RENYI DIVERGENCES LEARNING FOR EXPLAINABLE CLASSIFICATION OF SAR IMAGE PAIRS

Matthieu Gallet, Ammar Mian, Abdourrahmane Atto

RENYI DIVERGENCES LEARNING FOR EXPLAINABLE CLASSIFICATION OF SAR IMAGE PAIRS

Matthieu Gallet[†]*, Ammar Mian*[†]*, Abdourrahmane Atto*[†]

[†]: Université Savoie Mont Blanc, LISTIC, Annecy France
Contact: {firstname.lastname} at univ-smb.fr

ABSTRACT

We consider the problem of classifying a pair of Synthetic Aperture Radar (SAR) images by proposing an explainable and frugal algorithm that integrates a set of divergences. The approach relies on a statistical framework that takes standard probability distributions into account for modelling SAR data. Then, by learning a combination of parameterized Renyi divergences and their parameters from the data, we are able to classify the pair of images with fewer parameters than regular machine learning approaches while also allowing an interpretation of the results related to the priors used. Experiments on real multi-class data demonstrate the virtues of the suggested method when compared to both Random Forest and Convolutional Neural Networks (CNN) classifiers, showing its resilience to disturbances such as polluted labels and variations in the percentage of training data.

***Index Terms*—** SAR, Rényi divergence, explainable machine learning, classification

1. INTRODUCTION

In the recent decades, Synthethic Aperture Radar (SAR) images have shown a dramatic increase in temporal and spatial availability thanks to the numerous satellite acquisition missions. Measuring the distance between any pair of SAR observations is one among the challenges related to the exploitation of these images. Such a measure becomes intricate when studying images from different acquisition modalities. This scenario arises when using sensors with the same type but operating on different frequency bands [1] or when operating with multiple/various sensor categories and configurations [2] (optical, SAR, near infrared) while trying to merge information for classification. This scenario also occurs when considering physical interferometry matching (to assess coherence) in supervised [3] but also non-supervised [4] and semi-supervised [5] classification issues, in addition with the analysis of time-shifted observations (same sensor but different acquisition dates) for change detection using wavelet decompositions [6].

Classification of SAR images has been extensively developed with approaches ranging from texture-based feature extraction [7] to extravagant, over-parameterized and onerous deep learning approaches that are furthermore difficult to interpret [8]. Moreover the nature of SAR images (speckle multiplicative noise and very large spatial spanning) makes a ground truth difficult to obtain at large scales. This limits the performance and generalisation of learned models compared to computer-vision applications. One approach to circumvent this limitation is to integrate physics-based priors allowing to obtain representations that are robust to the speckle noise. Additionally, by keeping those priors at each step of the learning algorithm, it is possible to interpret the results by backtracking each step to understand the impacts of the models involved, yielding an explainable framework. Those approaches can rely on low-level descriptors linked to the physics of the problem [9] or by trying to inject a more complex knowledge into neural networks architectures and by interpreting the decisions [10].

One approach that has been successful in SAR applications is considering the probability distributions of the observations, either in a non-parametric approach (histograms for instance) or by using parametric probability models, and then comparing these distributions for the sake of clustering/classification. When dealing with classification over stastical models, statistical divergence based methods have been popular [11]. However the limitations of those approaches is that they either rely on prior and compute distances between empirical divergences [12] or rely on a single probability prior model [13]; while it is often difficult to model the variability of the data from a single probability distribution family. In the second case, neural networks are able to learn the characteristics of imaged fields thanks to providing a huge quantity of data. However, in SAR context, these networks often lack generalisation [14] and decision explainability given the important number of parameters involved. We propose to combine hereafter the strengths of the 2 approaches highlighted above in terms of deriving a shallow neural network operating on probability distribution models and their corresponding statistical divergences.

The contributions provided by the papers are as follows: *(i)* We propose a learning framework that is able to classify based upon a combination of Renyi divergences over prior distributions. *(ii)* Given SAR data distributions, we propose

to learn the α parameter from the data and derive the necessary gradients to perform an end-to-end optimisation, which to the best of our knowledge has not been studied yet in this scenario. Noticeably, we provide analytical expressions for the derivative for log-normal, gamma and Rayleigh distributions. *(iii)* We perform a study on a real dataset in mountainous regions and showcase both the explainable nature of the learned parameters and the better robustness of the approach, in comparison with CNN and Random Forest algorithms.

2. SAR DISTRIBUTION MODELLING

We consider amplitude SAR images meaning that the pixel values belongs in $\mathbb{R}^+$. To model such data, extensive probability models have been proposed [15, 16] to account for the varaibility in the scenes observed. In this paper, we consider the following probability models:

(i) Gamma: $\mathcal{G}(x;\mu,L) = e^{-\frac{xL}{\mu}} \cdot \left(\frac{L}{\mu}\right)^L \cdot \Gamma(L) \cdot x^{L-1}$, with shape and scale L and μ,

(ii) log-normal: $\mathcal{O}(x;\mu,\sigma) = e^{-\frac{(\log x-\mu)^2}{2\sigma^2}} \cdot \frac{1}{x\sigma\sqrt{2\pi}}$, with mean μ and variance σ,

(iii) Rayleigh: $\mathcal{R}(x;\mu) = \frac{x}{2\mu^2} \cdot e^{-(\frac{x}{2\mu})^2}$ with scale μ,

that have shown ability to accurately model amplitude SAR data with a few easily estimable parameters. To estimate the parameters of those probability models, we rely on an estimation of the log-cumulants of the distributions that are directly linked to the statistical parameters. This consideration allows to compromise between a good estimation and a very low computational cost [17]. The estimate of the log-cumulant of order r is given by the recurrence formula:

$$\hat{k}_r = \hat{m}_r - \sum_{i=1}^{r-1} \binom{r-1}{i-1} \hat{k}_i \hat{m}_{r-i}, \tag{1}$$

with $\hat{k}_1 = \hat{m}_1$, $\hat{m}_r = \frac{1}{n}\sum_{i=1}^n \ln z_i^r$, $r \in \mathbb{N}, r > 1$ and z_i, being a sample of a continuous random variable Z with a probability density function defined over $\mathbb{R}^+$.

Concerning divergences we rely on the Renyi parameterization for its high-generalisation capabilities with respect to a wide class of well-known divergences [18]. The Rényi divergence of order α between two probability distributions P, Q on $\mathbb{R}^n$ can be expressed as follows:

$$D_\alpha(P\|Q) = \frac{1}{\alpha-1} \ln \int P(x)^\alpha Q(x)^{1-\alpha} dx, \tag{2}$$

with $\alpha > 0$ and $\alpha \neq 1$. This parametric divergence is interesting due to its capacity to represent the Kullback-Leibler divergence as α tends to 1. It encompasses special cases such as the Bhattacharyya distance when $\alpha = 0.5$ and becomes proportional to a χ^2 divergence when $\alpha = 2$. In contrast with approaches that determine and fix α for subsequent use, we propose a framework where α is a learnable parameter with respect to the data and class distributions, hence learning the suitable metrics for the optimization problem (4). Thanks to [19], we have the analytical forms of the Rényi divergence for the three distributions considered, given their parameters and α. Due to a non-definition for the case $\alpha = 1$, we restrict the feasible set to $\alpha \in [0, 1[$.

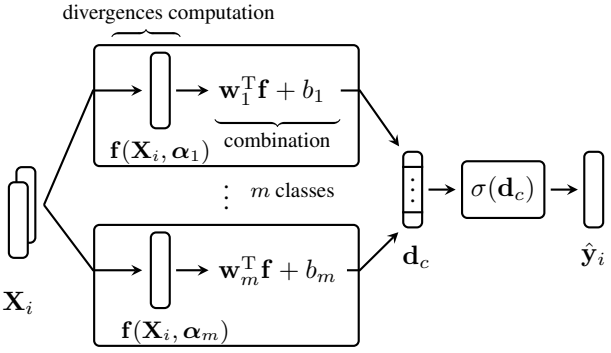

Fig. 1: Summary of the proposed classification approach.

3. DIVERGENCE LEARNING FRAMEWORK

We consider a problem of supervised multi-class classification. Given an ensemble of n labelled data $\mathcal{E} = \{(\mathbf{X}_i, \mathbf{y}) \in \mathbb{R}^{v\times 2} \times \{0,1\}^m : 1 \leq i \leq n\}$, where $\mathbf{X}_i$ corresponds to a pair of vectors of v features that are the estimated parameters of the prior distributions. They are associated with a one hot encoding label vector $\mathbf{y}_i(c) = [\delta_{1c}, \ldots, \delta_{mc}]^{\mathrm{T}}$, m being the number of classes, δ the kronecker symbol with $c \in \{1, \ldots, m\}$, corresponding to the class of the pair. One of those classes corresponds to a dissimilarity class between the two images of the pair. Let $\mathbf{f} = [D_1(\mathbf{X}_i, \alpha_1), ..., D_p(\mathbf{X}_i\alpha_p)]^{\mathrm{T}}$ be a vector composed by a set of p Renyi divergences with parameters $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_p]^{\mathrm{T}} \in (0,1)^p$. We construct a network, as shown in Figure 1, that computes for each class c, a vector of divergences, given a set of parameters $\boldsymbol{\alpha}_c$ between the pair and combines them with:

$$d_c(\mathbf{X}_i, \boldsymbol{\alpha}_c) = \mathbf{w}_c^{\mathrm{T}} \mathbf{f}(\mathbf{X}_i, \boldsymbol{\alpha}_c) + b_c, \tag{3}$$

where $\mathbf{w}_c \in \mathbb{R}_+^p$ and $b_c \in \mathbb{R}_+$. The positivity of the weights and the bias are enforced to preserve explainability of the final combination. Then all the obtained combinations are fed into a softmax operator ($\sigma(\mathbf{d}_c) = \frac{\mathrm{e}^{d_c}}{\sum_i^m \mathrm{e}^{d_i}}$) to output the class probabilities.

We consider the cross-entropy loss defined as $H(y, \hat{y}) = -\sum_c^m y_c \log(\hat{y}_c)$, with $\hat{y}, y \in \mathbb{R}^m$ respectively be the prediction of a classifier and its associated ground truth. This gives us the following minimization problem:

$$\underset{\substack{\forall c\in\{1,\ldots,m\},\\ \boldsymbol{\alpha}_c\in(0,1)^p,\\ \mathbf{w}_c\in\mathbb{R}_+^p, b_c\in\mathbb{R}_+}}{\arg\min} \frac{1}{n}\sum_{i=0}^{n} \underbrace{-\sum_c^m \mathbf{y}_i(c) \log\left[\sigma \circ d_c(\mathbf{X}_i, \boldsymbol{\alpha}_c)\right]}_{\mathcal{L}_i}. \tag{4}$$

Divergence	derivate $\partial D_{\hat{\alpha}_{(\cdot)}}/\partial\alpha_{(\cdot)}$
$D_{\hat{\alpha}_l}(\mathcal{G}_i\|\mathcal{G}_j)$	$e^{\alpha_l}\frac{L_i\mu_j-L_j\mu_i}{\lambda_{ij}\beta_{ij}} - e^{\alpha_l}L_i\log\left[\frac{(e^{\alpha_l}+1)\mu_i\mu_j}{\beta_{ij}}\right] - e^{\alpha_l}\log\left[\frac{\left(\frac{\mu_i}{L_i}\right)^{-L_i}\Gamma(\lambda_{ij})}{\Gamma(L_i)}\right] + e^{\alpha_l}\frac{(L_j-L_i)\psi^{(0)}(\lambda_{ij})}{e^{\alpha_l}+1}$
$D_{\hat{\alpha}_l}(\mathcal{O}_i\|\mathcal{O}_j)$	$\frac{e^{\alpha_l}}{2(e^{\alpha_l}\Sigma_{ij})^2}\left[e^{\alpha_l}\sigma_j^2(\sigma_j^2-\sigma_i^2)+\sigma_i^2\left[(\mu_i-\mu_j)^2-\Sigma_{ij}\right]-(e^{\alpha_l}\Sigma_{ij})^2\log\left(\frac{(e^{\alpha_l}+1)\sigma_j^2}{e^{\alpha_l}\Sigma_{ij}}\right)\right]$
$D_{\hat{\alpha}_l}(\mathcal{R}_i\|\mathcal{R}_j)$	$\frac{\gamma_{ij}-e^{\alpha_l}\mu_i^2}{\gamma_{ij}+\mu_i^2}-e^{\alpha_l}\log\left[\frac{\gamma_{ij}+\mu_j^2}{\gamma_{ij}+\mu_i^2}\right]$

Table 1: Rényi's derivate, with $\lambda_{ij}=(e^{\alpha_l}L_i+L_j)/(1+e^{\alpha_l})$, $\beta_{ij}=e^{\alpha_l}L_i\mu_j+L_j\mu_i$, $\Sigma_{ij}=\sigma_j^2+\sigma_i^2$ and $\gamma_{ij}=e^{\alpha_l}\mu_j^2$.

To solve this problem, we use a gradient descent approach on the weights $\mathbf{w}_c$ and bias b_c, as well as Renyi parameters α. In order to enforce the domain constraints on the variables, we use a reparameterization: $\hat{\mathbf{w}}=\mathrm{e}^{\mathbf{w}}$, $\hat{b}=\mathrm{e}^{b}$, and $\hat{\alpha}=1/(1-\mathrm{e}^{-\alpha})$[1]. To compute the gradient we compute the derivatives of the loss, thanks to the chain-rule[2]:

$$\frac{\partial\mathcal{L}_i}{\partial d_c}=\hat{y}_c-y_c,\ \frac{\partial\mathcal{L}_i}{\partial\mathbf{w}_c}=\mathbf{f}\cdot\mathrm{e}^{\mathbf{w}_c}\cdot(\hat{y}_c-y_c),$$
$$\frac{\partial\mathcal{L}_i}{\partial b_c}=\mathrm{e}^{b_c}\cdot(\hat{y}_c-y_c),$$

For Renyi parameters, we have:

$$\frac{\partial\mathcal{L}_i}{\partial\boldsymbol{\alpha}_c}=\frac{\partial\mathcal{L}_i}{\partial d_c}\frac{\partial d_c}{\partial\mathbf{f}}\frac{\partial\mathbf{f}}{\partial\boldsymbol{\alpha}_c},\ \text{where}\ \frac{\partial d_c}{\partial\mathbf{f}}=\mathbf{w}_c,$$

and $\frac{\partial\mathbf{f}}{\partial\boldsymbol{\alpha}_c}$ is matrix of derivatives computed thanks to the results given in Table 1.

4. EXPERIMENTS

4.1. Real SAR dataset

To evaluate our framework, we create an X-band SAR dataset from the PAZ satellite. Operating at a frequency of 9.65 GHz, this satellite delivers images at approximately 11-day intervals, offering a resolution of 3 meters per pixel in two polarizations, HH and HV. We specifically examine six linear amplitude acquisitions spanning from July 7, 2020, to August 26, 2020, encompassing the Mont Blanc massif. In total, we have extracted 645 patches, each with size 32×32 pixels, during this time interval. These patches are categorized into five distinct classes: glacier accumulation area, city, forest, plain or land and rocky area, respectively named ACC, CIT, FOR, PLA and ROC for the results. The labeling of the images was done using the CORINE land cover model provided by Copernicus[3]. We then consider the balanced version of all the pairs in this dataset, adding the dissimilar class (DIS), corresponding to the pair where the class of each image is different.

[1]For conciseness, we keep the notation $\mathcal{L}_i$ while replacing the parameters by their reparameterizations.

[2]The details of derivations are omitted due to page limitations.

[3]© European Union, Copernicus Land Monitoring Service 2018, European Environment Agency (EEA)

4.2. Model architectures and parameterization

	RF	CNN	Rényi
input size	200×1	$32\times32\times4$	10×2
parameters	$\sim152,000$	$226,406$	222

Table 2: Number of parameters and size of the inputs used

We compare our approach with two other methods: RandomForest (RF) and CNN. RandomForest is computationally efficient and offers decision transparency through threshold analysis but may require many nodes depending on features, making interpretation complex and sensitive to noise and outliers. In our case, we use Principal Component Analysis (PCA) to reduce the dimension followed by a random forest with 100 trees.

On the other hand, Convolutional Neural Networks capture detailed spatial information but require a large sample size and lack transparency. Due to the limited sample data available, we created a shallow CNN as in [20]. The CNN takes a $32\times32\times4$ image pair as input and consists of 3 successive blocks. Each block includes a 3×3 convolution followed by a 2×2 max-pooling operation. The network then involves a dense aggregation layer of size 256, followed by a final dense layer for class prediction. In total, this CNN comprises 226,406 parameters.

Our Rényi approach operates on divergences between image channel distributions (gamma, log normal, and Rayleigh). We define these combinations as $\{\|;=;\perp\},ij$, where ij specifies the images for which the divergence is computed, $\|$ and $=$ denote co-polarization and cross-polarization divergences (e.g., $D(HH_i,HH_j)$ and $D(HV_i,HV_j)$). $\perp$ represents cross divergences between co-polarization and cross-polarization (e.g., $D(HH_{i/j},HV_{i/j})$, $D(HH_i,HV_j)$ and $D(HV_i,HH_j)$). This provides 18 divergences for 6 classes, weighted by 6 vectors with 18 weights and 6 biases. The number of parameters and the size of the inputs are described in the Table 2.

4.3. Global evaluation

We first assess our results using a stratified K-fold with K=5 to maintain class balance while analyzing the entire dataset. The outcomes are presented in Table 3, which displays the average classification percentages for each class across the three methods. We observed that both the CNN and our approach yielded similar results, outperforming RF, but with distinct

	ACC	CIT	DIS	FOR	PLA	ROC
RF	65.3 ± 13.0	70.8 ± 9.2	82.7 ± 4.1	11.9 ± 1.2	33.6 ± 5.8	55.9 ± 9.0
CNN	$\mathbf{83.5 \pm 7.0}$	61.1 ± 16.5	$\mathbf{82.9 \pm 4.3}$	45.1 ± 8.1	49.5 ± 13.0	$\mathbf{72.3 \pm 1.2}$
Renyi	59.1 ± 11.1	$\mathbf{83.2 \pm 4.2}$	45.3 ± 1.2	$\mathbf{80.5 \pm 6.9}$	$\mathbf{67.3 \pm 3.7}$	62.7 ± 12.0

Table 3: Percentage of good classification with a stratified K-Fold with K=5

strengths. The CNN excelled in detecting glaciers, rocky areas, and dissimilar pairs, whereas our method demonstrated superiority in classifying cities, forests, and plain areas. Notably, our method showed slightly improved score consistency across the folds. The most substantial disparity in performance was evident in the dissimilar class, where the CNN performed approximately 2 times better. This discrepancy can be attributed to our use of divergences from the same distribution, which may not effectively model the diverse class combinations inherent to dissimilar pairs. However, one can note that for the same experiment, RF performs 7 times lesser and CNN 2 times lesser than the divergence learning approach proposed on the "forest" class: a CNN needs to be very deep and needs a large amount of larger size texture samples in general for being efficient on textures.

4.4. Learned parameters and explicability

We examine the learned parameters, $\mathbf{w}_c$ and α, and visualize them in Figure 2 and Figure 3, respectively. In Figure 2, we observe that class decisions primarily rely on 3 or 4 divergences, with variations among classes. Notably, most classes employ a Rayleigh distribution, validating its suitability for single-look amplitude images of homogeneous areas. Additionally, cross-divergences are utilized by all classes to emphasize the significance of polarimetric channels and their correlations. Figure 3 displays the distribution of learned Rényi α parameters for each class. Notably, there is a strong concentration around 0.5, homogeneous to the Bhattacharyya distance, which is consistently employed across all classes. Additionally, we note the use, mainly for the forest and plain classes, of a distance approaching the Kullback-Leibler.

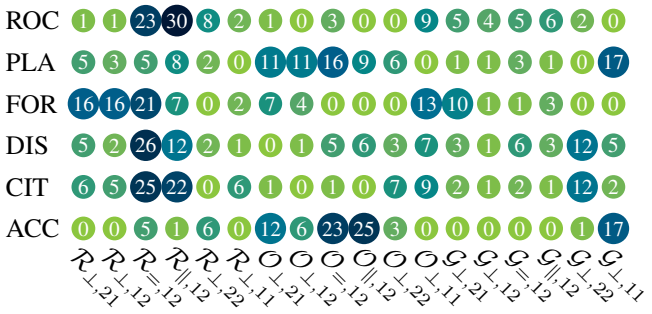

Fig. 2: Visualisation of associated weights (mean over the folds) in the decision process for each class. The values and color correspond to a percentage of the total weights sum.

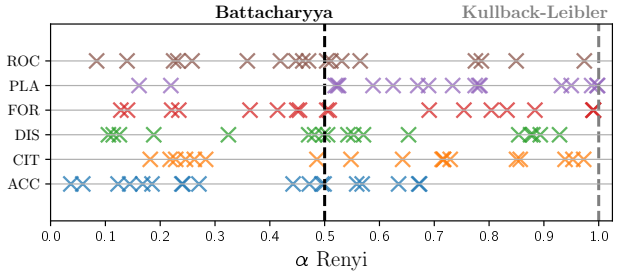

Fig. 3: α learned for each features and for each classes.

4.5. Sensitivity to perturbation

We assess our approach under two disturbance scenarios, as depicted in Figure 4. We gauge performance using the weighted F1 score averaged across all classes and employ four different seeds for score variation. In Figure 4.a, we examine methods sensitivity concerning training data variations. Our approach exhibits superior resilience, maintaining performance even when training data decreases by 20%, while the CNN's performance starts to decline at approximately 35% of data decrease. Figure 4.b illustrates changes in the percentage of samples used for training. Our approach demonstrates greater robustness when confronted with reduced training data, owing to the statistical assumptions we introduce that mitigate the impact of learning through data.

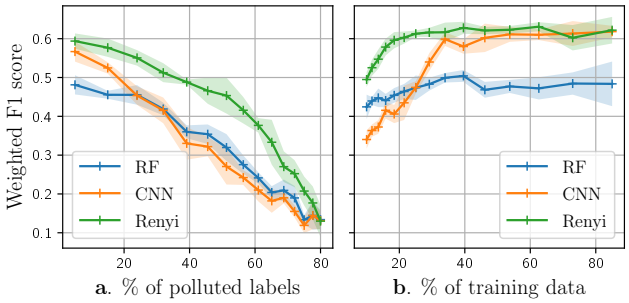

Fig. 4: Comparison of performance (mean of weighted f1 score over all class in function) of two perturbations, **a**. Percentage of label perturbation and **b**. Percentage of data training

5. CONCLUSION

We have explored the use of a shallow divergence based learning algorithm for pairs of SAR image classification. By modelling a physical prior thanks to probability distributions and incorporating divergences of those models into the learning process, we are able to obtain a comparable performance than regular approaches with no priors while keeping the number of parameters learned very low, as well as yielding explainable results. The proposed approach has also been shown more robust in case of data mislabelling and when fewer samples for training are applicable. Future directions of research will include the computation of divergences between different mono-variate distribution to enrich the dictionary of divergences used, dealing with unbalanced data and extend the domain on α parameter learning.

6. REFERENCES

[1] Laura Dingle Robertson, Andrew Davidson, Heather McNairn, Mehdi Hosseini, and Scott Mitchell, "Assessment of multi-frequency sar for crop type classification and mapping," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 489–492.

[2] A.V. Bogdanov, S. Sandven, O.M. Johannessen, V.Yu. Alexandrov, and L.P. Bobylev, "Multisensor approach to automated classification of sea ice image data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 7, pp. 1648–1664, 2005.

[3] Rudiger Gens and John L Van Genderen, "Review article sar interferometry—issues, techniques, applications," *International journal of remote sensing*, vol. 17, no. 10, pp. 1803–1835, 1996.

[4] B. Aiazzi, L. Alparone, and S. Baronti, "Land cover classification of urban and sub-urban areas via fuzzy nearest-mean reclustering of sar features," in *2003 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, 2003, pp. 62–66.

[5] Rong Duan, Wei Jiang, and Hong Man, "Semi-supervised image classification in likelihood space," in *2006 International Conference on Image Processing*, 2006, pp. 957–960.

[6] Rodney V. Fonseca, Rogério G. Negri, Aluísio Pinheiro, and Abdourrahmane Mahamane Atto, "Wavelet spatio-temporal change detection on multitemporal sar images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 4013–4023, 2023.

[7] Rizwan Ahmed Ansari, Krishna Mohan Buddhiraju, and Rakesh Malhotra, "Urban change detection analysis utilizing multiresolution texture features from polarimetric sar images," *Remote Sensing Applications: Society and Environment*, vol. 20, pp. 100418, 2020.

[8] Sizhe Chen, Haipeng Wang, Feng Xu, and Ya-Qiu Jin, "Target classification using the deep convolutional networks for sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4806–4817, 2016.

[9] Shanshan Mu, Xiaofeng Li, and Haoyu Wang, "The fusion of physical, textural, and morphological information in sar imagery for hurricane wind speed retrieval based on deep learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[10] Chandrabali Karmakar, Corneliu Octavian Dumitru, Gottfried Schwarz, and Mihai Datcu, "Feature-free explainable data mining in sar images using latent dirichlet allocation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 676–689, 2021.

[11] Michéle Basseville, "Divergence measures for statistical data processing—an annotated bibliography," *Signal Processing*, vol. 93, no. 4, pp. 621–633, 2013.

[12] Hatice Kubra Cilingir, Rachel Manzelli, and Brian Kulis, "Deep divergence learning," in *International Conference on Machine Learning*. PMLR, 2020, pp. 2027–2037.

[13] Wagner B Silva, Corina C Freitas, Sidnei JS Sant'Anna, and Alejandro C Frery, "Classification of segments in polsar imagery by minimum stochastic distances between wishart distributions," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 3, pp. 1263–1273, 2013.

[14] R. Hänsch, A. Ley, and O. Hellwich, "Correct and still wrong: The relationship between sampling strategies and the estimation of the generalization error," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017, pp. 3672–3675.

[15] Charles-Alban Deledalle, L. Denis, G. Ferraioli, and Florence Tupin, "Mathematical models for very high resolution SAR data and their applications," in *Mathematical Models for Remote Sensing Image Processing*. Springer, 2018.

[16] Dong-Xiao Yue, Feng Xu, Alejandro C. Frery, and Ya-Qiu Jin, "Synthetic aperture radar image statistical modeling: Part one-single-pixel statistical models," vol. 9, no. 1, pp. 82–114, Number: 1.

[17] Jean-Marie Nicolas, "Application de la transformée de mellin : étude des lois statistiques de l'imagerie cohérente," .

[18] Tim Van Erven and Peter Harremos, "Rényi divergence and kullback-leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, 2014.

[19] Manuel Gil, Fady Alajaji, and Tamas Linder, "Rényi divergence measures for commonly used univariate continuous distributions," *Information Sciences*, vol. 249, pp. 124–131, 2013.

[20] Hemani Parikh, Samir Patel, and Vibha Patel, "Classification of SAR and PolSAR images using deep learning: a review," vol. 11, no. 1, pp. 1–32, Number: 1.