Fractal Autoencoder-Based Unsupervised Hyperspectral Bands Selection for Remote Sensing Land-Cover Classification †

Sara Benali [1,*], Mohammed El Amin Larabi [2] and Dalila Attaf [2]

1   National Higher School of Geodesy Science and Spatial Techniques, Oran 31004, Algeria
2   Algerian Space Agency, Alger 16032, Algeria; arabi.mohamed.lamine@gmail.com (M.E.A.L.); dalila.attaf@gmail.com (D.A.)
*   Correspondence: benalisara931@gmail.com
†   Presented at the 4th International Electronic Conference on Applied Sciences, 27 October–10 November 2023; Available online: https://asec2023.sciforum.net/.

Abstract: Band selection is a frequently used dimension reduction technique for hyperspectral images (HSI) to address the "curse of dimensionality" phenomenon in machine learning (ML). This technique identifies and selects a subset of the most important bands from the original ones to remove redundancy and noisy information while maintaining optimal generalization ability. Band selection methods can be categorized into supervised and unsupervised techniques depending on whether labels are used or not. An unsupervised band selection and feature extraction framework is proposed in this study. The framework trains a sub-neural network to identify the most important and informative bands from the original data space, which is then projected to a reduced and more informative space. The classification performance of the selected bands' combination on the Pavia University HSI datasets has been verified using multiple machine learning algorithms. The proposed method not only enhances the classification results of HSI, but also reduces the computational time and data storage requirements compared to other state-of-the-art band selection approaches.

1. Introduction

High-dimensional datasets are common in various fields, such as image processing, genomics, finance, and more. These datasets have a wide range of features (attributes) and often surpass the number of samples available for analysis. This wealth of information is valuable, but it also presents numerous challenges, collectively known as the "curse of dimensionality". The latter involves issues such as increased computational complexity, overfitting, degraded model performance, and reduced interpretability. These challenges hinder the effectiveness of traditional data analysis methods [1].

Feature selection is a technique that identifies a subset of relevant features from a high-dimensional dataset. There are three main types of feature selection methods: filter methods, wrapper methods, and embedded methods. Filter methods assess feature importance independently of any specific learning algorithm, while wrapper methods use a specific learning algorithm to evaluate the impact of feature subsets on model performance [2]. Embedded methods combine feature selection seamlessly with the learning process itself. The choice of feature selection method depends on the specific application and the available resources. These methods cover a wide range of feature selection techniques, including Principal Feature Analysis (PFA) [3], which prioritizes key features through statistical measures; Multi-cluster Feature Selection (MCFS) [4], which leverages clustering techniques; Unsupervised Discriminative Feature Selection (UDFS), which seeks to maximize feature discrimination; and Principal Component Analysis (PCA), which focuses on orthogonal transformations. However, a significant challenge arises when the

selected features exhibit high correlations. This can potentially lead to the representation of only partial information and limit the global representativeness of the feature subset.

Multispectral imagery (MSI) captures a limited range of electromagnetic radiation in a few wide spectral bands. For some real-world applications, MSI can provide adequate information for the task. However, compared to other advanced satellite imagery, it offers less detailed data. In contrast, hyperspectral imagery (HSI) captures a broad range of electromagnetic radiation in hundreds of narrow bands, providing rich information about scene materials. This makes HSI ideal for applications like material identification, target detection, and environmental monitoring. However, HSI can be expensive and difficult to process. In HSI, feature or band selection and data compression are key techniques for managing large data volumes, improving analysis efficiency, and simplifying storage and transmission. This enables more effective HSI applications in areas like agriculture, mineral exploration, and environmental monitoring [5].

Deep learning-based feature selection methods, such as autoencoders (AE), use neural networks to automatically identify and extract the most important features from complex datasets. AE are a type of neural network that can learn compact representations of input data. This makes them well-suited for both feature selection and data compression tasks. By learning compact representations of input data, AE can enhance data analysis efficiency and preserve vital information across diverse domains [6].

A new framework of feature selection for HSI based on Fractal AE (FAE) [7] is introduced in this paper. FAE seeks to achieve optimal feature subsets that effectively balance the representation of information and diversity, which can enhance the performance of subsequent data analysis tasks. In the following sections, the details of FAE are delved into, its unique characteristics are showcased, and its effectiveness is demonstrated through experiments and comparisons with state-of-the-art methods.

This work is organized as follows: In Section 1, a detailed presentation of the architecture and formulation of AE and FAE is provided. Following that, in Section 2, our methodology for utilizing these techniques is elucidated, and our comparative analysis against several other methods is discussed.

2. Methodology

This section introduces an approach tailored for HSI analysis based on the concept of FAE. While building on the foundation of AE, the approach customizes the architecture to address the specific challenges of HSI data. Figure 1 shows the architecture of FAE, which forms the core component of this method. The design of the architecture aims to enable effective feature selection for HSI. In the following sections, an in-depth explanation of the architecture and its constituent parts is provided.

2.1. Formalization of AE

For HSI, we formalize the AE as follows:

$$min \; \|X - f(g(X))\|_F^2$$

Here, the encoder is represented by g, and the decoder is represented by f. The function $g(X)$ transforms the input HSI data X into a latent space $R^{n \times d}$, where d signifies the dimension of the bottleneck layer within the AE. To illustrate, the application of our approach to a HSI dataset is considered. In the context of HSI analysis, this formalization allows the essential spectral information to be effectively captured and represented within a reduced-dimensional latent space.
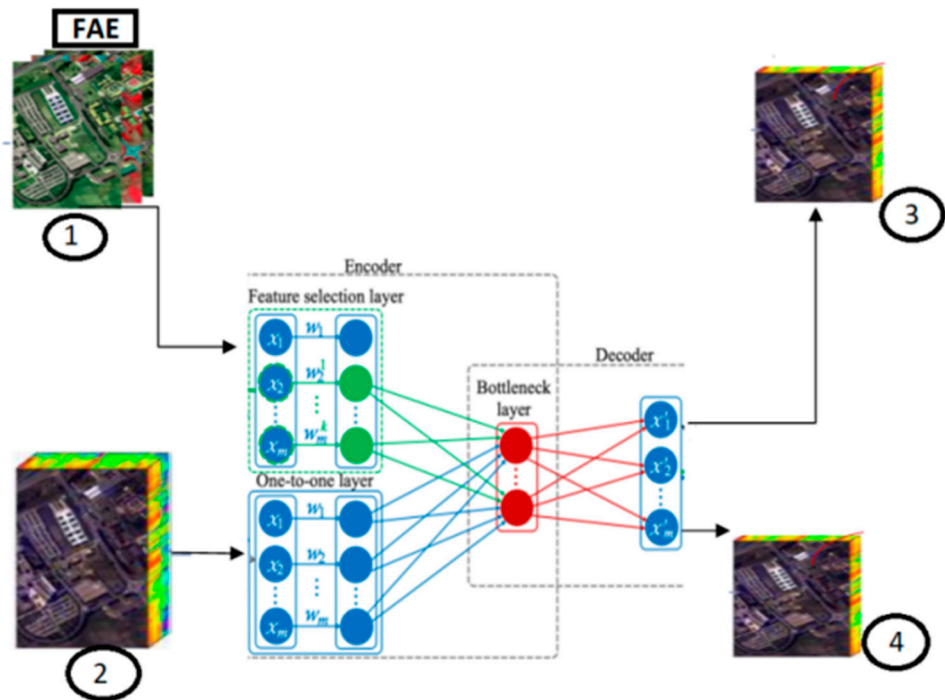
Figure 1. The architecture of FAE. The presented quantifies are (1) feature selection result, (2) input, (3) reconstruction based on the selected features, (4) reconstruction from the one-to-one layer.

2.2. Formalization of Feature Selection

Feature selection entails the process of pinpointing a subset of informative features within the original feature space, and it can be defined as follows:

$$min_{S^k,H} \| H(X_{S^k}) - X \|_F^2$$

Here, S^k represents the chosen subset of k features, X_{S^k} denotes the dataset derived from X by retaining only the features in S^k, and H signifies a mapping from the space defined by X_{S^k} to a new space, all performed without relying on any label or class information.

2.3. Formalization of FAE

FAE, a novel approach designed to tackle feature selection, introduces a concept akin to self-similarity in its operation. The primary objective of FAE is to select a subset of k informative features from a HSI X, such that the chosen features collectively retain as much information about the overall spectral content of the original samples as possible.

The operation of FAE is formalized as an optimization problem with two key components; the global reconstruction term minimizes the reconstruction error between the original HSI data X and the data reconstructed after passing through the encoder g and decoder f networks, considering the selected features represented by W_I. The diversity term is introduced to encourage the selected subset of features (W_I) to be diverse and not highly correlated with each other. This term ensures that the chosen features effectively capture various aspects of the HSI.

$$min_{w,g,f} \| X - f(g(W_I)) \|_F^2 + \lambda_1 \| X - f(g(W_I^{max\ k})) \|_F^2 + \lambda_2 \| W_I \|_1, \quad s.t.\ W_1 \geq 0$$

In the notation used here, a bold capital letter, for instance, W, represents a matrix, whereas a lowercase bold capital letter, like w, signifies a vector. The notation $Diag(w)$ represents the creation of a diagonal matrix with its diagonal elements derived from the vector w. The operation $W^{max\ k}$ is defined to retain the k largest entries of the vector w while setting the remaining entries to zero. Furthermore, $\|.\|_F$ denotes the Frobenius norm.

The overall objective function is balanced between these two terms, and is controlled by non-negative balancing parameters, λ_1, λ_2 and $W_I^{max\ k} = Diag\ (w^{max\ k})$. This approach is named FAE because of its intriguing characteristic: a small proportion of features selected in the second term can achieve performance similar to using the entire set of features in the first term when reconstructing the original HSI. This self-similarity trait becomes even more evident when FAE is applied to extract multiple feature subsets for different tasks. Firstly, FAE is utilized to perform feature selection on the HSI. FAE is tailored to select a subset of informative spectral bands from the original dataset while ensuring that the chosen features are diverse. This process aims to enhance the representativeness of the feature subset.

However, the innovation in FAE comes from its 'bottleneck' layer, which is part of the encoder. This bottleneck layer enables dimensionality reduction by compressing the HSI into a lower-dimensional latent space. This compression captures the most important information from the original data, emphasizing feature extraction. In this way, FAE skillfully balances between selecting informative yet diverse features. This complexity is represented by the "diversity term", which encourages the chosen features to be distinct and uncorrelated. This ensures the feature subset effectively covers a wide range of hyperspectral characteristics. In essence, FAE's bottleneck layer plays a critical role in extracting the essential spectral information while promoting feature diversity, making it a powerful tool for HSI feature selection.

In this neural network architecture, feature selection is an important part of the model's post-training process. Initially, all features are given equal weights, implying equal importance. As the model trains, these feature weights are updated. When the selection parameter is set to 'True', a feature selection mechanism is activated. This mechanism analyzes the feature weights to evaluate their significance. It determines a threshold based on the largest weight value, and any feature whose weight falls below this threshold is removed by setting it to zero. This process yields a reduced feature subset containing only the top *k* features ordered by their learned importance. Applying this post-training selection simplifies the model, lowers dimensionality, and improves generalization by focusing on the most relevant features identified by the neural network during training.

Once feature selection with FAE is completed, different supervised classification tasks are performed. Ensemble learning algorithms, such as Random Forest, LightGBM, XGBoost, and CatBoost, are also used for supervised classification. These classifiers are known for their robustness and ability to handle complex feature spaces. The selected FAE features are used as inputs for these classifiers, which improves classification accuracy and interpretability. This methodology enables a comprehensive evaluation of the effectiveness of FAE-based feature selection in supervised classification scenarios, contributing to a deeper understanding of HSI analysis techniques.

3. Experiments

3.1. Dataset Description

In this paper, the benchmarking dataset used is Pavia University. These data are commonly used in the HSI domain to assess and compare the performance of HSI processing and analysis algorithms.

3.2. Result and Discussion

In our study, the application of FAE for feature selection yielded notable improvements in classification performance. When compared to alternative feature selection methods, FAE consistently demonstrated superior results across various evaluation metrics, including accuracy, F1-score, recall, and reconstruction error, which is measured in mean squared error (MSE) for evaluating the model's reconstruction error.

As observed in Table 1, the FAE (Fractal autoencoder) approach yields excellent accuracy results across all classification methods. Furthermore, a closer examination of the data presented in Tables 2 and 3 reveals that FAE consistently outperforms other classification

methods, demonstrating superior F1 scores and recall values. These findings underscore the effectiveness and versatility of the FAE approach in enhancing classification model performance, highlighting its significant potential across various application contexts.

Table 1. Performance accuracy metric.

	Accuracy				
	UDFS	**MCFS**	**AE**	**PCA**	**FAE**
RF	0.48	0.80	0.85	0.79	0.85
LGBM	0.44	0.57	0.56	0.55	0.57
XGBOOST	0.55	0.81	**0.87**	0.80	0.85
CATBOOST	0.50	0.81	0.82	0.50	0.85

Table 2. Performance F1-score metric.

	F1-Score				
	UDFS	**MCFS**	**AE**	**PCA**	**FAE**
RF	0.26	0.87	0.86	0.82	**0.90**
LGBM	0.61	0.86	0.86	0.82	0.89
XGBOOST	0.29	0.82	0.82	0.83	0.88
CATBOOST	0.29	0.87	0.87	0.83	0.89

Table 3. Performance recall metric.

	Recall				
	UDFS	**MCFS**	**AE**	**PCA**	**FAE**
RF	0.24	0.88	0.86	0.80	**0.91**
LGBM	0.00	0.84	0.80	0.77	0.89
XGBOOST	0.30	0.85	0.86	0.83	0.88
CATBOOST	0.29	0.86	0.86	0.83	0.88

The success of FAE in obtaining the most pertinent and varied set of features from HSI underlines the observed improvements in classification accuracy and other performance metrics. This thorough feature selection procedure makes sure that the chosen features retain important spectral content information while also assisting in the reduction of dimensionality. FAE has an advantage over other feature selection methods since it can balance the preservation of important spectral information with the promotion of feature diversity. Because of this quality, FAE is especially well suited for HSI, where it's crucial to strike a careful balance between feature information and redundancy.

4. Conclusions and Future Work

Mixed results were obtained via supervised classification using different feature selection techniques. While some techniques outperformed others, the classification outcomes employing the Fractal Autoencoder (FAE) method's feature selection showed the most promise. This was completed in an effort to reduce the amount of time and money needed to process hyperspectral images (HSI) while still getting accurate classification results. With this decision, the workflow was streamlined and memory and computational needs were decreased, improving the process overall and lowering costs.

Author Contributions: Conceptualization, S.B., M.E.A.L. and D.A.; software, S.B. and M.E.A.L.; validation, S.B., M.E.A.L. and D.A.; formal analysis, S.B. and M.E.A.L.; investigation, S.B. and M.E.A.L.; data curation, S.B. and M.E.A.L.; writing—original draft preparation, S.B. and M.E.A.L.; writing—review and editing, S.B. and M.E.A.L.; visualization, S.B., M.E.A.L. and D.A.; supervision, M.E.A.L. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in: https://github.com/sara-benali/FAE.git (accessed on 1 November 2023).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bandos, T.V.; Bruzzone, L.; Camps-Valls, G. Classification of hyperspectral images with regularized linear discriminant analysis. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 862–873. [CrossRef]
2. Wah, Y.B.; Ibrahim, N.; Hamid, H.A.; Abdul-Rahman, S.; Fong, S. Feature selection methods: Case of filter and wrapper approaches for maximising classification accuracy. *Pertanika J. Sci. Technol.* **2018**, *26*, 329–340.
3. Lu, Y.; Cohen, I.; Zhou, X.S.; Tian, Q. Feature selection using principal feature analysis. In Proceedings of the 15th ACM International Conference on Multimedia, Augsburg, Germany, 24–29 September 2007; pp. 301–304.
4. Cai, D.; Zhang, C.; He, X. Unsupervised feature selection for multi-cluster data. In Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 25–28 July 2010; pp. 333–342.
5. Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–36. [CrossRef]
6. Balın, M.F.; Abid, A.; Zou, J. Concrete autoencoders: Differentiable feature selection and reconstruction. In Proceedings of the International Conference on Machine Learning, Beach, CA, USA, 10–15 June 2019; pp. 444–453.
7. Wu, X.; Cheng, Q. Fractal autoencoders for feature selection. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2021; Volume 35, pp. 10370–10378.