



Hyperspectral unmixing for Raman spectroscopy via physics-constrained autoencoders

Dimitar Georgiev^{a,b,c,d,e}, Álvaro Fernández-Galiana^{c,d,e,1}, Simon Vilms Pedersen^{c,d,e,1,2}, Georgios Papadopoulos^{a,b,c,d,e}, Ruoxiao Xie^{c,d,e}, Molly M. Stevens^{c,d,e,f,g,h,3}, and Mauricio Barahona^{i,3}

Affiliations are included on p. 10.

Edited by Alexis Bell, University of California, Berkeley, CA; received April 24, 2024; accepted September 10, 2024

Raman spectroscopy is widely used across scientific domains 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 based on autoencoder neural networks, and we systematically validate them using both 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 monocytic cell.

Raman spectroscopy | hyperspectral unmixing | machine learning | autoencoders | chemometrics

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 (1–5). 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 (6–10). As a result, RS has become an important analytical tool in a myriad of scientific domains, from chemistry (11, 12), biology (13–16), and medicine (17–23), to materials science (24, 25), pharmacology (26–28), environmental science (29–31), food quality control (32, 33), and even forensics (34–36).

Despite the wealth of information RS affords, the analysis and interpretation of experimental RS data remains a major challenge (37–39). 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 (39, 40).

Hyperspectral unmixing (also known as (hyper)spectral deconvolution or multivariate curve resolution) aims to resolve such mixed signals (41, 42) by identifying the individual components present (endmember identification) and/or quantifying their proportions (abundance estimation) (Fig. 1A). Popular approaches include N-FINDR (43) and vertex component analysis (VCA) (44) for endmember identification, and non-negative least squares (NNLS) (45) and fully constrained least squares (FCLS) (46) for abundance estimation (41, 47). However, such techniques, which originated in the field of remote sensing (48, 49), have limitations for the unmixing of RS data. Specifically, these methods are typically 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).

Over the past few years, machine learning (ML) has gained considerable traction for the analysis of Raman spectroscopy data (50–54). Nonetheless, despite significant advances, most methods remain focused on constrained supervised tasks where a model is learned assuming knowledge of target species and their values (55). In particular, approaches for compositional characterization of Raman measurements normally utilize

Significance

Hyperspectral unmixing methods are essential to exploit the capabilities of Raman spectroscopy for nondestructive, unbiased chemical characterization in a wide array of domains, from biology, chemistry, and materials to engineering and environmental science. Here, we take advantage of recent advances in machine learning and introduce a framework for Raman unmixing based on autoencoder neural networks. We demonstrate that such methods offer more versatile, robust, and data-driven Raman unmixing with improved performance compared to conventional methods in complex samples.

Competing interest statement: M.M.S. holds part-time appointments at Imperial College London and the Karolinska Institute. M.M.S. is founder of Sparta Bioscience Ltd, which commercializes a technology for single particle Raman spectroscopy. M.M.S. is inventor in a patent describing a technique for SPARTA, a technique for single particle Raman spectroscopy (1810010.7), and in a patent describing Raman tags (2314282.1/GB/PRV). M.M.S. invested in, consults for (or was on scientific advisory boards or boards of directors), and conducts sponsored research funded by companies related to the biomaterials field.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

¹Á.F.-G. and S.V.P. contributed equally to this work.

²Present address: Spaceline Labs, Odense 5270, Denmark.

³To whom correspondence may be addressed. Email: molly.stevens@dpag.ox.ac.uk or m.barahona@imperial.ac.uk.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2407439121/-/DCSupplemental>.

Published October 29, 2024.

explicit information provided during model training about the presence (or absence) of particular target species (56) or known concentrations (57–61). Therefore, the development of methods for general unsupervised hyperspectral unmixing that do not rely on known target substances and concentrations remains essential to realize the potential of Raman spectroscopy for label-free chemical characterization.

Autoencoder (AE) neural networks have recently emerged as a framework for hyperspectral unmixing in remote sensing, spurred by the availability of standardized benchmark datasets for model evaluation (e.g., *Urban*, *Samson*, *AVARIS Cuprite*) (62–66). Yet, despite initial explorations (67, 68), the utility of unmixing AEs for Raman spectroscopy data remains largely unexplored. Here, we develop a range of AEs for RS hyperspectral unmixing, which we systematically validate against conventional unmixing methods using synthetic and experimental Raman data.

Results

Raman Unmixing with Autoencoder Neural Networks. Autoencoders are a family of (deep) neural network models consisting of two subnetworks (encoder and decoder) connected sequentially (69). The encoder $\mathcal{E} : \mathbb{R}^b \rightarrow \mathbb{R}^m$, where $m \ll b$, transforms input data \mathbf{x} to a lower-dimensional latent space representation $\mathbf{z} = \mathcal{E}(\mathbf{x})$, which the decoder $\mathcal{D} : \mathbb{R}^m \rightarrow \mathbb{R}^b$ uses to produce reconstructions $\hat{\mathbf{x}} = \mathcal{D}(\mathbf{z})$ of the original input. AE models are typically trained in a self-supervised manner by minimizing a loss function $\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}})$ that measures the discrepancy between the input \mathbf{x} and the reconstruction $\hat{\mathbf{x}}$ (e.g., the mean squared error (MSE)); hence no ground-truth target values are provided during training. 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.

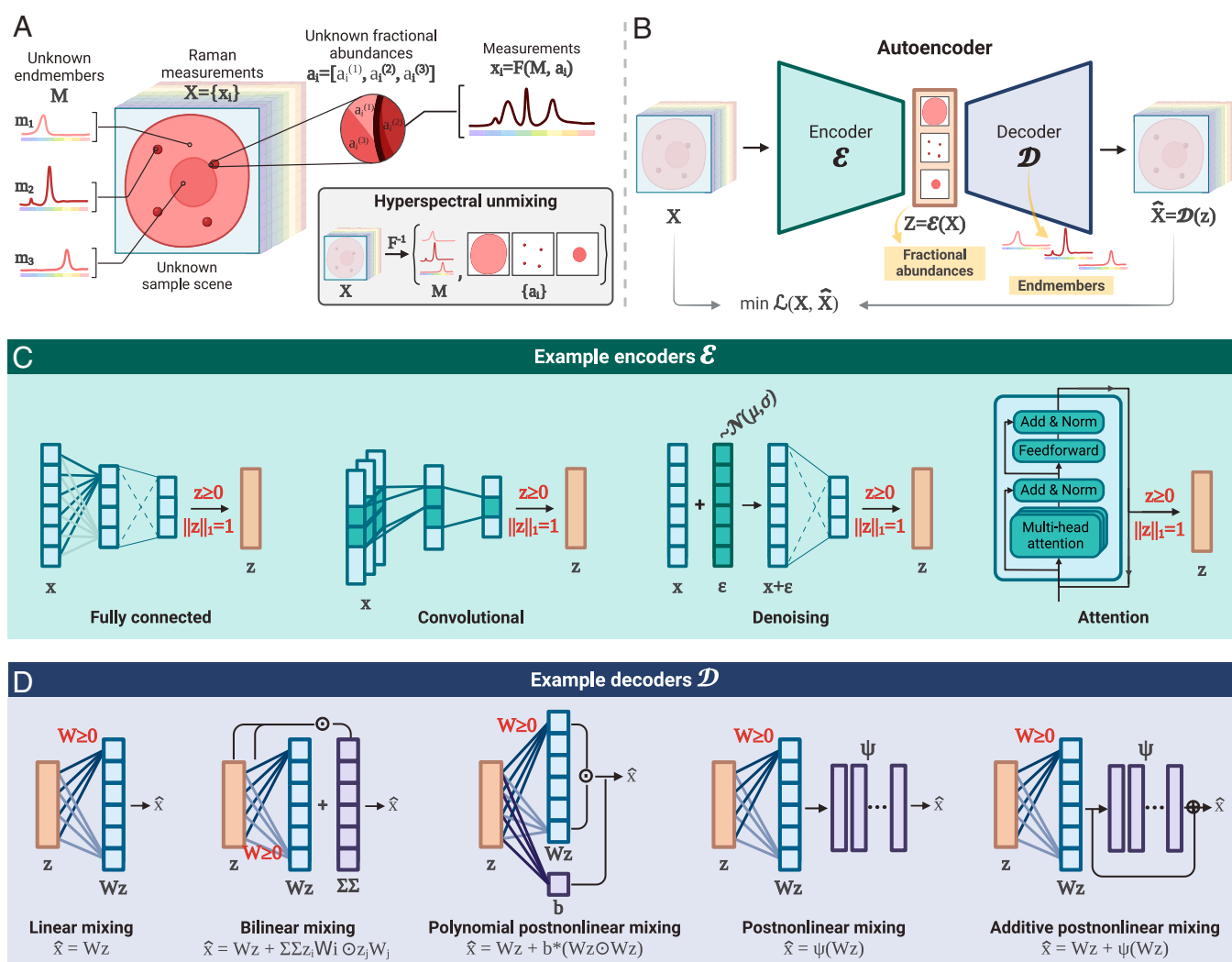


Fig. 1. Hyperspectral unmixing for Raman spectroscopy using autoencoder neural networks. (A) Illustration of hyperspectral unmixing. A set of Raman measurements $X = \{\mathbf{x}_i\}$, where each \mathbf{x}_i is a spectrum, are collected from a sample of interest. The spectra \mathbf{x}_i are viewed as mixtures of a set of Raman spectral signatures M (endmembers), related to the chemical species present, and their respective fractional abundances \mathbf{a}_i . Hyperspectral unmixing aims to infer the set of unknown underlying endmembers M and fractional abundances \mathbf{a}_i from the set of measurements X , thereby obtaining qualitative and quantitative compositional information about the sample. (B) Hyperspectral unmixing as a self-supervised autoencoder learning problem: the decoder \mathcal{D} learns endmember signatures during training and the encoder \mathcal{E} learns to derive 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 nonlinear mixing models. Physics-inspired constraints are indicated in red in C and D.

This dual functionality can be harnessed to design autoencoders for hyperspectral unmixing: the latent representations $\mathbf{z} = \mathcal{E}(\mathbf{x})$ can be interpreted as fractional abundances (with respect to the 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 “unmixing” where the decoder identifies endmember signatures and the encoder quantifies the fractional abundances of these learned endmembers in the input spectrum (Fig. 1B). To guide the learning, we incorporate physical constraints into the AE architecture to reflect the nature of hyperspectral unmixing, e.g., nonnegativity of endmembers and fractional abundances, and sum-to-one abundances (Fig. 1 and *Materials and Methods*).

Compared to standard methods for unmixing, autoencoders are more effective at capturing complex, noisy, and nonlinear relationships in the data, and do not rely on assumptions like the presence of pure pixels (see *SI Appendix, Table S1*, for extended discussions). Moreover, AEs offer a more flexible and versatile framework for unmixing, as illustrated by the wide design space of encoder and decoder architectures (Fig. 1C and D).

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 (70–72), or attention mechanisms to model long-range dependencies (73) (Fig. 1C). In addition, sparsity, part-based learning, and denoising objectives can be enforced during training to enhance explainability and robustness (74–78).

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 (Fig. 1D) (66, 79, 80). This is akin to introducing an inductive prior with respect to the mixture model directly via the AE architecture. Furthermore, the decoder can be preinitialized with a set of endmembers (e.g., an endmember library or signatures derived using methods such as VCA), or readily adapted to nonblind unmixing by fixing certain parameters to predefined endmember signatures.

Unmixing autoencoders. To assess the effectiveness of AEs as a framework for RS unmixing, we develop and evaluate a collection of AE models, each defined by a specific encoder and decoder. We consider four types of encoders encompassing a variety of architectures, from traditional dense layers to contemporary convolutional and attention mechanisms: 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*) (81); and 4) a transformer-based encoder with a 1D convolutional feature extractor (*Convolutional Transformer*). The two types of decoders we investigate are 1) a decoder designed for linear unmixing (Eq. 4) and 2) a decoder designed for bilinear unmixing (Eq. 5). The autoencoders are trained in a self-supervised fashion by minimizing a loss based on the spectral angle divergence (SAD) (82) that measures the cosine similarity between input and reconstructed spectra.

Baseline methods for comparison. 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. This is performed using the RamanSPy package (83) in Python. We also compared to principal component analysis (PCA), which, despite not being designed for unmixing, is commonly used in applications. We

omit the PCA results in the main text as they exhibit substantially lower performance (*SI Appendix, Tables S2–S4 and Figs. S1 and S2*).

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 (Fig. 2A). This allows us to quantitatively compare the performance of unmixing approaches (see Fig. 2B for unmixing of an example synthetic dataset).

Using our data generator, we produce 11 types of synthetic datasets of variable complexity, based on four mixture scenarios 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 (84) 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 10,000 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 semimixed 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). Therefore, our synthetic datasets cover increasingly complex 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 to 3). Such data comply with the linear mixing assumption of conventional methods, and for consistency, we equip the AE models with a decoder for linear unmixing. Fig. 2C summarizes 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 *SI Appendix, Table S2*, for calculated performance metrics). We measure the discrepancy between ground truth and derived endmembers (using SAD), and the discrepancy between ground truth and derived fractional abundances (using MSE). 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.

Nonlinear unmixing with autoencoders. We then proceed to our benchmark analysis on synthetic data generated using a nonlinear mixture model (i.e., *+bilinear* scenario). The results are displayed in Fig. 2D, where we equip our AEs with a decoder specific

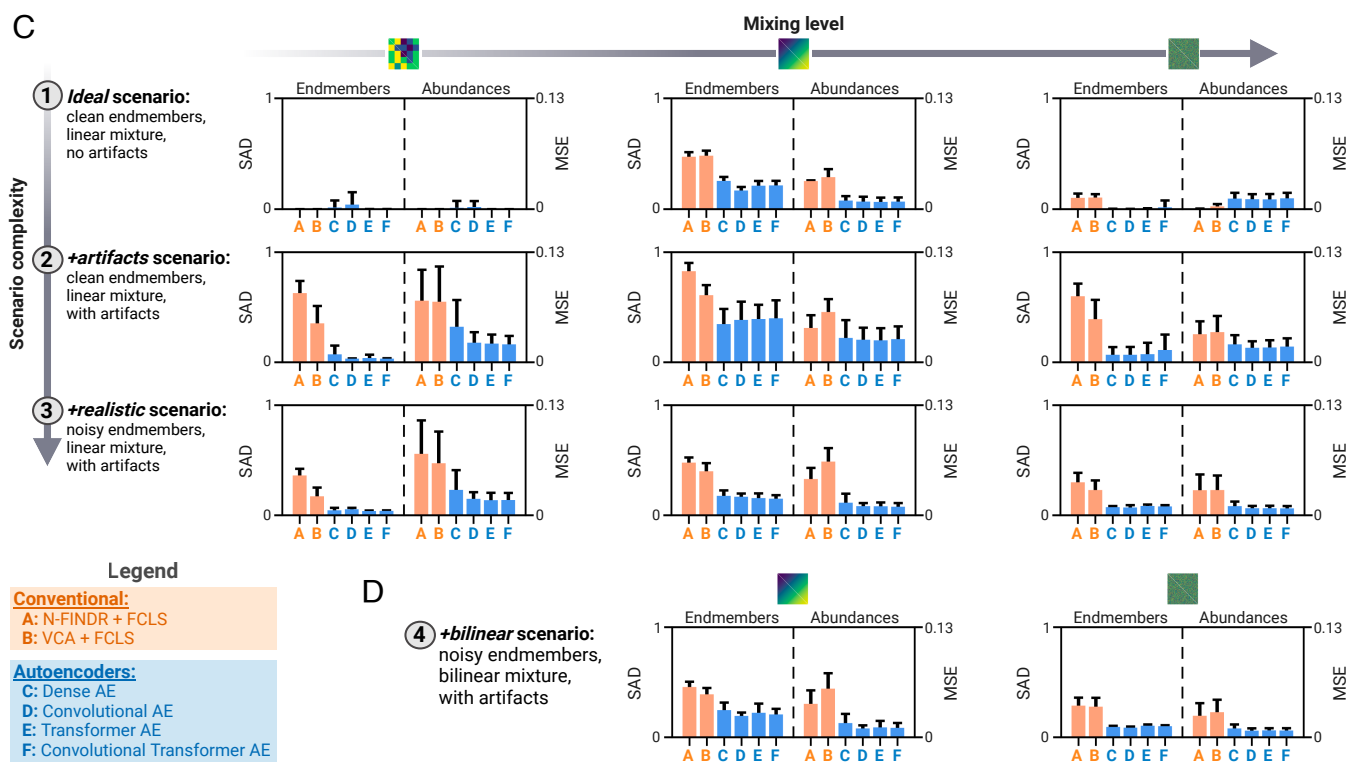
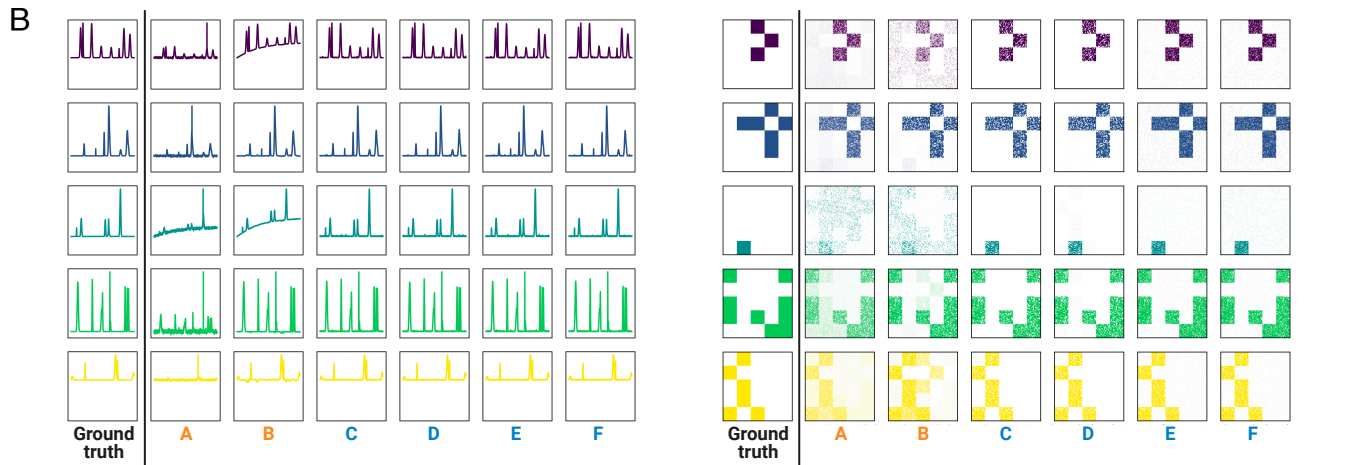
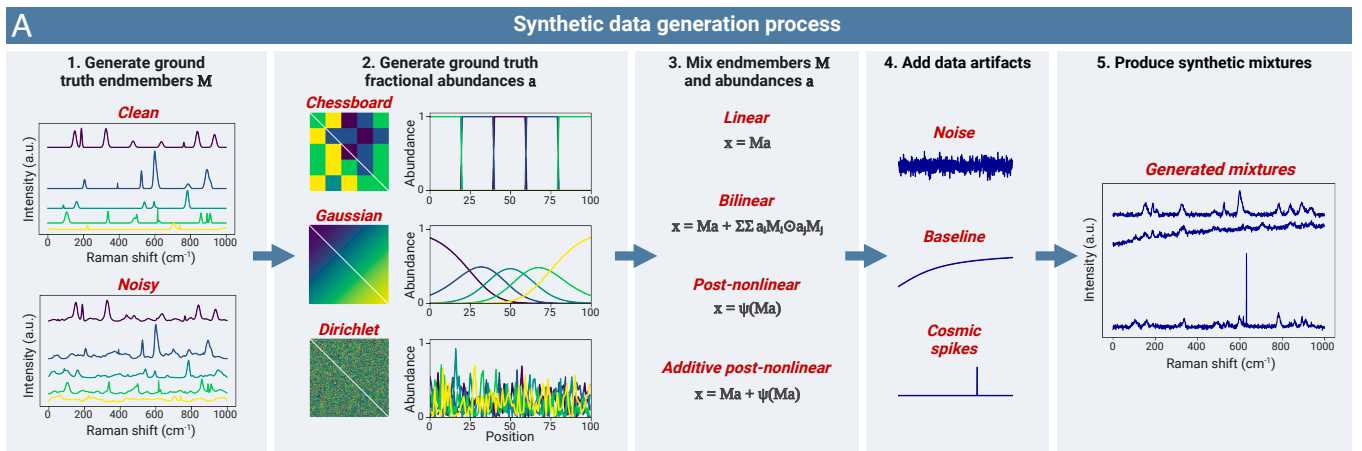


Fig. 2. Benchmarking autoencoders on synthetic Raman mixtures. (A) Schematic of our synthetic data generation workflow, which allows us to create synthetic mixtures with known ground truth. (B) Representative unmixing results (endmembers (Left), fractional abundances (Right)) obtained by each of the six algorithms considered here when applied to one of our synthetic datasets (Chessboard scene, +artifacts scenario). (C and D) Summary of the unmixing performance of the six algorithms on 11 types of synthetic datasets, which span variable mixing levels and complexity. Linear mixtures are shown in C, and bilinear mixtures in D. CI are given by one SD around the sample mean (each bar represents $n = 25$ evaluations; 5 datasets with 5 model repetitions each).

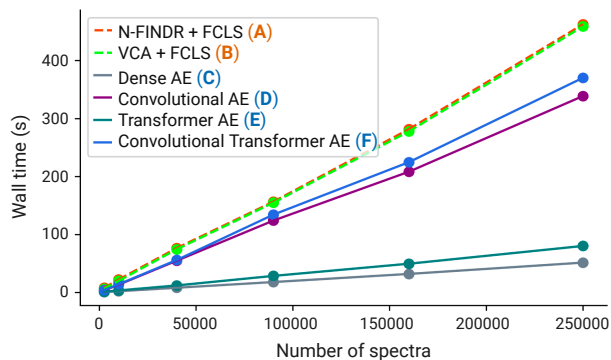


Fig. 3. Computational efficiency analysis. The computational cost (measured as wall time) of autoencoders and conventional methods on synthetic datasets (*Chessboard* scene, *+artifacts* scenario) as the number of spectra is increased. Each dot represents the average across three evaluations (CI based on one SD are small and not visible to the eye). AE models are equipped with decoders for linear unmixing.

to the bilinear mixture model by merely adapting the decoder architecture. Again, we observe that all four AE models provide a substantial improvement in unmixing accuracy compared to methods like N-FINDR+FCLS and VCA+FCLS for both endmember and abundance estimation.

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 250,000 spectra. To be fair to conventional algorithms, we include the full training time for autoencoders and we use standard CPU computation to avoid any advantage from GPU acceleration. Fig. 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 (85). Hence, AEs provide efficient unmixing, even without utilizing GPU acceleration and parallel processing, which can further enhance their performance.

Validation of Unmixing Autoencoders on Experimental Raman

Data from Sugar Mixtures. To validate the unmixing performance of AEs on real experimental data, we next performed 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 (Fig. 4A). To evaluate different signal-to-noise (SNR) conditions, we acquired high SNR (1,920 spectra) and low SNR (7,680 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 systematically evaluate unmixing algorithms under typical experimental artifacts, such as baseline shifts, environmental noise, and cosmic spikes.

We preprocess the data using RamanSPy (83) and then 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 endmember signatures we obtain from reference spectra measured from five additional pure solutions (Fig. 4B). As with the synthetic data above, we benchmark the performance of our

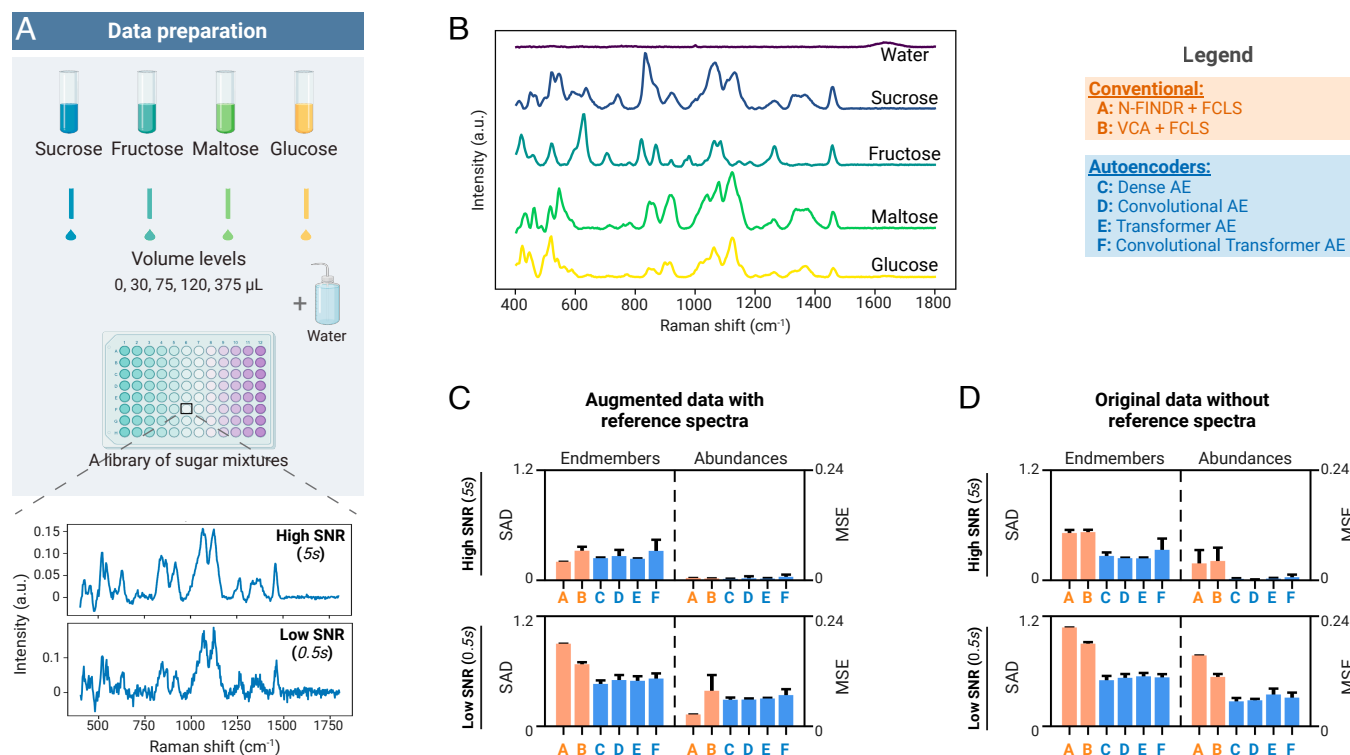


Fig. 4. Experimental validation on Raman spectroscopy data from sugar mixtures. (A) Schematic diagram of sugar mixture preparation. Two sets of Raman measurements are acquired from a library of 240 sugar mixtures—high and low signal-to-noise ratio (SNR) data collected using integration times of 5 s and 0.5 s, respectively. (B) Ground-truth endmember signatures were estimated by taking the median of high SNR reference spectra measured from pure solutions containing a single sugar species. (C and D) Summary of unmixing performance for: (C) idealized scenario where spectra collected from pure solutions are included; and (D) original data without the spectra from pure solutions (i.e., no pure pixels). CI are given as one SD around the sample mean ($n = 5$).

four AE models (linear decoders) against N-FINDR+FCLS and VCA+FCLS.

First, we consider an idealized scenario, purposefully devised to favor conventional methods, whereby endmembers are present in the data. To do this, we include in our analysis the additional reference spectra measured from pure solutions, which serve as pure pixels. When such pure pixels are available, we observe (as expected) that conventional methods (NFINDR+FCLS, VCA+FCLS) perform comparably to AEs on clean, high SNR data (Fig. 4C). Yet, AEs already provide improved performance in low SNR regimes.

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 datasets without including the reference spectra from pure solutions. Our results in Fig. 4D demonstrate that, in such situations, AEs substantially outperform conventional methods in both low and high SNR settings (see *SI Appendix*, Figs. S1 and S2, for additional data).

Application of Unmixing Autoencoders 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 (Fig. 5A) (86). Using Raman spectroscopy

and chemometric techniques, the composition of the cell can be probed to study its morphology in a nondestructive, label-free manner.

After loading and preprocessing the data using RamanSPy (83), we conduct hyperspectral 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 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 DNA, proteins, triglycerides (TAGs), phospholipids (PLPs) and cholesterol esters (see *Materials and Methods*).

Fig. 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 (86), albeit with more distinct endmember signatures and well-defined abundance features (see *SI Appendix*, Figs. S3–S5, for full results). The *Deep Dense AE* model provides cleaner endmember signatures that enable more precise spectral and compositional information (Fig. 5C). Notably, unlike the original VCA+NNLS approach, our AEs detect cholesterol, an important functional and structural component in

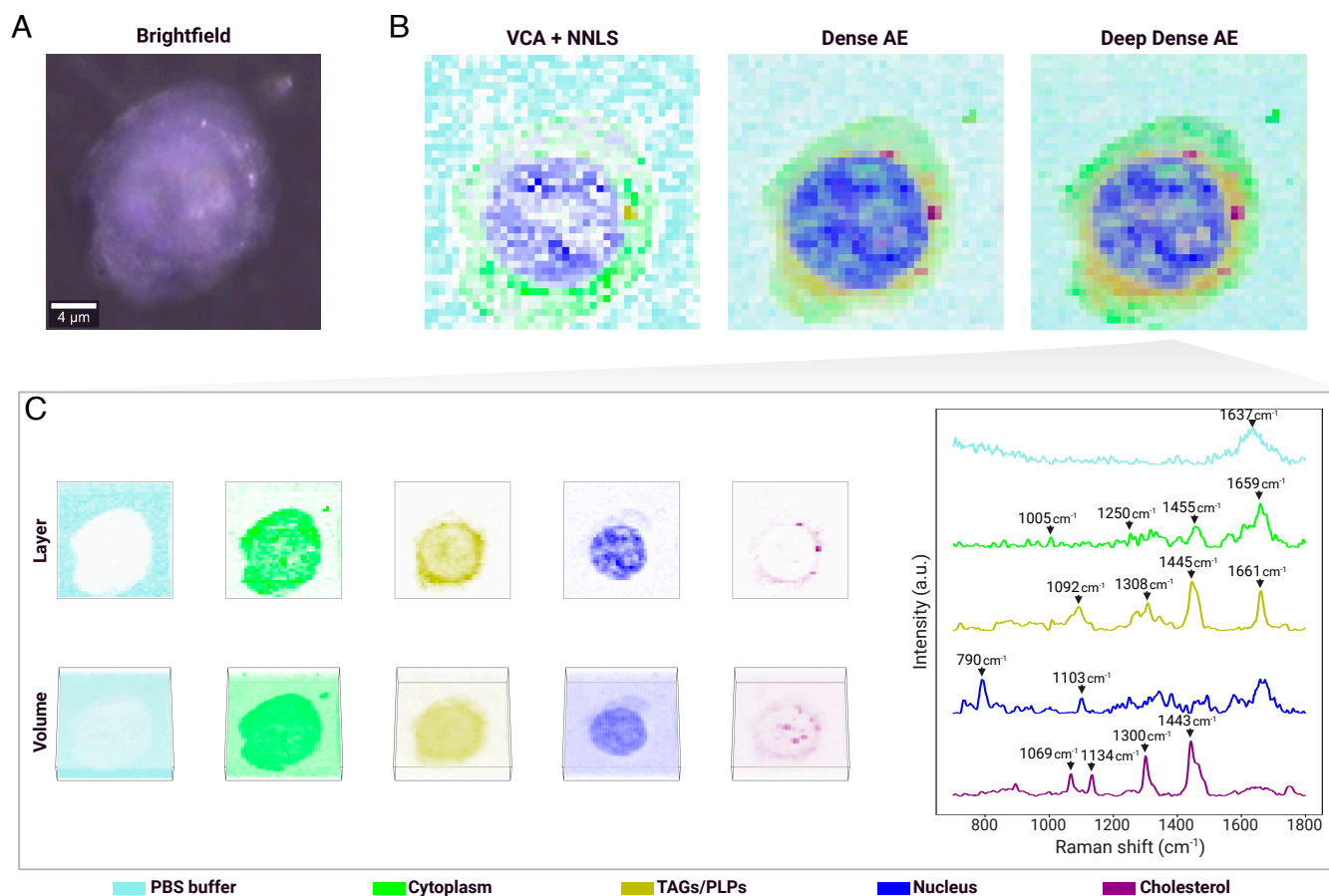


Fig. 5. Volumetric Raman imaging of a THP-1 cell. (A) Brightfield image of the studied THP-1 cell. (B) Reconstruction of a cross-section of the cell obtained by overlaying fractional abundances derived by: VCA+NNLS, our *Dense AE* model, and our *Deep Dense AE* model. (C) Detailed overview of the unmixing results obtained with our *Deep Dense AE* model, displaying the spatial distribution (layer and volume) of the individual fractional abundances (Left) and the associated endmember signatures (Right). Fractional abundance maps are normalized for consistent visualization. Brightfield image and raw Raman spectroscopy data are from Kallepitis et al. (86).

cells, where it plays a key role in membrane fluidity and stability, signaling pathways, and immune response (87–89).

Finally, we remark that our unsupervised autoencoder models do not utilize any ground truth information during training, in contrast with previous supervised machine learning approaches for Raman cell imaging and characterization where specific target concentrations are used as labels to train the model (e.g., using separately acquired fluorescence images) (90–93).

Conclusion

In this work, we have presented an autoencoder-based methodology for hyperspectral unmixing in Raman spectroscopy, which we validated on a wide array of synthetic and real 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 (66) and/or encoders [e.g., stacked and denoising AE architectures (76, 78)], as well as the use of training objectives that better capture spectral reconstruction. In particular, AE architectures that leverage spatial correlation (e.g., via spatial convolutional layers) merit further attention in imaging tasks where chemical species exhibit spatial organization. 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 (94)]. Building on this, the development of foundation models for Raman spectroscopy based on unmixing AEs trained on a multitude of experimental datasets would also warrant further investigation (95).

While our results demonstrate the potential of autoencoders for Raman unmixing, it is important to highlight some of the limitations. Among others, these include potentially higher data requirements for stable performance; the challenging interpretability of learned internal representations; and the sensitivity to hyperparameters requiring careful tuning. Future work should address these issues to enhance the robustness and applicability of autoencoder-based unmixing methods.

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

Materials and Methods

Hyperspectral Unmixing. 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. Raman measurements can be treated as the result of an underlying mixing of n “pure” components, defined by their Raman signatures (endmembers) $\mathbf{m}_i \in \mathbb{R}_+^b$, $i = 1, \dots, n$, and their respective proportions (fractional abundances) $\{\alpha_i\}_{i=1}^n$, $\alpha_i \in \mathbb{R}_+$. Hyperspectral unmixing is the inverse problem of recovering the endmembers and fractional abundances from a measurement \mathbf{x} . The unmixing can be performed with respect to a set of known endmembers (nonblind 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 nonblind unmixing.

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:

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

where $M = [\mathbf{m}_1 \ \mathbf{m}_2 \ \dots \ \mathbf{m}_n]$ is an $b \times n$ nonnegative matrix containing the n endmember signatures, and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)^\top$ is an $n \times 1$ vector storing the corresponding abundances. A random noise term $\epsilon \in \mathbb{R}^b$ is also often added to Eq. 1 to model stochastic variations. The abundances α_i are constrained to be nonnegative [i.e., the abundance nonnegativity 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), $\|\boldsymbol{\alpha}\|_1 = 1$].

The linear mixing simplification allows the development of tractable approaches for unmixing, such as N-FINDR and VCA for endmember identification, and NNLS and FCLS for abundance estimation. N-FINDR and VCA are geometric methods based on the concept of a simplex in Euclidean space. N-FINDR exploits the fact that, under Eq. 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.

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 nonlinear signal contributions can be neglected (48). However, when nonlinear interactions become significant, more intricate models are required (96). 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 (96, 97), including intimate mixture models (98), bilinear models (84, 99, 100), multilinear models (101), or post-nonlinear models (102), among others. A popular bilinear mixing model is the Fan model (84):

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

where \odot is the Hadamard product. However, accounting for nonlinear mixing interactions increases the complexity and computational cost of unmixing (96, 97), 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 (41, 85).

Unmixing Autoencoders. Consider an AE model with an encoder $\mathcal{E} : \mathbb{R}^b \rightarrow \mathbb{R}^m$ that transforms spectra \mathbf{x} into a latent space representation $\mathbf{z} = \mathcal{E}(\mathbf{x})$ of some predefined dimension $m \ll b$, which is subsequently passed through a decoder $\mathcal{D} : \mathbb{R}^m \rightarrow \mathbb{R}^b$ to produce the output reconstruction

$$\hat{\mathbf{x}} = \mathcal{D}(\mathbf{z}) = \mathcal{D}(\mathcal{E}(\mathbf{x})). \quad [3]$$

Notice that the decoder \mathcal{D} can be understood as playing the role of a mixing function on the representation \mathbf{z} . For instance, consider a linear decoder \mathcal{D}_{Lin} consisting of a single linear layer defined by a $b \times m$ weight matrix W then we have

$$\hat{\mathbf{x}} = \mathcal{D}_{\text{Lin}}(\mathbf{z}) = W\mathbf{z}. \quad [4]$$

It follows from the formulation of the LMM (Eq. 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.

To reinforce the physical interpretation of unmixing into the AE learning, we enforce relevant constraints, such as the nonnegativity of W and the nonnegativity (ANC) and sum-to-one constraint (ASC) of \mathbf{z} , by applying appropriate choices of activation functions and penalties during training.

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

$$\hat{\mathbf{x}} = \mathcal{D}_{\text{Bilin}}(\mathbf{z}) = W\mathbf{z} + \sum_{k=1}^m \sum_{\substack{l=1, \\ l \neq k}}^m z_k \mathbf{w}_k \odot z_l \mathbf{w}_l, \quad [5]$$

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 (66, 79, 80), 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. Finally, note that the AE unmixing framework can be directly adapted for nonblind unmixing by fixing the weight matrix W in the decoder to a given set of predefined endmembers.

Autoencoder Architectures.

Dense AE. This autoencoder employs an encoder comprising two 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 two-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 multihead attention block with two attention heads of size 32 (81), 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 (103) and dropout (10%) (104) after the multihead 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 (Eq. 4). Our bilinear Fan decoder has the same architecture as the linear decoder but also calculates the additional bilinear interaction terms in line with Eq. 5.

Physics-inspired constraints. Fractional abundance constraints are applied through the choice of latent space activation functions. To enforce both ANC and ASC, we apply a softmax activation function in the final layer in each encoder. When only ANC is used, the activation function is changed to a "softly rectified" hyperbolic tangent function given by $\frac{1}{\gamma} \log(1 + e^{\gamma * \tanh(x)})$, with $\gamma = 10$, which we design to ensure abundances are between 0 and 1 but do not necessarily add up to one. To ensure the nonnegativity of endmembers, we constrain the weight matrix of the linear layer in the decoder by clipping negative values to zero during training.

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 $b = 1,000$. 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 amplitudes, widths, and locations, 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 that better reflect subtle peaks and variations present in 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 $\alpha \in \mathbb{R}_+^n$. Here, we set $H = W = 100$, resulting in 10,000 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 (Eq. 1) and bilinear mixtures based on the Fan model (Eq. 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 $\epsilon \in \mathbb{R}^b$ 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 $S \sim h_S \mathcal{U}(0.75, 1.25)$ is added to each spectrum at a band $b_S \sim \mathcal{U}\{2, b - 2\}$. In our experiments: $\sigma_N = 0.1$, $p_B = 0.25$, $h_B = 2$, $p_S = 0.1$, $h_S = 5$.

Representativeness of synthetic datasets. To ensure that the synthetic datasets obtained with our data generator are representative, we conducted analyses comparing statistical and similarity metrics between the synthetic and experimental datasets. The results are summarized in *SI Appendix, Figs. S6 and S7*. We find that our synthetic datasets align well with the real Raman measurements (sugar solutions, cell line) across all considered metrics.

Although our synthetic data generation accounts for various sources of variability (e.g., dark noise, cosmic spikes, baseline variations), synthetic data inherently provide an idealized scenario. It is thus important to exercise caution when extrapolating findings from synthetic data alone. Nonetheless, synthetic data can serve as a valuable validation, particularly when ground truth data are unavailable or nonexistent, and provide a useful initial validation step during method development. As done here with the sugar mixtures, this validation can then be further supported by subsequent experiments using real data.

Model Training and Evaluation. Autoencoders were trained using the Adam optimizer (learning rate 0.001), with spectral angle distance (SAD) (82), see Eq. 7, as a loss function between input and reconstructed spectra.

In the cases where ground-truth information is available (synthetic data and sugar mixtures), the predictive accuracy is quantified according to two measures: i) the MSE between ground-truth and predicted fractional abundances α and $\hat{\alpha}$:

$$\text{MSE}(\alpha, \hat{\alpha}) = \frac{1}{n} \|\alpha - \hat{\alpha}\|_2, \quad [6]$$

and ii) the SAD between ground-truth and predicted endmembers \mathbf{m}_i and $\widehat{\mathbf{m}}_i$:

$$0 \leq \text{SAD}(\mathbf{m}_i, \widehat{\mathbf{m}}_i) = \arccos\left(\frac{\mathbf{m}_i \cdot \widehat{\mathbf{m}}_i}{\|\mathbf{m}_i\|_2 \|\widehat{\mathbf{m}}_i\|_2}\right) \leq 1. \quad [7]$$

Each evaluation is performed after first matching the derived and ground-truth endmembers (and corresponding fractional 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 corresponding fractional abundance estimates to compute the performance metrics.

Note that each dataset is analyzed independently by fitting separate models from scratch.

Analysis of Synthetic Raman Mixtures. 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 1,650 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. 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. AE models were trained for 10 epochs.

Measuring Computational Cost. We profile the computational cost of unmixing methods on synthetic datasets (*ideal* scenario, *Chessboard* scene) of increasing sizes, from 2,500 to 250,000 spectra. The number of endmembers to extract was set to $n = 5$ for all methods. For each experiment, we performed three 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 methods which, by design, do not utilize GPU acceleration.

Analysis of Experimental RS Data from Sugar Mixtures.

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

Sugar mixtures were prepared in standard 96-well plates, with a volume of 375 μL per well. A full factorial experiment was performed comprising 4 volume levels for each sugar (0, 30, 75, and 120 μL), 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 μL 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 μL 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 data collection. The excitation wavelength was 785 nm and the power incident to the samples was 36.3 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 to 3684.8 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 s and 0.5 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 = 1,920 high-SNR measurements (1,960 with reference spectra); and 240 solutions \times 8 measurements \times 4 repetitions = 7,680 low-SNR measurements (7,840 with reference spectra). Ground-truth endmembers signatures were obtained by taking the median (band-wise) of the reference spectra (8 reference spectra for each of the five components in the high-SNR setup, and 32 for each of the five components in the 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 and analysis of sugar data. First, we preprocess each sugar dataset: 1) cropping to the region 400 to 1,800 cm^{-1} ; 2) baseline correction with Adaptive Smoothness Parameter Penalized Least Squares (ASPLS) (105) – 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.

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:

$$\mathcal{L}(\mathbf{x}, \widehat{\mathbf{x}}) = \text{SAD}(\mathbf{x}, \widehat{\mathbf{x}}) + \lambda \text{MSE}(\mathbf{x}, \widehat{\mathbf{x}}), \quad [8]$$

with $\lambda = 1,000$. This term breaks the invariance to scale and leads to better abundance estimation given the weak water endmember (SI Appendix, Table S3). The standard SAD loss was used for experiments on low SNR. Each experiment is repeated for 5 model initializations.

SI Appendix, Table S4, 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.

Analysis of Volumetric RS Data from THP-1 Cell. The volumetric Raman scan of the THP-1 cell (86) was collected using 0.3 s integration time and comprises a z-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 to 1,800 cm^{-1} ; 2) cosmic spike removal using the algorithm in ref. 106 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 (107); 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 $t = 0.001$ (108); 5) global MinMax normalization to the interval $[0, 1]$.

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 (86). 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 Fig. 5B. Layer visualization in Fig. 5C was performed on the seventh z-layer.

Cell organelles were determined based on the following peaks: PBS buffer - 1,637 cm^{-1} (water peak); cytoplasm - 1,005 cm^{-1} (phenylalanine), 1,250 cm^{-1} (Amide III), 1,659 cm^{-1} (Amide I) and 1,445 cm^{-1} (CH deformations of proteins and lipids); TAGs/PLPs - 1,092 cm^{-1} (C-C stretching), 1,308 cm^{-1} (CH_2 twists), 1,445 cm^{-1} (CH deformation), and 1,661 cm^{-1}

(C=C stretching); nucleus/DNA - 790 cm^{-1} (symmetric phosphodiester stretch and ring breathing modes of pyrimidine bases) and 1,103 cm^{-1} (symmetric dioxy-stretch of the phosphate backbone); cholesterol - 1,069 cm^{-1} and 1,134 cm^{-1} (cholesteryl stearate), 1,300 cm^{-1} (CH_2 twists), and 1,443 cm^{-1} (CH deformation) (1, 86, 109).

Implementation. We conducted our analyses in Python, using TensorFlow (110) for autoencoder model development and training, and the RamanSPy package (83) for unmixing with conventional methods, data loading and management, preprocessing, and plotting.

Data, Materials, and Software Availability. The original data sets presented in this work, both synthetic data and experimental data from sugar mixtures, are available at <https://doi.org/10.5281/zenodo.10779223> (111). Previously published data from a THP-1 cell used in this work is available at <https://doi.org/10.5281/zenodo.256329> (scan '001') (86). The code and associated scripts developed for this work are available at <https://github.com/barahona-research-group/raman-unmixing-aes> (112). The synthetic data generator has also been integrated into the RamanSPy Python package (<https://github.com/barahona-research-group/RamanSPy>) (83).

ACKNOWLEDGMENTS. D.G. and G.P. are supported by UK Research and Innovation [UKRI Centre for Doctoral Training in AI for Healthcare Grant No. EP/S023283/1]. Á.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). We thank Dr Akemi Nogiwa Valdez for proofreading and data management support. Figures were assembled in BioRender.

Author affiliations: ^aDepartment of Computing, Faculty of Engineering, Imperial College London, London SW7 2AZ, United Kingdom; ^bUKRI Centre for Doctoral Training in AI for Healthcare, Imperial College London, London SW7 2AZ, United Kingdom; ^cDepartment of Materials, Faculty of Engineering, Imperial College London, London SW7 2AZ, United Kingdom; ^dDepartment of Bioengineering, Faculty of Engineering, Imperial College London, London SW7 2AZ, United Kingdom; ^eInstitute of Biomedical Engineering, Faculty of Engineering, Imperial College London, London SW7 2AZ, United Kingdom; ^fMedical Sciences Division, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford OX1 3QU, United Kingdom; ^gMathematical, Physical & Life Sciences Division, Department of Engineering Science, University of Oxford, Oxford OX1 3QU, United Kingdom; ^hMedical Sciences Division and Mathematical, Physical & Life Sciences Division, Kavli Institute for Nanoscience Discovery, University of Oxford, Oxford OX1 3QU, United Kingdom; and ⁱDepartment of Mathematics, Faculty of Natural Sciences, Imperial College London, London SW7 2AZ, United Kingdom

Author contributions: D.G., Á.F.-G., S.V.P., R.X., M.M.S., and M.B. designed research; D.G., Á.F.-G., S.V.P., and M.B. performed research; D.G. developed methods, implemented software, and analyzed data; Á.F.-G. and G.P. contributed new reagents/analytic tools; D.G. and M.B. wrote the paper, with input from all authors.

1. Z. Movasaghi, S. Rehman, I. U. Rehman, Raman spectroscopy of biological tissues. *Appl. Spectrosc. Rev.* **42**, 493–541 (2007).
2. A. C. S. Talari, Z. Movasaghi, S. Rehman, I. U. Rehman, Raman spectroscopy of biological tissues. *Appl. Spectrosc. Rev.* **50**, 46–111 (2015).
3. H. J. Butler *et al.*, Using Raman spectroscopy to characterize biological materials. *Nat. Protoc.* **11**, 664–687 (2016).
4. R. L. McCreery, *Raman Spectroscopy for Chemical Analysis* (John Wiley & Sons, 2005).
5. E. Smith, G. Dent, *Modern Raman Spectroscopy: A Practical Approach* (John Wiley & Sons, 2019).
6. J. A. Koningstein, *Introduction to the Theory of the Raman Effect* (Springer Science & Business Media, 2012).
7. H. A. Szymanski, *Raman Spectroscopy: Theory and Practice* (Springer Science & Business Media, 2012).
8. N. Colthup, *Introduction to Infrared and Raman Spectroscopy* (Introduction to infrared and Raman spectroscopy, 2012).
9. R. R. Jones, D. C. Hooper, L. Zhang, D. Wolverson, V. K. Valev, Raman techniques: Fundamentals and frontiers. *Nanoscale Res. Lett.* **14**, 1–34 (2019).
10. T. W. Bocklitz, S. Guo, O. Ryabchykov, N. Vogler, J. Popp, Raman based molecular imaging and analytics: A magic bullet for biomedical applications? *Anal. Chem.* **88**, 133–151 (2016).
11. S. Schlucker, Surface-enhanced Raman spectroscopy: Concepts and chemical applications. *Angew. Chem. Int. Ed.* **53**, 4756–4795 (2014).
12. K. Dodo, K. Fujita, M. Sodeoka, Raman spectroscopy for chemical biology research. *J. Am. Chem. Soc.* **144**, 19651–19667 (2022).
13. G. Pezzotti, Raman spectroscopy in cell biology and microbiology. *J. Raman Spectrosc.* **52**, 2348–2443 (2021).
14. R. Smith, K. L. Wright, L. Ashton, Raman spectroscopy: An evolving technique for live cell studies. *Analyst* **141**, 3590–3600 (2016).
15. D. W. Shipp, F. Sinjab, I. Notingher, Raman spectroscopy: Techniques and applications in the life sciences. *Adv. Opt. Photonics* **9**, 315–428 (2017).
16. D. Cialla-May, X. S. Zheng, K. Weber, J. Popp, Recent progress in surface-enhanced Raman spectroscopy for biological and biomedical applications: From cells to clinics. *Chem. Soc. Rev.* **46**, 3945–3961 (2017).
17. K. Kong, C. Kendall, N. Stone, I. Notingher, Raman spectroscopy for medical diagnostics—From in-vitro biofluid assays to in-vivo cancer detection. *Adv. Drug Deliv. Rev.* **89**, 121–134 (2015).
18. K. J. Ember *et al.*, Raman spectroscopy and regenerative medicine: A review. *npj Regen. Med.* **2**, 12 (2017).
19. I. Pence, A. Mahadevan-Jansen, Clinical instrumentation and applications of Raman spectroscopy. *Chem. Soc. Rev.* **45**, 1958–1979 (2016).
20. V. Balan *et al.*, Vibrational spectroscopy fingerprinting in medicine: From molecular to clinical practice. *Materials* **12**, 2884 (2019).
21. G. W. Auner *et al.*, Applications of Raman spectroscopy in cancer diagnosis. *Cancer Metastasis Rev.* **37**, 691–717 (2018).
22. A. Mahadevan-Jansen, R. R. Richards-Kortum, Raman spectroscopy for the detection of cancers and precancers. *J. Biomed. Opt.* **1**, 31–70 (1996).
23. S. Tanwar, S. K. Paidi, R. Prasad, R. Pandey, I. Barman, Advancing Raman spectroscopy from research to clinic: Translational potential and challenges. *Spectrochim. Acta Part A, Mol. Biomol. Spectrosc.* **260**, 119957 (2021).
24. Á. Fernández-Galiana, O. Bibikova, S. V. Pedersen, M. M. Stevens, Fundamentals and applications of Raman-based techniques for the design and development of active biomedical materials. *Adv. Mater.* **2210807** (2023). [10.1002/adma.202210807](https://doi.org/10.1002/adma.202210807).
25. C. S. Kumar, *Raman Spectroscopy for Nanomaterials Characterization* (Springer Science & Business Media, 2012).
26. W. Wang, H. Zhang, Y. Yuan, Y. Guo, S. He, Research progress of Raman spectroscopy in drug analysis. *AAPS PharmSciTech* **19**, 2921–2928 (2018).
27. A. Paudel, D. Rajjada, J. Rantanen, Raman spectroscopy in pharmaceutical product design. *Adv. Drug Deliv. Rev.* **89**, 3–20 (2015).
28. T. Vankeirsbilck *et al.*, Applications of Raman spectroscopy in pharmaceutical analysis. *Trends Anal. Chem.* **21**, 869–877 (2002).
29. R. A. Halvorson, P. J. Vikesland, Surface-enhanced Raman spectroscopy (SERS) for environmental analyses. *Environ. Sci. Technol.* **44**, 7749–7755 (2010).
30. T. T. Ong, E. W. Blanch, O. A. Jones, Surface enhanced Raman spectroscopy in environmental analysis, monitoring and assessment. *Sci. Total Environ.* **720**, 137601 (2020).
31. L. R. Terry *et al.*, Applications of surface-enhanced Raman spectroscopy in environmental detection. *Anal. Sci. Adv.* **3**, 113–145 (2022).
32. Y. S. Li, J. S. Church, Raman spectroscopy in the analysis of food and pharmaceutical nanomaterials. *J. Food Drug Anal.* **22**, 29–48 (2014).
33. S. Pang, T. Yang, L. He, Review of surface enhanced Raman spectroscopic (SERS) detection of synthetic chemical pesticides. *Trends Anal. Chem.* **85**, 73–82 (2016).
34. J. M. Chalmers, H. G. M. Edwards, M. D. Hargreaves, eds., *Infrared and Raman Spectroscopy in Forensic Science* (Wiley Online Library, 2012).
35. S. R. Khandasamy *et al.*, Bloodstains, paintings, and drugs: Raman spectroscopy applications in forensic science. *Forensic Chem.* **8**, 111–133 (2018).
36. E. L. Izake, Forensic and homeland security applications of modern portable Raman spectroscopy. *Forensic Sci. Int.* **202**, 1–8 (2010).
37. O. Ryabchykov, S. Guo, T. Bocklitz, Analyzing Raman spectroscopic data. *Phys. Sci. Rev.* **4**, 20170043 (2018).
38. S. Guo, J. Popp, T. Bocklitz, Chemometric analysis in Raman spectroscopy from experimental design to machine learning-based modeling. *Nat. Protoc.* **16**, 5426–5459 (2021).
39. R. Gautam, S. Vanga, F. Ariese, S. Umaphathy, Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Tech. Instrum.* **2**, 1–38 (2015).
40. H. J. Byrne, P. Knief, M. E. Keating, F. Bonnier, Spectral pre and post processing for infrared and Raman spectroscopy of biological tissues and cells. *Chem. Soc. Rev.* **45**, 1865–1878 (2016).
41. X. Li, R. Zhou, Y. Xu, X. Wei, Y. He, Spectral unmixing combined with Raman imaging, a preferable analytic technique for molecule visualization. *Appl. Spectrosc. Rev.* **52**, 417–438 (2017).
42. V. Olmos *et al.*, Relevant aspects of unmixing/resolution analysis for the interpretation of biological vibrational hyperspectral images. *Trends Anal. Chem.* **94**, 130–140 (2017).
43. M. E. Winter, N-FINDR, An algorithm for fast autonomous spectral end-member determination in hyperspectral data in Imaging Spectrometry V. *SPIE* **3753**, 266–275 (1999).
44. J. M. Nascimento, J. M. Dias, Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **43**, 898–910 (2005).
45. C. L. Lawson, R. J. Hanson, *Solving Least Squares Problems* (SIAM, 1995).
46. D. C. Heinz *et al.*, Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **39**, 529–545 (2001).
47. M. Hedegaard *et al.*, Spectral unmixing and clustering algorithms for assessment of single cells by Raman microscopic imaging. *Theor. Chem. Acc.* **130**, 1249–1260 (2011).
48. N. Keshava, J. F. Mustard, Spectral unmixing. *IEEE Signal Process. Mag.* **19**, 44–57 (2002).
49. A. T. Harris, Spectral mapping tools from the earth sciences applied to spectral microscopy data. *Cytometry, Part A* **69**, 872–879 (2006).
50. F. Lussier, V. Thibault, B. Charron, G. O. Wallace, J. F. Masson, Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering. *Trends Anal. Chem.* **124**, 115796 (2020).

51. J. Yang *et al.*, Deep learning for vibrational spectral analysis: Recent progress and a practical guide. *Anal. Chim. Acta* **1081**, 6–17 (2019).
52. R. Luo, J. Popp, T. Bocklitz, Deep learning for Raman spectroscopy: A review. *Analytica* **3**, 287–301 (2022).
53. C. A. Meza Ramirez, M. Greenop, L. Ashton, I. U. Rehman, Applications of machine learning in spectroscopy. *Appl. Spectrosc. Rev.* **56**, 733–763 (2021).
54. X. Bi, L. Lin, Z. Chen, J. Ye, Artificial intelligence for surface-enhanced Raman spectroscopy. *Small Methods* **8**, 2301243 (2024).
55. L. Pan, P. Zhang, C. Daengngam, S. Peng, M. Chongcheawchamnan, A review of artificial intelligence methods combined with Raman spectroscopy to identify the composition of substances. *J. Raman Spectrosc.* **53**, 6–19 (2022).
56. Z. Zhao *et al.*, ConInceDeep: A novel deep learning method for component identification of mixture based on Raman spectroscopy. *Chemom. Intell. Lab. Syst.* **234**, 104757 (2023).
57. J. Liu *et al.*, Deep convolutional neural networks for Raman spectrum recognition: A unified solution. *Analyst* **142**, 4067–4074 (2017).
58. L. Pan *et al.*, Identification of complex mixtures for Raman spectroscopy using a novel scheme based on a new multi-label deep neural network. *IEEE Sens. J.* **21**, 10834–10843 (2021).
59. W. J. Thrift, R. Ragan, Quantification of analyte concentration in the single molecule regime using convolutional neural networks. *Anal. Chem.* **91**, 13337–13342 (2019).
60. Z. Wang *et al.*, Deep learning-based Raman spectroscopy qualitative analysis algorithm: A convolutional neural network and transformer approach. *Talanta* **275**, 126138 (2024).
61. O. C. Koyun *et al.*, RamanFormer: A transformer-based quantification approach for Raman mixture components. *ACS Omega* **9**, 23241–23251 (2024).
62. B. Palsson, J. R. Sveinsson, M. O. Ulfarsson, Blind hyperspectral unmixing using autoencoders: A critical comparison. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **15**, 1340–1372 (2022).
63. S. Zhang *et al.*, "Recent advances in hyperspectral unmixing using sparse techniques and deep learning" in *Hyperspectral Image Analysis: Advances in Machine Learning and Signal Processing*, S. Prasad, J. Chanussot, Eds. (Springer, Cham, Switzerland, 2020).
64. Y. Wang, C. M. Albrecht, N. A. A. Braham, L. Mou, X. X. Zhu, Self-supervised learning in remote sensing: A review. *IEEE Geosci. Remote Sens. Mag.* **10**, 213–247 (2022).
65. J. S. Bhatt, M. V. Joshi, "Deep learning in hyperspectral unmixing: A review" in *IGARSS* (2020), pp. 2189–2192.
66. J. Chen, M. Zhao, X. Wang, C. Richard, S. Rahardja, Integration of physics-based and data-driven models for hyperspectral image unmixing: A summary of current methods. *IEEE Signal Process. Mag.* **40**, 61–74 (2023).
67. N. Burzynski *et al.*, "Deep learning techniques for unmixing of hyperspectral stimulated raman scattering images" in *2021 IEEE International Conference on Big Data (Big Data)* (IEEE, 2021), pp. 5862–5864.
68. D. Boildieu, D. Helbert, A. Magnaudeix, P. Leproux, P. Carré, "Multivariate curve resolution with autoencoders for CARS microspectroscopy" in *Computational Imaging Conference, IS&T Electronic Imaging* (2023).
69. I. Goodfellow, Y. Bengio, A. Courville, *Autoencoders in Deep Learning* (MIT Press, 2016), pp. 499–523.
70. X. Zhang, Y. Sun, J. Zhang, P. Wu, L. Jiao, Hyperspectral unmixing via deep convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **15**, 1755–1759 (2018).
71. B. Palsson, M. O. Ulfarsson, J. R. Sveinsson, Convolutional autoencoder for spectral-spatial hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **59**, 535–549 (2020).
72. M. M. Elkholy, M. Mostafa, H. M. Ebied, M. F. Tolba, Hyperspectral unmixing using deep convolutional autoencoder. *Int. J. Remote Sens.* **41**, 4799–4819 (2020).
73. P. Ghosh, S. K. Roy, B. Koirala, B. Rasti, P. Scheunders, Hyperspectral unmixing using transformer network. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–16 (2022).
74. S. Ozkan, B. Kaya, G. B. Akar, Endnet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **57**, 482–496 (2018).
75. Y. Ou, H. Qi, uDAS: An untied denoising autoencoder with sparsity for spectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **57**, 1698–1712 (2018).
76. Y. Su, A. Marinoni, J. Li, J. Plaza, P. Gamba, Stacked nonnegative sparse autoencoders for robust hyperspectral unmixing. *IEEE Geosci. Remote Sens. Lett.* **15**, 1427–1431 (2018).
77. Y. Su, A. Marinoni, J. Li, A. Plaza, P. Gamba, "Nonnegative sparse autoencoder for robust endmember extraction from remotely sensed hyperspectral images" in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (IEEE, 2017), pp. 205–208.
78. Y. Qu, R. Guo, H. Qi, "Spectral unmixing through part-based non-negative constraint denoising autoencoder" in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (IEEE, 2017), pp. 209–212.
79. K. T. Shahid, I. D. Schizas, Unsupervised hyperspectral unmixing via nonlinear autoencoders. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–13 (2021).
80. M. Zhao, M. Wang, J. Chen, S. Rahardja, Hyperspectral unmixing for additive nonlinear models with a 3-D-CNN autoencoder network. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–15 (2021).
81. A. Vaswani *et al.*, "Attention is all you need" in *Advances in Neural Information Processing Systems* **30**, I. Guyon *et al.*, Eds. (NIPS, 2017).
82. F. A. Kruse *et al.*, The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data. *Remote Sens. Environ.* **44**, 145–163 (1993).
83. D. Georgiev *et al.*, RamanSPy: An open-source Python package for integrative Raman spectroscopy data analysis. *Anal. Chem.* **96**, 8492–8500 (2024).
84. W. Fan, B. Hu, J. Miller, M. Li, Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data. *Int. J. Remote Sens.* **30**, 2951–2962 (2009).
85. J. M. Bioucas-Dias *et al.*, Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **5**, 354–379 (2012).
86. C. Kallepitis *et al.*, Quantitative volumetric Raman imaging of three dimensional cell cultures. *Nat. Commun.* **8**, 14843 (2017).
87. L. Krihtharides, A. Christian, G. Stoudt, D. Morel, G. H. Rothblat, Cholesterol metabolism and efflux in human THP-1 macrophages. *Arterioscler. Thromb. Vasc. Biol.* **18**, 1589–1599 (1998).
88. A. R. Tall, L. Yvan-Charvet, Cholesterol, inflammation and innate immunity. *Nat. Rev. Immunol.* **15**, 104–116 (2015).
89. A. K. Saha *et al.*, Cellular cholesterol regulates monocyte deformation. *J. Biomech.* **52**, 83–88 (2017).
90. B. Manifold, S. Men, R. Hu, D. Fu, A versatile deep learning architecture for classification and label-free prediction of hyperspectral images. *Nat. Mach. Intell.* **3**, 306–315 (2021).
91. J. Zhang, J. Zhao, H. Lin, Y. Tan, J. X. Cheng, High-speed chemical imaging by dense-net learning of femtosecond stimulated Raman scattering. *J. Phys. Chem. Lett.* **11**, 8573–8578 (2020).
92. Z. Wei *et al.*, Pixel-level multimodal fusion deep networks for predicting subcellular organelle localization from label-free live-cell imaging. *Front. Genet.* **13**, 1002327 (2022).
93. Y. Zhang *et al.*, From genotype to phenotype: Raman spectroscopy and machine learning for label-free single-cell analysis. *ACS Nano* **18**, 18101–18117 (2024).
94. C. C. Horgan *et al.*, High-throughput molecular imaging via deep-learning-enabled Raman spectroscopy. *Anal. Chem.* **93**, 15850–15860 (2021).
95. R. Bommasani *et al.*, On the opportunities and risks of foundation models. arXiv [Preprint] (2021). <https://doi.org/10.48550/arXiv.2108.07258> (Accessed 8 October 2024).
96. N. Dobigeon *et al.*, Nonlinear unmixing of hyperspectral images: Models and algorithms. *IEEE Signal Process. Mag.* **31**, 82–94 (2013).
97. R. Heylen, M. Parente, P. Gader, A review of nonlinear hyperspectral unmixing methods. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**, 1844–1868 (2014).
98. B. Hapke, Bidirectional reflectance spectroscopy: 1 theory. *J. Geophys. Res. Solid Earth* **86**, 3039–3054 (1981).
99. A. Halimi, Y. Altmann, N. Dobigeon, J. Y. Tourneret, Nonlinear unmixing of hyperspectral images using a generalized bilinear model. *IEEE Trans. Geosci. Remote Sens.* **49**, 4153–4162 (2011).
100. A. Halimi, Y. Altmann, N. Dobigeon, J. Y. Tourneret, "Unmixing hyperspectral images using the generalized bilinear model" in *2011 IEEE International Geoscience and Remote Sensing Symposium* (IEEE, 2011), pp. 1886–1889.
101. R. Heylen, P. Scheunders, A multilinear mixing model for nonlinear spectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **54**, 240–251 (2015).
102. Y. Altmann, A. Halimi, N. Dobigeon, J. Y. Tourneret, Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery. *IEEE Trans. Image Process.* **21**, 3017–3025 (2012).
103. J. L. Ba, J. R. Kiros, G. E. Hinton, Layer normalization. arXiv [Preprint] (2016). <https://arxiv.org/abs/1607.06450> (Accessed 8 October 2024).
104. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
105. F. Zhang *et al.*, Baseline correction for infrared spectra using adaptive smoothness parameter penalized least squares method. *Spectrosc. Lett.* **53**, 222–233 (2020).
106. D. A. Whitaker, K. Hayes, A simple algorithm for despiking Raman spectra. *Chemom. Intell. Lab. Syst. Syst.* **179**, 82–84 (2018).
107. A. Savitzky, M. J. Golay, Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627–1639 (1964).
108. P. H. Eilers, H. F. Boelens, Baseline correction with asymmetric least squares smoothing. *Leiden Univ. Med. Centre Rep.* **1**, 5 (2005).
109. X. Zhang *et al.*, Label-free live-cell imaging of nucleic acids using stimulated Raman scattering microscopy. *ChemPhysChem* **13**, 1054–1059 (2012).
110. M. Abadi *et al.*, "A system for large-scale machine learning" in *OSDI* (Savannah, GA, 2016), vol. 16, pp. 265–283.
111. D. Georgiev *et al.*, Research data supporting "Hyperspectral unmixing for Raman spectroscopy via physics-constrained autoencoders". Zenodo. <https://doi.org/10.5281/zenodo.10779223>. Deposited 16 October 2024.
112. D. Georgiev *et al.*, Codebase supporting "Hyperspectral unmixing for Raman spectroscopy via physics-constrained autoencoders". GitHub. <https://github.com/barahona-research-group/raman-unmixing-aes>. Deposited 16 October 2024.