IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 October 2023, accepted 10 October 2023, date of publication 30 October 2023, date of current version 10 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3328388

RESEARCH ARTICLE

Graph-Based Semi-Supervised Learning With Tensor Embeddings for Hyperspectral Data Classification

IOANNIS GEORGOULAS[1], EFTYCHIOS PROTOPAPADAKIS[2], KONSTANTINOS MAKANTASIS[3], DYLAN SEYCHELL[3], (Senior Member, IEEE), ANASTASIOS DOULAMIS[1], AND NIKOLAOS DOULAMIS[1]

[1]School of Rural and Surveying Engineering, National Technical University of Athens, 157 80 Athens, Greece
[2]School of Information Sciences, University of Macedonia, 546 36 Thessaloniki, Greece
[3]Department of Artificial Intelligence, University of Malta, MSD2080 Msida, Malta

Corresponding author: Ioannis Georgoulas (johnniegeo@mail.ntua.gr)

This work was supported by the H2020 Photonics Project "A CostEffective Photonics-Based Device for Early Prediction, Monitoring, and Management of Diabetic Foot Ulcers" funded through the Information and Communication Technologies (ICT) H2020 Framework, under Grant 871908.

ABSTRACT Hyperspectral data classification is one of the fundamental problems in remote sensing. Several algorithms based on supervised machine learning have been proposed to address it. The performance, however, of the proposed algorithms is inherently dependent on the amount and quality of annotated data. Due to recent advances in hyperspectral imaging and autonomous (unmanned) aerial vehicles collecting new hyperspectral data is an easy task. Annotating those data, however, is a tedious, time-consuming and costly task requiring the *in-situ* presence of human experts. One way to loosen the requirement of a large number of annotated data is the shift to *semi-supervised learning* combined with highly sample-efficient *tensor-based* neural networks. This study provides a comprehensive experimental analysis of the performance of a variety of graph-based semi-supervised learning techniques combined with tensor-based neural network embeddings for the problem of hyperspectral data classification. Experimental results suggest that the combination of tensor-based neural network embeddings with graph-based semi-supervised learning can significantly improve the classification results minimizing human annotation effort.

INDEX TERMS Graph-based, hyperspectral data, semi-supervised learning, tensor-based embedding.

NOMENCLATURE

n	Number of nodes contained in the graph.
m	Number of edges contained in the graph.
c	Number of available classes.
$\boldsymbol{W}$	Learnable weight matrix.
R	The rank for the CP decomposition.
$\odot$	Khatri-Rao product operator.
$\boldsymbol{X}_{x,y}$	Input tensor of size $p_1 \times p_2 \times p_3$ centred at (x, y) pixel position.

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy.

I. INTRODUCTION

One of the fundamental problems in Remote Sensing is the task of hyperspectral data classification. This problem is treated as a supervised machine learning task, where the objective is to build a machine learning model that classifies hyperspectral image pixels according to the material they depict, by effectively fusing their spectral and spatial information. The performance, however of the learnt models heavily depends on the amount and quality of annotated data.

Given the recent advances in hyperspectral imaging and autonomous aerial vehicles, one can argue that collecting hyperspectral data is a relatively easy task. A similar argument, however, for annotating the data would be unsubstantiated. Annotating hyperspectral data is a tedious,

time-consuming and costly task, usually requiring human experts to physically visit the area depicted in a hyperspectral image and manually label each of the image's pixels.

To loosen the requirement of a large number of annotated data, two different lines of research have been proposed. The first one focuses on the development of fully-supervised machine learning models that can efficiently handle small sample setting problems [23] (i.e., problems where the number of labelled data is limited). The second one, instead of following the (fully) supervised learning paradigm, focuses on developing algorithms that exploit information from both labelled and unlabelled data [28]. These techniques are under the umbrella of the *semi-supervised* learning paradigm.

The main objective of this paper is to investigate the degree to which the above two lines of research can be merged. Specifically, we investigate the development and efficiency of hyperspectral classification algorithms based on *graph-based* semi-supervised learning and highly sample-efficient *tensor-based* neural networks.

Graph-based semi-supervised learning methods have gained significant attention due to their effectiveness in capturing the underlying structure of high-dimensional data [30]. By representing the data as a graph, where each point corresponds to a node, and the edges represent the pairwise similarity between them, these methods can propagate the label information from labelled to unlabeled data in a smooth and coherent way. The performance, however, of graph-based semi-supervised learning methods heavily relies on the quality of the data points representation used to construct the graph. Data points representation, also called *embedding*, must capture in a compact way (low dimensional) the essential features of the data.

Tensor-based embedding techniques, such as tensor decomposition, can be used to extract low-dimensional representations of the data while preserving the underlying structure [20], [21]. Tensor-based embeddings can capture complex patterns and relationships in the data and can be particularly useful in applications employing small-sample setting problems.

Based on the discussion above, this paper focuses on exploring the benefits of a synergistic approach between tensor-based embeddings and graph-based semi-supervised learning. The idea lies in utilizing the output of the penultimate layer of a tensor-based neural network classifier as embeddings for graph-based semi-supervised learning. Tensor-based embeddings are expected to facilitate graph creation before employing a graph-based classifier.

The main hypothesis is that the tensor-based embeddings are able to capture the essential features of hyperspectral data, resulting in a high-quality graph and increasing, this way, the semi-supervised classifiers' performance. We validate this hypothesis using data from 4 publicly available datasets. We thoroughly evaluate the performance of 11 graph-based semi-supervised learning techniques by running more than 22 thousand experiments under different parameterizations of the employed components.

The rest of the paper is organised as follows. Section II presents the literature related to the problem of hyperspectral data classification, as well as embedding learning and semi-supervised learning for hyperspectral data. Section III presents the problem formulation for building semi-supervised learning classifiers using tensor-based embeddings. In Section IV we present the experimental setup and discuss the experimental results. Finally, Section V concludes our study.

II. RELATED WORK

The field of hyperspectral image analysis has witnessed significant advancements in recent years, driven by the exploration of various techniques for enhancing classification accuracy [18]. Yet, research on feature learning given an insufficient number of samples remains an active area [32], with multiple emerging schemes and techniques.

Among these techniques, semi-supervised learning has emerged as a powerful paradigm that leverages the abundance of unlabeled data to augment limited labelled samples. If the unlabelled data points provide additional information that is relevant for prediction, they can potentially be used to achieve improved classification performance [31].

Additionally, the use of tensor-based embeddings has shown promise in transforming hyperspectral data, enabling efficient representation learning for downstream analysis [20]. Such approaches can extract and model relationships between features that may not be apparent using traditional methods, leading to more accurate and meaningful insights.

In this section, we review the existing literature on various techniques applied to hyperspectral image analysis, focusing on two prominent areas related to this study: embedding learning and semi-supervised learning. The utilization of deep learning approaches for the creation of appropriate embeddings, prior to the application of multiple semi-supervised learning has been investigated before [28] provided several promising outcomes.

A. EMBEDDING LEARNING IN HYPERSPECTRAL DATA CLASSIFICATION

The accurate representation of available data plays a vital role in numerous applications, particularly in the field of hyperspectral image analysis research. To capture and describe the intricate relationships among different instances, a fundamental step is the transformation of the data into meaningful feature vectors. This process, known as embedding learning, aims to construct a vector space representation that effectively encapsulates the inherent characteristics of the data. By leveraging advanced techniques, embedding learning facilitates the extraction of informative features, enabling the development of more robust hyperspectral image analysis systems.

The work of Makantasis et al. [20] introduced a Rank-R Feedforward Neural Network (Rank-R FNN) for hyperspectral data classification tasks. The weights of Rank-R FNN are constrained to satisfy a rank-R Canonical Polyadic

Decomposition, which promotes learning capabilities by avoiding overfitting and underfitting problems related to limited data availability. Such models have significantly fewer trainable parameters and work well under noisy data inputs. Yet, the impact of the decomposition rank on the classifier's performance remains uncertain. The extent to which pursuing higher ranks may be beneficial or unnecessary for a given dataset warrants further investigation.

The work of Deng et al. [3] presented a network-based architecture designed to embed hyperspectral image pixels onto a specific metric space, where the similarity between any two pixels is clearly represented. The adopted approach can be applied in both same- and cross-scene classifications. For the latter case, an unsupervised (adversarial) domain adaptation technique is utilized, which preserves the distribution among instances. Multiple experiments run over various, but it's unclear if the relationships between labelled sample sizes and patch sizes were subject to the mutual influence, and the extent of their interdependence remains vague, potentially varying across different scenarios.

A graph construction based on locally linear embeddings for the classification of hyperspectral data has been proposed in [35] so that the spatial pixels' information is considered. To mitigate any drawback due to noise, weighted mean filtering is applied as a prepossessing step. The outcomes suggest the importance of preserving spatial information and emphasize the importance of using appropriate embeddings to fully exploit the potential of graph-based approaches. There are, however, limited details on how the filter's parameters may affect the adopted locally linear embedding schemes and if there are any rules over the pixels' values (e.g. distribution) that promote the utilization of specific filters.

The discriminant information of both source and target domains, i.e. different datasets or data distributions that are associated with different environments or conditions, is of major importance. The source domain typically represents the labeled dataset. The work of [16] proposed a domain adaptation method for HSI classification based on graph embedding and distribution alignment. The methodology uses graph embedding algorithms and pseudo-label learning methods to learn interclass and intraclass divergence matrices of source and target domains, which preserves the local discriminant information of both domains. However, large spectral drifts in such cross-scene classification problems can jeopardize the performance, even if spatial filtering is applied.

The utilization of graph convolutional networks (GCNs) has also been investigated in HSI classification GCNs allow for adaptive learning over the kernel parameters, based on the distribution of the land cover. The work of Ding et al. [8] introduced an aggregator fusion mechanism to present the different spatial node relationships and extract the spatial features of land covers. Yet, GCNs are computationally expensive and face the problem of oversmoothing [6], [7]. The work of Zhang et al. [34] focused on exploring a wider receptive field in a way that the local convolutional features of the nodes are preserved.

The prior knowledge deficiency, large spectral variability, and high dimension of HSI guided researchers to unsupervised [4], or self-supervised [5] approaches. The unsupervised methods have encoundered difficulties in feature extraction and expression for HSI clustering. In the self-supervised scenario, authors note that the unsupervised methods need to be improved in terms of clustering accuracy, especially for large-scale HSIs.

In our study, however, we focus on the problem of limited labelled samples in HSI classification. For this reason, we opt for using a simple yet sample-efficient model to learn hyperspectral data embeddings.

B. SEMI-SUPERVISED LEARNING IN HYPERSPECTRAL DATA CLASSIFICATION

Semi-supervised learning approaches can serve either as a classification model or provide additional pseudo-labels, which, then, will be used to improve the performance of the main classifier further. The work of [33] utilized a cluster-based approach, using Dirichlet Process Mixture Models, to produce pseudo-labels to train a convolutional neural network on all available data. Then, a transfer-learning approach was applied to establish a second model, which was fine-tuned over existing labelled data.

A process based on sample re-sampling and semi-supervised strategies to handle the problem of unbalanced datasets has been proposed in the work of [9]. The proposed methodology is a ''self-learning'' variation that incorporates pseudo-labelled data for retraining if consensus among different classifiers, i.e. high confidence score, is achieved. The adopted approach can address the problem of unbalanced classification but fails to distinguish among objects of similar spectral-spatial characteristics.

He et al. [15] proposed a graph-based semi-supervised algorithm combined with particle cooperation and competition, which can improve the model performance effectively by using unlabelled samples. The key idea lay in adopting a particle competition and cooperation mechanism into label propagation, which could detect and re-label misclassified samples dynamically, thus stopping the propagation of wrong labels and allowing the overall model to obtain better classification performance by using predicted labelled samples in hyperspectal datasets. There are no suggestions regarding possible bounds (minimum or maximum) regarding the labelled or unlabelled instances to be used.

Miao et al. [26] proposed a random multigraph and ensemble strategy to tackle hyperspectral imagery classification problems. The approach employs the construction of multiple anchor graphs for label propagation using cross-entropy regularization. The entire setup tries to address all problems that come from the utilization of a single classifier, i.e. creation of one adjacent graph, which fails to learn the complex structures and intrinsic properties of hyperspectral data effectively.

The scalability limitations of graph-based semi-supervised learning approaches, when handling large corpora were

considered in the work of [14]. The adopted approach considers a parameter-free, naturally sparse and scale-invariant graph consisting of anchor points. The gain over computational complexity justifies the utilization of such models, despite any minor performance loss. Their work, however, does not take the spatial information of hyperspectral data into account.

Different approaches for semi-supervised learning in hyperspectral data classification have been utilized. Graph-based models have the ability to model complex structures and intrinsic properties in hyperspectral data classification problems. Yet, there is no consensus on what approach should be utilized nor how many unlabeled data samples should be used, i.e. trade-off between classification performance gain and computational time required.

C. RESEARCH CHALLENGES

On the one hand, despite the advantages in temsor-based learning, there are limitations related to the appropriate decomposition rank, preprocessing steps, amount of labeled instances, and the handling of noisy data inputs, depenting on the adopted approach as described in sec. II-A.

On the other hand, in SSL, determining which technique to be used or an adequate upper bound for unlabeled data instance remains a challenge. The trade-off between incorporating more unlabeled samples and the computational complexity involved is a critical consideration, and general guidelines for effectively harnessing the potential of unlabeled data are yet to be established, as shown in sec. II-B.

The research landscape reveals diverse alternatives for data representations, varying viewpoints on the selection of appropriate semi-supervised learning techniques, considerations of optimal utilization bounds for unlabeled data, and trade-offs among different types of embeddings. While the integration of data representation and semi-supervised learning holds immense potential, the absence of a unified consensus underscores the complexity of the problem and open research questions in this domain. Resolving these challenges requires further investigation and analysis to identify the most effective and robust strategies for combining embedding learning and semi-supervised learning techniques in hyperspectral image analysis applications.

D. OUR CONTRIBUTION

The innovation of this paper lies in the proposed integration scheme of graph-based SSL with Rank-R FNN for embedding learning, aiming to address the unique challenges related to limited labelled data instances in HSI data.

Specifically, this study emphasize on conducting an extensive exploration between two disparate, yet resource-efficient techniques. This collaborative approach will motivate for additional research directions, rather than pursuing a direct confrontation with established state-of-the-art approaches.

The contributions of this work can be summarized as follows:

- The work addresses one key-problem in HSI semantic segmentation task, related to the limited availability of labelled samples.
- We synergistically combine two inherently sample-efficient techniques: a tensor-based embedding learner and graph-based SSL methods.
- We undertake an extensive empirical investigation, conducting over 22,000 experiments to evaluate the effectiveness of proposed scheme.
- The results provides rigorous and statistically sound conclusions regarding the performance of our methodology.

III. PROBLEM FORMULATION

Hyperspectral data classification is a multi-class classification task. Let us denote as $D_L = ((x_i, y_i))_{i=1}^{l}$ a collection of l labeled datapoints. The object $x_i \in \mathcal{X}$ stands for the explanatory variables (or predictors) while $y_i \in \mathcal{Y}$ is the target variable (i.e., class index) associated with x_i. On top of D_L, we assume that we have in our disposal another collection of u unlabelled data denoted as $D_U = (x_i)_{i=l+1}^{n}$, where $n = l + u$. Based on the above, the problem of hyperspectral data classification boils down to building a model able to exploit information from both D_L and D_U to make accurate predictions.

Exploiting information from both D_L and D_U can be achieved via graph-based semi-supervised learning techniques. Let us denote as $G = (V, E)$ a graph with $|V| = n$ nodes and edges $E \subseteq V \times V$. Each edge connecting the i-th and the j-th nodes is associated with a real-valued weight U_{ij}. The objective of graph-based SSL approaches, which are by nature transductive, is to learn a predictive function $f : G, \mathbf{Y}_L \rightarrow \mathbf{Y}_U$, to infer the missing labels $\mathbf{Y}_U$ for the unlabeled nodes in D_U based on the constructed graph and the labels $\mathbf{Y}_L$ of D_L [10]. Consequently, a graph-based semi-supervised learning approach starts by creating the graph and then propagating labels' information from labelled to unlabelled nodes.

The graph creation occurs in an unsupervised way, e.g. kernel-based or locally-linear reconstruction method, among many other approaches. The former case involves pairwise similarity measures, by using Gaussian or Laplacian kernels. The later case, focuses on local relationships by explicitly reconstructing each data point as a linear combination of its neighbors. Regardless of the adopted approach, the graph is constructed based on the provided input data.

In this study, instead of utilizing the raw data values, $\{x_i\}_{i=1}^{n}$, we exploit tensor-based embeddings $\{\Phi(x_i)\}_{i=1}^{n}$ generated by the penultimate layer of a tensor-based neural network classifier (see Section III-A). The label propagation has, also, many approaches; the adopted ones are explained in subsection III-B.

A. TENSOR-BASED EMBEDDINGS

In the context of this study, tensor embeddings correspond to one-dimensional vector representations of tensor data points

(three-dimensional objects) produced by the penultimate layer of a Rank-R FNN model. The Rank-R FNN model, proposed and experimentally and theoretically validated in [20], [22], [23], and [25], is a two-layer neural network that receives as input D-dimensional tensor objects. The inputs are propagated to the hidden layer via a set of weights on which the Canonical-Polyadic (CP) decomposition has been applied. CP decomposition can significantly reduce the number of trainable parameters, making the Rank-R FNN model very efficient for small sample setting problems that involve high dimensional data, like data points in tensor format.

Given an input $\boldsymbol{X} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$, the Rank-R FNN models the weights connecting the input layer with the hidden layer as:

$$\boldsymbol{W}^{(q)} = [\![\boldsymbol{W}_1^{(q)}, \cdots, \boldsymbol{W}_D^{(q)}]\!] \in \mathbb{R}^{I_1 \times \cdots \times I_D}, \tag{1}$$

or else

$$\text{vec}(\boldsymbol{W}^{(q)}) = (\boldsymbol{W}_D^{(q)} \odot \cdots \odot \boldsymbol{W}_1^{(q)})\mathbf{1}_R \in \mathbb{R}^{\prod_{d=1}^{d=D} I_d}, \tag{2}$$

where $\mathbf{1}_R$ stands for a vector with R ones, q stands for the number of hidden layer's neurons, and "$\odot$" operator for the Khatri-Rao product. The output of the q-th hidden neuron of the Rank-R FNN is given by

$$u_q = g(\langle \boldsymbol{W}^{(q)}, \boldsymbol{X} \rangle) = \text{trace}\Big(g\big((\boldsymbol{W}_d^{(q)})^T \boldsymbol{Z}_{\neq d}^{(q)}\big)\Big), \tag{3}$$

where $\boldsymbol{W}_d^{(q)} \in \mathbb{R}^{I_d \times R}$, R is the rank for the CP decomposition and

$$\boldsymbol{Z}_{\neq d}^{(q)} = \boldsymbol{X}_{(d)}(\boldsymbol{W}_D^{(q)} \odot \cdots \odot \boldsymbol{W}_{d+1}^{(q)} \odot \boldsymbol{W}_{d-1}^{(q)} \odot \cdots \odot \boldsymbol{W}_1^{(q)}) \tag{4}$$

with $\boldsymbol{X}_{(d)}$ to denote the mode-d matricization of tensor $\boldsymbol{X}$.

After the transformation of the tensor input into a vector at the first hidden layer, the information is propagated to the output layer of the Rank-R FNN. The parameters of the model are estimated based on the available labelled data using an appropriate loss function (e.g., cross-entropy for classification problems of mean squared error for regression), the back-propagation algorithm and gradient-based optimization (for more information about the training procedure, we refer the interested reader to [23]).

In this study, the tensor input embeddings correspond to the output of the hidden layer of the trained Rank-R FNN. In other words, the tensor inputs are represented by q-dimensional vectors that are obtained by the output of the penultimate layer of the Rank-R FNN.

B. GRAPH-BASED APPROACHES

Graph-based approaches are transductive by nature. As such, the optimization process, i.e. obtaining a solution, is solely concerned with obtaining predictions for the available unlabeled data points [31]. This trait was the main reason for selecting them, in the current study, against other approaches of inductive nature, i.e. building a classification model for the entire input space.

Multiple different graph-based semi-supervised learning approaches are utilized to investigate the impact of tensor-based embeddings on classification performance. Specifically, this research involves techniques based on the Laplace equation solution, introduced by [36], and variations [2], [29]. The game-theoretic p-Laplace equation was also considered [11]. Additional techniques involve the solution of the Poisson equation, using either a conjugate gradient or a spectral solver [2]. Other techniques utilized are the "lazy random walks" [37], the sparse label propagation [19], and Merriman-Bence-Osher scheme-based approaches [12], [17]. Table 2 compactly presents the semi-supervised learning techniques explored in this paper.

When utilizing a transductive learner, information is propagated among the data points through available connections (edges) on a graph. A common approach, when using graph-based approaches involves three steps: a) construction of a graph, b) edge weighting, and c) inference. The first two steps create an appropriate graph, which will be used for the label estimation during the inference process. The following paragraphs briefly summarize the basic concepts for the adopted label propagation techniques.

Label propagation over graphs is a general framework that encompasses various methods. These methods are able to infer the labels for the unlabeled instances by minimizing a loss function of the form [30]:

$$L(f) = L_s(f, D_L) + \lambda L_u(f, D_u) + \mu L_r(f, D) \tag{5}$$

where f represents the function to be learned (the labelling of data points), $L_s(f, D_L)$ and $L_u(f, D_u)$ are the supervised and unsupervised losses, based on labeled and unlabeled data respectively, $L_r(f, D)$ is the regularization loss, over all data, and λ, μ are hyperparameters to be optimized.

IV. EXPERIMENTAL SETUP

The proposed schemes are evaluated using four publicly available hyperspectral datasets; each dataset is a hyperspectral image. The objective of our methodology is to semantically segment the images into regions depicting different materials following a pixel-wise classification task. As such, for the performance evaluation, metrics like f1-score were considered. As shown in sec. IV-A class imbalance exist in the case studies, which motivate us to consider the F1 score for the statistical tests performed in sec. IV-G.

At this point we need to stress that the study focus on addressing the problem of limited labelled samples in HSI, when applying a newly introduced synergistic scheme. As such, the setup was arranged in a way to allow for providing robust conclusions about the impact of combining two, different in nature, sample-efficient techniques and applying them to HSI segmentation tasks. The optimal setup of the hyperparameters will be investigated in future work.

The following sections present the datasets employed in this study, and the approach followed to form training and evaluation sets. Then, we present the models'

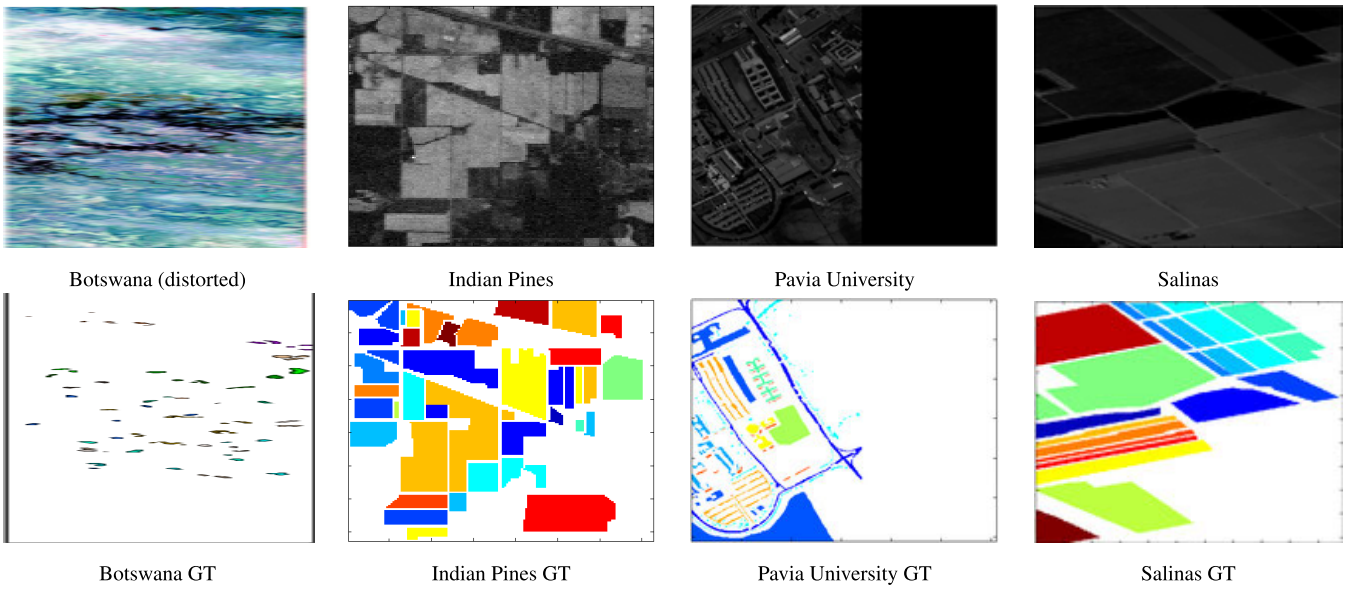


FIGURE 1. Utilized datasets.

parameterization and the evaluation protocol, and, finally, the obtained results.

A. DATASET DESCRIPTION

Four widely known and publicly available datasets, captured by different sensors, are used. Table 1, provides further details on the distribution of instances per class. A brief description is provided bellow.

- the Botswana dataset, which has been captured by Hyperion sensor and consists of 145 spectral bands and 3,248 labelled pixels assigned to 14 different classes.
- the Indian Pines dataset, which s captured by the AVIRIS sensor and contains a total of 145 spectral bands and 10,086 labeled pixels, assigned to 16 classes
- the Pavia University dataset, which has been captured by ROSIS sensor and consists of 103 spectral bands and 41,849 labelled pixels assigned to 9 different classes
- the Salinas (corrected) dataset, which has been captured by AVIRIS sensor and consists of 224 spectral bands, minus 20 water absorption bands, and 53370 labelled pixels assigned to 16 different classes

Figure 1 the employed datasets, along with their ground truth.

Each of the employed datasets corresponds to one hyperspectral image. Hyperspectral images are represented as third-order tensors of dimension $p_1 \times p_2 \times p_3$. Dimensions p_1 and p_2 correspond to the image's height and width while dimension p_3 corresponds to the number of the image's spectral bands.

To conduct pixel-wise classification, i.e., to classify each hyperspectral image pixel $I_{x,y}$ at location (x, y) according to the material it depicts, we follow the approach proposed in [20] and [23]. By following the approach in [20] and [23], we are able to form training and evaluation sets, enabling our models to exploit both the spectral and spatial pixels information.

Following [20] and [23], we start by assuming that the label of a square patch $\boldsymbol{X}_{x,y}$ of size $s \times s \times p_3$ centred at (x, y) has the same label with pixel $I_{x,y}$. Denoting as $\boldsymbol{t}_{x,y}$ the ground truth label of $I_{x,y}$, that is the class of the depicted material, we form the dataset $\mathcal{D} = \{(\boldsymbol{X}_{x,y}, \boldsymbol{t}_{x,y})\}$ for training and evaluation purposes. Each $\boldsymbol{X}_{x,y}$ patch inherently carries spectral and spatial information. Regarding parameter s, we conducted two sets of experiments using the values of five and seven. By using a relatively small value for the parameter s we can assure that the square patch $\boldsymbol{X}_{x,y}$ has the same label as $I_{x,y}$ for the majority of the pixels [24], satisfying this way our initial assumption. Finally, each $\boldsymbol{X}_{x,y} \in \mathcal{D}$ is min-max normalised to [0, 1].

B. ABLATION EXPERIMENTS

In this study we use a tensor-based deep learning approach as a core model. Then, based on the embedding provided by this approach we evaluated the appropriateness of various graph-based SSL approaches. The former case, i.e. embedding impact, involves the investigation of the impact of the decomposition rank R and the number q of hidden neurons on the performance of the Rank-R FNN model. Results are presented in detail in sec. IV-D. The latter case, i.e. SSL approaches we focused on the impact of the unlabeled samples size on ten different approaches, when using as inputs the Rank-R FNN model embeddings. Results are presented in detail in sec. III-B. Over 22,000 experiments run in total, to test different problem instances.

The partitioning of the dataset, for all experiments, was based on the stratified k-fold paradigm. The first step involves

TABLE 1. Available labeled instances per class for the different employed datasets.

Dataset	Botswana		Indian Pines		Pavia University		Salinas	
	Samples	Perc	Samples	Perc	Samples	Perc	Samples	Perc
C1	270	8%	46	0%	6522	16%	1945	4%
C2	101	3%	1428	14%	17907	43%	3726	7%
C3	251	8%	777	8%	2048	5%	1976	4%
C4	215	7%	237	2%	3039	7%	1394	3%
C5	269	8%	468	5%	1345	3%	2678	5%
C6	269	8%	730	7%	5029	12%	3929	7%
C7	259	8%	28	0%	1330	3%	3558	7%
C8	203	6%	478	5%	3682	9%	11157	21%
C9	314	10%	20	0%	947	2%	6190	12%
C10	248	8%	967	10%			3221	6%
C11	305	9%	2413	24%			1047	2%
C12	181	6%	593	6%			1890	4%
C13	268	8%	205	2%			901	2%
C14	95	3%	1265	13%			1051	2%
C15			338	3%			7061	13%
C16			93	1%			1646	3%
Total	3248	100%	10086	100%	41849	100%	53370	100%

loading the data instances per demo area. We note that we consider only pixels for which labels are available. Table 1 provides details on the available instances per class and the corresponding appearance frequencies.

Given the data over one demo area, a 6-fold approach is adopted, creating the corresponding training/test subsets, which preserve the percentage of samples for each class as in the initial set. These sets are adjusted, as described bellow, to follow the few-labeled samples paradigm, adopted in this work.

Then, for each fold, we randomly select, from the training set, a prespecified number α of samples per class. In this setup three different values for α were used: 10, 20 and 30 samples per class. Additionally, we randomly select additional 10 samples per class, from the remaining train instances, to form the validation set, which will be used by the Rank-R FNN for early-stopping purposes.

At this point we have a subset of $c\times(\alpha+10)$ samples, where c denotes the number of classes, from the initially provided training set. These are considered as labeled instances. The remaining training samples, i.e. those not selected, and the test data are considered as unlabelled instances. Given a set of labeled instances we used various portions from the unlabelled instances, ranging from 20% to 60% to evaluate the impact of the unlabeled instances availability[1].

C. MODEL PARAMETERIZATION AND EXPERIMENTATION PROTOCOL

In our approach, the performance can be affected by various hyperparameters involving two domains. Firstly, the tensor-based embedding process relies on carefully selecting the rank parameter (R) and hidden neuron count (q), striking a balance between classification power and resource efficiency. Secondly, the graph-based semi-supervised learning (SSL) approach is influenced by multiple hyperparameters, depending on the type of the adopted approach.

The performance of the employed Rank-R FNN tensor-based neural network is mainly affected by two hyperparameters: the rank R for the CP decomposition and the number q of neurons placed at the hidden layer. In this study, based on the findings of [22], we use two different values for the parameter R, that is 5 and 3, and two different values for the number of hidden neurons q, that is 50 and 75. We opt for these settings since these values balance the trade-off between high classification performance and computational requirements.

As far as the semi-supervised learning techniques are concerned, they were implemented using the same library as in [1], and unless stated differently, their hyperparameters take the default values (see Table 2). Please note that hyperparameter tuning, for the SSL approach, could further improve the performance. However, identifying the best possible sets of parameters was not in the scope of this study.

To test the efficiency of the employed semi-supervised learning techniques combined with tensor embeddings, we train the models using a small number of labelled examples. This way we are able to test our initial hypothesis by focusing on the models' capacity to learn small sample setting classification tasks. On top of that, restricting the number of labelled samples used for training reflects limitations present in real-world remote sensing applications. As mentioned in Section I, annotating large amounts of remote sensing data is a tedious, time-consuming and costly task. Therefore, many real-world remote sensing applications must deal with very small volumes of annotated data.

D. EMBEDDING IMPACT

For the sake of completeness, this section presents the behaviour of the Rank-R FNN using different hyperparameter settings and different training set sizes.

Figure 2 illustrates the impact of the decomposition rank R and the number q of hidden neurons on the performance of

[1]Experiments are conducted using Python and Google Colab. The code is provided in the following link: https://rb.gy/44c97.

TABLE 2. Evaluated graph-based SSL techniques.

Short Description	Short name	Proposed by	Hyperparameters setup value	Complexity
Laplace learning original method	Lap	[38]	No class priors provided, no re-weighting schemes, Zeroth order term in Laplace equation: 0, Tolerance for conjugate gradient solver: 1e-5, Parameters for properly reweighting α, z: 2, 1e7	$O(n)$
Laplace learning with Poisson reweighting schemes	LapRePo	[2]	No class priors provided, with Poisson re-weighting schemes, remaining parameters same as in Lap	$O(n^2)$
Weighted Nonlocal Laplacian on Interpolation from Sparse Data	LapWNL	[29]	No class priors provided, with Sparse Data interpolated re-weighting schemes, remaining parameters same as in Lap	$O(n^2)$
Random walk implementation	LaWa	[37]	No class priors provided, Parameter in model α: 0.95	$O(n)$
The Multiclass MBO method	MMBO	[12]	No class priors provided, # of inner/outer iterations: 6/10, time step: 0.15, fidelity penalty: 50, # of eigenvectors: 300	$O(nm)$
SSL via the solution of the game-theoretic p-Laplace equation	p-Lap	[11]	No class priors provided, Value of p in the p-Laplace equation: 10, Tolerance with which to solve the equation: 1e-3	$O(n^3)$
SSL via the solution of the Poisson equation, using a conjugate gradient solver	PoCG	[2]	No class priors provided, solver: conjugate gradient, Min/max number of iterations before stopping condition: 50/1000	$O(m)$
SSL via the solution of the Poisson equation, using a spectral solver	PoSpe	[2]	No class priors provided, solver: Spectral, Number of eigenvectors to use for spectral solver: 10	$O(m)$
The sparse label propagation method	SpLa	[19]	No class priors provided, Number of iterations: 100	$O(n\dot{m}\dot{m})$
The Volume MBO method	VMBO	[17]	Class priors were provided, Temperature value: 0.1, volume constraint: 0.5	$O(n\dot{m}\dot{m}log(m))$

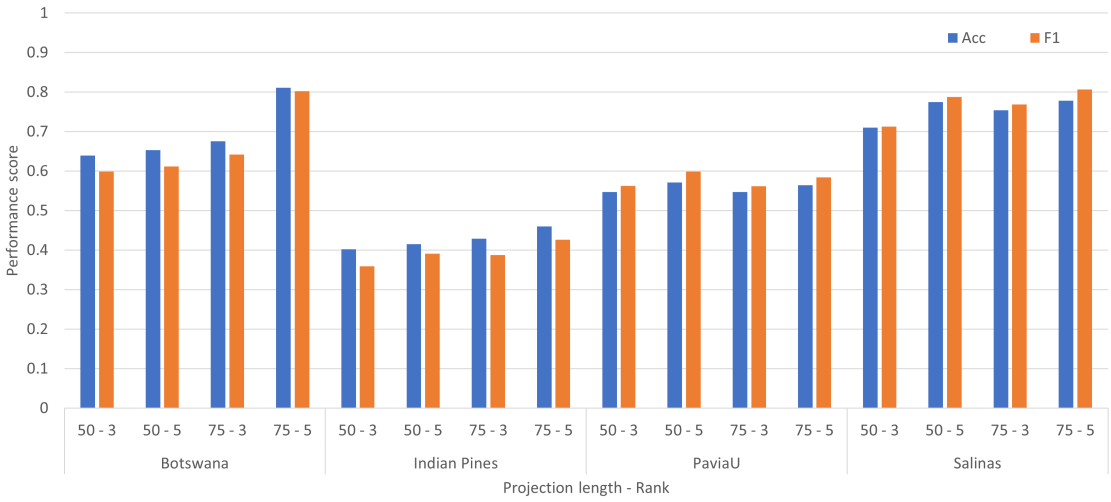


FIGURE 2. Effects of decomposition's rank and the projection size, for the different datasets.

the Rank-R FNN model across the four datasets. At first, as the rank of the decomposition increases while keeping the embedding size fixed, the performance of the model improves. However, the opposite trend is not consistently observed. In other words, higher embedding dimensions do not necessarily result in better performance. As shown by the results on the PaviaU dataset, using 50 hidden neurons is preferred compared to using 75. For the Indian Pines and the Salinas datasets, there is no significant improvement. Instead, we can observe only marginal gains. It is worth noting that the Botswana dataset exhibits an exception to this observation, where higher embedding dimensions do lead to improved performance.

As mentioned above, the Rank-R FNN model is trained using a limited number of samples per class, specifically ranging from 10 to 30 instances. This choice is made deliberately to investigate how labelled data scarcity affects the model's performance. As depicted in Figure 3, increasing the number of labelled instances per class in the training set appears to be beneficial up to a certain threshold. Across

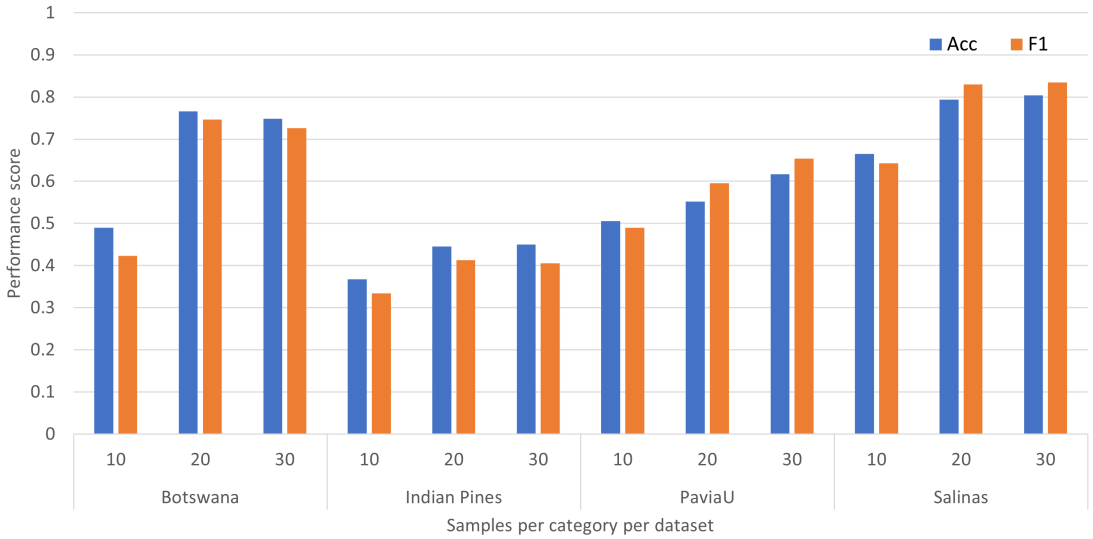


FIGURE 3. Illustrating the impact of the available samples per class, on the training set, for the embedding suitability.

all four datasets, employing 20 or 30 training samples per category yields superior classification results. However, it remains unclear whether there exists a notable distinction between using 20 or 30 samples. It is worth mentioning that other studies considered even higher numbers of labelled samples per class, e.g., Makantasis et al. in [13] and [22]. Nevertheless, no significant performance impact was observed.

E. GRAPH-BASED SSL PERFORMANCE

Based on the findings of the experiments presented in the previous section, we use a Rank-R FNN of decomposition rank $R = 5$ and seventyfive hidden neurons, $q = 75$, to create embeddings to be used for semi-supervised learning. Any semi-supervised learning technique use information both from labelled and unlabelled data. Therefore, a parameter of extremely high importance for investigating the behaviour of different semi-supervised techniques is the number of unlabelled samples used.

To investigate the behaviour of the semi-supervised learning techniques employed we evaluate their classification performance using a variable number of unlabelled data. Specifically, we use 20%, 40%, or 60% of the available unlabelled data to train the semi-supervised algorithms. Similarly to Rank-R FNN evaluation, we follow a holdout cross-validation protocol repeated six times. Finally, we evaluate the performance of the different techniques in terms of F1 score and classification accuracy averaged over the six iterations of the cross-validation scheme.

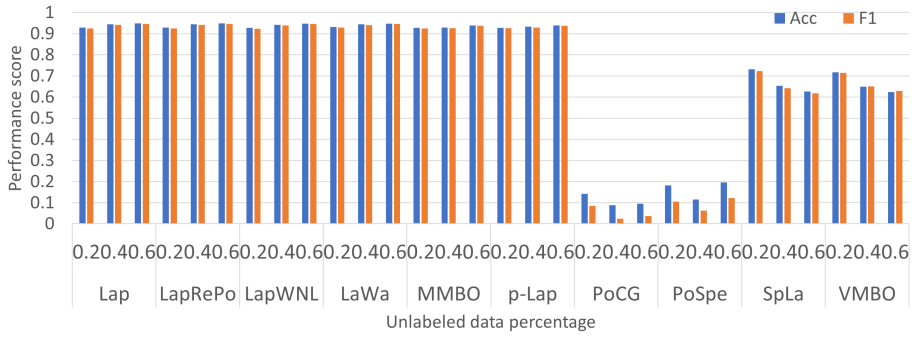
Figure 4 demonstrates how the volume of the unlabelled data used affects the performance of the employed semi-supervised algorithms. We observe that Laplace-based methods, such as Lap, LapRePo, LapWNL, and random walks, namely LaWa, exhibit consistent behavior across all four datasets. As the amount of unlabeled data increases, the performance of the algorithms improves as well. However, the marginal gains in performance, coupled with a substantial increase in processing time, indicates that the performance of the algorithms saturates when an adequate large number of unlabelled data is used. This suggests that there is an upper bound for the number of unlabelled data that can be effectively used to improve the algorithms' classification performance.

Quite surprisingly, the performance for several of the employed semi-supervised algorithms does not improve as the number of unlabelled data used increases. Nevertheless, this behaviour has also been reported in other studies, e.g. [27], and is partially attributed to the unbalanced class distributions. This is perfectly aligned with our application at hand since all the datasets employed in this study can be considered as highly imbalanced.

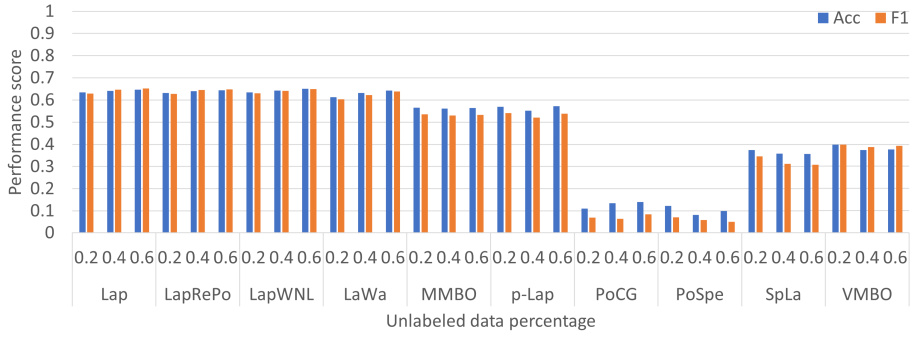
Specifically, MBO variations, Poisson-based implementations (e.g., PoCG, PoSpe), and sparse label approaches exhibit complex patterns when confronted with additional unlabeled data. The case is demonstrated in Figure 5, where marginal gains in performance are scarce, and the overall trend shows a decline.

Figure 6 demonstrates the performance, average scores over all datasets, of the most prominent SSL techniques against the best TNN architecture. Among these methods, Laplace based variations, i.e. Lap, LapRePo, LapWNL, and LaWa exhibited the highest accuracy and F1 scores,, indicating their ability to achieve both precise class identification and satisfactory overall segmentation performance.

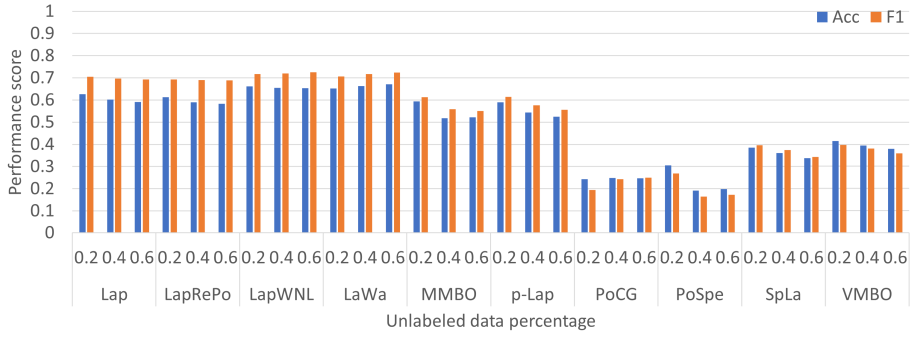
On the other hand, MMBO, p-Lap, SpLa, and VMBO showed relatively lower accuracy and F1 scores, suggesting limitations in their capability to accurately segment hyperspectral data. Nevertheless, the capabilities of the proposed combinatory TB, graph-based SSL schemes, appear as a solid step forward in hyperspectral image segmentation.



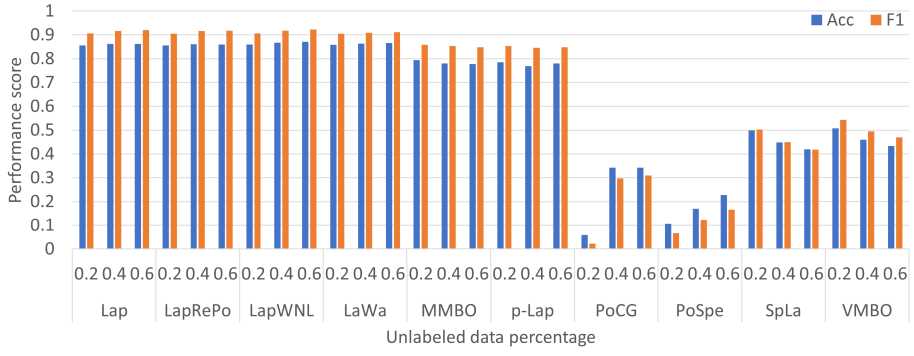
(a) Botswana



(b) Indian Pines



(c) Pavia U



(d) Salinas

FIGURE 4. **Variations in F1 scores when different percentages of unlabeled samples were used, given the adopted classification techniques.**

By leveraging tensor-based embeddings, these approaches effectively extract and utilize spatial and spectral information present in hyperspectral data, leading to improved segmentation results.

F. MODELS COMPLEXITY

Figure 7 presents the average execution times for the employed techniques for different sample sizes. Although graph-based SSL techniques improve classification results,

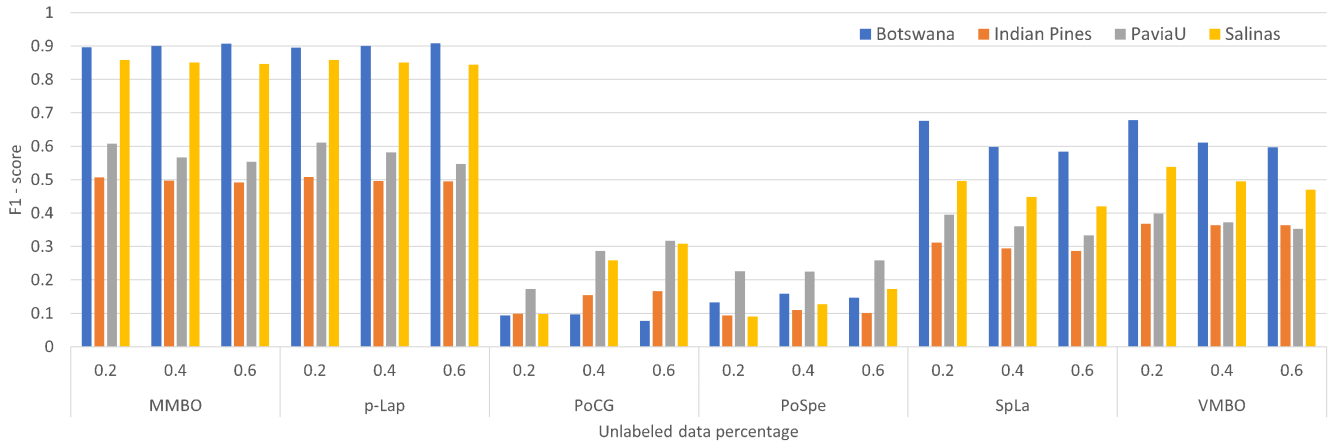


FIGURE 5. Scenarios where more unlabeled data did not guarantee higher performance.

TABLE 3. Updated p-values and t-statistics, in parenthesis, for the F1 scores comparison.

	Lap	LapWNL	LapRePo	MMBO	p-Lap	LaWa	SpLa	VMBO
TNN	0.00 (-17.52)	0.00 (-17.86)	0.00 (-17.10)	0.00 (-10.63)	0.00 (-10.64)	0.00 (-16.62)	0.00 (18.78)	0.00 (16.850)
Lap		0.778 (-0.28)	0.705 (0.38)	0.00 (5.81)	0.00 (5.81)	0.487 (0.69)	0.00 (40.14)	0.00 (39.46)
LapWNL			0.508 (0.66)	0.00 (6.09)	0.00 (6.09)	0.330 (0.98)	0.00 (40.65)	0.00 (40.03)
LapRePo				0.00 (5.44)	0.00 (5.44)	0.750 (0.32)	0.00 (39.53)	0.00 (38.80)
MMBO					0.998 (0.00)	0.00 (-5.10)	0.00 (30.24)	0.00 (28.89)
p-Lap						0.00 (-5.10)	0.00 (30.26)	0.00 (28.91)
LaWa							0.002 (38.71)	0.00 (37.97)
SpLa								-9.1 (-3.14)

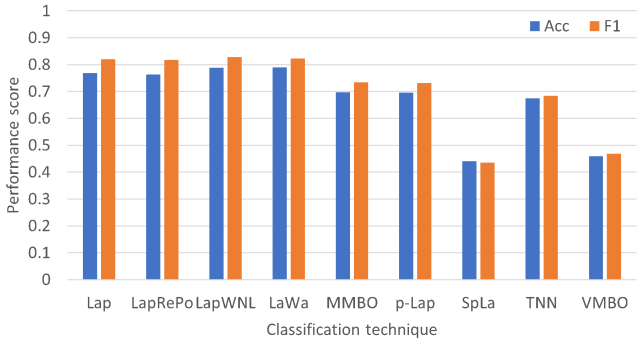


FIGURE 6. Average performance scores, for the best TNN topology.

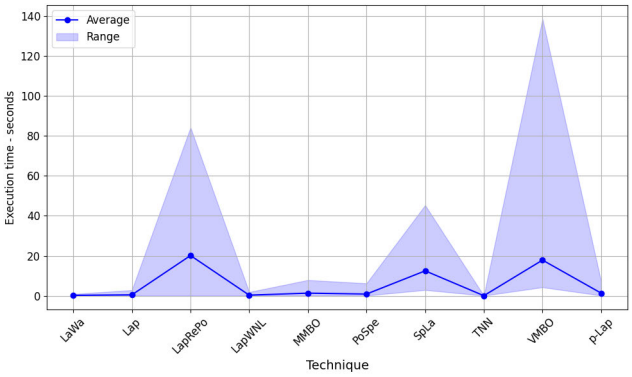


FIGURE 7. Execution times (seconds) for the proposed techniques, for different sample sizes ranging from 20% to 60% of the available data.

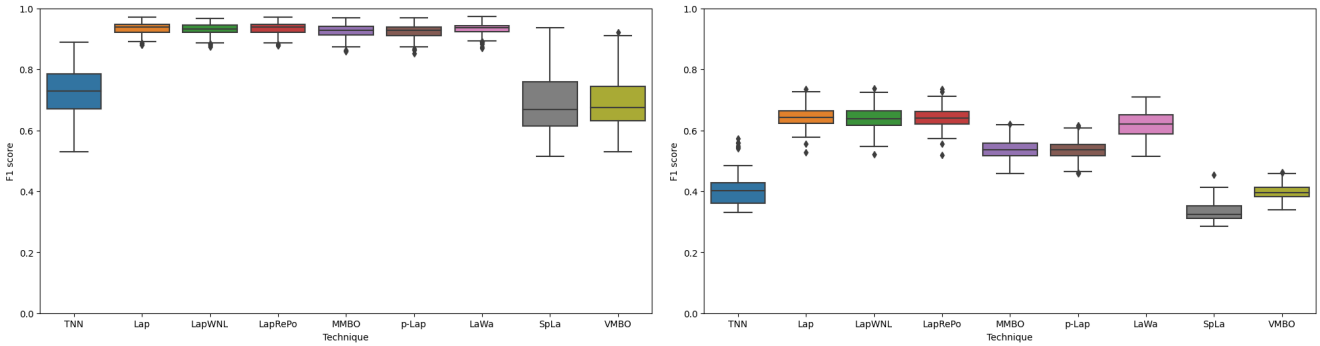
this comes at the cost of longer prediction times. However, given the performance improvement, the extra time required is not prohibitive in applying those techniques to HSI classification.

For the applied SSL approaches, the complexity is presented at the last column of table 2. Please note that most SSL techniques require the construction of a nearest neighbor matrix, prior to any label propagation. Applying an ϵ-neighborhood approach has a complexity of $O(n^2)$, where n stands for the number of available data, labeled and unlabeled. Further details on the complexity of graph-based SSL approaches are provided by Song et al. [30]

As far as the tensor-based neural network is concerned, its computational complexity both for training and inference is the same as the computational complexity of a fully connected neural network with inputs of size $s \times s \times p_3$, as the size of a square patch $\boldsymbol{X}_{x,y}$, and one hidden layer with the same number of hidden neurons as the tensor network. Assuming that we have h hidden neurons, complexity would be $O(h \cdot s^2 \cdot p_3)$.

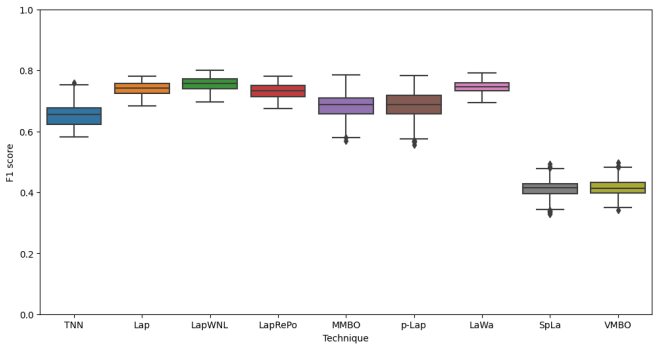
G. STATISTICAL EVALUATION

In this section, our aim is to investigate whether the employment of semi-supervised algorithms results into a statistically significant improvement, in terms of

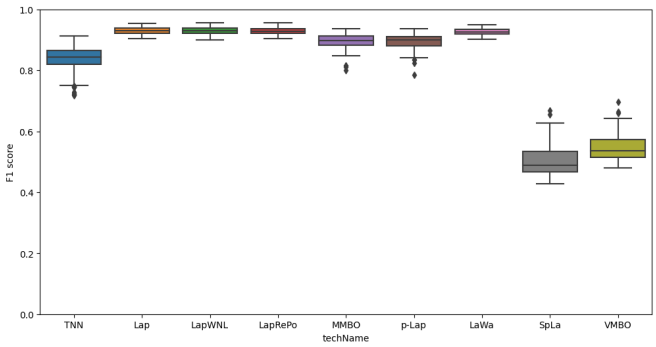


(a) Botwsana

(b) Indian Pines



(c) Pavia University



(d) Salinas

FIGURE 8. **Boxplots for the performance distribution, given the adopted classification techniques.**

classification performance, against the fully-supervised Rank-R FNN.

We start by presenting in Figure 8 the variations in classification performance of employed Rank-R FNN and the graph-based semi-supervised learning approaches used. A different boxplot was produced per dataset while the low performing semi-supervised techniques (see Section IV) were excluded.

These boxplots provide insights into the distribution and variability of the classification performance across the different techniques. Motivated by these figures, a paired t-test has been conducted to assess the significance of the differences between the classification techniques. The results are summarized in Table 3.

Based on the provided t-test results, we can make several observations about the performance comparisons between different algorithms. Firstly, when comparing Rank-R FNN and Lap, a highly significant difference was found with a negative t-statistic of -17.52 ($p < 0.001$). This indicates that Lap outperforms Rank-R FNN significantly in terms of the F1 score. Similar findings were observed when comparing Rank-R FNN with other algorithms such as LapWNL, LapRePo, MMBO, p-Lap, LaWa, and SpLa, with all resulting in highly significant differences and negative t-statistics.

On the other hand, when comparing Lap with MMBO, p-Lap, LaWa, SpLa, and VMBO, significant differences were observed, but this time with positive t-statistics. These results suggest that Lap performs significantly better in terms of the F1 score. Additionally, when comparing Lap with LapWNL, LapRePo, and VMBO, significant differences were observed, but with smaller t-statistics compared to the prreviously mentioned comparisons.

Furthermore, among the comparisons between LapWNL, LapRePo, MMBO, p-Lap, LaWa, SpLa, and VMBO, significant differences were found in most cases, indicating variations in performance levels. Notably, the comparison between LapRePo and LaWa yielded a t-statistic of 0.66 ($p = 0.508$), suggesting no significant difference between the two compared algorithms in terms of F1 score.

Overall, the above-mentioned t-test results provide evidence for performance variations between the different algorithms. It is important to consider these findings when selecting the most suitable classification algorithm for a specific application at hand, as they indicate the relative strengths and weaknesses of each approach.

V. CONCLUSION

Two different lines of research have been merged successful in this work: a) sample-efficient tensor-based neural networks and b) graph-based SSL. This is a new-approach to an established problem of segmentation over hyper spectral images. The experiments conducted in this study demonstrate the potential of such schemes.

This new approach provides multiple potentials on future work. At first, there are multiple other, i.e. non-graph-based, approaches that could be evaluated on similar schemes.

Secondly, applied SSL techniques were not optimized, either computationally or hyperparameter-related. Research work on the field could be provided. Thirdly, the paradigm of few-shot learning can be adopted using as base model tensor-based NN supproted by the outcomes of SSL approaches. On a theoretical part, approximation bounds for different types of tensor decomposition may be an interesting investigation area. A lot of work remains to be done in terms of rank evaluation impact.

By leveraging the spatial and spectral information present in hyperspectral data, tensor-based embeddings can effectively extract useful features for segmentation tasks. Graph-based SSL approaches benefit from high quality features and achieve a high classification performance. Overall, the proposed method has significant potential for various applications in remote sensing, agriculture, and environmental monitoring, among others, and could pave the way for more accurate and efficient hyperspectral image segmentation in the future.

REFERENCES

[1] J. Calder. *Graph Learning Python Package*. Accessed: Jan. 2022. [Online]. Available: https://github.com/jwcalder/GraphLearning

[2] J. Calder, B. Cook, M. Thorpe, and D. Slepcev, "Poisson learning: Graph based semi-supervised learning at very low label rates," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1306–1316.

[3] B. Deng, S. Jia, and D. Shi, "Deep metric learning-based feature embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1422–1435, Feb. 2020.

[4] Y. Ding, Z. Zhang, X. Zhao, W. Cai, N. Yang, H. Hu, X. Huang, Y. Cao, and W. Cai, "Unsupervised self-correlated learning smoothy enhanced locality preserving graph convolution embedding clustering for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5536716.

[5] Y. Ding, Z. Zhang, X. Zhao, Y. Cai, S. Li, B. Deng, and W. Cai, "Self-supervised locality preserving low-pass graph convolutional embedding for large-scale hyperspectral image clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5536016.

[6] Y. Ding, Z. Zhang, X. Zhao, D. Hong, W. Cai, N. Yang, and B. Wang, "Multi-scale receptive fields: Graph attention neural network for hyperspectral image classification," *Expert Syst. Appl.*, vol. 223, Aug. 2023, Art. no. 119858.

[7] Y. Ding, Z. Zhang, X. Zhao, D. Hong, W. Cai, C. Yu, N. Yang, and W. Cai, "Multi-feature fusion: Graph neural network and CNN combining for hyperspectral image classification," *Neurocomputing*, vol. 501, pp. 246–257, Aug. 2022.

[8] Y. Ding, Z. Zhang, X. Zhao, D. Hong, W. Li, W. Cai, and Y. Zhan, "AF2GNN: Graph convolution with adaptive filters and aggregator fusion for hyperspectral image classification," *Inf. Sci.*, vol. 602, pp. 201–219, Jul. 2022.

[9] S. Fang, D. Quan, S. Wang, L. Zhang, and L. Zhou, "A two-branch network with semi-supervised learning for hyperspectral classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 3860–3863.

[10] W. Feng, J. Zhang, Y. Dong, Y. Han, H. Luan, Q. Xu, Q. Yang, E. Kharlamov, and J. Tang, "Graph random neural networks for semi-supervised learning on graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 22092–22103.

[11] M. Flores, J. Calder, and G. Lerman, "Analysis and algorithms for ℓ_p-based semi-supervised learning on graphs," 2019, *arXiv:1901.05031*.

[12] C. Garcia-Cardona, E. Merkurjev, A. L. Bertozzi, A. Flenner, and A. G. Percus, "Multiclass data segmentation using diffuse interface methods on graphs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1600–1613, Aug. 2014.

[13] I. Georgoulas, E. Protopapadakis, K. Makantasis, and A. Doulamis, "Tensor-based embedding for graph-based semi-supervised approaches," in *Proc. 16th Int. Conf. Pervasive Technol. Rel. Assistive Environ. (PETRA)*, 2023, pp. 1–6, doi: 10.1145/3594806.3596550.

[14] F. He, R. Wang, and W. Jia, "Fast semi-supervised learning with anchor graph for large hyperspectral images," *Pattern Recognit. Lett.*, vol. 130, pp. 319–326, Feb. 2020.

[15] Z. He, K. Xia, T. Li, B. Zu, Z. Yin, and J. Zhang, "A constrained graph-based semi-supervised algorithm combined with particle cooperation and competition for hyperspectral image classification," *Remote Sens.*, vol. 13, no. 2, p. 193, Jan. 2021.

[16] Y. Huang, J. Peng, Y. Ning, W. Sun, and Q. Du, "Graph embedding and distribution alignment for domain adaptation in hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7654–7666, 2021.

[17] M. Jacobs, E. Merkurjev, and S. Esedoğlu, "Auction dynamics: A volume constrained MBO scheme," *J. Comput. Phys.*, vol. 354, pp. 288–310, Feb. 2018.

[18] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, "A survey: Deep learning for hyperspectral image classification with few labeled samples," *Neurocomputing*, vol. 448, pp. 179–204, Aug. 2021.

[19] A. Jung, A. O. Hero, III, A. Mara, and S. Jahromi, "Semi-supervised learning via sparse label propagation," 2016, *arXiv:1612.01414*.

[20] K. Makantasis, A. Doulamis, N. Doulamis, A. Nikitakis, and A. Voulodimos, "Tensor-based nonlinear classifier for high-order data analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2221–2225.

[21] K. Makantasis, A. Doulamis, N. Doulamis, and A. Voulodimos, "Common mode patterns for supervised tensor subspace learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2927–2931.

[22] K. Makantasis, A. D. Doulamis, N. D. Doulamis, and A. Nikitakis, "Tensor-based classification models for hyperspectral data analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 6884–6898, Dec. 2018.

[23] K. Makantasis, A. Georgogiannis, A. Voulodimos, I. Georgoulas, A. Doulamis, and N. Doulamis, "Rank-R FNN: A tensor-based learning model for high-order data classification," *IEEE Access*, vol. 9, pp. 58609–58620, 2021.

[24] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 4959–4962.

[25] K. Makantasis, A. Voulodimos, A. Doulamis, N. Doulamis, and I. Georgoulas, "Hyperspectral image classification with tensor-based rank-R learning models," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 3125–3148.

[26] Y. Miao, M. Chen, Y. Yuan, J. Chanussot, and Q. Wang, "Hyperspectral imagery classification via random multigraphs ensemble learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 641–653, 2022.

[27] A. Oliver, A. Odena, C. A. Raffel, E. D. Cubuk, and I. Goodfellow, "Realistic evaluation of deep semi-supervised learning algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–12.

[28] E. Protopapadakis, A. Doulamis, N. Doulamis, and E. Maltezos, "Stacked autoencoders driven by semi-supervised learning for building extraction from near infrared remote sensing imagery," *Remote Sens.*, vol. 13, no. 3, p. 371, Jan. 2021.

[29] Z. Shi, S. Osher, and W. Zhu, "Weighted nonlocal Laplacian on interpolation from sparse data," *J. Sci. Comput.*, vol. 73, nos. 2–3, pp. 1164–1177, Dec. 2017.

[30] Z. Song, X. Yang, Z. Xu, and I. King, "Graph-based semi-supervised learning: A comprehensive review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8174–8194, Nov. 2023.

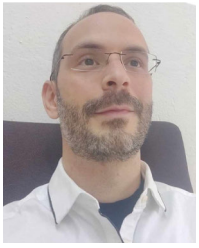
[31] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Mach. Learn.*, vol. 109, no. 2, pp. 373–440, 2020.

[32] N. Wambugu, Y. Chen, Z. Xiao, K. Tan, M. Wei, X. Liu, and J. Li, "Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks: A review," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, Dec. 2021, Art. no. 102603.

[33] H. Wu and S. Prasad, "Semi-supervised deep learning using pseudo labels for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1259–1270, Mar. 2018.

[34] Z. Zhang, Y. Ding, X. Zhao, L. Siye, N. Yang, Y. Cai, and Y. Zhan, "Multireceptive field: An adaptive path aggregation graph neural framework for hyperspectral image classification," *Expert Syst. Appl.*, vol. 217, May 2023, Art. no. 119508.
[35] Z. Zhang, "Semi-supervised hyperspectral image classification algorithm based on graph embedding and discriminative spatial information," *Microprocessors Microsyst.*, vol. 75, Jun. 2020, Art. no. 103070.
[36] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 16, 2003, pp. 1–8.
[37] D. Zhou and B. Schölkopf, "Learning from labeled and unlabeled data using random walks," in *Pattern Recognition*. Tübingen, Germany: Springer, Aug./Sep. 2004, pp. 237–244.
[38] X. Zhu and Z. Ghahramanin, "Learning from labeled and unlabeled data with label propagation," Carnegie-Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CALD-02-107, 2002.

IOANNIS GEORGOULAS received the degree in computer science from the Aristotle University of Thessaloniki, and the master's degree in the field of development and management of advanced information systems from the University of Piraeus. He has participated in several H2020 research and development projects, focusing on machine learning and computer vision applications.



EFTYCHIOS PROTOPAPADAKIS received the degree in production engineering and management, the M.S. degree in management and business administration, and the Ph.D. degree in decision systems from the Technical University of Crete, and the Post-Ph.D. degree in semi-supervised deep learning models from the National Technical University of Athens. He is an Assistant Professor with the Department of Applied Informatics, University of Macedonia. He has been an engineer with the European and Interegg projects, since 2010. His research outcomes resulted in more than 80 publications in international journals and conferences. His research interest includes machine learning applications. He received multiple awards, which include the University Scholarships for Excellence, the Technical Chamber of Greece Award for Excellence in Studies, the State Scholarship Foundation–SIEMENS Doctorate Scholarship 2012, and the Post-Ph.D. Scholarship of the National Strategic Reference Framework, from 2014 to 2020.

KONSTANTINOS MAKANTASIS received the Diploma, M.Sc., and Ph.D. degrees in computer engineering from the Technical University of Crete. He is a Lecturer with the Department of AI, University of Malta. He received the prestigious MSCA IF Widening Fellowship, to work on tensor-based machine learning methods for affect modeling. He has more than 50 publications in international journals and conferences on computer vision, signal and image processing, and machine learning. He is mostly involved and interested in computer vision, machine learning/pattern recognition, and probabilistic programming.

DYLAN SEYCHELL (Senior Member, IEEE) received the B.Sc.IT. degree (Hons.) in computer science and artificial intelligence and the M.Sc. degree in artificial intelligence from the University of Malta, in 2010 and 2011, respectively, and the Ph.D. degree in computer vision from the Department of Communications and Computer Engineering, University of Malta.

From 2011 to 2017, he was a Resident Academic with the Saint Martin's Institute of Higher Education, where he was the Head of the Computing Department, for five years. In 2017, he joined the Department of Artificial Intelligence, University of Malta, as a Resident Academic. In 2015, he was selected to lead Malta's Google Developers Group. He was a member of the Malta Neuroscience Network and the Malta National AI Taskforce, responsible for developing the national AI strategy. In 2023, he was appointed as a Technical Expert with the Malta Digital Innovation Authority. He was a technology advisor and a coordinator of several high-profile heritage projects. His research interests include saliency, AI in media, AI safety, and user experience design. He received several international awards for his work, such as the Gold Seal for e-Excellence at CeBit, in 2011; and the First Prize and the Runner-Up by the European Satellite Navigation Competition (Living Labs), in 2010 and 2017, respectively.

ANASTASIOS DOULAMIS received the Diploma and Ph.D. degrees (Hons.) in electrical and computer engineering from the National Technical University of Athens (NTUA). He was an Associate Professor with the Technical University of Crete. Currently, he is an Associate Professor with NTUA. He is the author of more than 350 papers in leading journals and conferences receiving more than 10,000 citations. He has received several awards in his studies, including the Best Greek Student Engineer and the Best Graduate Thesis Award. He has served as a program committee member for several major conferences of IEEE and ACM.

NIKOLAOS DOULAMIS received the Diploma and Ph.D. degrees (Hons.) in electrical and computer engineering from the National Technical University of Athens (NTUA), Athens, Greece, in 1995 and 2000, respectively. He is currently a Professor with NTUA. He has involved in several European research projects. He has authored more than 100 (300) journals (conference) papers in the field of signal processing and machine learning, many of them with applications in geoinformatics and Earth sciences. He has received more than 11,000 citations. He received many awards, including the Best Student Among All Engineers and the Best Paper Awards. He was an organizer and/or a TPC member of major IEEE conferences.