Analysis and Classification of SAR Textures using Information Theory

Eduarda T. C. Chagas, Alejandro C. Frery, *IEEE Senior Member,* Osvaldo A. Rosso, and Heitor S. Ramos, *IEEE Senior Member*

***Abstract*—The use of Bandt-Pompe probability distributions and descriptors of Information Theory has been presenting satisfactory results with low computational cost in the time series analysis literature [1], [2], [3]. However, these tools have limitations when applied to data without time dependency. Given this context, we present a newly proposed technique for texture analysis and classification based on the Bandt-Pompe symbolization for SAR data. It consists of (i) linearize a 2-D patch of the image using the Hilbert-Peano curve, (ii) build an Ordinal Pattern Transition Graph that considers the data amplitude encoded into the weight of the edges; (iii) obtain a probability distribution function derived from this graph; (iv) compute Information Theory descriptors (Permutation Entropy and Statistical Complexity) from this distribution and use them as features to feed a classifier. The ordinal pattern graph we propose considers that the weight of the edges is related to the absolute difference of observations, which encodes the information about the data amplitude. This modification takes into account the scattering properties of the target and leads to the characterization of several types of textures. Experiments with data from Munich urban areas, Guatemala forest regions and Cape Canaveral ocean samples show the effectiveness of our technique in homogeneous areas, which achieves satisfactory levels of separability. The two descriptors chosen in this work are easy and quick to calculate and are used as input for a k-nearest neighbor classifier. Experiments show that this technique presents results similar to state-of-the-art techniques that employ a much larger number of features and, consequently, require a higher computational cost.**

***Index Terms*—Synthetic Aperture Radar (SAR), Texture, Terrain Classification, Permutation Entropy, Ordinal Patterns Transition Graphs.**

I. Introduction

TEXTURE is an elusive trait. When dealing with remotely sensed images, the texture of different patches carries relevant information that, more often than not, is hard to quantify and to transform into useful and parsimonious features. This may be since textures, in this context, is a synesthesia phenomenon that triggers tactile responses from visual inputs. This paper presents a new way of extracting features from textures, both natural and resulting from anthropic processes, in SAR (Synthetic Aperture Radar) imagery.

E. T. C. Chagas and H. S. Ramos ar with Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil (e-mail: eduarda.chagas@dcc.ufmg.br, ramosh@dcc.ufmg.br).

A. C. Frery is with Laboratório de Computação Científica e Análise Numérica – LaCCAN, Universidade Federal de Alagoas, Brasil; (e-mail: acfrery@laccan.ufal.br)

O. A. Rosso is with Instituto de Física, Universidade Federal de Alagoas, Brasil (e-mail: oarosso@if.ufal.br)

Manuscript received XX YY, 20ZZ; revised WW UU, 20VV.

SAR systems are a vital source of data because they provide high-resolution images in almost all weather and day-night conditions. They provide basilar information, complementary to that offered by sensors that operate in other regions of the electromagnetic spectrum, for a variety of Earth Observation applications. Although they present rich information, such data have challenging characteristics. Most notably, they do not follow the usual Gaussian additive model, and the signal-to-noise ratio is usually low.

Yue et al. [4] provide a comprehensive account of how the physical properties of the target are translated into first- and second-order statistical properties of SAR intensity data. There is general agreement that non-deterministic textures are encoded in the second-order features, i.e., in the spatial correlation structure. Therefore is frequent the use of covariance matrix and other measures that assume that a linear dependence, namely the Pearson correlation coefficient, suffices to characterize natural textures. However, texture, in SAR imagery, is often visible only over large areas, and the multiplicative and non-Gaussian nature of speckle antagonizes with the additive assumption that underlies classical approaches, making complex the process of characterizing such data.

Surface classification and land use are among the most important applications of the Synthetic Aperture Radar (SAR) image [5]. In recent years, handcrafted features and representation learning (supervised and unsupervised) algorithms have been proposed [6], [7], [8]. Algorithms of the unsupervised generative adversarial network (GAN) have revolutionized the classification of SAR images, improving performance in small sample problems, and helping the interpretability of such data [9]. Among the supervised algorithms, support vector machine (SVM) [10], random forest (RF) [11], and neural network (NN) [12] have been frequently used in remote sensing. The Principle Component Analysis (PCA) [13], autoencoder [14] and the Boltzmann machine [15] can to extract non-local resources and classify non-labeled PolSAR pixels using an unsupervised approach. However, methods such as graph-based semi-supervised deep learning algorithms [16] can improve classification accuracy in problems with few labeled samples.

Handcrafted features in SAR textures can be studied following two complementary approaches, namely analyzing the marginal properties of the data (first-order statistics), and observing their spatial structure [4], [17]. In this work, we focus on the second approach, which has been showing relevant results using techniques from the image process-

ing literature, such as co-occurrence matrices and Haralick's descriptors [18]. Through the GLCM we can extract features that reflect statistical relationships of the pixel intensity values. On the other hand, Haralick's descriptors can capture information on intensity and amplitude based on global statistics of SAR images. Radford et al. [19] used textural information derived from gray-level co-occurrence matrices (GLCM), along with Random Forests, for geological mapping of remote and inaccessible localities; the authors obtained a classification accuracy of $\approx 90\,\%$, even when using limited training data ($\approx 0.15\,\%$ of the total data). Hagensieker and Waske [20] evaluated the synergistic contribution of multi-temporal L-, C-, and X-band data to tropical land cover mapping, comparing classification outcomes of ALOS-2 [21], RADARSAT-2 [22], and TerraSAR-X [23] datasets for a study site in the Brazilian Amazon using a wrapper approach. The wrapper utilizes the gray-level co-occurrence matrix texture information and a Random Forest classifier to estimate scene importance. Storie [24] proposed an open-source workflow for detecting and delineating the urban-rural boundary using Sentinel-1a SAR data. The author used a combination of GLCM information and a k-means classifier to produce a three-category map that distinguishes urban from rural areas. In higher resolution image classification activities, it is necessary to obtain more granular information from the data by extracting local characteristics of the data, such as scale and orientation. In this scenario, techniques such as Fourier power spectrum [25], random fields [26], Gabor filter [27] and wavelet transform [28] are usually applied.

In our approach, we opt to analyze the 1-D signals resulting from the linearization of the image samples, using non-parametric time series analysis techniques. With this approach, we reduce the dimensionality of the data while preserving the spatial correlation structure. Observations are then transformed into ordinal patterns with the Bandt-Pompe symbolization. We use Information Theory descriptors to analyze the distributions these patterns induce, both directly and by building transition graphs among subsequent patterns.

Those descriptors are the Entropy and the Statistical Complexity, which are easy to obtain and are interpretable. They reveal important features of the underlying process.

We are guided by the following question:

> Which is the best representation of a texture patch that allows extracting expressive Information Theory descriptors for the task of characterizing textures in the presence of speckle?

We verified that the both the histogram of Bandt-Pompe ordinal patterns, and of classical transition graphs do not convey enough information for suitable applications as, for instance, classification.

Henceforth, we propose the Weighted Amplitude Transition Graph (WATG). This graph incorporates the absolute difference among observations as weights of the edges between nodes transitions. Such weights take part in the computation of the probabilities and, thus, influence both Entropy and Statistical Complexity.

The main contribution of this work is the proposal of a new representation of SAR textures which allows a low-dimensional characterization useful for, among other applications, their classification. We compare its performance with the classical histograms of Bandt-Pompe ordinal patterns and the basic transition graph. Since the proposed approach has a low computational cost, the results obtained suggest that this technique has good potential in other applications, such as texture segmentation tools of SAR images.

The paper is structured as follows: Section II describes our proposed methodology. Section II-A presents the patch linearization process of the images. In the Section II-B we report the Bandt-Pompe symbolization process. Section II-C describes the ordinal patterns transition graphs. Section II-E shows our technique of ordinal amplitude transition graph weighting by amplitudes. In Section II-F we report the Information Theory descriptors used throughout this work. SectionřefMethods we summarize the weighted ordinal patterns methods used to compare with our proposal. Section III describe the SAR image datasets, the analysis of ordinal pattern methods, experiments of sliding window selection, and a quantitative assessment. Finally, Section IV concludes the paper.

II. Methodology

As outlined in Section I, we are interested in methods which employ Information Theory descriptors obtained from ordinal patterns. Such techniques are defined for time series, so the first step of our proposal consists of turning 2-D image patches into a 1-D signal; this is discussed in Section II-A. The second step, presented in Section II-B is obtaining the ordinal patterns. Such patterns can be used directly, or be the basis for building transition graphs; these graphs and their variants are described in Sections II-C and II-D. Our proposal is detailed in Section II-E, and is properties are assessed in Section II-G

All these transformations produce different empirical probability distributions. The final features computed on these distributions are the Entropy and Statistical Complexity, which are discussed in Section II-F.

A. Linearization of image patches

We perform a data dimensionality reduction by turning the 2-D patch into a 1-D signal. This could be accomplished by reading the data by lines, columns, or any transformation of 2-D indexes into a sequence of integers. In this work, we chose to use the Hilbert-Peano [29] curve, due to its low computational cost and its ability to preserve relevant properties of pixel spatial correlation.

Nguyen et al. [30] firstly employed Space-filling curves, to map texture into a one-dimensional signal. Carincotte et al. [31] used the Hilbert-Peano curve in the problem of change detection in pairs of SAR images. The authors noted that this transformation exploits the spatial locality and that its pseudo-randomness of direction changes works well for a large family of images, especially natural ones.

Assuming an image patch is supported by an $M \times N$ grid, we have the following definition.

Definition 1: An image scan is a bijective function $f\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ in the ordered pair set $\{(i,j) : 1 \leq i \leq M, 1 \leq$

$j \leq N\}$, which denotes the points in the domain, for the closed range of integers $\{1, \ldots, MN\}$. A scan rule is $\{f^{-1}(1), \ldots, f^{-1}(MN)\}$.
This Definition imposes that each pixel is visited only once, and that all pixels are visited.

Space-filling curves, such as raster-1, raster-2, and Hilbert-Peano scanning techniques, stipulate a proper function f. Hilbert-Peano curves scan an array of pixels of dimension $2^k \times 2^k$, $k \in \mathbb{N}$, never keeping the same direction for more than three consecutive points, as shown in Fig. 1. Using the Hilbert-Peano curve, we reduce the data dimensionality while maintaining the spatial dependence information of the patch. In this work, we use Hilbert-Peano patches of size 128×128.

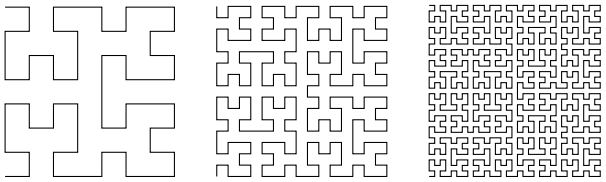


Figure 1: Hilbert-Peano curves in areas of: (a) 8×8, (b) 16×16, and (c) 32×32 pixels.

Figs. 6(a), 6(b), 6(c), 6(d), and 6(e) show five image patches with different textures. Figs. 6(f), 6(g), 6(h), 6(i), and 6(j) present their 1-D representation as signals.

B. Bandt-Pompe Symbolization

The representation of time series by ordinal patterns was introduced by Bandt and Pompe [32] as a transformation resistant to noise, and invariant to nonlinear monotonic transformations. The first step of the `WATG` subroutine is to calculate the ordinal patterns of the 1-D signal by Band-Pompe symbolization.

Consider $\mathcal{X} \equiv \{x_t\}_{t=1}^{T}$ a real valued time series of length T. Let $\mathfrak{A}_D$ (with $D \geq 2$ and $D \in \mathbb{N}$) be the symmetric group of order $D!$ formed by all possible permutation of order D, and the symbol component vector $\boldsymbol{\pi}^{(D)} = (\pi_1, \pi_2, \ldots, \pi_D)$ so every element $\boldsymbol{\pi}^{(D)}$ is unique ($\pi_j \neq \pi_k$ for every $j \neq k$). Consider for the time series $\mathcal{X} \equiv \{x_t\}_{t=1}^{T}$ its time delay embedding representation, with embedding dimension $D \geq 2$ and time delay $\tau \geq 1$ ($\tau \in \mathbb{N}$, also called "embedding time," "time delay", or "delay"):

$$\mathbf{X}_t^{(D,\tau)} = (x_t, x_{t+\tau}, \ldots, x_{t+(D-1)\tau}), \tag{1}$$

for $t = 1, 2, \ldots, N$ with $N = T - (D-1)\tau$. Then the vector $\mathbf{X}_t^{(D,\tau)}$ can be mapped to a symbol vector $\boldsymbol{\pi}_t^D \in \mathfrak{A}_D$. This mapping is such that preserves the desired relation between the elements $x_t \in \mathbf{X}_t^{(D,\tau)}$, and all $t \in \{1, \ldots, T-(D-1)\tau\}$ that share this pattern (also called "motif") are mapped to the same $\boldsymbol{\pi}_t^D$.

We define the mapping $\mathbf{X}_t^{(D,\tau)} \mapsto \pi_t^D$ by ordering the observations $x_t \in \mathbf{X}_t^{(D,\tau)}$ in increasing order. Consider the time series $\mathcal{X} = (1.8, 1.2, 3.2, 4.8, 4.2, 4.5, 2.3, 3.7, 1.2, .5)$ depicted in Fig. 2. Assume we are using patterns of length $D = 5$ with unitary time lag $\tau = 1$. The code associated to $\mathbf{X}_3^{(5,1)} = (x_3, \ldots, x_7) = (3.2, 4.8, 4.2, 4.5, 2.3)$, shown in black, is formed by the indexes in $\boldsymbol{\pi}_3^5 = (1, 2, 3, 4, 5)$ which sort the elements of $\mathbf{X}_3^{(5,1)}$ in increasing order: 51342. With this, $\widetilde{\pi}_3^5 = 51342$, and we increase the counting related to this motif in the histogram of all possible patterns of size $D = 5$.

The dash-dot line in Fig. 2 illustrates $\mathbf{X}_1^{(5,2)}$, i.e. the sequence of length $D = 5$ starting at x_1 with lag $\tau = 2$. In this case, $\mathbf{X}_1^{(5,2)} = (1.8, 3.2, 4.2, 2.3, 1.2)$, and the corresponding motif is $\widetilde{\pi}_1^5 = 51423$.

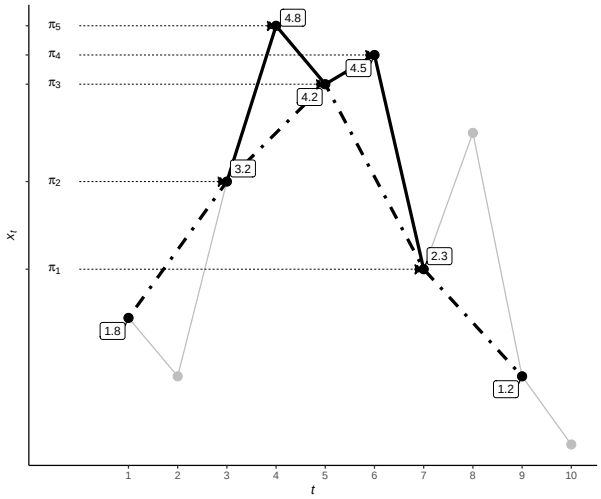


Figure 2: Illustration of the Bandt and Pompe coding

The classic approach to calculating the probability distribution of ordinal patterns is through the frequency histogram. Denote Π the sequence of symbols obtained by a given series $\mathbf{X}_t^{(D,\tau)}$. The Bandt-Pompe probability distribution is the relative frequency of symbols in the series against the $D!$ possible patterns $\{\widetilde{\pi}_t^D\}_{t=1}^{D!}$:

$$p(\widetilde{\pi}_t^D) = \frac{\#\left\{\mathbf{X}_t^{(D,\tau)} \text{ is of type } \widetilde{\pi}_t^D\right\}}{T-(D-1)\tau}, \tag{2}$$

where $t \in \{1, \ldots, T-(D-1)\tau\}$. These probabilities meet the conditions $p(\widetilde{\pi}_t^D) \geq 0$ and $\sum_{i=1}^{D!} p(\widetilde{\pi}_t^D) = 1$, and are invariant before monotonic transformations of the time series values. For example, the presence of α multiplicative noise in $\mathcal{X}$ does not change the results of the patterns produced.

C. Graph of Transitions between Ordinal Patterns

Alternatively, one may form an oriented graph with the transitions from $\widetilde{\pi}_t^D$ to $\widetilde{\pi}_{t+1}^D$. The Ordinal Pattern Transition Graph $G = (V, E)$ represents the transitions between two consecutive ordinal patterns over time t. The vertices are the patterns, and the edges the transitions between them: $V = \{v_{\widetilde{\pi}_t^D}\}$, and $E = \{(v_{\widetilde{\pi}_t^D}, v_{\widetilde{\pi}_{t+1}^D}) : v_{\widetilde{\pi}_t^D}, v_{\widetilde{\pi}_{t+1}^D} \in V\}$ [33].

The literature reports two approaches to compute the weight of edges. Some authors employ unweighted edges [34], [35] which represent only the existence of transitions, while others apply the frequency of transitions [36], [37]. The weights $\mathbb{W} = \{w_{v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D}} : v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D} \in V\}$ assigned to each edge describe the chance of transitions between the patterns $(v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D})$

The weights are calculated as the relative frequency of each transition, i.e.:

$$w_{v_{\widetilde{\pi}_i^D},v_{\widetilde{\pi}_j^D}} = \frac{|\Pi_{\widetilde{\pi}_i^D,\widetilde{\pi}_j^D}|}{T-(D-1)\tau-1}, \tag{3}$$

where $|\Pi_{\widetilde{\pi}_i^D,\widetilde{\pi}_j^D}|$ is the number of transitions from pattern $\widetilde{\pi}_i^D$ to pattern $\widetilde{\pi}_j^D$, $\sum_{v_{\widetilde{\pi}_i^D},v_{\widetilde{\pi}_j^D}} w_{v_{\widetilde{\pi}_i^D},v_{\widetilde{\pi}_j^D}} = 1$, and the denominator is the number of transitions between sequential patterns in the series of motifs of length $T-(D-1)\tau$.

D. Weighted Ordinal Patterns Methods

Recent works proposed using weights in the calculation of relative frequencies for ordinal patterns. They all aim at incorporating the information coded in the amplitude of the observations back into the Permutation Entropy. We summarize in the following those that we used for comparison with our proposal.

1) Weighted Permutation Entropy: The Weighted Permutation Entropy (WPE) was proposed by Fadlallah et al. [38].

Denote $\overline{X}_t^{(D,\tau)}$ the arithmetic mean:

$$\overline{X}_t^{(D,\tau)} = \frac{1}{D}\sum_{k=1}^{D} x_{t+(k-1)}. \tag{4}$$

The weight w_t is the sample variance of each vector $X_t^{(D,\tau)}$:

$$w_t = \frac{1}{D}\sum_{k=1}^{D}\left[x_{t+(k-1)} - \overline{X}_t^{(D,\tau)}\right]^2. \tag{5}$$

Then, the probability distribution is given from the weighted relative frequencies:

$$p(\widetilde{\pi}_t^D) = \frac{\sum_{i:\{\mathbf{X}_i^{(D,\tau)}\mapsto\widetilde{\pi}_t^D\}} w_i}{\sum_{i=1}^{T-(D-1)\tau} w_i}. \tag{6}$$

Fig. 7(b) shows the weighted graph produced by the urban area shown in Figs. 6(d) (as image) and 6(i) (as signal).

2) Fine-Grained Permutation Entropy: The Fine-Grained Permutation Entropy (FGPE) was introduced in Ref. [39].

Let β_t be the difference series:

$$\beta_t = \left\{|x_{t+1}-x_t|,\ldots,|x_{t+(D-1)}-x_{t+(D-2)}|\right\}. \tag{7}$$

The weight w_t quantifies such differences:

$$w_t = \left\lfloor \frac{\max\{\beta_t\}}{\alpha s(\beta_t)} \right\rfloor, \tag{8}$$

where s is the sample standard deviation, α is a user-defined parameter, and $\lfloor\cdot\rfloor$ is the floor function. Then, w_t is added as a symbol at the end of the corresponding pattern, leading to an update of Π:

$$\pi'^D_t = \{\widetilde{\pi}_t^D \cup w_t\}. \tag{9}$$

Finally, the probability distribution is calculated as:

$$p(\pi'^D_t) = \frac{\#\left\{\mathbf{X}_t^{(D,\tau)} \text{ is of type } \pi'^D_t\right\}}{T-(D-1)\tau}. \tag{10}$$

3) Amplitude-Aware Permutation Entropy: The Amplitude-Aware Permutation Entropy (AAPE) was proposed in Ref. [40]. It consists of weighting the amplitude of ordinal patterns by both the mean and the differences of the elements. For this, only an additional parameter $A \in [0,1]$ is required:

$$w_t = \frac{A|x_t|}{D} + \sum_{k=1}^{D-1}\left(\frac{A|x_{t+k}|}{D} + \frac{(1-A)|x_{t+k}-x_{t+k-1}|}{D-1}\right). \tag{11}$$

The probability distribution is given from the weighted relative frequencies:

$$p(\widetilde{\pi}_t^D) = \frac{\sum_{i:\{\mathbf{X}_i^{(D,\tau)}\mapsto\widetilde{\pi}_t^D\}} w_i}{\sum_{i=1}^{T-(D-1)\tau} w_i}. \tag{12}$$

E. Weighted Graph of Transitions between Ordinal Patterns – WATG

Our proposal for computing the probability distribution, henceforth referred to as Weighted Amplitude Transition Graph (WATG), takes into account the dispersion of the return values, leading to a good. We propose a modification of the current ordinal pattern transition graph, incorporating the absolute difference between the observations that produced the patterns.

First, each $\mathcal{X}$ time series is scaled to $[0,1]$, since we are interested in a metric able to compare datasets:

$$\frac{x_i - x_{\min}}{x_{\max}-x_{\min}} \longmapsto x_i, \tag{13}$$

where $x_{\min}$ and $x_{\max}$ are, respectively, the minimum and maximum values of the series. This transformation is relatively stable before contamination, e.g., if instead of $x_{\max}$ we observe $kx_{\max}$ with $k \geq 1$, the relative values are not altered. Nevertheless, other more resistant transformations as, for instance, z scores, might be considered.

Each $\mathbf{X}_t^{(D,\tau)}$ vector is associated with a weight β_t that measures the largest difference between its elements:

$$\beta_t = \max\{|x_i - x_j|\}, \tag{14}$$

where $x_i, x_j \in \mathbf{X}_t^{(D,\tau)}$.

We propose that the weight assigned to each edge is proportional to the amplitude difference observed in the transition:

$$w_{v_{\widetilde{\pi}_i^D},v_{\widetilde{\pi}_j^D}} = \sum_{i:\{\mathbf{X}_t^{(D,\tau)}\mapsto\widetilde{\pi}_i^D\}}\ \sum_{j:\{\mathbf{X}_t^{(D,\tau)}\mapsto\widetilde{\pi}_j^D\}} |\beta_i - \beta_j|. \tag{15}$$

Thus, the probability distribution taken from the weighted amplitude transition graph is given as follows:

$$\begin{cases} \lambda_{v_{\widetilde{\pi}_i^D},v_{\widetilde{\pi}_j^D}} = 1, \text{ if } (v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D}) \in E, \\ \lambda_{v_{\widetilde{\pi}_i^D},v_{\widetilde{\pi}_j^D}} = 0, \text{ otherwise.} \end{cases}, \text{ and} \tag{16}$$

$$p(\widetilde{\pi}_i^D, \widetilde{\pi}_j^D) = \frac{\lambda_{v_{\widetilde{\pi}_i^D},v_{\widetilde{\pi}_j^D}} \cdot w_{v_{\widetilde{\pi}_i^D},v_{\widetilde{\pi}_j^D}}}{\sum_{v_{\widetilde{\pi}_a^D},v_{\widetilde{\pi}_b^D}} w_{v_{\widetilde{\pi}_a^D},v_{\widetilde{\pi}_b^D}}}. \tag{17}$$

Note that the conditions $p(\widetilde{\pi}_i^D, \widetilde{\pi}_j^D) \geq 0$ and $\sum_{\widetilde{\pi}_i^D,\widetilde{\pi}_j^D} p(\widetilde{\pi}_i^D, \widetilde{\pi}_j^D) = 1$ are satisfied.

Thus, series with uniform amplitudes have edges with probability of occurrence well distributed along the graph, while those with large peaks have edges with probability of occurrence much higher than the others.

Fig. 7(b) shows the weighted graph produced by the urban area shown in Figs. 6(d) (as image) and 6(i) (as signal) using this approach. Notice the difference in weights with respect to Fig. 7(b), which was obtained with WPE.

F. Information-Theoretic Descriptors

We chose two Information Theory descriptors: Shannon Entropy and Statistical Complexity. Computing these quantities is the last step of the algorithm, i.e., obtaining the point in the $H \times C$ plane.

Entropy measures the disorder or unpredictability of a system characterized by a probability measure $\mathbb{P}$. Let $\mathbb{P} = \{p_{(\widetilde{\pi}_1^D, \widetilde{\pi}_1^D)}, p_{(\widetilde{\pi}_1^D, \widetilde{\pi}_2^D)}, \ldots, p_{(\widetilde{\pi}_{D!}^D, \widetilde{\pi}_{D!}^D)}\} = \{p_1, \ldots, p_{D!^2}\}$ be the probability function obtained from the 1-D signal weighted amplitude transition graph $\mathbb{X}$. Its normalized Shannon entropy is given by:

$$H(\mathbb{P}) = -\frac{1}{2 \log D!} \sum_{\ell=1}^{D!^2} p_\ell \log p_\ell. \tag{18}$$

The ability of the entropy to capture system properties is limited, so it is necessary to use it in conjunction with other descriptors to obtain a more complete analysis. Other interesting measures are the distances between $\mathbb{P}$ and a probability measure that describes a non-informative process, typically the uniform distribution.

The Jensen-Shannon distance to the uniform distribution $\mathbb{U} = (\frac{1}{D!^2}, \ldots, \frac{1}{D!^2})$ is a measure of how similar the underlying dynamics is to a non-informative process. It is calculated as:

$$Q'(\mathbb{P}, \mathbb{U}) = \sum_{\ell=1}^{D!^2} \Big(p_\ell \log \frac{p_\ell}{u_\ell} + u_\ell \log \frac{u_\ell}{p_\ell} \Big). \tag{19}$$

This quantity is also called "disequilibrium." The normalized disequilibrium is $Q = Q' / \max\{Q'\}$.

Conversely to entropy, the statistical complexity seeks to find interaction and dependence structures among the elements of a given series, being an extremely important factor in the study of dynamic systems. The Statistical Complexity is defined as [41]:

$$C(\mathbb{P}, \mathbb{U}) = H(\mathbb{P}) Q(\mathbb{P}, \mathbb{U}). \tag{20}$$

In our analysis, each time series can then be described by a point $(H(\mathbb{P}), C(\mathbb{P}, \mathbb{U}))$. The set of all pairs $(H(\mathbb{P}), C(\mathbb{P}, \mathbb{U}))$ for any time series described by patterns of length D lies in a compact subset of $\mathbb{R}^2$: the Entropy-Complexity plane. Martín et al [42] obtained explicit expressions for the boundaries of this closed manifold, which depend only on the dimension of the probability space considered, that is, $D!$ for the traditional Bandt-Pompe method, $D! \times D!$ in our case. Through such a tool it is possible to discover the nature of the series, determining if it corresponds to a chaotic (or other deterministic dynamics) or stochastic sequences

Algorithm 1 outlines our methodology. Line 1 transforms the input texture patch P in a 1-D signal with a Hilbert-Peano sequence. With this, the spatial information is encoded into a one-dimensional signal. Line 2 computes the probability distribution of the weighted transition graph induced by the 1-D signal. The WATG function is detailed in Lines 6–9. Lines 3 and 4 compute the two descriptors of the patch.

The `WATG` function consists of three steps: (i) each subsequence of size (dimension) D of observations at delay τ is transformed into an ordinal pattern using the Bandt-Pompe symbolization (function `BPSymbolization`, Line 6); (ii) function `transitions` (Line 7) calculates the sequence of alternations of the ordinal patterns; and (iii) function `weigthGraph` (Line 8) generates the incidence matrix of the graph using the as weights the amplitude differences between the time series elements. Finally, the probability distribution is obtained by turning the transition matrix into a vector (Line 9).

These steps are also depicted in Fig. 3, and are detailed in the following sections.

Algorithm 1 $H \times C$ point from a patch using WATG

Input: Patch of texture P, dimension D and time delay τ
Output: $H \times C$ feature

```
1: signal.1-D ← hilbertCurve(P)
2: Probs ← WATG(signal.1-D, D, τ)
3: H ← ShannonEntropy(Probs)
4: C ← StatisticalComplexity(Probs)
5: function WATG(signal.1-D, D, τ)
6:     patterns ← BPSymbolization(signal.1-D, D, τ)
7:     transitions ← transitions(patterns)
8:     graph ← weigthGraph(signal.1-D, transitions)
9:     Probs ← as.vector(graph)
10:    return Probs
11: end function
```

G. Properties

We conducted two experiments to analyze the response of WATG to different noise levels and image rotations. Our truth is the deterministic image generated by the function:

$$z(x, y) = \sin(4x + 0.5y),$$

where $x, y \in [-2\pi, 2\pi]$. Fig. 4(a) shows this function as a 128×128-pixel patch.

The speckle noise was modeled as outcomes of independent, identically distributed unitary-mean Gamma random variables with shape parameter L (the number of looks, which controls the signal-to-noise ratio) $W(L) \sim \Gamma(L, L)$, with $L \in \{1, 100, 150, 200, 250, 300, 350, 400, 450, 500\}$. The observed images $I(L)$ are the pixelwise product of z and $w(L)$. Fig. 4(b) shows the product $I(100)$.

Fig. 5 shows how the point in the $H \times C$ plane varies according to the level of noise introduced. The ground truth (identified as "0") has relatively low entropy and is close to the maximum complexity (the continuous line is the upper bound). This is typical of deterministic sequences. When we

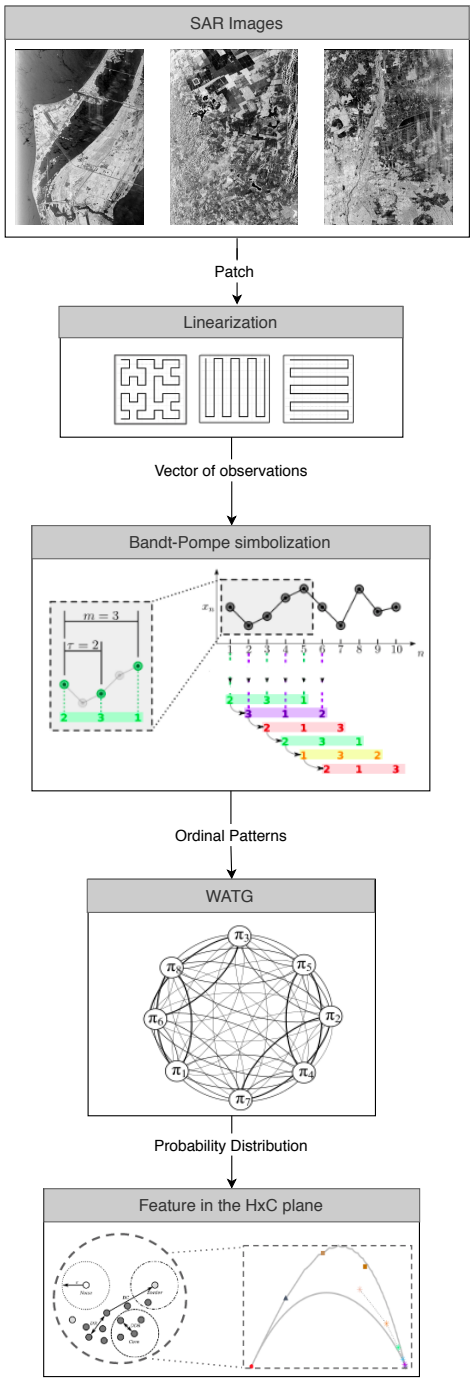


Figure 3: Outline of the methodology used for the classification of textures.

inject single-look speckle ($L = 1$), the entropy increases along with the complexity. The technique is, thus, able to identify the deterministic component even when it is embedded in the strongest possible speckle noise. The point "1" shifts towards "0" when the signal-to-noise progressively increases.

We also verified that the points in the $H \times C$ plane are almost insensitive to rotations. Fig. 4(c) shows the transpose of $I(500)$. In all cases, the coordinates (h, c) of the transposed noisy images were equal, up to the fourth decimal place, to those of the original version.

(a) z (b) $I(100)$ (c) $I^{\mathrm{T}}(500)$

Figure 4: Ground truth, speckled, and speckled transposed versions.

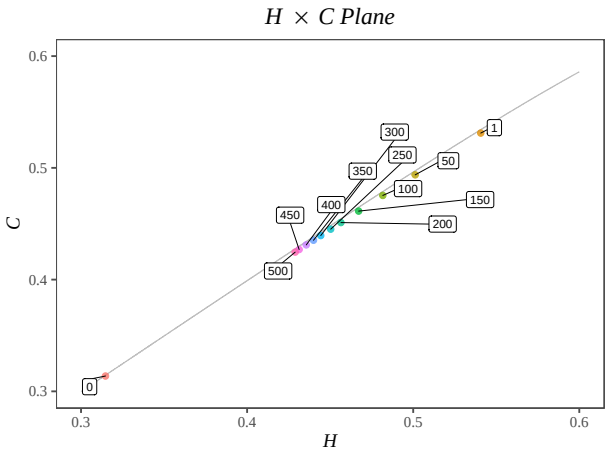


Figure 5: Modifications to the $H \times C$ Plane features by adding different multiplicative noises.

These experiments provide evidence that the WATG mapping is little sensitive to rotations (thanks to the use of Peano-Hillbert curves), and that it is able to identify the presence of underlying structural information in the presence of varying levels of speckle.

III. Results

In this section, we describe the datasets, the classification process, and the results of applying WATG.

To assess the performance of the technique here proposed, we analyze the impact of its parameters and compare its results in the classification with other methods.

A. Image Datasets

We used the HH backscatter magnitudes of three quad-polarimetric L-band SAR images from the NASA Jet Propulsion Laboratory's (JPL's) uninhabited aerial vehicle synthetic aperture radar (UAVSAR) sensor with $L = 36$ nominal looks:

- a forest region of Sierra del Lacandón National Park, Guatemala, (acquired on April 10, 2015)[1]. The image has 8917×3300 pixels with $10\,\mathrm{m} \times 2\,\mathrm{m}$ resolution.
- ocean regions from Cape Canaveral Ocean (acquired on September 22, 2016). The image has 7038×3300 pixels with $10\,\mathrm{m} \times 2\,\mathrm{m}$ resolution;

[1]https://uavsar.jpl.nasa.gov/cgi-bin/product.pl?jobName=Lacand_30202_15043_006_150410_L090_CX_01#dados

- an urban area of the city of Munich, Germany (acquired on June 5, 2015)[2]. The image has 5773×3300 pixels with $10\,\mathrm{m} \times 3\,\mathrm{m}$ resolution.

We manually selected 200 samples of size 128×128 to compose the dataset used in the experiments. It is organized as follows: 40 samples from Guatemalan forest regions; 40 samples from Guatemalan pasture regions; 80 samples from the oceanic regions of Cape Canaveral, divided into two types with different contrast; and 40 samples of urban regions of the city of Munich. Fig. 6 shows examples of each.

We randomly split the samples in training ($85\,\%$) and test ($15\,\%$) sets. We used the first set to train a k-nearest neighbor classifier algorithm with tenfold cross-validation.

B. Analysis of ordinal patterns methods

Fig. 6 shows examples of the ocean, forest, urban, and pasture samples as sequences values, after the linearization process.

The variation in the magnitude of the targets' backscatter and, consequently, in the intensity of the image pixels, depends on the intrinsic properties of the regions under analysis. Urban targets usually exhibit the strongest variation, followed by forest, pasture, forests, and finally, water bodies. By adding information related to amplitude, the proposed method is able to increase, compared to traditional methods, the granularity of information captured by ordinal patterns.

Fig. 8 shows the evolution of the descriptive power of these techniques.

The Bandt-Pompe symbolization was the first method based on ordinal patterns proposed in the literature. As shown in Fig. 8 left, it provides a limited separation of the textures. Transition graphs (Fig. 8 center) improve the spread of the features, but with some amount of confusion. Our proposal, shown in Fig. 8 right, produces well-separated features. In this way, we were able to obtain, for this experiment, a perfect characterization and, consequently, the high descriptive power of the regions.

As already described in Section II-E, our proposal weights the edges in terms of the difference of amplitudes. As expected, the greatest impact is observed on the transition graphs obtained from urban areas. The urban area 1-D signal shown in Fig. 6 has the largest dynamic range. Fig. 7 shows how this information alters the weights of the transition graph. Notice, in particular, that $(v_{\widetilde{\pi}^3_{123}}, v_{\widetilde{\pi}^3_{123}})$ almost doubled, while $(v_{\widetilde{\pi}^3_{312}}, v_{\widetilde{\pi}^3_{231}})$ and $(v_{\widetilde{\pi}^3_{213}}, v_{\widetilde{\pi}^3_{132}})$ became negligible. We highlight the impact of the weighting on the probability distribution in the two extreme cases observed:

- If the 1-D signal presents a low amplitude variation and intensity peaks between, then the transitions of ordinal patterns that represent the latter have larger weights. This contributes so that the probability distribution becomes less uniform among the symbols, since it will be more concentrated in these edges. This will also cause a drop in the Entropy, when compared to the traditional method.
- In 1-D signal that show a uniform amplitude variation, the weights are well distributed between their edges, giving rise to a more random probability distribution, thus obtaining larger entropy.

[2]https://uavsar.jpl.nasa.gov/cgi-bin/product.pl?jobName=munich_19417_15088_002_150605_L090_CX_01#data

C. Experiments on sliding window selection

In this section, we analyze the parameters of the proposed method and its impact on the classification of textures. McCullough et al. [34] report that inadequate values may hinder important characteristics of the phenomenon under analysis. The two parameters of the transition graph are the dimension D, and the delay τ. In the experiments below, we present the results in the classification using different values of these parameters.

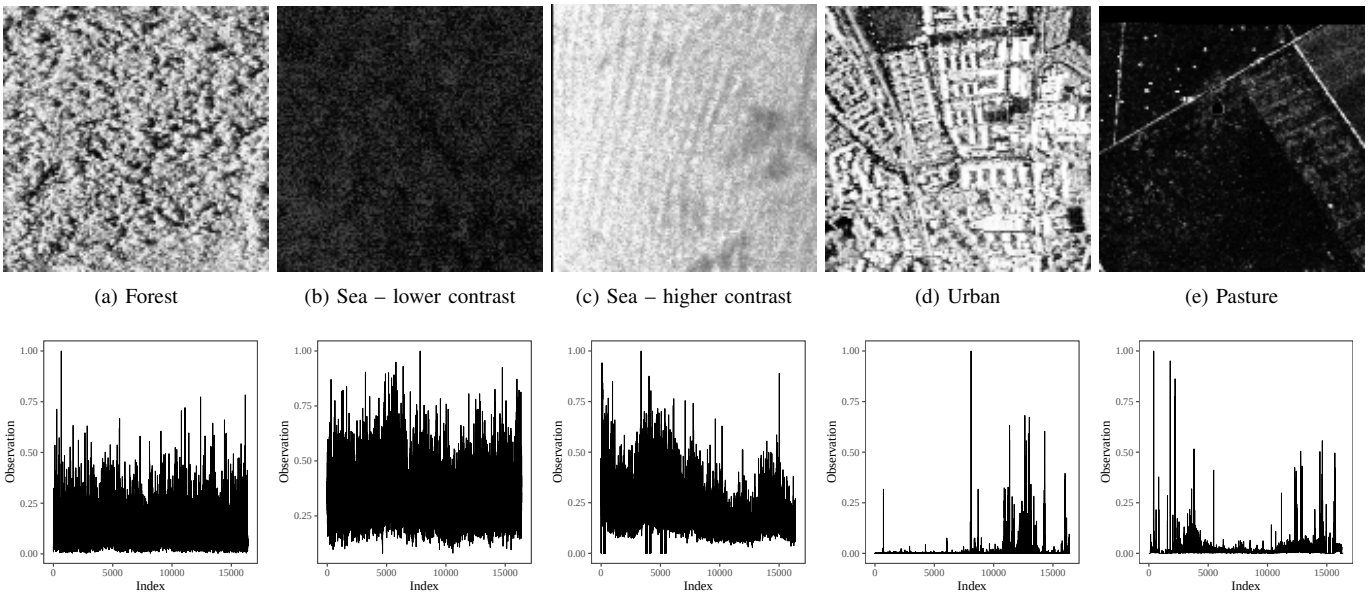
The performance of the classification method based on ordinal patterns is sensitive to window size, the embedding dimension, and the delay. In techniques based in Bandt-Pompe symbolization, for a fixed time series, as the size of embedding dimension decreases, more ordinal patterns are produced. Therefore, we acquire a greater granularity of information about the dynamics of the system and, consequently, we capture more spatial dependencies between the elements.

We used the ROC curve shown in Fig. 9 for different values of $D \in \{3, 4, 5, 6\}$ and $\tau \in \{1, 2, 3, 4, 5\}$ to select the best configuration. As we can see in Fig. 9, the configurations that extracted most information from the 1-D signal and, thus, that presented the best results in the experiments, are $(D = 3, \tau = 1)$ and $(D = 4, \tau = 1)$. The technique, thus, shows its best performance choosing the parameters with the lowest computational cost.

Figure 10 shows the points in the $H \times C$ produced by the same samples with all the parameters mentioned above. The spatial distribution of the points changes with the parameters, and certain configurations promote better separation. This figure shows that discrimination ability decreases with increasing τ. Larger values of delay dilute the spatial dependence, as neighboring points in the sample tend to be more distant in the image. For this reason, we will use $\tau = 1$. Considering $\tau = 1$ (first column of Fig. 10), we also notice that $D = 3$ produces the best separation among classes. Increasing D also increases the Statistical Complexity; this is noticeable for the Forest class. The other effect of considering larger values of D is an increased Entropy of Ocean and, with it, an undesirable overlap with Urban samples.

D. Quantitative Evaluation

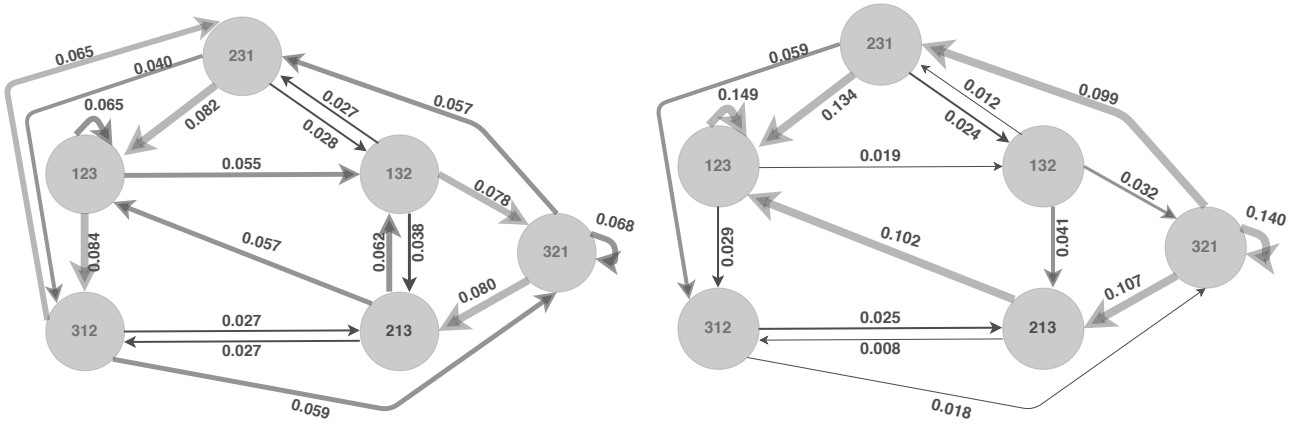
We present a comparison between our proposal and other methods for characterization and texture classification. We use the following ten methods: Gabor filters [43], Histogram of oriented gradients (HOG) [44], Gray-level co-occurrence matrices (GLCM) [45], Speeded-Up Robust Features (SURF) [46], Short Time Fourier Transform (STFT) [47] with SURF, Bandt-Pompe probability distribution [32], Ordinal patterns transition graphs [33], Weighted Permutation Entropy (WPE) [38], Fine-Grained Permutation Entropy (FGPE) [39] with $\alpha = 0.5$, and Amplitude-Aware Permutation Entropy (AAPE) [40] with $A = 0.5$. As in [48], we computed



(a) Forest (b) Sea – lower contrast (c) Sea – higher contrast (d) Urban (e) Pasture

(f) Signal – forest (g) Signal – sea (h) Signal – sea (i) Signal – sea (j) Signal – pasture

Figure 6: Types of regions (Guatemala forest, Canaveral ocean types 1 and 2, Munich urban area, and pasture area) and their signal representation.



(a) Transition Graph (b) WATG

Figure 7: Example of the difference in weights at the edges in a sample of Munich urban regions in the transition graph and in the weighted graph of ordinal patterns transition with dimension 3 and delay 1

four statistics from co-occurrence matrices: contrast, correlation, energy, and homogeneity. Likewise, we implemented the Gabor filters in five scales and eight orientations; using the energy, we obtained an 80-dimensional feature vector for each patch. For the HOG technique, we used image pixels divided into cells equal to 3×3 pixels, where for each cell 6-bin histograms ranging from 0 to 180 degrees or 0 to 360 degrees.

In the classification, we used the k-nearest neighbor algorithm with Euclidean distance, selecting the value of k with the automatic grid search method of the Caret R package [49]. For validation, we used 10-fold cross-validation. More details about the classifier and the sampling can be seen in [50]. Table I summarizes the number of features each method produces, as well as its performance of classifying the 200 samples. We assess the effectiveness of each approach using the following metrics:

- Recall or True Positive Rate (TPR): Total of the correctly detected class of regions over the total of regions for given ground truth.

$$\text{TPR} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}.$$

- Precision or Positive Predictive Value (PPV): Total of regions correctly predicted over the total of regions classified by the algorithm.

$$\text{PPV} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}.$$

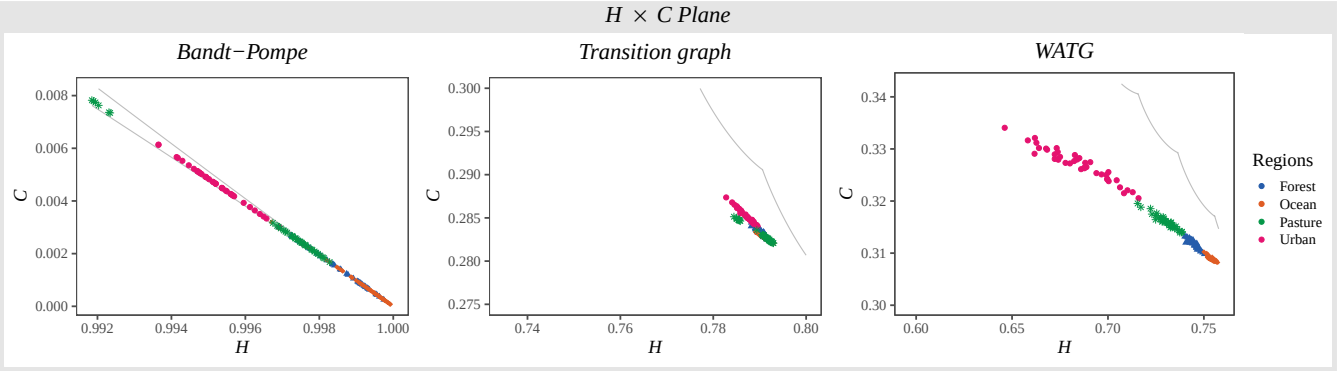


Figure 8: Location of Guatemala (forest), Cape Canaveral (ocean), and Munich (urban) in the $H \times C$ plane for dimension 3 and delay 1. The continuous curves correspond to the maximum and minimum values of C as a function of H.

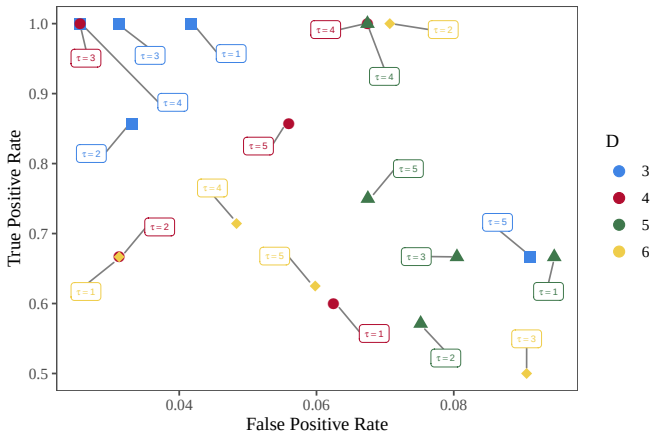


Figure 9: Evaluation of the sliding window parameters using ROC curve

- Overall Accuracy (OA): Total of regions correctly classified. Knowing the number of true positive (TP) for each region, we have to:

$$\text{OA} = \frac{\text{TP}_{\text{ocean}} + \text{TP}_{\text{forest}} + \text{TP}_{\text{urban}}}{\text{Total Regions}}.$$

- F1-score: The harmonic mean of precision and recall.

$$\text{F1-score} = 2 \times \frac{\text{PPV} \times \text{TPR}}{\text{PPV} + \text{TPR}}.$$

As we can observed in Table I, among the methods of weighting ordinal patterns, FGPE produced the worst results: OA $= 86.6\,\%$ and F1-score $= 72.7\,\%$. AAPE also produced a low F1-score result, but it produced more consistent results in the other metrics, presenting OA $= 93.3\,\%$. On the other hand, WPE achieved one of the best results of the developed analysis: OA $= 93.3\,\%$ and F1-score $= 92.3\,\%$. Therefore, we found that among such methods, WATG is the one that best describes the textures presented, obtaining the best performance.

The STFT + SURF produced the worst results: OA $= 40.0\,\%$ and F1-score $= 20.0\,\%$. On the other hand, the application of the SURF descriptor alone provided a better performance: OA $= 50.0\,\%$ and F1-score $= 66.6\,\%$. Since the problem analyzed consists of classification with few samples, although the classification method GLCM-based obtains good TPR and PPV, it has a low OA provided by the impact of the presence of false-positive and false-negative results. Among all handcraft methods analyzed, HOG and Gabor filters achievement the highest success rate in the classification tasks with OA $= 100\,\%$ and F1-score $= 100\,\%$. However, WATG achieves that same performance using only 2 features. This reduction implies less computational power requirement and avoids the curse of dimensionality [51].

IV. Conclusions

We presented and assessed a new method of analysis and classification of SAR image textures. This method consists of three steps: (1) linearization, (2) computing the Weighted Ordinal Pattern Transition Graph, and (3) obtaining Information Theory descriptors. With the use of the WATG technique, we can find a probability distribution function of ordinal patterns that is sensitive to amplitude information and then apply the k-NN algorithm to classify the descriptors. As a result, in addition to perfectly separating urban areas and pasture from the others analyzed by entropy values, we are still able to differentiate oceanic and forest areas through their different values of statistical complexity, which informs us of the degree of spatial dependence between their elements.

Experiments using patches from UAVSAR images showed that the proposal performs better than GLCM, Bandt-Pompe, Transition Graphs, SURF, STFT + SURF, and other techniques of amplitude information analysis in ordinal patterns and that it provides the same quality of results obtained with Gabor filters and HOG. However, while Gabor filters employ 80 features and HOG employ 54 features, our proposal requires only two. Since the proposed method has a lower number of descriptive features than its direct competitors, we were able to obtain some additional advantages under this. First, by reducing the dimension of the features to 2-D, we can accurately visualize the differences between the classes of regions analyzed. In addition, for machine learning algorithms, the smaller the number of dimensions, the faster the process, and less storage space are required. So, we managed to avoid

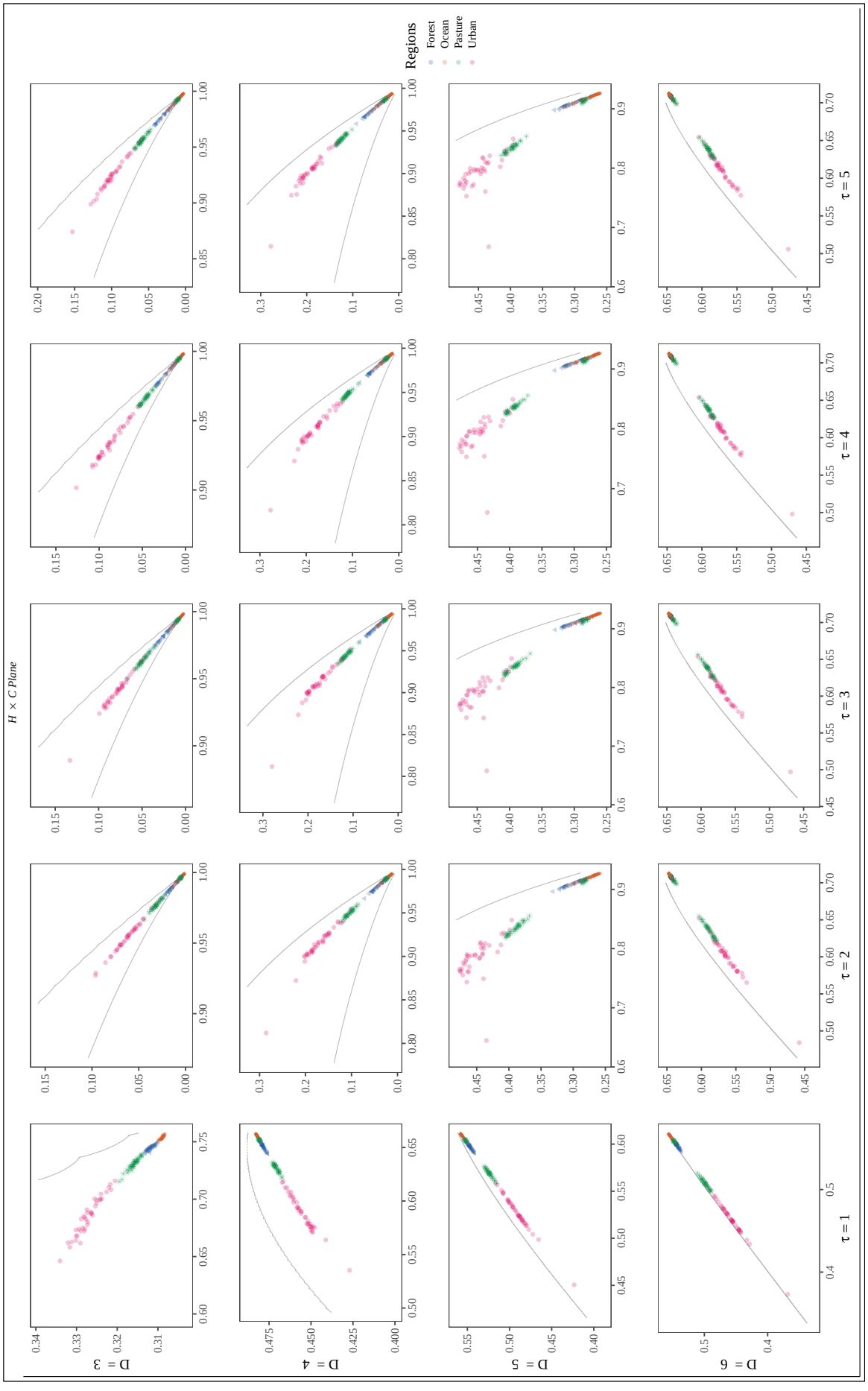


Figure 10: Characterization resulting in $H \times C$ Plane from the application of the Hilbert-Peano curve in WATG on textures of different regions: Guatemala (forest), Cape Canaveral (ocean) and Munich (urban). The continuous curves correspond to the maximum and minimum values of C as a function of H.

Table I: Experimental results using k-NN

Method	# features	TPR				PPV				OA	F1-Score
		Forest	Pasture	Ocean	Urban	Forest	Pasture	Ocean	Urban		
Gabor	80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
HOG	54	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GLCM	32	0.833	1.000	1.000	0.833	1.000	0.857	0.923	1.000	0.933	0.909
SURF	1856	0.500	0.000	1.000	0.000	1.000	0.000	0.444	0.000	0.500	0.666
STFT + SURF	1856	0.166	0.000	0.833	0.166	0.250	0.000	0.416	0.500	0.400	0.200
Bandt-Pompe	2	0.333	1.000	0.750	1.000	0.500	0.857	0.750	0.857	0.766	0.400
Transition Graph	2	0.833	0.666	0.833	1.000	0.833	0.800	0.769	1.000	0.833	0.833
WPE	2	1.000	0.833	1.000	0.833	0.857	0.833	1.000	1.000	0.933	0.923
AAPE	2	0.666	1.000	1.000	1.000	1.000	0.857	0.923	1.000	0.933	0.800
FGPE	2	0.666	0.666	1.000	1.000	0.800	0.666	0.923	1.000	0.866	0.727
WATG	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

overfitting classification models, a recurring problem in data with high dimensionality.

Since the application of this work is limited to texture patches from homogeneous regions, we aim in the future study the possible impacts of analysis on heterogeneous areas, such as cultivation zones and mixed urban regions. The use of transition detection techniques and the segmentation of these signals in a stage before the use of WATG may increase the power of its application and will be fruitful avenues for future work.

V. Reproducibility and Replicability

Following the guidelines presented in Ref. [**?**], the text, source code, and data used in this study are available at the *SAR-WATG* repository https://github.com/EduardaChagas/SAR-WATG.

References

[1] A. L. L. Aquino, H. S. Ramos, A. C. Frery, L. P. Viana, T. S. G. Cavalcante, and O. A. Rosso, "Characterization of electric load with information theory quantifiers," *Physica A*, vol. 465, pp. 277–284, 2017.

[2] O. A. Rosso, R. Ospina, and A. C. Frery, "Classification and verification of handwritten signatures with time causal information theory quantifiers," *PLoS ONE*, vol. 11, no. 12, p. e0166868, 2016.

[3] T. A. Schieber, L. Carpi, A. C. Frery, O. A. Rosso, P. M. Pardalos, and M. G. Ravetti, "Information theory perspective on network robustness," *Physics Letters A*, vol. 380, pp. 359–364, 2016.

[4] D.-X. Yue, F. Xu, A. C. Frery, and Y.-Q. Jin, "A generalized Gaussian coherent scatterer model for correlated SAR texture," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2947–2964, Apr. 2020.

[5] J.-S. Lee, M. R. Grunes, E. Pottier, and L. Ferro-Famil, "Unsupervised terrain classification preserving polarimetric scattering characteristics," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 4, pp. 722–731, 2004.

[6] P. Han, B. Han, X. Lu, R. Cong, and D. Sun, "Unsupervised classification for polsar images based on multi-level feature extraction," *International Journal of Remote Sensing*, vol. 41, no. 2, pp. 534–548, 2020.

[7] Z. Huang, C. O. Dumitru, Z. Pan, B. Lei, and M. Datcu, "Classification of large-scale high-resolution sar images with deep transfer learning," *arXiv preprint arXiv:2001.01425*, 2020.

[8] W. Xie, G. Ma, F. Zhao, H. Liu, and L. Zhang, "Polsar image classification via a novel semi-supervised recurrent complex-valued convolution neural network," *Neurocomputing*, 2020.

[9] F. Liu, L. Jiao, and X. Tang, "Task-oriented gan for polsar image classification and clustering," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 9, pp. 2707–2719, 2019.

[10] C. Sukawattanavijit, J. Chen, and H. Zhang, "Ga-svm algorithm for improving land-cover classification using sar and optical remote sensing data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 3, pp. 284–288, 2017.

[11] H. McNairn, A. Kross, D. Lapen, R. Caves, and J. Shang, "Early season monitoring of corn and soybeans with terrasar-x and radarsat-2," *International Journal of Applied Earth Observation and Geoinformation*, vol. 28, pp. 252–259, 2014.

[12] Z. Lin, K. Ji, M. Kang, X. Leng, and H. Zou, "Deep convolutional highway unit network for sar target classification with limited labeled training data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 1091–1095, 2017.

[13] R. Ressel, A. Frost, and S. Lehner, "A neural network-based classification for sea ice types on x-band sar images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 7, pp. 3672–3680, 2015.

[14] R. Wang and Y. Wang, "Classification of polsar image using neural nonlocal stacked sparse autoencoders with virtual adversarial regularization," *Remote Sensing*, vol. 11, no. 9, p. 1038, 2019.

[15] F. Qin, J. Guo, and W. Sun, "Object-oriented ensemble classification for polarimetric sar imagery using restricted boltzmann machines," *Remote Sensing Letters*, vol. 8, no. 3, pp. 204–213, 2017.

[16] H. Bi, J. Sun, and Z. Xu, "A graph-based semisupervised deep learning model for polsar image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2116–2132, 2018.

[17] F. N. Numbisi, F. Van Coillie, and R. De Wulf, "Multi-date Sentinel SAR image textures discriminate perennial agroforests in a tropical forest-savannah transition landscape," *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 42, no. 1, 2018.

[18] Q. Yu, M. Xing, X. Liu, L. Wang, K. Luo, and X. Quan, "Detection of land use type using multitemporal SAR images," in *IGARSS 2019 – IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 1534–1537.

[19] D. D. Radford, M. J. Cracknell, M. J. Roach, and G. V. Cumming, "Geological mapping in Western Tasmania using radar and random forests," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3075–3087, 2018.

[20] R. Hagensieker and B. Waske, "Evaluation of multi-frequency SAR images for tropical land cover mapping," *Remote Sensing*, vol. 10, no. 2, p. 257, 2018.

[21] Y. Kankaku, S. Suzuki, and Y. Osawa, "Alos-2 mission and development status," in *IGARSS 2013 – IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2013, pp. 2396–2399.

[22] L. Morena, K. James, and J. Beck, "An introduction to the RADARSAT-2 mission," *Canadian Journal of Remote Sensing*, vol. 30, no. 3, pp. 221–234, 2004.

[23] H. Breit, T. Fritz, U. Balss, M. Lachaise, A. Niedermeier, and M. Vonavka, "TerraSAR-X SAR processing and products," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 2, pp. 727–740, 2009.

[24] C. D. Storie, "Urban boundary mapping using Sentinel-1A SAR data," in *IGARSS 2018 – IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 2960–2963.

[25] J. B. Florindo and O. M. Bruno, "Fractal descriptors based on Fourier spectrum applied to texture analysis," *Physica A*, vol. 391, no. 20, pp. 4909–4922, 2012.

[26] T. Zhu, F. Li, G. Heygster, and S. Zhang, "Antarctic sea-ice classification based on conditional random fields from RADARSAT-2 dual-polarization satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2451–2467, 2016.

[27] C. O. Dumitru, S. Cui, G. Schwarz, and M. Datcu, "Information content of very-high-resolution sar images: Semantics, geospatial context, and ontologies," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 4, pp. 1635–1650, 2014.
[28] G. Akbarizadeh, "A new statistical-based kurtosis wavelet energy feature for texture recognition of sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4358–4368, 2012.
[29] J.-H. Lee and Y.-C. Hsueh, "Texture classification method using multiple space filling curves," *Pattern Recognit. Lett.*, vol. 15, no. 12, pp. 1241–1244, 1994.
[30] P. Nguyen and J. Quinqueton, "Space filling curves and texture analysis," in *IEEE Intl. Conf. Pattern Recognition*, 1982, pp. 282–285.
[31] C. Carincotte, S. Derrode, and S. Bourennane, "Unsupervised change detection on SAR images using fuzzy hidden Markov chains," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 432–441, Feb 2006.
[32] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, p. 174102, 2002.
[33] J. Borges, H. Ramos, R. Mini, O. A. Rosso, A. C. Frery, and A. A. F. Loureiro, "Learning and distinguishing time series dynamics via ordinal patterns transition graphs," *Appl. Math. Comput.*, vol. 362, p. UNSP 124554, 2019.
[34] M. McCullough, M. Small, T. Stemler, and H. H.-C. Iu, "Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 5, p. 053101, 2015.
[35] C. W. Kulp, J. M. Chobot, H. R. Freitas, and G. D. Sprechini, "Using ordinal partition transition networks to analyze ECG data," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 26, no. 7, p. 073114, 2016.
[36] T. Sorrentino, C. Quintero-Quiroz, A. Aragoneses, M. Torrent, and C. Masoller, "Effects of periodic forcing on the temporally correlated spikes of a semiconductor laser with feedback," *Opt. Express*, vol. 23, 2015.
[37] J. Zhang, J. Zhou, M. Tang, H. Guo, M. Small, and Y. Zou, "Constructing ordinal partition transition networks from multivariate time series," in *Scientific Reports*, 2017.
[38] B. H. Fadlallah, B. Chen, A. Keil, and J. C. Príncipe, "Weighted-permutation entropy: a complexity measure for time series incorporating amplitude information." *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, vol. 87 2, p. 022911, 2013.
[39] L. Xiao-Feng and W. Yue, "Fine-grained permutation entropy as a measure of natural complexity for time series," *Chin. Phys. B*, vol. 18, no. 7, p. 2690, 2009.
[40] H. Azami and J. Escudero, "Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation," *Comput. Methods Programs Biomed.*, vol. 128, pp. 40–51, 2016.
[41] P. W. Lamberti, M. T. Martín, A. Plastino, and O. A. Rosso, "Intensive entropic non-triviality measure," *Physica A*, vol. 334, no. 1–2, pp. 119–131, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378437103010963
[42] M. T. Martín, A. Plastino, and O. A. Rosso, "Generalized statistical complexity measures: Geometrical and analytical properties," *Physica A*, vol. 369, no. 2, pp. 439–462, 2006.
[43] T. P. Weldon, W. E. Higgins, and D. F. Dunn, "Efficient Gabor filter design for texture segmentation," *Pattern Recognit.*, vol. 29, no. 12, pp. 2005–2015, 1996.
[44] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 886–893.
[45] A. Kourgli, M. Ouarzeddine, Y. Oukil, and A. Belhadj-Aissa, "Texture modelling for land cover classification of fully polarimetric SAR images," *International Journal of Image and Data Fusion*, vol. 3, no. 2, pp. 129–148, 2012.
[46] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*. Springer, 2006, pp. 404–417.
[47] M. Portnoff, "Time-frequency representation of digital signals and systems based on short-time fourier analysis," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 1, pp. 55–69, 1980.
[48] D. Guan, D. Xiang, X. Tang, L. Wang, and G. Kuang, "Covariance of textural features: A new feature descriptor for SAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 10, pp. 3932–3942, 2019.
[49] M. Kuhn, "Building predictive models in R using the caret package," *Journal of Statistical Software*, vol. 28, no. 5, pp. 1–26, 2008.
[50] T. M. Mitchell, *Machine Learning*. McGraw-hill New York, 1997.
[51] N. Altman and M. Krzywinski, "The curse(s) of dimensionality," *Nat. Methods*, vol. 15, no. 6, pp. 399–400, May 2018.

ACKNOWLEDGEMENTS

This work was partially funded by the Coordination for the Improvement of Higher Education Personnel (CAPES) and National Council for Scientific and Technological Development (CNPq).