, $Q_{qualmeas}$ [30] is the quality of the received reference signal (RSRQ) measured, which indicates the signal-interference ratio plus the noise ratio of cell-specific reference signals. $Q_{qualmin}$ is the minimum value for the RSRQ, which the base station makes available in System Information Block 1 (SIB 1). This prevents a mobile phone from selecting a cell on a carrier frequency that is subject to high levels of interference.

D. CELL RESELECTION

For mobile terminals in the “IDLE” state, cell reselection management procedures have two main goals: maximizing mobile terminal battery life and minimizing signalling load on the network. From release 9 [31], a UE can also start taking measurements on neighbouring cells if the quality of the received reference signal (RSRQ) falls below a threshold:

$$S_{qual} \leq S_{IntraSearchQ} \tag{3}$$

where, $S_{IntraSearchQ}$ is another threshold made public by the base station in SIB 3. S_{qual} depends on the RSRQ of the service cell and is calculated using (2). After finding and measuring the neighbouring cells, the UE calculates the classification scores of the service cell and one of its neighbours. The UE then moves on to the cell with the best score, provided that three conditions are met. Firstly, the UE must have been hooked up to the service cell for at least a second. Secondly, the new cell must meet the criteria set out in "Cell Selection". Finally, the new cell must be better classified than the service cell for a certain period between 0 and 7 s.

IV. RAINFALL CLASSIFICATION BASED ON RADIO SIGNAL QUALITY PARAMETERS

The new method for estimating rainfall levels, proposed in this paper, is based on a nonlinear matching pattern recognition approach. As in previous studies carried out in audio biometrics [32 – 33], once the set of parameters that characterizes the radio channel is extracted, they are analyzed on time windows by measuring the statistical parameters of the first and second order. The parameter set is sent to a nonlinear matching block based on multi-layer perceptron neural networks (MLP). In particular, the radio parameters used for analyzing and creating the dataset in input to the neural network were measured in terms of average and variance, via the app GMON for each level of rainfall. The following radio parameters were therefore obtained:

$RSRP_{AVG}$, the average power of the RSRP signal received from the mobile terminal;

- $RSRQ_{AVG}$, the average signal power quality received from the mobile terminal;
- RXL_{AVG}, the average instant power of the RXL signal received from the mobile terminal;
- SNR_{AVG}, the average signal-to-noise ratio received from the mobile terminal;
- $RSRP_{VAR}$ indicates variations in signal strength of the RSRP signal received from the mobile terminal between two consecutive measurements;
- $RSRQ_{VAR}$ indicates variations in received signal strength quality from the mobile terminal between two consecutive measurements;
- RXL_{VAR} indicates variations in instant power of the RXL signal received from the mobile terminal between two consecutive measurements;
- SNR_{VAR} indicates variations in the signal-to-noise ratio received from the mobile terminal between two consecutive measurements.

In addition, with regard to the Cell Identification (CID), i.e. the identification number of the LTE cell the terminal is hooked to, several parameters have been calculated:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011375, IEEE Access

- The "reference CID", CID_{CSR}, the CID for which the terminal remains hooked to for longer;
- The "CID_{HOPS}", the number of connections to other cells the mobile terminal makes;
- The $CID_{PERMANENCE}$ that indicates how long the mobile terminal remains hooked to the reference cell;
- The CID_{CSR} indicates the number of times the mobile terminal disconnects from the reference cell, and then returns to it.

These parameters are used for statistical analysis and as input to an MLP neural network.

V. THE TESTBED SCENARIO

In order to assess the impact of rain on the main parameters that characterize cell selection phases, the scenario taken into account is that of an LTE mobile terminal in the "IDLE" state, equipped with an application (e.g. GMON) for the measurement of the parameters used by the cell selection mechanisms (Fig. 1). Nearby, a classic tipping bucket rain gauge records rainfall levels in mm/h using a processing, labelling and synchronizing board with radio-mobile channel data measured by GMON. The mobile terminal, as described in the previous section, detects the power level of the base radio stations nearby and hooks to the station with the greatest power.

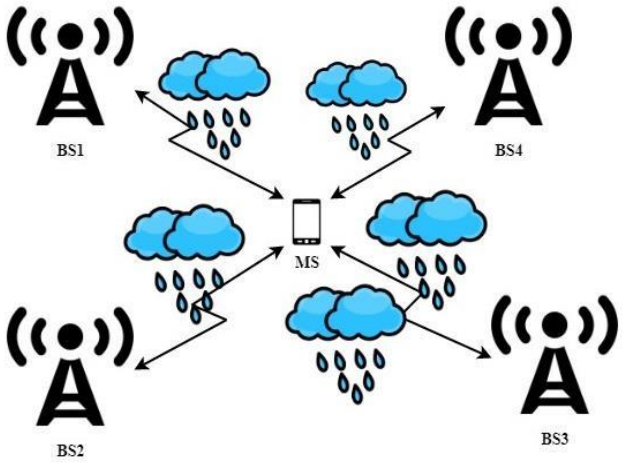

FIGURE 1. Testbed scenario.

In cases where atmospheric precipitation occurs, in an "IDLE" state the mobile terminal may be affected by small changes in the radio signal. The mobile terminal, in fact, receiving a lower power signal from the base station to which it is hooked, could pick up a higher power signal from another base station nearby. The impact of the rain can, thus, lead to greater frequency of the reselection phases of the base station to which the device is hooked. This data can have a major impact on the classification of rainfall levels as the higher the level of rainfall the higher the selection and reselection procedures of one or more neighbouring cells may be. The parameter used to describe this procedure is the CID, which distinguishes, as already seen, the cell to which the mobile terminal is hooked, the number of hops, i.e. jumps that it makes from one reference cell to another, and the timespan of continuous connection to a given reference cell; which is the one to which it is statistically hooked to the longest. Following, in the next subsections, we will describe the testbed and the database used to train the MLP neural network.

A. THE TEST CAMPAIGN

This section will describe RSRQ, RXL, SNR, RSRP, and CID (LTE defines it as PCI) data collection procedure. The data employed for the creation of the radio database, on which the analyses and tests were carried out, were collected through an ad hoc implemented acquisition system consisting of: a tilting tub, a processing unit and a 4G smartphone SIM inside a shaker.

A dedicated application, called GMON, is installed on the smartphone, able to export a full report for different network and signal parameters such as RSSI, RSRP and RSRQ (illustrated in section III) in CSV format. A), signal-to-noise ratio (SNR), Location Area Code (LAC), CID and connection type (LTE, HSPA, UMTS, etc.). As for the tipping bucket rain gauge, it acts as a rain indicator. It includes a rain-gathering funnel, two triangular tubs mounted on a fulcrum and an electronic switch.

The rain is channelled through the funnel to one of the trays. When the tub is full, it loses balance and flips over, emptying into the outer shell of the meter, while the other tub is lifted into place for later reading. The rain gauge is connected to the processing unit via an RJ11 cable and is managed ad hoc through a software interface that can detect and count the "interruptions" generated by the rain gauge tray whenever a tilt occurs.

TABLE I
RAIN CLASSIFICATION AND PRECIPITATION INTENSITY RANGE

Rain Classification	Acronyms	Precipitation Intensity
No rain	nr	< 0.5 mm/h
Moderate rain	m	[0.5 ÷ 6] mm/h
Heavy rain	h	[6 ÷ 10] mm/h
Shower	s	[10 ÷ 30] mm/h
Cloudburst	c	> 30 mm/h

To process the proposed classification approach, a database was created by recording the collected data. The database includes measurements of the parameters described above in five different weather conditions: "No rain", "Moderate rain", "Heavy rain", "Shower" and "Cloudburst". The database framework is displayed in Table 1.

B. DATABASE

The database is made with GMON for each terminal and comprises data from LTE radio parameter measurements and CID. This data is sent at once as the tipping bucket rain gauge is activated. Tips generated by the tipping bucket rain gauge are sent to an IoT platform, using the publisher/subscriber protocol. The values taken from the tray are used to label the radio signal with different rain intensity. A labeling algorithm is applied to obtain the estimate in mm/h

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011375, IEEE Access

(described in detail in [14]). Once the labelling algorithm is executed, the rainfall values, in mm/h, are compared and synchronized in time with the data in the CSV file created by GMON. This process enables data labelling in the CSV file. In fact, this allows obtaining files where the network parameters are related to each classification level. The database consists of five categories of precipitation intensity, defined in Table 1; for each category there is a CSV file containing network parameters: RSRP, RSRQ, RSSI and SNR; date and time from when the data was recorded; CID and LAC. The current database was created by recording network parameters on different days and in different areas of the territory during the precipitation (in various rain intensity conditions, including the "No rain" case). This implies a certain robustness of the system, as the results obtained take into account the use of base stations in different locations.

VI. NEURAL NETWORK DESCRIPTION

In this section we will describe the chosen neural network and supervised learning algorithm used in our study and applied to the statistical parameters of the first and second-order LTE radio channel. One of the first algorithms used in machine learning for supervised learning is the Single Layer Perceptron (SLP). Frank Rosemblat in [34] published the first concept of the Perceptron learning rule based on the McCullock-Pitts (MCP) neuron. This learning rule consists in an algorithm that automatically expresses the optimal weight coefficients to be multiplied by input characteristics. This multiplication operation allows making the decision on whether to activate the neuron or not; with the possibility, therefore, to define whether or not a given sample belongs to a particular class.

Mathematically, it is possible to have input signals x_i and weights w_i. One must define a function of activation, which works on a linear combination of certain X input values and a corresponding vector of W weights, where Z is the neural network input:

$$W = [w_i \ \dots \ w_m], \ X = [x_i \ \dots \ x_m] \tag{4}$$

$$Z = (w_1 x_1 + \ w_2 x_2 + \cdots + \ w_m x_m) \tag{5}$$

$$\Phi(z) = \{1 \ if \ z \geq 0 \ \ -1 \ otherwise \tag{6}$$

If the activation of a particular sample $x^{(i)}$, which is the output of the $\Phi(z)$, is greater than the given threshold, one can predict which class it belongs to. In the Perceptron algorithm, the activation function is defined in sections.

The input of the $z = W^T X$ network is reduced to a binary output by the Perceptron activation function, thus being used to linearly discriminate between the two classes.

SLP is a very simple neural network, which manages well to classify only if the variables are linearly separable. Based on the statistical analysis described in the following sections, it is possible to note that the variables involved in this study cannot be separated linearly.

For this reason, we have used a Multi-Layer Perceptron (MLP) network which, unlike the SLP which has a single hidden layer and a single neuron, contains multiple hidden layers and for each layer there are multiple neurons. The substantial difference is that sigmoid functions such as logistic regression are used in MLP networks, to activate neurons, according to (7).

$$\Phi(z) = \frac{1}{1-e^{-z}} \tag{7}$$

MLP is a typical example of a feedforward artificial neural network. The term feedforward refers to the fact that each level acts as input to the next level, without loops.

As for our study, it is possible to observe in Figure 2 that we used a five-layer MLP. The first is the input layer, the second, third and fourth are hidden layers with 150, 100 and 50 neurons, respectively, while the last layer is the output layer.

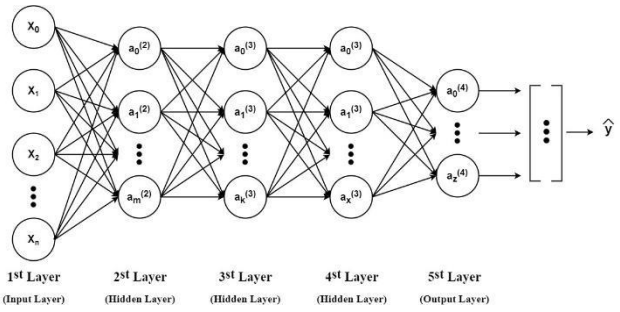

FIGURE 2. Diagram of how multilayer perceptron works.

VII. DATA ANALYSIS AND RESULTS

This section presents the data analysis and the main results obtained in this study. As previously mentioned in Section V, the data collected by GMON are stored, every second, in a CSV file and this file is composed of records containing the data related to the radio parameters indicated above.

In particular, the number of records for each level of rainfall is as follows:

- No rain: 124.210 records
- Moderate rain: 378.584 records
- Heavy rain: 18.667 records
- Shower: 15.595 records
- Cloudburst: 5.255 records

Initially, the first and second-order statistical parameters of the radio channel quality parameters described in section IV were calculated.

The statistical analysis was conducted by considering a 180-second sliding window with 15 second off-sets. The first order statistical parameters, represented by the averages of the values, are as follows:

- RXL_{AVG}
- $RSRP_{AVG}$
- $RSRQ_{AVG}$
- SNR_{AVG}

The statistical parameters of the second order, represented by the variations, are as follows:

- RXL_{VAR}
- $RSRP_{VAR}$
- $RSRQ_{VAR}$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011375, IEEE Access

- SNR_{VAR}

The number of databases for each class is as follows:

- No rain: 686 records
- Moderate rain: 2089 records
- Heavy rain: 103 records
- Shower: 86 records
- Cloudburst: 29 records

Due to the non-uniformity of the data in the dataset for each class, it was decided to reduce the number of samples in the "No rain", "Moderate rain", "Heavy rain" and "Shower" classes to 50 records each and to leave the "Cloudburst" class unchanged.

The three cell change parameters were also analysed: CID_{HOPS}, $CID_{PERMANENCE}$ and CID_{CSR}. Subsequently, the degrees of linear separation were assessed using the Fischer Discriminant Ratio (FDR), which allows measuring the degree of linear separation that the given parameter has [35]. Finally, a multi-layer perceptron-type neural network (MLP) was applied, providing the parameters used for statistical analysis as an input, and the 5 classes of rainfall levels as an output. After the training phase the neural network ranks the contribution made by each parameter considering that the best match is the one based on non-linear techniques. Finally, by analysing the confusion matrix, the accuracy of the system is determined.

A. STATISTICAL DATA ANALYSIS

First, mean and variance of radio parameters are studied to determine if there is a link between the values obtained from the statistical analysis of these parameters and the rainfall levels defined in Table 1. The statistics of the individual parameter with the representation of mean, minimum and maximum, standard deviation and typical distribution of values around the mean were represented for each class of rainfall.

Figure 3, 4, and 5 show that there are no parameters clearly distinguishing between classes, although some classes have low-overlapping value distributions with other classes (e.g. $RSRP_{AVG}$ and RXL_{AVG} have “No Rain” class values that overlap with those of the other classes). The same goes for the $RSRQ_{AVG}$ parameter that ensures a good separation between the “Shower” and “Cloudburst” classes. The same case applies to the variance of radio parameters (Figure 4).

Figure 5 shows the average jump that is made from one cell to another at each level of rainfall. In this case, there is a clear distinction between "s" and "c" classes. In general, it is noted that the data appears to be jagged and does not give a clear distinction between precipitation levels according to linear analysis criteria. A simple first and second level statistical analysis is therefore not enough to establish clear classification of rainfall levels. For this reason, section B, following an analysis of the Fisher Discriminant Ratio applied to radio parameters, presents the results of classification techniques suitable for those cases with non-linear separation in order to define the parameter that helps obtain better discrimination.

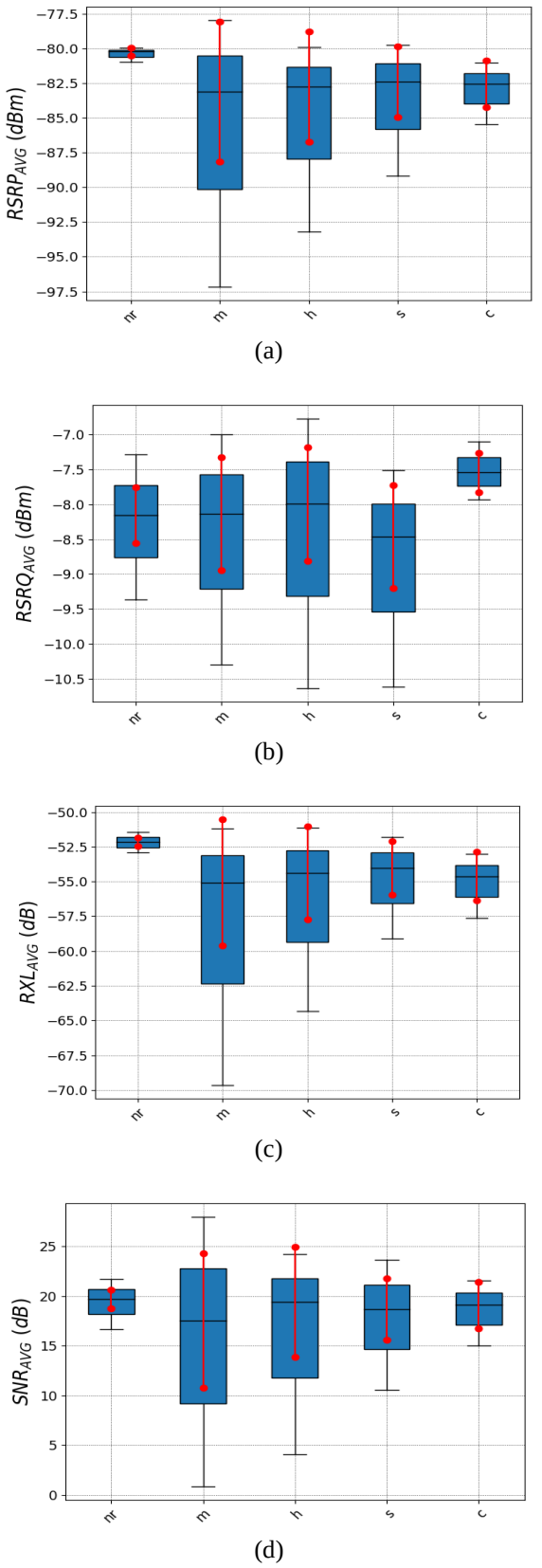

FIGURE 3. Representation of radio parameter averages statistics: (a) average RSRP, (b) average RSRQ, (c) average RXL, (d) average SNR.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011375, IEEE Access

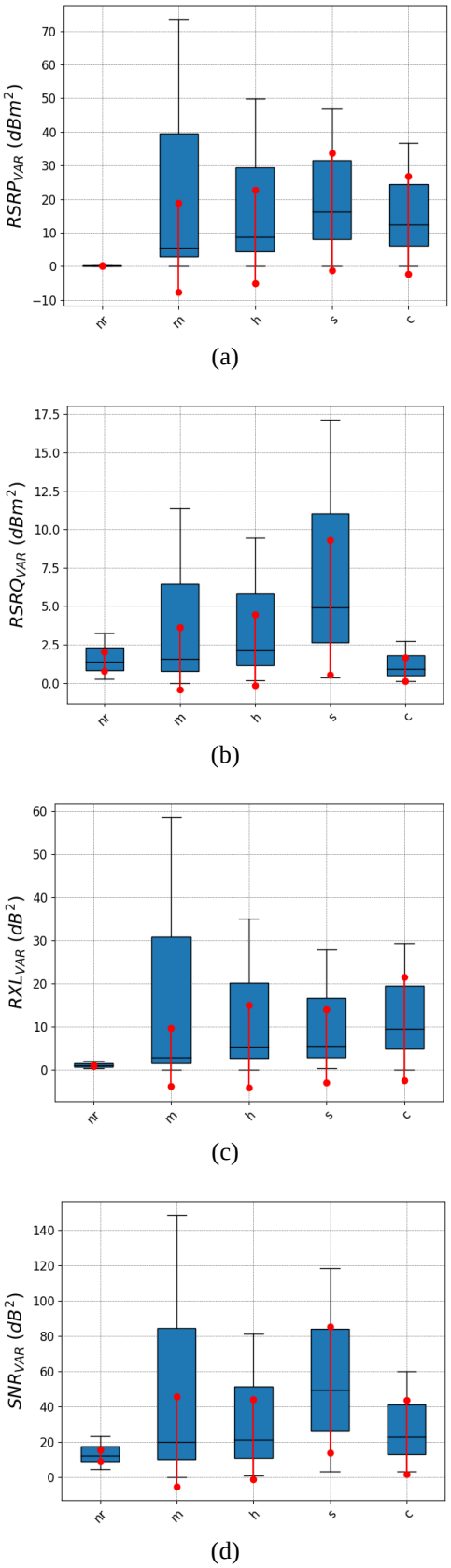

FIGURE 4. **Representation of radio parameter variance statistics: (a) RSRP variance, (b) RSRQ variance, (c) RXL variance, (d) SNR variance.**

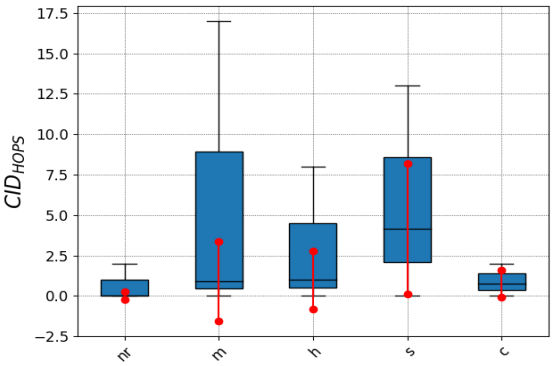

FIGURE 5. **Representation of CID_{HOPS} statistics.**

B. LINEAR AND NON-LINEAR SEPARABILITY BETWEEN THE CLASSES

In this section, FDR value between two adjacent classes, relative to each parameter described in the previous section, will be evaluated. As there are five levels of rainfall considered in this study, there will be five classes and, therefore, four FDR values, for each radio parameter considered, for the classes: NR-M, M-H, H-S, S-C. Figure 6 shows the total FDR values of all parameters for each pair of adjacent classes. The parameters that have the highest degree of linear separation are:

- the RXL_{AVG} and the $RSRP_{AVG}$ for classes NR to M (Figure 6.a);
- The SNR_{AVG} and RXL_{VAR} for classes M through H (Figure 6.b);
- he CID_{CSR} and CID_{HOPS} for classes H to S (Figure 6.c);
- $RSRQ_{AVG}$ and $RSRQ_{VAR}$ for classes S through C (Figure 6.d).

By analysing and calculating the degree of total linear separation (see Figure 7), i.e. applicable to all classes, we obtain the parameters with the highest linear separation index: the $RSRQ_{AVG}$, the CID_{CSR} and the CID_{HOPS}. Starting from these values, it was decided to carry out the Principal Component Analysis (PCA) [36] of these parameters. PCA is a technique aimed at deriving a smaller set of "artificial" orthogonal variables starting from a set of correlated numerical variables. The reduced set of linear orthogonal projections (known as "principal components", "PC") is obtained by appropriately combining the original variables linearly.

Figure 8 outlines the results obtained by the PCA. As evidenced, individual parameters contribute differently to the separability of the individual classes. To find a solution to this problem, we implemented a Multilayer Perceptron Neural Network (MLP) with multiple classes, where the above parameters are used as input, while the output will comprise the accuracy of the classification and the determination of parameters which are more important than others to neural network training.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011375, IEEE Access

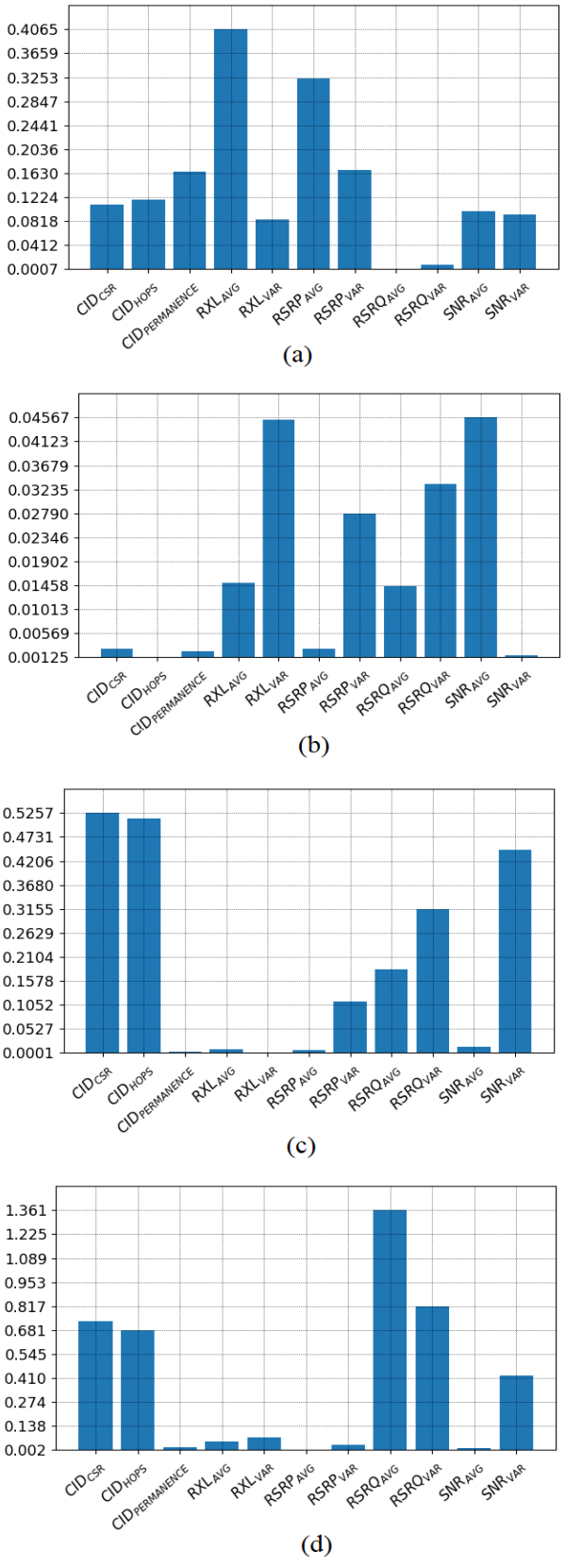

FIGURE 6. **Parameter FDR values between adjacent classes: (a) FDR between NR – M, (b) FDR between M – H, (c) FDR between H – S, (d) FDR between S – C.**

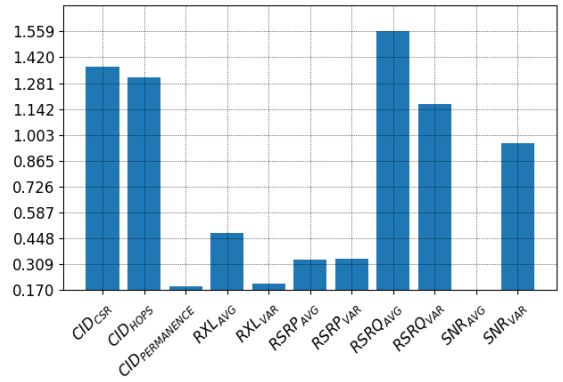

FIGURE 7. **Total FDR.**

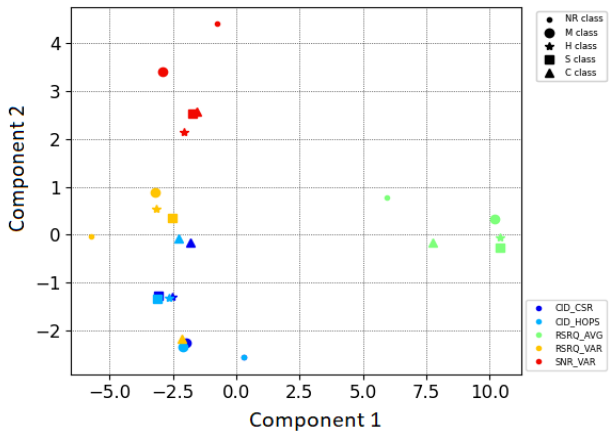

FIGURE 8. **PCA analysis.**

VIII. PERFORMANCE EVALUATION

Once the statistical analysis was completed and the FDR was calculated, as noticeable from Figure 8, the separation of classes is not linear. For this reason, the Perceptron neural network was applied to the dataset. The dataset used for training and testing the Perceptron network has been described in Chapter IV.B.

In particular, from the CSV files, containing the recordings of radio parameters made by GMON, relating to each precipitation level, the statistical parameters were extracted within a window of 180 seconds (data is recorded every second) and a 15 second offset. The dataset is therefore divided as follows:

- Fifty data samples for the "No rain" class;
- Fifty data samples for the "Moderate" class;
- Fifty data samples for the "Heavy" class;
- Fifty data samples for the "Shower" class;
- Twenty-nine data samples for the "Cloudburst" class.

The training data makes up 70% of the dataset, while the remaining 30% of the dataset is used to the testing phase, once the neural network was trained. The result of this phase is shown in Figure 9 with the confusion matrix. The statistical classification functions [37] are applied to evaluate the proposed method, based on the results obtained from the confusion matrix.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011375, IEEE Access

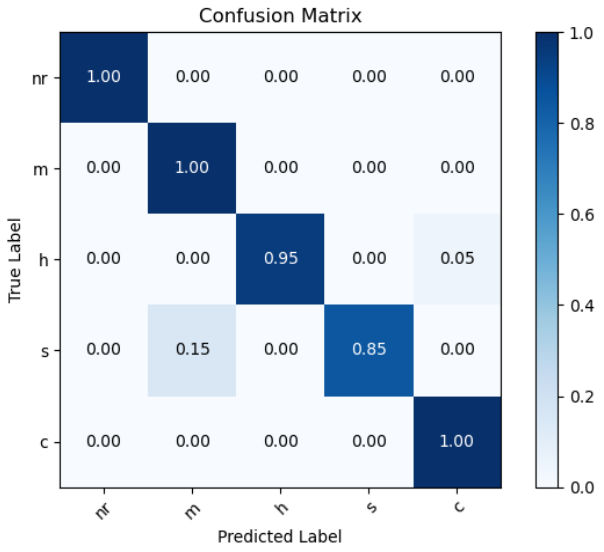

FIGURE 9. **Confusion matrix.**

In particular, we applied sensitivity, also known as True Positive Ratio (TPR); Fall – out, also known as False Positive Ratio (FPR); Precision (PRE) and Recall (REC), which are performance metrics closely related to the TPR and FPR values. In practice, a combination of precision and recall is often used, i.e. the so-called F1 score.

The equations below are related to the previously described classification functions

$$TPR = REC = \frac{TP}{TP+FN} \tag{8}$$

$$FPR = \frac{FP}{FP+TN} \tag{9}$$

$$PRE = \frac{TP}{FP+TP} \tag{10}$$

$$F_1 = 2\frac{PRE \times REC}{PRE + REC} \tag{11}$$

The results obtained in terms of statistical parameters are shown in Table II.

Table II suggests good performance of Accuracy (96.0%), Sensitivity/Recall (96.0%), Fall – out (13.32%), precision (91.4 %) and F1 score (93.6 %).

TABLE II
THE OVERALL VALUES OF ACCURACY, TPR, FPR, PRE AND F1 SCORE

Class	Mean Accuracy	TPR/ REC [%]	FPR [%]	PRE [%]	F1 Score [%]
nr		100	0	100	100
m		100	3.8	87	93
h		95	0	95	95
s		85	0	85	85
c		100	2.8	90	95
Total	96	96	13.32	91.4	93.6

Once these accuracy metrics are implemented and calculated, it is possible to use an additional model validation tool, namely the one based on the ROC (Receiver Operator Characteristic) and ROC AUC (Area Under the Curve) graphs to validate the model.

A perfect classifier would be located in the upper left corner of the graph, with a true positive rate equal to 1 and a false positive rate equal to 0. Based on the ROC curve, we can therefore calculate the area under the curve, AUC, to characterize the performance of the classification model.

Applying this concept to our classification method, in Figure 10 we observe the resulting ROC curve which indicates that a certain degree of variance between the various parts and the average ROC AUC lies between the perfect score (1.0) and the diagonal (0.5).

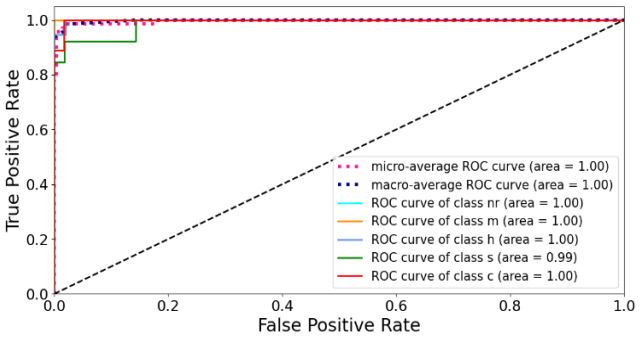

FIGURE 10. **ROC curve.**

Research on the performance of a classifier such as ROC AUC may provide additional information on its performance compared to unbalanced samples.

In [38] Bradley demonstrated that ROC AUC and accuracy metrics generally agree with each other.

The micro-average is calculated from the individual true positives, true negatives, false positives and false negatives of the system. In our case the micro-average of the precision score in a 5 (nr, m, h, s and c) class system can be calculated as follows:

$$PRE_{micro} = \frac{TP_{nr} + TP_m + TP_h + TP_s + TP_c}{TP_{nr} + \cdots + TP_c + FP_{nr} + \cdots + FP_c} \tag{12}$$

The macro-average is simply calculated as the average scores of the different systems:

$$PRE_{macro} = \frac{PRE_{nr} + PRE_m + PRE_h + PRE_s + PRE_c}{5} \tag{13}$$

Micro-media is useful if we want to weigh each instance or forecast, while macro-media weighs all classes equally to evaluate the overall performance of a classifier compared to the labels given to the most frequent classes. The graph shows that the area under the ROC curve is very large. This means that our model has excellent performance.

In general, tipping bucket rain gauge labelling does not allow establishing typical variations in rain intensity. By means of the causal forest technique [39], it is possible to gather useful information concerning the importance that the network assigns to various input parameters. The graph in Figure 11 shows that the five features in the order of importance for the neural network are: RXL_{AVG}, $RSRP_{AVG}$, SNR_{AVG}, $RSRQ_{AVG}$ and $RSRP_{VAR}$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011375, IEEE Access

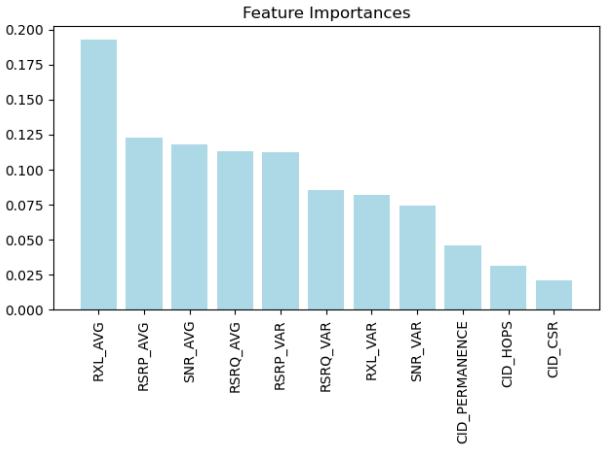

FIGURE 11. **Level of importance of features.**

IX. DISCUSSION

In this section the obtained results are discussed and compared with recent studies in the state of the art.

Our five-class classifier (nr, m, h, s and c) has managed to obtain classification accuracy of 96.0 % and an F1 score of 93.6 %. Table III shows the comparison in terms of radio technology in use, features, classification methods, classification levels and achieved performance.

Based on research carried out on the current state of knowledge, the study referenced in [22] and the one described in [40] are the only studies dealing with the classification of precipitation levels by radio signals at frequencies below 10 GHz. In [40] GSM technology is used, and the RSL (Received Signal Strength) parameter is considered as a feature. As for the classifier, a 3-class (nr, w, h) SVM (Support Vector Machine) is used. Compared to the study [22], it can be seen that overall it has lower performance, since in [22] we consider 4 rainfall levels (nr, w, m, h).

In [41] commercial microwave link is used to analyze the radio signal attenuation caused by rain at frequencies about 20 Ghz. In this study the authors adopted the KFD (Kernel Fisher Discriminant) method for intensity classification into only 3 classes: dry (no rain), rain and sleet (i.e. melted snow/freezing rain which, for simplicity, has been considered as weak rain).

It should be emphasized that the comparison proposed in Table III presents a congruity of classes only for the two No rain/Rain columns, from which there is already a clear performance improvement in the proposed method compared to the existing solutions. As indicated, the classes of rain levels vary in number and therefore it is not possible to make a direct comparison between the various methods. It should also be noted that the datasets used are different.

TABLE III
COMPARISON OF RAINFALL CLASSIFICATION METHODS

Researcher	Radio Technogies	Features	Methods	F1 Score [%]	Accuracy [%]						
					No rain	Rain	Rain levels				
							w	m	h	s	c
Brito et al. [40]	GSM	RSL	SVM	n.d.	82.84	84.79	75.78	n.d.	88.26	n.d.	n.d.
Cheakassky et al. [41]	Commercial Microwave link	Signal attenuation	KFD	n.d.	96.74	80.18	26	n.d.	n.d.	n.d.	n.d.
Beritelli et al. [22]	4G/LTE	RSL (value, mean and variance)	PNN	n.d.	100	n.d.	90	96.7	100	n.d.	n.d.
Proposed Method	4G/LTE	RSRP, RSRQ, RXL, SNR (mean and variance), CID	Multiclass MLP	93.6	100	95	n.d.	100	95	85	100

However, comparing this study with the previous one [22], it can be said that in this study several classification levels are proposed, adding the highest intensity classes (Shower and Cloudburst) and using more parameters of 4G radio technology to strengthen the system in the previous study [22] which exploits only the mean and the variance of the RSL parameter.

In fact, it is well known that the radio signal strength may undergo variations depending on the conditions of signal propagation and, therefore, on the surrounding environment.

With regard to the matching technique, this study exploits a very simple neural network with low computational complexity, compared to the study proposed in [22], achieving, however, overall good performances.

At high frequencies the impact of rain on attenuation is very big, resulting in greater performance in terms of classification accuracy. In any case, for further performance optimization, it is always useful to add the set of parameters proposed in this study in addition to that of simple signal strength.

X. CONCLUSIONS

The paper proposes an innovative approach to rainfall classification for smart city applications. The main idea is based on the impact that rain has on a set of parameters that characterize the radio-mobile channel quality. The paper highlights the link between rainfall levels and the trend of parameters adopted for the cell selection phase in the LTE mobile network. In particular, the system requires only the extraction of the parameters that a mobile terminal measure and the subsequent comparison with a nonlinear matching system based on MLP networks. The performance is very good in terms of accuracy and spatial resolution. Taking into account the typical micro-variances of rainfall intensity, it is

possible to consider an average accuracy of 96 %. The new rain gauge exceeds the limits of the traditional ones, as it has no mechanical parts and requires no maintenance.

REFERENCES

[1] J. Schlingensiepe, F. Nemtanu, R. Mehmood, and L. McCluskey, "Autonomic Transport Management Systems—Enabler for Smart Cities, Personalized Medicine, Participation and Industry Grid/Industry 4.0," in *Intelligent Transportation Systems – Problems and Perspectives* , 1st ed., vol 32, Springer, 2015, pp. 3-35.

[2] I. Pramanik, R. Y. K. Lau, H. Demirkan, and M. A. K. Azad, "Smart health: Big data enabled health paradigm within smart cities," *Expert Systems with Applications*, vol. 87, pp. 370 – 383, Nov. 2017.

[3] B. L. R. Stojkoska, and K. V. Trivodaliev, "A review of Internet of Things for smart home: Challenges and solutions," *Journal of Cleaner Production*, vol. 140, no. 3, pp. 1454 – 1464, Jan. 2017.

[4] T-H. Kim, C. Romos, and S. Mohammed, "Smart City and IoT," *Future Generation Computer Systems*, vol. 76, pp. 159 – 162, Nov. 2017.

[5] S.M. Alzahrani, "Sensing for the Internet of Things and Its Applications," in *Proc. 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, Prague, Czech Republic, 21-23 Aug. 2017.

[6] J. Cabra, D. Castro, J. Colorado, D. Mendez, and L. Trujillo "An IoT Approach for Wireless Sensor Networks Applied to e-Health Environmental Monitoring," in *Proc. EEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, Exeter, UK, 21-23 June 2017.

[7] F. Beritelli, A. Gallotta, and C. Rametta, "A dual streaming approach for speech quality enhancement of VoIP service over 3G networks," In *Proc. IEEE International Conference on Digital Signal Processing, DSP*, Fira, Greece, 1-3 July 2013.

[8] W. Fan, and A. Bifet, "Mining big data: current status, and forecast to the future," *ACM SIGKDD Explorations Newsletter*, vol. 14, no. 2, pp. 1-5, April 2013.

[9] J.M. Trabal, and D.J. McLaughlin, "Rainfall estimation and rain gauge comparison for x-band polarimetric CASA radars," in *Proc. IEEE International Geoscience and Remote Sensing Symposium*, Barcelona, Spain, 23-28 July 2007.

[10] D. Nagel, "Detection of Rain Areas with Airborne Radar," in *Proc. 18th International Radar Symposium (IRS)*, Prague, Czech Republic, 28-30 June 2017.

[11] A. K. Shukla, C.S.P. Ojha, and R. D. Garg, "Comparative study of TRMM satellite predicted rainfall data with rain gauge data over Himalayan basin," in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGRSS)*, Valencia, Spain, 22-27 July 2018.

[12] A. K. Varma, "Measurement of Precipitation from Satellite Radiometers (Visible, Infrared, and Microwave): Physical Basis, Methods, and Limitations," *Remote Sensing of Aerosols, Clouds, and Precipitation*, pp. 223-248, 2018.

[13] R. Avanzato, F. Beritelli, F. Di Franco, and V. Puglisi, "A Convolutional Neural Networks approach to Audio Classification for Rainfall Estimation," In *Proc. The 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, 18-21 September 2019, Metz, France.

[14] R. Avanzato, and F. Beritelli, "A Rainfall Classification Technique based on the Acoustic Timbre of Rain and Convolutional Neural Networks," *MDPI Information*, vol. 11, no. 4, 2020.

[15] A. Arroub, B. Zahi, E. Sabir, and M. Sadik, "A Literature Review on Smart Cities: Paradigms, Opportunities and Open Problems," In *Proc. International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Fez, Morocco, 2016.

[16] D. Cherkassky, J. Ostrometzky, and H. Messer, "Precipitation Classification Using Measurements from Commercial Microwave Links," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2350 – 2356, May 2014.

[17] A. Overeem, H. Leijnse, and R. Uijlenhoet, "Rainfall Monitoring Using Microwave Links from Cellular Communication Networks: The Dutch Experience," in *Proc. IEEE Statistical Signal Processing Workshop (SSP)*, Freiburg, Germany, 10-13 June 2018.

[18] C. Han, and S. Duan, "Impact of Atmospheric Parameters on the Propagated Signal Power of Millimeter-Wave Bands Based on Real Measurement Data," *IEEE Access*, vol. 7, pp. 113626 - 113641, Aug. 2019. DOI: 10.1109/ACCESS.2019.2933025.

[19] S. Sabu, S. Renimol, D. Abhiram, and B. Premlet, "Effect of rainfall on cellular signal strength: A study on the variation of RSSI at user end of smartphone during rainfall," in *Proc. IEEE Region 10 Symposium (TENSYMP)*, Cochin, India, 14-16 July 2017.

[20] V. Christofilakis, G. Tatsis, C.T. Votis, S.K. Chronopoulos, P. Kostarakis, C.J. Lolis, and A. Bartzokas, "Rainfall Measurements Due to Radio Frequency Signal Attenuation at 2 GHz," *Journal of Signal and Information Processing*, vol. 9, no. 3, pp. 192 – 201, Aug. 2018. DOI: 10.4236/jsip.2018.93011.

[21] F. Beritelli, G. Capizzi, G. Lo Sciuto, F. Scaglione, D. Połap, and M. Woźniak "A Neural Network Pattern Recognition Approach to Automatic Rainfall Classification by Using Signal Strength in LTE/4G Networks," in *Proc. International Joint Conference on Rough Sets (IJCRS)*, Olsztyn, Poland, July 3-7, 2017.

[22] F. Beritelli, G. Capizzi, G. Lo Sciuto, C. Napoli, and F. Scaglione, "Rainfall Estimation Based on the Intensity of the Received Signal in a LTE/4G Mobile Terminal by Using a Probabilistic Neural Network," *IEEE Access*, vol. 6, pp. 30865 – 30873, June 2018.

[23] M. Christopher Bishop, "Pattern Recognition and Machine Learning," in *Information Science and Statics*, 1st ed., vol. 1, Verlag, NY:Springer, 2006.

[24] A. Cimmino, T. Pecorella, R. Fantacci, F. Granelli, T. Faizur Rahman, C. Sacchi, et all., "The role of small cell technology in future Smart City applications," *Transactions on Emerging Telecommunications Technologies*, vol. 25, pp. 11-20, Nov. 2013.

[25] M. S. Ali, E. Hossain, and D. In Kim, "LTE/LTE-A Random Access for Massive Machine-Type Communications in Smart Cities," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 76 – 83, Jan. 2017.

[26] F. Malandra, L. O. Chiquette, L. P. Lafontaine-Bédard, and B. Sanso, "Traffic Characterization and LTE Performance Analysis for M2M Communications in Smart Cities," *Pervasive and Mobile Computing*, vol. 48, pp. 59-68, Aug. 2018.

[27] 3GPP TS 36.211 (2013) Physical Channels and Modulation, Release 11, Section 6.10, September 2013.

[28] 3GPP TS 36.304 (2013) User Equipment (UE) Procedures in Idle Mode, Release 11, Sections 5.2.1, 5.2.2, 5.2.3, 5.3, September 2013.

[29] 3GPP TS 36.214 (2012) Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer; Measurements, Release 11, Section 5.1.1, December 2012.

[30] 3GPP TS 36.304 (2013) User Equipment (UE) Procedures in Idle Mode, Release 11, Section 5.2.4, September 2013.

[31] 3GPP TS 36.133 (2013) Requirements for Support of Radio Resource Management, Release 11, Section 4.2, September 2013.

[32] F. Beritelli, and A. Spadaccini, "A Statistical Approach to Biometric Identity Verification based on Heart Sounds," *in Proc. of the Fourth International Conference on Emerging Security Information, Systems and Technologies*, Venice, Italy, 18-25 July 2010.

[33] F. Beritelli and A. Spadaccini, "The Role of Voice Activity Detection in Forensic Speaker Verification," in *Proc. of the 17th IEEE International Conference on Digital Signal Processing (DSP 2011)*, Corfu Island, Greece, 6-9 July 2011.

[34] Rosenblatt Frank, "The Perceptron, a Perceiving and Recognizing Automation," Cornell Aeronautical Laboratory, Buffalo, NY, USA, Rep. 85-460-1, Jan. 1957.

[35] Title: Fischer Discriminant Ratio, Access Date: 13 Feb. 2020. https://en.wikipedia.org/wiki/Linear_discriminant_analysis.

[36] Title: A Complete Guide to Principal Component Analysis — PCA in Machine Learning, Access Date: 25 March 2020. https://towardsdatascience.com/a-complete-guide-to-principal-component-analysis-pca-in-machine-learning-664f34fc3e5a.

[37] C. Beleites, R. Salzer, V. Sergo, "Validation of soft classification models using partial class memberships: An extended concept of sensitivity & co. applied to grading of astrocytoma tissues,"

Chemometrics and Intelligent Laboratory Systems, vol. 122, pp. 12–22, 2013.

[38] A. P. Bradley, "The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, July 1997.

[39] A. K. Nandi, and H. Ahmed, "Decision Trees and Random Forests," *Wiley-IEEE Press*, 1st ed., 2019, pp. 199-224.

[40] A. Brito, L. Fernando, and M. Keese Albertini, "Data Mining of Meteorological-related Attributes from Smartphone Data," *INFOCOMP: Journal of Computer Science*, vol. 15, no. 2, 2016.

[41] D. Cherkassky, J. Ostrometzky, and H. Messer, "Precipitation Classification Using Measurements from Commercial Microwave Links," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2350-2356, 2014.

ROBERTA AVANZATO received a master's degree in telecommunications engineering from the University of Catania (Italy) in October 2018. From November 2018, as part of the Avionic IoT (AvIoT) project, funded by the MISE (Ministry of Economic Development), she collaborates with the University of Catania on the creation of a platform dedicated to the search for missing persons through the employment of femtocells aboard drones. From October 2019 she is a PhD student in "Systems, Energy, Computer and Telecommunications Engineering" at the University of Catania (Italy). Her research interests include signal processing, positioning algorithms, analysis of performance parameters related to the 4G radio channel for high precision rainfall estimation and monitoring.

FRANCESCO BERITELLI received a master's degree in electronic engineering and a Ph.D. degree in electronics, computer science, and telecommunications engineering from the University of Catania, Catania, Italy, in 1993 and in 1997, respectively. From 1997 to 2000, he, in collaboration with CSELT (now Telecom Italia LAB), took an active part in international ITU-T standardization meetings. From 2002, he has since been Assistant Professor in the Department of Electric, Electronic and Computer Science Engineering at the University of Catania. His main research activities are in the area of robust audio and speech signal classification and recognition, variable bit-rate speech coding, and adaptive-rate voice and dual stream transmission for mobile IP telephony applications, QoS in mobile Internet access and drone communications. His interests also include the field of cardiac biometric identification, rainfall estimation and monitoring, post-earthquake geolocation. He has 130 scientific publications, mainly in international journals, books and conference proceedings.