Impact of Training Set Size on Representation Learning for Hyperspectral Image Classification

Victor Andres Ayma Quirita[1], Aramis Palacios[2], Victor Hugo Ayma Quirita[2], Walter Aliaga[2], Gilson A. O. P. Costa[3]

[1] Dept. of Engineering, Pontifical Catholic University of Peru, Lima, Peru - vaaymaq@pucp.pe
[2] Universidad del Pacifico, Lima, Peru - (aa.palaciosa, vh.aymaq, w.aliagaa)@up.edu.pe
[3] Dept. of Informatics and Computer Science, Rio de Janeiro State University, Rio de Janeiro, Brazil - gilson.costa@ime.uerj.br

Keywords: Dimensionality Reduction, Hyperspectral Image Classification, Autoencoders, Orthogonal Autoencoders, Representation Learning, Limited Training Data

Abstract

Nowadays, the ever-increasing amount of information provided by hyperspectral sensors requires efficient solutions for facilitating subsequent data analysis. Dimensionality reduction plays a central role in this context, as it allows the extraction of meaningful and compact representations from high-dimensional hyperspectral data. Existing methodologies address data representation problems through dimensionality reduction techniques, predominantly employing Principal Component Analysis (PCA), Autoencoders (AE), and more recently, Hyperspectral Orthogonal Autoencoders (HOAE). However, these approaches commonly rely on the entire image to build projection models, which may result in high computational costs. A pragmatic attempt to mitigate such computational challenge is to use a subset of the image data to construct accurate data representation models. In this work, we investigate the extent to which using a reduced number of training samples affects the quality of the latent space generated by AE and HOAE models, and how this impacts classification performance. Experiments conducted on the Pavia University hyperspectral dataset demonstrate that the representation efficacy of the AE and HOAE models significantly exceeds that of traditional hyperspectral dimensionality reduction algorithms, such as PCA. We also show that competitive classification results can be obtained even when the representation models are trained with a small portion of the image, which opens the door to more computationally efficient pipelines.

1. Introduction

According to Ghamisi et al., (2017), recent advances in hyperspectral imaging acquisition technologies provide rich and unlimited spatial and spectral information. The subsequent hyperspectral image analysis enables different Remote Sensing applications, such as Earth's observation, environmental and climate change monitoring, forest management, precision agriculture, and urban planning, among others, where segmentation and classification techniques are widely used to adequately exploit the large spectral data provided by hyperspectral sensors (Minaee et al., 2022; Paoletti et al., 2019).

Despite its importance in Remote Sensing, several factors such as high dimensionality, data complexity and volumes, and the presence of mixed pixels make the processing of hyperspectral data naturally complex, imposing major challenges to adequately store, process, and analyze such type of data (Wu et al., 2016), which hinder the deployment of effective hyperspectral image classification models. In this sense, a major challenge is to adequately handle the high spectral dimensionality associated with large volumes of data (Wu et al., 2021; Thilagavathi et al., 2018). To address this, dimensionality reduction (DR) techniques are therefore essential and widely adopted to transform the original data into more compact and discriminative representations while preserving the most relevant information; nevertheless, controlling the demand for computational power imposes a major requirement of diminishing the loss of essential features when performing this task.

Scientific literature reports several approaches for representing hyperspectral imaging data in lower-dimensional spaces, ranging from conventional feature selection to feature extraction processes (Wu et al., 2021; Zhang et al., 2015). Feature selection methods aim to choose relevant bands or pixels (e.g., mutual information, reliefF, sparse learning), while feature extraction methods create new representations, often using linear or nonlinear transformations. Among the latter, the most representative and currently used techniques are Principal Component Analysis (PCA) and Autoencoders (AE) (Haut et al., 2019; Su et al., 2017; Sellami et al., 2018; Xu et al., 2018), both based on unsupervised learning.

Traditional linear approaches, such as PCA, have long been used in hyperspectral analysis to reduce dimensionality while preserving the most relevant variance. However, PCA often fails to capture the nonlinear spectral-spatial structures that are critical to discriminating complex land cover classes, and requires processing of the entire image to project the data into lower-dimensional spaces, thus imposing high demands for hardware capabilities such as processing speed and memory (Wu et al., 2016).

Advances in deep learning have enabled the development of nonlinear DR approaches, such as Autoencoders (AEs) and their variants, such as Hyperspectral Orthogonal Autoencoders (HOAE), which are capable of learning compact but highly informative latent representations (Haut et al., 2019; Ayma et al., 2020). While AE and HOAE can better capture complex structures and have shown improved classification accuracy, they also involve substantial computational costs when trained on the full hyperspectral image. Other approaches such as manifold learning (e.g., LLE, Isomap) and tensor-based techniques have also been explored, although they are often less scalable or less robust in noisy conditions (Song et al., 2024; Ghamisi et

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-21-2026

al., 2017). Overall, the trade-off between representation quality and computational efficiency remains an open problem.

Recent works have emphasized few-shot and self-supervised strategies to alleviate the dependency on large annotated datasets while improving representation learning. For instance, Li et al. (2023) proposed a meta-learning framework with regularized finetuning for few-shot hyperspectral image classification, demonstrating competitive results when only a handful of labeled samples are available. Similarly, Chen et al. (2025) introduced a spectral–spatial self-supervised learning paradigm, where pretext tasks enable the extraction of transferable representations that generalize well to new classes under few-shot conditions.

Other contributions have further advanced the field by exploring ensemble meta-optimization and transformer-based architectures. An ensemble of meta-optimizers has been shown to improve classification robustness in hyperspectral scenarios with few shots (Hao et al., 2024), while transformer-based models such as S3L leverage spectral–spatial attention mechanisms for self-supervised pretraining (Guo and Liu, 2024). More recently, hybrid deep transformer models coupled with few-shot learning strategies have demonstrated promising performance for hyperspectral image classification under data scarcity, highlighting the growing synergy between scalable architectures and data-efficient learning paradigms (Ran et al., 2023).

In this context, training representation models using only a subset of the image emerges as a potential strategy to mitigate computational costs while preserving classification performance. However, the impact of such sampling on different dimensionality reduction methods and subsequent classification tasks has not been extensively quantified. Despite recent progress, most studies still emphasize improving classification accuracy under standard conditions, paying limited attention to how the size of the training set influences the quality and robustness of the learned latent representations.

In this work, we aim to assess the impact of taking different hyperspectral image proportions in generating data representation for hyperspectral image classification. Thus, we produced several representation models based on PCA, AE, and HOAE, and used them as input to train and test different classifiers, to assess their data representation capabilities. We conduct our experiments using the Pavia University dataset, which is widely adopted in the literature and provides a suitable benchmark to evaluate the influence of training size on model representation effectiveness. However, our goal is not to tailor solutions to a specific dataset, but to analyze general trends in the trade-off between representation cost and classification performance.

First, we randomly sampled several subsets from the Pavia University hyperspectral dataset. Next, we use each subset to build representation models of the AE and HOAE approaches with distinct configurations on the number of dimensions (2, 4, 8, 16, 32, and 64); note that PCA is applied to the entire image, as is common practice in the literature. Then, we apply each representation model to project the image into lower-dimensional spaces. Later, we used such representations as inputs for Support Vector Machine (SVM), Decision Trees (DT), Random Forest (RF), Naive Bayes (NB), and Xtreme Gradient Boost (XGB) classifiers to perform pixel-wise hyperspectral classification. Finally, we use the classification outcomes to assess the performance of each representation strategy.

The remainder of the paper is organized as follows. Section 2 presents the foundations for implementing our representation method with Autoencoders. Section 3 describes the complete methodology and provides information regarding the experimental procedure adopted to assess the performance of our method. Section 4 presents the analysis performed over the experimental results. Finally, Section 6 presents our final remarks and directions for future work.

2. Dimensionality Reduction via Autoencoders

Autoencoders are neural network architectures designed to learn compact latent space representations while reproducing input patterns at their outputs. A conventional Autoencoder consists of two main stages: encoding and decoding. It is typically organized into an input layer with d neurons, a hidden layer with h neurons, and an output layer with d neurons, as illustrated in Figure 1. In this configuration, the encoding stage $f(\cdot)$ maps an input vector $\mathbf{x} \in \mathbb{R}^d$ into a latent representation $\mathbf{z} \in \mathbb{R}^h$, while the decoding stage $g(\cdot)$ reconstructs an approximation $\widehat{\mathbf{x}} \in \mathbb{R}^d$ of the original input. The encoding and decoding processes are expressed in Equations 1 and 2, respectively, where $\mathbf{W}_f$ and $\mathbf{W}_g$ denote weight matrices, $\mathbf{b}_f$ and $\mathbf{b}_g$ are bias vectors, and σ_f and σ_g represent activation functions.

$$f(\mathbf{x}) = \sigma_f(\mathbf{W}_f\mathbf{x} + \mathbf{b}_f) = \mathbf{z} \tag{1}$$

$$g(\mathbf{z}) = \sigma_g(\mathbf{W}_g\mathbf{z} + \mathbf{b}_g) = \widehat{\mathbf{x}} \tag{2}$$

Generally, the parameters $\mathbf{W}_f$, $\mathbf{W}_g$, $\mathbf{b}_f$, and $\mathbf{b}_g$ are optimized by minimizing the reconstruction error between the original input $\mathbf{x}$ and its reconstruction $\widehat{\mathbf{x}}$, commonly computed using the squared Euclidean distance:

$$L(\mathbf{x}, \widehat{\mathbf{x}}) = \|\mathbf{x} - \widehat{\mathbf{x}}\|^2 \tag{3}$$

where $L(\cdot)$: reconstruction error
$\|\cdot\|$: Euclidean norm

During training, the Autoencoder learns a compressed representation $\mathbf{z}$ that captures the most salient features of the input $\mathbf{x}$. To improve the quality and interpretability of this latent representation, the Hyperspectral Orthogonal Autoencoder (HOAE) (Ayma et al., 2020) introduces an orthogonality constraint among the components of $\mathbf{z}$. This is achieved by incorporating a regularization term into the loss function, which

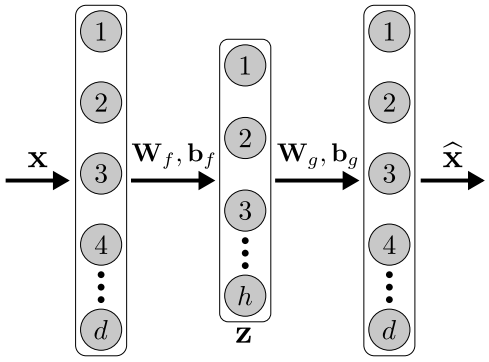


Figure 1. Single-layer Autoencoder architecture.

penalizes the deviation of the covariance matrix $\mathbf{z}^T\mathbf{z}$ from the identity matrix $\mathbf{I}$. The modified loss function is given by:

$$L'(\mathbf{x}, \widehat{\mathbf{x}}) = \|\mathbf{x} - \widehat{\mathbf{x}}\|^2 + \lambda \left\|\mathbf{z}^T\mathbf{z} - \mathbf{I}\right\|^2 \quad (4)$$

Here, λ is a regularization hyperparameter that controls the strength of the orthogonality penalty. A value of $\lambda = 0$ produces the conventional Autoencoder loss (Equation 3), whereas increasing λ enforces greater orthogonality among the latent components.

3. Methodology

The most frequently used approaches for representing image data in hyperspectral image processing methods are PCA and AE. In practice, such techniques require processing the entire image data, which may become computationally prohibitive given the computational power and time constraints when the dataset volume increases substantially.

A feasible solution to overcome this limitation would be to use a small portion of the image data to train a representation model. Nevertheless, an important question arises: How much image data is required to generate acceptable data representation models without significantly impacting the subsequent hyperspectral image processing tasks? To answer that question, in this work, we assess the impact of image data usage on generating AE and HOAE representation models. Our methodology, depicted in Figure 2, comprises six stages designed to assess the impact of dimensionality reduction representations on classification processes from a hyperspectral input image I of $M \times N \times B$ rows, columns and bands, respectively.

The first stage, *Data Preparation for Representation Learning* (DPRL), provides S_k training sets for learning compact representations of the hyperspectral input image I. Initially, we rearrange I into a bidimensional array J of size $N_J \times B$, where N_J and B are the number of pixels and hyperspectral bands in I, respectively. Next, we split J into subsets R and r to train and test different image representation models. To assess the influence of training sample size, we randomly sample the S_k subsets of R, each with different proportions of data k.

The second stage, *Representation Learning* (RL), aims at learning representations of the hyperspectral input image I given the S_k training sets from the previous stage. In this manner, this stage provides representation learning models $G_{k,d}(\cdot)$, each of which projects the hyperspectral input image I into a low d-dimensional space.

The third stage, *Data Preparation for Classification* (DPC), provides training and testing sets to fit and evaluate hyperspectral image classifiers. First, we select the labeled pixels from the input image I and arrange them into a new bidimensional array L. Next, we partition L into constant training (T) and testing (t) sets for the subsequent classification processes. Notice that L has a size of $N_L \times B$, where N_L corresponds to the number of labeled pixels in I.

In the fourth stage, *Feature Extraction* (FE), we project the training and testing sets, T and t, respectively, into a d-dimensional space defined by the model $G_{k,d}(\cdot)$ from the RL stage. We expect that the projected training and testing sets, $T'_{k,d}$ and $t'_{k,d}$, provide enough information to facilitate classifier learning. Here, each labeled pixel $\mathbf{p}$ of length B is thus represented by a d-dimensional vector $\mathbf{x}$ in the reduced space.

In the fifth stage, *Classification Learning* (CL), we take as inputs a tuple $(\mathbf{x}, y)$ of projected vectors and class labels to train a classification model $C_{k,d}(\cdot)$. In this manner, the classification model captures the influence of the proportion of training samples (k) used to train the representation model $G_{k,d}(\cdot)$.

Finally, the *Classifier's Evaluation* (CE) stage aims to evaluate the performance of a given hyperspectral image classifier. It accepts as input a classifier $C_{k,d}(\cdot)$ and a testing set $t'_{k,d}$ to produce a performance indicator m. The testing set $t'_{k,d}$ remains constant for all classifier evaluations to enable fair comparisons.

3.1 Implementation Details

We implemented the image data representation models and the five classifiers using TensorFlow v2.11 and scikit-learn v1.2.0, respectively. We used the `metrics` module within scikit-learn to evaluate the classifiers' performances.

In the *DPRL* stage, we set the training and testing subsets (R and r) to contain 80% and 20% of the image data J, respectively. We then varied k among 25%, 50%, 75%, and 100% to sample four S_k training subsets from R. In the *RL* stage, we considered PCA, AE, and HOAE as the representation learning methods, projecting the input image data S_k into low-dimensional spaces with $d \in \{2, 4, 8, 16, 32, 64\}$. Both AE and HOAE were implemented using a single hidden layer with d neurons. We followed the training configuration reported in Ayma et al., (2020) for AE and HOAE models. Furthermore, for HOAE, we experimented with orthogonality penalization parameters $\lambda = 0.1$ and $\lambda = 0.01$ across all d and k values.

In the *DPC* stage, we excluded pixels with zero classification label to build the labeled array L. We partitioned L into training (T) and testing (t) sets for classification using 80% and 20% splits, respectively. Then, both T and t were projected into the reduced d-dimensional space ($T'_{k,d}$ and $t'_{k,d}$) using the representation models $G_{k,d}(\cdot)$.

For the *CL* stage, we trained five representative classifiers using default parameters: Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting (XGB), and Naive Bayes (NB) . Finally, in the *CE* stage, we evaluated the classifiers using their overall accuracy as the main performance metric.

4. Experimental Results

To assess the impact of the number of training samples on image data representation, we evaluated the performance of PCA, AE, and HOAE techniques in the context of hyperspectral image classification. Specifically, we analyzed the performance of five classifiers trained on different image data representations generated by the aforementioned techniques. This section presents a detailed analysis of the experimental outcomes.

4.1 Dataset Description

The *Pavia University Dataset* (**PaviaU**) is a hyperspectral image acquired by the Reflective Optics Spectrographic Imaging System (ROSIS-3) during a flight campaign over the University

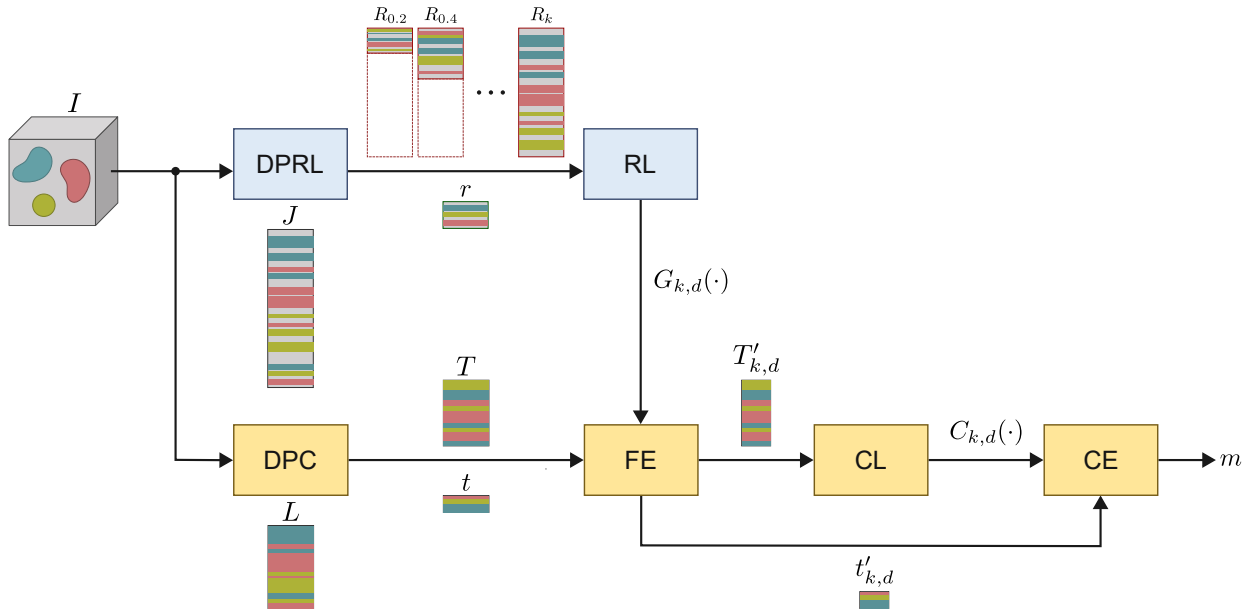


Figure 2. Proposed methodology for assessing the influence of using a reduced number of samples for creating different representation models for hyperspectral image classification

of Pavia, in northern Italy. The dataset contains 103 spectral bands covering the spectral range from 0.43 μm to 0.86 μm. Each spectral band has a 610×610 image resolution and a spatial resolution of 1.3 meters per pixel. The image includes various land use classes, such as asphalt, meadow, gravel, tree, metal sheets, bare soil, bitumen, bricks, and shadows.

4.2 Results

To evaluate the influence of data volume on representation learning, we conducted a series of experiments using the Pavia University dataset. We compared classification performance based on representations generated by AE and HOAE models against PCA, which was used as the baseline.

Figure 3 shows the overall accuracy obtained by the SVM, RF, XGB, DT, and NB classifiers, for different numbers of dimensions extracted by each representation model (AE, HOAE, and PCA). A visual inspection reveals that AE and HOAE produce remarkably similar results and significantly outperform PCA across all dimensional settings. These results suggest that nonlinear methods, such as AE and HOAE, capture more relevant structures from hyperspectral data than PCA.

Each accuracy value for AE and HOAE in Figure 3 corresponds to the mean overall accuracy across the four training subsets S_k (with different proportions of data k), and the vertical bars at each value indicate the standard deviation. Notably, the standard deviation is relatively low, indicating stable performance across different data volumes. In contrast, PCA results were computed using the full image dataset, as dimensionality reduction via PCA requires access to all data to compute the principal components.

To further explore the relationship between training data volume (required to create the representation models) and classification performance, we analyzed the XGB classifier (identified as the top-performing classifier) in more detail. Figure 4 shows the classification results for two extreme cases: $k = 25\%$ and $k = 100\%$ of the training data R. Remarkably, the classification performance is similar in both cases, especially when the number of representation dimensions exceeds 16. In these settings, both AE and HOAE achieve overall accuracies above 90%.

Furthermore, we observe that with a low number of dimensions, the regular AE model ($\lambda = 0$) often outperforms the HOAE. However, as the number of dimensions increases, the performance gap between AE and HOAE diminishes. This suggests that orthogonality constraints may benefit compact representations, but have limited effect when higher-dimensional latent spaces are used.

These findings highlight that the dimensionality of the latent space has a more significant impact on classification performance than the percentage of training data used for representation learning. Even when using only 25% of the data to train the AE or HOAE models, the resulting representations led to competitive and often high classification accuracy, especially when using 16 or more latent dimensions. This suggests that it is not necessary to use the entire hyperspectral image to obtain effective representations for classification tasks, which in turn reduces the computational cost of the representation learning process.

5. Discussion

The results obtained in this study provide new information on the relationship between the size of the training set and the quality of the representations learned for hyperspectral image classification. Unlike most conventional studies that primarily emphasize maximizing classification accuracy under standard conditions, our work explicitly addresses how different dimensionality reduction techniques behave under data scarcity scenarios.

Our findings show that AE and HOAE consistently outperform PCA in terms of classification accuracy, even when trained with relatively small datasets. This suggests that deep representation learning approaches are capable of extracting more discriminative latent features while preserving relevant spectral

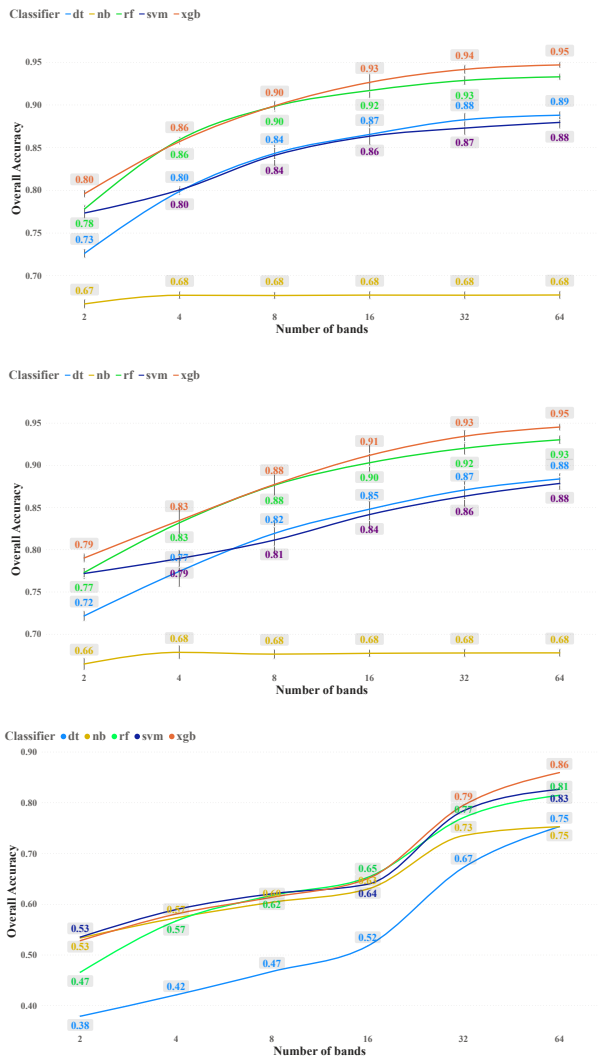


Figure 3. Overall Accuracy vs Number of dimensions for the: (top) AE, (middle) HOAE, and (bottom) PCA representation models.

information, making them promising alternatives to classical linear methods when annotated data are limited. These observations complement recent advances in few-shot and self-supervised paradigms, which also seek to alleviate the dependency on large labeled datasets by promoting data-efficient representation learning.

By quantifying the trade-offs between training set size and representation quality, our work highlights that nonlinear models are less sensitive to data scarcity compared to linear projections. This finding resonates with recent contributions that leverage meta-learning, self-supervised pretraining, and transformer-based strategies, all of which converge on the need to design scalable yet data-efficient solutions. In this sense, training with subsets of the image emerges as a simple but effective strategy to mitigate computational costs while maintaining competitive classification performance, bridging a gap not extensively explored in prior literature.

The efficiency and robustness of the AE and HOAE representations make them suitable candidates for large-scale operational scenarios, such as the forthcoming generation of hyperspectral satellite missions and UAV-based surveys. In these contexts, the ability to achieve competitive performance without requiring extensive labeled datasets is particularly relevant, as it reduces both annotation costs and computational requirements.

However, some limitations of this work must be acknowledged. The current analysis is restricted to a single benchmark dataset, and although the observed trends are consistent, additional validation is necessary on more diverse and challenging benchmarks to confirm the generalizability of our conclusions.

Furthermore, while we focus on PCA, AE, and HOAE, other paradigms such as contrastive learning, ensemble meta-optimization, and transformer-based DR architectures could be integrated in future studies to expand the scope of comparison. These directions open promising opportunities to further investigate the scalability and robustness of representation learning in data-constrained hyperspectral settings.

6. Conclusion

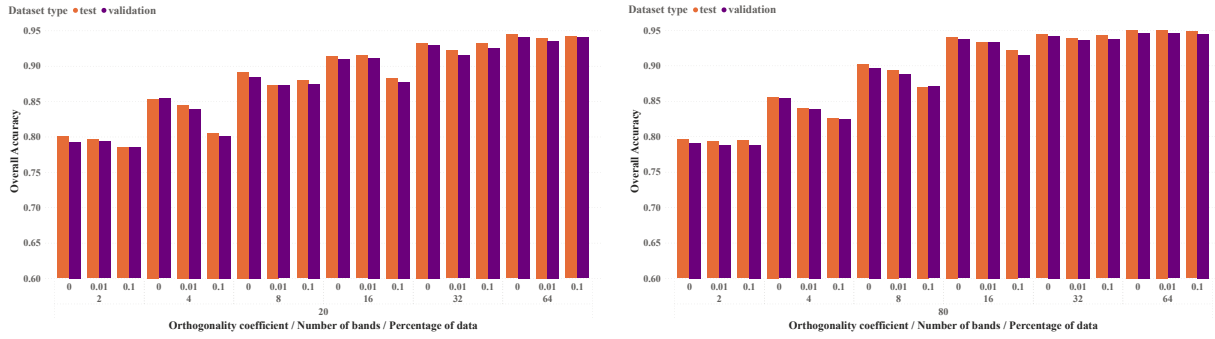
In this work, we investigated the impact of hyperspectral data representation on classification tasks by employing Autoencoders (AE) and Hyperspectral Orthogonal Autoencoders (HOAE), particularly when only a reduced number of training samples are available for learning the representation models. We compared these approaches with the traditional Principal Component Analysis (PCA), conducting experiments on the Pavia University hyperspectral dataset. As discussed in Section 5, our findings contribute to the broader discourse on efficient representation learning under data scarcity. Future work will further extend these insights with additional datasets and methodological refinements.

Our results demonstrated that both AE and HOAE significantly outperform PCA in terms of classification accuracy. We used five classifiers: Support Vector Machines (SVM), Decision Trees (DT), Random Forest (RF), Naive Bayes (NB), and Extreme Gradient Boosting (XGB); and evaluated their performance across different numbers of latent dimensions and different data usage proportions for training the AE and HOAE representation models. We observed that increasing the number of latent dimensions consistently led to improved classification accuracy. Moreover, HOAE achieved accuracy levels comparable to those of AE, particularly at higher latent dimensionalities.

Importantly, our findings show that the number of latent dimensions has a stronger influence on classification performance than the percentage of data used to train the representation models. Even when only a quarter of the training data was used, AE and HOAE still produced effective representations that enabled high classification accuracy, especially when sufficient latent dimensions were retained. This finding suggests that it is feasible to reduce computational cost without compromising performance by training representation models on smaller data subsets.

Finally, our results highlight the potential of the HOAE approach for hyperspectral data representation, especially in scenarios with limited training data. Future work may focus on enhancing the orthogonality constraint, exploring hyperparameter optimization strategies, and extending the HOAE framework to other remote sensing tasks or multimodal representation learning.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-21-2026 | © Author(s) 2026. CC BY 4.0 License.



(a) $k = 25\%$ (b) $k = 100\%$

Figure 4. Overall Accuracy of the XGB classifier as a function of the λ orthogonality regularization parameter and the number of latent dimensions, for training subsets of (a) $k = 25\%$ and (b) $k = 100\%$ of the training data R.

References

Ayma, V. H., Ayma, V. A., Gutierrez, J., 2020. Dimensionality Reduction via an Orthogonal Autoencoder Approach for Hyperspectral Image Classification. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2020, 357–362. https://isprs-archives.copernicus.org/articles/XLIII-B3-2020/357/2020/.

Chen, W., Zhang, Y., Xiao, Z., Chu, J., Wang, X., 2025. Spectral-spatial self-supervised learning for few-shot hyperspectral image classification.

Ghamisi, P., Yokoya, N., Li, J., Liao, W., Liu, S., Plaza, J., Rasti, B., Plaza, A., 2017. Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 37-78.

Guo, H., Liu, W., 2024. S3L: Spectrum Transformer for Self-Supervised Learning in Hyperspectral Image Classification. *Remote Sensing*, 16(6). https://www.mdpi.com/2072-4292/16/6/970.

Hao, T., Zhang, Z., Crabbe, M. J. C., 2024. Few-Shot Hyperspectral Remote Sensing Image Classification via an Ensemble of Meta-Optimizers with Update Integration. *Remote Sensing*, 16(16). https://www.mdpi.com/2072-4292/16/16/2988.

Haut, J. M., Gallardo, J. A., Paoletti, M. E., Cavallaro, G., Plaza, J., Plaza, A., Riedel, M., 2019. Cloud Deep Networks for Hyperspectral Image Analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12), 9832-9848.

Li, W., Liu, Q., Zhang, Y., Wang, Y., Yuan, Y., Jia, Y., He, Y., 2023. Few-Shot Hyperspectral Image Classification Using Meta Learning and Regularized Finetuning. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1-14.

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D., 2022. Image Segmentation Using Deep Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7), 3523-3542.

Paoletti, M., Haut, J., Plaza, J., Plaza, A., 2019. Deep learning classifiers for hyperspectral imaging: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158, 279-317. https://www.sciencedirect.com/science/article/pii/S0924271619302187.

Ran, Q., Zhou, Y., Hong, D., Bi, M., Ni, L., Li, X., Ahmad, M., 2023. Deep transformer and few-shot learning for hyperspectral image classification. *CAAI Transactions on Intelligence Technology*, 8, 1323-1336. https://digital-library.theiet.org/doi/abs/10.1049/cit2.12181.

Sellami, A., Farah, M., , 2018. Comparative study of dimensionality reduction methods for remote sensing images interpretation. *2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 1–6.

Song, W., Zhang, X., Yang, G., Chen, Y., Wang, L., Xu, H., 2024. A Study on Dimensionality Reduction and Parameters for Hyperspectral Imagery Based on Manifold Learning. *Sensors*, 24(7). https://www.mdpi.com/1424-8220/24/7/2089.

Su, J., Yi, D., Liu, C., Guo, L., Chen, W., 2017. Dimension Reduction Aided Hyperspectral Image Classification with a Small-sized Training Dataset: Experimental Comparisons. *Sensors*, 17(12). https://www.mdpi.com/1424-8220/17/12/2726.

Thilagavathi, K., Vasuki, A., , 2018. Dimension reduction methods for hyperspectral image: A survey. *International Journal of Engineering and Advanced Technology*, 8(2), 160-167.

Wu, Z., Li, Y., Plaza, A., Li, J., Xiao, F., Wei, Z., 2016. Parallel and Distributed Dimensionality Reduction of Hyperspectral Data on Cloud Computing Architectures. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 9(6), 2270-2278. https://ieeexplore.ieee.org/document/7446277.

Wu, Z., Sun, J., Zhang, Y., Wei, Z., Chanussot, J., 2021. Recent Developments in Parallel and Distributed Computing for Remotely Sensed Big Data Processing. *Proceedings of the IEEE*, 109(8), 1282-1305. https://ieeexplore.ieee.org/document/9458233.

Xu, J.-L., Esquerre, C., Sun, D.-W., 2018. Methods for performing dimensionality reduction in hyperspectral image classification. *J. Near Infrared Spectrosc.*, 26(1), 61–75. http://www.osapublishing.org/jnirs/abstract.cfm?URI=jnirs-26-1-61.

Zhang, Q., Tian, Y., Yang, Y., Pan, C., 2015. Automatic Spatial–Spectral Feature Selection for Hyperspectral Image via Discriminative Sparse Multimodal Learning. *IEEE Transactions on Geoscience and Remote Sensing*, 53(1), 261-279.