Hyperspectral Unmixing for Raman Spectroscopy via Physics-Constrained Autoencoders

Dimitar Georgiev [1 2] **Álvaro Fernández-Galiana** [2 *] **Simon Vilms Pedersen** [2 *] **Georgios Papadopoulos** [1 2]
Ruoxiao Xie [2] **Molly M. Stevens** [2 3] **Mauricio Barahona** [4]

Abstract

Raman spectroscopy is widely used across science and industry to characterize the chemical composition of samples in a nondestructive, label-free manner. Many applications entail the unmixing of signals from mixtures of molecular species to identify the individual components present and their proportions, yet conventional methods for chemometrics often struggle with complex mixture scenarios encountered in practice. Here, we develop hyperspectral unmixing algorithms for Raman spectroscopy based on autoencoder neural networks, which we systematically validate using synthetic and experimental benchmark datasets created in-house. Our results demonstrate that unmixing autoencoders provide improved accuracy, robustness and efficiency compared to standard unmixing methods. We also showcase the applicability of autoencoders to complex biological settings by showing improved biochemical characterization of volumetric Raman imaging data from a human leukemia monocytic cell.

1. Introduction

Raman spectroscopy (RS) is a powerful optical modality that facilitates the identification, characterization and quantification of the molecular composition of chemical and biological specimens, offering in-depth insights into their structure and functionality (Movasaghi et al., 2007; Talari et al., 2015; Butler et al., 2016; McCreery, 2005; Smith & Dent, 2019). RS interrogates the vibrational modes of molecules through the analysis of inelastic scattering of monochromatic light from matter, thereby enabling the nondestructive, label-free fingerprinting of chemical species (Koningstein, 2012; Szymanski, 2012; Colthup, 2012; Jones et al., 2019; Bocklitz et al., 2016). As a result, RS has become an important analytical tool in a myriad of scientific domains, from chemistry (Schlücker, 2014; Dodo et al., 2022), biology (Pezzotti, 2021; Smith et al., 2016; Shipp et al., 2017; Cialla-May et al., 2017), and medicine (Kong et al., 2015; Ember et al., 2017; Pence & Mahadevan-Jansen, 2016; Balan et al., 2019; Auner et al., 2018; Mahadevan-Jansen & Richards-Kortum, 1996; Tanwar et al., 2021), to materials science (Fernández-Galiana et al., 2023; Weber & Merlin, 2013; Kumar, 2012), pharmacology (Wang et al., 2018; Paudel et al., 2015; Vankeirsbilck et al., 2002), environmental science (Halvorson & Vikesland, 2010; Ong et al., 2020; Terry et al., 2022), food quality control (Li & Church, 2014; Pang et al., 2016), and even forensics (Chalmers et al., 2012; Khandasammy et al., 2018; Izake, 2010).

Despite the wealth of information RS affords, the analysis and interpretation of experimental RS data remains a major challenge (Ryabchykov et al., 2018; Guo et al., 2021; Gautam et al., 2015). Many important applications entail the analysis of complex mixtures of molecular species coexisting and interacting at micro- and nanoscales. Such complexity can hinder the qualitative and quantitative investigation of RS measurements, especially when dealing with the biomolecular diversity of biological samples (Byrne et al., 2016; Gautam et al., 2015).

Hyperspectral unmixing (also known as (hyper)spectral deconvolution or multivariate curve resolution) aims to resolve such mixed signals (Li et al., 2017; Olmos et al., 2017) by identifying the individual components present (*endmember identification*) and/or quantifying their proportions (*abundance estimation*). Popular approaches include N-FINDR (Winter, 1999) and Vertex Component Analysis (VCA) (Nascimento & Dias, 2005) for endmember iden-

[*]Equal contribution [1]Department of Computing & UKRI Centre for Doctoral Training in AI for Healthcare, Imperial College London, London, United Kingdom, SW7 2AZ [2]Department of Materials, Department of Bioengineering & Institute of Biomedical Engineering, Imperial College London, London, United Kingdom, SW7 2AZ [3]Department of Physiology, Anatomy and Genetics, Department of Engineering Science, and Kavli Institute for Nanoscience Discovery, University of Oxford, Oxford, United Kingdom, OX1 3QU [4]Department of Mathematics, Imperial College London, London, United Kingdom, SW7 2AZ. Correspondence to: Molly M. Stevens <molly.stevens@dpag.ox.ac.uk>, Mauricio Barahona <m.barahona@imperial.ac.uk>.

Accepted at the 5^{th} AI for Science Workshop at ICML 2024.

tification, and Non-negative Least Squares (NNLS) (Lawson & Hanson, 1995) and Fully Constrained Least Squares (FCLS) (Heinz et al., 2001) for abundance estimation (Li et al., 2017; Hedegaard et al., 2011). However, such techniques, which originated in the field of remote sensing (Keshava & Mustard, 2002; Harris, 2006), have limitations for the unmixing of RS data. Specifically, these methods are restricted to linear mixing; lack robustness to data artifacts abundant in RS data (e.g., dark noise, baseline variations, cosmic spikes); rely on additional assumptions (e.g., endmembers present as 'pure pixels' in the data) or require additional information (e.g., number of endmembers, underlying mixture model, endmember library); and are computationally demanding for large datasets (e.g., imaging and volumetric Raman raster scans).

In this work, we introduce a new approach for RS hyperspectral unmixing based on autoencoder (AE) neural networks, which we systematically validate against conventional methods for unmixing using a comprehensive array of synthetic and experimental Raman spectroscopy data.

2. Background

Hyperspectral unmixing in Raman spectroscopy. Raman spectra can be represented as vectors $\mathbf{x} \in \mathbb{R}_+^b$, whose components correspond to the intensity of inelastically scattered light binned over b wavelength/wavenumber bands. Such measurement can be treated as the result of an underlying mixing of n 'pure' components, defined by their Raman signatures $\mathbf{m}_i \in \mathbb{R}_+^b$, $i = 1, \ldots, n$ (*endmembers*), and their respective proportions $\{\alpha_i\}_{i=1}^n$, $\alpha_i \in \mathbb{R}_+$ (*fractional abundances*). Hyperspectral unmixing is the inverse problem of recovering the endmembers and fractional abundances from a (collection of) measurement(s) $\mathbf{x}$ (Figure 1a). The unmixing can be performed with respect to a set of known endmembers (non-blind unmixing) or without knowing the endmembers (blind unmixing). Here, we focus on blind unmixing but we also discuss how to adapt the framework to the simpler problem of non-blind unmixing.

Linear mixing model. A major hurdle for unmixing is the lack of information about the underlying mixing function. The simplest and most common model is the linear mixing model (LMM), where measurements are assumed to be a linear combination of the endmembers:

$$\mathbf{x} = M\boldsymbol{\alpha} = \sum_{i=1}^{n} \alpha_i \mathbf{m}_i, \tag{1}$$

where $M = \begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \cdots & \mathbf{m}_n \end{bmatrix}$ is an $b \times n$ non-negative matrix containing the n endmember signatures, and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_n)^\mathsf{T}$ is an $n \times 1$ vector storing the corresponding abundances. A random noise term $\boldsymbol{\epsilon} \in \mathbb{R}^b$ is also often added to Equation (1) to model stochastic variations. The abundances α_i are constrained to be non-negative (i.e., the abundance non-negativity constraint (ANC), $\alpha_i \geq 0, \forall i$), and are forced to sum to 1 when corresponding to proportions (i.e., the abundance sum-to-one constraint (ASC), $||\boldsymbol{\alpha}||_1 = 1$). Standard methods for unmixing, such as N-FINDR and VCA for endmember identification, and NNLS and FCLS for abundance estimation, operate under the LMM (see Appendix A for more information).

Nonlinear mixing models. The LMM is a good approximation when endmember species are spatially well-separated with respect to the focal volume, and complex light interactions that cause non-linear signal contributions can be neglected (Keshava & Mustard, 2002). However, when non-linear interactions become significant, more intricate models are required (Dobigeon et al., 2013). To represent phenomena such as multiple scattering events, topographic variances and shadowing effects, the LMM has been extended in remote sensing to more complex variants (Dobigeon et al., 2013; Heylen et al., 2014), including intimate mixture models (Hapke, 1981), bilinear models (Halimi et al., 2011a; Fan et al., 2009; Halimi et al., 2011b), multilinear models (Heylen & Scheunders, 2015), or post-nonlinear models (Altmann et al., 2012), among others. A popular bilinear mixing model is the Fan model (Fan et al., 2009):

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{m}_i + \sum_{k=1}^{n} \sum_{\substack{l=1, \\ l \neq k}}^{n} \alpha_k \mathbf{m}_k \odot \alpha_l \mathbf{m}_l, \tag{2}$$

where $\odot$ is the Hadamard product. Nonetheless, accounting for non-linear mixing interactions increases the complexity and computational cost of unmixing (Dobigeon et al., 2013; Heylen et al., 2014), an issue of especial relevance in Raman spectroscopy where datasets (e.g., imaging/volumetric scans) are typically larger than in remote sensing. Hence, despite its limitations, the LMM remains a cornerstone of hyperspectral unmixing in most practical settings (Bioucas-Dias et al., 2012; Li et al., 2017).

Autoencoders. Autoencoders are a family of (deep) neural network models consisting of two sub-networks (*encoder* and *decoder*) connected sequentially (Goodfellow et al., 2016). The encoder $\mathcal{E} : \mathbb{R}^b \to \mathbb{R}^m$, usually $m \ll b$, transforms input data $\mathbf{x}$ to a lower-dimensional latent space representation $\mathbf{z} = \mathcal{E}(\mathbf{x})$, which the decoder $\mathcal{D} : \mathbb{R}^m \to \mathbb{R}^b$ uses to produce reconstructions $\widehat{\mathbf{x}} = \mathcal{D}(\mathbf{z})$ of the original input. AE models are typically trained in a *self-supervised* manner by minimizing a loss function $\mathcal{L}(\mathbf{x}, \widehat{\mathbf{x}})$ that measures the discrepancy between the input $\mathbf{x}$ and the reconstruction $\widehat{\mathbf{x}}$ (e.g., the mean squared error (MSE)). As the training of the model proceeds, the encoder progressively learns a latent representation that captures the most salient features of the input data, whereas the decoder learns how to recover the data back from the latent representation.

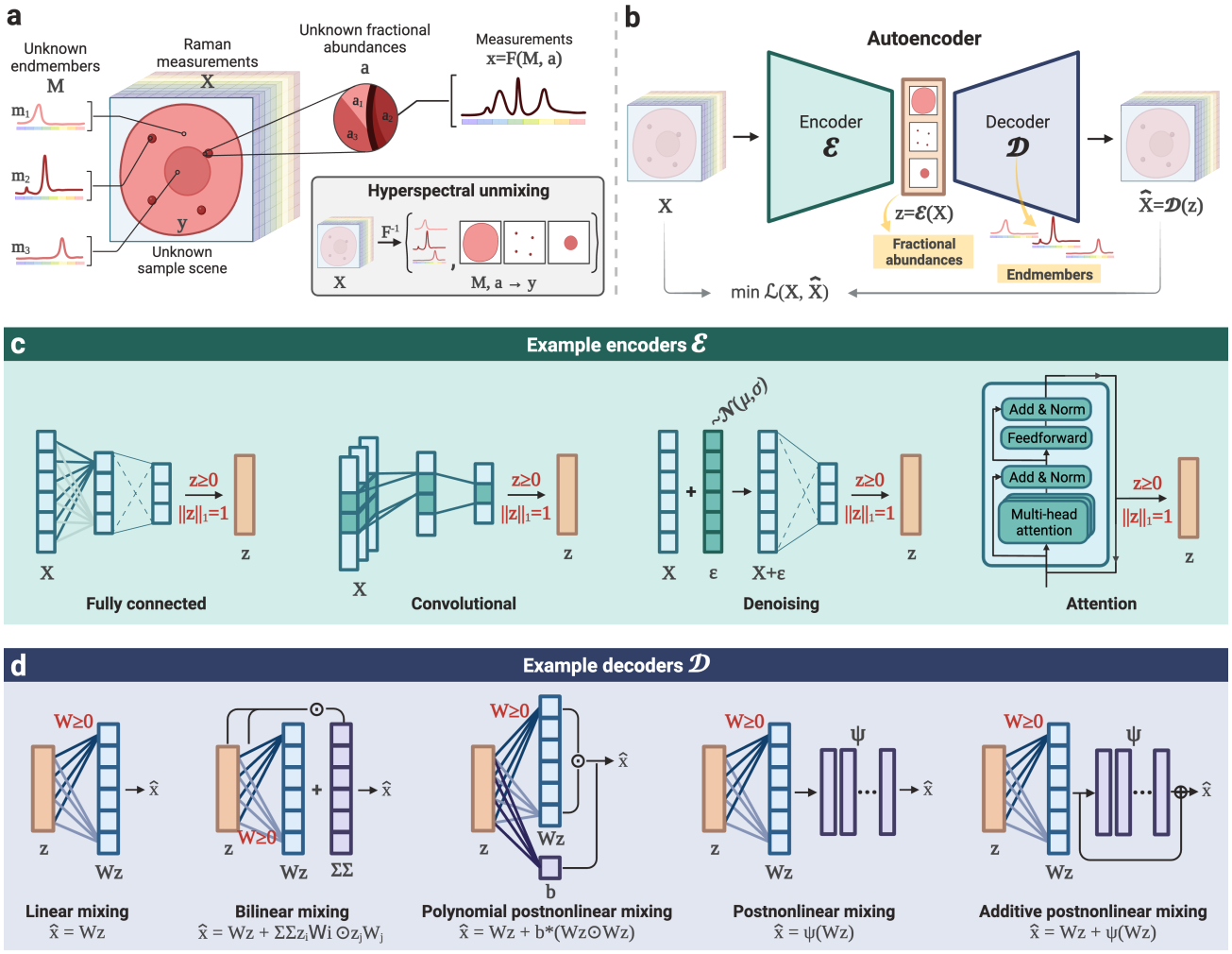

Figure 1. Hyperspectral unmixing for Raman spectroscopy using autoencoder neural networks. a, Diagram of the task of hyperspectral unmixing. b, Hyperspectral unmixing as a self-supervised autoencoder learning problem: the decoder learns to derive endmembers and the encoder learns the corresponding fractional abundances. c, Encoders can accommodate different concepts from representation learning, such as convolutional layers and attention, to improve feature extraction and provide more accurate and robust unmixing. d, Decoders can be structured to model different linear and non-linear mixing models. c-d, Labels in red indicate physics-inspired constraints.

Unmixing autoencoders. AEs have recently emerged as a framework to enhance the precision of hyperspectral unmixing in remote sensing, spurred by the availability of standardized benchmark datasets for model evaluation (e.g., *Urban*, *Samson*, *AVARIS Cuprite*) (Palsson et al., 2022; Zhang et al., 2020c; Wang et al., 2022; Bhatt & Joshi, 2020; Chen et al., 2022). Yet, despite initial explorations (Burzynski et al., 2021; Boildieu et al., 2023), the utility of unmixing AEs for Raman spectroscopy data remains largely unexplored.

3. Raman unmixing autoencoders

The dual functionality of autoencoders can be harnessed to design AE models for hyperspectral unmixing: the latent representations $\mathbf{z} = \mathcal{E}(\mathbf{x})$ can be interpreted as fractional abundances (with respect to an input spectrum $\mathbf{x}$), and the decoder $\mathcal{D}(\cdot)$ acts as a mixing function on these representations by encoding endmember signatures and other interactions. Hence, AE models learn to perform an 'unmixing' where the decoder identifies endmember signatures and the encoder quantifies the fractional abundances of these learned endmembers in the input spectrum (Figure 1b).

This setup provides a highly adaptable and versatile framework for unmixing, which can address many of the limitations of conventional techniques (see Table 1 in Appendix).

Encoder design. On the one hand, the learning of physical and biochemical features in the encoder can be enhanced by adopting strategies from representation learning, such as

convolutional layers to capture spectral and/or spatial correlations among neighboring bands and/or pixels (Zhang et al., 2018; Palsson et al., 2020; Elkholy et al., 2020), or attention mechanisms to model long-range dependencies (Ghosh et al., 2022) (Figure 1c). In addition, sparsity, part-based learning and denoising objectives can be adopted to enhance explainability and robustness (Ozkan et al., 2018; Qu & Qi, 2018; Su et al., 2018; 2017; Qu et al., 2017).

In this work, we consider four types of encoders encompassing a variety of architectures, from standard dense layers to more contemporary convolutional and attention mechanisms (see Appendix B for details): 1) an encoder consisting of fully connected layers (*Dense*); 2) an encoder with a 1D convolutional feature extractor block, followed by a fully connected part (*Convolutional*); 3) a transformer-based encoder that uses multi-head attention (*Transformer*) (Vaswani et al., 2017); and 4) a transformer-based encoder with a 1D convolutional feature extractor (*Convolutional Transformer*).

Decoder design. On the other hand, the design of the decoder allows for flexible modeling of input data, specifically to account for various mixture models, e.g., linear, bilinear and post-nonlinear (Figure 1d) (Chen et al., 2022; Shahid & Schizas, 2021; Zhao et al., 2021). This is akin to introducing an inductive prior with respect to the mixture model directly via the AE architecture.

To see that, consider a linear decoder $\mathcal{D}_{\text{Lin}}$ consisting of a single linear layer defined by a $b \times m$ weight matrix W. Under this setup, output reconstructions $\widehat{\mathbf{x}}$ are given by

$$\widehat{\mathbf{x}} = \mathcal{D}_{\text{Lin}}(\mathbf{z}) = W\mathbf{z}\,. \tag{3}$$

It follows from the formulation of the LMM (Equation (1)) that the latent representations $\mathbf{z}$ resemble the abundances $\boldsymbol{\alpha}$, the weight matrix W resembles the matrix of endmembers M, and the dimensionality m of the latent space defines the number n of endmembers to learn.

This framework can accommodate non-linear mixture models through the design of the decoder. For instance, the bilinear Fan model (Equation (2)) can be implemented by extending $\mathcal{D}_{\text{Lin}}$ to account for the additional bilinear terms:

$$\widehat{\mathbf{x}} = \mathcal{D}_{\text{Bilin}}(\mathbf{z}) = W\mathbf{z} + \sum_{k=1}^{m}\sum_{\substack{l=1,\\ l\neq k}}^{m} z_k\mathbf{w}_k \odot z_l\mathbf{w}_l, \tag{4}$$

where z_k, z_l are components of $\mathbf{z}$, and $\mathbf{w}_k, \mathbf{w}_l$ are column vectors of W. Similarly, one can devise decoders suited for other mixture models (Chen et al., 2022; Shahid & Schizas, 2021; Zhao et al., 2021), or adopt a general decoder that learns the underlying mixing model in a more data-driven manner, at the cost of interpretability of the extracted endmembers and fractional abundances.

Notice that decoders can be pre-initialized with a set of endmembers (e.g., an endmember library, or signatures derived using methods such as VCA), or readily adapted to non-blind unmixing by, respectively, pre-initializing or fixing the weight matrix W to predefined endmember signatures.

The two types of decoders we investigate are (see Appendix B for details): 1) a decoder designed for linear unmixing (Equation (3)); and 2) a decoder designed for bilinear unmixing (Equation (4)), both trained for non-blind unmixing without pre-initialization.

Physics-inspired constraints. To guide the AE learning and reinforce the physical interpretation of unmixing, we incorporate appropriate physical constraints into the AE architectures, e.g., non-negativity of endmembers and fractional abundances, and sum-to-one abundances. We enforce fractional abundance constraints through the choice of a latent space activation function. We use softmax to enforce both ANC and ASC; or, when interested in ANC only, a 'softly-rectified' hyperbolic tangent function given by $f(x) = \frac{1}{\gamma}\ln(1 + e^{\gamma * tanh(x)})$, $\gamma = 10$, designed to ensure abundances are non-negative (between 0 and 1) but do not necessarily add up to one. To ensure the non-negativity of endmembers, we constrain the weight matrix W in our decoders by clipping negative values to zero.

Model training. We train our AE models in a self-supervised fashion by minimizing a loss based on spectral angle divergence (SAD) (Kruse et al., 1993) that measures the cosine similarity between input and reconstructed spectra $\mathbf{m}_i$ and $\widehat{\mathbf{m}}_i$:

$$0 \leq \text{SAD}(\mathbf{m}_i, \widehat{\mathbf{m}}_i) = \arccos\left(\frac{\mathbf{m}_i \cdot \widehat{\mathbf{m}}_i}{\|\mathbf{m}_i\|_2\,\|\widehat{\mathbf{m}}_i\|_2}\right) \leq 1. \tag{5}$$

Notice that no ground truth information is provided during training as opposed to previously reported machine learning approaches for Raman hyperspectral analysis and characterization based on supervised learning (Manifold et al., 2021; Zhang et al., 2020b; Wei et al., 2022).

Model evaluation. We compare AE performance to conventional unmixing approaches: N-FINDR and VCA as endmember extraction algorithms followed by NNLS or FCLS to derive fractional abundances. When ground truth information is available, accuracy is quantified with respect to two measures - MSE for fractional abundances, and SAD for endmembers. For each evaluation, we first assign derived and ground truth endmembers (and corresponding abundances) via the Hungarian algorithm with SAD as the objective to minimize. When the number of extracted endmembers n is higher than the number of ground truth endmembers n_{true}, we only use n_{true} endmember and fractional abundance estimates to compute the performance metrics.

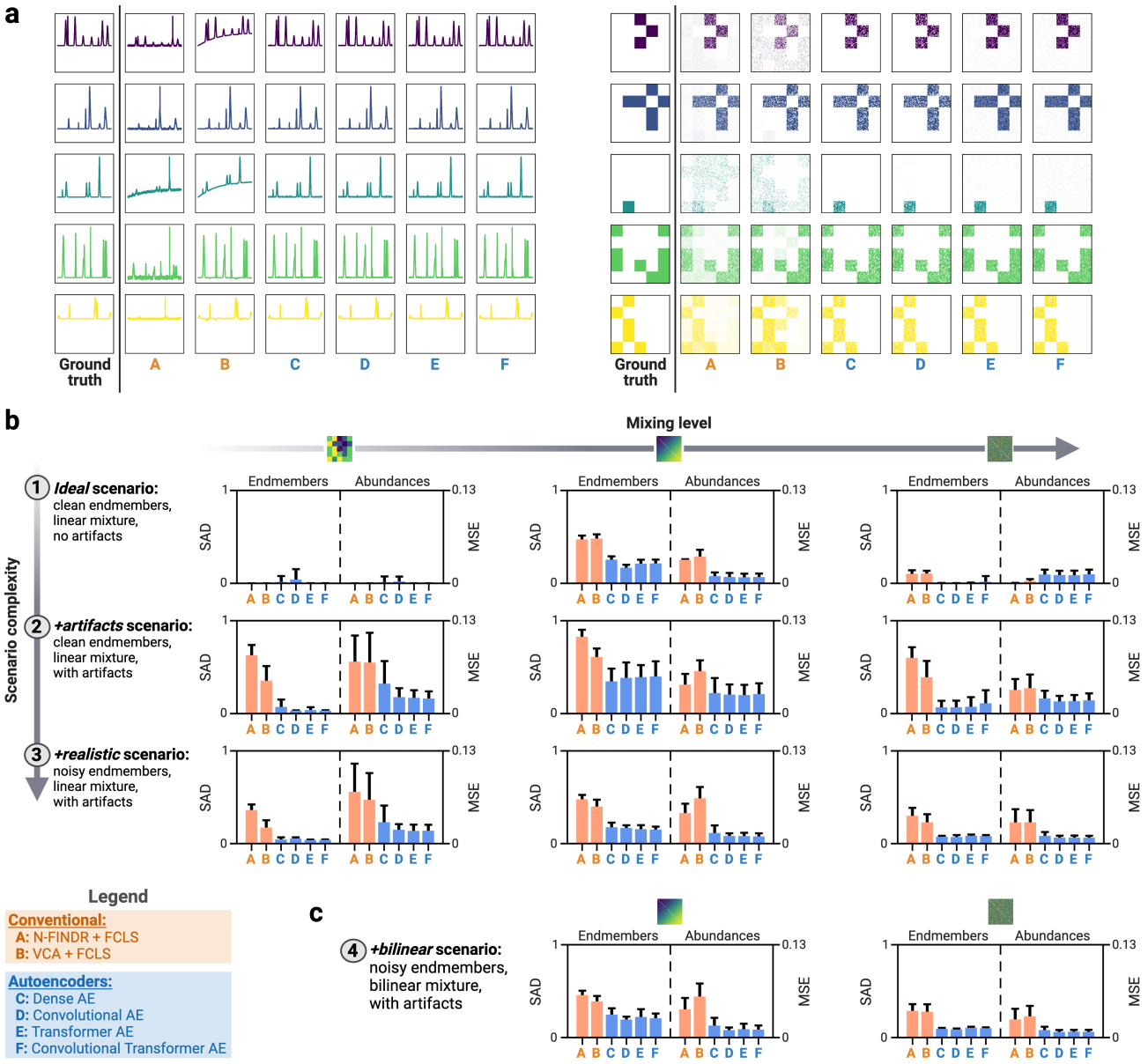

Figure 2. Benchmarking autoencoders on synthetic Raman mixtures. a, Representative results for the six algorithms (two conventional and four AEs) on an example synthetic dataset (*Chessboard+artifacts* scenario): endmembers (*left*), and fractional abundances (*right*). b-c, Summary of unmixing performance on synthetic datasets of variable mixing level and complexity: linear mixtures (b), bilinear mixtures (c). Confidence intervals are given by one standard deviation around the sample mean ($n = 25$ samples: 5 datasets with 5 model repetitions each).

4. Benchmarking unmixing autoencoders on synthetic Raman mixtures

We first benchmark the performance of our AE architectures on synthetic datasets created in-house.

Synthetic data generation. We developed a custom data generator that produces synthetic Raman mixtures with different characteristics (e.g., number and type of endmembers, abundance profiles, mixture model, data artifacts) with full knowledge of the 'ground truth' endmembers and fractional abundances (see Appendix C for data generation). This allows us to quantify and compare the performance of unmixing approaches (see Figure 2a for unmixing of an example synthetic dataset and Appendix D for experimental details).

Using our data generator, we produce 11 types of synthetic datasets of variable complexity, based on four mixture sce-

narios over three fractional abundance scenes. In order of complexity, the four mixture scenarios are: 1) a linear mixture with *clean* endmembers and no data artifacts (*ideal*); 2) a linear mixture with *clean* endmembers, but contaminated with artifacts representing dark noise, baseline variations and cosmic spikes (+*artifacts*); 3) a linear mixture with *noisy* endmembers (i.e., containing additional smaller noise peaks) and artifacts (+*realistic*); and 4) a bilinear mixture based on the Fan model (Equation (2)) with *noisy* endmembers and artifacts (+*bilinear*). For each of the four mixture scenarios, we generate three dataset variants (two for the +*bilinear* scenario since no bilinear interactions are present in our *Chessboard* scene) based on custom 100×100 fractional abundance scenes. This produces 10k spectra per dataset, organized into two-dimensional scenes for visualization purposes. In increasing level of mixing, we have: 1) a scene comprising well-separated patches, each containing a single species (*Chessboard* scene); 2) a semi-mixed scene given by a Gaussian mixture of species (*Gaussian* scene); and 3) a highly-mixed scene where each pixel represents a random sample of species drawn from a Dirichlet distribution (*Dirichlet* scene). Thus, our synthetic datasets cover varied mixing scenarios, from the *ideal Chessboard* dataset, which is trivial for conventional methods, to noisier, more complex mixtures containing different types of artifacts.

Benchmark results on linear mixtures. We first discuss our results on the nine dataset variants created through the linear mixture scenarios (1-3). Such data complies with the linear mixing assumption of conventional methods and, for consistency, we equip the AE models with a decoder for linear unmixing. Figure 2b summarises the performance of the six models (two conventional and four AEs) across the nine dataset variants, with experiments performed over 5 distinct datasets and 5 model initializations for each variant (refer to Table 2 in Appendix for calculated performance metrics). We find that the AE models outperform the two conventional methods, providing more accurate endmembers and fractional abundances across virtually all scenarios and abundance scenes. The AEs recover the performance of the conventional methods on the simple *ideal Chessboard* datasets, and the improvement in AE performance becomes increasingly prominent for mixture scenarios with higher levels of noise and data artifacts.

Non-linear unmixing with autoencoders. We next proceed to our benchmark analysis on synthetic data generated using a non-linear mixture model (i.e., +*bilinear* scenario). This time, we equip AEs with a decoder specific to the bilinear mixture model, which is achieved by merely adapting the decoder architecture. Our experimental results are displayed in Figure 2c. Again, we observe that all four AE models provide a substantial improvement in unmixing accuracy compared to standard unmixing methods for both endmember and abundance estimation.

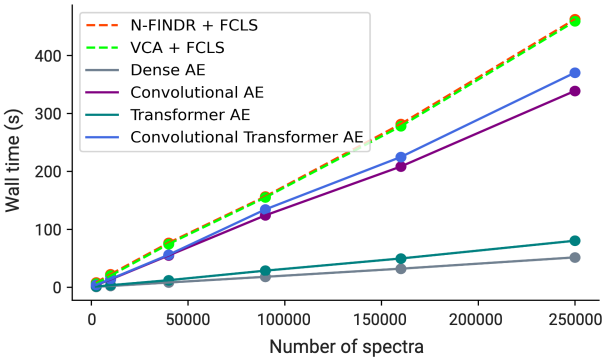

Figure 3. Computational efficiency of autoencoders and conventional methods on synthetic datasets with an increasing number of spectra. Each dot represents the average across 3 evaluations (confidence intervals based on one standard deviation are small and not visible to the eye). AE models are equipped with decoders for linear unmixing. Data generated under *Chessboard +artifacts*.

Computational efficiency. The computational complexity and scalability of unmixing methods can become a significant bottleneck in real-world applications, particularly for imaging and volumetric Raman scans, which can contain hundreds of thousands of spectra. To examine this issue, we profile the computational cost of our four AE methods (linear decoders) and the two conventional methods on synthetic datasets of increasing size up to 250000 spectra (refer to Appendix E for experimental details). To be fair to conventional algorithms, we include the full training time for autoencoders and use CPU computation to avoid any advantage from GPU acceleration. Figure 3 shows that all AE models are faster than N-FINDR+FCLS and VCA+FCLS, which are already among the most computationally lightweight conventional unmixing techniques (Bioucas-Dias et al., 2012).

5. Validation on experimental Raman data from sugar mixtures

To validate the unmixing performance of AEs on real experimental data, we next conduct benchmark analyses on data from a library of 240 sugar mixtures prepared in-house with four types of sugar (glucose, sucrose, fructose, maltose) at different concentrations (Figure 4a). To consider different signal-to-noise (SNR) conditions, we acquired high SNR (1920 spectra) and low SNR (7680 spectra) measurements using a custom Raman microspectroscopy platform at integration times of 5 s and 0.5 s, respectively. We used these experimental datasets with ground truth to evaluate the performance of unmixing algorithms under typical experimental artifacts, such as baseline shifts, environmental

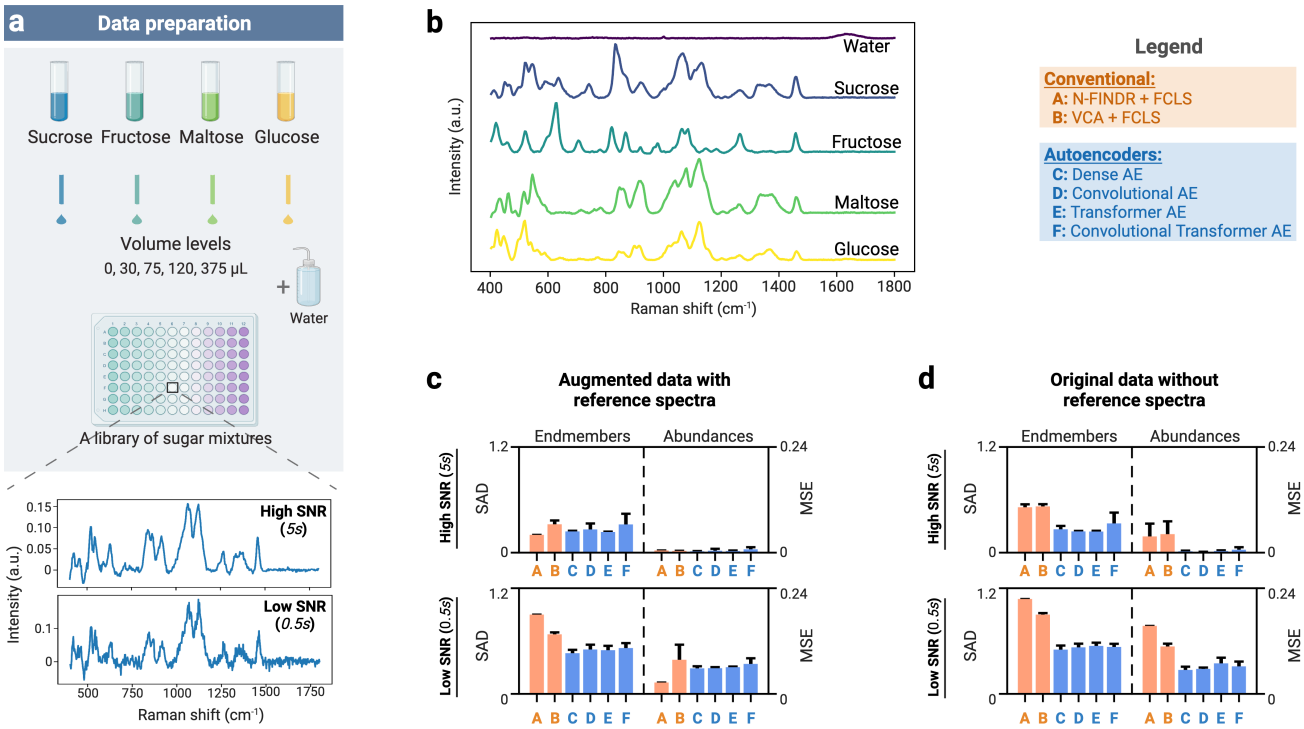

Figure 4. Experimental validation on Raman spectroscopy data from sugar solutions. a, Schematic diagram of sugar mixture preparation. Two sets of data are acquired—high and low signal-to-noise ratio (SNR) data, by using integration times of 5 s and 0.5 s, respectively. b, Endmember signatures estimated from reference spectra (high SNR) additionally collected from pure solutions. c-d, Summary of unmixing performance for: an idealized scenario with augmented data including reference spectra (c); and original data without augmentation (d). Confidence intervals are given as one standard deviation around the sample mean ($n = 5$).

noise, and cosmic spikes. Full experimental details, including sample preparation and data acquisition, are provided in Appendix F.

We perform unmixing on these data to identify the content of each mixture, i.e., types of sugar and their concentrations. The ground truth is defined by the experimental concentrations and the endmember signatures we obtain from reference spectra measured from 5 additional pure solutions (Figure 4b). As with the synthetic data above, we benchmark the performance of our four AE models (linear decoders) against N-FINDR+FCLS and VCA+FCLS.

Data with reference spectra. First, we consider an idealized scenario, purposefully devised to favor conventional methods, whereby endmembers are present in the data. To do this, we augment our data with the additional reference spectra we measured. When such 'pure pixels' are available, we observe that conventional methods (NFINDR+FCLS, VCA+FCLS) perform comparably to AEs on clean, high SNR data (Figure 4c). Yet, AEs already provide improved performance in low SNR regimes.

Data without reference spectra. In many experimental applications, however, the underlying endmembers are not present in the data and cannot be separately obtained (e.g., target-agnostic applications, or unknown species). To consider such cases, we analyzed our original data without augmentation. Our results in Figure 4d demonstrate that, in such situations, AEs substantially outperform conventional methods in both low and high SNR settings (see Figures 7-8 in Appendix for qualitative results, and Table 3 for performance metrics).

6. Application to biological data: volumetric Raman imaging of a THP-1 cell

As an application to biological research, we use unmixing autoencoders to analyze a low-SNR volumetric RS raster scan of a human leukemia monocytic (THP-1) cell (Figure 5a) (Kallepitis et al., 2017). Using Raman chemometrics, the composition of the cell is probed to study its morphology in a nondestructive, label-free manner.

After loading and preprocessing the data using *RamanSPy* (Georgiev et al., 2024), we conduct unmixing with: 1) VCA+NNLS - as in the original paper; 2) *Dense AE* - our simplest and most computationally efficient AE model; and 3) *Deep Dense AE* - an extension of *Dense AE* with a deeper

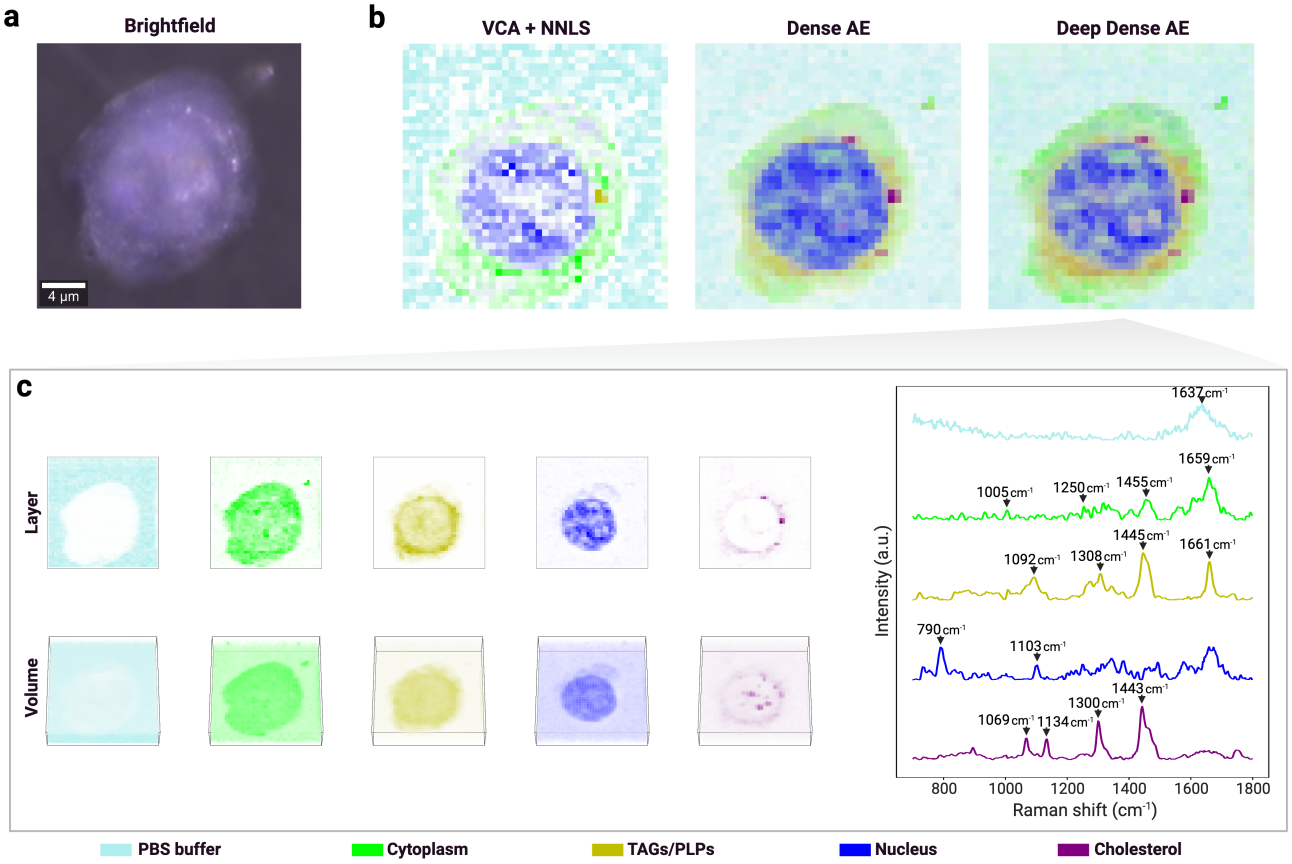

Figure 5. Analysis of volumetric Raman imaging of a THP-1 cell with unmixing autoencoders. a, A brightfield image of the studied THP-1 cell. b, A cross-section reconstruction of the cell (layer $z = 7$) obtained by overlaying the fractional abundances derived by: VCA+NNLS, our *Dense AE*, and a deeper *Dense AE*. c, Results obtained with our deeper *Dense AE* model, displaying the spatial distribution of the individual fractional abundances and the associated endmember signatures. Fractional abundance maps normalized for consistent visualization. Brightfield and Raman data from Kallepitis et al. (2017).

encoder with five layers. We derive 20 endmembers, which we characterize via peak assignment to identify biochemical species present in the scanned cell, such as deoxyribonucleic acid (DNA), proteins, triglycerides (TAGs), phospholipids (PLPs) and cholesterol esters (see Appendix G for experimental details and characterization).

Figure 5b shows the reconstructions of the cell created by overlaying selected fractional abundances derived by each method, revealing the spatial organization of key cellular organelles, including the nucleus, cytoplasm, lipid bodies and membranes. Although direct comparisons are challenging due to the lack of ground truth, the unmixing results of our AE models are aligned with the original findings (Kallepitis et al., 2017), albeit with more distinct endmember signatures and well-defined abundance features (see Figures 9-11 in Appendix for full results). In particular, we observe that our *Deep Dense AE* model provides cleaner endmember signatures and enables more precise spectral and compositional information (Figure 5c). Notably, unlike the original VCA+NNLS approach, our AEs detect cholesterol, an important functional and structural component in cells, which plays a key role in membrane fluidity and stability, signaling pathways, and immune response (Kritharides et al., 1998; Tall & Yvan-Charvet, 2015; Saha et al., 2017).

7. Conclusion

In summary, we have presented an autoencoder-based methodology for hyperspectral unmixing in Raman spectroscopy, which we validated on a wide array of synthetic and experimental datasets. Our results demonstrate that autoencoders are adept at handling diverse mixture scenarios and exhibit robustness against data artifacts, offering an effective, versatile and efficient framework for RS unmixing.

The potential of autoencoders for RS unmixing opens several avenues for future research. One direction is the investigation of AE architectures with more complex decoders (Chen et al., 2022) and/or encoders, as well as the

development of training objectives that further improve robustness to spectral variations (e.g., stacked and denoising AE architectures (Qu et al., 2017; Su et al., 2018)). Another promising area is the use of AEs as a pretraining procedure in downstream tasks, potentially combined with other AI-based approaches (e.g., deep learning models for denoising (Horgan et al., 2021)).

Finally, while our focus here is on RS, we wish to underscore the applicability of our work to other spectroscopic modalities, including infrared spectroscopy.

Acknowledgements

D.G. and G.P. are supported by UK Research and Innovation [UKRI Centre for Doctoral Training in AI for Healthcare grant number EP/S023283/1]. A.F.G. acknowledges support from the Schmidt Science Fellows, in partnership with the Rhodes Trust. S.V.P. acknowledges support from the Independent Research Fund Denmark (0170-00011B). R.X. and M.M.S. acknowledge support from the Engineering and Physical Sciences Research Council (EP/P00114/1 and EP/T020792/1). M.M.S. acknowledges support from the Royal Academy of Engineering Chair in Emerging Technologies award (CiET2021\94) and the Bio Innovation Institute. M.B. acknowledges support from the Engineering and Physical Sciences Research Council (EP/N014529/1, funding the EPSRC Centre for Mathematics of Precision Healthcare at Imperial, and EP/T027258/1). The authors thank Dr Akemi Nogiwa Valdez for proofreading and data management support. Figures were assembled in BioRender.

References

Altmann, Y., Halimi, A., Dobigeon, N., and Tourneret, J.-Y. Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery. *IEEE Transactions on Image Processing*, 21(6):3017–3025, 2012.

Auner, G. W., Koya, S. K., Huang, C., Broadbent, B., Trexler, M., Auner, Z., Elias, A., Mehne, K. C., and Brusatori, M. A. Applications of Raman spectroscopy in cancer diagnosis. *Cancer and Metastasis Reviews*, 37(4): 691–717, 2018.

Ba, J. L., Kiros, J. R., and Hinton, G. E. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

Balan, V., Mihai, C.-T., Cojocaru, F.-D., Uritu, C.-M., Dodi, G., Botezat, D., and Gardikiotis, I. Vibrational spectroscopy fingerprinting in medicine: from molecular to clinical practice. *Materials*, 12(18):2884, 2019.

Bhatt, J. S. and Joshi, M. V. Deep learning in hyperspectral unmixing: A review. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2189–2192. IEEE, 2020.

Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., and Chanussot, J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(2):354–379, 2012.

Bocklitz, T. W., Guo, S., Ryabchykov, O., Vogler, N., and Popp, J. Raman Based Molecular Imaging and Analytics: A Magic Bullet for Biomedical Applications!? *Analytical chemistry*, 88(1):133–151, 2016.

Boildieu, D., Helbert, D., Magnaudeix, A., Leproux, P., and Carré, P. Multivariate curve resolution with autoencoders for CARS microspectroscopy. In *Computational Imaging Conference, IS&T Electronic Imaging*, 2023.

Burzynski, N., Yuan, Y., Felsen, A., Reitano, D., Wang, Z., Sethi, K. A., Lu, F., and Chiu, K. Deep Learning Techniques for Unmixing of Hyperspectral Stimulated Raman Scattering Images. In *2021 IEEE International Conference on Big Data (Big Data)*, pp. 5862–5864. IEEE, 2021.

Butler, H. J., Ashton, L., Bird, B., Cinque, G., Curtis, K., Dorney, J., Esmonde-White, K., Fullwood, N. J., Gardner, B., Martin-Hirsch, P. L., et al. Using Raman spectroscopy to characterize biological materials. *Nature protocols*, 11 (4):664–687, 2016.

Byrne, H. J., Knief, P., Keating, M. E., and Bonnier, F. Spectral pre and post processing for infrared and Raman spectroscopy of biological tissues and cells. *Chemical Society Reviews*, 45(7):1865–1878, 2016.

Chalmers, J. M., Edwards, H. G. M., and Hargreaves, M. D. (eds.). *Infrared and Raman Spectroscopy in Forensic Science*. Wiley Online Library, 2012.

Chen, J., Zhao, M., Wang, X., Richard, C., and Rahardja, S. Integration of physics-based and data-driven models for hyperspectral image unmixing. *arXiv preprint arXiv:2206.05508*, 2022.

Cialla-May, D., Zheng, X.-S., Weber, K., and Popp, J. Recent progress in surface-enhanced Raman spectroscopy for biological and biomedical applications: from cells to clinics. *Chemical Society Reviews*, 46(13):3945–3961, 2017.

Colthup, N. *Introduction to infrared and Raman spectroscopy*. Elsevier, 2012.

Dobigeon, N., Tourneret, J.-Y., Richard, C., Bermudez, J. C. M., McLaughlin, S., and Hero, A. O. Nonlinear unmixing of hyperspectral images: Models and algorithms. *IEEE Signal processing magazine*, 31(1):82–94, 2013.

Dodo, K., Fujita, K., and Sodeoka, M. Raman spectroscopy for chemical biology research. *Journal of the American Chemical Society*, 144(43):19651–19667, 2022.

Eilers, P. H. and Boelens, H. F. Baseline correction with asymmetric least squares smoothing. *Leiden University Medical Centre Report*, 1(1):5, 2005.

Elkholy, M. M., Mostafa, M., Ebied, H. M., and Tolba, M. F. Hyperspectral unmixing using deep convolutional autoencoder. *International Journal of Remote Sensing*, 41(12):4799–4819, 2020.

Ember, K. J., Hoeve, M. A., McAughtrie, S. L., Bergholt, M. S., Dwyer, B. J., Stevens, M. M., Faulds, K., Forbes, S. J., and Campbell, C. J. Raman spectroscopy and regenerative medicine: a review. *NPJ Regenerative medicine*, 2(1):12, 2017.

Fan, W., Hu, B., Miller, J., and Li, M. Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data. *International Journal of Remote Sensing*, 30(11): 2951–2962, 2009.

Fernández-Galiana, Á., Bibikova, O., Pedersen, S. V., and Stevens, M. M. Fundamentals and Applications of Raman-Based Techniques for the Design and Development of Active Biomedical Materials. *Advanced Materials*, pp. 2210807, 2023.

Gautam, R., Vanga, S., Ariese, F., and Umapathy, S. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Techniques and Instrumentation*, 2:1–38, 2015.

Georgiev, D., Pedersen, S. V., Xie, R., Fernández-Galiana, A., Stevens, M. M., and Barahona, M. RamanSPy: An Open-Source Python Package for Integrative Raman Spectroscopy Data Analysis. *Analytical Chemistry*, 2024.

Ghosh, P., Roy, S. K., Koirala, B., Rasti, B., and Scheunders, P. Hyperspectral unmixing using transformer network. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2022.

Goodfellow, I., Bengio, Y., and Courville, A. Autoencoders. In *Deep learning*, chapter 14, pp. 499–523. MIT Press, 2016.

Guo, S., Popp, J., and Bocklitz, T. Chemometric analysis in Raman spectroscopy from experimental design to machine learning–based modeling. *Nature Protocols*, 16 (12):5426–5459, 2021.

Halimi, A., Altmann, Y., Dobigeon, N., and Tourneret, J.-Y. Nonlinear Unmixing of Hyperspectral Images Using a Generalized Bilinear Model. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4153–4162, 2011a. doi: 10.1109/TGRS.2010.2098414.

Halimi, A., Altmann, Y., Dobigeon, N., and Tourneret, J.-Y. Unmixing hyperspectral images using the generalized bilinear model. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1886–1889. IEEE, 2011b.

Halvorson, R. A. and Vikesland, P. J. Surface-enhanced Raman spectroscopy (SERS) for environmental analyses, 2010.

Hapke, B. Bidirectional reflectance spectroscopy: 1. theory. *Journal of Geophysical Research: Solid Earth*, 86(B4): 3039–3054, 1981.

Harris, A. T. Spectral mapping tools from the earth sciences applied to spectral microscopy data. *Cytometry Part A: The Journal of the International Society for Analytical Cytology*, 69(8):872–879, 2006.

Hedegaard, M., Matthäus, C., Hassing, S., Krafft, C., Diem, M., and Popp, J. Spectral unmixing and clustering algorithms for assessment of single cells by Raman microscopic imaging. *Theoretical Chemistry Accounts*, 130: 1249–1260, 2011.

Heinz, D. C. et al. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE transactions on geoscience and remote sensing*, 39(3):529–545, 2001.

Heylen, R. and Scheunders, P. A multilinear mixing model for nonlinear spectral unmixing. *IEEE transactions on geoscience and remote sensing*, 54(1):240–251, 2015.

Heylen, R., Parente, M., and Gader, P. A review of nonlinear hyperspectral unmixing methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):1844–1868, 2014.

Horgan, C. C., Jensen, M., Nagelkerke, A., St-Pierre, J.-P., Vercauteren, T., Stevens, M. M., and Bergholt, M. S. High-throughput molecular imaging via deep-learning-enabled Raman spectroscopy. *Analytical chemistry*, 93 (48):15850–15860, 2021.

Izake, E. L. Forensic and homeland security applications of modern portable Raman spectroscopy. *Forensic science international*, 202(1-3):1–8, 2010.

Jones, R. R., Hooper, D. C., Zhang, L., Wolverson, D., and Valev, V. K. Raman techniques: fundamentals and frontiers. *Nanoscale research letters*, 14:1–34, 2019.

Kallepitis, C., Bergholt, M. S., Mazo, M. M., Leonardo, V., Skaalure, S. C., Maynard, S. A., and Stevens, M. M. Quantitative volumetric Raman imaging of three dimensional cell cultures. *Nature communications*, 8(1):14843, 2017.

Keshava, N. and Mustard, J. F. Spectral unmixing. *IEEE signal processing magazine*, 19(1):44–57, 2002.

Khandasammy, S. R., Fikiet, M. A., Mistek, E., Ahmed, Y., Halámková, L., Bueno, J., and Lednev, I. K. Bloodstains, paintings, and drugs: Raman spectroscopy applications in forensic science. *Forensic Chemistry*, 8:111–133, 2018.

Kong, K., Kendall, C., Stone, N., and Notingher, I. Raman spectroscopy for medical diagnostics—From in-vitro biofluid assays to in-vivo cancer detection. *Advanced drug delivery reviews*, 89:121–134, 2015.

Koningstein, J. A. *Introduction to the Theory of the Raman Effect*. Springer Science & Business Media, 2012.

Kritharides, L., Christian, A., Stoudt, G., Morel, D., and Rothblat, G. H. Cholesterol metabolism and efflux in human THP-1 macrophages. *Arteriosclerosis, thrombosis, and vascular biology*, 18(10):1589–1599, 1998.

Kruse, F. A., Lefkoff, A., Boardman, J., Heidebrecht, K., Shapiro, A., Barloon, P., and Goetz, A. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, 44(2-3):145–163, 1993.

Kumar, C. S. *Raman spectroscopy for nanomaterials characterization*. Springer Science & Business Media, 2012.

Lawson, C. L. and Hanson, R. J. *Solving least squares problems*. SIAM, 1995.

Li, X., Zhou, R., Xu, Y., Wei, X., and He, Y. Spectral unmixing combined with Raman imaging, a preferable analytic technique for molecule visualization. *Applied Spectroscopy Reviews*, 52(5):417–438, 2017.

Li, Y.-S. and Church, J. S. Raman spectroscopy in the analysis of food and pharmaceutical nanomaterials. *Journal of food and drug analysis*, 22(1):29–48, 2014.

Mahadevan-Jansen, A. and Richards-Kortum, R. R. Raman spectroscopy for the detection of cancers and precancers. *Journal of biomedical optics*, 1(1):31–70, 1996.

Manifold, B., Men, S., Hu, R., and Fu, D. A versatile deep learning architecture for classification and label-free prediction of hyperspectral images. *Nature machine intelligence*, 3(4):306–315, 2021.

McCreery, R. L. *Raman spectroscopy for chemical analysis*. John Wiley & Sons, 2005.

Movasaghi, Z., Rehman, S., and Rehman, I. U. Raman spectroscopy of biological tissues. *Applied Spectroscopy Reviews*, 42(5):493–541, 2007.

Nascimento, J. M. and Dias, J. M. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE transactions on Geoscience and Remote Sensing*, 43(4): 898–910, 2005.

Olmos, V., Benitez, L., Marro, M., Loza-Alvarez, P., Pina, B., Tauler, R., and De Juan, A. Relevant aspects of unmixing/resolution analysis for the interpretation of biological vibrational hyperspectral images. *TrAC Trends in Analytical Chemistry*, 94:130–140, 2017.

Ong, T. T., Blanch, E. W., and Jones, O. A. Surface Enhanced Raman Spectroscopy in environmental analysis, monitoring and assessment. *Science of The Total Environment*, 720:137601, 2020.

Ozkan, S., Kaya, B., and Akar, G. B. Endnet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1):482–496, 2018.

Palsson, B., Ulfarsson, M. O., and Sveinsson, J. R. Convolutional autoencoder for spectral–spatial hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 59(1):535–549, 2020.

Palsson, B., Sveinsson, J. R., and Ulfarsson, M. O. Blind hyperspectral unmixing using autoencoders: A critical comparison. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:1340–1372, 2022.

Pang, S., Yang, T., and He, L. Review of surface enhanced Raman spectroscopic (SERS) detection of synthetic chemical pesticides. *TrAC Trends in Analytical Chemistry*, 85: 73–82, 2016.

Paudel, A., Raijada, D., and Rantanen, J. Raman spectroscopy in pharmaceutical product design. *Advanced drug delivery reviews*, 89:3–20, 2015.

Pence, I. and Mahadevan-Jansen, A. Clinical instrumentation and applications of Raman spectroscopy. *Chemical Society Reviews*, 45(7):1958–1979, 2016.

Pezzotti, G. Raman spectroscopy in cell biology and microbiology. *Journal of Raman Spectroscopy*, 52(12):2348–2443, 2021.

Qu, Y. and Qi, H. uDAS: An untied denoising autoencoder with sparsity for spectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 57(3):1698–1712, 2018.

Qu, Y., Guo, R., and Qi, H. Spectral unmixing through part-based non-negative constraint denoising autoencoder. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 209–212. IEEE, 2017.

Ryabchykov, O., Guo, S., and Bocklitz, T. Analyzing Raman spectroscopic data. *Physical Sciences Reviews*, 4(2): 20170043, 2018.

Saha, A. K., Dallo, S. F., Detmar, A. L., Osmulski, P., Gaczynska, M., Huang, T. H.-M., and Ramasubramanian, A. K. Cellular cholesterol regulates monocyte deformation. *Journal of biomechanics*, 52:83–88, 2017.

Savitzky, A. and Golay, M. J. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.

Schlücker, S. Surface-Enhanced Raman spectroscopy: Concepts and chemical applications. *Angewandte Chemie International Edition*, 53(19):4756–4795, 2014.

Shahid, K. T. and Schizas, I. D. Unsupervised hyperspectral unmixing via nonlinear autoencoders. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2021.

Shipp, D. W., Sinjab, F., and Notingher, I. Raman spectroscopy: techniques and applications in the life sciences. *Advances in Optics and Photonics*, 9(2):315–428, 2017.

Smith, E. and Dent, G. *Modern Raman spectroscopy: a practical approach*. John Wiley & Sons, 2019.

Smith, R., Wright, K. L., and Ashton, L. Raman spectroscopy: an evolving technique for live cell studies. *Analyst*, 141(12):3590–3600, 2016.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

Su, Y., Marinoni, A., Li, J., Plaza, A., and Gamba, P. Nonnegative sparse autoencoder for robust endmember extraction from remotely sensed hyperspectral images. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 205–208. IEEE, 2017.

Su, Y., Marinoni, A., Li, J., Plaza, J., and Gamba, P. Stacked nonnegative sparse autoencoders for robust hyperspectral unmixing. *IEEE Geoscience and Remote Sensing Letters*, 15(9):1427–1431, 2018.

Szymanski, H. A. *Raman spectroscopy: theory and practice*. Springer Science & Business Media, 2012.

Talari, A. C. S., Movasaghi, Z., Rehman, S., and Rehman, I. U. Raman spectroscopy of biological tissues. *Applied spectroscopy reviews*, 50(1):46–111, 2015.

Tall, A. R. and Yvan-Charvet, L. Cholesterol, inflammation and innate immunity. *Nature Reviews Immunology*, 15 (2):104–116, 2015.

Tanwar, S., Paidi, S. K., Prasad, R., Pandey, R., and Barman, I. Advancing Raman spectroscopy from research to clinic: Translational potential and challenges. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 260:119957, 2021.

Terry, L. R., Sanders, S., Potoff, R. H., Kruel, J. W., Jain, M., and Guo, H. Applications of surface-enhanced Raman spectroscopy in environmental detection. *Analytical Science Advances*, 3(3-4):113–145, 2022.

Vankeirsbilck, T., Vercauteren, A., Baeyens, W., Van der Weken, G., Verpoort, F., Vergote, G., and Remon, J. Applications of Raman spectroscopy in pharmaceutical analysis. *TrAC Trends in Analytical Chemistry*, 21(12):869–877, 2002. ISSN 0165-9936. doi: https://doi.org/10.1016/S0165-9936(02)01208-6. URL https://www.sciencedirect.com/science/article/pii/S0165993602012086.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Wang, W.-t., Zhang, H., Yuan, Y., Guo, Y., and He, S.-x. Research progress of Raman spectroscopy in drug analysis. *Aaps Pharmscitech*, 19(7):2921–2928, 2018.

Wang, Y., Albrecht, C. M., Braham, N. A. A., Mou, L., and Zhu, X. X. Self-supervised learning in remote sensing: A review. *arXiv preprint arXiv:2206.13188*, 2022.

Weber, W. H. and Merlin, R. *Raman scattering in materials science*, volume 42. Springer Science & Business Media, 2013.

Wei, Z., Liu, X., Yan, R., Sun, G., Yu, W., Liu, Q., and Guo, Q. Pixel-level multimodal fusion deep networks for predicting subcellular organelle localization from label-free live-cell imaging. *Frontiers in Genetics*, 13:1002327, 2022.

Whitaker, D. A. and Hayes, K. A simple algorithm for despiking Raman spectra. *Chemometrics and Intelligent Laboratory Systems*, 179:82–84, 2018.

Winter, M. E. N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *Imaging Spectrometry V*, volume 3753, pp. 266–275. SPIE, 1999.

Zhang, F., Tang, X., Tong, A., Wang, B., Wang, J., Lv, Y., Tang, C., and Wang, J. Baseline correction for infrared

spectra using adaptive smoothness parameter penalized least squares method. *Spectroscopy Letters*, 53(3):222–233, 2020a.

Zhang, J., Zhao, J., Lin, H., Tan, Y., and Cheng, J.-X. High-Speed Chemical Imaging by Dense-Net Learning of Femtosecond Stimulated Raman Scattering. *The journal of physical chemistry letters*, 11(20):8573–8578, 2020b.

Zhang, S., Su, Y., Xu, X., Li, J., Deng, C., and Plaza, A. Recent advances in hyperspectral unmixing using sparse techniques and deep learning. *Hyperspectral Image Analysis: Advances in Machine Learning and Signal Processing*, pp. 377–405, 2020c.

Zhang, X., Roeffaers, M. B., Basu, S., Daniele, J. R., Fu, D., Freudiger, C. W., Holtom, G. R., and Xie, X. S. Label-free live-cell imaging of nucleic acids using stimulated Raman scattering microscopy. *ChemPhysChem*, 13(4): 1054–1059, 2012.

Zhang, X., Sun, Y., Zhang, J., Wu, P., and Jiao, L. Hyperspectral unmixing via deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 15(11): 1755–1759, 2018.

Zhao, M., Wang, M., Chen, J., and Rahardja, S. Hyperspectral Unmixing for Additive Nonlinear Models With a 3-D-CNN Autoencoder Network. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2021.

A. Standard methods for hyperspectral unmixing

N-FINDR and VCA are geometric methods based on the concept of a simplex in Euclidean space. N-FINDR exploits the fact that, under the linear mixing model (Equation (1)), endmembers represent vertices of a simplex spanning the data, and operates by iteratively finding a set of points (endmembers) that maximizes the volume of the simplex they form. In contrast, VCA finds endmembers by projecting the data onto directions orthogonal to the subspace spanned by previously found endmembers and identifying new endmembers as the farthest points in these directions, effectively constructing a simplex that encompasses all data points. In both methods, the number of endmembers to extract is specified *a priori* by the user. Once endmember signatures M are derived, optimization-based algorithms such as NNLS and FCLS are employed to estimate the fractional abundances $\boldsymbol{\alpha}$ for a given spectrum $\mathbf{x}$ by minimizing the reconstruction error between the observed data and the model $\min_{\boldsymbol{\alpha}} \|M\boldsymbol{\alpha} - \mathbf{x}\|^2$. NNLS imposes the ANC, whereas FCLS imposes both the ANC and ASC.

B. Autoencoder architectures

Dense AE. This autoencoder employs an encoder comprising 2 fully connected (or dense) layers. The first layer projects spectra of dimension b to hidden features of dimension 128 (Leaky ReLU activation with a slope of 0.02), which the second layer further projects to latent representations of dimension n (n is the number of endmembers to extract). In the *Deep Dense AE* model used in the analysis of the THP-1 cell, we increase the number of hidden layers to five, comprising 512, 256, 128, 64 and 32 neurons, respectively, before the final layer of size n.

Convolutional AE. This model extends the *Dense AE* by adding a convolutional block before the dense layers. The convolutional block consists of two layers of 1D convolutions connected in parallel, each comprising 16 filters of size 3 and 16 filters of size 5 (ReLU activation; input padded with zeroes). The outputs from these two layers are concatenated and merged (channel-wise) via a 2-dimensional dense layer to produce representations of dimension b, which are then fed to the *Dense* encoder described above.

Transformer AE. In this transformer-based encoder, input spectra are first projected to features of size 32 through a fully connected layer, and then fed to a transformer encoder layer comprising a multi-head attention block with 2 attention heads of size 32 (Vaswani et al., 2017), followed by two fully connected layers expanding the features to size 64 (ReLU activation) and condensing back to 32 (no activation). We apply layer normalization (Ba et al., 2016) and dropout (10%) (Srivastava et al., 2014) after the multi-head attention block and the fully connected layers. The output of the transformer block is then channeled into the last fully connected layer of size n.

Convolutional Transformer AE. In this model, the *Transformer AE* architecture is extended with the same convolutional block used in the *Convolutional AE*, here added before the transformer-based encoder block.

Decoder choice. Our linear unmixing decoder architecture consists of a single fully connected layer using the identity activation function without bias (Equation (3)). Our bilinear Fan decoder has the same architecture as the linear decoder but also calculates the additional bilinear interaction terms in line with Equation (4) during each forward pass.

C. Generating synthetic Raman mixtures

Generating endmembers. For each synthetic dataset, we first generate n endmembers spanning b spectral bands. For the scope of this work, $n = 5$ and $b = 1000$. Each endmember $\mathbf{m}_i \in \mathbb{R}_+^b$ is created by a superposition of a set of $n_{\text{peaks},i}$ Gaussian peaks of different amplitude, width and location, randomly sampled as follows. The number of peaks is sampled from a discrete uniform distribution $n_{\text{peaks},i} \sim \mathcal{U}(5, 9)$. Each peak p is described by $p = h_p \sigma_p \sqrt{2\pi}\, \mathcal{N}(b_p, \sigma_p)$, where $\mathcal{N}(\cdot)$ represents a Gaussian distribution. The height of the peak is defined as $h_p = h_1 \cdot h_2$, where $h_1 = 1 + 5\, h_\beta$ with $h_\beta \sim \text{Beta}(1, 3)$ and $h_2 \sim \mathcal{U}(0.1, 1)$. The center of the peak is sampled from $b_p \sim \mathcal{U}(10, b - 10)$, and the width of the peak is defined as $\sigma_p = w_p \sigma$, with $\sigma \sim \mathcal{U}(0.1, 1)$. We create two types of endmembers: *clean* and *noisy*. For the former, we produce peaks with $w_p = 1$. For the latter, we augment *clean* endmembers by adding $n_{\text{peaks},i}^{\text{small}} \sim \mathcal{U}(50, 99)$ smaller peaks sampled with $h_1 = 1/3$ and $w_p = 2$, thus making *noisy* endmembers better resemble experimental Raman signatures.

Generating fractional abundances. For visualization purposes, we present the fractional abundance profiles in the form of two-dimensional scenes comprising $H \times W$ pixels, where each pixel represents a fractional abundance vector $\boldsymbol{\alpha} \in \mathbb{R}_+^n$.

Table 1. Advantages of autoencoders over standard unmixing techniques.

	Commonly used methods	**Autoencoders**
Assumptions	Strict - e.g. endmembers present in the data; fixed mixture model; number of endmembers is known.	Autoencoders can facilitate data-driven unmixing without many strong assumptions about the data and task.
Mixture model	A specific model is assumed *a priori*. Usually, a linear mixing only, as alternatives can become too complex and computationally expensive.	Autoencoders can be designed to capture various mixture models, including non-linear mixtures. This can be enforced by appropriate architectural constraints or learned in a data-driven manner.
Input modality & capturing spatial information	Spectra are processed individually. If imaging or volumetric data are provided, these are typically unfolded and used pixel-wise, thereby discarding any spatial information.	Autoencoders can be extended to utilize the spatial information available in imaging or volumetric data by incorporating (2D, 3D, and/or 4D) convolutional layers.
Simultaneously deriving endmembers and fractional abundances?	Typically, two separate algorithms are applied.	Autoencoders simultaneously extract both endmembers and fractional abundances by default. They can also be adjusted for non-blind unmixing by fixing a pre-defined endmember matrix in the decoder.
Robustness to non-specific signals and preprocessing – e.g. noise, baseline	Performance is usually extremely dependent on the quality of the data, as well as on preprocessing, which can vary greatly from application to application.	Autoencoders can incorporate feature selection blocks and training loss to promote invariance to scaling, baselines, and noise, making them more robust to artifacts and outliers.
Variable spectral axis	Observations are assumed to share a common spectral axis, and the spectral axis is discarded. This impedes the integration of data across experimental setups unless they share the same spectral axis.	The axis can be integrated as an input to the model (e.g. as a positional encoding), allowing the direct integration of diverse data sources.
Scalability	Even the simplest models can become computationally prohibitive for larger datasets. More advanced options are intractable in real-world applications involving imaging or volumetric Raman scans.	Deep neural networks are designed to be scalable and parallelizable out-of-the-box.
Number of endmembers	Typically, fixed *a priori*.	Normally fixed *a priori*, too, but can potentially be learned by introducing sparsity and information-theoretic criteria in the latent space, for instance.
Extensibility	Most conventional techniques are based on specialized optimization-based algorithms tailored for the particular unmixing task (e.g. mixture model, scene type, number of endmembers), which makes their extension challenging.	Easy extension and adaptation. One can readily expand unmixing autoencoders into more advanced variants, as well as integrate models into established AI and ML pipelines.
Downstream applications	Conventional techniques are generally constrained within hyperspectral unmixing, with algorithms only outputting the derived unmixing results.	Autoencoders can be used as feature extractors in a variety of modeling and predictive downstream tasks, such as classification, clustering and anomaly detection. They can also leverage transfer learning techniques to improve performance in situations with limited labeled data by utilizing unmixing autoencoders for self-supervised pre-training.

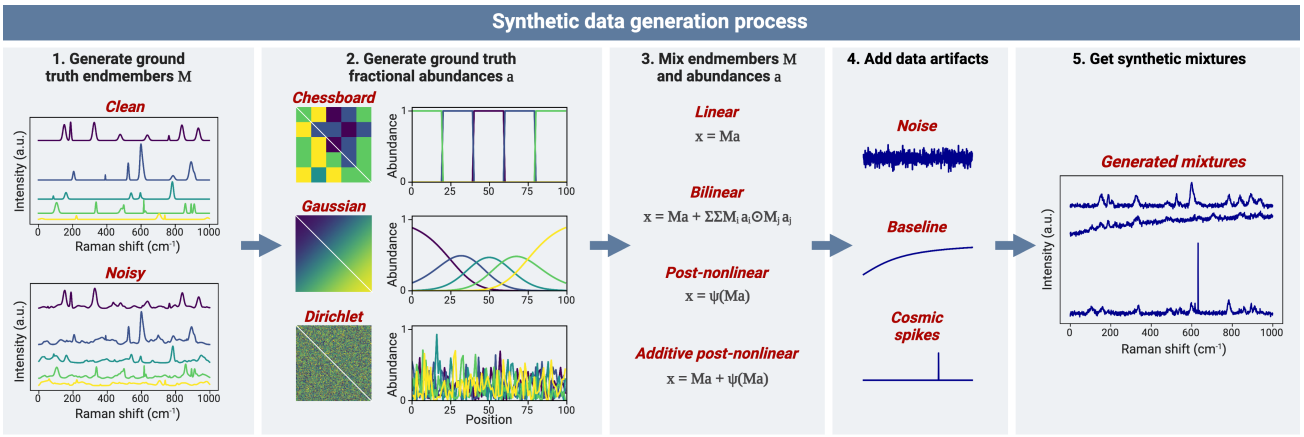

Figure 6. Synthetic data generation workflow.

Here, we set $H = W = 100$, resulting in 10000 spectra per scene/dataset. In the simplest scene (*Chessboard*), we split the scene into 20×20 square patches, each containing a single randomly assigned endmember (i.e., all 400 pixels in each patch are the same one-hot vector). Our second scene (*Gaussian*) consists of n Gaussian functions equally spaced along the diagonal of the scene. After each pixel is normalized to comply with the ASC, we obtain abundance profiles representing different levels of overlap of components. Our last fractional abundance scene (*Dirichlet*) corresponds to a highly mixed scene, where each pixel is individually sampled from a n-dimensional Dirichlet distribution, producing a random mixture of all endmembers. Note that the fractional abundance profile of each pixel in all three scenes complies with both ANC and ASC.

Mixing model. Having generated a set of endmembers and an underlying fractional abundance scene, mixed data measurements $\mathbf{x} \in \mathbb{R}^b$ are created based on a mixing model chosen by the user. In this study, we consider linear mixtures (Equation (1)) and bilinear mixtures based on the Fan model (Equation (2)).

Adding data artifacts. Finally, data artifacts (noise, baseline, cosmic spikes) can be optionally added to create more realistic synthetic Raman spectra. Here, we add Gaussian noise $\boldsymbol{\epsilon} \in \mathbb{R}^b$ to each spectrum, with independent and identically distributed components $\epsilon_i \sim \mathcal{N}(0, \sigma_N)$. Further, we add a baseline signal $\mathbf{B} = h_B \arctan(\pi[1 : b]/b) \in \mathbb{R}^b$ to each spectrum with probability p_B. Finally, with probability p_S, a cosmic spike of intensity $S \sim h_S U(0.75, 1.25)$ is added to each spectrum at a band $b_S \sim \mathcal{U}\{2, b-2\}$. In our experiments: $\sigma_N = 0.1$, $p_B = 0.25$ $h_B = 2$, , $p_S = 0.1$, $h_S = 5$.

D. Model training and evaluation on synthetic Raman mixtures

Autoencoders were trained on synthetic data using the Adam optimizer (learning rate 0.001) over 10 epochs, with spectral angle distance (Equation (5)) as a loss function between input and reconstructed spectra. The latent dimensionality m of each AE model is set to 5 for the *ideal* mixture scenario, and 6 for the other mixture scenarios with data artifacts. Both ANC and ASC are enforced for all experiments on synthetic data. Each experiment on the synthetic data was performed on 5 datasets and 5 model initializations using different random seeds, resulting in $5 \times 5 = 25$ replicates per evaluation, or 1650 experiments in total: 6 models (2 conventional, 4 AEs) $\times$ 11 dataset variants $\times$ 25 replicates. Random seeds were kept the same across mixture scenarios to allow direct comparison.

E. Measuring computational cost

We profile the computational cost of unmixing methods on synthetic datasets (*ideal* scenario, *Chessboard* scene) of increasing sizes, from 2500 to 250000 spectra. The number of endmembers to extract was set to $n = 5$ for all methods. For each experiment, we performed 3 separate evaluations, measuring the wall time of each method (including the training time for autoencoders). All experiments were conducted on a MacBook Air laptop with an Apple M2 chip (8-core CPU, 10-core GPU, and 16-core Neural Engine). We only employed CPU computations to ensure a fair comparison with traditional

Table 2. Benchmark results on diverse synthetic Raman mixture datasets. Each value represents the average result of 25 replicates, including 5 dataset and 5 model initialisations. Confidence intervals are calculated as one standard deviation around the sample mean. Values rounded to 3 decimal places. Best and second best results (mean value) are given in **underlined bold** and **bold**, respectively.

		Chessboard		*Gaussian*		*Dirichlet*	
	Method	Endmembers (SAD) ↓	Abundances (MSE) ↓	Endmembers (SAD) ↓	Abundances (MSE) ↓	Endmembers (SAD) ↓	Abundances (MSE) ↓
ideal	PCA	0.874 ± 0.139	22.592 ± 8.969	1.051 ± 0.081	5.387 ± 1.876	0.847 ± 0.092	3.811 ± 1.291
	N-FINDR + FCLS	**0.000** ± **0.000**	**0.000** ± **0.000**	0.473 ± 0.042	0.033 ± 0.001	0.101 ± 0.040	**0.001** ± **0.000**
	VCA + FCLS	**0.000** ± **0.000**	**0.000** ± **0.000**	0.482 ± 0.047	0.038 ± 0.009	0.105 ± 0.031	**0.003** ± **0.003**
	Dense AE	0.015 ± 0.065	0.002 ± 0.008	0.255 ± 0.037	**0.010** ± **0.005**	**0.003** ± **0.001**	0.012 ± 0.007
	Convolutional AE	0.042 ± 0.111	0.002 ± 0.007	**0.168** ± **0.033**	**0.009** ± **0.006**	**0.003** ± **0.001**	0.012 ± 0.006
	Transformer AE	**0.002** ± **0.000**	**0.000** ± **0.000**	**0.212** ± **0.043**	**0.009** ± **0.005**	**0.007** ± **0.002**	0.012 ± 0.006
	Convolutional Transformer AE	**0.002** ± **0.000**	**0.000** ± **0.000**	0.215 ± 0.042	**0.009** ± **0.005**	0.017 ± 0.062	0.013 ± 0.007
+ artifacts	PCA	0.819 ± 0.072	79.38 ± 49.727	1.009 ± 0.082	61.488 ± 48.586	0.869 ± 0.069	60.154 ± 48.330
	N-FINDR + FCLS	0.629 ± 0.111	0.073 ± 0.036	0.828 ± 0.075	0.041 ± 0.015	0.599 ± 0.116	0.033 ± 0.015
	VCA + FCLS	0.353 ± 0.159	0.071 ± 0.042	0.609 ± 0.093	0.059 ± 0.015	0.392 ± 0.175	0.036 ± 0.019
	Dense AE	0.072 ± 0.081	0.042 ± 0.032	**0.347** ± **0.137**	0.029 ± 0.021	**0.068** ± **0.070**	0.021 ± 0.011
	Convolutional AE	**0.033** ± **0.003**	0.023 ± 0.013	**0.385** ± **0.166**	**0.027** ± **0.014**	**0.068** ± **0.072**	**0.017** ± **0.008**
	Transformer AE	**0.039** ± **0.031**	**0.022** ± **0.011**	0.393 ± 0.129	**0.026** ± **0.014**	**0.073** ± **0.105**	**0.017** ± **0.009**
	Convolutional Transformer AE	**0.033** ± **0.004**	**0.021** ± **0.010**	0.399 ± 0.164	**0.027** ± **0.015**	0.112 ± 0.140	**0.019** ± **0.010**
+ realistic	PCA	0.967 ± 0.074	88.132 ± 50.053	1.079 ± 0.045	63.735 ± 48.513	0.993 ± 0.089	61.244 ± 48.733
	N-FINDR + FCLS	0.361 ± 0.062	0.072 ± 0.040	0.478 ± 0.047	0.043 ± 0.013	0.299 ± 0.087	0.030 ± 0.019
	VCA + FCLS	0.173 ± 0.079	0.061 ± 0.038	0.400 ± 0.075	0.063 ± 0.016	0.229 ± 0.088	0.030 ± 0.017
	Dense AE	0.045 ± 0.024	0.030 ± 0.023	0.177 ± 0.050	0.015 ± 0.011	**0.073** ± **0.012**	**0.011** ± **0.006**
	Convolutional AE	0.054 ± 0.013	**0.019** ± **0.008**	0.168 ± 0.030	**0.011** ± **0.004**	**0.071** ± **0.022**	**0.008** ± **0.003**
	Transformer AE	**0.039** ± **0.005**	**0.018** ± **0.009**	**0.156** ± **0.044**	**0.011** ± **0.005**	0.085 ± 0.014	**0.008** ± **0.003**
	Convolutional Transformer AE	**0.040** ± **0.005**	**0.018** ± **0.008**	**0.151** ± **0.033**	**0.010** ± **0.004**	0.081 ± 0.012	**0.008** ± **0.003**
+ bilinear	PCA	—"—	—"—	1.117 ± 0.071	67.189 ± 48.404	1.038 ± 0.112	63.915 ± 48.423
	N-FINDR + FCLS	—"—	—"—	0.456 ± 0.049	0.039 ± 0.016	0.287 ± 0.074	0.025 ± 0.015
	VCA + FCLS	—"—	—"—	0.391 ± 0.058	0.057 ± 0.018	0.277 ± 0.082	0.030 ± 0.015
	Dense AE (bilinear)	—"—	—"—	0.247 ± 0.070	0.017 ± 0.011	**0.094** ± **0.011**	**0.010** ± **0.005**
	Convolutional AE (bilinear)	—"—	—"—	***0.194*** ± **0.030**	**0.010** ± **0.004**	**0.087** ± **0.010**	**0.008** ± **0.003**
	Transformer AE (bilinear)	—"—	—"—	0.222 ± 0.084	0.012 ± 0.008	0.105 ± 0.012	**0.008** ± **0.003**
	Convolutional Transformer AE (bilinear)	—"—	—"—	**0.208** ± **0.051**	**0.011** ± **0.006**	0.100 ± 0.010	**0.008** ± **0.003**

methods which, by design, do not utilize GPU acceleration.

F. Analysis of experimental RS data from sugar mixtures

Preparation of sugar solutions. We prepared $1\,\mathrm{mol/L}$ solutions of each type of sugar (sucrose, fructose, maltose, and glucose) by dissolving the appropriate weight of sugar into $40\,\mathrm{mL}$ of ultrapure distilled water (Invitrogen™ – UltraPure™ DNase/RNase-Free Distilled Water). The weights of sugars dissolved were $13.83\,\mathrm{g}$ for sucrose (Thermo Scientific Chemicals – Sucrose, 99%), $7.279\,\mathrm{g}$ for fructose (Thermo Scientific Chemicals – D-Fructose, 99%), $15.171\,\mathrm{g}$ for maltose (Thermo Scientific Chemicals – D-(+)-Maltose monohydrate, 95%) and $7.279\,\mathrm{g}$ for glucose (D-(+)-Glucose, AnalaR NORMAPUR® analytical reagent). All solutions were mixed and vortexed in standard $50\,\mathrm{mL}$ centrifuge tubes until no solute was visible.

Sugar mixtures were prepared in standard 96-well plates, with a volume of $375\,\mu\mathrm{L}$ per well. A full factorial experiment was performed comprising 4 volume levels for each sugar ($0\,\mu\mathrm{L}$, $30\,\mu\mathrm{L}$, $75\,\mu\mathrm{L}$ and $120\,\mu\mathrm{L}$), filled with distilled water where necessary. Discarding the mixtures exceeding the volume of the well and the one that contains no sugar, 240 distinct mixtures were prepared. In addition, 5 extra 'pure' solutions (i.e., $375\,\mu\mathrm{L}$ of water, sucrose, fructose, maltose, or glucose) were prepared, which we used to extract reference spectra for each chemical species. This resulted in a total of 245 wells distributed in three standard 96-well plates. Mixtures were stirred using standard $200\,\mu\mathrm{L}$ pipettes before spectral acquisition to ensure good mixing.

Raman measurements from sugar solutions. All spectra were acquired using a custom Raman microspectroscopy platform designed for high-throughput analysis known as B-Raman. This platform is based on the Thorlabs Cerna® and features the BWTek BRM-785-0.55-100-0.22-SMA laser excitation source and the Ibsen EAGLE Raman-S spectrometer. The instrument was calibrated using an Argon wavelength calibration source (AR-2 – Ocean Insight) reference lamp before

Table 3. Full unmixing results on Raman spectroscopy data from sugar solutions. Each value represents the average result of 5 replicates. Confidence intervals are calculated as one standard deviation around the sample mean. Values rounded to 3 decimal places. Best and second best results (mean value) are given in **underlined bold** and **bold**, respectively.

		High SNR				*Low SNR*			
		with endmembers		*without endmembers*		*with endmembers*		*without endmembers*	
	Method	Endmembers (SAD) ↓	Abundances (MSE) ↓	Endmembers (SAD) ↓	Abundances (MSE) ↓	Endmembers (SAD) ↓	Abundances (MSE) ↓	Endmembers (SAD) ↓	Abundances (MSE) ↓
	PCA	1.018 ± 0.000	0.073 ± 0.000	1.026 ± 0.000	0.072 ± 0.000	1.173 ± 0.000	0.075 ± 0.000	1.210 ± 0.000	0.073 ± 0.000
	N-FINDR+FCLS	**$\underline{0.202} \pm \underline{0.000}$**	0.005 ± 0.000	0.515 ± 0.035	0.037 ± 0.029	0.900 ± 0.000	**$\underline{0.026} \pm \underline{0.000}$**	1.077 ± 0.000	0.154 ± 0.000
	VCA+FCLS	0.322 ± 0.044	**0.004 ± 0.000**	0.525 ± 0.026	0.042 ± 0.029	0.677 ± 0.022	0.077 ± 0.035	0.901 ± 0.017	0.108 ± 0.006
SAD	Dense AE	0.212 ± 0.004	0.057 ± 0.001	0.219 ± 0.001	0.058 ± 0.001	**$\underline{0.462} \pm \underline{0.038}$**	**0.058 ± 0.005**	**$\underline{0.503} \pm \underline{0.046}$**	**$\underline{0.054} \pm \underline{0.007}$**
SAD	Convolutional AE	**0.203 ± 0.001**	0.058 ± 0.001	**$\underline{0.215} \pm \underline{0.002}$**	0.058 ± 0.001	0.503 ± 0.055	**0.058 ± 0.002**	0.529 ± 0.041	**0.057 ± 0.003**
SAD	Transformer AE	0.206 ± 0.001	0.058 ± 0.001	**0.218 ± 0.002**	0.028 ± 0.002	**0.496 ± 0.050**	0.061 ± 0.002	0.545 ± 0.037	0.069 ± 0.013
SAD	Conv. Trans. AE	0.208 ± 0.002	0.057 ± 0.001	**0.218 ± 0.001**	0.057 ± 0.001	0.521 ± 0.059	0.068 ± 0.013	0.533 ± 0.035	0.062 ± 0.011
MSE + SAD	Dense AE	0.242 ± 0.008	**$\underline{0.003} \pm \underline{0.001}$**	0.265 ± 0.040	**0.003 ± 0.002**	0.498 ± 0.038	0.070 ± 0.006	**0.509 ± 0.045**	0.069 ± 0.008
MSE + SAD	Convolutional AE	0.264 ± 0.068	**0.004 ± 0.004**	0.240 ± 0.007	**$\underline{0.002} \pm \underline{0.000}$**	0.568 ± 0.057	0.083 ± 0.017	0.593 ± 0.052	0.080 ± 0.013
MSE + SAD	Transformer AE	0.234 ± 0.006	**0.004 ± 0.001**	0.240 ± 0.009	0.004 ± 0.001	0.606 ± 0.034	0.101 ± 0.026	0.632 ± 0.014	0.097 ± 0.028
MSE + SAD	Conv. Trans. AE	0.321 ± 0.121	0.008 ± 0.005	0.333 ± 0.123	0.007 ± 0.006	0.570 ± 0.068	0.113 ± 0.012	0.665 ± 0.179	0.164 ± 0.037

Table 4. Full unmixing results on Raman spectroscopy data from sugar solutions with respect to an alternative endmember similarity metric PCC based on the Pearson correlation coefficient. Each value represents the average result of 5 replicates. Confidence intervals are calculated as one standard deviation around the sample mean. Values rounded to 3 decimal places. Best and second best results (mean value) are given in **underlined bold** and **bold**, respectively. The PCC between two endmember signatures x and y is defined as $\mathrm{PCC}(x, y) = 1 - \dfrac{\sum (x_i - \bar{x}) \sum (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$.

		High SNR				*Low SNR*			
		with endmembers		*without endmembers*		*with endmembers*		*without endmembers*	
	Method	Endmembers (PCC) ↓	Abundances (MSE) ↓	Endmembers (PCC) ↓	Abundances (MSE) ↓	Endmembers (PCC) ↓	Abundances (MSE) ↓	Endmembers (PCC) ↓	Abundances (MSE) ↓
	PCA	0.547 ± 0.000	0.073 ± 0.000	0.541 ± 0.000	0.072 ± 0.000	0.641 ± 0.000	0.075 ± 0.000	0.691 ± 0.000	0.073 ± 0.000
	N-FINDR+FCLS	0.039 ± 0.000	0.005 ± 0.000	0.203 ± 0.055	0.037 ± 0.029	0.520 ± 0.000	**$\underline{0.026} \pm \underline{0.000}$**	0.641 ± 0.000	0.154 ± 0.000
	VCA+FCLS	0.099 ± 0.047	**0.004 ± 0.000**	0.221 ± 0.056	0.042 ± 0.029	0.325 ± 0.019	0.077 ± 0.035	0.482 ± 0.005	0.108 ± 0.006
SAD	Dense AE	0.029 ± 0.001	0.057 ± 0.001	**0.031 ± 0.000**	0.058 ± 0.001	**$\underline{0.164} \pm \underline{0.042}$**	**0.058 ± 0.005**	**0.198 ± 0.046**	**$\underline{0.054} \pm \underline{0.007}$**
SAD	Convolutional AE	**$\underline{0.027} \pm \underline{0.000}$**	0.058 ± 0.001	**$\underline{0.030} \pm \underline{0.002}$**	0.058 ± 0.001	0.207 ± 0.052	**0.058 ± 0.002**	0.228 ± 0.030	**0.057 ± 0.003**
SAD	Transformer AE	**0.028 ± 0.000**	0.058 ± 0.001	**0.031 ± 0.000**	0.028 ± 0.002	0.200 ± 0.048	0.061 ± 0.002	0.243 ± 0.032	0.069 ± 0.013
SAD	Conv. Trans. AE	**0.028 ± 0.000**	0.057 ± 0.001	**0.031 ± 0.000**	0.057 ± 0.001	0.226 ± 0.059	0.068 ± 0.013	0.236 ± 0.034	0.062 ± 0.011
MSE + SAD	Dense AE	0.037 ± 0.002	**$\underline{0.003} \pm \underline{0.001}$**	0.047 ± 0.018	**0.003 ± 0.002**	**0.174 ± 0.032**	0.070 ± 0.006	**$\underline{0.185} \pm \underline{0.042}$**	0.069 ± 0.008
MSE + SAD	Convolutional AE	0.056 ± 0.044	**0.004 ± 0.004**	0.037 ± 0.002	**$\underline{0.002} \pm \underline{0.000}$**	0.243 ± 0.044	0.083 ± 0.017	0.272 ± 0.040	0.080 ± 0.013
MSE + SAD	Transformer AE	0.035 ± 0.002	**0.004 ± 0.001**	0.037 ± 0.002	0.004 ± 0.001	0.296 ± 0.041	0.101 ± 0.026	0.311 ± 0.020	0.097 ± 0.028
MSE + SAD	Conv. Trans. AE	0.104 ± 0.096	0.008 ± 0.005	0.108 ± 0.099	0.007 ± 0.006	0.250 ± 0.072	0.113 ± 0.012	0.348 ± 0.176	0.164 ± 0.037

data collection. The excitation wavelength was $785\,\mathrm{nm}$ and the power incident to the samples was $36.3\,\mathrm{mW}$. The Raman scattering was collected in reflection via a Leica N PLAN 10x/0.25 objective with 0.25 numerical aperture. The raw spectra were acquired over the spectral wavenumber range of 142–$3684.8\,\mathrm{cm}^{-1}$.

Spectra were measured from the center (horizontal) of each well at a fixed depth that was established to provide the highest signal. Two sets of data were collected from each well, at $5\,\mathrm{s}$ and $0.5\,\mathrm{s}$ integration times, to compare unmixing performance on low and high signal-to-noise ratio (SNR) data. Several measurements were collected from each well, resulting in a total of 240 solutions $\times$ 2 measurements $\times$ 4 repetitions $=$ 1920 high-SNR measurements (1960 with reference spectra); and 240 solutions $\times$ 8 measurements $\times$ 4 repetitions $=$ 7680 low-SNR measurements (7840 with reference spectra). Ground-truth endmembers signatures were obtained by taking the median (band-wise) of the reference spectra (40 in high SNR setup, and 160 in low SNR setup) collected from the 5 additional wells containing pure solutions. Ground truth fractional abundances were determined by calculating the ratio of the components present in each mixture.

Preprocessing of sugar data. First, we preprocess each sugar dataset: 1) cropping to the region 400–$1800\,\mathrm{cm}^{-1}$; 2) baseline correction with Adaptive Smoothness Parameter Penalized Least Squares (ASPLS) (Zhang et al., 2020a)—

smoothing parameter $\lambda = 10^5$, differential matrix of order 2, maximum iterations set to 100, exit criteria with tolerance $t = 0.001$; 3) global vector normalization, where each observation is divided by the highest magnitude observed in the data. Baseline removal is important to ensure models extract relevant features (i.e., characteristic peaks) as opposed to merely capturing the trend.

Analysis of sugar data. To perform hyperspectral unmixing, we set the number of endmembers to extract to $n = 5$, and we follow similar training and evaluation protocols to those employed for the synthetic data, but we increase the number of epochs to 15 for low SNR data and 50 for high SNR data given the more limited number of spectra collected. We also incorporate an additional MSE term in the training loss $\mathcal{L}$ of autoencoders on high SNR data:

$$\mathcal{L}(\mathbf{x}, \widehat{\mathbf{x}}) = \text{SAD}(\mathbf{x}, \widehat{\mathbf{x}}) + \lambda \, \text{MSE}(\mathbf{x}, \widehat{\mathbf{x}}), \tag{6}$$

with $\lambda = 1000$. This term breaks the invariance to scale and leads to better abundance estimation given the weak water endmember (see Table 3). The standard SAD loss was used for experiments on low SNR. Each experiment is repeated for 5 model initializations.

Table 4 presents performance evaluation using an alternative endmember distance based on Pearson's correlation coefficient (PCC), showing an even more pronounced improvement in endmember estimation accuracy.

G. Analysis of volumetric RS data from THP-1 cell

Acquisition and preprocessing of cell data. The volumetric Raman scan of the THP-1 cell (Kallepitis et al., 2017) is collected using 0.3 s integration time and comprises a $z = 1, \ldots, 10$ stack of ten 40×40 raster scans, organized into a single volumetric hypercube for analysis. We preprocess the data before unmixing using the following protocol: 1) spectral cropping to the fingerprint region 700–1800 cm^{-1}; 2) cosmic spike removal using the algorithm in (Whitaker & Hayes, 2018) with kernel of size 3 and z-value threshold of 8; 3) denoising with Savitzky-Golay filter using a cubic polynomial kernel of size 7 (Savitzky & Golay, 1964); 4) baseline correction using Asymmetric Least Squares (AsLS) with smoothing parameter $\lambda = 10^6$, penalizing weighting factor $p = 0.01$, differential matrix of order 2, maximum iterations set to 50, exit criteria with tolerance threshold of $t = 0.001$ (Eilers & Boelens, 2005); 5) global MinMax normalization to the interval $[0, 1]$.

Analysis of cell data. Unmixing is performed following the same AE training protocol as in other analyses, with the number of training epochs set to 20, and the number of endmembers to extract to $n = 20$. Here, we also discard the constraint that fractional abundances must sum to one. Out of the 20 endmembers we obtain, we display the 5 deemed most biologically relevant following peak assignment as per the original paper (Kallepitis et al., 2017). For VCA+NNLS, two of those five endmembers corresponded to the same cell organelle, namely cytoplasm, and were visualized using the same color in the merged reconstruction displayed in Figure 5b.

Peak characterization. Cell organelles were determined based on the following peaks: PBS buffer - 1637 cm^{-1} (water peak); cytoplasm - 1005 cm^{-1} (phenylalanine), 1250 cm^{-1} (Amide III), 1659 cm^{-1} (Amide I) and 1445 cm^{-1} (CH deformations of proteins and lipids); TAGs/PLPs - 1092 cm^{-1} (C–C stretching), 1308 cm^{-1} (CH_2 twists), 1445 cm^{-1} (CH deformation), and 1661 cm^{-1} (C=C stretching); nucleus/DNA - 790 cm^{-1} (symmetric phosphodiester stretch and ring breathing modes of pyrimidine bases) and 1103 cm^{-1} (symmetric dioxy-stretch of the phosphate backbone); cholesterol - 1069 cm^{-1} and 1134 cm^{-1} (cholesteryl stearate), 1300 cm^{-1} (CH_2 twists), and 1443 cm^{-1} (CH deformation) (Kallepitis et al., 2017; Zhang et al., 2012; Movasaghi et al., 2007).

Figure 7. Endmember estimates on high SNR sugar data. Qualitative comparison of derived endmembers on the high SNR sugar datasets - with (a), and without reference spectra (b). Endmembers are scaled such that their maximum intensity is equal to 1 for visualization purposes. x-axes represent the Raman shift region 400–1800 cm^{-1}, and y-axes, which are shared for each row, represents normalized intensity (a.u.).

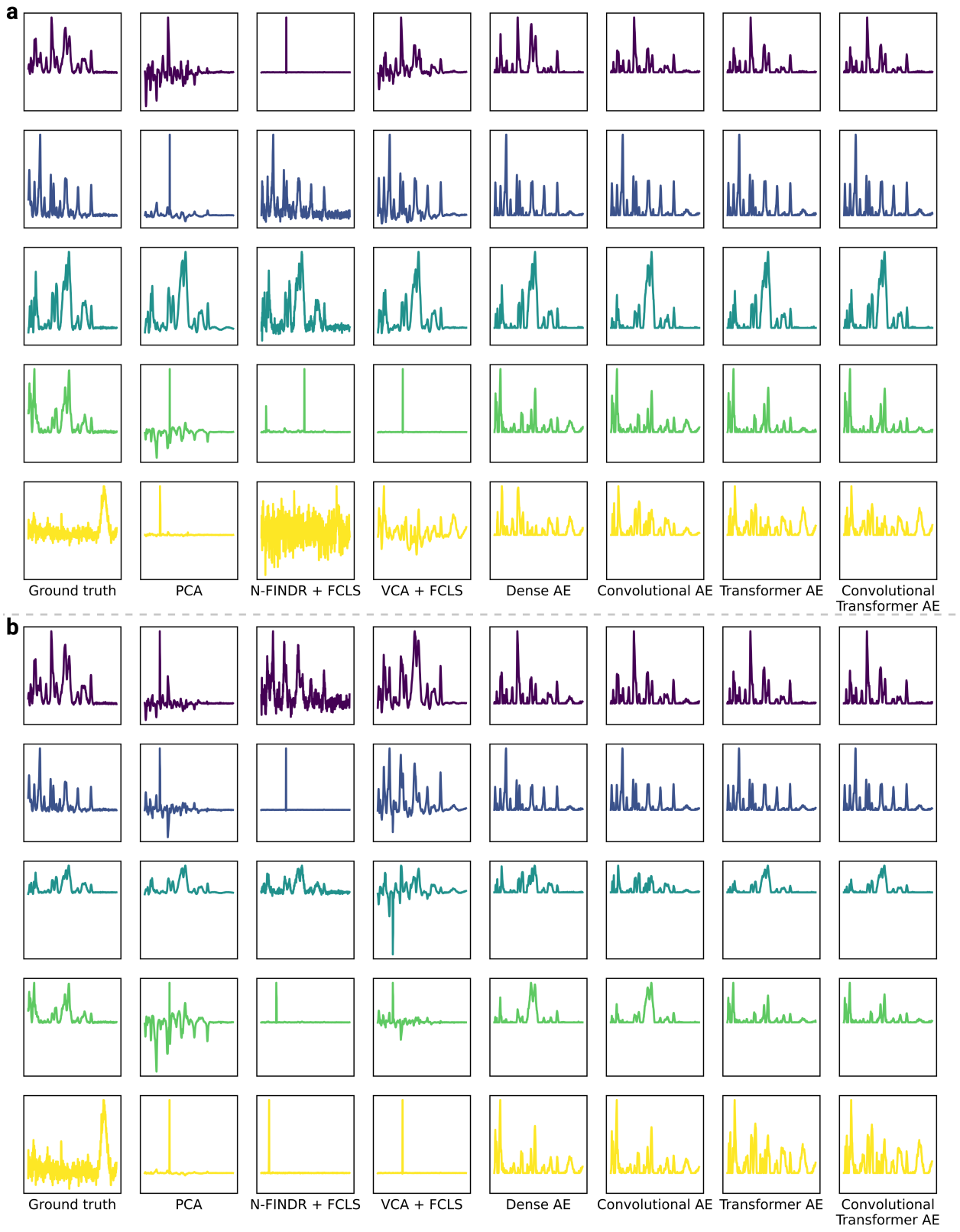

Figure 8. Endmember estimates on low SNR sugar data. Qualitative comparison of derived endmembers on the low SNR sugar datasets - with (a), and without reference spectra (b). Endmembers are scaled such that their maximum intensity is equal to 1 for visualization purposes. x-axes represent the Raman shift region 400–1800 cm^{-1}, and y-axes, which are shared for each row, represents normalized intensity (a.u.).

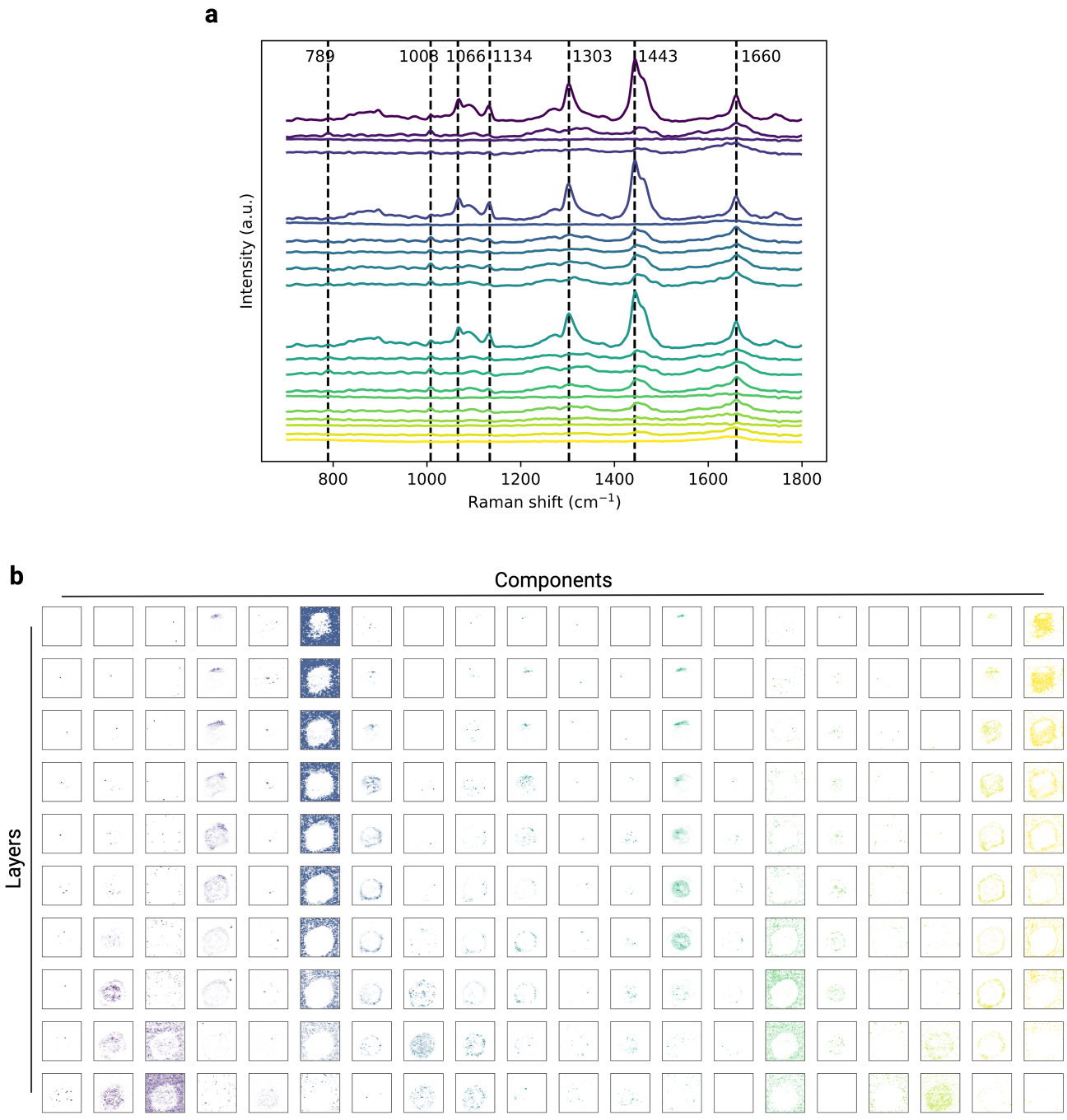

Figure 9. Full unmixing results obtained with VCA + NNLS on the THP-1 cell data. a, Derived endmembers. b, Derived fractional abundances.

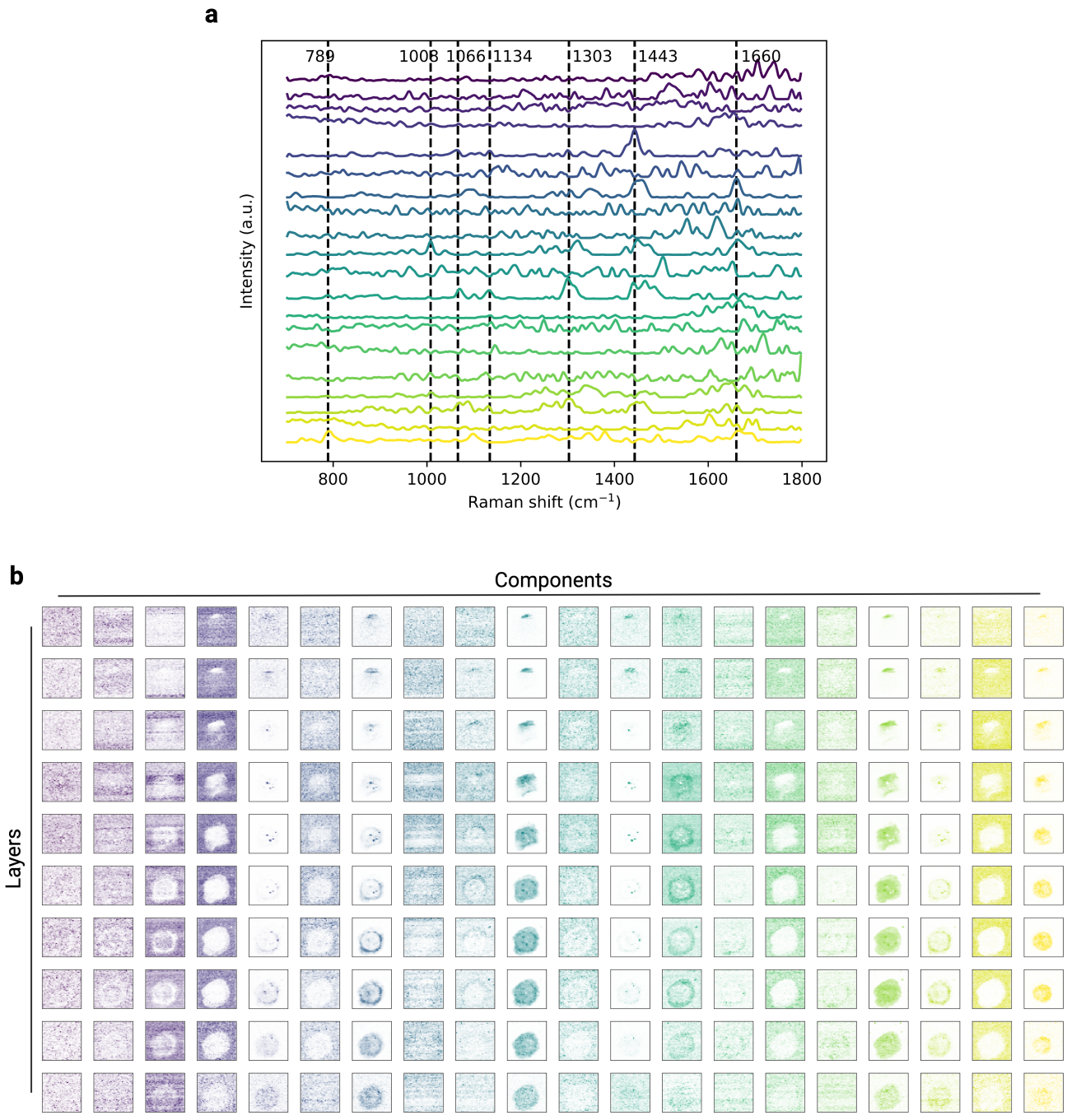

Figure 10. Full unmixing results obtained with our *Dense AE* model on the THP-1 cell data. a, Derived endmembers. b, Derived fractional abundances.

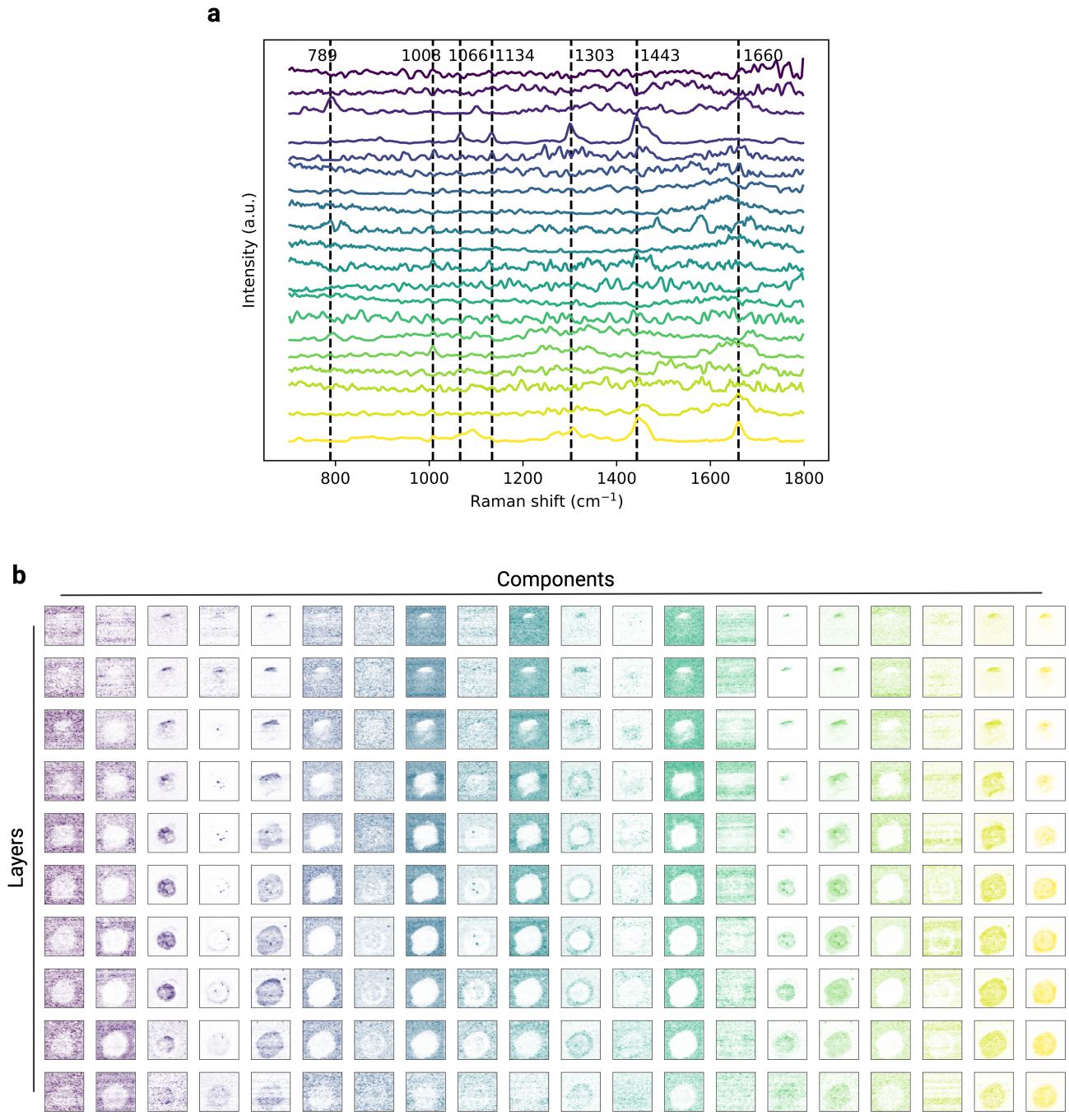

Figure 11. Full unmixing results obtained with our *Deep Dense AE* model on the preprocessed THP-1 cell data. a, Derived endmembers. b, Derived fractional abundances.