INDUSTRIAL AND SYSTEMS ENGINEERING

ISE

Harnessing Nonnegative Matrix Factorization for Computational Materials Modeling

THAKSHEEL N. ALLECK[1], AKWUM ONWUNTA[1], AND CHINEDU E. EKUMA[2]

[1]Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, USA

[2]Department of Physics, Lehigh University, Bethlehem, PA 18015, USA

ISE Technical Report 26T-023

LEHIGH UNIVERSITY

Harnessing Nonnegative Matrix Factorization for Computational Materials Modeling

Thaksheel N. Alleck[1] Akwum Onwunta[1] Chinedu E. Ekuma[2]

[1] Department of Industrial and Systems Engineering,
Lehigh University, Bethlehem, PA 18015, USA

[2] Department of Physics,
Lehigh University, Bethlehem, PA 18015, USA

Abstract

Interpretable representation learning is becoming an essential requirement for materials modeling and discovery, where accurate predictions alone are often insufficient for guiding scientific understanding or experimental design. Nonnegative Matrix Factorization (NMF) offers a natural framework for this purpose because many materials descriptors, including elemental compositions, spectral intensities, structural counts, and engineered feature groups, are nonnegative and can be meaningfully decomposed into additive latent factors. In this work, we develop and evaluate `nimd`, a computational framework for Nonnegative Interpretable Matrix Decomposition in materials informatics. The framework brings together standard NMF, Frobenius- and beta-divergence formulations, hierarchical and multilayer NMF, deep NMF, min-volume regularized deep NMF, and semi-supervised NMF variants within a unified workflow for prediction, comparison, and interpretation. Using a high-dimensional Heuslerene monolayer dataset as a representative materials discovery test case, we compare NMF-based representations with PCA across regression and classification tasks. The results show that NMF models can achieve predictive performance comparable to PCA while providing substantially more interpretable latent factors through nonnegative, parts-based decompositions. Deep and multilayer NMF further expose hierarchical structure in the descriptor space, while semi-supervised NMF provides a route for aligning latent representations with target properties. Beyond this demonstration, the framework is designed to support broader materials datasets where interpretability, factor stability, and chemically or physically meaningful feature attribution are central to model usefulness. These results establish NMF-based representation learning as a transparent and scientifically grounded alternative to conventional dimensionality reduction for materials modeling and discovery. The code and data are available through `https://github.com/thaksheel/nimd`.

1 Introduction

Modern materials modeling increasingly depends on high-dimensional descriptors that encode composition, structure, spectra, electronic response, and other experimentally or computationally derived quantities. These representations are powerful, but they also create a central challenge for materials discovery: the models built from them must be not only predictive, but also interpretable enough to expose the structural and physical features that drive a prediction. In many scientific settings, a low-dimensional representation is useful only if it helps answer mechanistic questions, identify meaningful feature groups, reveal stable latent motifs, or guide subsequent design. This need makes interpretable matrix factorization especially important for materials informatics.

Matrix factorization methods have long provided a foundation for extracting low-rank structure from complex data, with applications spanning clustering [15], machine learning [56], and

large-scale data analysis [58, 50]. Among these approaches, Nonnegative Matrix Factorization (NMF) is particularly well suited to scientific datasets because many materials descriptors are naturally nonnegative. Elemental fractions, spectral intensities, diffraction counts, spatial densities, and engineered chemical descriptors often represent additive physical quantities. By constraining both the basis and coefficient matrices to be nonnegative, NMF produces part-based decompositions in which latent factors can often be interpreted as physically meaningful mixtures, motifs, phases, or feature groups. This distinguishes NMF from conventional signed decompositions and gives it practical advantages for interpretability, storage, computational efficiency, and analysis of large scientific datasets [18, 47, 26, 37, 57].

Dimensionality reduction plays a central role in materials characterization and informatics, where high-dimensional spectral, structural, and compositional data must be transformed into compact representations. Principal Component Analysis (PCA) [38, 55, 32, 10] remains a widely used baseline because it is simple, efficient, and analytically convenient. However, PCA components contain positive and negative coefficients, and their orthogonality constraint can mix physically unrelated contributions. In measurements such as electron counting, spectroscopy, diffraction, and imaging, negative intensities or signed mixtures may be mathematically useful but physically difficult to interpret. NMF addresses this limitation by enforcing additive, nonnegative components that better align with the structure of many materials datasets [59, 44, 51].

In addition to PCA, Uniform Manifold Approximation and Projection (UMAP) [7] is widely used in materials informatics, so we include it as a second baseline. UMAP is a nonlinear manifold learning method valued for capturing structure that linear methods miss, balancing local and global data structure, scaling efficiently to large datasets, and producing low-dimensional embeddings that reveal clusters and property trends. It has been applied to visualize composition–property relationships across alloy design spaces [9] and to analyze learned structural embeddings of crystalline materials [8]. Like PCA, however, its coordinates lack an additive, parts-based interpretation tied to elemental features, which motivates the nonnegative approach developed here.

The value of NMF in materials science has been demonstrated across several characterization problems. Modified alternating least-squares algorithms [60] have been used to map phase evolution in degrading Li-ion battery cathodes [40, 29, 43, 42, 24] and to identify nitrogen chemical states in N-doped TiO_2 [62, 49]. NMF has also enabled the extraction of atom-site-specific core-loss spectra from EELS datasets influenced by diffraction and electron channeling [53, 61, 54]. In these studies, the nonnegative basis spectra and spatial intensity maps produced components consistent with first-principles calculations and chemically meaningful assignments [40, 43, 29, 42, 24, 53, 61]. Beyond spectroscopy, NMF has been applied to crystallographic signal separation, in situ diffraction, time-resolved diffraction, spatially resolved electron diffraction, nanosheets, multimodal measurements, and metal-organic frameworks [52, 16, 4, 46, 17, 45, 32, 60].

Recent work has extended NMF to settings where partial supervision and multiple data sources are available. Li et al. [30] proposed a semi-supervised non-negative matrix trifactorization in which the similarity graph is learned adaptively from the data and then combined with a block-diagonal structure built from a small set of labeled samples, so label information shapes the graph used during factorization with only one regularization parameter. Extending this idea to multi-view data, Yuan et al. [31] introduced UJS2NMF, which embeds soft label constraints in a relational matrix and hard label constraints through a factorized label-constraint matrix, eliminating the balancing parameters required by earlier multi-view methods. Both studies show that embedding prior knowledge directly into the factorization structure, rather than adding separate regularization terms, can improve clustering on benchmark image datasets. Both methods, however, remain sensitive to graph construction choices such as neighborhood size, and both were evaluated on clustering tasks rather than supervised prediction. These developments motivate extending NMF-based representations to supervised

materials modeling, where labeled property data such as stability labels are often available.

Despite this strong foundation, NMF remains underused as a general framework for modern materials informatics. Much of its use has focused on individual characterization tasks, while less attention has been given to systematic workflows that connect factorization, supervised prediction, hierarchical representation learning, attribution, stability, and identifiability. Recent developments in multilayer and deep NMF create an opportunity to move beyond flat low-rank decompositions toward hierarchical representations that can expose structure across multiple levels of abstraction [56, 21]. For materials modeling, this is especially important because chemical information is often organized hierarchically: individual descriptors combine into elemental motifs, motifs combine into compositional families, and these families influence target properties through nonlinear or multiscale relationships.

In this work, we develop and evaluate an interpretable NMF-based framework for materials modeling and discovery. The goal is not to claim that NMF replaces state-of-the-art predictive architectures, but to establish how NMF and its modern extensions can provide scientifically meaningful representations while retaining competitive predictive performance. We examine standard NMF, Frobenius- and beta-divergence formulations, hierarchical NMF, multilayer NMF, deep NMF, min-volume regularized deep NMF, and semi-supervised NMF within a unified workflow for regression, classification, and interpretation. The implementation is released as `nimd`, a package for Nonnegative Interpretable Matrix Decomposition. Using a high-dimensional Heuslerene monolayer dataset as a representative materials discovery test case, we compare NMF-based representations with PCA across supervised classification and regression tasks. The Heuslerene system provides a useful benchmark because it contains composition-derived and electronic descriptors with chemically meaningful structure. However, the broader aim is to evaluate whether nonnegative latent factors can support prediction while providing more transparent explanations than signed low-dimensional embeddings. Our results show that standard and deep NMF models match or improve upon PCA in the studied prediction tasks while producing latent factors that are easier to connect to elemental groups, compositional motifs, and structurally interpretable feature patterns. We further show that deep and multilayer NMF reveal hierarchical organization in the descriptor space, enabling relationships between low-level factors and higher-level representations to be analyzed through layer mapping and factor attribution. Min-volume regularized deep NMF improves stability and interpretability under noisy or perturbed initializations, which is important for scientific workflows where reproducibility and physical meaning are as valuable as raw predictive accuracy. Semi-supervised NMF provides an additional mechanism for aligning latent factors with target properties, linking representation learning more directly to supervised materials modeling. Together, these results position NMF-based representation learning as more than a transparent alternative to PCA. They establish it as a practical and scientifically grounded framework for interpretable materials informatics, capable of supporting prediction, revealing chemically meaningful structure, analyzing hierarchical feature organization, and improving trust in data-driven materials discovery.

2 Background and Related Work

2.1 Standard NMF

NMF is one of the foundational techniques for extracting interpretable, parts-based representations from high-dimensional scientific data. Extensive methodological advances have been documented in the literature, including detailed algorithmic reports on NMF optimization methods [34, 3], a seminal textbook by the original pioneers of the method [12], a comprehensive chemometrics-oriented review [48], and a modern treatment of matrix factorization techniques [19]. In its basic form, NMF approximates an experimental data matrix $X \in \mathbb{R}_{+}^{m \times n}$ as the

product of two lower-rank, nonnegative matrices W and H:

$$X \approx WH, \tag{1}$$

where $W \in \mathbb{R}^{m \times r}$ contains basis patterns and $H \in \mathbb{R}^{r \times n}$ contains their corresponding nonnegative coefficients with m data samples, n features, and r representing the factorization rank. In many classical NMF applications, such as object recognition [34] or text mining, the ordering of the columns of X is irrelevant. In scientific measurements, however, the column structure of X carries physical meaning: columns may correspond to spatial positions (e.g., hyperspectral imaging) or temporal evolution (e.g., acoustic or spectroscopic time-series). Although X can be transposed without changing the mathematical formulation, in this study the rows and columns of X represent reciprocal-space (spectral) and real-space coordinates, respectively, consistent with conventions in several analyses such as 4D-STEM [28].

To obtain W and H, NMF solves an optimization problem that minimizes a reconstruction cost function D, typically the Frobenius-norm error:

$$\min_{W,\, H \geq 0} D(X, WH). \tag{2}$$

This is optimized iteratively. Two major algorithmic families are widely used, namely multiplicative update (MU) and alternative least squares (ALS) [12]. The MU algorithm, derived from a majorization-minimization framework [34], updates W and H using elementwise multiplication and division:

$$W \leftarrow W \odot \frac{XH^T}{WHH^T}, \tag{3}$$

$$H \leftarrow H \odot \frac{W^T X}{W^T WH}, \tag{4}$$

where $\odot$ denotes elementwise operations. The ALS algorithm alternates between solving two constrained least-squares problems:

$$H \leftarrow \min_{H \geq 0} ||X - WH||_F^2, \quad W \leftarrow \min_{W \geq 0} ||X - WH||_F^2. \tag{5}$$

In practice, each step is computed by an unconstrained least-squares followed by projection onto the nonnegative orthant:

$$H \leftarrow [W^\dagger X]_+, \quad W \leftarrow [XH^\dagger]_+, \tag{6}$$

where $W^\dagger$ and $H^\dagger$ denotes the Moore-Penrose pseudoinverse and $[\cdot]_+$ applies elementwise truncation to enforce nonnegativity.

Both MU and ALS are implemented in widely used scientific software, such as MATLAB and scikit-learn. As noted in Gillis [19], ALS is not mathematically rigorous due to the projection step, but it remains effective in practice. Primitive NMF algorithms often produce sparse components [25], which may not always reflect physically meaningful structure. When the true underlying signals are not sparse, such as continuous baselines with superimposed peaks, standard NMF can introduce artifacts, including downward convex distortions or unnatural intensity drops [51, 41, 27]. Moreover, primitive NMF lacks mechanisms for frequency domain discrimination and therefore cannot suppress high-frequency noise, sometimes yielding components that are physically implausible. These limitations motivate the development of constrained or regularized NMF variants, discussed in the following sections, which incorporate domain knowledge to better capture the resolution, smoothness, and intensity-profile characteristics of scientific measurements.

This study employs a broad family of NMF methods, each designed to capture different aspects of compositional structure, supervision, and hierarchical organization. We use

β-divergence NMF to model non-Gaussian noise and heterogeneous feature scales [35]. A generalized objective that models β-divergence NMF is given by:

$$\min_{W,H\geq 0} D_\beta(X||WH), \tag{7}$$

where

$$D_\beta(x\|y) = \begin{cases} \frac{1}{\beta(\beta-1)}\left(x^\beta + (\beta-1)y^\beta - \beta x y^{\beta-1}\right), & \beta \neq 0,1, \\ x\log\left(\frac{x}{y}\right) - x + y, & \beta = 1, \\ \frac{x}{y} - \log\left(\frac{x}{y}\right) - 1, & \beta = 0. \end{cases} \tag{8}$$

This includes Frobenius ($\beta = 2$), KL ($\beta = 1$), and IS ($\beta = 0$) as special cases. Also, we include Frobenius-norm NMF, which we evaluate using several widely used solvers such as MU, ADMM, ALSH, and FPGM [19], to assess stability and solver-dependent variability. The Frobenius-norm NMF objective is given by:

$$\min_{W,H\geq 0} ||X - WH||_F^2. \tag{9}$$

To incorporate label information directly into the factorization, we employ semi-supervised NMF (SSNMF), which jointly factorizes the data and label matrices so that the learned representations align more closely with supervised targets [22]:

$$\min_{W,H\geq 0} ||X - WH||_F^2 + \lambda||Y - UH||_F^2, \tag{10}$$

where Y is the label matrix, U is the supervised mapping from latent factors to labels, and λ controls the strength of supervision. This encourages the latent factors to align with the predictive structure. To capture hierarchical structure, we use hierarchical rank-2 NMF [20], a recursive method that repeatedly splits the data into rank-2 subproblems, producing coarse-to-fine motifs that resemble a binary tree of chemical patterns. Each split solves a rank-2 NMF subproblem:

$$\min_{W_2,H_2\geq 0} ||X - W_2H_2||_F^2, \tag{11}$$

and then recursively applies the same objective to one of the two branches. The global model is a binary tree of rank-2 factorizations.

2.2 Deep NMF

Deep NMF generalizes classical NMF by decomposing the data matrix through a sequence of nonnegative transformations, enabling the model to learn features at progressively higher levels of abstraction [36]. Instead of a single factorization $X \approx WH$, deep NMF constructs a hierarchy of layers:

$$X \approx W_1H_1, \quad W_1 \approx W_2H_2 \quad ..., \quad W_{L-1} \approx W_LH_L, \tag{12}$$

which together yield the multi-scale representation

$$X \approx W_1H_1 \approx W_2H_2H_1 \approx \quad ... \quad \approx W_LH_L...H_1. \tag{13}$$

Each layer W_ℓ captures increasingly abstract motifs where early layers isolate fine-scale elemental patterns, while deeper layers merge related motifs into broader chemical families. This hierarchical structure is optimized using either a data-centric loss function or a layer-centric loss function, which minimizes the reconstruction error at each layer. On the other hand, Deep NMF uses a layer-centric loss function, which is given by:

$$L_{LCLF} = \sum_{\ell=1}^{L} \lambda_\ell\, D_\beta(W_{\ell-1},\, W_\ell H_\ell) \quad \text{s.t.} \quad H_\ell\, e = e \text{ for } \ell = 1, 2, ..., L, \tag{14}$$

where D_β is a β-divergence [14], e denotes the vector of all ones of appropriate dimensions, $H_\ell e = e$ is the normalized constraint on H_ℓ, and λ_ℓ's are positive penalty parameters. This formulation improves consistency across layers and enhances identifiability, making deep NMF particularly well-suited for materials informatics, where chemical structure is inherently hierarchical.

Min-volume deep NMF augments the standard deep NMF formulation with a geometric constraint that encourages the basis vectors at each layer to form a minimum-volume simplex. Leplat et al. [36] achieved this by adding a log determinant penalty:

$$\min_{\{W_\ell,\, H_\ell \geq 0\}_{\ell=1}^{L}} \sum_{\ell=1}^{L} \lambda_\ell \, D_\beta(W_{\ell-1},\, W_\ell H_\ell) \;+\; \alpha_\ell \log\det\!\big(W_\ell^T W_\ell + \delta I\big) \quad \text{s.t.} \quad W_\ell^T \, e = e, \tag{15}$$

where $W_0 = X$, W_ℓ has r_ℓ columns with $r_0 = n$, δ is a small positive scalar it prevents the log det from going to $-\infty$, and α_ℓ's are positive penalty parameters. The min-volume term biases the solution toward identifiable factorizations, improves robustness to noise, and encourages sparsity in the coefficient matrices H_ℓ. In materials informatics, these properties translate directly into more interpretable chemical motifs: columns of W_ℓ align more closely with real composition patterns, and deeper layers remain coherent rather than collapsing into dense mixtures. This makes min-volume deep NMF particularly effective for scientific applications where reproducibility and physical meaning are essential.

A final variant in the NMF family is multilayer NMF, which constructs a hierarchy of representations using a data-centric loss function. In this formulation, each layer attempts to reconstruct the original data matrix rather than only the output of the previous layer. The objective is:

$$L_{DCLF} = \sum_{\ell=1}^{L} \lambda_\ell \, D(X, W_\ell \prod_{i=0}^{\ell-1} H_{\ell-i}), \tag{16}$$

where $D(A, B)$ is the measure of distance between A and B, and the λ_ℓ's are positive penalty parameters. This loss function minimizes a weighted sum of the errors in the decompositions of X at each of the L layers, encouraging each layer to preserve information from the original data while progressively refining the learned representations.

3 Methods

This study integrates NMF-based models, supervised learning workflows, and interpretability analyses to evaluate how nonnegative factorizations capture chemically meaningful structure in high-dimensional materials data. The full computational framework, including data preprocessing, model training, hyperparameter selection, supervised evaluation, and interpretability analysis, is implemented in the reproducible `nimd` package.

3.1 Materials Dataset

The experiments are conducted on a curated dataset of Heuslerenes, a recently proposed class of atomically thin X_2YZ intermetallic monolayers generated through a high-throughput first-principles workflow [23]. The dataset originates from a systematic exfoliation study in which bulk Heusler alloys were cleaved into monolayers using a symmetry-guided algorithm designed to preserve the X_2YZ stoichiometry. The resulting structures were relaxed and characterized using density functional theory (DFT) within the PBE-GGA framework, with mechanical stability evaluated from in-plane elastic constants and Born stability criteria. In total, the dataset contains 864 Heuslerene monolayers spanning diverse chemistries, crystallographic motifs, and magnetic configurations.

Although the dataset is restricted to a single family of two-dimensional materials, it provides a physically grounded instance of materials datasets characterized by high-dimensional, sparse,

nonnegative compositional features together with physically derived electronic descriptors and both categorical and continuous material properties. The objective is therefore not to treat Heuslerenes as representative of the entire materials space, but to use them as a chemically diverse test case for the types of compositional data on which the proposed factorization methods are intended to operate. This study is intended as a proof of concept: it tests whether NMF-based representations yield interpretable compositional structure in this setting before the approach is extended to other material classes. Evaluation on additional datasets, including those whose descriptors are not predominantly nonnegative and additive, is an important direction for future work.

This chemical and structural diversity makes it a useful testbed for evaluating whether NMF-based representations can recover interpretable compositional structure from high-dimensional materials descriptors. Each sample corresponds to a unique monolayer and is associated with two supervised targets. The first is a binary thermodynamic stability label, used for classification, indicating whether the monolayer satisfies the DFT-based mechanical and energetic stability criteria. The second is a continuous lattice constant, used for regression, representing the relaxed in-plane equilibrium geometry.

The feature matrix contains compositional descriptors for more than 60 metallic elements, where each entry represents the fractional contribution of an element to the compound formula. These features are nonnegative and naturally additive, making them well aligned with the mathematical structure of NMF. The dataset also includes nine electronic and quantum-mechanical descriptors: average and fractional s, p, d, and f valence electron counts, together with molecular weight. These additional descriptors encode periodic-table trends and provide information beyond stoichiometry alone. The dataset is partitioned into a fixed 80/20 train–test split that is used consistently across all supervised experiments. For interpretability analyses focused on the global compositional structure of the dataset, NMF models are fit to the full feature matrix so that the learned factors reflect the complete chemical variability present in the Heuslerene space. For supervised evaluation, target leakage is avoided by training the random forest models only on the training split and evaluating performance exclusively on the held-out test split. This fixed protocol provides a common basis for comparing PCA, standard NMF, multilayer NMF, deep NMF, and min-volume deep NMF across both predictive performance and interpretability.

3.2 Supervised Learning Tasks

To evaluate how well NMF-derived representations support downstream prediction, we conducted two supervised learning tasks aligned with common objectives in materials informatics, namely binary classification of thermodynamic stability and regression of lattice constants. For the classification task, we employed logistic regression, a widely used linear classifier that provides a transparent baseline for assessing the separability of materials classes in the latent feature space. Logistic regression is particularly useful in this context because its decision boundary directly reflects how well the learned NMF features encode stability-relevant chemical patterns. For the regression task, we used a random forest regressor to predict lattice constants from the same latent representations. Preliminary experiments with linear regression showed that the relationship between composition-derived features and lattice parameters is strongly nonlinear; therefore, random forests is a better choice since they consistently achieved higher performance across multiple evaluation metrics (including R^2, RMSE, and MAE).

The supervised learning results provide a quantitative anchor for the interpretability work that follows. They allow us to identify which latent factors are most predictive, assess the stability of these factors across layers of deep NMF, and connect mathematical decompositions to chemical intuition. In doing so, the supervised results help bridge the gap between unsupervised feature learning and domain-relevant explanations.

3.3 NMF Initialization Strategies

Initialization is a critical component of NMF because the optimization problem is nonconvex and highly sensitive to the starting values of the factor matrices. Several initialization strategies have been developed to balance convergence speed, reconstruction accuracy, sparsity, and computational cost. The most basic approach is random initialization, which generates nonnegative random entries scaled to the magnitude of the data matrix. Although simple and widely used, random initialization often requires many iterations to converge and can lead to inconsistent solutions across runs. To provide more structured starting points, several SVD-based initialization methods have emerged. The classical NNDSVD method constructs deterministic, nonnegative approximations of the leading singular components of the data matrix, producing sparse and interpretable initial factors that significantly reduce early reconstruction error. Two practical variants, NNDSVDa and NNDSVDar, modify the zero entries produced by NNDSVD. NNDSVDa fills zeros with the global mean of the data to produce denser factors when sparsity is not desired, while NNDSVDar replaces zeros with small random values to accelerate convergence while maintaining moderate density [6].

More recently, NNSVD-LRC has been introduced as an improved SVD-based initialization that incorporates a low-rank correction step [2]. Instead of discarding information from the negative components of the SVD, NNSVD-LRC uses this discarded structure to refine the initial factors through a small number of inexpensive updates on a truncated low-rank approximation of the data. This approach preserves the desirable sparsity of NNDSVD, ensures that initialization error decreases as the target rank increases, and often yields substantially lower initial reconstruction error at negligible additional cost. Together, these initialization strategies provide a spectrum of options, ranging from fast and stochastic to deterministic and structure-aware, allowing practitioners to tailor NMF initialization to the sparsity, interpretability, and performance requirements of their specific application.

3.4 Metrics for Structural and Hierarchical Interpretation

To systematically evaluate the structural interpretability and stability of the factors produced by our NMF models, we employ three complementary metrics, namely Normalized Mutual Information (NMI), the Jaccard overlap matrix, and entropy. Together, these measures capture different aspects of factor behavior, where NMI quantifies global similarity between support patterns, Jaccard measures direct element-level overlap, and entropy characterizes the sharpness or diffuseness of each factor's distribution. Using all three provides a multi-resolution view of how motifs emerge, propagate, and specialize across layers, enabling a rigorous and interpretable analysis of both shallow and deep NMF representations.

Normalized Mutual Information (NMI) measures the global similarity between two clustering membership patterns by quantifying how much knowing one labeling reduces uncertainty about the other [33]. Consider two binary membership sets A and B; NMI can be computed as follows:

$$NIM(A,B) = \frac{2I(A;B)}{H(A)+H(B)}, \tag{17}$$

where the mutual information is

$$I(A;B) = \sum_{a\in\{0,1\}} \sum_{b\in\{0,1\}} p(a,b) \log \frac{p(a,b)}{p(a)p(b)}, \tag{18}$$

with

- $p(a,b)$ the joint probability that an element takes value a in A and b in B,

- $p(a)$ and $p(b)$ are the marginal probabilities, computed from empirical frequencies, for example

$$p(a) = \frac{\{i : A_i = a\}}{n}. \tag{19}$$

Moreover, $H(X)$ represents entropy of an arbitrary binary membership set X, and is given by

$$H(X) = -\sum_{x \in \{0,1\}} p(x) \log p(x). \tag{20}$$

It quantifies the uncertainty in the membership pattern of X.

In the context of NMF interpretation, each factor is converted into a binary mask indicating which elements belong to its top-k support, and NMI evaluates how structurally similar two such masks are across all elements. Unlike overlap-based metrics, NMI is sensitive to the full pattern of zeros and ones, making it particularly effective for detecting inherited structure across layers in deep NMF. High NMI values indicate that two factors share nearly identical support patterns, even if they do not share the same top-k elements, while low values signal independence or divergence in how the factors partition the dataset. This makes NMI a powerful tool for assessing hierarchical consistency, factor redundancy, and pattern level inheritance across layers.

The Jaccard matrix provides a complementary, fine-grained measure of similarity by quantifying the literal overlap between the top-k elements of two factors.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}. \tag{21}$$

For any pair of factors, the Jaccard index computes the ratio of shared elements to the total number of unique elements in their combined supports [13]. Unlike NMI, which evaluates global structural similarity, the Jaccard index focuses exclusively on direct element-level inheritance. A high Jaccard value indicates that two factors select many of the same elements, while a value of zero means they share none. This makes the Jaccard matrix especially useful for tracing how specific chemical motifs propagate, split, or merge across layers in deep NMF.

Entropy quantifies the concentration or dispersion of a factor's weight distribution, offering insight into how focused or diffuse a component is. For a normalized factor vector h, entropy is computed as

$$H(h) = -\sum_i h_i \log h_i, \tag{22}$$

with low entropy indicating a sharply peaked distribution (a highly specific motif) and high entropy indicating a broad, diffuse mixture [1]. In NMF interpretation, entropy serves as a diagnostic for factor sharpness, specialization, and noise sensitivity. Low entropy factors typically correspond to well-defined chemical motifs or localized compositional groups, while high entropy factors may reflect blended patterns, unresolved structure, or noise-driven artifacts. When applied across layers, entropy helps reveal how deep NMF compresses or refines information, that is, whether deeper layers produce more abstract but still coherent motifs, or whether they collapse into overly diffuse representations. As a result, entropy provides an essential scalar summary of factor quality and interpretability.

4 Results and Discussion

To ensure statistical robustness, we evaluate PCA, UMAP, and all NMF variants across ranks $r \in [20, 70]$, reporting mean and standard deviation computed over this rank range. All experiments use a fixed train–test split and a fixed random seed, so the only source of variation in the reported metrics is the change in rank, not randomness in training or data partitioning.

Each model is tuned individually using a random grid search to find optimized configurations rather than default settings. However, for UMAP, we left the configurations at their default values in the Python implementation. For Deep NMF models in particular, we tuned the layer depth to balance interpretability and predictive accuracy. The full depth selection procedure is described in Section 4.4. Additionally, we report results across five initialization strategies. Our experiments consistently show that random initialization yields the weakest performance on this dataset, underscoring the importance of testing multiple initialization schemes when applying NMF in scientific settings.

All computational experiments were conducted on a Windows desktop equipped with 48 GB of RAM, a 12th Gen Intel® Core™ i7-12700F CPU, and an NVIDIA GeForce RTX 4060 GPU with 8 GB of dedicated memory. However, all evaluations and reported wall-time measurements were performed exclusively on the CPU, without GPU acceleration.

4.1 Lattice Constant Regression

Regression performance is substantially affected by initialization strategies; notably, random initialization is never the best approach for any NMF variant. For five of the six variants, it is lower than all four initialization strategies, for example, beta (0.536 ± 0.044 under random versus 0.666 ± 0.042 under NNDSVD-LRC). The exception is ssnmf, where random initialization (0.438 ± 0.206) exceeds NNDSVDa (0.355 ± 0.172) but remains below the other three strategies. Under random initialization, beta (0.536 ± 0.044) performs best among NMF variants. Under NNDSVD, multilayer (0.672 ± 0.031) leads, followed by fronorm (0.663 ± 0.037) and minvol (0.662 ± 0.040). Under NNDSVDa, beta (0.622 ± 0.055) leads, followed by fronorm (0.601 ± 0.053). Under NNDSVDar, beta (0.659 ± 0.037) leads, followed by multilayer (0.656 ± 0.046) and deep (0.653 ± 0.038). Under NNDSVD-LRC, beta (0.666 ± 0.042) leads, followed by fronorm (0.652 ± 0.035). Each variant's best initialization is plain NNDSVD for fronorm, SSNMF, multilayer, and minvol, NNDSVD-LRC for beta, and NNDSVDar for deep. NNDSVDa is the weakest initialization strategy for every variant. With their best initialization, five of the six NMF variants exceed the PCA baseline (0.599 ± 0.019), while ssnmf does not (best 0.540 ± 0.034). PCA has no explicit initialization, so it is reported under the Random column as a common reference point. Hierarchical NMF (0.699 ± 0.037, no explicit initialization) follows a fixed rank-two recursion with no initialization parameter, so it is not strictly controlled against PCA and is reported separately; however, it has the highest R^2 value across all the NMF variants and initializations. UMAP is the weakest method (0.241 ± 0.059), which is expected for a method designed mainly for visualization and neighborhood preservation rather than supervised feature extraction.

Model	Random	NNDSVD	NNDSVDa	NNDSVDar	NNDSVD-LRC	Hier.[†]
pca	0.599 ± 0.019	–	–	–	–	–
umap	0.241 ± 0.059	–	–	–	–	–
hier	–	–	–	–	–	0.699 ± 0.037
beta	0.536 ± 0.044	0.654 ± 0.031	0.622 ± 0.055	0.659 ± 0.037	0.666 ± 0.042	–
fronorm	0.489 ± 0.143	0.663 ± 0.037	0.601 ± 0.053	0.634 ± 0.074	0.652 ± 0.035	–
ssnmf	0.438 ± 0.206	0.540 ± 0.034	0.355 ± 0.172	0.506 ± 0.283	0.529 ± 0.073	–
multilayer	0.437 ± 0.070	0.672 ± 0.031	0.554 ± 0.035	0.656 ± 0.046	0.637 ± 0.044	–
deep	0.426 ± 0.070	0.636 ± 0.056	0.516 ± 0.055	0.653 ± 0.038	0.626 ± 0.058	–
minvol	0.456 ± 0.075	0.662 ± 0.040	0.579 ± 0.048	0.645 ± 0.046	0.643 ± 0.047	–

Table 1: Mean regression R^2 results using Frobenius norm NMF methods with standard deviations, with ranks ranging from 20 to 70. Each cell reports the mean over ranks for the given initialization. The Frobenius norm NMF methods include the highest score from a range of five different methods.

[†]Hierarchical NMF follows a fixed rank-two recursive procedure and has no initialization parameter, so it is not strictly controlled against PCA and is reported separately. Its result is 0.699 ± 0.037 (no explicit initialization).

4.2 Thermodynamic Stability Classification

The classification results in Table 2 are broadly similar across methods. PCA gives RMSE 0.444 ± 0.012, MAE 0.197 ± 0.008, accuracy $80.3\% \pm 1.2\%$, and F1-macro $68.6\% \pm 2.2\%$, and most NMF variants fall close to these values. UMAP is the weakest method on every metric (RMSE 0.500 ± 0.003, accuracy $75.0\% \pm 0.3\%$, F1-macro $46.1\% \pm 1.9\%$). The clearest RMSE improvement over PCA comes from β-NMF under NNDSVDar (0.436 ± 0.012), with accuracy ($80.9\% \pm 1.0\%$) and F1-macro ($69.7\% \pm 2.0\%$) also close to the PCA values. Frobenius-norm NMF matches PCA only under random initialization (0.443 ± 0.014), and its other initializations give slightly higher RMSE. Deep NMF reaches its best RMSE under NNDSVD-LRC (0.440 ± 0.019), and min-volume deep NMF reaches 0.437 ± 0.021 under the same scheme. Multilayer NMF and hierarchical NMF (0.443 ± 0.013, no explicit initialization) remain close to PCA on all metrics. SSNMF has the largest variance among the NMF variants. Its best RMSE is 0.473 ± 0.016 under NNDSVDar, and its best accuracy is $77.6\% \pm 1.5\%$, below PCA. Min-volume deep NMF under NNDSVD has the highest F1-macro in the table ($79.8\% \pm 2.6\%$), but its RMSE under that scheme is 0.450 ± 0.017, and its accuracy is $79.8\% \pm 1.5\%$. The F1 gain therefore does not carry over to the other metrics.

4.3 Computational Cost

Modern materials datasets are increasingly large and high-dimensional, so computational efficiency matters as much as predictive performance. As NMF-based models scale with data size and factorization rank, their runtime behavior affects whether they can be deployed in materials informatics workflows. To evaluate the cost of the methods in this study, we measured the wall time of each NMF variant as the factorization rank increased from 20 to 70. This analysis places the trade-offs between interpretability, accuracy, and computational cost on a common footing.

Table 3 summarizes mean wall time and standard deviation for each method. PCA is the fastest method by a wide margin (0.008 ± 0.003 s), and min-volume deep NMF is the slowest (115.328 ± 48.757 s). Among the NMF variants, fronorm has the highest mean runtime (2.941 ± 1.316 s), followed by deep NMF (2.441 ± 0.610 s), while multilayer NMF is the fastest (0.543 ± 0.137 s). The standard deviations for min-volume deep NMF and fronorm are large relative to their means, indicating that runtime varies substantially across ranks and runs.

As shown in Figure 1, deep NMF with min-volume regularization is by far the most computationally expensive method, with runtime exceeding 150 seconds at higher ranks. Its mean runtime is approximately 47 times that of unconstrained deep NMF and approximately 70–71 times the mean of the other NMF variants (excluding PCA). The min-volume constraint is the only difference between these two models, so the additional cost reflects that constraint. Excluding min-volume deep NMF, fronorm NMF consistently incurs the highest cost among the remaining variants, especially as rank grows. Its solver dynamics are heavier than those of the other methods. In contrast, PCA is the most efficient method, completing in roughly 0.004–0.021 seconds across all ranks. The other NMF methods, including β-divergence NMF, hierarchical rank-2 NMF, multilayer NMF, deep NMF, and SSNMF, show only modest increases in runtime with rank, which is consistent with expectations for larger materials datasets. Overall, these results underscore the need to balance interpretability, model complexity, and computational cost when selecting NMF architectures for large-scale materials informatics.

4.4 Tuning Number of Layers in Deep NMF

Determining the appropriate number of layers is a central challenge in deep NMF, as the depth of the model directly influences both interpretability and predictive performance. Unlike standard NMF, where the factorization is shallow and fixed, deep NMF introduces a hierarchy of latent representations whose quality depends on how information is propagated across layers.

Model	Random	NNDSVD	NNDSVDa	NNDSVDar	NNDSVD-LRC	Hier.†
RMSE						
pca	0.444 ± 0.012	–	–	–	–	–
umap	0.500 ± 0.003	–	–	–	–	–
hier	–	–	–	–	–	0.443 ± 0.013
beta	0.449 ± 0.018	0.438 ± 0.012	0.456 ± 0.019	0.436 ± 0.012	0.444 ± 0.012	–
fronorm	0.443 ± 0.014	0.447 ± 0.011	0.451 ± 0.018	0.446 ± 0.012	0.455 ± 0.014	–
ssnmf	0.578 ± 0.054	0.497 ± 0.009	0.520 ± 0.062	0.473 ± 0.016	0.484 ± 0.016	–
multilayer	0.442 ± 0.017	0.444 ± 0.014	0.461 ± 0.020	0.441 ± 0.015	0.444 ± 0.014	–
deep	0.441 ± 0.016	0.451 ± 0.014	0.464 ± 0.014	0.449 ± 0.013	0.440 ± 0.019	–
minvol	0.448 ± 0.012	0.450 ± 0.017	0.454 ± 0.015	0.450 ± 0.011	0.437 ± 0.021	–
MAE						
pca	0.197 ± 0.008	–	–	–	–	–
umap	0.250 ± 0.003	–	–	–	–	–
hier	–	–	–	–	–	0.197 ± 0.012
beta	0.202 ± 0.017	0.192 ± 0.011	0.209 ± 0.017	0.191 ± 0.010	0.198 ± 0.011	–
fronorm	0.196 ± 0.012	0.200 ± 0.010	0.204 ± 0.017	0.199 ± 0.010	0.207 ± 0.013	–
ssnmf	0.336 ± 0.064	0.247 ± 0.009	0.274 ± 0.065	0.224 ± 0.015	0.234 ± 0.015	–
multilayer	0.196 ± 0.015	0.197 ± 0.013	0.213 ± 0.018	0.194 ± 0.013	0.197 ± 0.013	–
deep	0.194 ± 0.014	0.204 ± 0.013	0.215 ± 0.013	0.202 ± 0.012	0.194 ± 0.017	–
minvol	0.200 ± 0.011	0.202 ± 0.015	0.206 ± 0.014	0.202 ± 0.010	0.192 ± 0.018	–
Accuracy (%)						
pca	80.3 ± 1.2	–	–	–	–	–
umap	75.0 ± 0.3	–	–	–	–	–
hier	–	–	–	–	–	80.3 ± 1.2
beta	79.8 ± 1.7	80.8 ± 1.1	79.1 ± 1.7	80.9 ± 1.0	80.2 ± 1.1	–
fronorm	80.4 ± 1.2	80.0 ± 1.0	79.6 ± 1.7	80.1 ± 1.0	79.3 ± 1.3	–
ssnmf	66.4 ± 6.4	75.3 ± 0.9	72.6 ± 6.5	77.6 ± 1.5	76.6 ± 1.5	–
multilayer	80.4 ± 1.5	80.3 ± 1.3	78.7 ± 1.8	80.6 ± 1.3	80.3 ± 1.3	–
deep	80.6 ± 1.4	79.6 ± 1.3	78.5 ± 1.3	79.8 ± 1.2	80.6 ± 1.7	–
minvol	80.0 ± 1.1	79.8 ± 1.5	79.4 ± 1.4	79.8 ± 1.0	80.8 ± 1.8	–
F1-macro (%)						
pca	68.6 ± 2.2	–	–	–	–	–
umap	46.1 ± 1.9	–	–	–	–	–
hier	–	–	–	–	–	68.9 ± 2.3
beta	66.6 ± 3.8	69.3 ± 2.2	65.9 ± 3.6	69.7 ± 2.0	68.9 ± 2.1	–
fronorm	66.9 ± 2.5	67.5 ± 2.2	66.4 ± 4.5	67.5 ± 2.3	66.5 ± 2.6	–
ssnmf	56.0 ± 5.9	75.3 ± 3.7	65.0 ± 6.1	59.2 ± 5.9	54.9 ± 5.4	–
multilayer	67.7 ± 3.0	68.1 ± 3.0	64.9 ± 3.6	68.9 ± 2.6	68.5 ± 2.4	–
deep	67.2 ± 3.6	66.5 ± 2.8	63.5 ± 2.9	67.2 ± 2.4	68.9 ± 3.3	–
minvol	67.0 ± 2.3	79.8 ± 2.6	66.0 ± 2.8	67.2 ± 1.9	69.1 ± 3.7	–

Table 2: Mean binary classification results (RMSE, MAE, Accuracy, and F1-macro) with standard deviations across ranks 20–70. Each cell reports the mean over ranks for the given initialization.

†Hierarchical NMF follows a fixed rank-two recursive procedure and has no initialization parameter, so it is not strictly controlled against PCA and is reported separately. Its results are given in the Hier.† column (no explicit initialization).

Too few layers may fail to capture meaningful hierarchical structure, while too many can lead to motif collapse, instability, or diminishing returns in accuracy. To address this, we develop a principled framework for selecting the number of layers by evaluating how well deeper architectures preserve structural information from the first layer and how they contribute to downstream supervised learning performance. This transforms depth selection from an arbitrary hyperparameter choice into a model selection problem, ensuring that deep NMF operates at a depth where hierarchical chemical motifs remain stable, interpretable, and predictive.

To determine the appropriate depth for deep NMF, we employ a geometric mean metric that jointly captures predictive performance and hierarchical interpretability. For a given depth

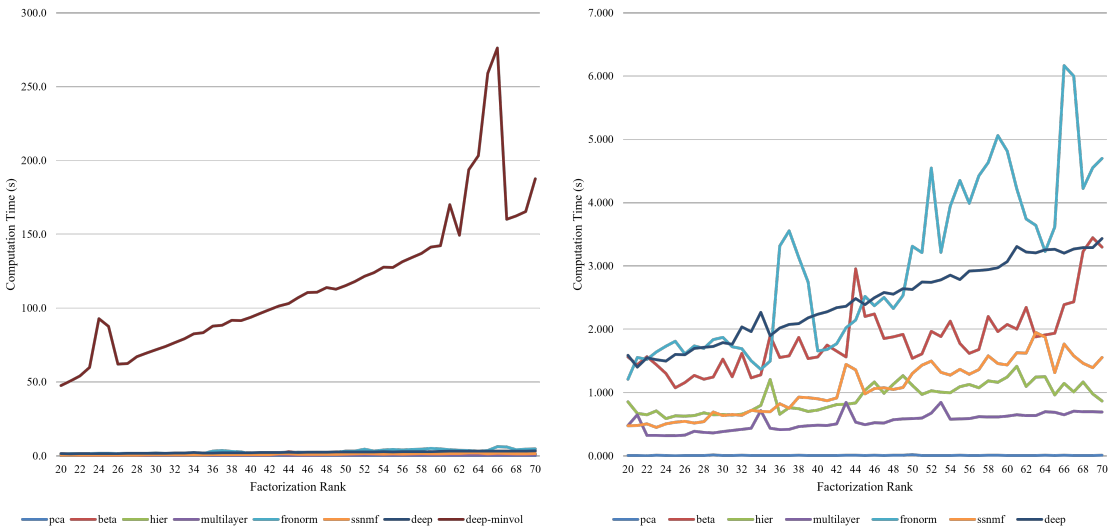


Figure 1: Computational time in seconds of each NMF variant from factorization rank 20 to 70. **Left**: computational results with minvol-deep NMF. **Right**: computational results without minvol-deep NMF as it inflates the y-axis making differentiating between other NMF variants difficult to visualize.

Method	**Mean runtime (s)**	**Standard deviation (s)**
pca	0.008	0.003
beta	1.835	0.532
hier	0.921	0.226
multilayer	0.543	0.137
fronorm	2.941	1.316
ssnmf	1.078	0.426
deep	2.441	0.610
minvol	115.328	48.757

Table 3: Mean wall-clock runtime and standard deviation in seconds for each method. Values are averaged over ranks 20–70.

k, we compute

$$GM(k) = \sqrt{score(k)_{norm} \cdot NMI(H_0, H_k)}, \tag{23}$$

where $NMI(H_0, H_k)$ measures the information preserved between the first layer representation H_0 (acting as a baseline) and the final layer representation H_k. The $score(k)_{norm}$ is a normalized supervised learning score. Because supervised metrics differ in scale and direction, R^2 may be negative, and RMSE is a lower-is-better quantity, we transform each metric into a bounded, higher-is-better score in [0, 1] before combining it with NMI. For regression, we map R^2 to R^2_+ $= \max(0, R^2)$, ensuring that negative values do not artificially inflate the geometric mean. For classification RMSE, we apply an inverse transformation,

$$RMSE_{norm} = 1 - \frac{RMSE - RMSE_{\min}}{RMSE_{\max} - RMSE_{\min}}, \tag{24}$$

using dataset-specific bounds to place all values on a comparable [0,1] scale. This normalization ensures that both interpretability and predictive performance contribute meaningfully and comparably to the depth-selection criterion.

The threshold $GM(k) \geq 0.5$ is chosen because it corresponds to a balanced regime in which both supervised performance and structural consistency are at least moderate. Since both terms

Rank	$GM(k) = 0.4$	$GM(k) = 0.5$	$GM(k) = 0.6$
20	2	2	2
21	2	2	2
22	2	2	2
23	3	3	3
24	2	2	2
25	2	2	2
26	2	2	2
27	2	2	2
28	2	2	2
29	5	5	5

Table 4: Sensitivity analysis of $GM(k)$ across different ranks. The selected depths remain stable across the evaluated values of $GM(k)$, with depth selection generally occurring early. This stability indicates that the method is not sensitive to moderate changes in $GM(k)$ and supports the use of $GM(k) = 0.5$ as an appropriate choice for our experiments.

lie in [0,1], their geometric mean provides a joint criterion that penalizes solutions for which either component becomes small. We use 0.5 as an interpretable operating threshold rather than as a strict guarantee that each individual component exceeds 0.5. For example, when the two components are comparable, $GM(k) = 0.5$ corresponds to both attaining a value of approximately 0.5; when they differ, a stronger value in one component can compensate only multiplicatively for a weaker value in the other. Consequently, depths with $GMK(k) < 0.5$ indicate an unfavorable joint trade-off, typically arising from degraded predictive performance, reduced preservation of hierarchical structure or both. In contrast, $GMK(k) \geq 0.5$ identifies depths that retain a practically meaningful compromise between these objectives. Empirically, this criterion produces stable and chemically interpretable depth selections across evaluated ranks (see Table 4 for sensitivity analysis). Although a geometric mean involving normalized reconstruction error can also be used, we find that incorporating supervised learning scores is more appropriate for this application. Reconstruction error decreases monotonically with depth and therefore provides limited guidance on interpretability; in contrast, supervised scores directly reflect the usefulness of the learned representations for downstream tasks.

To quantify interpretability across layers, we compute the NMI between H_0 and a propagated version of H_k. Because H_0 and H_k generally differ in shape, we recursively propagate H_k through the intermediate transformations so that both matrices share the same dimensionality before computing NMI. This choice reflects the layer-centric formulation of deep NMF, where as depth increases, all layers are optimized jointly through a global objective, and changes in H_0 reflect coordinated adjustments across the hierarchy. Using H_0 as the reference therefore provides a consistent measure of how much of the original structure is preserved as depth increases. An illustrative example is shown in Table 5, where a deep NMF model with rank 50 and depth 5 is analyzed by tracing the dominant metal class labels across layers. As depth increases from L_1 to L_5, several classes collapse (e.g., lanthanides and transition metals), and the corresponding NMI values decrease from 0.949 to 0.678. This decline reflects a loss of hierarchical diversity that is when a factor becomes dominated by a single elemental class. Therefore, it ceases to provide meaningful structural refinement. For this reason, we consider NMI values above 0.5 to indicate that sufficient hierarchical structure is preserved for scientific interpretation.

Overall, this framework provides a rigorous and interpretable method for selecting the number of layers in deep NMF. By combining supervised performance and structural hierarchy considerations, we ensure that the chosen depth yields representations that are both predictive and scientifically meaningful. Using Eq. (23), we determine the number of layers in deep NMF through a two-stage procedure. First, we set the depth to a large value (e.g., $k = 30$) and exam-

Metal class	$L1$	$L2$	$L3$	$L4$	$L5$
transition	190	130	105	100	25
basic	25	20	25	5	5
semimetal	20	30	30	5	20
alkaline	20	55	20	20	15
lanthanide	10	70	30	0	0
nonmetal	5	10	30	10	10
actinide	5	0	0	0	0
alkali	5	20	10	10	5
nmi	–	0.949	0.9151	0.8152	0.678

Table 5: Metal class distribution across layers with corresponding NMI values.

ine the resulting NMI values across all intermediate layers. Empirically, we observe that deep NMF produces highly consistent NMI trajectories regardless of whether the maximum depth is large or small. This makes a large depth run an efficient proxy for identifying the region in which NMI begins to deteriorate below a set threshold. In practice, this allows us to restrict the search to a significantly smaller depth interval without sacrificing accuracy or interpretability.

Across extensive experimentation, we find that the optimal depth is typically small, often $k \leq 5$. We therefore define a reduced search space of length k consisting of all depths whose NMI values exceed a threshold of 0.5, ensuring that hierarchical structure is still preserved. An exhaustive search over k is then performed, and the depth that maximizes $GM(k)$ is selected for each rank. Table 6 summarizes the results of this procedure across ranks 10–60. For each rank, we report the depth k that maximizes $GM(k)$, along with the corresponding supervised learning score, NMI value, and geometric mean. The results show that high predictive performance and high NMI can be simultaneously achieved, and that the optimal depth is consistently small, most often $k = 2$ for this sample. Despite the fact that each rank on average has an initial search space of approximately 15 candidate depths, (23) reliably selects shallow architectures that balance interpretability and performance. This demonstrates that deep NMF does not require large depth to achieve strong hierarchical structure and that the proposed selection method is both efficient and robust across ranks.

Rank	**Depth (k)**	**Score**	**NMI**	**GM(k)**
10	2	0.667	0.868	0.722
15	2	0.670	0.907	0.734
20	3	0.683	0.856	0.731
25	2	0.635	0.944	0.715
30	2	0.669	0.943	0.741
35	2	0.687	0.942	0.755
40	4	0.649	0.875	0.709
45	2	0.718	0.945	0.780
50	2	0.704	0.936	0.767
55	2	0.692	0.963	0.764
60	2	0.709	0.947	0.773

Table 6: Deep NMF depth selection metrics across ranks.

5 Interpretation

Understanding the latent structure learned by NMF and deep NMF is central to this work. While supervised benchmarks establish that NMF-based representations can match or exceed

PCA in predictive performance, their primary value lies in the additive, nonnegative factors that provide a natural basis for examining recurring compositional motifs. We distinguish between structural interpretation, which characterizes the organization and evolution of the learned representation, and physical interpretation, which relates these structures to chemically meaningful patterns. Our interpretability analysis is focused on structural interpretability, restricted to elemental composition descriptors, allowing individual factors to be traced to dominant elements or element groups. Element-level attribution, hierarchical layer mapping, Normalized Mutual Information (NMI), Jaccard overlap, and entropy are used to examine how compositional structures are organized and whether deeper layers preserve, refine, split, or merge structures learned in earlier layers. These analyses characterize latent structural organization rather than directly assigning physical meaning to individual factors.

Physical interpretation can be supported by examining whether persistent or dominant structural patterns align with recognizable chemical relationships. Such associations provide cues toward physical interpretation and hypothesis generation, rather than evidence that an individual NMF factor corresponds to a unique physical mechanism. We further assess identifiability and stability through min-volume regularization, since scientifically useful interpretations should be reproducible across changes in initialization or small perturbations. By reducing ambiguity in the learned components, min-volume constraints help determine whether observed structures reflect consistent properties of the data and decomposition rather than artifacts of a particular optimization run. Together, elemental attribution, hierarchical structural analysis, and stability testing provide a framework for determining whether NMF and deep NMF learn coherent, reproducible, and chemically grounded latent structures that can subsequently support chemical reasoning and future materials investigation.

5.1 Element Level Attribution

Element-level attribution provides the first layer of interpretability in our framework by linking each learned NMF factor to chemically meaningful elemental structure. For a factorization $X \approx WH$, each row of H represents a latent factor, and each column corresponds to an elemental feature. The largest coefficients in a row of H therefore identify the elements that contribute most strongly to that factor. We use top-k attribution to assign each factor to dominant elemental contributors and periodic-table classes, such as transition metals, lanthanides, alkali metals, or other chemically meaningful groups. This creates a direct map between the algebraic decomposition and the materials chemistry encoded in the feature matrix. Stable, consistently labeled factors indicate that the model is not merely fitting numerical structure, but recovering reproducible compositional motifs that can be related to stability, lattice constants, and other materials properties.

To evaluate whether these motifs persist across model complexity, we perform element-level attribution over factorization ranks from 20 to 70. For each rank and each factor, the ten elements with the largest H-matrix coefficients are selected as the factor label set. This provides a consistent rank-dependent description of the latent space in terms of original elemental features. To connect the learned factors to downstream predictive behavior, we train a Random Forest model on the NMF-derived representation and compute SHAP values using TreeSHAP [39]. SHAP values are computed only on the held-out test set to avoid data leakage and to ensure that factor importance reflects generalization rather than training-set fit. For each rank, we aggregate the mean absolute SHAP value across test samples to obtain a global importance score for every NMF factor. The highest-importance factors are then paired with their top-k elemental labels, yielding the attribution subsets used for coherence, diversity, and chemical-structure analysis. Table 7 provides an illustrative example of this element-level attribution procedure.

To illustrate the interpretability of the learned NMF representations, Figure 2 visualizes the relationship between elements, NMF ranks, and NMF factors for a selected subset of the dataset.

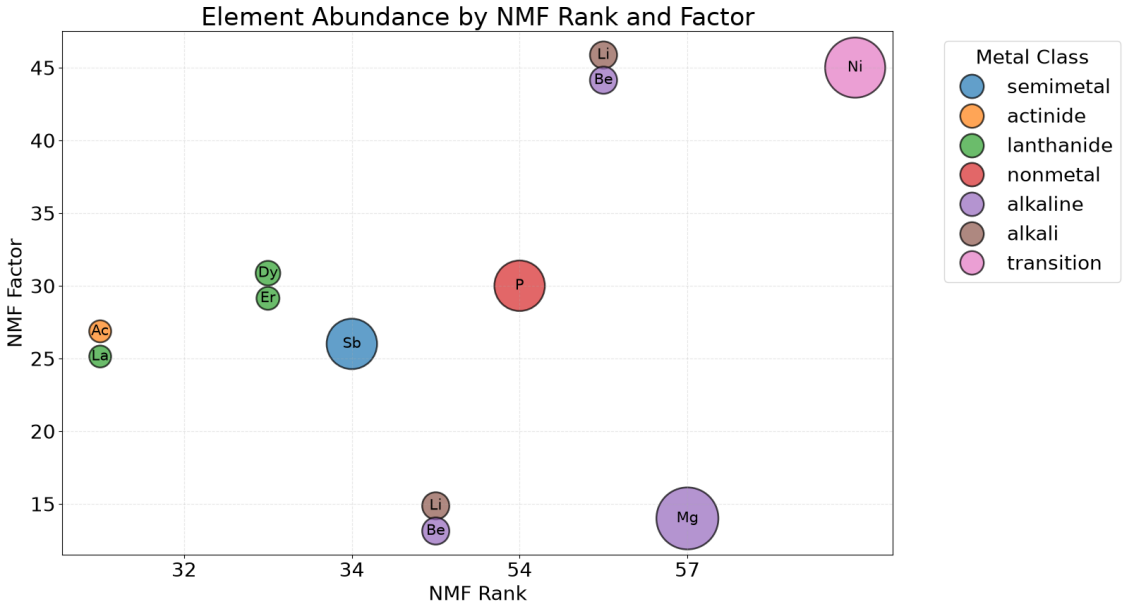


Figure 2: Elemental composition of selected NMF factors across different decomposition ranks. Each bubble represents an element associated with an NMF factor, where bubble size is proportional to elemental abundance and color identifies the corresponding elemental class. The figure uses a selected subset of the dataset for illustration, demonstrating how learned NMF factors can range from distributed representations involving multiple elements to highly concentrated representations dominated by a single element.

The horizontal and vertical axes represent the NMF rank and factor, respectively, while marker size indicates elemental abundance and marker color denotes the corresponding elemental class. The visualization reveals that certain factors are strongly dominated by individual elements. For example, Sb and P exhibit abundances of approximately 0.67 for factors 26 and 30 at ranks 32 and 34, respectively, whereas Mg and Ni exhibit substantially stronger associations at ranks 54 and 57, with abundances of approximately 1.00 and 0.94. This illustrative subset demonstrates how the NMF decomposition can expose the elemental composition underlying individual latent factors and distinguish between factors characterized by distributed elemental contributions and those dominated by a single element.

Factor	Abundance	Rank	Element #	Element	Metal Class	Importance	R^2	NMF Model
3	0.131	20	16	Ni	transition	0.303	0.568	deep
3	0.040	20	5	Si	semimetal	0.303	0.568	deep
3	0.026	20	6	P	nonmetal	0.303	0.568	deep
16	0.159	21	16	Ni	transition	0.298	0.605	deep
16	0.032	21	6	P	nonmetal	0.298	0.605	deep
16	0.011	21	42	Nd	lanthanide	0.298	0.605	deep
9	0.170	22	16	Ni	transition	0.297	0.594	deep
9	0.034	22	6	P	nonmetal	0.297	0.594	deep
9	0.012	22	42	Nd	lanthanide	0.297	0.594	deep

Table 7: Element-level attribution sample table across ranks 20–22 in the deep NMF model. Each rank of a Random Forest regressor is used on the lattice constant, and the SHAP values are computed to obtain the Importance. Lastly, the H matrix of each rank is used to retrieve the top elements with the highest abundance.

Although our evaluation is inspired by coherence and diversity metrics from topic modeling, direct analogues such as pointwise mutual information (PMI) [11] and normalized PMI (NPMI)

[5] are not appropriate for elemental-composition data. These metrics rely on semantic co-occurrence patterns between words in a corpus, whereas co-occurrence among elements does not carry the same linguistic meaning and can be dominated by dataset composition, stoichiometric constraints, or sampling bias. We therefore define chemistry-aware metrics that operate on the class structure of the attributed elements rather than on word-like co-occurrence statistics. For coherence, we measure the class purity of each factor's top-k elements. Let the top-k elements associated with a factor induce a class-label distribution $p = \{p_1, p_2, \ldots, p_C\}$ over C metal classes. We define the coherence score as

$$\text{Coherence} = 1 - \frac{H(p)}{\log C}, \tag{25}$$

where $H(p)$ is the entropy defined in Eq. (22). This score approaches one when the top-k elements are concentrated within a single metal class and decreases as the attribution becomes distributed across multiple classes. Thus, coherence captures whether a factor corresponds to a chemically concentrated motif rather than a diffuse mixture of unrelated element groups. Diversity provides a complementary measure of how broadly the attribution set spans the available elemental space. For an attribution slice containing the important factors selected for a given model and rank, let S denote the set of unique element labels appearing among their top-k attributions, and let E denote the total set of element labels available in the dataset. Then, we define diversity as follows:

$$\text{Diversity} = \frac{|S|}{|E|}. \tag{26}$$

High diversity indicates that the model distributes importance across a broader range of elements, whereas low diversity indicates that the attribution is concentrated on a smaller subset of the elemental space. Together, entropy-based coherence and element-label diversity provide chemically meaningful alternatives to linguistic topic-modeling metrics, allowing us to evaluate whether the learned factors are both internally interpretable and sufficiently broad to capture meaningful variation across the materials dataset.

NMF Model	**Element Coherence**	**Element Diversity**
hier	0.4016	0.0816
beta	0.3773	0.1200
multilayer	0.3554	0.1000
deep	0.3165	0.1400
deep-minvol	0.2855	0.1230
fronorm	0.2537	0.1733

Table 8: Element coherence and diversity across NMF models sorted by highest coherence.

The element coherence and diversity metrics provide a complementary view of how different NMF variants organize the latent space and how interpretable their learned factors are. As shown in Table 8, the models exhibit distinct structural tendencies. Hierarchical rank-2 NMF achieves the highest coherence score, 0.402, and the lowest diversity score, 0.082, indicating that its factors are highly concentrated within a small set of elemental classes. In contrast, Frobenius-norm NMF shows lower coherence, 0.254, and higher diversity, 0.173, suggesting a more distributed representation in which factor importance is spread across a broader range of elemental labels. Beta-divergence NMF, multilayer NMF, deep NMF, and min-volume deep NMF occupy intermediate regimes, balancing factor purity against chemical coverage in different ways. In particular, min-volume deep NMF yields a coherence of 0.286 and diversity of 0.123, indicating moderate specialization without severe collapse of the attribution space. These results show that the NMF variants do not differ only in predictive performance; they also impose different forms of structure on the learned representation. The appropriate choice

of factorization method should therefore depend on the scientific objective: highly coherent factors may be preferred when the goal is compact motif discovery, whereas higher diversity may be useful when broader chemical coverage is needed.

5.2 Layer Mapping

Layer mapping provides a structured view of how deep NMF reorganizes information across successive layers, allowing us to trace how compositional motifs evolve with model depth. To quantify these relationships, we use the Jaccard matrix, which measures the overlap between the top-k elemental contributors of factors in adjacent layers. High Jaccard values indicate motif preservation, where a deeper-layer factor inherits elemental structure from one or more factors in the preceding layer. Moderate overlaps distributed across several predecessors indicate motif merging, where the deeper representation integrates related chemical patterns into a broader abstraction. Near-zero overlap identifies motifs that are pruned, weakened, or no longer retained as the hierarchy deepens. In this way, the Jaccard matrix makes the layer-to-layer evolution of deep NMF explicit, showing whether deeper layers preserve, merge, refine, or discard earlier compositional patterns while maintaining a direct connection to interpretable chemical structure.

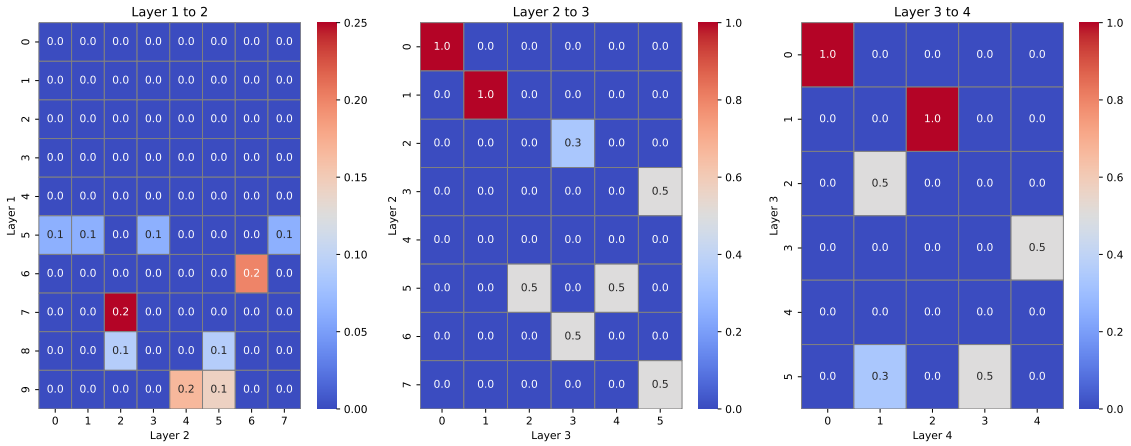


Figure 3: Jaccard matrices comparing adjacent layers of a 4-layer-deep NMF with ranks 10, 8, 6, and 5, respectively. Each entry shows the Jaccard overlap between the top-k inheritance of two factors. Most values are zero, with occasional small overlaps (0.1–0.2) and rare 1.0 entries indicating identical inheritance. Layer 1 (rank 10) to Layer 2 (rank 8) shows how initial motifs begin to split or merge; Layer 2 (rank 8) to Layer 3 (rank 6) captures mid-level reorganization; and Layer 3 (rank 6) to Layer 4 (rank 5) highlights the final refinement of chemical patterns across the hierarchy.

The layer mapping results in Fig. 3, visualized through the Jaccard heatmaps, reveal a consistent pattern of hierarchical refinement in the deep NMF decomposition. Each transition between layers contains a mixture of preserved motifs, merged motifs, and pruned motifs, illustrating how the model reorganizes compositional structure as depth increases. The Jaccard matrix highlights which earlier layer factors meaningfully contribute to each deeper layer factor, allowing researchers to trace how deeper representations remain grounded in the motifs learned at shallower depths. Between Layer 2 and Layer 3, we observe several clear inheritance relationships where Layer 3 Factor 0 (L3-F0) and Layer 3 Factor 1 (L3-F1) are pure continuations of L2-F0 and L2-F1, respectively, while L3-F5 emerges as a merged motif drawing equally from L2-F7 and L2-F3. A similar structure appears between Layer 3 and Layer 4: L4-F0 directly inherits L3-F0, L4-F2 inherits cleanly from L3-F1, and L3-F5 undergoes hierarchical refinement,

splitting into L4-F1 and L4-F3. By tracing the element-level attributions, we confirmed these relationships; for instance, the Li–Be motif dominating L3-F0 is preserved with high abundance in L4-F0, demonstrating how deeper layers reorganize but do not distort chemically meaningful patterns. Earlier layers show more diffuse relationships where L1-F5 distributes its mass across four Layer 2 factors with similar Jaccard values, which indicates broader refinement before motifs stabilize at deeper layers. Overall, the Jaccard-based layer mapping provides a transparent view of how deep NMF preserves stable motifs, merges related structures, and prunes weak patterns, enabling interpretable analysis even at the deepest layers of the hierarchy.

5.3 Min-Volume Identifiability

Min-Volume (minvol) identifiability provides a crucial lens for understanding how deep NMF behaves under constraints designed to enforce uniqueness and stability of the learned factors. Unlike standard deep NMF, where scaling and permutation perturbations can cause different evaluations to produce substantially different factor structures, the minvol formulation encourages solutions that are unique up to these trivial transformations. This behavior is consistent with the theoretical properties of minvol regularization, where minimizing the volume of the columns of W is a well-established strategy for enforcing identifiability in NMF, yielding solutions that are unique up to scaling and permutation. In practice, this constraint suppresses degenerate factorizations and prevents the model from drifting toward unstable regions of the optimization landscape [36]. This allows us to directly compare the latent representations learned by deep NMF with and without the min-volume constraint and to assess whether the chemical motifs identified are preserved or reorganized under a more identifiable model. By examining how minvol deep NMF reshapes or stabilizes factor identities across layers, we can determine whether the hierarchical structures observed in the unconstrained model reflect genuine chemical organization or artifacts of non-identifiable optimization. These comparisons can yield meaningful scientific insights: if minvol consistently recovers the same motifs regardless of initialization or rank, then those motifs are likely intrinsic to the dataset. Otherwise, the differences highlight where deep NMF may be unstable.

To evaluate robustness, we systematically perturb the initialization of each model by injecting controlled Gaussian noise directly into the initialization matrices before any factorization is performed. For each run, noise is added entry-wise to the initial W, H, and intermediate layer initializations. The perturbed matrices are then truncated to enforce nonnegativity, ensuring that all models begin from valid NMF-compatible starting points. To maintain consistency with our supervised learning experiments, noisy initializations are subsequently min–max normalized using the same preprocessing pipeline applied to the raw feature matrix. This ensures that the magnitude of the perturbation is comparable across ranks and models.

Our robustness grid spans ten random seeds, five ranks in the high-performing region ($r = 45 - 50$) and ten noise levels logarithmically spaced between minimal ($\eta = 0.0001$) and severe perturbation ($\eta = 16$). This yields 50 independent evaluations per method, providing reliable standard deviation estimates. Additionally, the same perturbation protocol is applied to both deep NMF and min-volume deep NMF, allowing a fair comparison of their sensitivity to initialization noise. This controlled stress test probes identifiability by examining whether small changes in initialization lead to consistent latent factors, predictive performance, and element-level attributions.

We quantify performance robustness using the mean absolute deviation (MAD) in RMSE relative to the noise-free baseline. For each noise level η, we compute

$$MAD(\eta) = \frac{1}{N} \sum_{i=1}^{N} |RMSE_i(\eta) - RMSE_i(0)|, \tag{27}$$

where N indexes the combinations of seeds and ranks. This metric provides and intuitive

measure of stability where larger deviations indicate that the model's predictive performance is sensitive to perturbations in initialization.

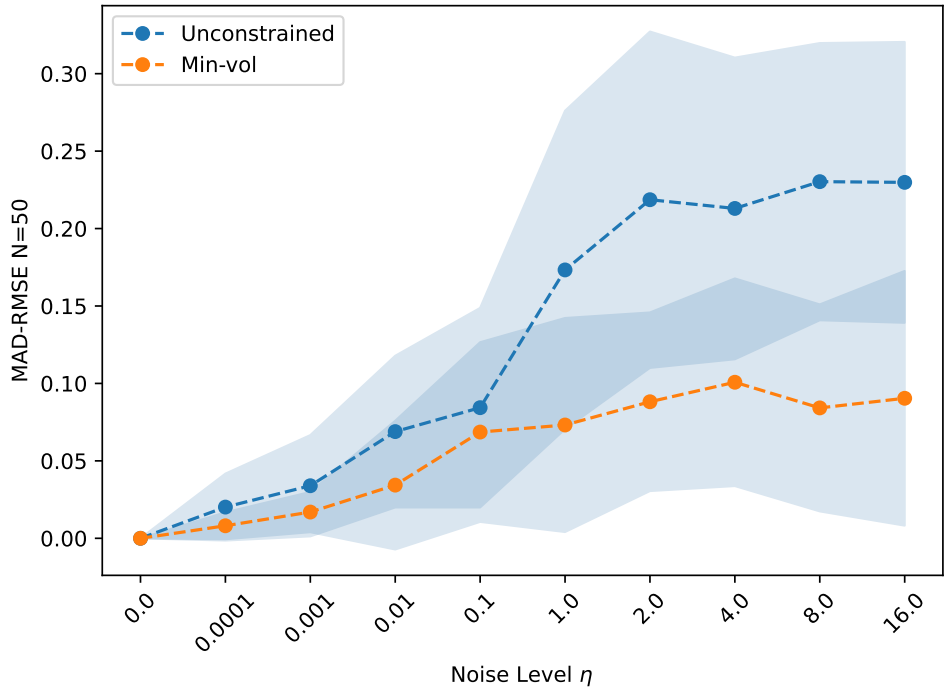


Figure 4: Mean absolute deviation–root mean square error (MAD–RMSE) as a function of noise level. The pale blue shaded region indicates one standard deviation from the mean.

The results in Fig. 4 show a clear separation between the two methods. Once the noise level exceeds $\eta > 1$, the unconstrained deep NMF model exhibits a sharp increase in MAD-RMSE, indicating substantial degradation in predictive accuracy. In contrast, the minvol deep NMF curve remains nearly flat across all noise levels, demonstrating exceptional robustness. These findings confirm that the minvol constraint stabilizes the optimization trajectory and prevents the model from collapsing under small perturbations.

Figure 4 shows MAD-RMSE as a function of noise level for the unconstrained and min-volume deep NMF models. The unconstrained model's error increases sharply once $\eta > 1$. The min-volume curve rises more gradually, reaching about 0.09–0.10 at the highest noise levels. Its one-standard-deviation band overlaps substantially with that of the unconstrained model at most noise levels, so the separation between the two methods is partially visible after $\eta > 1$. The results are consistent with min-volume regularization limiting error growth under large perturbations and helping the model avoid complete collapse under small perturbations.

Beyond predictive performance, we assess the stability of physical interpretability by examining how factor labels change under noise. Using the element-level attribution framework introduced earlier, we map each factor to its dominant elemental class and track how these assignments vary across seeds, ranks, and noise levels. A stable model should consistently recover the same small set of elemental labels, and instability manifests as rapid switching among many labels.

To quantify this behavior, we compute the entropy of the label distribution for each factor and again measure the mean absolute deviation from the noise-free baseline. Low entropy corresponds to concentrated, predictable factor mappings, while high entropy indicates diffuse or inconsistent assignments. Heatmaps of factor label distributions across ranks 45–50 reveal that minvol deep NMF consistently maps most factors to a small set of elements (e.g., Li and

a few others), whereas unconstrained deep NMF produces highly variable mappings that shift dramatically with noise, seed, and rank.

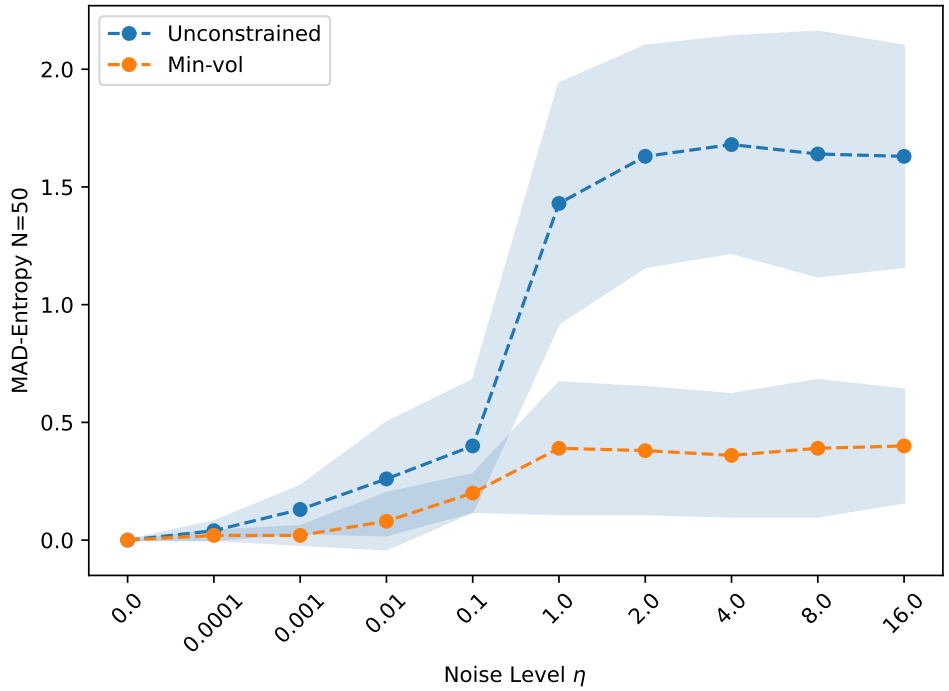


Figure 5: Mean absolute deviation–entropy (MAD–entropy) as a function of noise level.

The MAD-Entropy curves in Fig. 5 mirror the performance results. For noise levels $\eta > 1$, the entropy of the unconstrained model increases sharply, reflecting unstable and inconsistent factor interpretations. Minvol deep NMF, however, maintains low entropy across all noise levels, indicating that its factorization preserves meaningful and reproducible physical structure. Altogether, these results demonstrate that minvol deep NMF is substantially more robust than unconstrained deep NMF, both in terms of predictive performance and interpretability. The minvol constraint stabilizes the factorization against noise, preserves identifiable chemical motifs, and yields consistent hierarchical structure across seeds and ranks. These benefits, however, come at a substantial computational cost, with minvol deep NMF requiring approximately 70-71× greater average runtime than the other NMF variants while providing only marginal improvements in predictive performance. Thus, minvol deep NMF is most appropriate for scientific applications in which factor identifiability, reproducibility, and physical interpretability are sufficiently important to justify the additional computational expense. When computational efficiency or predictive performance alone is the primary consideration, the unconstrained model may provide a more favorable balance between accuracy and computational cost.

6 Conclusion

This work establishes NMF-based representation learning as a practical and scientifically interpretable framework for high-dimensional materials modeling. The central question was not whether NMF can outperform every available predictive architecture, but whether nonnegative matrix factorization and its modern extensions can provide representations that remain predictive while exposing chemically meaningful structure. This distinction is important for materials discovery, where model usefulness depends not only on numerical accuracy but also

on whether the learned representation can support scientific reasoning, feature attribution, and physically grounded interpretation. We show that standard NMF, hierarchical NMF, multilayer NMF, deep NMF, and min-volume regularized deep NMF provide competitive alternatives to PCA for supervised materials-property prediction. Across regression and classification tasks, NMF-based representations achieve performance broadly comparable to PCA, demonstrating that interpretability does not require a substantial sacrifice in predictive utility. Unlike PCA, however, NMF produces additive, nonnegative factors that are naturally aligned with compositional and descriptor-based materials data. These factors can be mapped to elemental groups, compositional motifs, and feature families, allowing the latent representation to be examined as part of the scientific analysis rather than treated as an opaque preprocessing step.

The primary methodological contribution is not the introduction of new NMF algorithms, but their systematic integration and analysis within a unified Python and PyTorch framework for high-dimensional materials datasets and GPU-accelerated computation. This implementation brings NMF variants previously distributed across separate formulations and software implementations into a consistent workflow for evaluation and comparison. A second contribution is a unified interpretability pipeline for translating NMF representations, particularly hierarchical and deep representations, into materials-specific scientific insights. The pipeline combines factor attribution, cross-layer mapping, overlap analysis, and stability diagnostics to examine how elemental contributions, descriptor patterns, and compositional motifs are organized within and across latent layers. Together, the computational framework and interpretation methodology provide a systematic approach for evaluating NMF-based representations while investigating what the learned factors represent and how their internal structure relates to materials chemistry.

The min-volume regularized deep NMF results further highlight the importance of identifiability in scientific machine learning. In materials discovery, a latent factorization is most useful when its components are stable enough to support interpretation across initializations and perturbations. We find that min-volume regularization improves the stability and interpretability of deep NMF factors under noisy initialization, suggesting that geometric constraints can make deep nonnegative representations more reproducible and scientifically meaningful. This result is important because interpretability without stability is fragile: a factor that changes substantially under small perturbations cannot reliably support chemical insight.

The stability benefits of min-volume regularization carry a computational cost. Min-volume deep NMF required approximately 70–71 times the runtime of the other variants, while its predictive gains over unconstrained deep NMF were modest and varied by initialization and metric (Table 1; Table 2). Min-volume deep NMF is therefore most appropriate when factor identifiability and reproducibility justify the additional expense. When computational cost or predictive accuracy is the primary concern, unconstrained deep NMF offers a more favorable trade-off.

Taken together, these findings position NMF and its modern extensions as more than dimensionality-reduction tools. They provide a transparent representation-learning framework for materials informatics, linking predictive modeling with chemically interpretable latent structure, hierarchical organization, and robustness analysis. The broader implication is that interpretable factorization methods can serve as a bridge between data-driven prediction and domain understanding, especially in settings where nonnegative, additive descriptors are physically meaningful. Future work should extend this framework to additional materials classes, multimodal datasets, and experimentally measured descriptors such as spectra, diffraction patterns, microscopy data, and time-resolved signals. Further development of constrained and semi-supervised NMF variants may also improve the alignment between latent factors and target properties, enabling representations that are simultaneously predictive, stable, and chemically interpretable. By emphasizing interpretability as part of the modeling objective rather than an afterthought, NMF-based methods offer a promising route toward more transparent

and scientifically useful materials discovery workflows. The accompanying `nimd` package provides a reproducible implementation of this framework for future benchmarking, extension, and application to other materials datasets where interpretable, nonnegative representations are scientifically valuable.

Code Availability

The implementation developed in this work is available as `nimd` (Nonnegative Interpretable Matrix Decomposition), a Python package for benchmarking, interpreting, and extending NMF-based representation learning methods for materials informatics. The repository includes the source code, installation instructions, example workflows, packaged benchmark data, and configuration files required to reproduce and extend the analyses presented here. The code is available at `https://github.com/thaksheel/nimd`.

Acknowledgement

This work was supported in part by the U.S. Department of Energy under Grant No. DE-SC0025801.

References

[1] Aldana-Bobadilla, E., Kuri-Morales, A., 2015. A clustering method based on the maximum entropy principle. Entropy 17, 151–180. URL: `https://www.mdpi.com/1099-4300/17/1/151`, doi:`10.3390/e17010151`.

[2] Atif, S.M., Qazi, S., Gillis, N., 2017. Improved svd-based initialization for nonnegative matrix factorization using low-rank correction. Pattern Recognition Letters 97, 96–102. doi:`10.1016/j.patrec.2017.07.017`.

[3] Berry, M.W., Browne, M., Langville, A.N., Pauca, V.P., Plemmons, R.J., 2007. Algorithms and applications for approximate nonnegative matrix factorization. Computational statistics & data analysis 52, 155–173.

[4] Bonnet, N., Nuzillard, D., 2005. Ultramicroscopy. Ultramicroscopy 102, 327. doi:`10.1016/j.ultramic.2004.10.003`.

[5] Bouma, G., 2009. Normalized (pointwise) mutual information in collocation extraction, in: Proceedings of the Biennial GSCL Conference.

[6] Boutsidis, C., Gallopoulos, E., 2008. Svd-based initialization: A head start for nonnegative matrix factorization. Pattern Recognition 41, 1350–1362. doi:`10.1016/j.patcog.2007.09.010`.

[7] Healy, J., McInnes, L., 2024. Uniform manifold approximation and projection. Nature Reviews Methods Primers 4, 82.

[8] Van, M.H., Verma, P., Zhao, C., Wu, X., 2026. A survey of ai for materials science: foundation models, llm agents, datasets, and tools. ACM Computing Surveys 58, 1–37.

[9] Vela, B., Hastings, T., Allen, M., Arróyave, R., 2025. Visualizing high entropy alloy spaces: methods and best practices. Digital Discovery 4, 181–194.

[10] Burke, M., Watanabe, M., Williams, D., Hyde, J., 2006. Quantitative characterization of nanoprecipitates in irradiated low-alloy steels: Advances in the application of feg-stem quantitative microanalysis to real materials. Journal of materials science 41, 4512–4522.

[11] Church, K.W., Hanks, P., 1989. Word association norms, mutual information, and lexicography. Computational Linguistics 16, 22–29.

[12] Cichocki, A., Zdunek, R., Phan, A.H., Amari, S.i. (Eds.), 2009. Nonnegative Matrix and Tensor Factorizations. Mathematics of Data Science: A Computational Approach to Clustering and Classification, Wiley, Chichester, UK.

[13] Costa, L.d.F., 2021. Further generalizations of the jaccard index. arXiv preprint arXiv:2110.09619 URL: https://doi.org/10.48550/arXiv.2110.09619, doi:10.48550/arXiv.2110.09619.

[14] Csiszár, I., 1975. I-divergence geometry of probability distributions and minimization problems. The annals of probability , 146–158.

[15] Ding, C., Li, T., Peng, W., Park, H., 2006. Orthogonal nonnegative matrix t-factorizations for clustering, in: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 126–135.

[16] Dobigeon, N., Brun, N., 2012. Ultramicroscopy. Ultramicroscopy 120, 25. doi:10.1016/j.ultramic.2012.06.002.

[17] Dobigeon, N., Moussaoui, S., Coulon, M., Tourneret, J.Y., Hero, A.O., 2009. Ieee transactions on signal processing. IEEE Transactions on Signal Processing 57, 4355. doi:10.1109/TSP.2009.2022893.

[18] Fagot, D., Wendt, H., Févotte, C., Smaragdis, P., 2019. Majorization-minimization algorithms for convolutive nmf with the beta-divergence, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. pp. 8202–8206.

[19] Gillis, N., 2021. Nonnegative Matrix Factorization. SIAM Data Science, Society for Industrial and Applied Mathematics, Philadelphia, PA. doi:10.1137/1.9781611976419.

[20] Gillis, N., Kuang, D., Park, H., 2014. Hierarchical clustering of hyperspectral images using rank-two nonnegative matrix factorization. IEEE Transactions on Geoscience and Remote Sensing 52, 1–12.

[21] Guo, Y.T., Li, Q.Q., Liang, C.S., 2024. The rise of nonnegative matrix factorization: Algorithms and applications. Information Systems 123, 102379.

[22] Haddock, J., Kassab, L., Li, S., Kryshchenko, A., Grotheer, R., Sizikova, E., Wang, C., Merkh, T., Madushani, R.W.M.A., Ahn, M., Needell, D., Leonard, K., 2020. Semi-supervised nmf models for topic modeling in learning tasks. URL: https://arxiv.org/abs/2010.07956, arXiv:2010.07956.

[23] Hart, J.B., Kastuar, S.M., Iloanya, A.C., Ekuma, C.E., 2025. High-throughput discovery of topological and magnetic phases in heuslerenes. APL Materials 13, 111107. URL: https://doi.org/10.1063/5.0297707, doi:10.1063/5.0297707.

[24] Honda, Y., Muto, S., Tatsumi, K., Kondo, H., Horibuchi, K., Kobayashi, T., Sasaki, T., 2015. Microscopic mechanism of path-dependence on charge–discharge history in lithium iron phosphate cathode analysis using scanning transmission electron microscopy and electron energy-loss spectroscopy spectral imaging. Journal of Power Sources 291, 85–94.

[25] Hoyer, P.O., 2004. Non-negative matrix factorization with sparseness constraints. Journal of machine learning research 5, 1457–1469.

[26] Jiménez-Sánchez, D., Ariz, M., Morgado, J.M., Cortés-Domínguez, I., Ortiz-de Solórzano, C., 2020. Nmf-ri: blind spectral unmixing of highly mixed multispectral flow and image cytometry data. Bioinformatics 36, 1590–1598.

[27] Kimoto, K., Kikkawa, J., Harano, K., Cretu, O., Shibazaki, Y., Uesugi, F., 2024. Unsupervised machine learning combined with 4d scanning transmission electron microscopy for bimodal nanostructural analysis. Scientific Reports 14, 2901.

[28] Kimoto, K., Uesugi, F., Harano, K., et al., 2025. Nonnegative matrix factorization incorporating domain specific constraints for four dimensional scanning transmission electron microscopy. Scientific Reports 15, 39143. URL: https://doi.org/10.1038/s41598-025-23541-7, doi:10.1038/s41598-025-23541-7.

[29] Kojima, Y., Muto, S., Tatsumi, K., Kondo, H., Oka, H., Horibuchi, K., Ukyo, Y., 2011. Degradation analysis of a ni-based layered positive-electrode active material cycled at elevated temperatures studied by scanning transmission electron microscopy and electron energy-loss spectroscopy. Journal of Power Sources 196, 7721–7727.

[30] Li, S., Li, W., Lu, H., Li, Y., 2023. Semi-supervised non-negative matrix tri-factorization with adaptive neighbors and block-diagonal learning. Engineering Applications of Artificial Intelligence 121, 106043. doi:10.1016/j.engappai.2023.106043.

[31] Yuan, J., Li, S., Wu, S., 2026. Unified joint semi-supervised non-negative matrix factorization for multi-view data clustering. IEEE Transactions on Mobile Computing 25, 11, 19202–19218. doi:10.1109/TMC.2026.3709098.

[32] Kotula, P.G., Keenan, M.R., 2006. Application of multivariate statistical analysis to stem x-ray spectral images: Interfacial analysis in microelectronics. Microscopy and Microanalysis 12, 538–544.

[33] Kvålseth, T.O., 2017. On normalized mutual information: Measure derivations and properties. Entropy 19. URL: https://www.mdpi.com/1099-4300/19/11/631, doi:10.3390/e19110631.

[34] Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. nature 401, 788–791.

[35] Leplat, V., Gillis, N., Févotte, C., 2022a. Multi-resolution beta-divergence nmf for blind spectral unmixing. Signal Processing 193, 108407.

[36] Leplat, V., Le Thi Khanh, H., Onwunta, A., Gillis, N., 2024. Deep nonnegative matrix factorization with beta divergences. Neural Computation. URL: https://doi.org/10.1162/neco_a_01679, doi:10.1162/neco_a_01679.

[37] López-Serrano, P., Dittmar, C., Özer, Y., Müller, M., 2019. Nmf toolbox: Music processing applications of nonnegative matrix factorization, in: Proceedings of the International Conference on Digital Audio Effects DAFx, pp. 2–6.

[38] Lucas, G., Burdet, P., Cantoni, M., Hébert, C., 2013. Multivariate statistical analysis as a tool for the segmentation of 3d spectral data. Micron 52, 49–56.

[39] Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S.,

Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

[40] Muto, S., Sasano, Y., Tatsumi, K., Sasaki, T., Horibuchi, K., Takeuchi, Y., Ukyo, Y., 2009a. Capacity-fading mechanisms of linio2-based lithium-ion batteries: Ii. diagnostic analysis by electron microscopy and spectroscopy. Journal of the electrochemical society 156, A371–A377.

[41] Muto, S., Shiga, M., 2020. Application of machine learning techniques to electron microscopic/spectroscopic image data analysis. Microscopy 69, 110–122.

[42] Muto, S., Tatsumi, K., Kojima, Y., Oka, H., Kondo, H., Horibuchi, K., Ukyo, Y., 2012. Effect of mg-doping on the degradation of linio2-based cathode materials by combined spectroscopic methods. Journal of Power Sources 205, 449–455.

[43] Muto, S., Tatsumi, K., Sasaki, T., Kondo, H., Ohsuna, T., Horibuchi, K., Takeuchi, Y., 2010. Mapping of heterogeneous chemical states of lithium in a linio2-based active material by electron energy-loss spectroscopy. Electrochemical and Solid-State Letters 13, A115–A117.

[44] Muto, S., Yoshida, T., Tatsumi, K., 2009b. Diagnostic nano-analysis of materials properties by multivariate curve resolution applied to spectrum images by s/tem-eels. Materials transactions 50, 964–969.

[45] Muto, S., Yoshida, T., Tatsumi, K., 2009c. Materials transactions. Materials Transactions 50, 964. doi:10.2320/matertrans.MRA2008245.

[46] Nascimento, J., Bioucas-Dias, J., 2005. Ieee transactions on geoscience and remote sensing. IEEE Transactions on Geoscience and Remote Sensing 43, 898. doi:10.1109/TGRS.2004.842468.

[47] Pu, Y., Tang, G., Wang, W., Savage, H., Schantz, S., Alfano, R., 2011. Native fluorescence spectroscopic evaluation of chemotherapeutic effects on malignant cells using nonnegative matrix factorization analysis. Technology in cancer research & treatment 10, 113–120.

[48] Ruckebusch, C., 2016. Data handling in science and technology, in: Walczak, B., Buydens, L. (Eds.), Data Handling in Science and Technology. Elsevier B.V., Amsterdam.

[49] Senga, J., Tatsumi, K., Muto, S., Yoshida, T., 2015. Effects of nitrogen-related defects on visible light photocatalytic response in n+ implanted tio2: A first-principles study. Journal of Applied Physics 118.

[50] Seyedi, A., He, L., Zhao, S., Onwunta, A., Gillis, N., 2026. Supervised deep multimodal matrix factorization for interpretable brain network analysis. https://arxiv.org/abs/2605.13312.

[51] Shiga, M., Tatsumi, K., Muto, S., Tsuda, K., Yamamoto, Y., Mori, T., Tanji, T., 2016. Sparse modeling of eels and edx spectral imaging data by nonnegative matrix factorization. Ultramicroscopy 170, 43–59.

[52] Spiegelberg, J., Rusz, J., 2017. Ultramicroscopy. Ultramicroscopy 175, 40. doi:10.1016/j.ultramic.2016.12.004.

[53] Tatsumi, K., Muto, S., 2009. Local electronic structure analysis by site-selective elnes using electron channeling and first-principles calculations. Journal of Physics: Condensed Matter 21, 104213.

[54] Tatsumi, K., Muto, S., Yamamoto, Y., Ikeno, H., Yoshioka, S., Tanaka, I., 2006. Site-specific electronic structure analysis by channeling eels and first-principles calculations. Ultramicroscopy 106, 1019–1023.

[55] Trebbia, P., Bonnet, N., 1990. Eels elemental mapping with unconventional methods i. theoretical basis: Image analysis with multivariate statistics and entropy concepts. Ultramicroscopy 34, 165–178.

[56] Trigeorgis, G., Bousmalis, K., Zafeiriou, S., Schuller, B., 2014. A deep semi-nmf model for learning hidden representations, in: International conference on machine learning, PMLR. pp. 1692–1700.

[57] Vesselinov, V.V., Ahmmed, B., Mudunuru, M.K., Pepin, J.D., Burns, E.R., Siler, D.L., Karra, S., Middleton, R.S., 2022. Discovering hidden geothermal signatures using non-negative matrix factorization with customized k-means clustering. Geothermics 106, 102576.

[58] Wang, D., Li, T., Deng, P., Liu, J., Huang, W., Zhang, F., 2022. A generalized deep learning algorithm based on nmf for multi-view clustering. IEEE Transactions on Big Data 9, 328–340.

[59] Wang, J.H., Hopke, P.K., Hancewicz, T.M., Zhang, S.L., 2003a. Application of modified alternating least squares regression to spectroscopic image analysis. Analytica Chimica Acta 476, 93–109.

[60] Wang, J.H., Hopke, P.K., Hancewicz, T.M., Zhang, S.L., 2003b. Application of modified alternating least squares regression to spectroscopic image analysis. Analytica Chimica Acta 476, 93–109. URL: https://www.sciencedirect.com/science/article/pii/S0003267002013697, doi:https://doi.org/10.1016/S0003-2670(02)01369-7.

[61] Yamamoto, Y., Tatsumi, K., Muto, S., 2007. Site-selective electronic structure of aluminum in oxide ceramics obtained by tem-eels analysis using the electron standing-wave method. Materials Transactions 48, 2590–2594.

[62] Yoshida, T., Muto, S., Wakabayashi, J., 2007. Depth-resolved eels and chemical state mapping of n+-implanted tio2 photocatalyst. Materials Transactions 48, 2580–2584.