On Measures of Uncertainty in Classification

Saloua Chlaily ⬤, Debanshu Ratha ⬤, Pigi Lozou ⬤, and
Andrea Marinoni ⬤, *Senior Member, IEEE*

Abstract—**Uncertainty is unavoidable in classification tasks and might originate from data (e.g., due to noise or wrong labeling), or the model (e.g., due to erroneous assumptions, etc). Providing an assessment of uncertainty associated with each outcome is of paramount importance in assessing the reliability of classification algorithms, especially on unseen data. In this work, we propose two measures of uncertainty in classification. One of the measures is developed from a geometrical perspective and quantifies a classifier's distance from a random guess. In contrast, the second proposed uncertainty measure is homophily-based since it takes into account the similarity between the classes. Accordingly, it reflects the type of mistaken classes. The proposed measures are not aggregated, i.e., they provide an uncertainty assessment to each data point. Moreover, they do not require label information. Using several datasets, we demonstrate the proposed measures' differences and merit in assessing uncertainty in classification. The source code is available at github.com/pioui/uncertainty.**

Index Terms—**Classification, uncertainty, evaluation measure, categorical distribution, geometry-based uncertainty, homophily-based uncertainty.**

I. INTRODUCTION

THE most commonly used approaches to evaluate the performance of a classifier rely on comparing the outcome of the classifier with the ground truth. Several measures might be used for this purpose, such as: overall accuracy, kappa coefficient score, and confusion matrix, to name a few [1], [2], [3]. These measures report how well the classifier has learned from the training data. However, these quantities heavily rely on the correctness of the labels. Accordingly, in the case of erroneous labels, a bad classifier might show a better performance than a good classifier that overcomes wrong labeling [4]. Moreover, as rich and dense as it can be, the labeled data cannot represent all possible variations among and across the classes. This is especially true when the data represent very complex phenomena such as remote sensing data, medical images, etc. In fact, the labeled data within these data sets are usually scarce, i.e., the portion of unlabeled pixels is more extensive than the labeled pixels. As such, there is a need for a metric to evaluate the performance of a classifier on unseen data. This measure is more beneficial if it is specific to each prediction and not aggregated as the commonly used measures [1], [2], [3]. In fact, aggregated measurements cannot convey which particular samples are challenging for the classifier to characterize. Accordingly, the properties of such samples and what made them problematic cannot be deduced. A key to this shortcoming, especially in the absence of ground truth, is assessing the classification's uncertainty.

Despite the various ways in which scientists interpret the concept of uncertainty, it is generally associated with probability [5]. Moreover, variance and entropy are the standard measures to reflect the uncertainty from probability distributions [6]. Variance is typically used as an uncertainty measure in the case of regression, while entropy is more suited for classification [5]. In fact, since the predictions in classification are categorical, i.e., the labels are nominal, their mean is unquantifiable. Accordingly, the variance, conventionally defined as a function of the mean, is inapplicable. Yet, in the binary classification, the predictions follow a Bernoulli distribution, and accordingly, the variance of this distribution applies [7]. A straightforward remedy to this issue is to treat a multiclass scenario as binary. However, in this case, a classifier of four categories that outputs probabilities $[0.51, 0.49, 0, 0]$ is equivalent to a classifier that outputs $[0.51, 0.19, 0.15, 0.15]$, although it is clear that the second model is closer to a random classifier and should produce a higher uncertainty accordingly. Different attempts exist in the literature for computing the variance of categorical data. Nonetheless, the general practice is substituting probabilities by relative frequencies in the Gini-Simpson index or Shannon entropy [8].

Many factors can contribute to uncertainty in classification. Noise, for instance, is an unavoidable source of uncertainty. It can stem from measurement instruments, observation conditions (such as clouds in optical data), or incorrect labeling. Moreover, the data in many applications are scarce and only represent partial information about the considered phenomena. This lack of information also adds up to the uncertainty. Uncertainty can also emanate from the chosen model. Generally, several assumptions that might be inaccurate are postulated for the sake of simplicity, such as independence, linearity, or gaussianity. Even complex models introduce uncertainty, given that they comprise a large number of parameters and given the underlying risk of overfitting.

The source of uncertainty is, however, irrelevant in this work. Whatever the source of uncertainty, in classification, it always comes down to *separability*. The separability of data refers to the extent to which different classes or categories in a dataset can be clearly distinguished. This can be hindered by all the reasons mentioned above. The separability becomes less pronounced when data points lie in ambiguous regions or fall near the boundary between categories. In such cases, the model's uncertainty in making predictions increases. The model may struggle to make confident predictions because the input data in those regions may share similarities with different classes compared to instances representative of a single category. Furthermore, untrained areas refer to regions of the feature space where the model has not encountered any examples during training. Since the model lacks exposure to these areas, it has no prior knowledge or reference points to make accurate predictions. Consequently, predictions made in these untrained areas are also expected to be inaccurate and uncertain.

Therefore, we are particularly interested in assessing the *confidence*, i.e., the certainty of a model in its prediction, and the *confusion*, i.e., the margin of this confidence compared to other possible classes, of a classifier. Such a measure won't only assess the quality of a classifier but will also provide additional input to the end-user and help make informed decisions. In some applications, certain misclassifications might be very costly. Sea-ice classification, for instance, is challenging and vulnerable to many errors that might put polar navigation at risk [9]. Likewise, in medical applications, not detecting cancer might be life-threatening. Providing an uncertainty measure along with the classification maps will help make better decisions by giving up on uncertain routes or running additional tests to recheck the uncertain results when needed.

Classification uncertainty can be derived from the posterior probability. The higher the probability that a data point belongs to a specific class, the lower the uncertainty. Conversely, the most uncertain scenario is where the classifier cannot make a decision, and a uniform distribution produces it. Accordingly, we define a *geometry-based* measure of uncertainty as a function of the distance between the posterior probability and the discrete uniform distribution in the feasible space of probability distributions. Several measures of uncertainty can be derived depending on the distance defined in this space. In particular, entropy is a special case when the distance considered in this space is given by the Kullback-Leibler divergence.

Based on this definition of uncertainty, we can quantify how far is the classifier from a random guess. Moreover, we can assess the quality of a classifier by examining its confidence in correct and wrong outcomes. Although this uncertainty unveils some properties of the classifier, it is lacking. It only reflects confidence and confusion of a classifier but not the type of confusion. Confusing close classes from a features point of view emanates from inseparability. However, confusing distant classes from a features perspective is alarming and might signal a lousy model, noisy data point or label, etc. In this work, in addition to the *geometry-based* uncertainty, we introduce a *homophily-based* uncertainty that incorporates

information from the class distributions and reflects the types of confused classes.

The remainder of this manuscript is organized as follows. Firstly, some related works are discussed in Section II. Then, the theory of uncertainty is given in Section III. In Section IV, we show how the proposed measures assess uncertainty in classification and reveal valuable information about the considered data and model. Section V concludes this article.

II. RELATED WORK

Uncertainty is generally associated with probability and, accordingly, with Bayesian approaches. For instance, Gaussian processes are well known for accurately estimating uncertainty [10]. This motivated the development of deep Bayesian networks. The Bayesian approaches enabled the characterization of uncertainty into aleatoric or epistemic [5]. Aleatoric uncertainty arises from randomness, while epistemic one arises from lack of information. As such, epistemic refers to the uncertainty that can be improved as opposed to the aleatoric one. However, it should be noted that these two notions might have different definitions under different circumstances, models, and domains of application [5]. More details on techniques for quantifying these uncertainties in machine learning and deep learning can be found in [5], [11].

To circumvent the computational cost of Bayesian methods, non-Bayesian approaches have also been proposed to quantify uncertainty. These approaches consist of approximating the Bayesian inference through dropout [12], ensemble neural networks [13], [14], or modeling the neural network outputs by a probability distribution [15], [16].

Other approaches propose associating probabilistic models with performance measures. Brodersen et al. model the overall accuracy with beta-binomial distribution [17]. Caelen models the confusion matrix by a Dirichlet-multinomial distribution [18], while Tötsch et al. model it by three beta-binomials [19]. Using the size of the dataset, these approaches evaluate how the classifier is better than random assignment.

While these methods address estimating a well-calibrated probability distribution, we are interested in quantifying uncertainty from the estimated probabilities. The related works include variance and entropy [6]. Variance quantifies the dispersion of values from their mean. The greater the variance, the more imprecise and, therefore, the more uncertain the estimate. Entropy is a measure of information and, respectively, lack of information. The more probable an event, the lower the entropy and, therefore, the lower the uncertainty. Alternative measures of information have also been proposed. We might cite, for instance, Rényi-entropy, which is a generalization of Shannon's entropy [20], Tsallis entropy which is one of the non-additive entropies [21] and from the new proposed measures we cite t-entropy [22], which is a function of the inverse tangent.

Several measures of uncertainty were also developed under the framework of credal sets [23], [24]. In fact, despite the powerful capabilities of probability, it has been criticized for its inability to model the lack of information (ignorance). Several generalizations have been proposed to overcome this limitation,

including imprecise probability and Dempster–Shafer theory, also known as evidence theory [25], [26]. The commonality of these theories consists of considering a set of distributions. For instance, a set of candidate priors is considered instead of a single prior probability in imprecise probability theory. However, it should be noted that these methods are out of the scope of this article since our article assumes a trained classifier.

III. THEORY OF UNCERTAINTY

We consider a classification problem with C classes. Given a training data with N instances, $\mathcal{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$, the aim of the classifier given a new input $\mathbf{x}_*$ is to infer its corresponding class y_*, i.e., estimate $p(y_*|\mathcal{D}, \mathbf{x}_*)$. Accordingly, the classifier generates C estimates, $\mathbf{p}_* = [p_{*1}, \ldots, p_{*C}]^T$, that correspond to the posterior probability of each of the C classes, $p_{*c} = p(y_* = c|\mathcal{D}, \mathbf{x}_*)$, with $c \in \{1, \ldots, C\}$. The minimum error of classification is achieved by choosing the class with the highest posterior probability. In the following, we develop two measures of uncertainty associated with the outcome y_*, geometry-based and homophily-based.

A. Geometry-Based Uncertainty

The most uncertain outcome of a classifier is that of no information when the classifier cannot make a decision and gives equal probabilities for all classes. Conversely, the most certain case is when all belief is put on only one class.

The space of the possible probabilities, $\mathbf{p}_*$, form a standard $(C-1)$-simplex, Δ^{C-1}, i.e.,

$$\Delta^{C-1} = \left\{ \mathbf{p}_* \in \mathbb{R}^C : \sum_{c=1}^{C} p_{*c} = 1, p_{*c} \geq 0 \text{ for } c = 1, \ldots C \right\} \tag{1}$$

The simplex center is the discrete uniform distribution, $\left[\frac{1}{C}, \ldots, \frac{1}{C}\right]^T$, corresponding to the most uncertain scenario, and vertices correspond to the cases of certainty, i.e., the permutations of the point $[1, 0, \ldots, 0]^T$. In Fig. 1, we present a standard 2-simplex.

Given that the case of certainty is the farthest point from the center and the total uncertainty is at the center, we quantify uncertainty by how far a classifier's outcome is from the center of the simplex. Accordingly, we define a geometry-based uncertainty as follows.

Definition 1: (Geometry-based uncertainty). Let $\mathbf{p}_*$ be a probability vector associated with a classifier's outcome y_*, and d a distance measure defined on the standard $(C-1)$-simplex, Δ^{C-1}, where C denotes the number of classes. The geometry-based uncertainty of y_* is,

$$\mathrm{GU}_{n|d}(y_*) \triangleq 1 - \left(\frac{d(\mathbf{p}_*, \mathbf{u}_C)}{d(\mathbf{e}, \mathbf{u}_C)} \right)^n \tag{2}$$

where n is a non-negative integer, $\mathbf{u}_C = \left[\frac{1}{C}, \ldots, \frac{1}{C}\right]^T$, and $\mathbf{e} = [1, 0, \ldots, 0]^T$.

This definition of uncertainty can be interpreted as the measure of how far is the classification output from a random
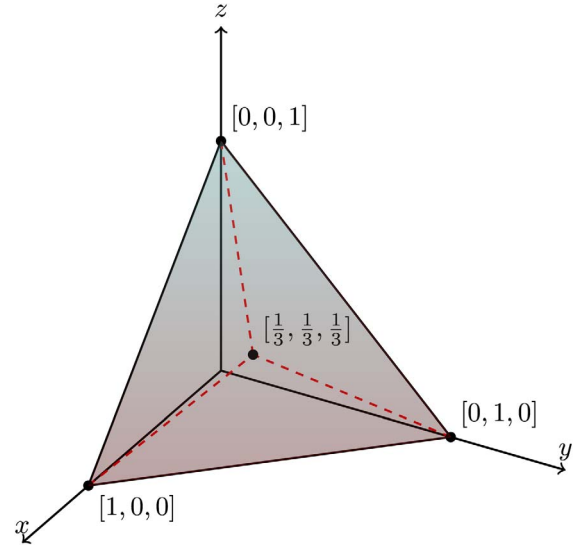

Fig. 1. The standard 2-simplex whose vertices are $[1, 0, 0]$, $[0, 1, 0]$, and $[0, 0, 1]$, and whose center is $\left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$.

classification (random guess). In the following, we discuss some of the properties of the geometric uncertainty.

Property 1: Let $\mathbf{p}_*$ be the probability vector associated with a classifier's outcome y_*, with n a non-negative integer, and d the distance measure defined on the standard $(C-1)$-simplex Δ^{C-1}, then $\mathrm{GU}_{n|d}(y_*)$
- is non-negative;
- is upper bounded by 1, i.e., $\mathrm{GU}_{n|d}(\mathbf{p}_*) \leq 1$;
- is maximized by the discrete uniform distribution;
- is minimized by the permutations of $[1, 0, \ldots, 0]^T$;
- increases with n.

Proof: The proof is straightforward ☐

In the definition of geometry-based uncertainty in (2), the distance d and n are the free parameters, thus corresponding to their individual assignments, we can obtain a particular geometric-based uncertainty. For instance, we can consider the Riemannian metric given by the Fisher-Rao distance, $d_{\mathrm{FR}}(\mathbf{p}_1, \mathbf{p}_2) = 2 \arccos\left(\sum_{c=1}^{C} \sqrt{\frac{p_{1c}}{p_{2c}}} \right)$, the Euclidean distance, $d_{\mathrm{E}}(\mathbf{p}_1, \mathbf{p}_2) = \sqrt{\sum_{c=1}^{C} (p_{1c} - p_{2c})^2}$, or the Kullback-Leibler divergence, $d_{\mathrm{KL}}(\mathbf{p}_1 \| \mathbf{p}_2) = \sum_{c=1}^{C} p_{1c} \log\left(\frac{p_{1c}}{p_{2c}}\right)$ [27], [28]. Based on the former distances, we present the following measures of geometric uncertainties,

Example 1: (Fisher-Rao uncertainty). Let $\mathbf{p}_*$ a probability vector associated with a classifier's outcome y_*, where $\mathbf{p}_* = [p_{*1}, \ldots, p_{*C}]^T$, C is the number of classes. The Fisher-Rao uncertainty of y_* is,

$$\mathrm{GU}_{n|\mathrm{FR}}(y_*) = 1 - \left(\frac{\arccos\left(\sum_{c=1}^{C} \sqrt{\frac{p_{*c}}{C}} \right)}{\arccos\left(\sqrt{\frac{1}{C}} \right)} \right)^n \tag{3}$$

where n is a non-negative integer.

Example 2: (Euclidean uncertainty). Let $\mathbf{p}_*$ a probability vector associated with a classifier's outcome y_*, where

$\mathbf{p}_* = [p_{*1}, \ldots, p_{*C}]^T$, C is the number of classes. The Euclidean uncertainty of y_* is,

$$\mathrm{GU}_{n|\mathrm{E}}(y_*) = 1 - \left(\frac{\sqrt{\sum_{c=1}^{C}(p_{*c} - \frac{1}{C})^2}}{\sqrt{1 - \frac{1}{C}}} \right)^n \qquad (4)$$

where n is a non-negative integer.

Remark 1: For $n = 2$, $\mathrm{GU}_{2|\mathrm{E}}$ is proportional to the Gini-Simpson index, also called Gini impurity [7]. Accordingly, it is also related to Tsallis entropy [21]. In fact, $\mathrm{GU}_{2|\mathrm{E}}(y_*) = \frac{C}{C-1}\left(1 - \sum_{c=1}^{C} p_{*c}^2\right)$, where the multiplicand of the right-hand side is the Gini-Simpson index which is also a special case of Tsallis entropy.

Example 3: (Kullback-Leibler uncertainty). Let $\mathbf{p}_*$ a probability vector associated with a classifier's outcome y_*, where $\mathbf{p}_* = [p_{*1}, \ldots, p_{*C}]^T$, C is the number of classes. The Kullback-Leibler uncertainty of y_* is,

$$\mathrm{GU}_{n|\mathrm{KL}}(y_*) = 1 - \left(\frac{\sum_{c=1}^{C} p_{*c} \log (C p_{*c})}{\log (C)} \right)^n \qquad (5)$$

where n is a non-negative integer.

Remark 2: It is straightforward to show that, for $n = 1$, the $\mathrm{GU}_{1|\mathrm{KL}}$ is the normalized Shannon entropy.

Please note that distances other than those considered in this work can be defined on Δ^{C-1}. Accordingly, other measures of uncertainty can be formulated. For instance, when we define the Rényi divergence, $d_{\mathrm{R}}(\mathbf{p}_1 \| \mathbf{p}_2) = \frac{1}{\alpha-1} \log \left(\sum_{c=1}^{C} \frac{p_{1c}^\alpha}{p_{2c}^{\alpha-1}} \right)$, on the simplex Δ^{C-1}, we identify the Rényi-entropy as a geometric uncertainty measure for $n = 1$ [20].

B. Homophily-Based Uncertainty

We derive the homophily-based uncertainty from the definition of variance, generally considered the standard measure of uncertainty. The variance of $p(y_* | \mathcal{D}, \mathbf{x}_*)$ can be written as,

$$\mathrm{Var}(y_*) = \frac{1}{2} \sum_{i=1}^{C} \sum_{j=1}^{C} p_{*i} p_{*j} (i - j)^2 \qquad (6)$$

where i and j refer to the ordinal encoding assigned to the C classes (ref. proof in Appendix A). Since the ordinal encoding is arbitrary and does not have a meaningful ranking of the classes, the difference $(i - j)^2$ in (6) is not adequate. In fact, for $C = 3$, the posterior probabilities $[0.5, 0.5, 0]$ and $[0.5, 0, 0.5]$ have variances 0.25 and 1, respectively. However, the difference $(i - j)^2$ can be interpreted as the squared Euclidean distance between classes i and j. Alternatively, and more generally, we model this difference by the square of the $d(i, j)$, where $d(i, j)$ denotes the distance between classes i and j.

In classification, several samples from each class are available, which can be used to compute $d(i, j)$. For instance, the pairwise distance between the classes can be quantified through similarity measures between their corresponding probability density functions, such as Wasserstein distance [29] or Energy distance [30]. Accordingly, we propose a homophily-based measure of uncertainty as follows,

Definition 2: (Homophily-based uncertainty). Let $\mathbf{p}_*$ a probability vector associated with a classifier's outcome y_*, where $\mathbf{p}_* = [p_{*1}, \ldots, p_{*C}]^T$, C is the number of classes. The homophily-based uncertainty of y_* is,

$$\mathrm{HU}(y_*) \triangleq \frac{\mathbf{p}_*^T (\mathbf{H} \odot \mathbf{H}) \mathbf{p}_*}{\mathbf{p}_{\max}^T (\mathbf{H} \odot \mathbf{H}) \mathbf{p}_{\max}} = \frac{\sum_{i=1}^{C} \sum_{j=1}^{C} p_{*i} p_{*j} d(q_i, q_j)^2}{\mathbf{p}_{\max}^T (\mathbf{H} \odot \mathbf{H}) \mathbf{p}_{\max}} \qquad (7)$$

where $.^T$ and $\odot$ denote the transpose operator and Hadamard product, respectively. Where $\mathbf{H} = (d(q_i, q_j))_{1 \leq i, j \leq C}$ and $d(q_i, q_j) \geq 0$ denotes the distance/similarity measure between the probability distributions q_i and q_j corresponding to classes i and j, respectively. And where $\mathbf{p}_{\max} = argmax\,(\mathbf{p}_*^T (\mathbf{H} \odot \mathbf{H}) \mathbf{p}_*)$.

Please note that $\mathbf{H}$, by definition, is a symmetric matrix with non-negative elements and zero values on the main diagonal. The denominator in (7) is a scaling factor so that $\mathrm{HU}(y_*)$ is in the interval $[0, 1]$. The probability distribution q_i is estimated using the data samples of class i. Please note that $\mathbf{p}_{\max}$ can be determined by quadratic optimization algorithms [31]. In the following, some of the properties of the homophily-based uncertainty.

Property 2: Let $\mathbf{p}_*$ a probability vector associated with a classifier's outcome y_*, then $\mathrm{HU}(y_*)$ is
- concave;
- positive;
- upper bounded by 1, i.e., $\mathrm{HU} \leq 1$;
- minimized by the permutations of $[1, 0, \ldots, 0]^T$.

Proof: ref. Appendix B □

The uncertainty in (7) is said *homophily*-based since it considers how close two classes are from a distribution point of view and, hence, how likely they are to be mistaken by a classifier. Note that the homophily-based uncertainty HU can be viewed as a weighted sum, where larger weights are given to distant classes. As such, mistaking distant/dissimilar classes will be assigned a higher uncertainty than mistaking close/similar classes. Accordingly, the homophily-based uncertainty quantifies how far a classifier is from a classifier that confuses distant classes.

Remark 3: If all classes are assumed equidistant, i.e., $d(q_i, q_j) = d(q_k, q_l)$, $\forall i, j, k, l \in \{1, \ldots, C\}$, $(i \neq j)$ and $(k \neq l)$, the homophily-based uncertainty equals $\mathrm{GU}_{2|\mathrm{E}}(y_*)$ (ref. proof in Appendix C).

The flowchart in Fig. 2 depicts the proposed measures of uncertainty and their relation to the existing ones in literature.

C. Gaussian Example

In order to understand the effect of classes' separability on uncertainty, we consider a two-dimensional dataset consisting of three Gaussian classes with means $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and $\boldsymbol{\mu}_3$. We assume that all classes share the same covariance matrix, $\boldsymbol{\Sigma}$, for the sake of simplicity. In this case, the posterior probability density of $\mathbf{x}$ is given by [7],
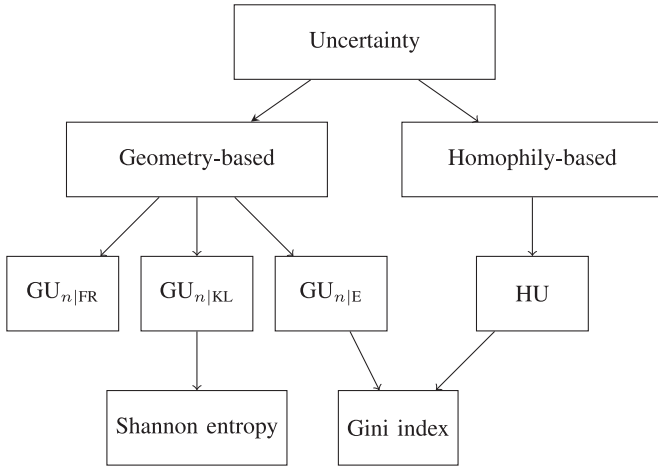
Fig. 2. Flowchart of the proposed uncertainty measures and their links to the existing ones in literature.

$$p(y = k|\mathbf{x}) = \left(1 + \exp\left((\boldsymbol{\mu}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{x} - \frac{\boldsymbol{\mu}_i + \boldsymbol{\mu}_k}{2}\right)\right)\right.$$
$$\left. + \exp\left((\boldsymbol{\mu}_j - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{x} - \frac{\boldsymbol{\mu}_j + \boldsymbol{\mu}_k}{2}\right)\right)\right)^{-1}$$

where $i, j, k \in \{1, 2, 3\}$ and $i \neq j \neq k$. The posterior probability depends on the differences between the means and $\boldsymbol{\Sigma}$ that characterize the class separation. It also depends on the position of the data point $\mathbf{x}$ with respect to the means, which assesses its closeness to the classes.

Let us consider the case $\boldsymbol{\mu}_1 = [0, 0]^T$, $\boldsymbol{\mu}_2 = [2, 0]^T$, $\boldsymbol{\mu}_3 = \alpha \times [1, 1]^T$ and $\boldsymbol{\Sigma} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix. In Fig. 3, we show the uncertainty values obtained using different uncertainty measures, $\text{GU}_{2|\text{FR}}$, $\text{GU}_{2|\text{E}}$ or normalized Gini-index, $\text{GU}_{1|\text{KL}}$ or normalized entropy, and HU for two data points, $\mathbf{x} = [1, 1]^T$ and $\mathbf{x} = [2.5, 2.5]^T$, as a function of α. The distance between classes required to calculate the homophily-based uncertainty in (7) is calculated using the infinity-norm between the means, i.e., $d(q_i, q_j) = ||\boldsymbol{\mu}_i - \boldsymbol{\mu}_j||_{\infty}$[1] for $i, j \in \{1, 2, 3\}$.

It may be observed that all geometric uncertainties have similar shapes but reflect different levels of uncertainty. The two peaks of the geometric uncertainties correspond to α values for which the data point is at an equal distance from the classes. The data point $\mathbf{x} = [1, 1]^T$ is at an equal distance from the three classes for $\alpha = 0$ and $\alpha = 2$, leading to a posterior probability $[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$. While $\mathbf{x} = [2.5, 2.5]^T$ is at an equal distance from the classes 2 and 3 for $\alpha = \frac{5 \pm \sqrt{13}}{2}$, leading to a posterior probability $[0, \frac{1}{2}, \frac{1}{2}]$. Please note that the point $[2.5, 2.5]^T$ is always closer to two of the three classes. It cannot be at an equal distance from the three classes. The uncertainty levels decrease when the data points get closer to one of the classes. Moreover, the curves flatten when the third class does not affect the estimation of the data point, i.e., the confusion only involves classes 1 and 2.

Compared to the geometric uncertainties, the homophily-based uncertainty shows different behavior. HU has two peaks but with varying levels of uncertainty. The first peak that implies

[1] $||\mathbf{z}||_{\infty} = ||[z_1, \ldots, z_N]^T||_{\infty} = \max(|z_1|, \ldots, |z_N|)$

closer classes, i.e., a low value of α, reflects a lower uncertainty than the second peak. Moreover, when class three deviates further from the two other classes, instead of flattening as the geometric uncertainties, HU decreases since the closeness of classes 1 and 2 relative to class 3 increases.

In summary, the maximum uncertainty is reached when the classifier cannot make a choice. This is the case when a data point is equidistant from two or more classes. This uncertainty increases with the number of classes involved. Moreover, the uncertainty based on homophily reflects different values for the same case of confusion, i.e., equal posterior probabilities, depending on whether this confusion concerns close or distant classes.

IV. EXPERIMENTAL EVALUATION AND ANALYSIS

This section demonstrates how geometry-based and homophily-based uncertainty measures can be used to assess classification quality. The measures of uncertainty we investigate are the binary variance, three of the geometry-based uncertainties $\text{GU}_{2|\text{FR}}$, the normalized entropy, i.e., $\text{GU}_{1|\text{KL}}$, and the normalized Gini-index, i.e., $\text{GU}_{2|\text{E}}$, in addition to the homophily-based uncertainty based on the Energy distance HU.

We explore several scenarios. In Subsection IV-A, we consider a remote sensing dataset and study the uncertainty of three classifiers. Subsection IV-B examines the uncertainty behavior across different noise levels in signal modulation data. Furthermore, Subsection IV-C investigates the impact of classes' separability on uncertainty measures in a medical imaging scenario. Finally, Subsection IV-D compares the proposed uncertainty measures and some of the existing ones.

The homophily-based uncertainty in (7) requires the calculation of $\mathbf{H}$, and we consider the Energy distance to this aim. Inspired by Newton's potential energy, Székely introduced the Energy distance in 1984 [30]. Let Q_i and Q_j be the cumulative distribution functions for classes i and j, respectively. In the case of a one-dimensional dataset, $\mathbf{H}$ based on the Energy distance is defined as:

$$\mathbf{H} = \left(\sqrt{\int ||Q_i(x) - Q_j(x)||^2 \, dx}\right)_{(i,j) \in \{1, \ldots, C\}^2} \quad (8)$$

This distance is implemented using the SciPy package [32]. In the case of a multidimensional dataset, we report the mean of the Energy distance over different channels plus the corresponding standard deviation to account for the variability between channels.

A. Trento

We consider the Trento dataset acquired over a rural area near the city of Trento, Italy [33]. It is composed of measurements from LiDAR and Hyperspectral imaging. The Optech ALTM 3100EA acquired the LiDAR data, while the hyperspectral data, consisting of 63 bands ranging from 402.89 to 989.09 nm, were obtained via the AISA Eagle sensor. Both datasets have a spatial resolution of 1 m and a size of 600×166 pixels.

(a) Visualization of the three classes.

(b) Entropy
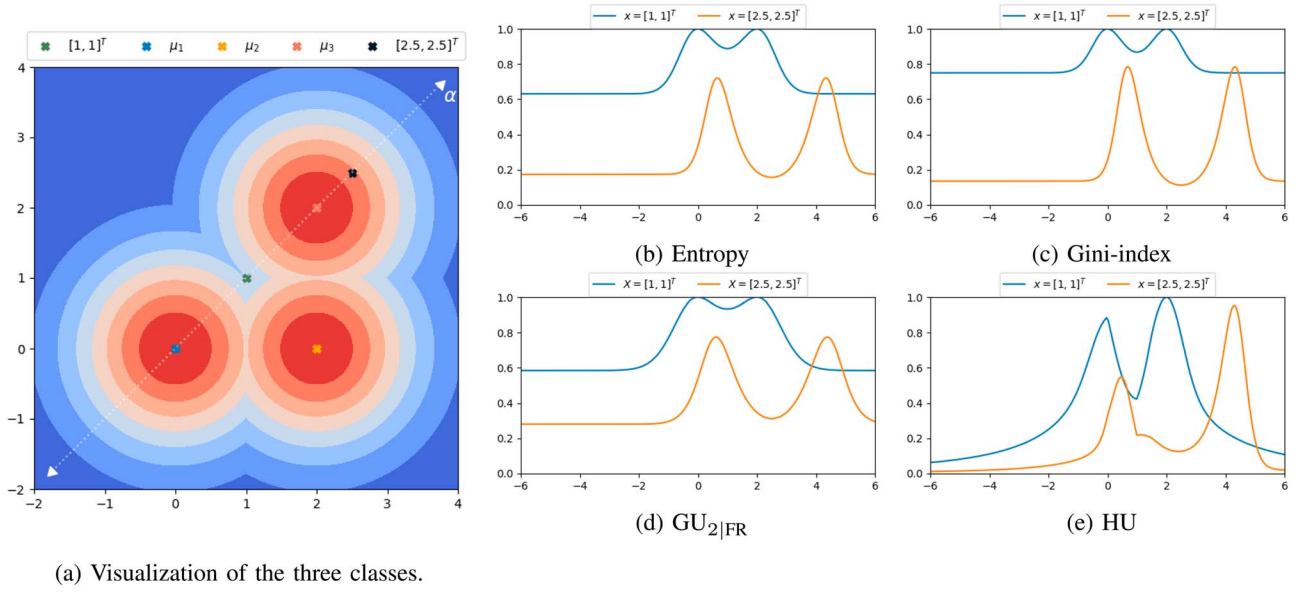
(c) Gini-index

(d) $GU_{2|FR}$

(e) HU

Fig. 3. Effect of class separation on Uncertainty. (a) Visualization of classes, their means, and the two data points of interest. (b)–(e) Uncertainty values obtained using the different metrics for two data points of interest as a function of α.

(a)

(b)

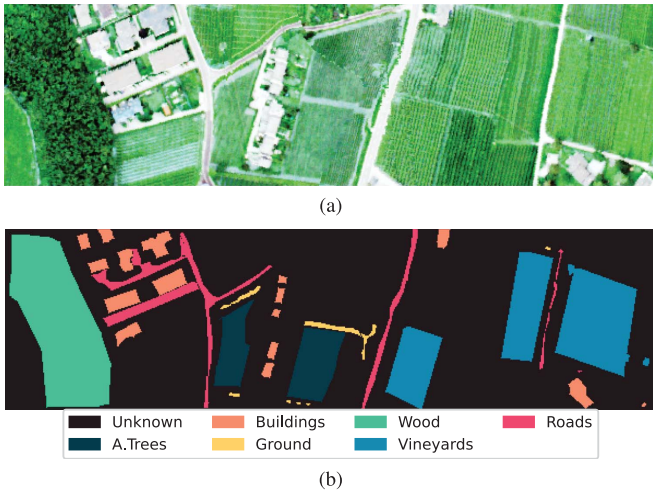| Unknown | Buildings | Wood |
| A.Trees | Ground | Vineyards | Roads |

Fig. 4. False color composite of Trento dataset (a), and its corresponding ground truth (b).

Six classes of interest were identified: Apple trees, Buildings, Ground, Wood, Vineyards, and Roads. A false-color composite of the hyperspectral data and the corresponding ground truth are shown in Fig. 4. The identified classes are summarized in Table V.

Equation (9) represents the normalized matrix obtained using Energy distance, defined in (8), for the Trento dataset.

$$
\mathbf{H}_{\text{Trento}} = \begin{bmatrix}
 & c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & \\
 & 0 & 0.89 & 0.58 & 0.35 & 0.36 & 0.88 & c_1 \\
 & 0.89 & 0 & 0.56 & 0.85 & \mathbf{1} & \mathbf{0.33} & c_2 \\
 & 0.58 & 0.56 & 0 & 0.6 & 0.73 & 0.65 & c_3 \\
 & 0.35 & 0.85 & 0.6 & 0 & 0.51 & 0.91 & c_4 \\
 & 0.36 & \mathbf{1} & 0.73 & 0.51 & 0 & 0.95 & c_5 \\
 & 0.88 & \mathbf{0.33} & 0.65 & 0.91 & 0.95 & 0 & c_6
\end{bmatrix}
$$

(9)

The maximum and minimum (non-zero) values of (9) are shown in blue and red, respectively. We observe that classes from the same category report a low cost, vegetation classes (c_4-Wood, c_1-Apple Trees, and c_5-Vineyards), and urban classes (c_2-Buildings and c_6-Roads). The maximum distance was achieved between c_2-Buildings and c_5-Vineyards.

We evaluate two classifiers: a kernel-based classifier, Support Vector Machine (SVM) [34], and a decision tree classifier, Random Forest (RF) [35]. The RF and SVM classifiers were implemented using the Sklearn Python package [36]. In what follows, we compare

- a sub-optimal SVM for which the hyperparameters have been set to their default value;
- the optimal SVM (OptSVM) for which the best hyperparameters have been identified using a search grid;
- RF with optimal hyperparameters (OptRF).

In Table I, we report the overall accuracy and kappa coefficient obtained for the three classifiers. The corresponding confusion matrices are conveyed in Tables II–IV.

Classification maps. Figs. 5–7 represent the classification maps predicted by the three considered models, SVM, OptSVM and OptRF, respectively along with the uncertainty maps obtained by the different considered measures.

In Fig. 5, we notice that c_2-Buildings are confused with c_6-Roads, c_1-Apple Trees with the c_3-Ground, and vegetation classes are mistaken for each other. The uncertainty measures reflect high confidence for some regions classified as c_4-Wood or c_5-Vineyards. Moreover, the binary variance exhibits the highest uncertainty for other pixels, while geometry-based uncertainties demonstrate moderate values. However, compared to the other measures, homophily-based uncertainty shows mostly a low uncertainty of the order of 0.2.

In Fig. 6, we observe that the optimal SVM fixed some confusion between classes, especially between c_2-Buildings and c_6-Roads. This is also reflected by the uncertainty measures, which

TABLE I
PERFORMANCE MEASURES OBTAINED ON
DIFFERENT CLASSIFIERS FOR TRENTO DATASET

	SVM		OptSVM		OptRF	
	OA	k	OA	k	OA	k
	87.3	83.1	92.9	90.5	96.4	95.2

TABLE II
CONFUSION MATRIX FOR TRENTO DATASET USING SVM

		Predicted Class					
		c_1	c_2	c_3	c_4	c_5	c_6
True class	c_1	80.6	0	0	0.1	19.3	0
	c_2	0	81.0	0	0	0	19.0
	c_3	3.3	0.2	95.2	0.2	0.6	0.4
	c_4	0.1	0	0	96.8	3.0	0
	c_5	8.7	0	0	3.1	88.2	0
	c_6	0.4	26.6	0.2	1.8	0.8	70.1

TABLE III
CONFUSION MATRIX FOR TRENTO DATASET USING
OPTSVM

		Predicted Class					
		c_1	c_2	c_3	c_4	c_5	c_6
True class	c_1	87.0	0	0	0	13.0	0
	c_2	0	87.9	0	0	0	12.0
	c_3	2.7	0	96.7	0.2	0.4	0
	c_4	0.2	0	0	98.6	1.2	0
	c_5	4.8	0	0	1.2	94.0	0
	c_6	0.1	14	0	1.4	0.6	83.8

TABLE IV
CONFUSION MATRIX FOR TRENTO DATASET USING OPTRF

		Predicted Class					
		c_1	c_2	c_3	c_4	c_5	c_6
True class	c_1	89.3	0	0	0	10.7	0
	c_2	0	97.8	0	0.1	0	2.1
	c_3	0.4	0	95.4	0	1.9	2.3
	c_4	0.1	0	0	99.9	0	0
	c_5	3.9	0	0	1.6	95.9	0
	c_6	0	1.7	0.2	0.9	1.1	95.9

TABLE V
THE IDENTIFIED CLASSES WITHIN TRENTO AND
BREAST CANCER DATASETS

	Trento Dataset	Breast Cancer Dataset
c_1	Apple trees	Tumor
c_2	Buildings	Stroma
c_3	Ground	Lymphocytic-infiltrate
c_4	Wood	Necrosis
c_5	Vineyards	Other
c_6	Roads	–

show more confidence in results than the suboptimal SVM. However, the homophily-based measure displays a very high uncertainty for pixels at the borders of classes.

In Fig. 7, we observe that the OptRF classifier shows more accurate predictions than the models based on SVM with less confusion between classes. Indeed, it reports the highest overall accuracy and kappa coefficient, followed by the OptSVM (ref. Table I). Nevertheless, it exhibits higher uncertainty, especially compared to OptSVM. Moreover, it exhibits more pixels with higher homophily-based uncertainty.

Classes similarity. In order to understand the behavior of homophily-based uncertainty, we compare it to its counterpart HU_{eq} that assumes equidistant classes. Accordingly, HU_{eq} only reflects the confusion between classes while HU includes the classes' similarity according to (9). Please recall that HU_{eq} corresponds to the normalized Gini-index.

In Fig. 8, we show the mean and standard deviation of the homophily-based uncertainties, HU_{eq} and HU obtained for the predictions misclassified as c_2-*Buildings*, c_4-*Wood* or c_5-*Vineyards* while the correct class c_6-*Roads* comes as the second best class.

We notice that HU reflects lower values than HU_{eq} when misclassifying c_6-*Roads* as c_2-*Buildings*. Conversely, it reflects slightly higher values than HU_{eq} when misclassifying c_6-*Roads* as c_4-*Wood* or c_5-*Vineyards*. Indeed, according to equation (9), c_2-*Buildings* is the closest to c_6-*Roads* while c_4-*Wood* or c_5-*Vineyards* are the more distant to c_6-*Roads*. This confirms that HU gives more weight to distant classes, reflecting lower values than HU_{eq} for closer classes and higher values than HU_{eq} for distant classes.

Number of confused Classes. The number of confused classes reflects the indecisiveness of the classifier. The greater this number, the closer we get to a random guess, and the higher should be the uncertainty. In Fig. 9, we report the different uncertainty measures as a function of the confused classes for the three considered models. The number of confused classes is determined as the number of classes with a probability greater than $\frac{1}{C}$. All uncertainty measures increase with the number of confused classes except for variance. Variance decreases in the case of four confused classes. In fact, variance decreases if the probabilities of the less likely classes sum up to a value greater than the probability of the most likely class. For instance, the posterior probabilities $[0.35, 0.25, 0.25, 0.15]$ and $[0.65, 0.35, 0, 0]$ report the same binary variance, while it is clear that the first case is more uncertain. Indeed, variance only considers the classifier's confidence, i.e., the highest posterior probability. Accordingly, it rejects information on confusion and only evaluates a binary scenario. As a conclusion of this test, variance is not a good measure of uncertainty.

Empirical cumulative distributions. We compare the distribution of uncertainty measures in the case of correct and incorrect predictions. In Fig. 10, we represent the empirical cumulative distribution functions using different uncertainty measures obtained on the predictions using SVM, OptSVM, and OptRF. We observe that all uncertainty measures report lower values for correct predictions than for wrong predictions. Moreover, at first glance, variance seems to be the best representation of uncertainty since it yields the lowest values for erroneous predictions. However, as explained earlier, variance does not incorporate information on the confusion, which might yield a wrong estimation of uncertainty. Accordingly, since Gini-index reports the highest values for the wrong predictions for all models, we deem it the best measure of uncertainty among the geometry-based measures.
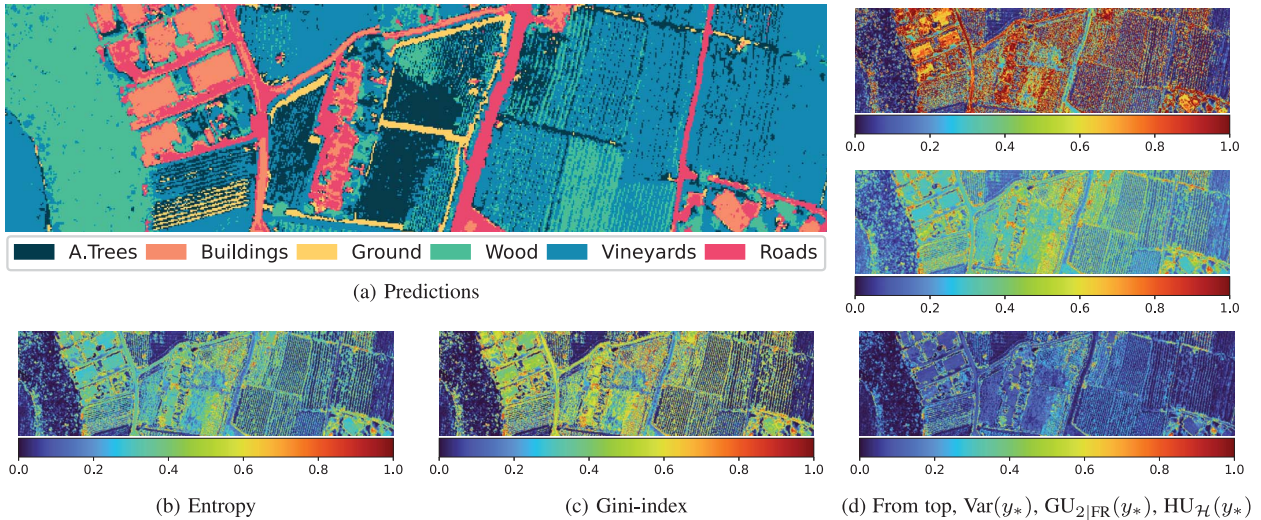
(a) Predictions

(b) Entropy

(c) Gini-index

(d) From top, $\text{Var}(y_*)$, $\text{GU}_{2|\text{FR}}(y_*)$, $\text{HU}_{\mathcal{H}}(y_*)$

Fig. 5. Classification results of Trento dataset using SVM (a) and its corresponding uncertainties (b)–(d).

(a) Predictions

(b) $\text{GU}_{1|\text{KL}}(y_*)$ (entropy)

(c) $\text{GU}_{2|\text{E}}(y_*) = \text{HU}_{\text{eq}}(y_*)$

(d) From top, $\text{Var}(y_*)$, $\text{GU}_{2|\text{FR}}(y_*)$, $\text{HU}_{\mathcal{H}}(y_*)$
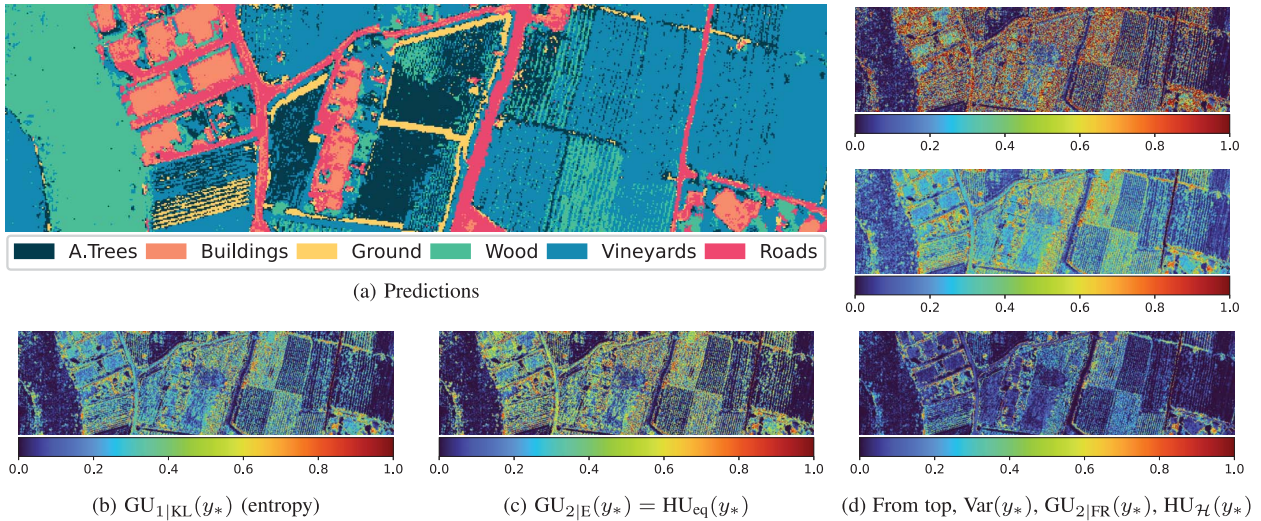
Fig. 6. Classification results of Trento dataset using OptSVM (a) and its corresponding uncertainties (b)–(d).

Model quality. We investigate the quality of the three considered models based on uncertainty. By comparing the geometric-based measures of uncertainty in Fig. 10, we notice that all models report lower uncertainties for their correct predictions than their incorrect predictions. Moreover, OptRF conveys higher uncertainties for its wrong predictions, followed by SVM and OptSVM. Accordingly, OptRF can be deemed as more trustworthy.

However, please note that OptRF reports a homophily-based uncertainty larger than 0.5 for almost 8% of the wrong estimates as opposed to SVM and OptSVM, which convey this high uncertainty for only 1% and 1.5% of the estimates, respectively. Accordingly, we can assume that random forest, as opposed to SVM, confuses between distant classes. Indeed, we find some pixels classified as c_6-*Roads* in the regions classified as c_5-*Vineyards*. This presumption is confirmed by the confusion matrices of the three models presented in Tables II–IV. But overall, since homophily-based uncertainty reports low values

for a large proportion of the estimates, we can conclude that most of the confusions involve close classes.

According to the aforementioned analysis, we can conclude that OptSVM is a classifier that confuses similar classes, while OptRF confuses between distant classes. However, OptRF shows higher uncertainty regarding misclassifications. The confusion in OptSVM is probably due to the under-exploitation of the height information from Lidar data, given the imbalance in the dimensionalities of both modalities. Please note that OptRF automatically overcomes this limitation since it uses a single feature for node splitting.

Outliers. As mentioned before, confusing two close classes is not alarming since they are more challenging to separate from a data perspective. However, mistaking two distant classes is a red flag and might be the symptom of a bad classifier, an instance of wrong labeling, or the presence of outliers, etc. In order to understand how homophily-based uncertainty can be used to detect such problematic instances, we consider a data
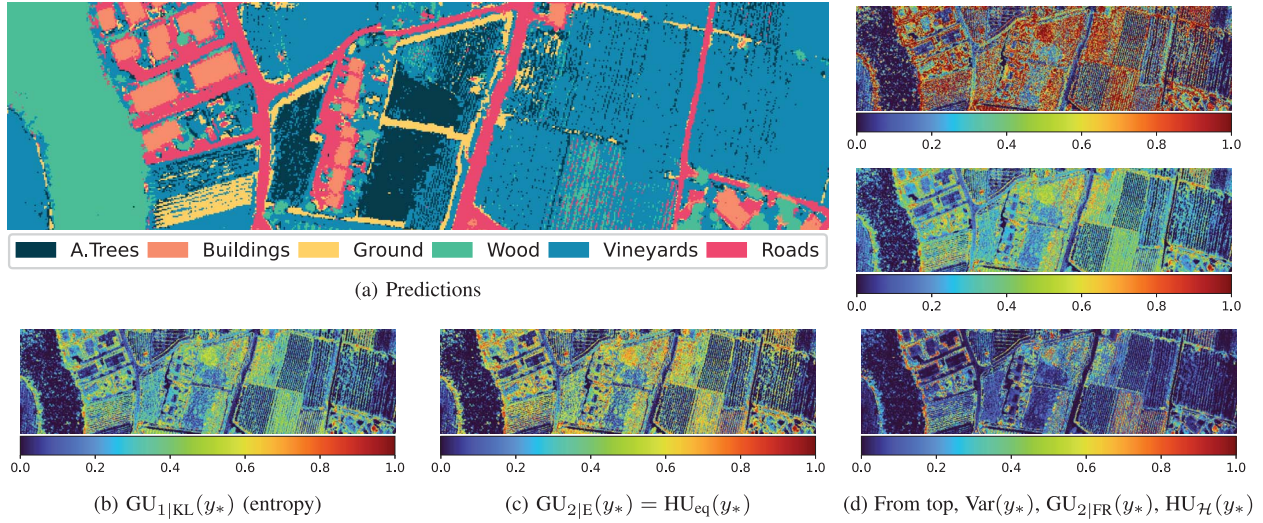
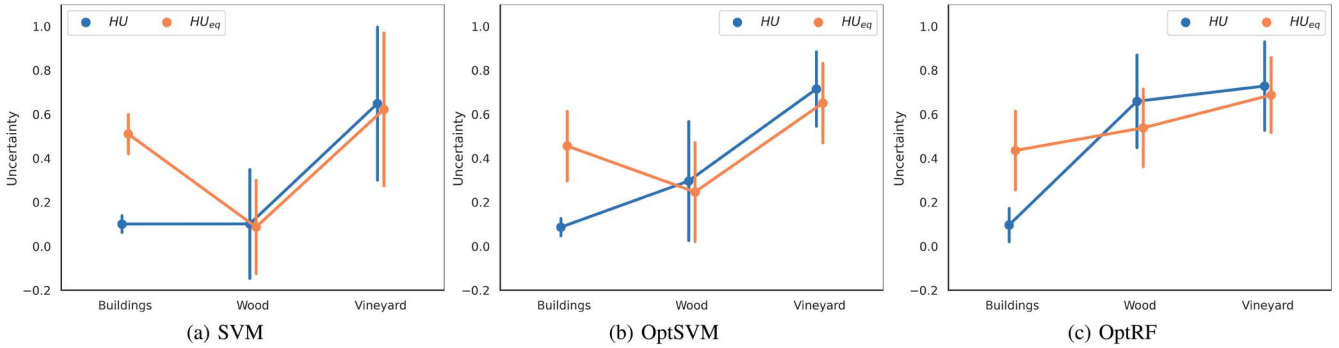Fig. 7. Classification results of Trento dataset using OptRF (a) and its corresponding uncertainties (b)–(d).

Fig. 8. Mean and standard deviation of homophily-based uncertainties obtained for data points labeled as c_6-*Roads* that were misclassified as c_2-*Buildings*, c_4-*Wood* or c_5-*Vineyards* while the correct class comes as the second best class.
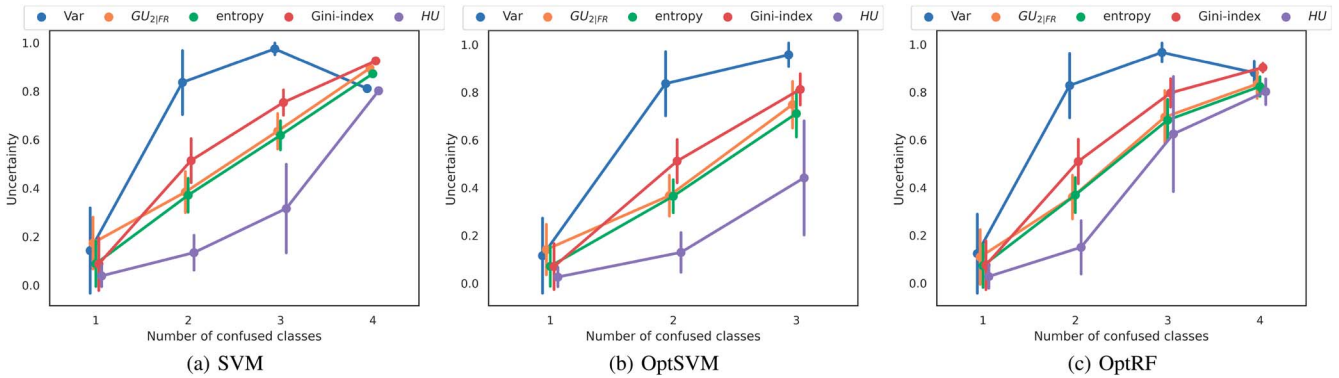

Fig. 9. Different measures of uncertainty as a function of the number of confused classes. The confused classes are identified as the classes with a posterior probability larger than $\frac{1}{C}$.

point that shows a homophily-based uncertainty greater than 0.8 by the three considered models. Recall that a high homophily-based uncertainty reflects confusion between distant classes. In Fig. 11, we show the mean of each class's hyperspectral signature and the corresponding standard deviation presented by a shaded region. The mean and standard deviation are calculated using only data points of high density to exclude outliers, if any. Fig. 11(b)–(d) represents the spectral signature of the considered data point. The signatures of the classes are

presented with different levels of opacity that reflect the magnitude of the estimated posterior probability by each classifier. The higher the posterior probability, the more opaque is the shaded region. Please note that the spectral signature of the considered data point differs from the classes' signatures, which might signal a new class or an outlier. This point's spectral signature is closer to the vegetation classes in the visible spectrum and closer to the urban classes in the infrared spectrum. Indeed, SVM and OptSVM confuse mainly between
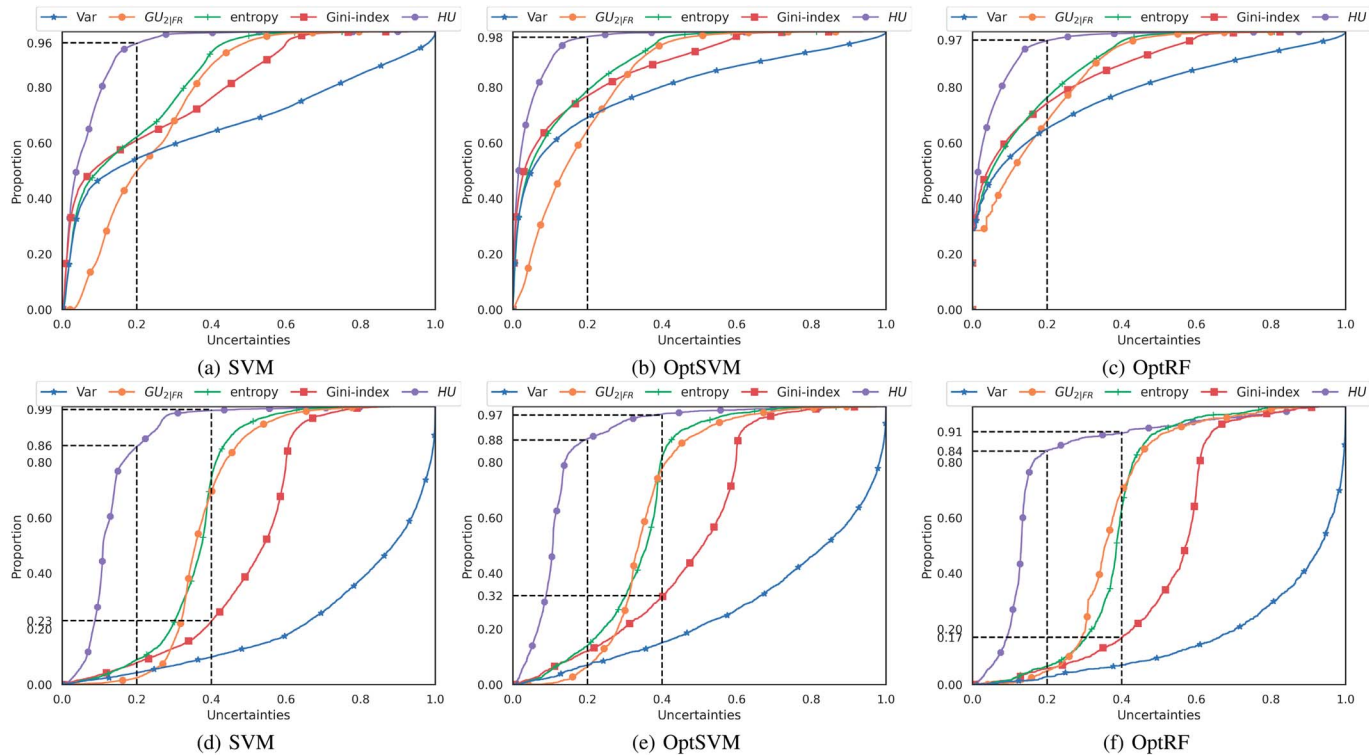
Fig. 10. Empirical cumulative distribution functions of different uncertainty measures of the classification estimates obtained using SVM, OptSVM, and OptRF: (top) right predictions, (bottom) wrong predictions.

TABLE VI
THE IDENTIFIED CLASSES WITHIN THE SMC DATASET

SMC Dataset			
c_1	16QAM	c_7	DSB-AM
c_2	64QAM	c_8	GFSK
c_3	8PSK	c_9	PAM4
c_4	B-FM	c_{10}	QPSK
c_5	BPSK	c_{11}	SSB-AM
c_6	CPFSK		

c_6-*Roads* and c_4-*Wood*, while OptRF confuses between c_6-*Roads*, c_4-*Wood* and c_2-*Vineyard*. Please note that the considered data point is mislabeled as c_6-*Roads*.

B. Signal Modulation

Previously, we studied the impact of the number and type of confused classes on uncertainty measures. We have also demonstrated how these measures can assess a classifier's quality and detect problematic instances, such as outliers. Now, we consider the Signal Modulation Classification (SMC) task and investigate the impact of noise and data drift on uncertainty.

We generate a synthetic dataset using the MATLAB code from [37]. The generated waveforms are impaired with additive Gaussian noise, with signal-to-noise ratio (SNR) taking either 15dB or 50dB values. The waveforms are passed through a Rician multipath fading channel with a path delay of $[0, 1.8, 3.4]$ samples with the corresponding average path gains of $[0, -2, -10]$ dB. The K-factor equals 4, and the maximum Doppler shift is set to 4Hz, equivalent to a walking speed at 906 MHz carrier frequency. This dataset includes eleven classes,

eight digital and three analog modulation types. Each waveform is represented by a frame that consists of 1024 samples and has a sample rate of 200 kHz. We consider center frequencies of 902 MHz and 100 MHz for the digital and analog modulation types, respectively. The classes include Binary Phase Shift Keying (BPSK), Quadrature Phase Shift Keying (QPSK), 8-ary Phase Shift Keying (8-PSK), 16-ary Quadrature Amplitude Modulation (16-QAM), 64-ary Quadrature Amplitude Modulation (64-QAM), 4-ary Pulse Amplitude Modulation (PAM4), Gaussian Frequency Shift Keying (GFSK), Continuous Phase Frequency Shift Keying (CPFSK), Broadcast FM (B-FM), Double Sideband Amplitude Modulation (DSB-AM), and Single Sideband Amplitude Modulation (SSB-AM). Table VI summarizes the identified classes within this dataset. For more details on SCM, please refer to the review articles [38], [39], [40].

A convolutional neural network (CNN) is utilized for modulation classification as suggested in [37]. The CNN architecture comprises six convolution layers and one fully connected layer. Except for the last convolution layer, each layer is followed by a batch normalization layer, a rectified linear unit activation layer, and a max pooling layer. An average pooling layer replaces the max pooling layer in the final convolution layer. The output layer incorporates softmax activation to provide scores for each label. To obtain probabilities for the respective labels, we apply isotonic calibration [41].

Noise effect. In order to study the effect of noise on the uncertainty measures, we consider two SMC datasets with different SNRs, specifically, 50dB and 15dB.

In Fig. 12(a) and 12(b), we represent the T-distributed Stochastic Neighbor Embedding (t-SNE) of the features

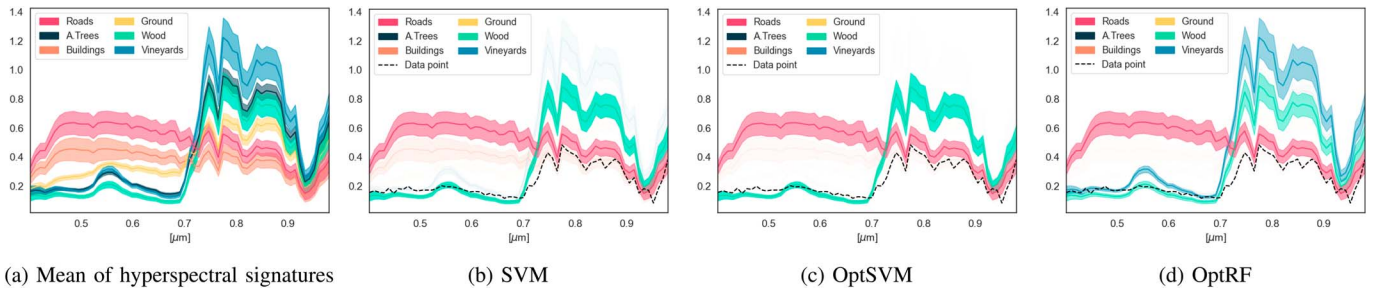(a) Mean of hyperspectral signatures   (b) SVM   (c) OptSVM   (d) OptRF

Fig. 11. Mean, bold lines, and standard deviation, shaded area, of the different classes' spectral signature. The opacity of the spectral signatures in subfigures (b)–(d) reflects the posterior probability estimates obtained by different classifiers considering the data point in the dashed line.

extracted using the considered deep learning model on both datasets [42]. We notice that, as expected, a reduced noise level tends to enhance the distinctiveness and separability of the classes. Indeed, the dataset with an SNR of 15dB shows an overlap between several classes where some totally overlap, such as c_3-*8PSK* and c_{10}-*QPSK*. Conversely, the dataset with an SNR of 50dB shows better separation between classes where few classes partially overlap, such as c_1-*16QAM* and c_2-*64QAM*.

Fig. 12(c)–(f) depicts the estimated empirical density functions of various uncertainty measures associated with the wrong and correct predictions of the modulation test data, considering two different noise levels. The uncertainty measures are calculated based on the classifiers' predictions, specifically on the test set, using the same noise level they were trained on. For correct and wrong predictions, we notice that a larger portion of data points exhibits higher uncertainty at an SNR of 15 dB compared to the 50dB scenario. For instance, almost 15% of the misclassified data points exhibit a Gini-index strictly larger than 0.6 for an SNR of 15dB as opposed to only 3% for an SNR of 50dB. Moreover, 66% of the misclassified data points exhibit a homophily-based uncertainty strictly larger than 0.15 for an SNR of 15dB as opposed to only 0.3% for an SNR of 50dB. Accordingly, the considered measures effectively capture the effect of noise on classification quality. Moreover, the homophily-based uncertainty properly highlights the decline in classes' separability due to the noise.

Data drift. We consider the case of data drift when the test data is different from the training data. We study a scenario where the CNN is trained only on data with an SNR of 50 dB (resp. 15dB) and which has been tested with data having SNR values of 15 dB (resp. 50dB).

In Fig. 13(a) and 13(b), we represent the t-SNE of features extracted using the considered deep learning model for both scenarios. t-SNE was calculated using training and test datasets. It is obvious that the scenario where the data drifts from 50dB to 15dB is more difficult because it shows a lot of overlap between the classes. This outcome is anticipated since the test set is more problematic than the training set. Conversely, in the case of a data drift from 15dB in training to 50dB in testing, the classifier has been trained on more complicated data.

The uncertainty measures shown in Fig. 13(c)–(f) for correct and incorrect predictions appropriately reflect the data drift impact. For example, the Gini index has an uncertainty greater

(a) t-SNE   (b) t-SNE

(c) Correct predictions   (d) Correct predictions
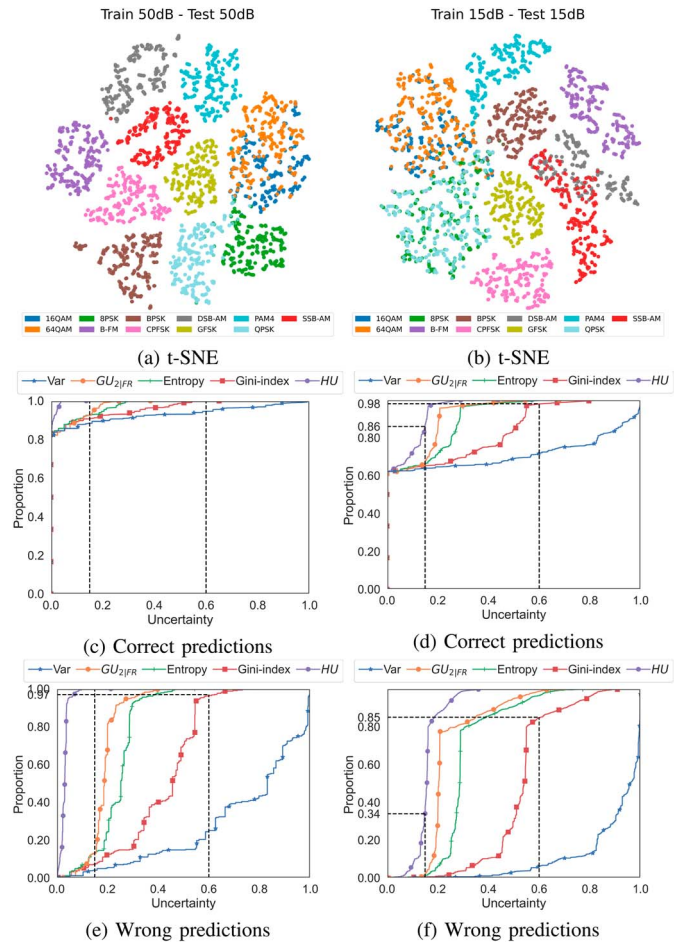
(e) Wrong predictions   (f) Wrong predictions

Fig. 12. Effect of noise. (Top) t-SNE of the features extracted by the CNN model. (Middle) Empirical cumulative distribution functions of different uncertainty measures obtained on the correct predictions. (Bottom) Empirical cumulative distribution functions of different uncertainty measures obtained on the wrong predictions. The left column corresponds to the model trained and tested on 50dB data. The right column corresponds to the model trained and tested on 15dB data.

than 0.6 for more than 97% of the data points in the complicated scenario, compared to only 13% in the case of a drift from 15dB in training to 50dB in testing. Moreover, the uncertainty based on homophily shows an uncertainty greater than 0.2 for almost 99% of the data points, compared to only 5% in the more manageable scenario.

(a) t-SNE

(b) t-SNE

(c) Correct predictions

(d) Correct predictions

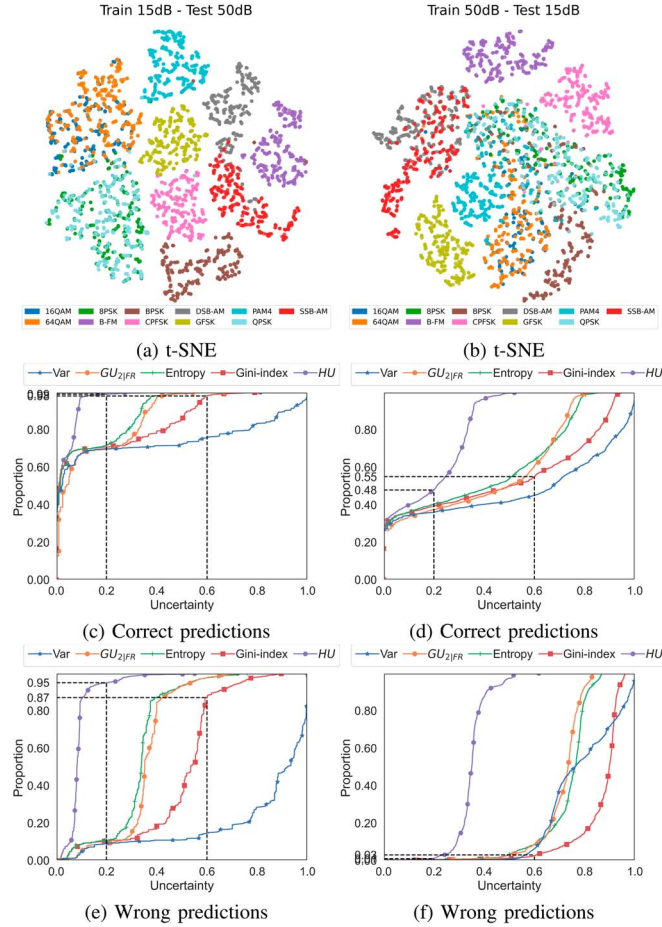(e) Wrong predictions

(f) Wrong predictions

Fig. 13. Data drift. (Top) t-SNE of the features extracted by the CNN model. (Middle) Empirical cumulative distribution functions of different uncertainty measures obtained on the correct predictions. (Bottom) Empirical cumulative distribution functions of different uncertainty measures obtained on the wrong predictions. The left column corresponds to the model trained on 15dB data and tested on 50dB data. The right column corresponds to the model trained on 50dB data and tested on 15dB data.
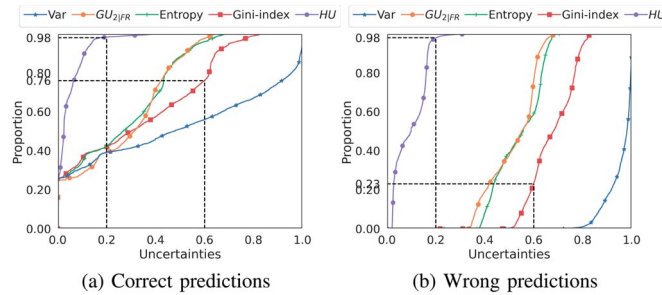

(a) Correct predictions

(b) Wrong predictions

Fig. 14. Empirical cumulative distribution functions of different uncertainty measures obtained on the correct predictions (left) and incorrect predictions (right) for the data points labeled as c_3-Lymphocytic-infiltrate.

C. BCSS

Using the Breast Cancer Semantic Segmentation (BCSS) dataset, we investigate classes' separability effect on uncertainty. This dataset consists of hematoxylin and eosin-stained whole-slide images (WSIs) corresponding to a case of
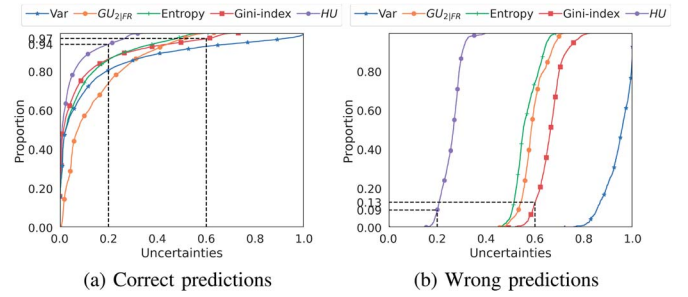

(a) Correct predictions

(b) Wrong predictions

Fig. 15. Empirical cumulative distribution functions of different uncertainty measures obtained on the correct predictions (left) and incorrect predictions (right) for the data points labeled as c_4-Necrosis.

histologically-confirmed breast cancer. This image of formalin-fixed paraffin-embedded tissues was acquired from the Cancer Genome Atlas, with triple-negative status determined from clinical data files. The original dataset consists of 151 images; in this work, we consider the region of interest within the image "TCGA-D8-A1JG-01Z-00-DX1.BA6D5CC7-3A9B-4D17-A86A-B159D345A216" [43]. Five classes of interest were identified: Tumor, Stroma, Lymphocytic-infiltrate, Necrosis, and a class Other comprising other tissue types of no interest. Table V summarizes the identified classes within this dataset. Equation (10) represents the normalized classes' distance matrix obtained using Energy distance on the BCSS dataset.

$$\mathbf{H}_{\text{BCSS}} = \begin{matrix} & c_1 & c_2 & c_3 & c_4 & c_5 \\ \begin{bmatrix} 0 & 0.46 & 0.34 & 0.51 & 0.74 \\ 0.46 & 0 & 0.26 & 0.8 & 0.33 \\ 0.34 & 0.26 & 0 & 0.55 & 0.47 \\ 0.51 & 0.8 & 0.55 & 0 & 1 \\ 0.74 & 0.36 & 0.47 & 1 & 0 \end{bmatrix} & \begin{matrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{matrix} \end{matrix} \quad (10)$$

The maximum and minimum (non-zero) values of (10) are shown in blue and red, respectively. The minimum distance is between c_2-Stroma and c_3-Lymphocytic-infiltrate classes while the largest distance is between c_4-Necrosis and c_5-Other.

In the following, we use OptRF to classify the BCSS dataset. We analyze the uncertainty distributions of correct and wrong predictions of classes c_3-Lymphocytic-infiltrate and c_4-Necrosis. We represent the corresponding empirical cumulative functions in Figs. 14 and 15.

We observe that 24% of the accurate predictions for the c_3-Lymphocytic-infiltrate class exhibit a Gini index greater than 0.6. In contrast, only 3% of accurate predictions for the c_4-Necrosis class surpass this threshold. Furthermore, 77% of the incorrect predictions for the c_3-Lymphocytic-infiltrate class have a Gini index greater than 0.6, while 87% of misclassifications for the c_4-Necrosis class fall into this category.

What is particularly interesting to note is that while the homophily-based uncertainties of accurate predictions for both classes follow similar trends, the homophily-based uncertainties for misclassifications are significantly different. In fact, only 6% of misclassifications for c_3-Lymphocytic-infiltrate have a homophily uncertainty larger than 0.2, whereas more than 91% of misclassifications for c_4-Necrosis have a homophily-based

TABLE VII
COMPARISON OF UNCERTAINTY MEASURES

		Gini-Index		Entropy		$GU_{2|FR}$		HU		t-Entropy $\alpha=1$		Rényi-Entropy $\alpha=2$		Tsallis-Entropy $\alpha=1.5$	
		μ	γ_1	μ	γ_1	μ	γ_1	μ	γ_1	μ	γ_1	μ	γ_1	μ	γ_1
								Trento dataset							
OptSVM	Correct	0.12	1.58	0.1	1.38	0.17	0.96	0.04	4.6	0.11	1.62	**0.07**	**1.93**	0.11	1.53
	Wrong	**0.45**	**-0.71**	0.33	0.08	0.35	1.19	0.12	3.45	0.4	-0.54	0.28	-0.1	0.39	-0.51
OptRF	Correct	0.13	1.43	0.11	1.27	0.14	0.96	0.04	4.53	0.12	1.47	**0.07**	**1.81**	0.12	1.39
	Wrong	**0.52**	**-1.06**	0.38	0.4	0.37	1.25	0.18	2.71	0.46	-0.69	0.33	-0.04	0.45	-0.57
ResNet50	Correct	0.005	11.1	0.005	9.39	0.01	5.76	0.002	21.27	0.004	11.13	**0.003**	**12.28**	0.005	10.65
	Wrong	**0.33**	-0.11	0.24	-0.27	0.24	**-0.44**	0.1	3.17	0.29	-0.1	0.2	0.07	0.28	-0.16
								SCM dataset - 50dB							
CNN	Correct	0.04	3.21	**0.02**	2.95	**0.02**	2.52	0.003	5.99	0.03	3.23	**0.02**	**3.5**	0.03	3.14
	Wrong	**0.41**	**-0.8**	0.24	-0.27	0.19	0.1	0.03	4.26	0.35	-0.65	0.2	-0.24	0.32	-0.67
								SCM dataset - 15dB							
CNN	Correct	0.16	0.93	0.09	1.14	**0.08**	**1.4**	0.05	1.15	0.14	0.99	**0.08**	1.25	0.13	0.99
	Wrong	**0.53**	**0.1**	0.31	1.81	0.24	2.1	0.16	1.19	0.46	0.77	0.28	1.42	0.42	1
								BCSS dataset							
OptRF	Correct	0.47	-0.74	0.39	-0.67	0.41	-0.69	0.11	0.17	0.43	-0.67	**0.33**	**-0.48**	0.43	-0.7
	Wrong	**0.68**	-0.37	0.55	-0.53	0.53	**- 0.82**	0.15	-0.22	0.62	-0.28	0.49	-0.13	0.61	-0.35

uncertainty larger than 0.2. This indicates that c_4-*Necrosis* is often confused with distant classes, in contrast to c_3-*Lymphocytic-infiltrate*. This discrepancy can be attributed to the fact that c_3-*Lymphocytic-infiltrate* is closely related to other classes, resulting in low distances between them, while c_4-*Necrosis* exhibits relatively larger distances. This demonstrates the capacity of homophily-based uncertainty to capture the type of confusion as opposed to the geometry-based metrics that show relatively comparable trends for both classes in the case of misclassifications.

D. Comparison

We compare the proposed measures with some of the existing ones. Namely, the Rényi entropy $\frac{\log\left(\sum_{c=1}^{C} p_{*c}\right)}{1-\alpha}$ [20], the Tsallis entropy, $\frac{k}{1-\alpha}\left(1 - \sum_{c=1}^{C} p_{*c}^{\alpha}\right)$ [21], and the t-entropy $\sum_{c=1}^{C} p_{*c} \tan^{-1}\left(p_{*c}^{-\alpha}\right) - \frac{\pi}{4}$ [22]. All the measures are scaled to the interval $[0, 1]$.

In Table VII, we report the average, μ, and skewness, γ_1, of different uncertainty measures corresponding to correct and erroneous predictions of several models and different datasets. We observe that all measures report a larger uncertainty for the wrong predictions than the correct ones. Specifically, Rényi entropy reports the lowest uncertainty for the correct predictions, while Gini-index reflects the largest values for the incorrect ones. Moreover, the largest margin between the averages corresponding to the correct and wrong predictions is ensured by the Gini-index followed by t-entropy.

The skewness coefficient γ_1 quantifies the asymmetry of a probability distribution. A positive skewness implies that a large portion of the data points lies on the left. Conversely, a negative skewness indicates that a large portion of the data points lies on the right. Accordingly, we believe a good measure of uncertainty should convey large absolute skewness values. Moreover, a positive skew is preferred for correct predictions since it implies that numerous data points have low uncertainty. Conversely, a negative skew is desired for the wrong predictions, indicating that most data points exhibit more significant uncertainty. The uncertainty measures that respect these requirements mostly are Gini-index, t-entropy, and Tsallis entropy. In the case of the SMC dataset with 15dB SNR, all measures report a positive skew for incorrect predictions. However, Gini-index reports the lowest value. Therefore, Gini-index is deemed the best measure of uncertainty among the ones compared here.

Furthermore, we observe that the homophily-based measure shows high skew values for all models and datasets except for the wrong predictions of BCSS. High skew values imply that the confusion mainly involves close classes. Moreover, the homophily-based uncertainty shows the highest average for the correct predictions of the BCSS dataset. Accordingly, BCSS dataset classification confuses distant classes in case of correct and wrong predictions, implying a low separability of the classes. This is also reflected by the other uncertainty measures exhibiting larger values than for other datasets.

By comparing the SMC datasets with 50dB and 15DB, we observe that the measures that show a significant increase in uncertainty are Gini-index, homophily-based uncertainty, t-entropy, and Tsallis entropy. Moreover, By comparing all models, we notice that ResNet50 shows the lowest uncertainty values for its correct and wrong predictions, making this classifier less reliable.

V. CONCLUSION

In this article, we proposed two measures of classification uncertainty: geometry-based and homophily-based. The geometry-based uncertainty is a function of the distance of class probabilities to the center of the feasible space of probabilities, given by the uniform distribution. In contrast, homophily-based uncertainty is a function of the average pairwise distances between classes.

We derive different measures of geometry-based uncertainty depending on different distances on the standard simplex (space of feasible probabilities), the Euclidean distance, the Kullback-Leibler distance, and the Fisher-Rao distance. The derived uncertainties quantify how far the classifier is from the random guess but behaves differently. According to our analysis, Euclidean-based uncertainty, i.e., Gini-index, is more suitable for assessing uncertainty in classification.

Moreover, we derived a homophily-based uncertainty that accounts for the separability of the classes. Accordingly, for the same set of probabilities, it reflects lower or larger uncertainties whether the confusion involves close or separate classes.

By combining homophily-based and geometry-based uncertainties, we demonstrated how to reveal relevant details on the model and data in the experimental section. Instead of evaluating the number of correct classifications, uncertainty measures assess the quality of a classifier by understanding,

- How *certain* is the classifier about the correct classifications?
- How *uncertain* is the classifier about misclassifications?
- What *type* of classes is the classifier mistaking?

Mistaking close classes is a sign of low separability. In contrast, mistaking distant classes is a sign of a bad model, noisy data points, or the existence of outliers. The proposed uncertainties are not aggregated, i.e., they are specific to each prediction and do not require a ground truth. The conclusions that can be drawn using the proposed uncertainties are confirmed by the confusion matrices resulting from the classification operation.

APPENDIX A
REFORMULATION OF VARIANCE

Gini in [44] noted that variance can be formulated as distance between data points,

$$\text{Var} = \sum_{i=1}^{C} i^2 p_{*i} - \sum_{i,j=1}^{C} p_{*i} p_{*j} ij = \frac{1}{2} \sum_{i,j=1}^{N} p_{*i} p_{*j} \left(i - j \right)^2$$

APPENDIX B
CONCAVITY OF HU

Consider the function $f(\mathbf{p}) = \mathbf{p}^T \left(\mathbf{H} \odot \mathbf{H} \right) \mathbf{p}$, where $\mathbf{p}$ is a probability vector, i.e., all its elements are non-negative and sum up to one. Using that f is non-negative (sum of non-negative terms, $f(\mathbf{p}) = \sum_{i=1}^{C} \sum_{j=1}^{C} p_i p_j \omega_{ij}^2$), it is straightforward to prove that $f((1-\alpha)\mathbf{p}_1 + \alpha\mathbf{p}_2) \geq (1-\alpha)f(\mathbf{p}_1) + \alpha f(\mathbf{p}_2)$, $\forall \alpha \leq 1$. Using Brauer minimum principle [45], and since the simplex Δ^{C-1} is convex and compact by construction, the minimum of HU is attained at the vertices, i.e., permutations of $[1, 0, \ldots, 0]^T$.

APPENDIX C
CONNECTION BETWEEN GEOMETRY-BASED AND HOMOPHILY BASED UNCERTAINTIES

In the case of equidistant classes, the homophily-based uncertainty writes,

$$\text{HU}_{\text{eq}}(y_*) \triangleq \frac{2C}{C-1} \sum_{i=1}^{C} \sum_{j>i}^{C} p_{*i} p_{*j} \qquad (11)$$

By substituting p_{*1} by $(1 - \sum_{i=2}^{C} p_{*i})$ in (4) and (11), we find that

$$\text{GU}_{2|\text{E}}(y_*) = \frac{2C}{C-1} \left(\sum_{i=2}^{C} p_{*i} - \sum_{i=2}^{C} p_{*i}^2 - \sum_{i=2}^{C} \sum_{j>i}^{C} p_{*i} p_{*j} \right)$$
$$= \text{HU}_{\text{eq}}(y_*)$$

REFERENCES

[1] P. Bharatkar and R. Patel, "Approach to accuracy assessment tor RS image classification techniques," *Int. J. Scientific Eng. Res.*, vol. 4, no. 12, pp. 79–86, 2013. [Online]. Available: https://www.ijser.org/research-paper/Approach-to-Accuracy-Assessment-tor-RS-Image-Classification-Techniques.pdf

[2] K. M. Ting, *Confusion Matrix*. Boston, MA, USA: Springer, 2010, pp. 209–209, doi: 10.1007/978-0-387-30164-8_157.

[3] C. Ferri, J. Hernández-Orallo, and R. Modroiu, "An experimental comparison of performance measures for classification," *Pattern Recognit. Lett.*, vol. 30, no. 1, pp. 27–38, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865508002687

[4] C. G. Northcutt, A. Athalye, and J. Mueller, "Pervasive label errors in test sets destabilize machine learning benchmarks," in *Proc. 35th Conf. Neural Inf. Process. Syst. Datasets Benchmarks Track (Round 1)*, 2021. [Online]. Available: https://openreview.net/forum?id=XccDXrDNLek

[5] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods," *Mach. Learn.*, vol. 110, no. 3, pp. 457–506, 2021, doi: 10.1007/s10994-021-05946-3.

[6] J. V. Zidek and C. van Eeden, "Uncertainty, entropy, variance and the effect of partial information," *Lecture Notes-Monograph Ser.*, vol. 42, pp. 155–167, 2003. [Online]. Available: http://www.jstor.org/stable/4356236

[7] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 1st ed. Berlin, Germany: Springer-Verlag, 2007.

[8] E. Allaj, "Measuring variability and association for categorical data," *Fuzzy Sets Syst.*, vol. 421, pp. 29–43, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165011420304620

[9] T. Reid, T. Walter, P. Enge, and A. Fowler, "Crowdsourcing arctic navigation using multispectral ice classification & GNSS," in *Proc. 27th Int. Tech. Meeting Satell. Division Inst. Navig. (ION GNSS+)*, 2014, pp. 707–721.

[10] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning.*, (Adaptive Computation and Machine Learning Series). Cambridge, MA, USA: MIT Press, 2006.

[11] M. Abdar et al., "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Inf. Fusion*, vol. 76, pp. 243–297, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253521001081

[12] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. 33rd Int. Conf. Mach. Learn., Proc. Mach. Learn. Res.*, M. F. Balcan and K. Q. Weinberger, Eds., vol. 48. New York, New York, USA: PMLR, Jun. 20–22, 2016, pp. 1050–1059. [Online]. Available: https://proceedings.mlr.press/v48/gal16.html

[13] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural network," in *Proc. 32nd Int. Conf. Mach. Learn.*, Proc. Mach. Learn. Res., F. Bach and D. Blei, Eds., vol. 37. Lille, France: PMLR, Jul. 7–9, 2015, pp. 1613–1622. [Online]. Available: https://proceedings.mlr.press/v37/blundell15.html

[14] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Proc. Adv. Neural Inf. Process. Syst.*, I. Guyon et al., Eds., vol. 30. Red Hook, NY, USA: Curran Associates, 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/9ef2ed4b7fd2c810847ffa5fa85bce38-Paper.pdf

[15] M. Sensoy, L. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," in *Proc. Adv. in Neural Inf. Process.*

Syst., S. Bengio et al., Eds., vol. 31. Red Hook, NY, USA: Curran Associates, 2018. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2018/file/a981f2b708044d6fb4a71a1463242520-Paper.pdf

[16] H. Wu and D. Klabjan, "Logit-based uncertainty measure in classification," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, 2021, pp. 948–956.

[17] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *Proc. 20th Int. Conf. Pattern Recognit.*, 2010, pp. 3121–3124.

[18] O. Caelen, "A Bayesian interpretation of the confusion matrix," *Ann. Math. Artif. Intell.*, vol. 81, nos. 3–4, pp. 429–450, Dec. 2017, doi: 10.1007/s10472-017-9564-8.

[19] N. Tötsch and H. Daniel, "Classifier uncertainty: Evidence, potential impact, and probabilistic treatment," *PeerJ. Comput. Sci.*, vol. 7, 2021, Art. no. e398.

[20] *On Measures of Entropy and Information*, vol. I. Berkeley, CA, USA: Univ. California Press, 1960, pp. 547–561.

[21] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *J. Statist. Phys.*, vol. 52, pp. 479–487, 1988.

[22] S. Chakraborty, D. Paul, and S. Das, "t-Entropy: A new measure of uncertainty with some applications," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2021, pp. 1475–1480.

[23] Y. Zhang, F. Huang, X. Deng, and W. Jiang, "A new total uncertainty measure from a perspective of maximum entropy requirement," *Entropy*, vol. 23, no. 8, 2021, Art. no. 1061. [Online]. Available: https://www.mdpi.com/1099-4300/23/8/1061

[24] E. Hüllermeier, S. Destercke, and M. H. Shaker, "Quantification of credal uncertainty in machine learning: A critical analysis and empirical comparison," in *Proc. 38th Conf. Uncertainty Artif. Intell.*, Proc. Mach. Learn. Res., J. Cussens and K. Zhang, Eds., vol. 180. PMLR, Aug. 1–5, 2022, pp. 548–557. [Online]. Available: https://proceedings.mlr.press/v180/hullermeier22a.html

[25] R. R. Yager and L. Liu, Eds., *Classic Works of the Dempster-Shafer Theory of Belief Functions*, Series Studies in Fuzziness and Soft Computing, vol. 219. Berlin, Germany: Springer, 2008.

[26] F. P. A. Coolen, M. C. M. Troffaes, and T. Augustin, *Imprecise Probability*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 645–648, doi: 10.1007/978-3-642-04898-2_296.

[27] F. Nielsen and K. Sun, *Clustering in Hilbert's Projective Geometry: The Case Studies of the Probability Simplex and the Elliptope of Correlation Matrices*. Cham, Switzerland: Springer International Publishing, 2019, pp. 297–331, doi: 10.1007/978-3-030-02520-5_11.

[28] R. Kass and P. Vos, *Geometrical Foundations of Asymptotic Inference*, Series Wiley Series in Probability and Statistics. New York, NY, USA: Wiley, 2011. [Online]. Available: https://books.google.no/books?id=e43EAIfUPCwC

[29] C. Villani, *Optimal Transport: Old and New*, Series Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

[30] M. L. Rizzo and G. J. Székely, "Energy distance," *WIREs Comput. Statist.*, vol. 8, no. 1, pp. 27–38, 2016. [Online]. Available: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wics.1375

[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, Mar. 2004. [Online]. Available: http://www.amazon.com/exec/obidos/redirect?tag=citeulike-20&path=ASIN/0521833787

[32] P. Virtanen et al., "SciPy 1.0: Fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[33] P. Ghamisi, B. Höfle, and X. X. Zhu, "Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 3011–3024, Jun. 2017.

[34] S. R. Gunn, "Support vector machines for classification and regression," in *Proc. Support Vector Mach.*, 1998. [Online]. Available: https://api.semanticscholar.org/CorpusID:120347962

[35] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002. [Online]. Available: https://CRAN.R-project.org/doc/Rnews/

[36] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[37] "Modulation classification with deep learning," The MathWorks, Inc., Natick, MA, USA, 2022. [Online]. Available: https://www.mathworks.com/help/deeplearning/ug/modulation-classification-with-deep-learning.html

[38] B. Jdid, K. Hassan, I. Dayoub, W. H. Lim, and M. Mokayef, "Machine learning based automatic modulation recognition for wireless communications: A comprehensive survey," *IEEE Access*, vol. 9, pp. 57851–57873, 2021.

[39] R. Zhou, F. Liu, and C. W. Gravelle, "Deep learning for modulation recognition: A survey with a demonstration," *IEEE Access*, vol. 8, pp. 67366–67376, 2020.

[40] O. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: Classical approaches and new trends," *IET Commun.*, vol. 1, no. 2, pp. 137–156, Apr. 2007. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/iet-com_20050176

[41] B. Zadrozny and C. Elkan, "Transforming classifier scores into accurate multiclass probability estimates," in *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, (KDD). New York, NY, USA: ACM, 2002, pp. 694–699, doi: 10.1145/775047.775151.

[42] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008. [Online]. Available: http://www.jmlr.org/papers/v9/vandermaaten08a.html

[43] M. Amgad et al., "Structured crowdsourcing enables convolutional segmentation of histology images," *Bioinformatics*, vol. 35, no. 18, pp. 3461–3467, Feb. 2019, doi: 10.1093/bioinformatics/btz083.

[44] C. Gini, *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche. [Fasc.* I.]. Bologna, Italy: Tipografia di Paolo Cuppin, 1912.

[45] M. Kružík, "Bauer's maximum principle and hulls of sets," *Calculus Variations Partial Differ. Equ.*, vol. 11, pp. 321–332, 2000.

Saloua Chlaily received the double M.Sc. degrees in 2013, from École Nationale Supérieure d'Électrotechnique, d'Électronique, d'Informatique, d'Hydraulique et des Télécommunications (EN-SEEIHT), Toulouse, France, and from Hassania School of Public Works (EHTP), Casablanca, Morocco, in electronics engineering and electrical engineering respectively, and the Ph.D. degree in signal, image, speech, and telecommunication from Grenoble Alpes University, Grenoble, France, in 2018. She is currently a Researcher at UiT The Arctic University of Norway. She is conducting her research at the Center of Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA). Her research interests include multimodal data analysis, signal processing, and image processing. She is a Reviewer of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Debanshu Ratha received the five-year integrated M.Sc. degree in mathematics from the School of Mathematical Sciences, NISER, Bhubaneswar, in 2013, and the Ph.D. degree from CSRE, IIT Bombay, Mumbai, India, in 2020. Currently, he is a Postdoctoral Fellow in earth observation with the Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø. He has been a Visiting Researcher with IETR, Université de Rennes1 and Tèlècom Paris, France, during Ph.D. degree (Raman–Charpak Fellowship 2017) and Postdoctoral (COSMIC project) tenures, respectively. He is interested in Synthetic Aperture Radar (theory and applications), image processing, data science, and climate change and sustainability. He is a Reviewer for most of the reputed geoscience and remote sensing journals, including journals from IEEE. Currently, he serves as the Chief Guest Editor for the Special Issue of MDPI Remote Sensing titled, "SAR Aware Methods towards Sustainable Development Goals."

Pigi Lozou is an undergraduate student working toward the Diploma in electrical engineering and computer science from the Technical University of Athens, specializing in signals, systems, and robotics. She is interested in applying deep learning techniques to the fields of remote sensing and medical diagnostics.

Andrea Marinoni (Senior Member, IEEE) is an Associate Professor in pattern recognition with the Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway. From 2013 to 2018, he has been a Research Fellow with the Telecommunications and Remote Sensing Laboratory, Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy. He has been a Visiting Researcher with the Communications Systems Laboratory, Department of Electrical Engineering, University of California, Los Angeles (UCLA), Los Angeles, CA, USA, in 2009; with the Earth and Planetary Image Facility, Ben-Gurion University of the Negev, Beer Sheva, Israel, in 2015; with the School of Geography and Planning, Sun Yat-sen University, Guangzhou, P.R.C., in 2016–2017; with the School of Computer Science, Fudan University, Shanghai, P.R.C., in 2016; with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, P.R.C., in 2016–2017; and with the Instituto de Telecomunicações, Instituto Superior Tecnico, Universidade de Lisboa, Lisbon, Portugal, in 2017. He has been a Visiting Professor with the Department of Electrical, Computer, and Biomedical Engineering, University of Pavia, Pavia, Italy, in 2020 and 2021; with the Department of Environmental Science, Policy and Management, University of California at Berkeley, Berkeley, CA, USA, in 2022; and with the Earth Observatory of Singapore, Nanyang Technological University, Singapore, in 2023. Since 2020, he is a Visiting Academic Fellow with the Department of Engineering, University of Cambridge, U.K. His main goal is to assess, infer, and predict the interactions between human and environment. To this aim, his research is focused on the development of machine learning and data analysis methods for reliable information extraction on these interactions and their implications for human welfare from multimodal datasets, as well as on the implementation of frameworks for reliable decision-making processes for sustainable development and climate change adaptation and mitigation.