# A Machine Learning Approach for Rainfall Estimation Integrating Heterogeneous Data Sources

Massimo Guarascio, Gianluigi Folino, Francesco Chiaravalloti, Salvatore Gabriele, Antonio Procopio, and Pietro Sabatino

*Abstract*—**Providing an accurate rainfall estimate at individual points is a challenging problem in order to mitigate risks derived from severe rainfall events, such as floods and landslides. Dense networks of sensors, named rain gauges (RGs), are typically used to obtain direct measurements of precipitation intensity in these points. These measurements are usually interpolated by using spatial interpolation methods for estimating the precipitation field over the entire area of interest. However, these methods are computationally expensive, and to improve the estimation of the variable of interest in unknown points, it is necessary to integrate further information. To overcome these issues, this work proposes a machine learning-based methodology that exploits a classifier based on ensemble methods for rainfall estimation and is able to integrate information from different remote sensing measurements. The proposed approach supplies an accurate estimate of the rainfall where RGs are not available, permits the integration of heterogeneous data sources exploiting both the high quantitative precision of RGs and the spatial pattern recognition ensured by radars and satellites, and is computationally less expensive than the interpolation methods. Experimental results, conducted on real data concerning an Italian region, Calabria, show a significant improvement in comparison with Kriging with external drift (KED), a well-recognized method in the field of rainfall estimation, both in terms of the probability of detection (0.58 versus 0.48) and mean-square error (0.11 versus 0.15).**

*Index Terms*—**Computational infrastructure, geophysical data, GIS, oceans and water, radar data.**

## I. Introduction

ACCURATE rainfall estimate is crucial for flood hazards protection, river basins management, erosion modeling, and other applications for hydrological impact modeling. To this aim, rain gauges (RGs) are used to obtain a direct measurement of intensity and duration of precipitations at individual sites.

In order to estimate rainfall events in areas not covered by RGs, interpolation methods computed on the basis of the values recorded by these RGs are used. Many variants of these methods have been proposed in the literature, and among them, the Kriging geostatistical method [1], [2] is one of the most used and recognized in the field.

An accurate spatial reconstruction of the rainfall field is a critical issue when dealing with heavy convective meteorological events. In particular, convective precipitations can produce highly localized heavy precipitation, not detected by sparse RGs, and floods can arise without a rainfall being detected [3]. To overcome this issue, a recent trend in the literature is to integrate heterogeneous rainfall data sources to obtain a more accurate estimate by using interpolation methods [4].

Unfortunately, the largely used ordinary Kriging (OK) can exploit only one source of data as input; therefore, Kriging with external drift (KED) was one of the most popular approaches adopted to overcome this limitation [5], [6]. Indeed, KED allows a random field to be interpolated, and different from the OK, it is able to take into account secondary information. The main problem is that these methods are computationally expensive and require a large number of resources to work properly.

A different approach relies on exploiting machine learning (ML) techniques. However, using these methods requires coping with different hard issues, i.e., unbalancing of the classes, a large number of missing attributes, and the need for working incrementally as soon as new data are available. Typically, ensemble methods are used to address these issues. Ensemble [7] is a classification technique, in which several models, first trained by using different classification algorithms or samples of data, are then combined to classify new unseen instances. In comparison with the case of using a single classification model, the ensemble paradigm permits handling the problem of unbalanced classes and reducing the variance and the bias of the error. Especially, ensemble-based techniques can be used to address the issues concerning the rainfall estimation and to support the monitoring of meteorological (intense) events. These methods are also able to capture nonlinear correlations (e.g., relations between sensor data, cloud properties, and rainfall estimate).

In order to address the main issues of rainfall estimation, in this article, an ML-based methodology, adopting a hierarchical probabilistic ensemble classifier (HPEC) for rainfall

Manuscript received December 23, 2019; revised April 16, 2020, May 20, 2020, July 13, 2020, and September 15, 2020; accepted October 31, 2020. This work was supported in part by RFI SpA, the Italian Railway Network, with the RAMSES (RAilway Meteorological SEcurity System) Project. *(Corresponding author: Gianluigi Folino.)*

Massimo Guarascio, Gianluigi Folino, and Pietro Sabatino are with ICAR-CNR, 87036 Rende, Italy (e-mail: gianluigi.folino@icar.cnr.it).

Francesco Chiaravalloti, Salvatore Gabriele, and Antonio Procopio are with IRPI-CNR, 87036 Rende, Italy.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TGRS.2020.3037776.

Digital Object Identifier 10.1109/TGRS.2020.3037776

estimation, is introduced. The proposed approach, by integrating data coming from different sources (i.e., RGs, radars, and satellites) and exploiting an undersampling technique for handling the unbalanced classes problem typical of this scenario, permits accurate estimation of the rainfall where RGs are not available.

Our approach is an effective solution for real scenarios, as in the case of an officer of the Department of Civil Protection (DCP), who has to analyze the rainfall in a specific zone presenting risks of landslides or floods. The experimental evaluation is conducted on real data concerning Calabria, a region located in the South of Italy, and provided by the DCP. Calabria is an effective test ground because of its strong climate variability and its complex orography.

Our contributions can be summarized as follows.

1) Three heterogeneous data sources (i.e., RGs, radar, and Meteosat) are integrated to generate more accurate estimates of rainfall events.
2) Different classification methods are compared on a real case concerning Calabria, a southern region in Italy, and a hierarchical probabilistic ensemble approach is proposed.
3) Different ML-based methods, pretrained only on historical data, with a widely used interpolation method in the hydrological field (i.e., KED) are compared.

The rest of this article is organized as follows. In Section II, some related works are analyzed, and the main differences with our approach are noted. Section III illustrates the case study and describes the main sources of data used by the framework. In Section IV, the methodology used to estimate the rainfall is specified. Section V is devoted to some experimental results and discussion. Finally, Section VI concludes this article and shows some interesting future developments.

## II. Related Works

The term "weather nowcasting" refers to weather forecasts concerning the near-future, typically a few hours, by using data coming from radars, satellites, and other sources, and usually, it is adopted to prevent many risks, such as landslides and floods. This field of research shares similar techniques and data sets with the task of rainfall estimation; therefore, we decided to analyze also some works concerning this field of research in the first part of this section.

Schroeter [8], similar to our work, also integrates input data coming from radars, RGs, and satellites and uses artificial neural networks (ANNs) as a forecasting method. Their experiments are conducted on real data concerning Australia and the integration of the data is necessary because, in this country, radar coverage is not optimal, particularly in regional areas. The experimental results show that their method overestimates the rainfall. In [9], an approach based on neural networks that keep track of spatiotemporal relationships is proposed to cope with the problem of rainfall nowcasting. However, only data provided by radars are considered. In [10], a hybrid approach based on recurrent neural networks and support vector machine (SVM) is used to provide rainfall forecasts from typical meteorological parameters, such as humidity, pressure, and temperature. An interesting review of the field of rainfall prediction can be found in [11].

Other works based on the ensemble paradigm include the work in [12], which, similar to our work, employs a probabilistic ensemble and merges two sources of data (i.e., rain gauges and radar) even if the aim of this work is to develop a run-off analysis. Afterward, a blending technique is applied to the results of the runoff hydrologic models to determine a single runoff hydrograph. Experimental results show that the hydrologic models are accurate and can help to make more effective decisions in the flood warning. Frei and Isotta [13] define a technique for deriving a probabilistic spatial analysis of daily precipitation from rain gauges. The final model represents an ensemble of possible fields, conditional on the observations, which can be explained as a Bayesian predictive distribution measuring the uncertainty due to the data sampling from the station network. An evaluation of a real case study, located in the European Alps, proves the capability of the approach in providing accurate predictions for a hydrological partitioning of the region. The work in [14] proposes an interesting study of the daily precipitations for Australia and several regions of South and East Asia, based only on high-resolution gauges. Basically, the adopted model can be figured out as a mean of the analyses generated for each source. The authors highlight how the ensemble approach outperforms the single members composing the model in terms of global accuracy. Moreover, the proposed model is also able to capture additional information from different precipitation products. Both the last two works exploit an ensemble scheme to provide more accurate predictions, proving the capability of ensemble methods to ensure good results also in a rainfall estimate scenario. However, different from our work, the adopted combination strategies are quite simple, and a combination of heterogeneous data sources is not considered.

The rest of this section is devoted to the analysis of works specifically designed for rainfall estimation. An extensive survey on these types of work can be found in [15]. For the same Italian region studied in our work, Calabria, Chiaravalloti *et al.* [16] studied the performance of three recently developed satellite-based products, i.e., IMERG, SM2RASC, and a clever combination of SM2RASC and IMERG using as a benchmark both RG only data and the integrated RG-radar product. Experiments permit to establish that IMERG has good performance at time resolutions higher than 6 h, and the combination of IMERG and SM2RASC obtains a higher quality satellite rainfall product. Most of the other approaches integrate data from different sources, i.e., satellite channels and radars. Some of them are based on the identification of suitable models that exploit the relation between optical and microphysical properties of clouds and use the data to find the appropriate parameters for these models [17], [18]. Other works individuate the models by using statistical techniques [19]–[21]. For instance, Bayesian estimation is used in [22] in order to provide precipitation estimations based on satellite multispectral data; reference estimates are provided by methods that use radar data as input. Verdin *et al.* [23] also adopt Bayesian estimation in order to estimate the parameters of the model; their system integrates RG observations and satellite data and adopts an interpolation technique based on the Kriging method.

All these techniques are able to provide interesting results, but they require a rather delicate phase of parameters estimation of the particular model; therefore, as a side effect, usually, their flexibility and effectiveness tend to be hampered.

As the relations between sensors data, cloud properties, and rainfall estimates are highly nonlinear, more flexible approaches based on ML techniques have been investigated recently. For instance, the problem of detecting convective events and closely related rainy areas is addressed in [24] by using ANNs combined with support vector machines. Data sets are obtained by processing data coming from optical channels of the multispectral instrument onboard of Meteosat Second Generation (MSG) satellites; different from our work, RG measures are used only as a reference but not in the training phase of the algorithm. Sehad *et al.* [25] propose an approach to rainfall estimation based on SVMs; the input data are integrated from multispectral channels on MSG; and two models are developed for daytime and nighttime respectively. Results are compared to similar approaches based on ANNs, and random forest (RF) and RGs are used only to validate the approach. Another approach based on ANNs is described in [26]; in this work, given as an image matrix, radar data are used as reference in detecting rainy pixels. Kuhnlein *et al.* [27] also adopt the ensemble paradigm and, in particular, employ RFs to infer rainfall rates from data coming from multispectral channels on MSG satellites.

In [28], a hybrid architecture that combines support vector regression (SVR) models, genetic programming, and recurrent networks is presented. The experimental results, performed on a real data set concerning the typhoon season in northern Taiwan, exhibit good performance in estimating the height of the rainfall. Backpropagation neural networks are used in [29] to combine data coming from multispectral channels on MSG satellites and parameters usually considered in numerical weather predictions. Rainfall rates estimated from the radars are used in the training phase of the neural network. Deep neural networks are used to improve rainfall rates estimation using only radar data in [30] and [31].

Finally, in [32], different data mining techniques are compared on the problem of rainfall estimation from multispectral satellite data. Techniques considered include RF, neural networks, and support vector machines. Experiments show that no particular algorithm performed dramatically better than the others; the authors conclude that further research is necessary to investigate whether different base learners could improve the results.

Our approach is also similar to some of the analyzed works on the ensemble paradigm; however, we do not rely on a single data source, as do most of the works present in the literature, but our approach integrates information coming from different sources, i.e., radars, satellites, and RGs. In addition, as the classes in this problem are intrinsically unbalanced, we also perform an undersampling during the training phase, which it is not used in the other works.

## III. Case Study: Rainfall Estimation in Calabria

In this section, the peculiarities of the Calabria region and the main characteristics of the three sources of data integrated to estimate the rainfall events are described.

### A. Data Description

Calabria covers an area of 15 000 km$^2$. This region exhibits high climatic variability [33], mainly due to its particular orography (i.e., the proximity of mountains and seas) and to the influence of the Mediterranean Sea. Calabria is characterized by the presence of small and very small basins that may be prone to landslides and flash floods during localized rainfall [34], [35]. Thus, an accurate spatial rainfall field is necessary for hydrological impact studies and for a correct analysis of hydrological scenarios producing floods and landslides affecting this fragile territory [36]. For these reasons, RG, weather radar, and MSG data collected during 2016 for Calabria will be used to test the framework developed in this work.

In particular, the Calabrian DCP supplies **RG data**, extracted from a real-time monitoring system consisting of a network of 156 telemetered sensors with an average distance of about 10 km. These data consist of precipitation heights measured in mm sampled every minute.

Radar data are supplied by the DCP, by exploiting the Italian meteorological radar network (20 weather radars spread over the whole Italian territory). Data are collected in near real time and processed by merging high-resolution volumetric data using the methodology described in [37]. The final data have a time resolution of 10 min over a grid with a spatial resolution of 1 km and include Constant Altitude Plan Position Indicator (CAPPI) that gives a horizontal cross section of reflectivity at 2000, 3000, and 5000 m above sea level (a.s.l.); vertical maximum intensity (VMI) that represents the maximum reflectivity value present on every point's vertical axis; and the surface rain intensity (SRI) that estimates the ground rain rate. The SRI product is obtained applying the Marshall–Palmer reflectivity-rainfall (Z-R) relationship to the lowest radar beam [38]. The study area is covered by two C-band weather radars, and their range is about 200 km (see Fig. 1).

The Meteosat Second Generation (MSG) data are provided by the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT). The Spinning Enhanced Visible and Infrared Imager (SEVIRI), onboard the geostationary MSG satellite, supplies detailed images of the full Earth disk at 12 different wavelengths and monitors the dynamic evolution of cloud structures at a high temporal resolution due to the rapid scan service (RSS), which provides an image every 5 min [39]. These characteristics allow the detection of rapidly developing high impact weather. Moreover, data from different channels correspond to different physical properties of the observed cloudy structures. For example, in the thermal infrared band (10–12 $\mu$m), satellites supply indications on

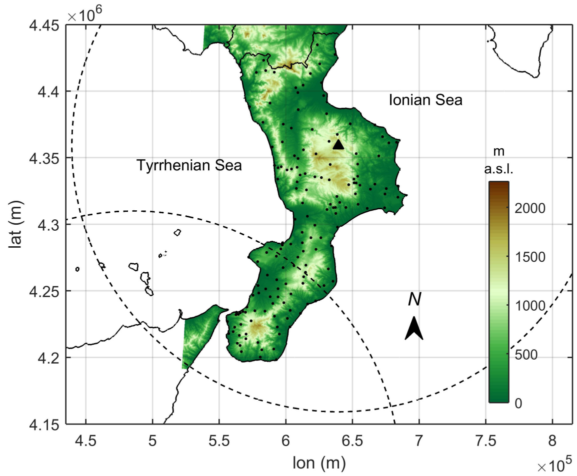



Fig. 1. Map of Calabria, projection system UTM WGS84, zone 33N. The points represent the RG network, while the large dashed circles represent the radar ranges (the location of the Calabrian radar is indicated by the solid triangle).

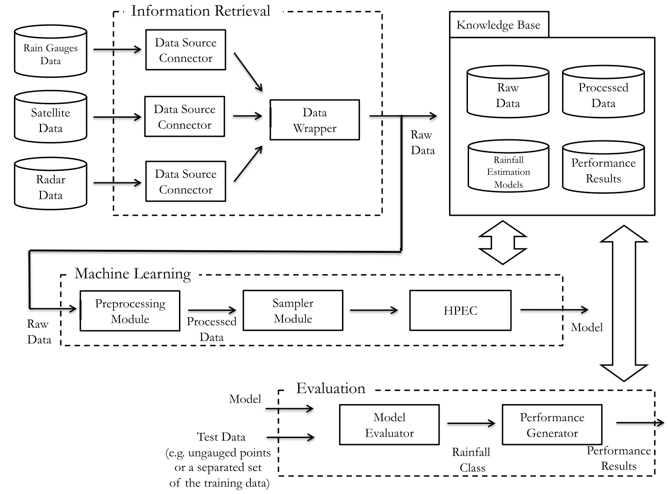



Fig. 2. Overall system diagram of the learning and evaluation processes.

the temperature of the cloudy area, which is related to the cloud height and to the convection development. The spatial resolution for the region considered in this work is about 4 km.

## IV. Methodology

This section describes the methodology proposed to estimate the rainfall in a specific zone exploiting an ML-based approach.

In Fig. 2, the overall learning and evaluation process, composed of three main phases (i.e., information retrieval, ML, and evaluation), is sketched. Information retrieval focuses on gathering the data required for the analysis. In this phase, different types of information, from several data sources, are extracted and integrated. Especially, a data source connector is used to establish the connection with a specific data source; then, the gathered data are provided as input for the data wrapper that combines these data in a single view suitable for the next analysis phase. Raw data are stored in the knowledge base (KB), which is exploited for the data exchange among the framework components.

In the ML phase, raw data are preprocessed to make them suitable for the analysis, and an undersampling strategy is adopted to address the class unbalanced problem. Then, a rainfall estimation model can be trained from the preprocessed data. To this aim, in this work, we introduce a novel metaensemble probabilistic classifier (named HPEC in the figure) and described in Section IV-C. However, the architecture is modular, and also another different estimation model could also be adopted.

In the final phase, i.e., evaluation, the rainfall estimations are computed for a set of preprocessed data by exploiting the trained model. This phase can be performed on both: 1) ungauged points (in a real usage scenario) in order to estimate the severity class of these points and 2) a separated set of training data (i.e., not used in the learning phase). For this latter case, how specified in the experimental section, we know the real class (as an RG is present); however, this value is removed in order to obtain the estimated class, and it is used only to compute the performance measures also reported in the experiments.

In the following, the preprocessing phase (the way in which the information extracted by several data sources is combined) and the proposed ensemble-based approach are detailed.

### A. Data Preprocessing

One of the main novelties of the approach is the integration of three main sources of data to better estimate rainfall events. However, the integration needs to overcome the issues concerning the heterogeneity of the data, so some preprocessing operations have to be performed to adjust the different scales both in terms of time and space.

A 30-min temporal sampling of the measurements was deemed appropriate for the purposes of this article, considering an acceptable balance between the need for a high temporal resolution and the need to reduce the statistical noise of the data (as pointed out in [40]); therefore, the RGs and the SRI data were cumulated every 30 min. Furthermore, as, in many cases, near points exhibit related rain conditions, it is useful to take into account the distance among the RGs. Thus, for every point, the coordinates of the four nearest points containing an RG, together with the distance, are considered.

To summarize, for every point in which an RG is present, for a 30-min time step, 35 features were extracted (as also illustrated in Table I), as detailed in the following.

1) *From the Radars (5 Features):* SRI, cumulated over 30 min, VMI, and three values of CAPPI (horizontal cross section of reflectivity at 2000, 3000, and 5000 m a.s.l.); the last value was taken for all these four features.
2) *From the MSG Satellites (11 Features):* The last value of the 11 Meteosat channels (excluding the high-resolution visible channel number 12).
3) *From the RGs (16 Features):* The coordinates of the four nearest points containing an RG, together with the distance and the value measured by the RG.

TABLE I

FEATURES EXTRACTED FROM THE THREE SOURCES OF DATA: RADAR, SATELLITE, AND RGS ($i = 1, \ldots, 4$)

| Source | Name | Time res. | Space res. | Unit of meas. | Description |
|---|---|---|---|---|---|
| Radar | SRI | 10 min. | 1 km | mm/h | Surface rainfall intensity |
| | VMI | 10 min. | 1 km | dBZ | Maximum reflectivity on the vertical |
| | CAPPI2000 | 10 min. | 1 km | dBZ | Reflectivity at the heights of 2000 m |
| | CAPPI3000 | 10 min. | 1 km | dBZ | Reflectivity at the heights of 3000 m |
| | CAPPI5000 | 10 min. | 1 km | dBZ | Reflectivity at the heights of 5000 m |
| MSG | Ch1 | 5 min. | 4 km | K | $0.635\mu m$ channel brightness temperature |
| | Ch2 | 5 min. | 4 km | K | $0.81\mu m$ channel brightness temperature |
| | Ch3 | 5 min. | 4 km | K | $1.64\mu m$ channel brightness temperature |
| | Ch4 | 5 min. | 4 km | K | $3.9\mu m$ channel brightness temperature |
| | Ch5 | 5 min. | 4 km | K | $6.25\mu m$ channel brightness temperature |
| | Ch6 | 5 min. | 4 km | K | $7.35\mu m$ channel brightness temperature |
| | Ch7 | 5 min. | 4 km | K | $8.7\mu m$ channel brightness temperature |
| | Ch8 | 5 min. | 4 km | K | $9.66\mu m$ channel brightness temperature |
| | Ch9 | 5 min. | 4 km | K | $10.8\mu m$ channel brightness temperature |
| | Ch10 | 5 min. | 4 km | K | $12\mu m$ channel brightness temperature |
| | Ch11 | 5 min. | 4 km | K | $13.4\mu m$ channel brightness temperature |
| Rain gauge | $x_i$ | - | punctual | m | $x$ coordinate of the $i^{th}$ rain gauge |
| | $y_i$ | - | punctual | m | $y$ coordinate of the $i^{th}$ rain gauge |
| | $dist_i$ | - | continuous | m | distance from the $i^{th}$ rain gauge |
| | $rainfall_i$ | 1 min | punctual | mm | mm of precipitation |

4) *Other Data (Three Features):* Longitude and latitude of the point and the month in which the data are detected.

It is important to remark that different spatial resolutions have been used in the data fusion process (respectively, 1 km for the radar and 4 km for the MSG satellite); in practice, the features, respectively coming from of the resolution cell (pixel) of the satellite or of the radar, are determined by the point in which the RG falls.

The final data set covers a period concerning the second half of 2016 in Calabria. After the phase of undersampling and preprocessing, it consists of 117 600 tuples and 35 features. Each tuple represents an observation for a predetermined period of time of 30 min, in a point of space corresponding to one of the 156 RGs in Calabria. The class to be estimated is measured by the RGs, and consider the mm of rain fell in the considered range of 30 min and is discretized into five classes according to these ranges: [$0-0.5$, $0.5-2.5$, $2.5-7.5$, $7.5-15$, $15-$]. The numbers of tuples for each class are, respectively, 94 080, 18 514, 4221, 654, and 131. The correct classification of the latter two classes is particularly important because they represent heavy rainfall events, which must be handled adequately. However, the number of tuples belonging to these two classes is really unbalanced as the ratio between each of them and the majority class ($0-0.5$, named in the following no-rain for the sake of simplicity) is very low, and therefore, correctly classifying these two rare classes of events is a really challenging problem.

### B. Undersampling Strategy

After the integration of the three data sources, i.e., RG data, weather radar, and MSG data, a sampling strategy must be adopted to avoid the class unbalance issue. In the class unbalance problem [41], the minority classes are overcome by the numerousness of the majority class. In the literature, different strategies are proposed to tackle this problem; in particular, the techniques based on undersampling obtain good results in terms of accuracy, while oversampling the other classes usually causes overfitting. Therefore, we adopt an undersampling strategy [42] operating only on the no-rain class, while the other four classes, representing different levels of rain, are not affected by the method. A random uniform undersampling strategy is adopted, which operates on two levels: first, only temporal periods presenting all the points without rain events are chosen; then, always uniformly, a random number of spatial points are removed from this period. In order to avoid loss of information, nothing is done in days presenting rain and no-rain points.

### C. Ensemble-Based Approach

In our framework, the rainfall estimation problem is addressed as a classification task. A classifier permits to divide the data into predefined categories (also called classes). Frequently, classifiers are also used to predict unknown categories for new unseen instances.

Formally, let $S = \{(x_i, y_i)|i = 1, \ldots, N\}$ be a training set where $x_i$, called example or tuple or instance, is an attribute vector with $m$ attributes and $y_i$ is the class label associated with $x_i$, where $x_i \in \{L_1, \ldots, L_C\}$ and $C$ is the number of classes. A classifier, given a new example, has the task to assign the class label for it, i.e., the task consists of computing a function $h(x_i)$ that is able to estimate $y_i^*$ (with $y_i^* \approx y_i$).

Ensemble [7], [43] is a learning paradigm where multiple base learners are trained for the same task by a learning algorithm, and the classifications of the component learners are combined for dealing with new unseen instances. In our framework, a set of classifiers are adopted as base learners.

Formally, as shown in Fig. 3, ensemble techniques build $S$ classifiers $T_1, \ldots, T_S$, each one trained on a different sampling of the training set (or also on differently weighted tuples of

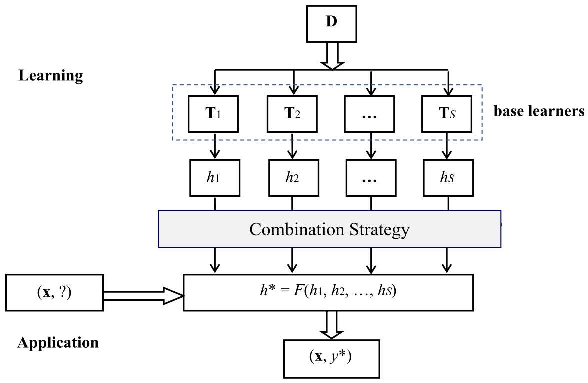



Fig. 3. Ensemble learning scheme.

the same training set or adopting different learning algorithms always on the same training set); then, they are combined according to a suitable combination strategy to classify the test set. Basically, given as input the classifications of the base classifiers $T_1, \ldots, T_s$, ensemble learning techniques allow combining them and computing a function $h^*$ that is able to produce the final classification.

In the literature, three standard combination strategies have been widely used: boosting, Bootstrap aggregation (bagging), and stacking.

Boosting was introduced by Schapire [44] and Freund [45] to improve the performance of any "weak" learning algorithm, i.e., an algorithm that "generates classifiers that need only be a little bit better than random guessing" [45]. The boosting algorithm, called AdaBoost, adaptively changes the distribution of the training set according to how difficult each example is to classify.

Basically, in bagging-based approaches [7], different subsets of the training data (extracted via a bootstrap resampling strategy) are used to train the base learners, while the final assignment is computed via a simple voting scheme.

Stacking [46] (stacked generalization) is an effective ensemble-based approach able to exploit the output of several classifiers to learn a more accurate metalearner. The whole training set is used to learn each single base model. Basically, the algorithm uses the trained models to build a stacked view, i.e., a table, in which each row corresponds to the classification of each base model for the corresponding tuple of the initial training data set; therefore, each column contains the assignment of each classifier. Finally, this view is used to train a metalearner in order to compute the final classification. Alternatively, the class probabilities of the same classifiers can be directly used to train the metalearner. In our approach, we adopt this latter strategy.

RF [47] is an efficient and particularly accurate ensemble-based technique. This algorithm combines simultaneously two different strategies: a bagging algorithm and a random selection of the features. Especially, the bagging approach is applied to a set of tree-based classifiers (base models). The main difference with respect to a simple bagging method relies on the building procedure of the decision trees, which are trained and evaluated only on a randomly chosen subset of the features.

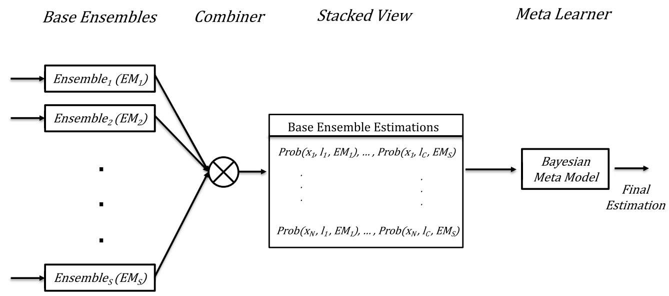



Fig. 4. HPEC model.

In our methodology, after the preprocessing and undersampling steps, an HPEC is adopted to learn a model able to estimate the rainfall events. The source code of HPEC, used to run the experimental results shown in Section V, is available at https://github.com/massimo-guarascio/ml-rainfall.

Although the idea of combining different learners in a hierarchical way is not new in itself, to the best of our knowledge, our specific hierarchical model has not been previously used to tackle the rainfall estimation problem.

Usually, when the number of examples for the minority classes is low in comparison with the majority class, stacking-based techniques behave better, as they avoid the overfitting problem typical in unbalanced data sets. Therefore, we adopted a stacking method for HPEC, which follows the architecture reported in Fig. 4. As a base learner, in the first stage, we adopt RFs (note that an RF is also an ensemble), as this technique exhibits good performance in unbalanced scenarios, reducing the problem of the overfitting, as different subsets of features are exploited in the training phase.

Especially, our approach exploits a two-level ensemble classifier: 1) in the first level, a set of ensembles (i.e., the RF classifiers) are trained on the same data set, but initialized with different random seeds, in order to vary the feature subset generation during the learning of the decision trees composing the forest and 2) in the second level, a probabilistic metalearner is used to combine the estimates provided by the first level classifiers according to a stacking schema [46]. In more detail, their estimates are combined in a stacked view, in which each row $i$ contains an estimate (i.e., a vector composed of the class probabilities, provided by each base model) for the class of the tuple $i$ of the training data set. This view is provided as input for a probabilistic Bayesian model. Finally, the output of the metalearner is the estimate of the class of the rainfall event, obtained on the basis of the probabilities provided by the base models.

## V. Experimental Results

In order to assess the quality of our approach in estimating rainfall, different experiments on the real data concerning Calabria are conducted. First, different ensemble-based techniques (RF, boosting, and HPEC) are compared. In addition, these ensemble algorithms are also compared with the decision tree algorithm and with the support vector regression model, which has been successfully used in the field of rainfall forecasting

[28]. Then, the ensemble approach is compared with the baseline method (Kriging with external drift), largely used and well-recognized in the field of rainfall estimation. Finally, the different contributions of neighboring RGs, satellites, and radar measures are studied.

The ensemble-based algorithms, the decision tree, and the SVR model, adopted in this article, are based on the well-known scikit-learn ML implementation.[1] If not differently specified, the algorithms were run using standard parameters. No tuning of the parameters was conducted. As for the KED method, we used the RG data as the primary variable and the radar data as the auxiliary information. The KED interpolation is performed by using the "autokrige" method described in [48] and implemented in the Automap and Gstat libraries of the statistical software R. Using this method, the theoretical variogram is automatically determined by picking the best fitting model. This procedure results in fully automated, and it allows, in near real-time, the estimation of the areal rainfall field over a $1 \times 1$ km$^2$ spatial grid, with a time resolution of 10 min.

All the algorithms were trained on a training set composed of $2/3$ of the original data set and then evaluated on the test set, consisting of $1/3$ of the original data set. In practice, for each period of time, $2/3$ of the observation points (RGs) were randomly chosen from the original data set, and they form the training set, while the remaining points will constitute the test set (and are excluded also for the interpolation method of Kriging). The original data are continuous; however, our classification approach needs discretized classes to work, and in addition, we are interested in retrieving the fine spatial structure of the precipitation field in order to correctly interpret the level of hazard for a given area (and not in the estimating the high-resolution value of rainfall). Therefore, the HPEC classification process is conducted after discretization of the original data into five classes, as described in Section IV-A. On the contrary, the Kriging method works with the continuous original data, and only at the end of the interpolation process, the final result is approximated to the class, whose range contains the obtained continuous value.

All the experiments were averaged over 30 runs. We adopt metrics taking into account the rarity of the minority classes, in order to avoid the problem of overestimating the accuracy of a classifier to recognize the instances of these classes correctly. Notably, true positive (TP), false positive (FP), true negative (TN), and false negative (FN) can be easily computed from the confusion matrix.

Therefore, in addition to the well-known metrics of Precision and Recall (in the field of rainfall estimation, usually named POD, i.e., probability of detection), computed, respectively, as $\text{Precision} = (\text{TP}/\text{FP} + \text{TP})$ and $\text{Recall} = (\text{TP}/\text{FN} + \text{TP})$, we adopted some measures, deriving from these metrics, largely used in the field of rainfall estimation [49], i.e., the false-alarm ratio (FAR), computed as $\text{FAR} = (FP/(\text{TP} + \text{FP}))$, the critical success index (CSI), computed as $\text{CSI} = (\text{TP}/(\text{TP}+\text{FP}+\text{FN}))$, and the mean square error (MSE), computed as $\text{MSE} = (1/n) \sum_{i=1}^{n} (y^{(i)} - \tilde{y}^{(i)})^2$,

[1]http://scikit-learn.org/stable/

TABLE II

CSI, FAR, POD, AND MSE FOR THE SVR, DECISION TREE, BOOSTING, RF, AND HPEC

| Algorithm | CSI | FAR | POD | MSE |
|---|---|---|---|---|
| **SVR** | $0.37 \pm 0.010$ | $0.40 \pm 0.027$ | $0.43 \pm 0.011$ | $0.11 \pm 0.002$ |
| **Decision Tree** | $0.41 \pm 0.009$ | $0.38 \pm 0.033$ | $0.47 \pm 0.011$ | $0.10 \pm 0.002$ |
| **Boosting** | $0.43 \pm 0.008$ | $0.33 \pm 0.026$ | $0.49 \pm 0.008$ | $0.09 \pm 0.002$ |
| **Random Forest** | $0.43 \pm 0.010$ | $0.31 \pm 0.024$ | $0.49 \pm 0.011$ | $0.09 \pm 0.002$ |
| **HPEC** | $0.44 \pm 0.011$ | $0.46 \pm 0.016$ | $0.58 \pm 0.016$ | $0.11 \pm 0.002$ |

TABLE III

PRECISION, RECALL, AND F-MEASURE FOR THE SVR, DECISION TREE, BOOSTING, RF, AND HPEC FOR THE MINORITY CLASSES 4 AND 5

| Algorithm | Precision | Recall | F-measure |
|---|---|---|---|
| Class 4 | | | |
| **SVR** | $0.31 \pm 0.04$ | $0.07 \pm 0.01$ | $0.11 \pm 0.02$ |
| **Decision Tree** | $0.32 \pm 0.08$ | $0.08 \pm 0.02$ | $0.13 \pm 0.03$ |
| **Boosting** | $0.43 \pm 0.07$ | $0.09 \pm 0.02$ | $0.15 \pm 0.03$ |
| **Random Forest** | $0.47 \pm 0.07$ | $0.09 \pm 0.01$ | $0.15 \pm 0.02$ |
| **HPEC** | $0.24 \pm 0.03$ | $0.28 \pm 0.03$ | $0.26 \pm 0.03$ |
| Class 5 | | | |
| **SVR** | $0.54 \pm 0.14$ | $0.11 \pm 0.05$ | $0.18 \pm 0.08$ |
| **Decision Tree** | $0.50 \pm 0.14$ | $0.18 \pm 0.07$ | $0.26 \pm 0.08$ |
| **Boosting** | $0.69 \pm 0.11$ | $0.24 \pm 0.04$ | $0.35 \pm 0.05$ |
| **Random Forest** | $0.68 \pm 0.10$ | $0.23 \pm 0.05$ | $0.35 \pm 0.06$ |
| **HPEC** | $0.33 \pm 0.06$ | $0.40 \pm 0.07$ | $0.36 \pm 0.06$ |

where $n$ is the number of points and $y^{(i)}$ and $\tilde{y}^{(i)}$ are, respectively, the real and the estimated class. Finally, we also considered F-measure, i.e., the weighted harmonic mean of precision and recall (or POD), computed only for the minority classes (4 and 5), which regard intense rainfall events.

### *A. Comparison of Ensemble-Based Techniques*

We performed the Friedman test for all the evaluation measures (columns) of Tables II and III, while the Nemenyi is adopted as *post hoc* test, as suggested in [50].

Table II reports the values of CSI (higher is better), FAR (lower is better), POD (higher is better), and MSE (lower is better) for the five algorithms, while, in Table III, Precision, Recall, and F-measure for the minority (and presenting a high-intensity of rainfall) classes 4 and 5 are shown.

For all the measures, the Friedman test is verified, and consequently, the differences among the algorithms are significant. Therefore, the null hypothesis is rejected, and then, in the following, we analyze the rankings and the *post hoc* test for the most important measures for our particular application, i.e., FAR, POD, and MSE (see Table II) and Recall and F-Measure (see Table III) for the minority classes.

In terms of POD, the test indicates that HPEC is significantly better than the other approaches. No significant differences are reported between SVR and decision tree and also for boosting and RF.

As for the mean-square error (MSE), according to the significance test, the RF algorithm is significantly better than the other ones, while HPEC is not significantly different from the other algorithms.

As for the F-measure, for the minority class 4, both in terms of Recall and F-measure, HPEC is significantly better than all the other approaches. On the contrary, RF and boosting are not significantly different, and in the same way, SVR and

TABLE IV

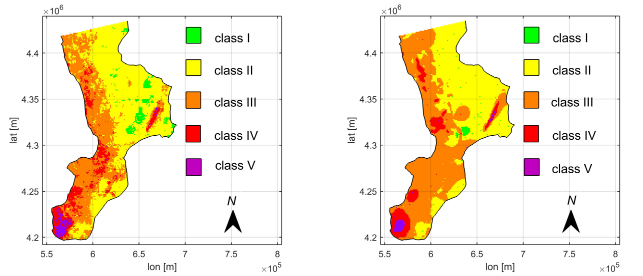
CSI, FAR, POD, AND MSE FOR THE KRIGING, RF, AND HPEC. THE VALUES IN BOLD (LIGHT GRAY) ARE SIGNIFICANTLY BETTER (WORSE) THAN THE KRIGING METHOD

| Algorithm | CSI | FAR | POD | MSE |
|---|---|---|---|---|
| **Kriging** | $0.40 \pm 0.010$ | $0.39 \pm 0.018$ | $0.48 \pm 0.013$ | $0.15 \pm 0.002$ |
| **Random Forest** | $\mathbf{0.43 \pm 0.010}$ | $\mathbf{0.31 \pm 0.024}$ | $\mathbf{0.49 \pm 0.011}$ | $\mathbf{0.09 \pm 0.002}$ |
| **HPEC** | $\mathbf{0.44 \pm 0.011}$ | $0.46 \pm 0.016$ | $\mathbf{0.58 \pm 0.016}$ | $\mathbf{0.11 \pm 0.002}$ |

TABLE V

PRECISION, RECALL, AND F-MEASURE FOR THE KRIGING, RF, AND HPEC FOR THE MINORITY CLASSES 4 AND 5. THE VALUES IN BOLD (LIGHT GRAY) ARE SIGNIFICANTLY BETTER (WORSE) THAN THE KRIGING METHOD

| Algorithm | Precision | Recall | F-measure |
|---|---|---|---|
| Class 4 | | | |
| **Kriging** | $0.37 \pm 0.042$ | $0.20 \pm 0.026$ | $0.26 \pm 0.031$ |
| **Random Forest** | $\mathbf{0.47 \pm 0.067}$ | $0.09 \pm 0.014$ | $0.15 \pm 0.020$ |
| **HPEC** | $0.24 \pm 0.031$ | $\mathbf{0.28 \pm 0.030}$ | $0.26 \pm 0.027$ |
| Class 5 | | | |
| **Kriging** | $0.45 \pm 0.078$ | $0.32 \pm 0.059$ | $0.38 \pm 0.062$ |
| **Random Forest** | $\mathbf{0.68 \pm 0.101}$ | $0.23 \pm 0.047$ | $0.35 \pm 0.057$ |
| **HPEC** | $0.33 \pm 0.062$ | $\mathbf{0.40 \pm 0.066}$ | $0.36 \pm 0.057$ |




(a) (b)

Fig. 5. Kriging with external drift. (a) HPEC. (b) Kriging with external drift.

decision tree are not significantly different. Finally, for class 5, HPEC is significantly better than all the other algorithms for the recall measure, while the ensemble-based algorithms overcome the other algorithms for the F-measure, but they are not significantly different among themselves. Therefore, by considering that we are particularly interested in detecting a larger number of heavy rainfall events (class 4 and 5), HPEC appears to be a good choice.

### *B. Comparison of Ensemble-Based Algorithms and Kriging*

This section aims to evaluate whether ensemble-based algorithms improve the performance in comparison with largely used traditional algorithms designed by experts of the domain, as the Kriging algorithm. More in detail, the adopted algorithm is Kriging with external drift (KED), using the RG and the radar data, respectively, as primary and secondary variables.

The significance of the differences between Kriging and each ensemble algorithm is evaluated by using the Wilcoxon signed-ranked test [confidence level equal to 0.95 ($\alpha = 0.05$)]. This test is conducted both for Kriging versus RF and Kriging versus HPEC, as we are mainly interested in the comparison with the Kriging algorithm. The values that are significantly better are reported in bold, while the values that are significantly worst are reported in light gray.

Table IV reports, for the three algorithms, the values of CSI, FAR, POD, and MSE, while Table V reports Precision, Recall, and F-measure for the minority (and high-rate rainfall) classes 4 and 5. For all the measures, the ensemble-based algorithms perform significantly better (see Table IV). In particular, HPEC reaches a high value for the POD measure (0.58).

As for the minority classes, in spite of having a worse precision, HPEC exhibits a better recall. In addition, by considering the F-measure, there are no significant differences between Kriging and HPEC both for classes 4 and 5, while the RF algorithm behaves considerably worse. However, the MSE is considerably better for the ensemble-based techniques in comparison with KED. Then, the misclassification regards, in many cases, classes not so dissimilar for the amount of rainfall, i.e., a rainfall of class 4 could be assigned to class 3 or 5.

The results of this section suggest that a clever combination of Kriging and ML algorithms could further improve the capacity of the framework to detect heavy rainfall events.

An example of the differences in reconstructing the rainfall is shown in Fig. 5, respectively, (a) for the HPEC and (b) for KED. We would like to point out that this figure gives just a qualitative idea of the behavior of the two techniques for a significant event. We choose a very rainy event outside of the training period considered (November 6, 2017), in which three RGs registered their annual maximum of rainfall (about 42.4 mm for an hour). In order to compare the results, we consider the rain cumulated every 30 min, and we discretize the estimate of the Kriging method on the same classes considered for the ML approach. From this figure, it seems that the HPEC approach gives better detail of the rainfall events; however, further investigations should be conducted to draw a conclusion. For a more accurate comparison, it would be better to refer to the results reported in the previous tables.

### *C. Effect of Integrating RG, Satellite, and Radar Measurements*

Similar to other works [8], [25], in our framework data coming from RGs, satellites, and radars are integrated in order to classify rainfall events better. In this section, we want to investigate the effect of this integration, by understanding whether all the three types of measurement are necessary to the classification. To this aim, by using the HPEC, we run three different suites of experiments by excluding, for each suite, one of the three types of measurement.

TABLE VI

CSI, FAR, POD, AND MSE FOR THE HPEC, USING ALL THE FEATURES VERSUS NOT CONSIDERING, RESPECTIVELY, RG, RADAR, AND SATELLITE DATA. THE VALUES IN LIGHT GRAY ARE SIGNIFICANTLY WORSE THAN THE METHOD USING ALL THE FEATURES

| Algorithm | CSI | FAR | POD | MSE |
|---|---|---|---|---|
| **All** | $0.44 \pm 0.011$ | $0.46 \pm 0.016$ | $0.58 \pm 0.016$ | $0.11 \pm 0.002$ |
| **No rain gauge** | $0.40 \pm 0.005$ | $0.53 \pm 0.008$ | $0.54 \pm 0.010$ | $0.16 \pm 0.010$ |
| **No radar** | $0.43 \pm 0.007$ | $0.47 \pm 0.011$ | $0.58 \pm 0.012$ | $0.12 \pm 0.002$ |
| **No meteosat** | $0.39 \pm 0.006$ | $0.52 \pm 0.008$ | $0.54 \pm 0.012$ | $0.16 \pm 0.003$ |

TABLE VII

PRECISION, RECALL, AND F-MEASURE FOR THE HPEC, USING ALL THE FEATURES VERSUS NOT CONSIDERING, RESPECTIVELY, RG, RADAR, AND SATELLITE DATA, FOR THE MINORITY CLASSES 4 AND 5. THE VALUES IN LIGHT GRAY ARE SIGNIFICANTLY WORSE THAN THE METHOD USING ALL THE FEATURES

| Algorithm | Precision | Recall | F-measure |
|---|---|---|---|
| Class 4 | | | |
| **All** | $0.24 \pm 0.031$ | $0.28 \pm 0.030$ | $0.26 \pm 0.027$ |
| **No gauge** | $0.13 \pm 0.018$ | $0.22 \pm 0.027$ | $0.16 \pm 0.017$ |
| **No radar** | $0.22 \pm 0.019$ | $0.28 \pm 0.028$ | $0.25 \pm 0.021$ |
| **No meteosat** | $0.19 \pm 0.017$ | $0.25 \pm 0.026$ | $0.22 \pm 0.017$ |
| Class 5 | | | |
| **All** | $0.33 \pm 0.062$ | $0.40 \pm 0.066$ | $0.36 \pm 0.057$ |
| **No gauge** | $0.17 \pm 0.036$ | $0.31 \pm 0.050$ | $0.22 \pm 0.038$ |
| **No radar** | $0.30 \pm 0.055$ | $0.40 \pm 0.058$ | $0.34 \pm 0.049$ |
| **No meteosat** | $0.29 \pm 0.043$ | $0.41 \pm 0.063$ | $0.34 \pm 0.040$ |

Tables VI and VII report the results for the cases of using all the features, and, respectively, in the case of the removal of the RG features, of the radar features and the satellite features.

From Table VI, it is clear that, by removing the RG information, the performance of the algorithm worsens for all the measures. On the contrary, by removing one of the other two types of data, the performance is less affected, even if the lower value (0.11) of the MSE, obtained when all the data are used, confirms the utility in using all the sources of data.

The benefit derived from the integration of the three sources is also confirmed by the F-measure in Table VII.

## VI. CONCLUSION AND DISCUSSION

An ML-based approach for the spatial rainfall field estimation has been defined. By integrating heterogeneous data sources, such as RGs, radars, and satellites, this methodology permits estimation of the rainfall, where RGs are not present, also exploiting the spatial pattern recognition ensured by radars and satellites. After a phase of preprocessing, a random uniform undersampling strategy is adopted, and finally, an HPEC permits the model used to be built to estimate the severity of the rainfall events. This ensemble is based on two levels: in the first level, a set of RF classifiers are trained, while, in the second level, a probabilistic metalearner is used to combine the estimated probabilities provided by the base classifiers according to a stacking schema.

Experimental results conducted on real data provided by the Department of Civil Protection show significant improvements in comparison with Kriging with external drift, a largely used and well-recognized method in the field of rainfall estimation. In particular, the ensemble method exhibits a better capacity in detecting the rainfall events. Indeed, both the POD (0.58) and the MSE (0.11) measures obtained by HPEC are significantly better than the values obtained by KED (0.48 and 0.15, respectively). As for the last two classes, representing intense rainfall events, the difference between the Kriging method and HPEC is not significant (in terms of F-measure) although HPEC is computationally more efficient.

Indeed, the complexity of the Kriging method is cubic in the number of the samples [51], which makes the procedure really expensive from the computational point of view, when a large number of points are analyzed. On the contrary, the ML algorithms (i.e., RF) exhibit a quadratic complexity. Moreover, ensemble methods are highly scalable and parallelizable. Therefore, we believe that our approach has some relevant advantages in this field of application.

In addition, by analyzing the effect of the integration of the different sources of data, it is evident that all the data sources contribute to the good performance of the technique. In particular, by removing the RG information, the performance of the algorithm worsens the sensibly for all the measures. In the cases of the removal of one of the other two types of data, the degradation is less evident; however, the lowest value (0.11) of the MSE is obtained when all the data are used, which confirms that it is necessary to use all the sources of data to obtain better results.

As future work, we plan to validate the method on a larger time interval, in order to consider effects due to seasonal and yearly variability, also considering the possibility of incrementally building the flexible ensemble model with the new data. In addition, we want to evaluate the effectiveness of the algorithm in individuate highly localized heavy precipitation events, also by adopting time series analysis to analyze the individual contributions of the different features for radar and Meteosat.

## ACKNOWLEDGMENT

The authors thank the Calabrian Regional Environmental Protection Agency (Functional Multirisk Center of the region Calabria) for providing the used data.

## REFERENCES

[1] J. E. Ball and K. C. Luk, “Modeling spatial variability of rainfall over a catchment,” *J. Hydrologic Eng.*, vol. 3, no. 2, pp. 122–130, Apr. 1998.

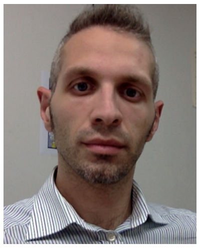
[2] S. Ly, C. Charles, and A. Degré, “Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. a review,” *Biotechnologie, Agronomie, Société et Environnement*, vol. 17, no. 2, p. 392, 2013.

[3] H. S. Wheater *et al.*, “Spatial-temporal rainfall fields: Modelling and statistical aspects,” *Hydrol. Earth Syst. Sci.*, vol. 4, no. 4, pp. 581–601, Dec. 2000.

[4] J. L. McKee and A. D. Binns, “A review of gauge–radar merging methods for quantitative precipitation estimation in hydrology,” *Can. Water Resour. J./Revue Canadienne des Ressources Hydriques*, vol. 41, nos. 1–2, pp. 186–203, 2016.

[5] F. Cecinati, O. Wani, and M. A. Rico-Ramirez, “Comparing approaches to deal with non-gaussianity of rainfall data in Kriging-based radar-gauge rainfall merging,” *Water Resour. Res.*, vol. 53, no. 11, pp. 8999–9018, Nov. 2017.

[6] H. Wackernagel, *Multivariate Geostatistics: An Introduction With Applications*. Berlin, Germany: Springer, 2003.
[7] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.
[8] B. J. E. Schroeter, *Artificial Neural Networks in Precipitation Now-Casting: An Australian Case Study*. Cham, Switzerland: Springer, 2016, pp. 325–339.
[9] X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, vol. 1, Dec. 2015, pp. 802–810.
[10] W.-C. Hong, "Rainfall forecasting by technological machine learning models," *Appl. Math. Comput.*, vol. 200, no. 1, pp. 41–57, Jun. 2008.
[11] A. Parmar, K. Mistree, and M. Sompura, "Machine learning techniques for rainfall prediction: A review," in *Proc. 4th Int. Conf. Innov. Inf. Embedded Commun. Syst. (ICIIECS)*, Mar. 2017, pp. 152–162.
[12] M. Lee, N. Kang, H. Joo, H. Kim, S. Kim, and J. Lee, "Hydrological modeling approach using radar-rainfall ensemble and multi-runoff-model blending technique," *Water*, vol. 11, no. 4, pp. 1–18, 2019.
[13] C. Frei and F. A. Isotta, "Ensemble spatial precipitation analysis from rain gauge data: Methodology and application in the European alps," *J. Geophys. Res., Atmos.*, vol. 124, no. 11, pp. 5757–5778, Jun. 2019.
[14] J. L. Peña-Arancibia, A. I. J. M. van Dijk, L. J. Renzullo, and M. Mulligan, "Evaluation of precipitation estimation accuracy in reanalyses, satellite products, and an ensemble method for regions in Australia and south and East Asia," *J. Hydrometeorology*, vol. 14, no. 4, pp. 1323–1333, Aug. 2013.
[15] V. Levizzani, P. Bauer, and F. J. Turk, *Measuring precipitation from Space: EURAINSAT and the Future*, vol. 28. Berlin, Germany: Springer, 2007.
[16] F. Chiaravalloti, L. Brocca, A. Procopio, C. Massari, and S. Gabriele, "Assessment of GPM and SM2RAIN-ASCAT rainfall products over complex terrain in southern italy," *Atmos. Res.*, vol. 206, pp. 64–74, Jul. 2018.
[17] B. Thies, T. Nauss, and J. Bendix, "Discriminating raining from non-raining cloud areas at mid-latitudes using Meteosat Second Generation SEVIRI night-time data," *Meteorolog. Appl.*, vol. 15, no. 2, pp. 219–230, 2008.
[18] M. Lazri, Z. Ameur, S. Ameur, Y. Mohia, J. M. Brucker, and J. Testud, "Rainfall estimation over a mediterranean region using a method based on various spectral parameters of SEVIRI-MSG," *Adv. Space Res.*, vol. 52, no. 8, pp. 1450–1466, Oct. 2013.
[19] H. Feidas and A. Giannakos, "Identifying precipitating clouds in greece using multispectral infrared Meteosat Second Generation satellite data," *Theor. Appl. Climatol.*, vol. 104, nos. 1–2, pp. 25–42, May 2011.
[20] H. Feidas, G. Kokolatos, A. Negri, M. Manyin, N. Chrysoulakis, and Y. Kamarianakis, "Validation of an infrared-based satellite algorithm to estimate accumulated rainfall over the mediterranean basin," *Theor. Appl. Climatol.*, vol. 95, nos. 1–2, pp. 91–109, Jan. 2009.
[21] A. Giannakos and H. Feidas, "Classification of convective and stratiform rain based on the spectral and textural features of meteosat second Generation infrared data," *Theor. Appl. Climatol.*, vol. 113, nos. 3–4, pp. 495–510, Aug. 2013.
[22] M. Grecu and W. S. Olson, "Bayesian estimation of precipitation from satellite passive microwave observations using combined radar–radiometer retrievals," *J. Appl. Meteorol. Climatol.*, vol. 45, no. 3, pp. 416–433, Mar. 2006.
[23] A. Verdin, B. Rajagopalan, W. Kleiber, and C. Funk, "A Bayesian Kriging approach for blending satellite and ground precipitation observations," *Water Resour. Res.*, vol. 51, no. 2, pp. 908–921, Feb. 2015.
[24] M. A. Tebbi and B. Haddad, "Artificial intelligence systems for rainy areas detection and convective cells' delineation for the south shore of Mediterranean Sea during day and nighttime using MSG satellite images," *Atmos. Res.*, vols. 178–179, pp. 380–392, Sep. 2016.
[25] M. Sehad, M. Lazri, and S. Ameur, "Novel SVM-based technique to improve rainfall estimation over the mediterranean region (North of Algeria) using the multispectral MSG SEVIRI imagery," *Adv. Space Res.*, vol. 59, no. 5, pp. 1381–1394, Mar. 2017.
[26] M. Lazri, S. Ameur, and Y. Mohia, "Instantaneous rainfall estimation using neural network from multispectral observations of SEVIRI radiometer and its application in estimation of daily and monthly rainfall," *Adv. Space Res.*, vol. 53, no. 1, pp. 138—155, 2014.
[27] M. Kühnlein, T. Appelhans, B. Thies, and T. Nauß, "Precipitation estimates from MSG SEVIRI daytime, nighttime, and twilight data with random forests," *J. Appl. Meteorol. Climatol.*, vol. 53, no. 11, pp. 2457–2480, Nov. 2014.
[28] P.-F. Pai and W.-C. Hong, "A recurrent support vector regression model in rainfall forecasting," *Hydrological Processes*, vol. 21, no. 6, pp. 819–827, 2007.
[29] T. Wardah, S. H. Abu Bakar, A. Bardossy, and M. Maznorizan, "Use of geostationary Meteorological Satellite images in convective rain estimation for flash-flood forecasting," *J. Hydrol.*, vol. 356, nos. 3–4, pp. 283–298, Jul. 2008.
[30] H. Tan, V. Chandrasekar, and H. Chen, "A machine learning model for radar rainfall estimation based on gauge observations," in *Proc. United States Nat. Committee URSI Nat. Radio Sci. Meeting (USNC-URSI NRSM)*, Jan. 2017, pp. 1–2.
[31] T. Zea Tan, G. Kee Khoon Lee, S.-Y. Liong, T. Kuay Lim, J. Chu, and T. Hung, "Rainfall intensity prediction by a spatial-temporal ensemble," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 1721–1727.
[32] H. Meyer, M. Kühnlein, T. Appelhans, and T. Nauss, "Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals," *Atmos. Res.*, vol. 169, pp. 424–433, Mar. 2016.
[33] T. Caloiero, G. Buttafuoco, R. Coscarelli, and E. Ferrari, "Spatial and temporal characterization of climate at regional scale using homogeneous monthly precipitation and air temperature data: An application in Calabria (Southern Italy)," *Hydrol. Res.*, vol. 46, no. 4, pp. 629–646, Aug. 2015.
[34] S. Gabriele, F. Chiaravalloti, and A. Procopio, "Radar–rain-gauge rainfall estimation for hydrological applications in small catchments," *Adv. Geosci.*, vol. 44, p. 61, 2017.
[35] E. Avolio, O. Cavalcanti, L. Furnari, A. Senatore, and G. Mendicino, "Brief communication: Preliminary hydro-meteorological analysis of the flash flood of 20 august 2018 in raganello gorge, southern italy," *Natural Hazards Earth Syst. Sci.*, vol. 19, no. 8, pp. 1619–1627, Aug. 2019.
[36] V. Rago *et al.*, "Geomorphic effects caused by heavy rainfall in southern calabria (Italy) on 30 october–1 november 2015," *J. Maps*, vol. 13, no. 2, pp. 836–843, Nov. 2017.
[37] G. Vulpiani *et al.*, "The Italian radar network within the national early-warning system for multi-risks management," in *Proc. 5th Eur. Conf. Radar Meteorol. Hydrol. (ERAD)*, vol. 184, 2008, pp. 1–6.
[38] D. Cimini *et al.*, "Validation of satellite OPEMW precipitation product with ground-based weather radar and rain gauge networks," *Atmos. Meas. Techn.*, vol. 6, no. 11, p. 3181, 2013.
[39] M. Vázquez-Navarro, H. Mannstein, and S. Kox, "Contrail life cycle and properties from 1 year of MSG/SEVIRI rapid-scan images," *Atmos. Chem. Phys.*, vol. 15, no. 15, pp. 8739–8749, Aug. 2015.
[40] F. Cecinati, A. de Niet, K. Sawicka, and M. Rico-Ramirez, "Optimal temporal resolution of rainfall for urban applications and uncertainty propagation," *Water*, vol. 9, no. 10, p. 0762, Oct. 2017.
[41] N. Japkowicz, "The class imbalance problem: Significance and strategies," in *Proc. Int. Conf. Artif. Intell. (ICAI)*, 2000, pp. 111–117.
[42] A. Dal Pozzolo, O. Caelen, and G. Bontempi, "When is undersampling effective in unbalanced classification tasks," in *Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, Porto, Portugal. Cham, Switzerland: Springer, 2015, pp. 200–215.
[43] Y. Freund and R. Scapire, "Experiments with a new boosting algorithm," in *Proc. 13th Int. Conf. Mach. Learn.*, 1996, pp. 148–156.
[44] R. E. Schapire, "The strength of weak learnability," *Mach. Learn.*, vol. 5, no. 2, pp. 197–227, 1990.
[45] R. E. Schapire, "Boosting a weak learning by majority," *Inf. Comput.*, vol. 121, no. 2, pp. 256–285, 1996.
[46] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992.
[47] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[48] P. H. Hiemstra, E. J. Pebesma, C. J. W. Twenhöfel, and G. B. M. Heuvelink, "Real-time automatic interpolation of ambient gamma dose rates from the dutch radioactivity monitoring network," *Comput. Geosci.*, vol. 35, no. 8, pp. 1711–1721, Aug. 2009.
[49] H. Stanski, L. Wilson, and W. Burrows, *Survey of Common Verification Methods in Meteorology*. Geneva, Switzerland: World Meteorological Organization, 1989.
[50] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Jan. 2006.
[51] B. V. Srinivasan, R. Duraiswami, and R. Murtugudde, "Efficient Kriging for real-time Spatio-temporal interpolation," in *Proc. 20th Conf. Probab. Statist. Atmos. Sci.*, 2010, pp. 228–235.



**Massimo Guarascio** received the Ph.D. degree in system and computer engineering from the University of Calabria, Calabria, Italy, in 2011.

He is a Researcher with the Institute for High Performance Computing and Networking (ICAR-CNR), National Research Council and shareholder of OKT s.r.l., a spin-off of the University of Calabria. He has participated in European and National research projects concerning machine learning and cybersecurity. He has coauthored over 40 articles published in international conference proceedings, chapters, and journals. His research mainly focuses on machine learning, anomaly detection and explanation, process mining, data analytics methods for geosciences and remote sensing, knowledge discovery and data mining for cybersecurity and fraud detection.

**Gianluigi Folino** received the Ph.D. degree in physics, mathematics, and computer science, Radboud University, Nijmegen (Holland), The Netherlands, in 2010.

Since 2001, he has been a Senior Researcher with the Institute of High Performance Computing and Networking, Italian National Research Council (ICAR-CNR), Rende, Italy. He is also a Lecturer with the University of Calabria. He was a Visiting Researcher with the University of Nottingham, Nottingham, U.K., Radbound University, Nijmegen, The Netherlands, and the University of California (UCLA). He has authored or coauthored more than 100 articles in international conferences and journals among which the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, *Parallel Computing, Information Sciences and Bioinformatics*. His research interests focus on applications of distributed computing in the area of data mining, bio-inspired algorithms, big data, bioinformatics, and cybersecurity.

Dr. Folino is on the Editorial Board of *Applied Soft Computing* (Elsevier).

**Francesco Chiaravalloti** received the master's degree *(Summa Cum Laude)* in physics and the Ph.D. degree in science and technology for complex systems from the University of Calabria, Calabria, Italy, in 2002 and 2015, respectively.

He was a Research Fellow with the Italian National Research Council (CNR), from 2002 to 2009. From November 2013 to May 2014, he was a Visiting Student with The Complex Systems and Networks Lab (COSNET), Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Zaragoza, Spain. In 2015, he was a Research Fellow with the Department of Mechanical, Energy and Management Engineering, University of Calabria. He is a Post-Doctoral Fellow with the Research Institute of Geo-Hydrological Protection (CNR-IRPI), Rende, Italy. He coauthored over 30 articles published in international conference proceedings and journals. His recent research activity is focused on the study of meteorological extreme events, rain field estimation, and short nowcasting of convective events.

**Salvatore Gabriele** received the master's degree *(Summa Cum Laude)* in civil engineering from the University of Calabria (UNICAL), in 1978.

Since 1982, he has been a Senior Research Scientist with the Italian National Research Council, Research Institute of Geo-Hydrological Protection (CNR-IRPI), Rende, Italy. He was a Coordinator with the Remote Sensing laboratory, CNR-IRPI, and the Italian Supervisor for the Bilateral Project Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil (CNR-CNPq) for TCEV—Two Component Extreme Value application in the basins of Minas Gerais, Brazil. He was a Supervisor and a Project Manager of different projects for hydrogeological risk monitoring in the Calabria Region. Since 2014, he has been a Supervisor for the research agreement between CNR-IRPI and RFI—Rete Ferroviaria Italiana (the Italian railway infrastructure) on hydrogeological disaster risk prevention for the South Italy railway network. He coauthored over 100 articles published in international conferences and journals. His recent research activity mainly concerns the statistical analysis of extreme hydrological events and short nowcasting of extreme events, which affect the security of railways, roads, and airports.

**Antonio Procopio** received the master's degree in mathematics and the Ph.D. degree in science and technologies of complex systems from the University of Calabria, in 2012 and 2017, respectively.

From February 2015 to July 2015, he was a Visiting Student with the School of Mathematical Science of Queen Mary University of London, U.K. He was a Post-Doctoral Fellow with the Research Institute of Geo-Hydrological Protection (CNR-IRPI), Rende, Italy. He is employed as a Junior Consultant with NTT DATA, Italy.

**Pietro Sabatino** received the master's degree in mathematics from the University of Calabria, in 1998, and the Ph.D. degree in mathematics from the University of Rome, "Tor Vergata," Rome, Italy, in 2003.

He worked as a Research Fellow within the Department of Mathematics, University of Calabria, subsequently, he was a Research Fellow with the Institute of High Performance Computing and Networking of the Italian National Research Council (ICAR-CNR). He was also a Lecturer with the University of Calabria, and at the University of Rome, "Tor Vergata." His research interests range from pure mathematics, in particular algebraic and differential geometry, to data mining and machine learning, including topological and geometric data analysis and algebraic statistics.