

# EXPLORING THE HIGH DIMENSIONAL GEOMETRY OF HSI FEATURES

Wojciech Czaja<sup>1,2</sup>, Ilya Kavalero<sup>3</sup>, Weilin Li<sup>4</sup>

<sup>1</sup>Department of Mathematics, University of Maryland, College Park

<sup>2</sup>Center for Scientific Computation and Mathematical Modeling, University of Maryland, College Park

<sup>3</sup>UMIACS, University of Maryland, College Park

<sup>4</sup>Courant Institute of Mathematical Sciences, New York University

## ABSTRACT

We explore feature space geometries induced by the 3-D Fourier scattering transform and deep neural network with extended attribute profiles on four standard hyperspectral images. We examine the distances and angles of class means, the variability of classes, and their low-dimensional structures. These statistics are compared to that of raw features, and our results provide insight into the vastly different properties of these two methods. We also explore a connection with the newly observed deep learning phenomenon of neural collapse.

**Index Terms**— hyperspectral images, features, Fourier scattering transform, extended attributes profile, deep learning, neural collapse

## 1. INTRODUCTION

Some of the fundamental methods in pattern recognition and machine learning are based on simple geometric notions. Support vector machines split data points into two classes by a maximum margin hyperplane,  $k$ -means clustering attempts to group points into isotropic balls, and nearest neighbor classifiers exploit notions of proximity and spatial regularity. Modern manifold learning algorithms such as Laplacian Eigenmaps and diffusion maps attempt to find hidden geometric structures within data. Without a question, geometry of the features play an important role in both supervised and unsupervised learning.

While it was thought that the ideal geometries or configuration of features should vary by dataset, data type, number of dimensions, and feature transform, there is growing empirical evidence that there are universal and optimal configurations,

---

Wojciech Czaja acknowledges support from the NSF under the DMS grant # 1738003 and from the ARO under the grant # W911NF-17-1-0304.

This work was supported in part by the MURI from the Army Research Office under the Grant No. W911NF-17-1-0304. This is part of the collaboration between US DOD, UK MOD and UK Engineering and Physical Research Council (EPSRC) under the Multidisciplinary University Research Initiative.

Weilin Li thanks the AMS-Simons Travel Grant for support.

and that appropriately trained neural networks yield such configurations [1]. More precisely, that paper coined the term “neural collapse” to refer to a phenomenon where, informally speaking, after completion of the training process, the neural network maps all data points from a single class to the same point in feature space, and that these points after recentering to their mean form the vertices of a (possibly rotated and scaled) simplex ETF (equiangular tight frame). There are several theoretical explanations for why neural collapse occurs – typically arguing that a simplex ETF frame minimizes a certain functional under constraints on the features [1, 2, 3] – but there are many open questions as to why neural networks do this to begin with and what the generalization advantages are.

Inspired by recent innovations, we examine the geometry of the features induced by two state-of-the-art methods for HSI classification.

- 3DFST (three-dimensional Fourier scattering transform) is a feature extractor formed by cascading the complex modulus with convolutions against functions generated by time-frequency shifts. This is a data-independent transform that combines the strengths of spectral decomposition with hierarchical methods [4, 5], and was used for HSI classification [6, 7] and radioactive anomaly detection [8].
- EAP (deep learning with extended attribute profile) [9, 10] projects the HSI spectra to several PCA components preserving most of the spectral energy and then creates attribute profiles (APs) for the 2D images corresponding to each component. The APs connect similar 2D structures through the morphological operations of erosion and dilation and create a more homogeneous version of the original 2D principle component image. Finally the cube of EAPs is passed to a 5 layer Neural Network which performs the classification.

Due to space constraints, we refer the reader to above references for further explanations.

We examine the features generated by both methods on four standard hyperspectral datasets: Indian pines, Kennedy Space Center, Botswana, and Pavia University. We emphasize that the purpose of this proceeding is to examine the ge-

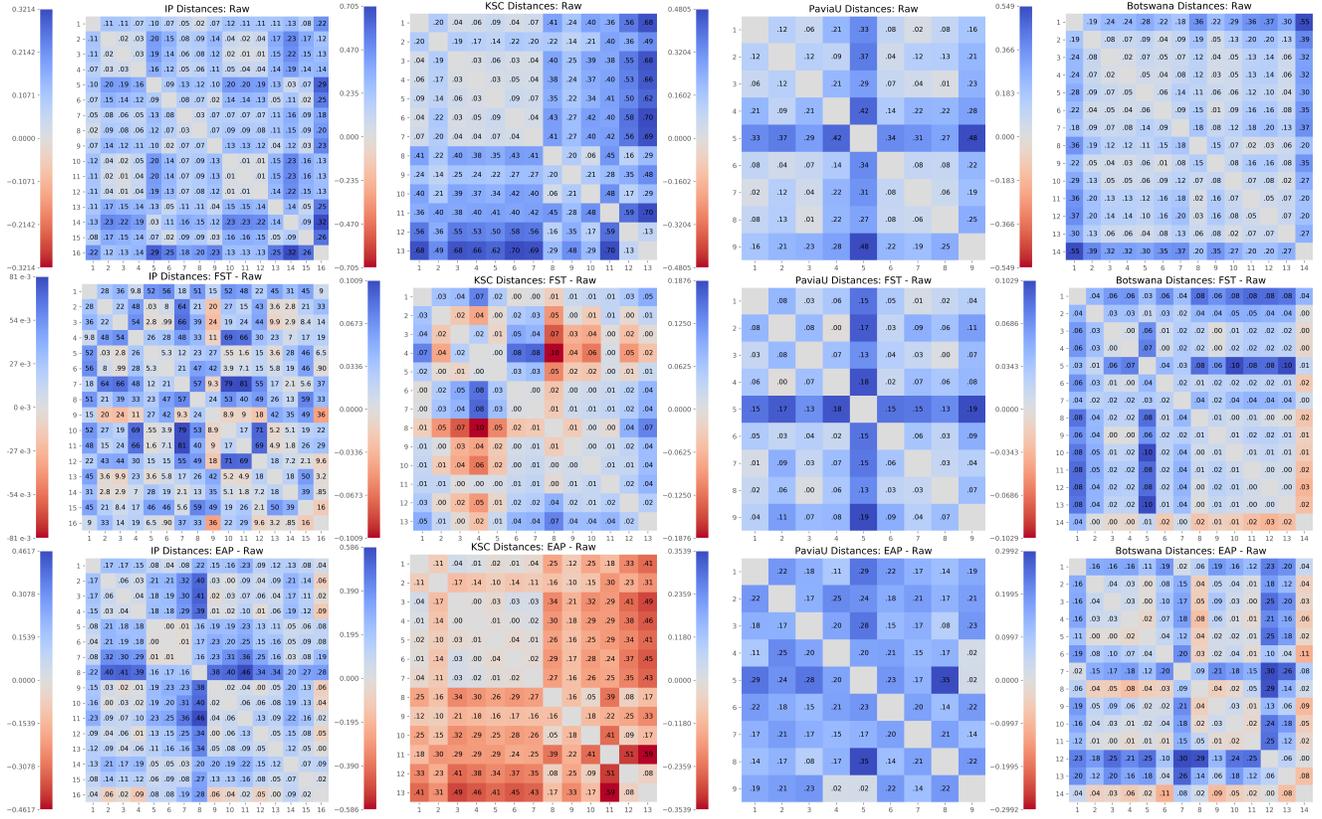


Fig. 1: Top row: class mean distances for raw features. Middle row: 3DFST minus raw. Bottom row: EAP minus raw.

ometries of the features, not to compare their classification performances. For the later, we refer the reader to [6].

## 2. METHODOLOGY

For each hyperspectral dataset and choice of features (raw, 3DFST, or EAP), we have access to a collection of vectors  $\{\mathbf{x}_k\}_{k=1}^n$  in  $\mathbb{R}^p$  with associated labels  $\{y_k\}_{k=1}^n$ . It is worth mentioning that the focus of this paper is not on classification performance – there is no training or test split in this paper. Here  $n$  is the total number of labeled points in the HSI.

The dimension  $p$  of the feature domain varies by dataset and algorithm, so in order to have a fair comparison, we normalize each  $\mathbf{x}_k$  by dividing by the maximum Euclidean norm over all the features. Doing so, we can assume each  $\mathbf{x}_k$  is contained within the unit Euclidean ball of their respective domains. This has no impact on the results we present below, since we will examine quantities that are invariant to global scaling. Throughout, we only consider the Euclidean distance  $\|\cdot\|$  between points.

We briefly summarize the parameters of the methods we use. All code for our feature extraction methods is open sourced at <https://github.com/ilyakava/pyfst>.

3D FST performs 3 stages of 3D filters, interspersed with downsampling and non linearities. For these hyperparameters

we follow [6] which found the best sizes of 3D filters to use for HSI 3D Scattering via a grid search on small validation sets. These 3D filter sizes vary by dataset. Classification is then performed with a linear svm with regularization parameter fixed to  $C = 1000$ .

The EAP method performs PCA, finds morphological profiles, and then uses a neural network to classify. For the parameters of these we follow [9, 10], which uses 4 PCA components and generates 9 morphological profiles for each component. Then, a square window of size 9 in the spatial dimension is treated as a single input into the Neural Network, which generates a class prediction. This  $9 \times 9 \times 36$  cube is passed to a neural network containing 3 2D convolutional layers (with spatial sizes 5, 3, 3 and channel lengths 48, 96, 96) and 2 Fully Connected layers of size 1024 before a softmax classifier. Thus the penultimate feature size before classification is always 1024.

## 3. GEOMETRY OF CLASS MEANS

HSI features live in a high-dimensional domain and their geometric arrangement carries important information that affects classification performance. In our quest to explore such a wild landscape, we first investigate the behavior of features near their class means, which can be viewed as the simplest thing to examine, but not necessarily the most illuminating or

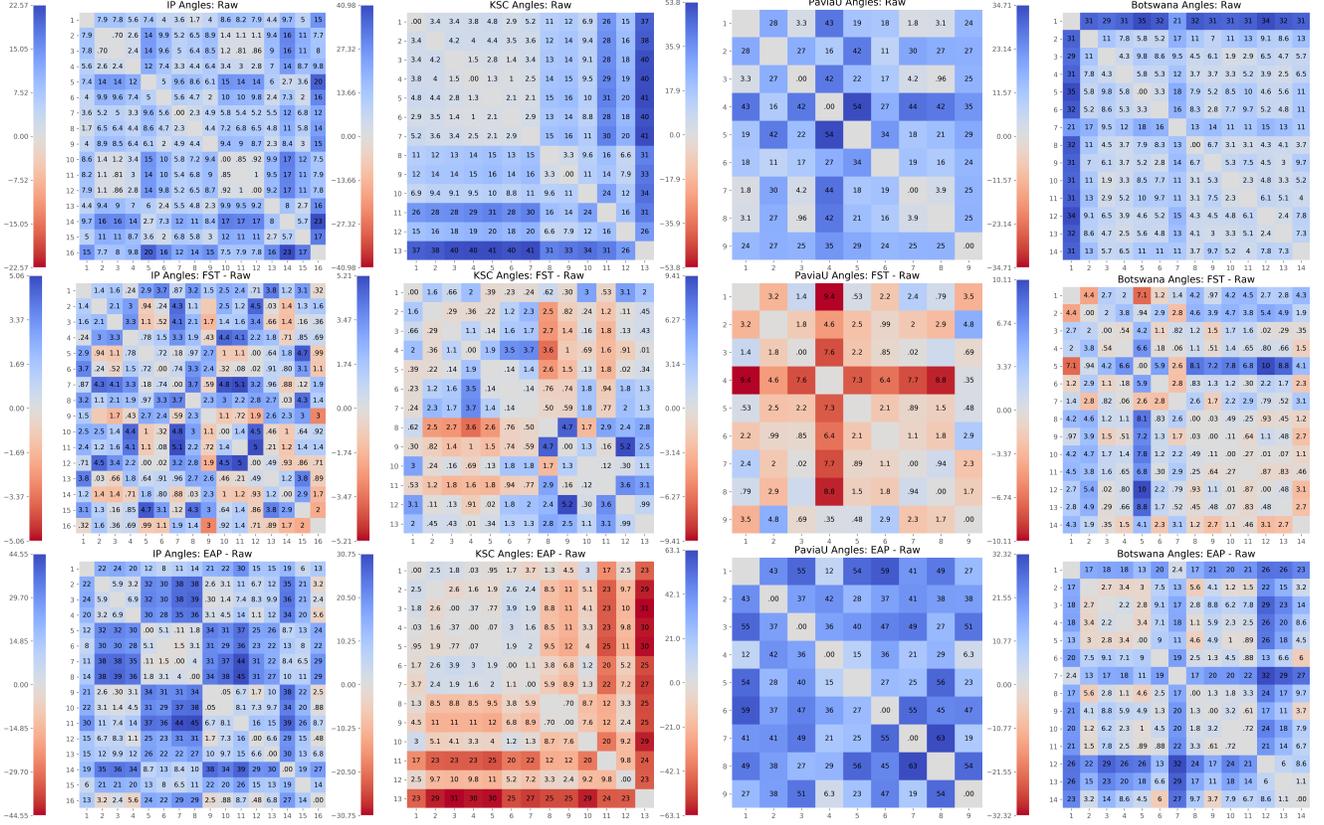


Fig. 2: Top row: class mean angles for raw features. Middle row: 3DFST minus raw. Bottom row: EAP minus raw.

comprehensive one. More formally, the mean  $\mathbf{m}_j$  of the  $j$ -th class is simply the average of all features  $\mathbf{x}_{k_c}$  belonging to that class. To measure angles between class means, we first subtract the center of all class means (found by averaging all  $\mathbf{m}_j$ ) from each class mean  $\mathbf{m}_j$ , and use the usual definition for the angle between two vectors,  $\cos \theta(\mathbf{u}, \mathbf{v}) := (\mathbf{u} \cdot \mathbf{v}) / (\|\mathbf{u}\| \|\mathbf{v}\|)$ .

Raw class mean distances and their changes are shown in Figure 1. Compared to distances between class means of raw data, 3DFST expands the distances for most pairs of classes on most datasets. There are several instances, such as in KSC, where the transform reduces the distances between class means. On the other hand, EAP being a trained neural network, is more expressive and significantly alters the arrangement of the class means. The results on KSC are visibly different from the others, primarily because classes 1–7 have means that are all tightly clustered together.

Raw class mean angles and their changes are shown in Figure 2. We see that for most pairs of classes, 3DFST expands class mean angles. Consistent with our earlier findings on class mean changes, EAP results in larger changes to the angles between class means. The results on KSC can be explained by the raw features of classes 1–7 being tightly clustered together. Interestingly though, rather than expand the angles between these seven classes, EAP contracts the angles between the remaining classes.

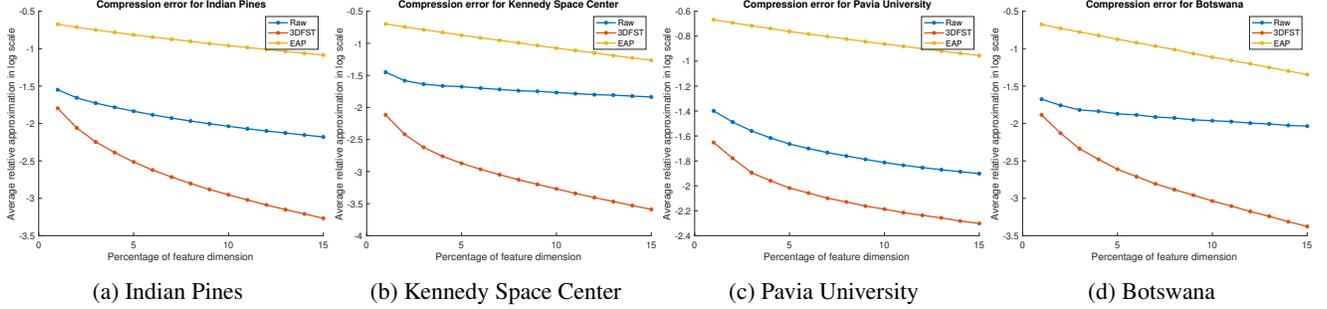
#### 4. COMPRESSION AND LOW-DIMENSIONALITY

The geometry of class means provides a first-order statistical description of the high-dimensional geometry of HSI features. Intra-class variability can be quantified by second order statistics such as squared deviation away from the mean. Table 1 displays the average Euclidean distance between feature vectors and their means. Across all four data sets, both feature transforms result in significant changes to the average radius of each class. However, EAP expands the average class radius by a significantly greater amount than 3DFST. This illustrates an important trade-off: while classes are generally mapped further apart and with greater angles by EAP, the cluster sizes themselves appear to expand.

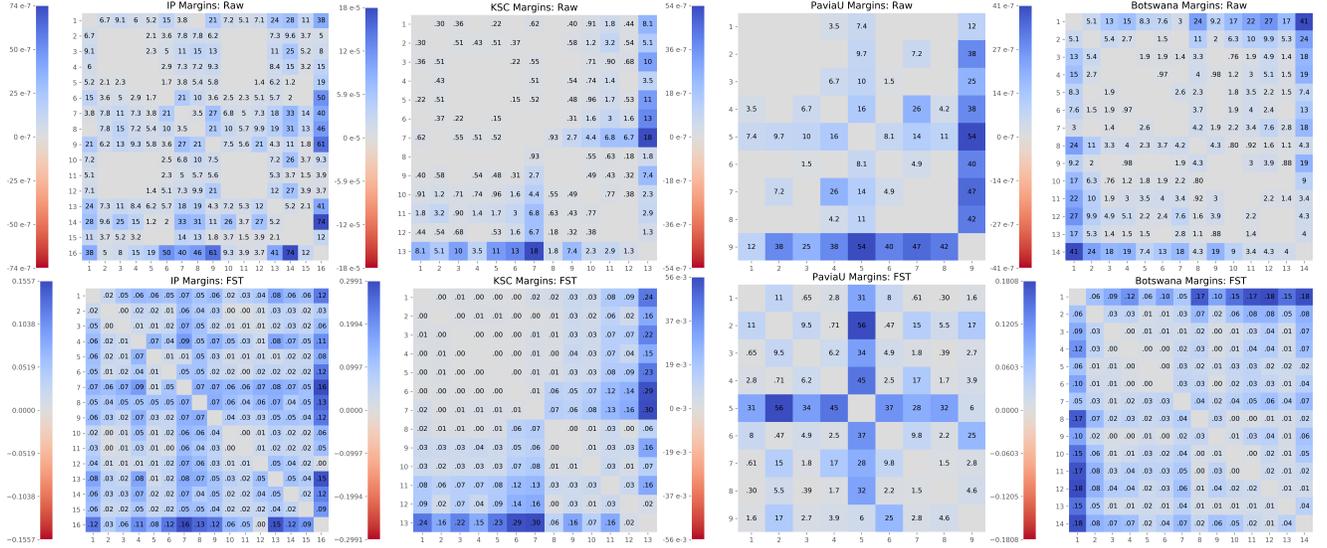
Our next set of experiments explores whether class features are isotropic (look like balls) or anisotropic (look like ellipsoids). To do this, we approximate each class with the best possible low-dimensional hyperplane, whose dimension is variable. The total approximation or compression error measured in the Euclidean norm decreases as the hyperplane dimension increases. This provides insight into class geometries: for anisotropic classes, the error decreases faster than that of isotropic ones. Our results summarized in Figure 3 indicate an important distinction between the tendencies of 3DFST and EAP: the former creates anisotropic features rel-

	Indian Pines	KSC	PaviaU	Botswana
Raw	$0.0551 \pm 0.0347$	$0.0722 \pm 0.0616$	$0.0465 \pm 0.0377$	$0.0283 \pm 0.0202$
3DFST	$0.0817 \pm 0.0372$	$0.0629 \pm 0.0469$	$0.0714 \pm 0.0522$	$0.0406 \pm 0.0281$
EAP	$0.1735 \pm 0.0449$	$0.1774 \pm 0.0246$	$0.1394 \pm 0.0752$	$0.1659 \pm 0.0371$

**Table 1:** Average class variability  $\pm$  standard deviations



**Fig. 3:** Average approximation error



**Fig. 4:** Maximum margins between classes. Top row: raw features. Bottom row: 3DFST features. Blank spots in the top row indicate class pairs that are not linearly separable.

	Indian Pines	KSC	PaviaU	Botswana
Neural collapse angle	93.8226	94.7802	97.1808	94.4117
EAP class distances	$0.2488 \pm 0.1281$	$0.1015 \pm 0.0614$	$0.3710 \pm 0.1358$	$0.2294 \pm 0.1342$
EAP class angles	$25.9985 \pm 14.9650$	$6.0478 \pm 3.3516$	$61.7212 \pm 12.3580$	$21.2320 \pm 15.7333$

**Table 2:** Average distances and angles between classes (in degrees)  $\pm$  standard deviations

ative to the feature space dimension, whereas the latter creates more isotropic classes. In fact, our results indicate that EAP features are more isotropic than the raw spectra.

## 5. 3DFST AND LINEAR SEPARABILITY

Our empirical observations indicate that 3DFST takes a more conservative approach. Rather than transform classes across large distances, it contracts variability along the normal directions. This has significant on the linear separability between classes. As shown in Figure 4, even if raw spectra belong-

ing to different classes are linearly separable, their distances to the maximum margin hyper-plane is small and on the order of  $10^{-4}$ . On the other hand, for every pair of classes in each dataset, their 3DFST features are linearly separable by hyperplanes whose margins are orders of magnitudes larger. The margin of a hyper-plane has important implications to the generalization error when a subset of the data points are used to determine a hyperplane split [11, 12].

## 6. EAP AND NEURAL COLLAPSE?

It is natural to wonder if EAP induces neural collapse or similar. While there are many parallels, such as EAP generally amplifying the distances and angles between class means, we point out several inconsistencies. In neural collapse, the distances between class means would be constant and the class means (after re-centering) would form a (scaled and rotated) simplex ETF. The angles between vectors in neural collapse are  $\cos^{-1}(-\frac{1}{m-1})$ , where  $m$  is the number of classes. For the four datasets used in this article, Table 2 summarizes the distances and angles between classes for EAP features (recall that the class means are shifted by the center of these means). These numbers are not consistent with what qualifies as neural collapse.

## 7. INSIGHTS AND DISCUSSION

In this paper we examined features generated by two state-of-the-art algorithms: 3DFST and EAP. We have demonstrated that they operate on drastically different geometric principles even though they yield comparable classification results. The results in this article naturally raises two questions.

The first question is what is the desired architecture for processing HSI? Compared with the state-of-the-art image classification networks, the ones considered here are rather shallow. The second question is whether neural collapse holds for other distributions, like those present in the HSI setting? We saw that both features extractors, while significantly different from each other, do not generate features consistent with neural collapse.

## 8. REFERENCES

- [1] Vardan Papyan, X. Y. Han, and David L Donoho, “Prevalence of neural collapse during the terminal phase of deep learning training,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 40, pp. 24652–24663, 2020.
- [2] Weinan E and Stephan Wojtowytsch, “On the emergence of tetrahedral symmetry in the final and penultimate layers of neural network classifiers,” *arXiv preprint arXiv:2012.05420*, 2020.
- [3] Dustin G. Mixon, Hans Parshall, and Jianzong Pi, “Neural collapse with unconstrained features,” *arXiv preprint arXiv:2011.11619*, 2020.
- [4] Wojciech Czaja and Weilin Li, “Analysis of time-frequency scattering transforms,” *Applied and Computational Harmonic Analysis*, vol. 47, no. 1, pp. 149–171, 2019.
- [5] Wojciech Czaja and Weilin Li, “Rotationally invariant time–frequency scattering transforms,” *Journal of Fourier Analysis and Applications*, vol. 26, no. 1, pp. 1–23, 2020.
- [6] Ilya Kavalero, Weilin Li, Wojciech Czaja, and Rama Chellappa, “3-d fourier scattering transform and classification of hyperspectral images,” *IEEE Transactions on Geoscience and Remote Sensing*, to appear.
- [7] Wojciech Czaja, Ilya Kavalero, and Weilin Li, “Scattering transforms and classification of hyperspectral images,” in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXIV*. International Society for Optics and Photonics, 2018, vol. 10644, p. 106440H.
- [8] John Peterson, Weilin Li, Brian Cesar-Tondreau, John Bird, Kevin Kochersberger, Wojciech Czaja, and Morgan McLean, “Experiments in unmanned aerial vehicle/unmanned ground vehicle radiation search,” *Journal of Field Robotics*, vol. 36, no. 4, pp. 818–845, 2019.
- [9] Erchan Aptoula, Murat Can Ozdemir, and Berrin Yanikoglu, “Deep learning with attribute profiles for hyperspectral image classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1970–1974, 2016.
- [10] “Remote sensed hyperspectral image classification with the extended morphological profiles and support vector machines,” <https://github.com/andreybicalho/ExtendedMorphologicalProfiles>, Accessed: 2020-02-15.
- [11] Vladimir N. Vapnik, *Statistical Learning Theory*, Wiley, 1998.
- [12] Peter Bartlett and John Shawe-Taylor, “Generalization performance of support vector machines and other pattern classifiers,” *Advances in Kernel methods—support vector learning*, pp. 43–54, 1999.