IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 2 August 2024, accepted 8 September 2024, date of publication 11 September 2024, date of current version 25 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3458407

RESEARCH ARTICLE

Rainfall Classification in Genoa: Machine Learning Versus Adaptive Statistical Models Using Satellite Microwave Links

CHRISTIAN GIANOGLIO[1], (Member, IEEE), SARA ZANI[2], MATTEO COLLI[2], AND DANIELE D. CAVIGLIA[1], (Life Member, IEEE)

[1]DITEN Department, University of Genoa, 16145 Genoa, Italy
[2]Artys, Darts Engineering Srl, 16121 Genoa, Italy

Corresponding author: Christian Gianoglio (christian.gianoglio@unige.it)

This work was supported by Artificial Intelligence for a Smart Rainfall System (AI4SRS) Horizon 2020 Framework Programme of the European Union, under Grant 951847.

ABSTRACT Monitoring rainfall is becoming increasingly important due to the impacts of climate change. In this context, opportunistic sensing based on satellite microwave links (SMLs) is gaining significant interest for its potential to provide real-time, low-cost, and valuable information. In this work, we deployed a network of 26 SML sensors over a survey territory corresponding to a catchment area of 140 km^2, and the data collected from it were used to train two machine learning (ML) algorithms and a statistical method to predict wet/dry conditions. Additionally, a voting mechanism based on the majority vote of the three classifiers enhances the performance of the individual methods. The results show that ML algorithms are valuable solutions for real-time monitoring of rainfall, even in an extended sensor network. The voting mechanism improves the prediction of low precipitation at the expense of a higher false positive rate.

INDEX TERMS Satellite microwave links, rainfall prediction, machine learning, smart rainfall system.

I. INTRODUCTION

The accuracy of precipitation monitoring and forecasting is of vital importance for water resource management, agriculture, urban planning, and flood prevention, in which smart city decision-makers, civil protection bodies, and infrastructure managers need timely information to undertake effective measures for risk mitigation. In addition, the lack of real-time information on the spatial distribution of rainstorms hinders environmental protection agencies from accurately predicting and responding to severe rainfall [1], [2], [3]. The development of rainfall monitoring systems capable of observing the spatial evolution of rain fields in real-time aims to enhance decision-makers ability to take informed actions in favor of the safety of people, property, and infrastructure. In this context, opportunistic monitoring systems based on microwave links are gaining increasing attention within the meteorological services community, due to their low implementation and operational costs, as well as their potential to implement extensive territorial coverage, which is enabled by the widespread availability of telecommunication infrastructures [4], [5], [6], [7], [8], [9]. These systems utilize terrestrial or satellite signals typically transmitted in the microwave bands for commercial telecommunication purposes. As a sub-category of commercial microwave links (CML), the satellite microwave downlinks (SML) typically operate in the Ku and Ka frequency bands, which experience notable attenuation during precipitation, resulting in measurable reductions in received power [10], [11], [12], [13], [14]. For example, to estimate rainfall intensity, the Smart Rainfall System (SRS) described in [14] and [15] analyzes the downlink signals of the digital television broadcast satellite (DVB-S) by measuring the RF power at the output of a universal LNB (Low Noise Block). Similarly, the NEFOCAST system, proposed in [13], exploits two-way interactive satellite terminals (called SmartLNBs) for the

The associate editor coordinating the review of this manuscript and approving it for publication was Wen-Sheng Zhao.

same purpose, and to send their measurements to a satellite hub on the return link. The NEFOCAST system was recently enhanced by updating the acquisition mode and upgrading the prediction algorithms [16].

Machine learning (ML) and signal processing algorithms have been explored in the literature to cope with instantaneous recognition of rain/non-rain conditions, as summarized below. In [10], an artificial neural network (ANN) identified rain/non-rain periods from an SML in the Ku-band, showing promising results to estimate rainfall from signal attenuation. In [17], authors measured the instantaneous bit error rate (BER) by computing the average and the standard deviation of the received signals. A logistic regression analysis based on BER measurements is used to identify wet and dry periods. In [13], two Kalman filters were used to detect precipitation events and estimate rainfall from attenuation in the downlink channel of a commercial DVB satellite signal. Reference [18] used a support vector machine (SVM) to identify rain/non-rain periods from SML signals and determined the attenuation baseline during rainy periods by adopting a long-short-term memory network (LSTM). Reference [19] trained a randomized trees classifier to assess precipitation presence and estimated the rain attenuation from the received power signal to compute the rain rate with a 1-minute time resolution. Reference [20] compared four ML algorithms with reduced computational complexity to classify rainy and non-rainy periods based on SML data. The results demonstrated the potential of using shallow ML algorithms to improve rain monitoring by SMLs. In [21], the authors proposed a two-step methodology to train ML algorithms classifying rain/non-rain events to avoid collecting a large amount of data before training the models. This methodology reduces the data amount being stored and overcomes the problem of seasonality of data that can significantly affect the prediction. In general, ML algorithms require training on large datasets to capture the diverse signal variations, often caused by spurious factors, which can impact the accuracy of rain/no-rain classification [14].

Some examples of ML algorithms for predicting wet/dry conditions from commercial microwave links can be found in [22], [23], [24], and [25] where an SVM, a convolutional neural network, and LSTMs were adopted respectively. In [26], a one-class autoencoder was proposed predicting wet/dry periods only using dry conditions acting as an anomaly detector. In general, ML techniques enhanced the classification accuracy in predicting rain/non-rain conditions from microwave link signals with respect to model-based approaches.

However, there is still a lack of research in the literature that thoroughly examines the potential benefits of reconstructing rainfall fields from an SML network across extensive territory using the machine learning methods described above. Further investigation is needed to understand the effectiveness of these approaches in large-scale applications. In addition, the performance of the proposed methods needs to be evaluated across different SML configurations, including satellite selection and signal features such as frequency and polarization. As a consequence, this study aims to experimentally evaluate the benefits of using machine learning-based algorithms for the real-time classification of rain/non-rain conditions by multiple SML sensors distributed over a wide area. This work extends the results presented in [20] and [21] where only two sensors, mounted in the same site, were considered. The testing field utilizes a network of 26 SMLs, part of the Smart Rainfall System (SRS) managed by Artys, distributed across the representative territory of the Polcevera river basin (Genoa, Italy). The advantages of the AI-based method are quantified by comparing the performance of AI techniques with the statistical algorithm currently used by SRS. In particular, two algorithms, namely a fully connected neural network with a single hidden layer and a random forest, are compared with the statistical algorithm. Furthermore, a voting mechanism, which provides the classification based on the voting majority, is adopted to enhance the performance of the individual methodologies.

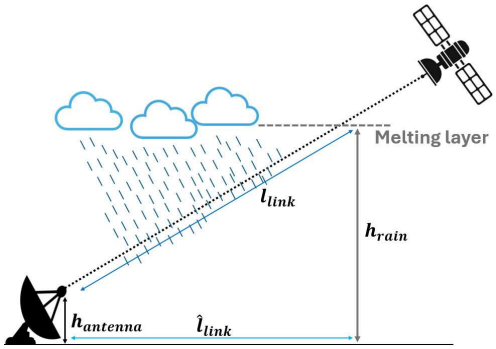

FIGURE 1. Example of a satellite microwave link, indicating the melting layer height h_{rain}, the link length l_{link}, and the projected length on the Earth surface $\hat{l}_{link}$.

The paper is organized as follows: Sec. II presents the data acquisition system and the collected dataset, Sec. III describes the classification algorithms and the training procedure, Sec. IV shows the results, and Sec. V concludes the paper.

II. DATA ACQUISITION SYSTEM

Data were collected from January 2021 to June 2023 employing 26 SRS sensors, installed in 11 sites, that received the signal power from satellites in the Ku-band. A detailed discussion of the SRS system is provided by [14]. Briefly, a SRS sensor consists of a receiving antenna located on the Earth and positioned at a height $h_{antenna}$, pointing to a satellite as shown in Fig. 1. The power of the received signal is affected by the occurrence of precipitation along the link, as explained in [14]. In Fig. 1, l_{link} is the length of the part of the path crossing the rain which extends from the receiving antenna to the altitude of the melting layer h_{rain}.

The block diagram of a sensor is depicted in Fig. 2 of [14]. The received microwave input signal is downconverted from the Ku-band to the L-band using a universal low noise block

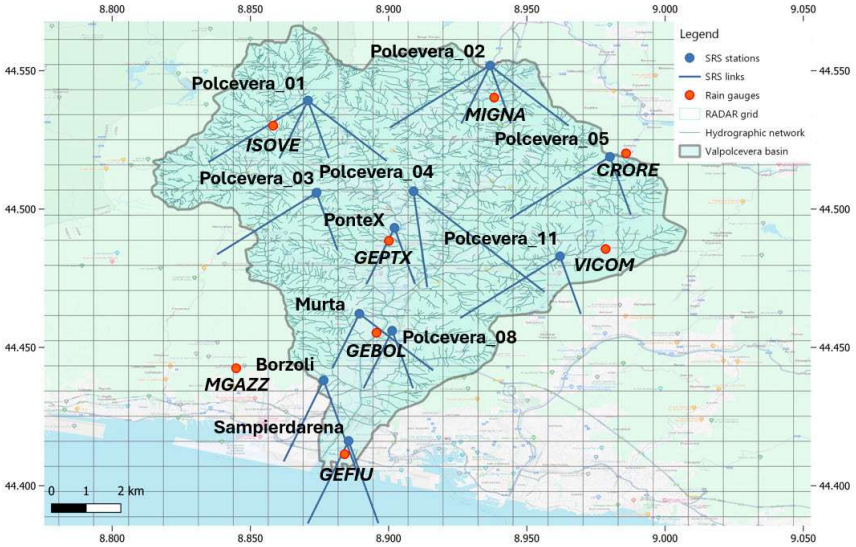

FIGURE 2. 26 SRS links in the city of Genoa used in data gathering.

(LNB) mounted on a parabolic dish. The LNB interface amplifies and filters the signal feeding an RF detector [27] which consists of a multi-stage logarithmic amplifier, operating as an RF power measurement device. It converts the RF input power, covering a range from −45 dBm to 0 dBm, into a voltage ranging from approximately 100 mV to 1.05 V. A low-cost, low-power 8-bit microcontroller unit (MCU) [28] controls the entire operation of the SRS sensor. Specifically, 64 times per minute, it performs an analog-to-digital conversion (ADC) of the voltage corresponding to the RF power level using its 10-bit ADC. Each conversion result accumulates in a 16-bit unsigned integer variable, and at the end of each minute, a UDP data packet containing this and auxiliary information is composed and sent via the WAN interface to a central server for further processing and archiving.

The field of investigation consists of 26 sensors installed at 11 different sites in the Polcevera stream basin within the city of Genoa. It extends over an area of 140 km^2 with an elevation ranging from the sea level to a maximum height of 1113 m. Table 1 presents the SRS links main information, with the site names where the sensors are installed and the satellite to which each sensor is pointed. Fig. 2 shows an example of the horizontal projection of the SRS links ($\hat{l}_{link}$ in Fig. 1 and here represented with blue lines) associated with the SRS sensors (blue circles); the red dots represent the rain gauges operated by ARPAL, the Regional Agency for the Protection of the Ligurian Environment, an instrumental body of Regione Liguria [15]. Twenty-two rainfall-significant days were selected based on data collected from rain gauges. Rainfall intensity (RI) over the area was obtained from maps collected by the regional weather radar installed on Mt. Settepani, about 60 km away from the area [15]. The radar provides a matrix of RIs for areas of one square kilometer, represented by the grey grid in Fig. 2. Table 2 shows the maximum one-hour rainfall accumulation and total rainfall accumulation for each of the 22 observed days. Values were averaged among all data acquired from rain gauges indicated by the red dots in Fig. 2. Table 3 shows the apparent positions in the sky of the satellites selected for our survey (relative to the center of the area) and the number of sensors pointed at each satellite.

TABLE 1. Information about the SRS links.

ID site	Site name	Satellites
1	Borzoli	Astra (19.2 E) Eutelsat (5 W)
2	Murta	Turksat (42 E) Eutelsat (5 W)
3	Polcevera_01_East	Turksat (42 E) Astra (19.2 E)
	Polcevera_01_West	Eutelsat (5 W) Hispasat (30 W)
4	Polcevera_02_West	Eutelsat (5 W) Hispasat (30 W)
	Polcevera_02_East	Turksat (42 E) Astra (19.2 E)
5	Polcevera_03	Astra (19.2 E) Hispasat (30 W)
6	Polcevera_04	Turksat (42 E) Hot Bird (13 E)
7	Polcevera_05	Astra (19.2 E) Hispasat (30 W)
8	Polcevera_08	Astra (19.2 E) Eutelsat (5 W)
9	Polcevera_11	Astra (19.2 E) Hispasat (30 W)
10	PonteX	Astra (19.2 E) Eutelsat (5 W)
11	Sampierdarena	Eutelsat (5 W) Astra (19.2 E)

A. DATASET

To train the algorithms for rain/rainlessness prediction, it was first necessary to build a suitable dataset by preprocessing

TABLE 2. Maximum one-hour rain accumulation (H_{max}) and total accumulation (H_{tot}) for the observed days, measured by the rain gauges.

Event Date	H_{max} [mm]	H_{tot} [mm]
22 Jan 2021	22.5	110.2
30 Jan 2021	12.4	33.1
10 Feb 2021	5.7	17.3
10 Apr 2021	10.0	59.6
11 Apr 2021	5.8	32.0
11 May 2021	7.6	47.6
14 May 2021	19.2	51.8
03 Oct 2021	34.2	121.7
13 Nov 2021	17.1	49.7
08 Dec 2021	5.0	27.7
24 Apr 2022	8.2	15.7
18 Aug 2022	8.1	30.8
02 Sep 2022	9.6	9.6
24 Sep 2022	27.0	47.1
03 Nov 2022	17.2	33.9
09 Nov 2022	3.5	12.7
15 Nov 2022	4.4	31.6
21 Nov 2022	7.6	31.7
15 Dec 2022	11.1	43.1
21 Dec 2022	5.0	36.2
13 Apr 2023	9.6	26.8
30 Jun 2023	8.7	22.3

TABLE 3. Satellite links pointed by the SRS sensors.

Satellite name	Azimuth [deg]	Elevation [deg]	N. Sensors
Astra (19.2 E)	165.5° N	37.8°	9
Eutelsat (5 W)	199.5° N	36.9°	7
Hispasat (30 W)	229.1° N	25.9°	5
Hot Bird (13 E)	174.2° N	38.6°	1
Turksat (42 E)	137.1° N	29.2°	4

the information acquired from the various sensors. Several steps and different kinds of sensors were involved in this procedure, which is reported in Procedure 1. The inputs of the procedure are the data collected by TBRG, radar, and SRS sensors. While TBRG and SRS data are available with a time resolution of 1 minute, radar data were sampled at a rate of 1 sample/5 minutes, corresponding to a total of 288 maps per observed day. Consequently, the SRS data, used for algorithm training, were reframed to fit the frequency of the radar maps. After that, for each link illustrated in Fig. 2 and described in Table 1 and for each observation during a day, the intersections of the SRS links with the radar grid were first determined to retrieve the RI along each link, which typically spans multiple radar cells. Knowing the length $\hat{l}_{link}$ of the projection of l_{link} (Fig. 1) on the Earth's surface, the lengths of the link segments within each grid cell were calculated. The total RI ($\hat{RI}$) was then calculated by multiplying the RI in each cell by the portion of the link intersecting that cell, expressed as a percentage, and summing these weighted intensities. Since the goal of this study is to recognize wet/dry intervals, thus addressing a binary classification problem, the new RIs were thresholded to map them into {0; 1} following the rule:

$$y = \begin{cases} 0 & \text{if } \hat{RI} < th_{RI} \\ 1 & \text{if } \hat{RI} \geq th_{RI} \end{cases} \tag{1}$$

where $th_{RI} = 0.1$ mm/h as in [20].

Procedure 1 Dataset Collection

Input:

- **TBRG data**: time resolution 1 minute, they provide information about the RI only in specific points and they are used to determine interesting days from the precipitations profile.
- **Radar data**: time resolution 5 minutes, they provide the ground truth of the RI for cells with an area of 1 squared km in all the monitored territory.
- **SRS data**: time resolution 1 minute, used to train the classification algorithms.

1. **Dataset creation**
   1) 22 interesting days d chosen with TBRG data
   2) SRS data resampled with a time resolution of 5 minutes to cope with radar data $\longrightarrow \boldsymbol{x}_d \in \mathbb{R}^{288\times 1}$
   3) choose a threshold on RI (th_{RI}) to determine if it is raining or not at a certain observation
   4) **for** each SRS link **do**
      for each observation x_i in a day **do**
      - Find the intersections of the link with the radar grid
      - Determine the link length in each intersected cell
      - Extract the RI for each intersected cell and multiply by a factor depending on the link portion that lays on that cell
      - Sum all the contributions of the weighted RIs $\longrightarrow \hat{RI}$
      - If the new RI value ($\hat{RI}$) is greater than th_{RI} it is raining ($y_i = 1$), otherwise not ($y_i = 0$)

      end for
   end for
2. **Output:** $\mathcal{D}_{link}$

As a result, the dataset utilized in this study can be formalized as:

$$\mathcal{D}_{link} = \Big\{(\boldsymbol{x}, \boldsymbol{y})_d,\ \boldsymbol{x} \in \mathbb{R}^{288\times 1},\ \boldsymbol{y} \in \{0; 1\}^{288\times 1};\ d = 1, \ldots, 22\} \tag{2}$$

where *link* is one of the SRS links in Table 1.

III. METHODOLOGY

A. CLASSIFICATION ALGORITHMS

Based on previous studies [20], [21], two ML algorithms, a fully connected neural network (FC) with one hidden layer, and a random forest (RF), were compared with the statistical adaptive model (ADA) described in [21] based on a thresholding mechanism. The ADA algorithm is currently used by the SRS system to predict wet/dry intervals: it directly processes raw data, whereas FC and RF require feature extraction from raw data.

Following the approach of [21], a moving window with a fixed length $l = 12$ (corresponding to 1 hour of

observations since data were reassembled over 5-minutes intervals) is applied to the SRS signals in $\mathcal{D}_{link}$. For each observation $x_i \in \mathcal{D}_{link}$, an array is created from which the features are extracted. This array, denoted by $\boldsymbol{w}_i = [x_{i-l+1}, x_{i-l+2}, \ldots, x_i]$, includes the previous $l-1$ observations as well as the current one. The extracted features from each $\boldsymbol{w}_i$ are: mean($\boldsymbol{w}_i$), std($\boldsymbol{w}_i$), skewness($\boldsymbol{w}_i$), kurtosis($\boldsymbol{w}_i$), min($\boldsymbol{w}_i$), max($\boldsymbol{w}_i$), sum($\Delta\boldsymbol{w}_i$), mean($\Delta\boldsymbol{w}_i$), std($\Delta\boldsymbol{w}_i$), min($\Delta\boldsymbol{w}_i$), max($\Delta\boldsymbol{w}_i$), where Δ indicates the difference between two consecutive measurements. The resulting dataset used for training ML algorithms can be formalized as:

$$\mathcal{T}_{link}^{ML} = \Big\{(\mathcal{X}, \boldsymbol{y})_d,\ \mathcal{X} \in \mathbb{R}^{N_w \times N_f},\ \boldsymbol{y} \in \{0;1\}^{N_w \times 1};\ d = 1, \ldots, 22\} \tag{3}$$

where $N_w = 276$ represents the number of windows because of the choice of $l = 12$, and $N_f = 11$ represents the number of features. For the sake of comparison, ADA data are also windowed with the same fixed length $l = 12$ resulting in:

$$\mathcal{T}_{link}^{ADA} = \Big\{(\mathcal{X}, \boldsymbol{y})_d,\ \mathcal{X} \in \mathbb{R}^{N_w \times l},\ \boldsymbol{y} \in \{0;1\}^{N_w \times 1};\ d = 1, \ldots, 22\} \tag{4}$$

B. TRAINING PROCEDURE

Sequentially, one out of every 22 days in $\mathcal{T}_{link}^{ML}$ was used for testing, while the remaining days constituted the training set. For the ML algorithms, given the imbalance in the data, with a higher number of non-rainy observations, an oversampling technique known as SMOTE [29] was applied to the training set to balance the number of samples in the two classes. The training set was divided into a training set and a validation set, with split percentages of 80% and 20%, respectively. The validation set was used to tune the hyperparameters of FC and RF. For the FC, the hyperparameters include the number of neurons in the hidden layer $N = \{50, 75, 100\}$ and the L2 regularizer $\lambda = \{10^i$ with $i = -4, -3, \ldots, 4\}$. For the RF, the hyperparameters chosen are the number of trees $N_{tree} = \{50, 75, 100\}$, the maximum depth of the trees $max\_depth = \{5, 10, None\}$, the minimum number of samples to make a split $min\_sam = \{2, 5, 10\}$, and the minimum number of samples to consider a node as a leaf $min\_leaf = \{1, 5, 10\}$. For ADA, adjusting hyperparameters consists of calculating thresholds on the five nonrainy days preceding the one tested, as described in [21].

IV. RESULTS

Due to the significant imbalance between wet and dry observations in the dataset, only metrics that are not influenced by this imbalance were considered in this study. The following metrics were computed to evaluate the performance of the classification algorithms:

- Specificity (Spe) $= \frac{TN}{TN+FP}$, which evaluates the effectiveness of a classifier in predicting non-rainy observations.
- Recall (Rec) $= \frac{TP}{TP+FN}$, which evaluates the effectiveness of a classifier in predicting rainy observations.
- Harmonic Mean (HM) $= \frac{2 \cdot \text{Spe} \cdot \text{Rec}}{\text{Spe}+\text{Rec}}$, which balances the two previous metrics and evaluates the overall effectiveness of a classifier in predicting both events.

Here, TP and TN represent the correctly classified wet and dry observations, respectively, while FN and FP represent the incorrectly classified wet and dry samples. All the metrics range between 0 and 1, where 0 indicates that the classifier fails all predictions and 1 represents a perfect classifier.

Based on the previous study [21], three plots were created for each algorithm to represent the results: (1) boxplots of the three metrics calculated over the 22 observed days, (2) the Rec metric computed on three intervals of $\hat{RI}$ along with the Spe and HM metrics computed on all observations, and (3) the classification results on the SRS signals. In the following, we refer to $\hat{RI}$ as RI for simplicity.

Furthermore, the results were grouped into two sets. The first set includes the results achieved by each link, as represented in Fig. 2 and described in Table 1. The second set groups the results of the satellites (Table 3) pointed by the sensors.

A. BOXPLOTS

Figures 3 and 4 represent the boxplots of the metrics (Spe in blue, Rec in black, HM in green) for three links and three satellites, respectively. Each plot reports the results for the three algorithms presented in Sec. III-A. The lower and upper parts of the boxes represent the first and third quartiles, respectively, while the horizontal line within the box indicates the median, i.e., the second quartile. The whiskers describe the variability outside the quartiles. Finally, the mean is represented by a diamond shape.

Additionally, a voting mechanism is employed over the three algorithms, where the label is assigned based on the majority of votes. This voting mechanism aims to enhance the robustness and accuracy of the classification by combining the strengths of each algorithm, potentially reducing the impact of any individual algorithm's weaknesses. The results are reported as 'Voting' in the figures. In both figures, ADA consistently achieves the highest Rec but also presents the lowest Spe, indicating that it correctly classifies most of the rain observations while having a high false-positive rate. Consequently, ADA shows the lowest HM metric. FC and RF exhibit a similar trend, with Spe and Rec showing a balanced behavior, having close median values and similar distributions. Finally, the voting mechanism yields a higher Rec compared to FC and RF but lower than ADA, and a higher Spe compared to ADA but lower than FC and RF. As a result, the Voting HM is similar to the FC and RF ones.

B. BARPLOTS

Figure 5 and Fig. 6 (a)(b)(c) present the Rec metric calculated for three intervals of RI for the three links and satellites: low-level rain ($0 < \text{RI} < 2$ mm/h), intermediate-level rain

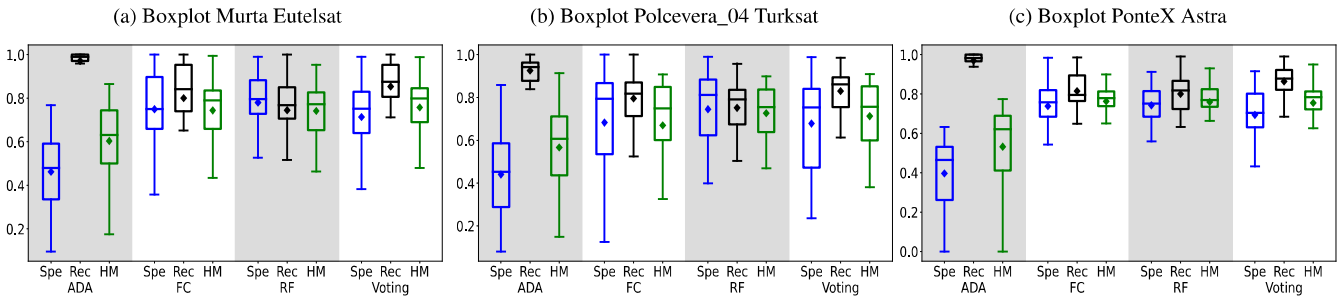

FIGURE 3. Spe (blue boxes), Rec (black boxes), and HM (green boxes) for rainy days for three different links.

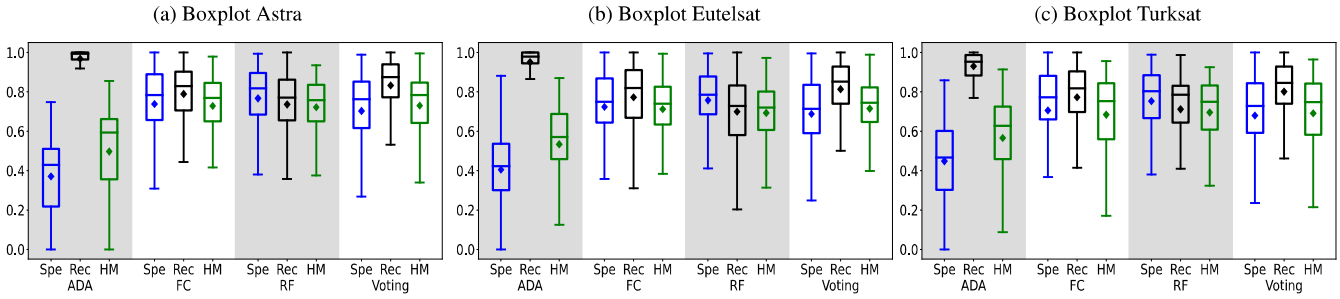

FIGURE 4. Spe (blue boxes), Rec (black boxes), and HM (green boxes) for rainy days for three satellites.

($2 \leq$ RI < 6 mm/h), and high-level rain (RI ≥ 6 mm/h). Furthermore, Fig. 5 and Fig. 6 (d)(e)(f) display the Spe and HM metrics.

In both figures, the ADA consistently outperforms all other algorithms across the three RIs. However, it also has the lowest Spe, resulting in the poorest HM performance. Interestingly, FC exceeds RF in terms of Rec for all RIs, but RF shows a higher Spe. This leads to similar HMs, with a difference of less than 2%, except in Fig. 6, where the gap is 2.6%. Finally, with Voting, there is an increase in Rec performance across all intervals and a decrease in the Spe metric. Overall, the HM metrics show a slight improvement compared to FC and RF. In conclusion, by applying the voting mechanism, the performance in predicting rain observations improves compared to using a single algorithm, albeit with a drawback in classifying non-rainy samples. Overall, it achieves the best trade-off by attaining the highest HM.

C. CLASSIFICATION

Figures 7, 8, and 9 present three days of SRS signals classified by the three algorithms and the voting mechanism for the three links. In each plot, 276 observations were classified. The x-axis represents the number of observations, whereas the y-axis shows the RI in mm/h on the left and the SRS received power in dBm on the right. The blue dots represent TN samples, the purple dots represent FP samples, the green dots represent TP samples, and the red dots represent FN samples. The cyan bars represent RI measurements higher than 0 mm/h.

Figure 7 contains 127 rainy and 149 non-rainy samples with a maximum RI of 23.9 mm/h. All the algorithms present a good trade-off in classifying wet/dry conditions. In particular, RF attains the highest Spe score, while voting presents the highest Rec, meaning that more rainy data are correctly classified if the results of all the algorithms are considered. However, Spe is 11% lower than that achieved by RF, which leads to a lower HM. Consistent with the results presented in Sec. IV-B, the rainy samples corresponding to a higher RI are well classified, while observations with a low RI have a higher misclassification rate.

Figure 8 contains 8 rainy and 268 non-rainy samples with a maximum RI of 46.6 mm/h. ADA presents the highest Rec, classifying correctly one sample more at the beginning of the rain period compared to the other algorithms. However, FC classifies all non-rainy data correctly, thus presenting the highest HM. The voting attains the same scores as RF, having the same Rec as FC but with a slightly lower Spe since 4 samples after the rain period are considered as rain samples.

Figure 9 contains 206 rainy and 70 non-rainy events with a maximum RI of 64.6 mm/h. ADA presents a very high Rec score, but it poorly classifies non-rainy data, resulting in the lowest HM score. FC has the highest Spe and HM, thus presenting the best trade-off performance in classifying wet/dry conditions. FC and RF poorly recognize rainy observations associated with low RI events. Voting mitigates this behavior, presenting the second-highest Rec

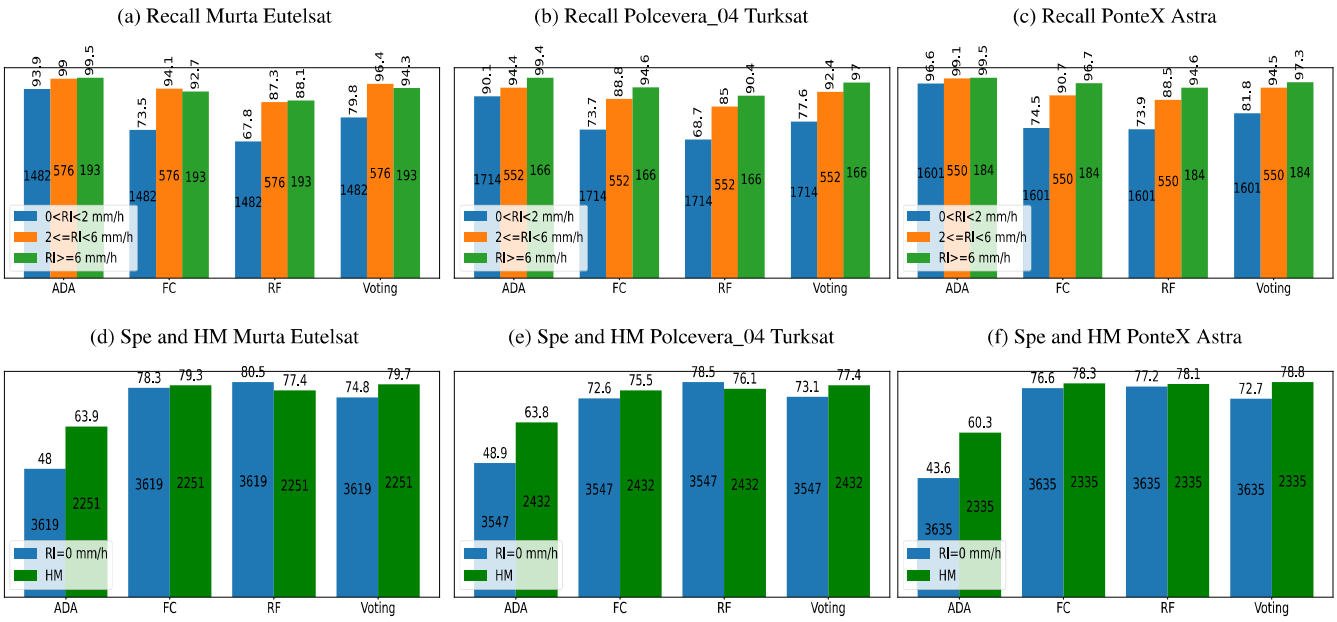

FIGURE 5. (a)(b)(c) Rec divided into three rain intensity intervals for three different links. (d)(e)(f) Spe and HM calculated for the three links.

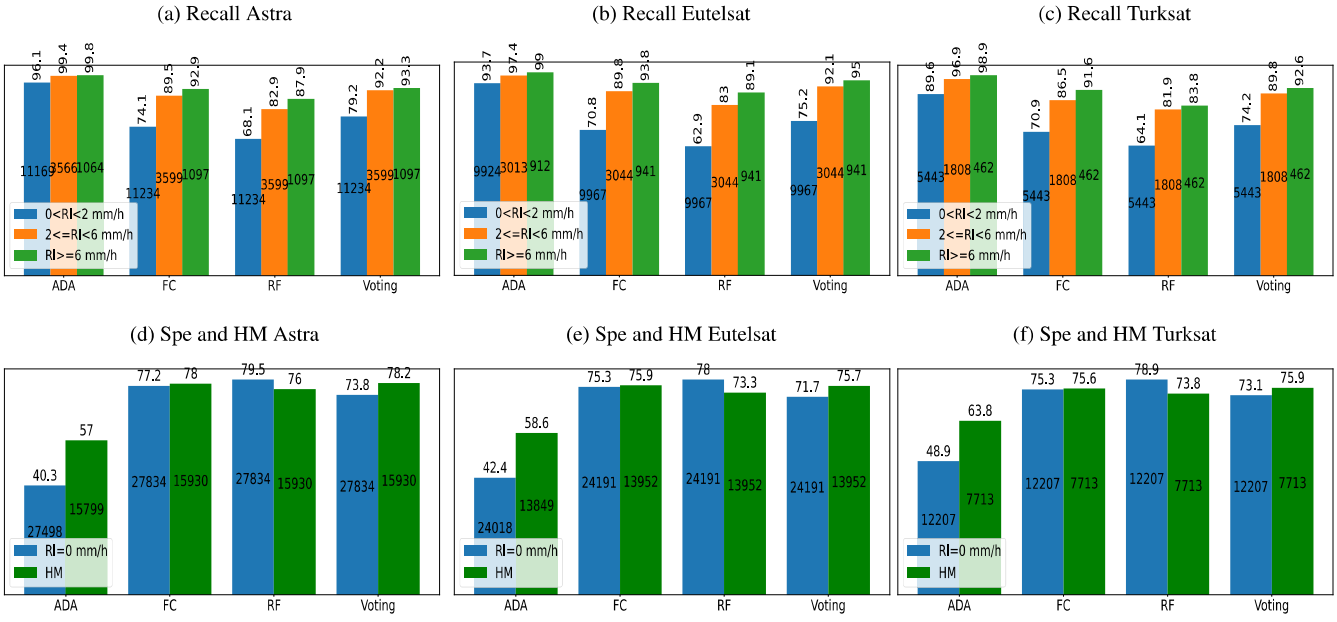

FIGURE 6. (a)(b)(c) Rec divided into three rain intensity intervals for three satellites. (d)(e)(f) Spe and HM calculated for the three satellites.

score, which is 11% higher than FC but 15% lower than ADA. Furthermore, the voting Spe is affected by the ADA results, resulting in a drop of almost 15% compared to FC.

D. DISCUSSION

In general, ML algorithms have improved the ADA method currently in use in the SRS system by providing a better trade-off in recognizing wet and dry periods. Although ADA showed a very high TP rate, it also had a high FP rate, resulting in a high Rec even for low RI values but a poor Spe. In contrast, ML algorithms achieved higher Spe and lower Rec, struggling in classifying events with low RI but resulting in a better HM. The introduction of a voting mechanism improved the Rec metric with a small decrease in Spe compared to individual ML algorithms, indicating a similar trade-off.

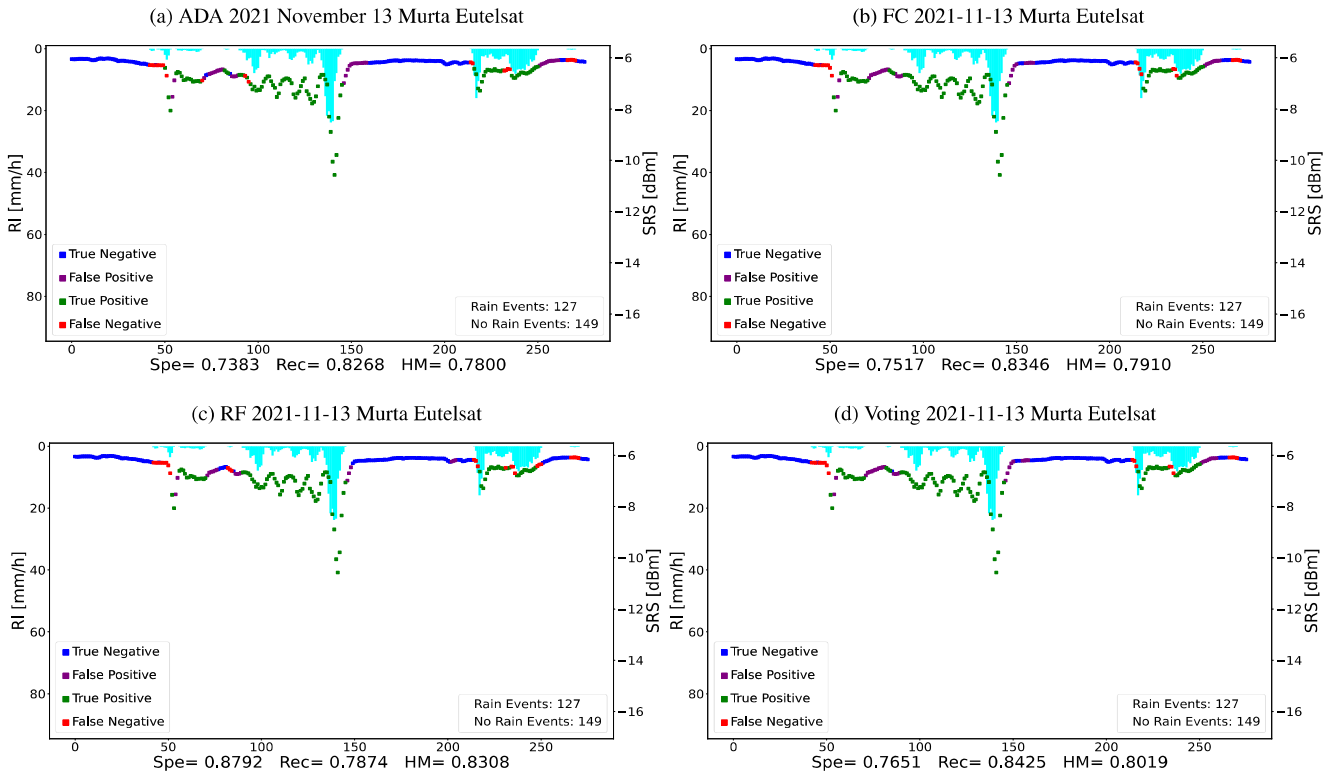

FIGURE 7. Classification results with ADA (a), FC (b), RF (c), and Voting (d) on the day 2021-11-13 Murta Eutelsat link.

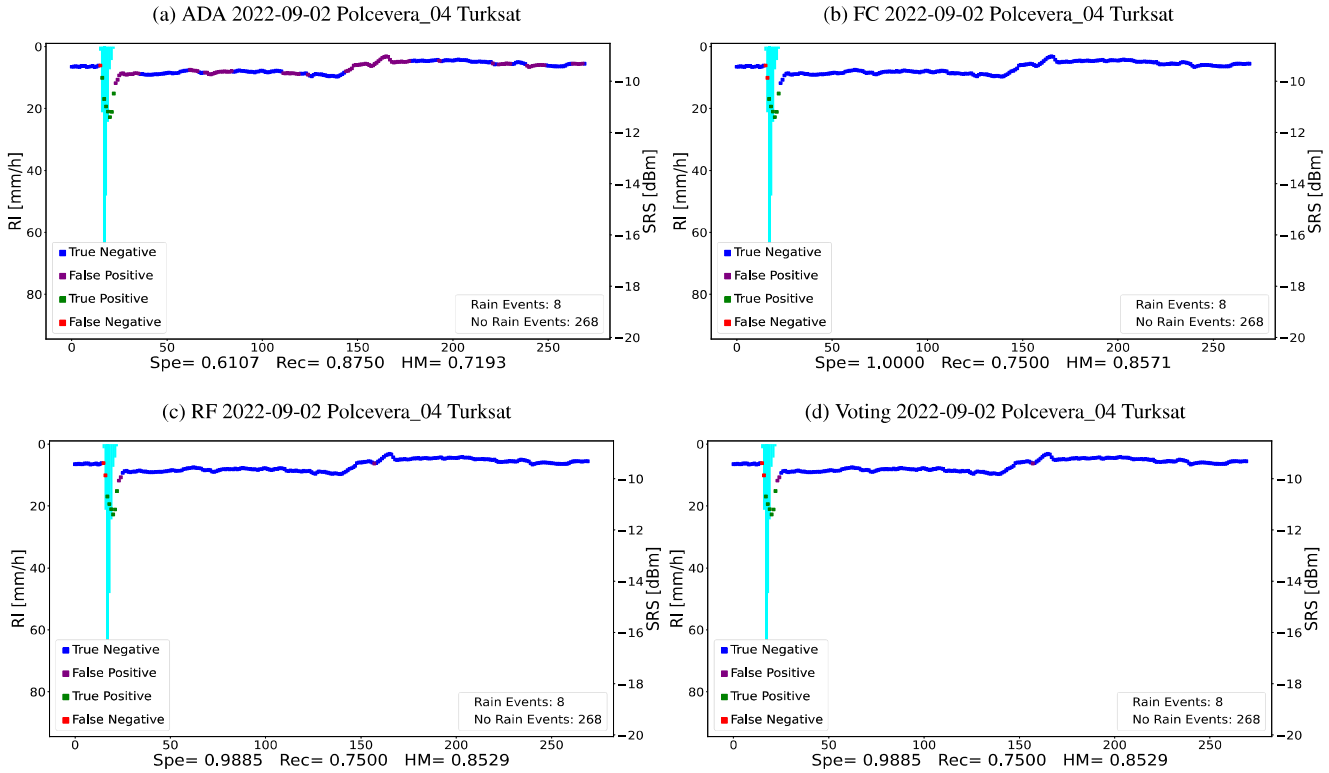

FIGURE 8. Classification results with ADA (a), FC (b), RF (c), and Voting (d) on the day 2022-09-02 Polcevera Turksat link.

In conclusion, employing a voting mechanism could improve classification accuracy, potentially by applying a weighting system. This system would assign higher importance to ML algorithm predictions when it is not raining

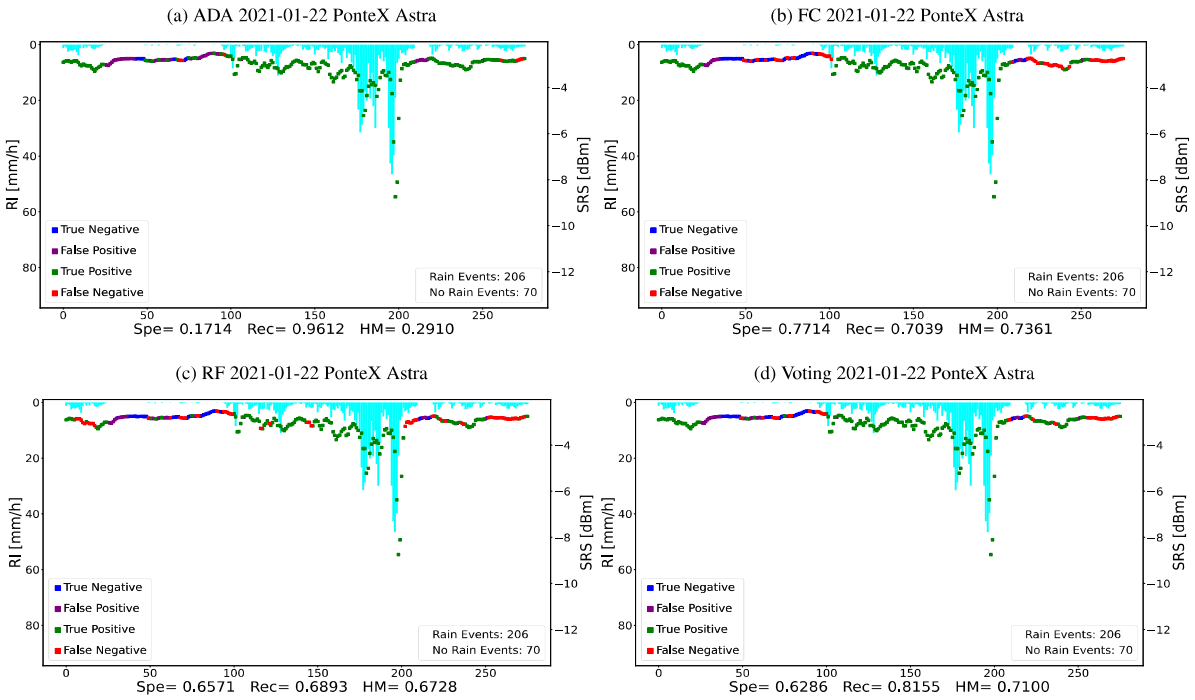

FIGURE 9. Classification results with ADA (a), FC (b), RF (c), and Voting (d) on the day 2021-01-22 PonteX Astra.

yet, thereby reducing the FP rate introduced by ADA, and giving more weight to ADA predictions once rainfall has started.

V. CONCLUSION

In this study, two machine learning algorithms were compared with an adaptive statistical model for classifying rainy and dry periods based on data derived from the processing of satellite microwave link signals. Data were acquired from a network of 26 sensors located across a river basin in Genoa, Italy, pointing to five different satellites. The study period covered twenty-two days between 2021 and 2023, selected based on data extracted from a network of rain gauges. Rainfall intensity data from a weather radar served as the ground truth to identify rain/non-rain observations for each link. For performance evaluation, a leave-one-out cross-validation approach was employed, with one out of the twenty-two days used as the test set. Performance was measured using recall, specificity, and harmonic mean. In addition, a voting mechanism was explored in which the majority vote of the three algorithms determined the prediction of wet/dry periods. The results indicated that ML algorithms provided the best trade-off in classifying rain/not-rain events but struggled with recognizing rainy observations with low rainfall intensity. In contrast, the statistical algorithm exhibited a very high recall but low specificity, indicating a high false positive rate. The voting mechanism showed a similar trade-off to the ML algorithms, with a higher recall but lower specificity. Future work will focus on improving the performance of the two metrics by incorporating a weighting mechanism into the voting process and possibly integrating other classification algorithms.

ACKNOWLEDGMENT

The authors would like to thank the Regional Agency for the Protection of the Ligurian Environment ARPAL, Italy, for providing the radar and rain gauges data.

REFERENCES

[1] B. B. Mirus, R. E. Becker, R. L. Baum, and J. B. Smith, "Integrating real-time subsurface hydrologic monitoring with empirical rainfall thresholds to improve landslide early warning," *Landslides*, vol. 15, no. 10, pp. 1909–1919, Oct. 2018, doi: 10.1007/s10346-018-0995-z.

[2] A. Fortelli, N. Scafetta, and A. Mazzarella, "Nowcasting and real-time monitoring of heavy rainfall events inducing flash-floods: An application to Phlegraean area (Central-Southern Italy)," *Natural Hazards*, vol. 97, no. 2, pp. 861–889, Jun. 2019, doi: 10.1007/s11069-019-03680-7.

[3] Y. Zhao, X. Meng, T. Qi, Y. Li, G. Chen, D. Yue, and F. Qing, "AI-based rainfall prediction model for debris flows," *Eng. Geol.*, vol. 296, Jan. 2022, Art. no. 106456, doi: 10.1016/j.enggeo.2021.106456.

[4] C. Chwala, A. Gmeiner, W. Qiu, S. Hipp, D. Nienaber, U. Siart, T. Eibert, M. Pohl, J. Seltmann, J. Fritz, and H. Kunstmann, "Precipitation observation using microwave backhaul links in the Alpine and pre-alpine region of Southern Germany," *Hydrol. Earth Syst. Sci.*, vol. 16, no. 8, pp. 2647–2661, Aug. 2012, doi: 10.5194/hess-16-2647-2012.

[5] C. Mugnai, F. Sermi, F. Cuccoli, and L. Facheris, "Rainfall estimation with a commercial tool for satellite Internet in KA band: Model evolution and results," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 890–893, doi: 10.1109/IGARSS.2015.7325908.

[6] R. Uijlenhoet, A. Overeem, and H. Leijnse, "Opportunistic remote sensing of rainfall using microwave links from cellular communication networks," *WIREs Water*, vol. 5, no. 4, Jul. 2018, Art. no. e1289, doi: 10.1002/wat2.1289.

[7] C. Chwala and H. Kunstmann, "Commercial microwave link networks for rainfall observation: Assessment of the current status and future challenges," *WIREs Water*, vol. 6, no. 2, Mar. 2019, Art. no. e1337, doi: 10.1002/wat2.1337.

[8] F. Giannetti and R. Reggiannini, "Opportunistic rain rate estimation from measurements of satellite downlink attenuation: A survey," *Sensors*, vol. 21, no. 17, p. 5872, Aug. 2021, doi: 10.3390/s21175872.

[9] F. Giannetti, V. Lottici, F. Sapienza, F. Porcú, G. Roversi, P. P. Alberoni, E. Covi, R. Nebuloni, G. Cazzaniga, C. De Michele, C. Deidda, M. Colli, S. Zani, C. Gianoglio, D. D. Caviglia, and E. Adirosi, "Opportunistic rainfall sensing: State of the art and perspectives in Italy," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. Workshops (ICASSPW)*, vol. 37, Jun. 2023, pp. 1–5, doi: 10.1109/icasspw59220.2023.10193014.

[10] L. Barthès and C. Mallet, "Rainfall measurement from the opportunistic use of an Earth–space link in the Ku band," *Atmos. Meas. Techn.*, vol. 6, no. 8, pp. 2181–2193, Aug. 2013, doi: 10.5194/amt-6-2181-2013.

[11] F. Mercier, L. Barthès, and C. Mallet, "Estimation of finescale rainfall fields using broadcast TV satellite links and a 4DVAR assimilation method," *J. Atmos. Ocean. Technol.*, vol. 32, no. 10, pp. 1709–1728, Oct. 2015, doi: 10.1175/jtech-d-14-00125.1.

[12] F. Giannetti, R. Reggiannini, M. Moretti, E. Adirosi, L. Baldini, L. Facheris, A. Antonini, S. Melani, G. Bacci, A. Petrolino, and A. Vaccaro, "Real-time rain rate evaluation via satellite downlink signal attenuation measurement," *Sensors*, vol. 17, no. 8, p. 1864, Aug. 2017, doi: 10.3390/s17081864.

[13] F. Giannetti, M. Moretti, R. Reggiannini, and A. Vaccaro, "The NEFOCAST system for detection and estimation of rainfall fields by the opportunistic use of broadcast satellite signals," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 34, no. 6, pp. 16–27, Jun. 2019, doi: 10.1109/MAES.2019.2916292.

[14] M. Colli, M. Stagnaro, A. Caridi, L. G. Lanza, A. Randazzo, M. Pastorino, D. D. Caviglia, and A. Delucchi, "A field assessment of a rain estimation system based on satellite-to-Earth microwave links," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2864–2875, May 2019, doi: 10.1109/TGRS.2018.2878338.

[15] M. Colli, F. Cassola, F. Martina, E. Trovatore, A. Delucchi, S. Maggiolo, and D. D. Caviglia, "Rainfall fields monitoring based on satellite microwave down-links and traditional techniques in the city of Genoa," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6266–6280, Sep. 2020, doi: 10.1109/TGRS.2020.2976137.

[16] S. Angeloni, E. Adirosi, F. Sapienza, F. Giannetti, F. Francini, L. Magherini, A. Valgimigli, A. Vaccaro, S. Melani, A. Antonini, and L. Baldini, "Enhanced estimation of rainfall from opportunistic microwave satellite signals," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4101312, doi: 10.1109/TGRS.2023.3349100.

[17] C. H. Arslan, K. Aydin, J. V. Urbina, and L. Dyrud, "Satellite-link attenuation measurement technique for estimating rainfall accumulation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 681–693, Feb. 2018, doi: 10.1109/TGRS.2017.2753045.

[18] M. Xian, X. Liu, M. Yin, K. Song, S. Zhao, and T. Gao, "Rainfall monitoring based on machine learning by Earth-space link in the Ku band," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3656–3668, 2020, doi: 10.1109/JSTARS.2020.3004375.

[19] R. A. Giro, L. Luini, C. G. Riva, D. Pimienta-del-Valle, and J. M. Riera Salis, "Real-time rainfall estimation using satellite signals: Development and assessment of a new procedure," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022, doi: 10.1109/TIM.2022.3165840.

[20] C. Gianoglio, A. Alyosef, M. Colli, S. Zani, and D. D. Caviglia, "Rain discrimination with machine learning classifiers for opportunistic rain detection system using satellite micro-wave links," *Sensors*, vol. 23, no. 3, p. 1202, Jan. 2023, doi: 10.3390/s23031202.

[21] C. Gianoglio, M. Colli, S. Zani, and D. D. Caviglia, "An online training procedure for rain detection models applied to satellite microwave links," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2023.3328718.

[22] K. Song, X. Liu, M. Zou, D. Zhou, H. Wu, and F. Ji, "Experimental study of detecting rainfall using microwave links: Classification of wet and dry periods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5264–5271, 2020, doi: 10.1109/JSTARS.2020.3021555.

[23] J. Polz, C. Chwala, M. Graf, and H. Kunstmann, "Rain event detection in commercial microwave link attenuation data using convolutional neural networks," *Atmos. Meas. Techn.*, vol. 13, no. 7, pp. 3835–3853, Jul. 2020, doi: 10.5194/amt-13-3835-2020.

[24] B. He, X. Liu, S. Hu, K. Song, and T. Gao, "Use of the C-band microwave link to distinguish between rainy and dry periods," *Adv. Meteorol.*, vol. 2019, pp. 1–9, Jul. 2019, doi: 10.1155/2019/3428786.

[25] H. V. Habi and H. Messer, "Wet-dry classification using LSTM and commercial microwave links," in *Proc. IEEE 10th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jul. 2018, pp. 149–153, doi: 10.1109/SAM.2018.8448679.

[26] P. Zhang, X. Liu, and K. Pu, "Classification of dry and wet periods using commercial microwave links: A one-class classification machine learning approach based on autoencoders," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 2000313, doi: 10.1109/TGRS.2023.3339682.

[27] *AD8314–100 MHz to 2.7 GHz, 45 dB RF Detector/Controller*, Analog Devices, Wilmington, MA, USA, 2018.

[28] *PIC18F14K22–20-Pin Flash Microcontrollers With XLP Technology*, Microchip Technology Inc., Chandler, AZ, USA, 2016.

[29] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

CHRISTIAN GIANOGLIO (Member, IEEE) received the master's degree in electronic engineering and the Ph.D. degree in electrical engineering from the University of Genoa, Genoa, Italy, in 2015 and 2018, respectively. He is currently a Research Fellow with the Department of Naval, Electrical, Electronic, and Telecommunications Engineering (DITEN), University of Genoa. His research interests include machine learning for resource-constrained devices and real-time applications, signal processing and machine learning for tactile sensing systems and rainfall classification, and pattern recognition for quality assessment of insulation systems in electrical apparatuses.

SARA ZANI received the Graduate degree in civil and environmental engineering and the degree in environment and protection of the territory from the University of Genoa, in 2015 and 2018, respectively. She is currently an Environmental Engineer with Artys, Darts Engineering, Genoa. Her main activities concern the analysis and processing of environmental data and numerical modeling in the field of hydrology.

MATTEO COLLI received the M.S. degree in water and soil defense engineering and the Ph.D. degree in fluid dynamics and environmental engineering processes from the University of Genoa, Genoa, Italy, in 2010 and 2014, respectively. From 2014 to 2017, he was a Research Assistant with the Department of Civil, Chemical and Environmental Engineering, University of Genoa, where he was a Research Assistant with the Department of Electrical, Electronics and Telecommunication Engineering and Naval Architecture, from 2017 to 2019. Since 2019, he has been the CTO of Artys, a business unit of Darts Engineering (Genoa).

DANIELE D. CAVIGLIA (Life Member, IEEE) is currently a Full Professor in electronics with the Department of Electrical, Electronics and Telecommunications Engineering and Naval Architecture (DITEN), University of Genoa, Italy, where he teaches the courses of electronic systems for telecommunications with the Laboratory of Electronics and Embedded Systems. His current research interests include the design of electronic circuits and systems for telecommunications, electronic equipment for health and safety, and energy harvesting techniques for Internet of Things (IoT) applications. He is also active in the development of innovative solutions for environmental monitoring and this activity led to the foundation of Artys srl as a spin-off of the University of Genoa, in 2014.

Open Access funding provided by 'Università degli Studi di Genova' within the CRUI CARE Agreement