# Deep NMF and Autoencoder: A Comparative Analysis for Hyperspectral Unmixing Using Prisma Real Images

**Conference Paper** · July 2024

DOI: 10.1109/IGARSS53475.2024.10642930



**4 authors:**

Gaetano Settembre
University of Bari Aldo Moro
**10** PUBLICATIONS **39** CITATIONS
SEE PROFILE

Nicoletta Del Buono
University of Bari Aldo Moro
**89** PUBLICATIONS **721** CITATIONS
SEE PROFILE

Flavia Esposito
University of Bari Aldo Moro
**52** PUBLICATIONS **339** CITATIONS
SEE PROFILE

Taggio Nicolo
Planetek
**19** PUBLICATIONS **232** CITATIONS
SEE PROFILE

# DEEP NMF AND AUTOENCODER: A COMPARATIVE ANALYSIS FOR HYPERSPECTRAL UNMIXING USING PRISMA REAL IMAGES

*Gaetano Settembre, Nicoletta Del Buono, Flavia Esposito*

Department of Mathematics
University of Bari Aldo Moro, Bari, Italy

*Nicolò Taggio*

Planetek Italia S.r.l.
Bari, Italy

## ABSTRACT

In hyperspectral data, mixed pixels are frequent due to the low-medium spatial resolution of the imaging spectrometer, or to intimate mixing effects. Hence the process of blind hyperspectral unmixing, which separates the pixel spectra into a collection of spectral signatures and a set of fractional abundances, is a mandatory task in hyperspectral image processing. In this study, among models capable of performing linear spectral unmixing, we present a comparative analysis performed on real data acquired from the PRISMA (PRecursore IperSpettrale della Missione Applicativa) hyperspectral satellite between deep nonnegative matrix factorization and artificial neural network autoencoder based methods.

***Index Terms*—** Hyperspectral unmixing, deep nonnegative matrix factorization, autoencoder, PRISMA satellite data

## 1. INTRODUCTION

Remote sensing imaging known as hyperspectral images (HSIs) records the Earth's surface in numerous narrow and contiguous bands across the electromagnetic spectrum. Due to the physical structure of the sensors, HSIs are characterized by high spectral but low-medium spatial resolution. This suggests that a single pixel in HSIs represents a mixture of many distinct materials. In remote sensing, blind hyperspectral unmixing (HU) is the process of finding a set of spectral signatures for different pure materials (*endmembers*) that can be used to unmix all other pixels in the data and their corresponding fractional coverage (*abundances*) all at once [1]. Some physical constraints characterize the unmixing problem: nonnegativity of the endmember spectra (since a negative reflectance is not possible), and nonnegativity of abundances that also sum to one for each pixel. Currently, unmixing methods depend on the kind of mixing that is expected: they can be classified into linear mixing models (LMMs) and nonlinear mixing models. LMMs are most widely used because of their simplicity and effectiveness; they are based on the assumption that the photons reaching the hyperspectral sensor must interact with only one material so that each mixed pixel can be expressed as a linear combination of a finite number of endmembers weighted by the corresponding abundances [1].

In this study, among LMMs, we present a comparative analysis between two particular fully unsupervised approaches: deep nonnegative matrix factorization (DNMF) [2, 3] and artificial neural network autoencoder (AE)-based methods [4]. In the peculiar context of HU, DNMF represents the recent deep extension of nonnegative matrix factorizations (NMF), which largely demonstrated their capabilities of automatically extracting latent feature representation from HSIs also preserving the physical nonnegativity of data [5]. DNMF benefits from linearity's interpretability power, enabling a hierarchical and lower-dimensional representation of the input nonnegative latent characteristics.

On the other hand, autoencoders are a powerful tool for extracting meaningful features and reducing the dimensionality of HSIs. AE are neural networks that aim to reconstruct their input data, typically by learning a compact representation (the "encoding"). This process forces the network to capture essential features in data [4]. Similar ideas inspired both DNMF and AE; furthermore, for both classes of algorithms, the learning process can be improved by adding additional physics-inspired constraints, e.g., abundance non-negativity constraint, abundance sum-to-one constraint, sparsity, and smoothness of the obtained abundance maps.

This work presents initial comparisons between some AE and DNMF models that emphasize the same spectral and spatial properties. Their effectiveness is evaluated on some well-known hyperspectral datasets and on HSIs acquired by PRISMA hyperspectral satellite, a medium-resolution hyperspectral imaging mission developed, owned, and operated by the Italian Space Agency from 2020 [6]. PRISMA (Hyperspectral PRecursor of the Application Mission) represents a cutting-edge Earth observation system, it has 239 bands at a spatial resolution of 30m and a spectral resolution better than

G.S. is funded by a PhD fellowship within the framework of the Italian "D.M. n. 352, April 9, 2022" - under the National Recovery and Resilience Plan, Msn. 4, Comp. 2, Inv. 3.3 - PhD Project "Low-rank models for the analysis of Earth Observation data focusing on coastal and marine environments", co-supported by "Planetek Italia S.r.l." (CUP H91I22000410007).

F.E. is supported by ERC Seeds Uniba project "Biomes Data Integration with Low-Rank Models" (CUP H93C23000720001).

12 nm in a spectral range of 400-2505 nm (VNIR and SWIR regions). It can distinguish both the geometric characteristics of observed objects and the chemical-physical composition of the surface of the Earth thanks to its electro-optical equipment that analyzes the spectral fingerprint of each material.

## 2. LMM, DEEP NMF AND AUTOENCODERS

A generic HSI with $B$ bands, $r$ rows, and $c$ columns, can be represented as a matrix $X \in \mathbb{R}_+^{B\times rc}$, through mode-3 unfolding operation. With this representation, a generic row vector encodes the ground information in the corresponding band, while a column represents the spectrum of the selected pixel.

The LMM assumes that $X$ is an approximated linear combination of $K$ endmembers and their corresponding abundances, that is $X \approx EA$, where $E = [e_1, e_2, \ldots, e_K] \in \mathbb{R}_+^{B\times K}$ denotes the endmember matrix, with $K$ endmembers vectors as columns, $A \in \mathbb{R}^{K\times rc}$ is the abundance matrix with abundances fractions for each pixel as columns. Generally, this formulation must satisfy two physically-inspired constraints, the abundance nonnegativity constraint ($a_{ij} \geq 0$), and the abundance sum-to-one constraint that is $\mathbb{1}_K^\top A = \mathbb{1}_{rc}$, where $\mathbb{1}$ denotes all-one column vectors with conformal dimension.

**Deep NMF.** In the context of blind HU, NMF furnishes the approximation of LMM, $X \approx EA$, adding the abundance nonnegativity constraint. It is usually obtained by minimizing a Frobenius-based cost function penalized by additional constraints on the endmembers and abundance factors, demonstrating in this way improved performance when unmixing [5]. However, NMF can be viewed as a single-layer learning process that missed to learn more complex and hierarchical latent feature representations from the nonnegative input data.

DNMF models represent a recent extension of NMF for blind HU tasks, that integrate the multi-layer architecture enabling feature extraction within the matrix factorization to obtain more flexible and expressive modeling of the underlying patterns in $X$. The DNMF interpretability power – due to linearity – allows us to understand how the features are combined and catch the propagation of the information from the first, more abstract layer, to the last, more refined layer. This approach provides numerous interpretations of the semantics hidden in the data.In the first layer, DNMF factorized $X$ into two matrices $E_1$ and $A_1$; then, in the next layer, $A_1$ is further factorized into other two matrices $E_2$ and $A_2$. This process continues until the factorization of the $L$ layers is done. In this way, $X \approx E_1E_2 \ldots E_LA_L$ with $E_l \in \mathbb{R}_+^{d_{l-1}\times d_l}$ for $l = 1, \ldots, L$, $d_0 = B$ and $A_L \in \mathbb{R}_+^{d_L\times rc}$. The latent factors are computed solving a regularized optimization problem:

$$\min_{E_l, A_l \geq 0} \|X - E_1E_2 \ldots E_LA_L\|_F^2 + \textstyle\sum_l \alpha_l \mathcal{R}_l(E_l) + \beta_l \mathcal{R}_l(A_l)$$

where $\mathcal{R}_l$ are regularization terms that enforce the endmembers and abundance factors to possess additional properties besides nonnegativity (i.e. sparseness, orthogonality, symmetry, convexity, etc.) [2, 3].

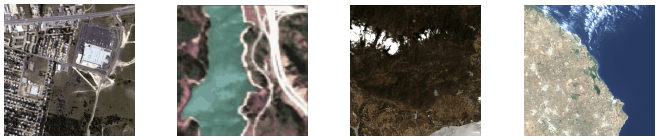
**Autoencoders.** Recent artificial neural network literature proposes the use of autoencoder models to solve the blind HU and estimate abundances that respect the nonnegativity and sum-to-one constraints. AE is a feed-forward neural network trained to replicate its input by learning the identity function. It has two parts: an encoder $\mathcal{G}_E$: $\mathbb{R}^{B\times 1} \to \mathbb{R}^{K\times 1}$, that encodes the spectra input of the $j$-th pixel, $x_j$ for $j = 1, \ldots, rs$, to a hidden representation, and a decoder, $\mathcal{G}_D$: $\mathbb{R}^{K\times 1} \to \mathbb{R}^{B\times 1}$, that reconstructs an approximation of input data, from the latent feature provided by the encoder. The final layer of the encoder has a small number of units to transform spectral vectors of length $B$ into latent space of dimension $K$. Subsequently, this latent representation is used by the decoder to reconstruct the input spectrum as accurately as possible. Generally, AE is trained solving a constrained minimization task on some discrepancy measure $\mathcal{D}$ between the inputs and the approximations provided by the model, i.e., $\mathcal{D}(x_j, \mathcal{G}_D(\mathcal{G}_E(x_j)))$, often regularized with additional terms taking into account additional property regarding the network parameters and the latent representation. AE-based linear unmixing frameworks assume that the endmembers are part of the set of decoder parameters, and the latent representation directly provides an estimate of the abundance matrix $A$.

Both DNMF and AE were inspired by similar ideas, but only a few instances of these two groups of unsupervised HU learning models have been compared in the literature. The following section provides some initial results from a comprehensive comparison between particular DNMF and AE models that impose the same additional constraints on $A$ and $E$ using benchmark datasets and PRISMA images.

## 3. EXPERIMENTS

This section presents experiments conducted to evaluate the performance of selected AE and DNMF methods. In particular, the used AE are: a deep encoder with no regularization (DAEU), a deep encoder with orthogonal sparse prior (OSPAEU), a fully CNN encoder and decoder (CNNAEU), and a multitask deep encoder with shared decoder (MTAEU). All these AE are spectral–spatial methods with no regularizations [4]. DNMF models used are multi-layer NMF (ML-NMF), imposing sparsity constraint on both endmembers and abundances matrices in each layer via $\ell_{1/2}$-norm [7], and orthogonal deep NMF (ODNMF) forcing orthogonality on abundances [8].

The AEs were trained on the TensorFlow framework using Python 3.10, while the NMF-based methods were run in MATLAB R2023b. For a fair comparison, all the experiments are conducted on the same workstation with an AMD Ryzen



(a) (b) (c) (d)

**Fig. 1**: RGB images of the datasets used in experiments (a) Urban, (b) Jasper Ridge, (d) Alimini, (c) Limassol Fire.

7 CPU, 64-GB memory, and without external GPU.
Experiments were performed on two benchmark datasets [9]: **(a)** 307 $\times$ 307 Urban dataset, **(b)** a subimage of 100 $\times$ 100 pixels of Jasper Ridge dataset with 162 and 198 selected bands respectively and two PRISMA images:

- **(c)** Limassol Fire: image recorded on the Limassol fires occurred in July 2021 in Cyprus, with 230 spectral bands (after corrupted and noisy band removal) and spatial dimension of 1000 $\times$ 1000.
- **(d)** Alimini: image recorded on Alimini Lakes, located north of Otranto, a natural area of Salento and Apulia, Italy. The entire image 1000 $\times$ 1000 with 230 spectral bands was used.

It is important to note that the estimation of the number of endmembers $K$ in the scene is an open question; it is known a priori only for the most recent state-of-the-art test datasets used (e.g., Urban, Samson, Cuprite, Jasper Bridge), and this essential but elusive information is not always available in the case of real data making the comparisons of AE and DNMF models a more challenging issue. For the benchmark datasets the accepted ground truth is available (6 endmembers: Asphalt, Grass, Tree, Roof, Metal, and Soil and 4 endmembers: Tree, Water, Soil, and Road for datasets for dataset **(a)** and **(b)**, respectively). While for PRISMA images, a priori identification of the number of endmembers $K$ and their reference spectra has been provided by domain experts trying to find pure pixels with the help of very high-resolution satellite data (4 constituent materials: Water, Clouds, Vegetation, and Soil/Burned area, for datasets **(c)** and **(d)**, respectively). Figure 1 illustrates the four used images.

Unmixing performances are evaluated using the average spectral angle distance (mSAD) that measures the angle between spectra in their signal space in radians, and is computed as $\mathrm{SAD}(\mathbf{e}, \hat{\mathbf{e}}) = \cos^{-1}\left(\frac{\mathbf{e}\cdot\hat{\mathbf{e}}}{\|\mathbf{e}\|_2\|\hat{\mathbf{e}}\|_2}\right)$. Smaller values indicate better performances. For Urban and Jasper Ridge datasets, approximations provided by the deep algorithms have been compared with ground truth endmembers and abundances. For reliable assessments, numerical experiments were repeated twenty-five times, and the mean and standard deviation of the results are reported in Table 1, 2, and 3. Computation time (in seconds for a single run) is also reported for a complete comparison between all tested methods in the last column of the tables.

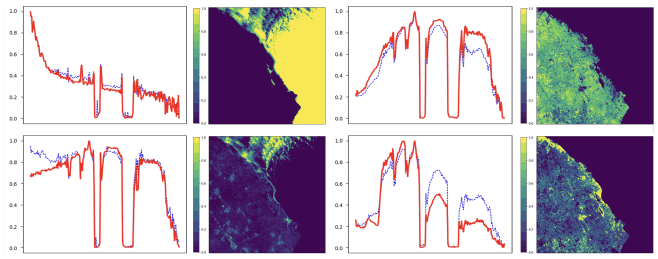


**Fig. 2**: Endmembers (blue dash line) and abundance maps extracted by OSPAEU for Alimini dataset. Red solid lines represent reference endmembers.

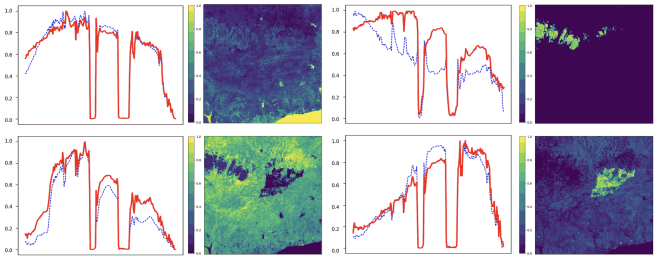


**Fig. 3**: Endmembers (blue dash line) and abundance maps extracted by CNNAEU for Limassol Fire dataset. Red solid lines represent reference endmembers.

It should be observed that for the two benchmark datasets, ODNMF presents the better performance in extracting some particular endmembers (i.e., Roof, Metal, and Soil) and the lowest computational time. High mSAD values for Asphalt in Urban and Water in Jasper Bridge greatly affect the overall average performance of DNMF algorithms.
Similar considerations can be made for PRISMA datasets. According to the time comparison of the analyzed methods, it can be noted that AE-based methods (i.e., OSPAEU, DAEU e MTAEU) seem to be independent by the size of the considered input data, whereas DNMF-based methods appear to suffer from this issue. For PRISMA images, the results are also qualitative illustrated as reported in Figures 2 and 3, depicting the endmembers and abundance maps obtained by OSPAEU and CNNAEU, respectively. It should further point out that the number of endmembers defined a priori for PRISMA images could be underestimated considering the extent of the area observed using the entire image. In particular, for Limassol Fire dataset, both AEs and DNMF often generate one endmember with high abundance values for both water and cloud pixels during repeated runs.

## 4. CONCLUSIONS AND PERSPECTIVES

An initial comparison of some AE and DeepNMF-based models in the blind HU task of real-world benchmark datasets and PRISMA images is presented in this study. These preliminary results prove the usefulness of both approaches characterized by the interpretability capability offered by linear-

**Table 1**: Average SAD from references endmembers in radians along with the standard deviation for all methods for the Urban dataset with six reference endmembers. Computing times are reported in the last column.

| Method | Asphalt | Grass | Tree | Roof | Metal | Soil | Avg | Time(s) |
|---|---|---|---|---|---|---|---|---|
| OSPAEU | 0.524±0.268 | 0.102±0.014 | 0.117±0.070 | 0.253±0.020 | 0.452±0.246 | 0.200±0.044 | 0.275±0.066 | 30 |
| DAEU | 0.055±0.016 | 0.083±0.082 | 0.122±0.028 | 0.289±0.094 | 0.260±0.252 | 0.210±0.123 | 0.170±0.048 | 10 |
| MTAEU | 0.055±0.019 | 0.047±0.018 | 0.145±0.013 | 0.114±0.036 | 0.396±0.122 | 0.062±0.029 | 0.136±0.028 | 12 |
| CNNAEU | **0.042±0.007** | **0.035±0.011** | 0.135±0.007 | 0.037±0.003 | 0.393±0.028 | 0.090±0.012 | 0.122±0.005 | 440 |
| MLNMF | 0.205±0.050 | 0.116±0.054 | 0.214±0.048 | 0.189±0.116 | 0.366±0.175 | 0.983±0.081 | 0.345±0.254 | 44 |
| ODNMF | 0.849±0.125 | 0.191±0.026 | **0.108±0.027** | **0.059±0.056** | **0.235±0.112** | **0.025±0.017** | 0.244±0.121 | 6 |

**Table 2**: Average SAD from references endmembers in radians along with the standard deviation for all methods for the Jasper Bridge dataset with four reference endmembers. Computing times are reported in the last column.

| Method | Tree | Water | Soil | Road | Avg | Times(s) |
|---|---|---|---|---|---|---|
| OSPAEU | 0.067±0.023 | 0.236±0.030 | 0.090±0.036 | 0.269±0.200 | 0.166±0.050 | 17 |
| DAEU | 0.046±0.021 | **0.046±0.018** | 0.090±0.087 | 0.186±0.088 | 0.092±0.033 | 10 |
| MTAEU | 0.051±0.014 | 0.064±0.009 | 0.117±0.053 | 0.393±0.220 | 0.158±0.045 | 12 |
| CNNAEU | 0.089±0.022 | 0.059±0.007 | 0.098±0.037 | 0.459±0.257 | 0.176±0.056 | 427 |
| MLNMF | 0.163±0.024 | 0.157±0.048 | 0.168±0.084 | 0.165±0.121 | 0.163±0.069 | 6 |
| ODNMF | **0.021±0.012** | 0.886±0.117 | **0.036±0.017** | **0.055±0.092** | 0.249±0.087 | 2 |

**Table 3**: Average SAD from references endmembers in radians along with the standard deviation for all methods for the Alimini and Limassol Fire dataset with four reference endmembers. Computing times are reported in last column.

| Alimini dataset | | | | | | |
|---|---|---|---|---|---|---|
| Method | Water | Soil | Clouds | Vegetation | Avg | Time(s) |
| OSPAEU | 0.101±0.006 | 0.091±0.008 | **0.110±0.044** | 0.206±0.058 | 0.127±0.021 | 22 |
| DAEU | 0.091±0.019 | 0.076±0.042 | 0.282±0.078 | 0.201±0.185 | 0.163±0.056 | 10 |
| MTAEU | 0.091±0.003 | **0.057±0.020** | 0.279±0.125 | 0.156±0.032 | 0.146±0.036 | 12 |
| CNNAEU | **0.084±0.002** | 0.070±0.004 | 0.278±0.044 | 0.565±0.163 | 0.249±0.036 | 1030 |
| MLNMF | 0.209±0.041 | 0.063±0.024 | 0.743±0.029 | **0.103±0.018** | 0.283±0.013 | 1500 |
| ODNMF | 0.733±0.078 | 0.099±0.011 | 0.184±0.063 | 0.266±0.096 | 0.320±0.044 | 146 |
| Limassol Fire dataset | | | | | | |
| Method | Water | Clouds | Vegetation | Burned area | Avg | Time(s) |
| OSPAEU | 0.115±0.071 | 0.383±0.082 | 0.116±0.012 | 0.269±0.240 | 0.221±0.070 | 22 |
| DAEU | 0.183±0.177 | 0.361±0.180 | 0.153±0.025 | 0.145±0.056 | 0.211±0.019 | 10 |
| MTAEU | **0.091±0.011** | 0.498±0.043 | 0.162±0.013 | **0.078±0.023** | 0.206±0.011 | 12 |
| CNNAEU | 0.139±0.125 | 0.364±0.146 | 0.142±0.002 | 0.159±0.027 | 0.201±0.013 | 780 |
| MLNMF | 0.531±0.126 | 0.488±0.099 | 0.140±0.067 | 0.244±0.098 | 0.349±0.027 | 1367 |
| ODNMF | 0.246±0.027 | **0.085±0.033** | **0.082±0.047** | 0.487±0.076 | 0.225±0.064 | 119 |

ity and the capability of determining a lower-dimensional and hierarchical representation of the nonnegative latent features. The methods demonstrated quite similar performances except for the computational time of AE which is faster than DNMF. Further research can be carried out in the future, building on current work to solve some well-known open issues such as the simultaneous use of spectral and spatial information, initialization, and hyperparameters fine-tuning. In particular, we plan to focus on the estimation of number of endmembers $K$ using very large HSIs captured by PRISMA-like sensors. This can be useful for future PRISMA applications in environmental monitoring with particular emphasis to plastic litter detection and to the assessment of wildfire-affected and burned areas.

## 5. ACKNOWLEDGEMENTS

G.S., N.D.B., F.E. are members of the Gruppo Nazionale Calcolo Scientifico - Istituto Nazionale di Alta Matematica (GNCS-INdAM). The PRISMA images were kindly provided by Planetek Italia S.r.l.

## 6. REFERENCES


[1] Bioucas-Dias et al., "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[2] Wen-Sheng Chen, Qianwen Zeng, and Binbin Pan, "A survey of deep nonnegative matrix factorization," *Neurocomp*, vol. 491, pp. 305–320, 2022.

[3] Valentin Leplat, Le Thi Khanh Hien, Akwum Onwunta, and Nicolas Gillis, "Deep nonnegative matrix factorization with beta divergences," 2023.

[4] Burkni Palsson, Johannes R. Sveinsson, and Magnus O. Ulfarsson, "Blind hyperspectral unmixing using autoencoders: A critical comparison," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 15, pp. 1340–1372, 2022.

[5] Xin-Ru Feng, Hengchao Li, Rui Wang, Qian Du, Xiuping Jia, and Antonio J. Plaza, "Hyperspectral unmixing based on nonnegative matrix factorization: A comprehensive review," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 15, pp. 4414–4436, 2022.

[6] R. Loizzo et al., "Prisma: The italian hyperspectral mission," in *IGARSS 2018 - 2018 IEEE Int Geoscience and Remote Sens Symp*. July 2018, IEEE.

[7] Roozbeh Rajabi and Hassan Ghassemian, "Spectral unmixing of hyperspectral imagery using multilayer nmf," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 1, pp. 38–42, Jan. 2015.

[8] Pierre De Handschutter and Nicolas Gillis, "Deep orthogonal matrix factorization as a hierarchical clustering technique," in *2021 29th European Signal Processing Conference (EUSIPCO)*. Aug. 2021, IEEE.

[9] Vytautas Paura and Virginijus Marcinkevičius, "Benchmark for hyperspectral unmixing algorithm evaluation," *Informatica*, p. 285–315, 2023.