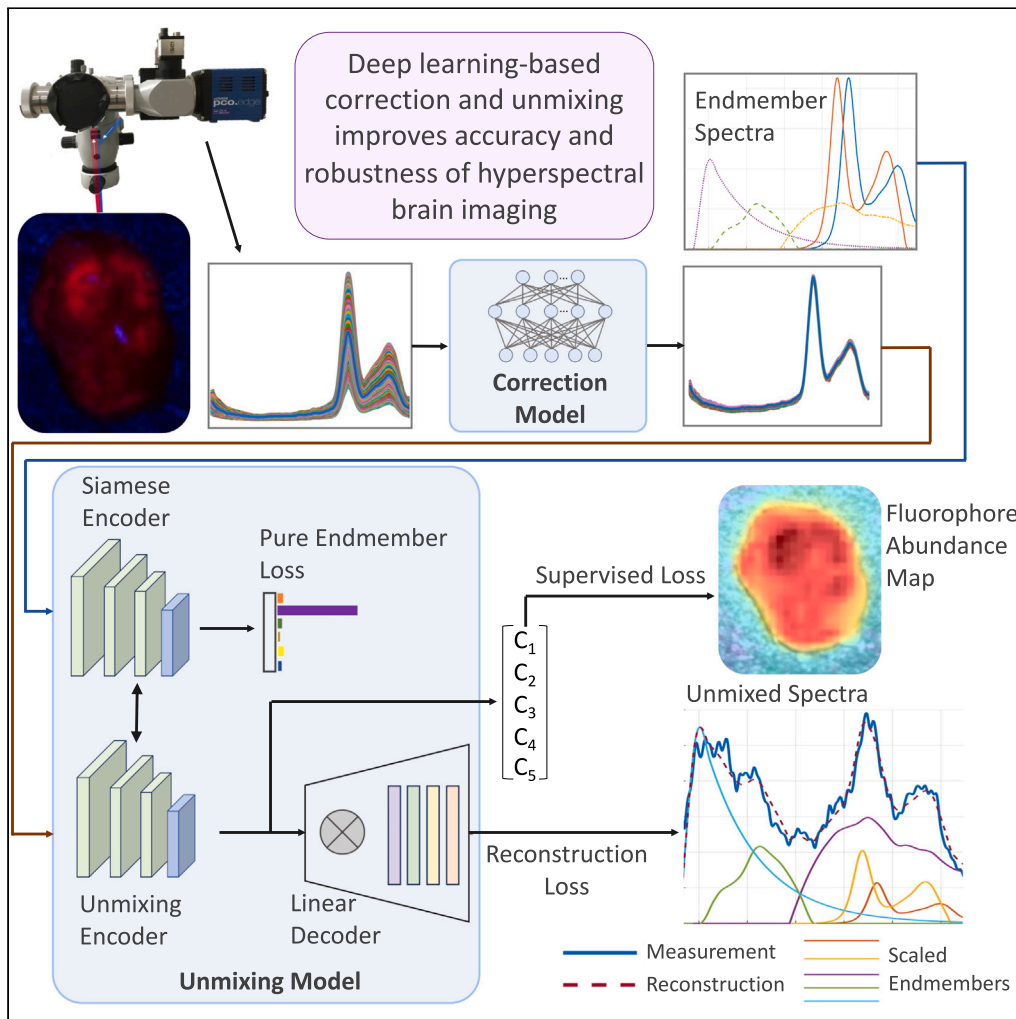**Article**

# Deep learning-based hyperspectral image correction and unmixing for brain tumor surgery

David Black,
Jaidev Gill,
Andrew Xie, Benoit
Liquet, Antonio Di
Ieva, Walter
Stummer, Eric
Suero Molina

eric.suero@ukmuenster.de

**Highlights**

Hyperspectral brain images are affected by tissue geometric and optical properties

This can disrupt quantization of fluorophore content and invalidate measurements

We introduce two semi-supervised deep learning models to correct for these effects

The models outperform existing methods and can thus improve brain tumor surgery

## Article

# Deep learning-based hyperspectral image correction and unmixing for brain tumor surgery

David Black,[1,8] Jaidev Gill,[2,8] Andrew Xie,[2,8] Benoit Liquet,[3,4] Antonio Di Ieva,[5,6] Walter Stummer,[7] and Eric Suero Molina[5,6,7,9,*]

### SUMMARY

**Hyperspectral imaging for fluorescence-guided brain tumor resection improves visualization of tissue differences, which can ameliorate patient outcomes. However, current methods do not effectively correct for heterogeneous optical and geometric tissue properties, leading to less accurate results. We propose two deep learning models for correction and unmixing that can capture these effects. While one is trained with protoporphyrin IX (PpIX) concentration labels, the other is semi-supervised. The models were evaluated on phantom and pig brain data with known PpIX concentration; the supervised and semi-supervised models achieved Pearson correlation coefficients (phantom, pig brain) between known and computed PpIX concentrations of (0.997, 0.990) and (0.98, 0.91), respectively. The classical approach achieved (0.93, 0.82). The semi-supervised approach also generalizes better to human data, achieving a 36% lower false-positive rate for PpIX detection and giving qualitatively more realistic results than existing methods. These results show promise for using deep learning to improve hyperspectral fluorescence-guided neurosurgery.**

### INTRODUCTION

Due to their infiltrative growth, identifying glioma margins during brain surgery is extremely difficult, if not impossible. However, surgical adjuncts such as fluorescence guidance can maximize resection rates, thus improving patient outcomes.[1,2] 5-Aminolevulinic acid (5-ALA) is a Food and Drug Administration-approved tissue marker for high-grade glioma.[3] 5-ALA is administered orally 4 h before induction of anesthesia for fluorescence-guided resection of malignant gliomas; this drug is metabolized preferentially in tumor cells to protoporphyrin IX (PpIX), a precursor on the heme synthesis pathway.[4] PpIX fluoresces bright red, with a primary peak at 634 nm, when excited with blue light at 405 nm. In this way, tumors that are otherwise difficult to distinguish from healthy tissue can sometimes be identified by their red glow under blue illumination. This allows for a more complete resection and thus improved progression and overall survival.[2,5] However, the fluorescence is often not visible in lower-grade glioma or in infiltrating margins of tumors.[6,7] In these cases, the PpIX fluoresces at a similar intensity to other endogenous fluorophores, known as autofluorescence, and remains indistinguishable.

Hyperspectral imaging (HSI) is, therefore, an active research area, as it allows the PpIX content to be isolated from autofluorescence. HSI systems capture three-dimensional data cubes in which each 2D slice is an image of the scene at a particular wavelength. A fluorescence intensity spectrum is obtained by tracing a pixel through the cube's third dimension. Thus, the emission spectrum of light is measured at every pixel.[8] This technology is used in many fields, including food safety and research,[9] materials science,[10] agriculture,[11] and space exploration,[12] as it provides rich spatial and spectral information without disturbing the system. In medical HSI, each spatial pixel contains a combination of fluorescing molecules or fluorophores. Assuming a linear model that neglects multiple scattering[13] and other effects, the measured fluorescence spectrum at that pixel ($\Phi_{Fluo}$) is thus a linear combination of the emission spectra of $K$ potentially present fluorophores ($\Phi_{Spec,k}$), also called endmember spectra,[14] as shown in Equation 1 (ignoring noise). With a priori knowledge of the endmember spectra, linear regression techniques have been employed to determine the relative abundances ($c_k$) of the endmembers[15] in a given spectrum.

$$\Phi_{Fluo} = \sum_{k=1}^{K} c_k \Phi_{Spec,k}$$

(Equation 1)

[1]Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada
[2]Engineering Physics, University of British Columbia, Vancouver, BC, Canada
[3]School of Mathematical and Physical Sciences, Macquarie University, Sydney, NSW, Australia
[4]Laboratoire de Mathématiques et de ses Applications, E2S-UPPA, Université de Pau & Pays de L'Adour, Pau, France
[5]Computational NeuroSurgery (CNS) Lab, Macquarie University, Sydney, NSW, Australia
[6]Macquarie Medical School, Macquarie University, Sydney, NSW, Australia
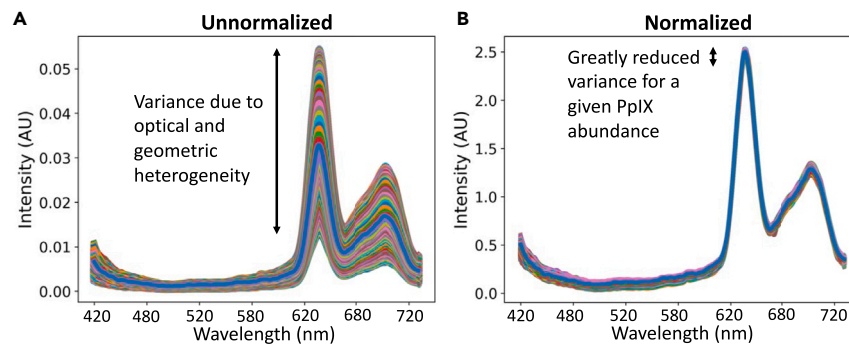[7]Department of Neurosurgery, University Hospital Münster, Münster, Germany
[8]These authors contributed equally
[9]Lead contact
*Correspondence: eric.suero@ukmuenster.de
https://doi.org/10.1016/j.isci.2024.111273

**Figure 1. Typical attenuation correction of measured spectra from a phantom of constant PpIX concentration**
(A) shows the raw spectra with large variance, and (B) shows the normalized ones after correction. The variance in the magnitudes is greatly decreased.

Recent advances in HSI for fluorescence-guided surgery have increased our ability to detect tumor regions[8,16] and even classify tissue types based on the endmember abundances.[17,18] They have also been used to study 5-ALA dosage[7] and timing of application,[6] and to improve the imaging devices.[19–21] However, these computations are extremely sensitive to autofluorescence, as well as to artifacts from the optical and topographic properties of the tissue and camera system. To mitigate the latter issue, the measured fluorescence spectra are corrected to account for heterogeneous absorption and scattering by using the measured spectra under white light illumination at the same location. With white light excitation, there is no fluorescence, so the measured spectra depend only on the heterogeneous tissue properties. The spectra can thus be used to correct for these variations. One common method for attenuation correction, called dual-band normalization, involves integrating over two portions of these spectra, raising one to an empirical exponent, and multiplying them to determine a scaling factor.[22] While effective in phantoms,[23] we have found this method to be of limited use in patient data.[16] The pixels are also corrected for their distance from the objective lens since further pixels appear dimmer than closer ones.[24,25] Other methods are also relatively simplistic, linear, and not based on human data.[26] They are thus unable to account for nonlinear effects such as multiple scattering,[13] the dual photostates of PpIX,[4,27] and fluorescence variation due to pH and tumor microenvironment,[16] nor can they entirely correct for the inhomogeneous optical properties of the tissue.[16] These effects may also include wavelength-dependent absorption and scattering variations, which are unmodeled when using a single scaling factor. An example of attenuation correction is shown in Figure 1.

Once the spectra are corrected for optical and topological variations, they must be unmixed into the endmember abundances. In 5-ALA-mediated fluorescence-guided tumor surgery, these likely include the two photostates of PpIX,[4,27] called $PpIX_{620}$ and $PpIX_{634}$, as well as autofluorescence from flavins, lipofuscin, NADH, melanin, collagen, and elastin,[15,28] though there are usually only 3 or 4 endmembers present in any given spectrum.[29] Previous work has commonly used non-negative least squares (NNLS) regression.[4,15–17,30,31] This is simple and fast and guarantees non-negative abundances. Three example unmixings using NNLS are shown in Figure 2. Other papers have proposed Poisson regression[32] to account for the theoretically Poisson-distributed photon emissions[33] or various sparse methods to reduce overfitting and enforce the fact that there are usually only a few fluorophores present in each pixel.[14] However, as mentioned before, all these methods assume linearity in the combination of the endmember spectra. Furthermore, they rely on the attenuation correction to be accurate.

Thus, performing the correction and unmixing in a single-step process that can handle the nonlinearity and complexity of the physical, optical, and biological systems described earlier would be beneficial. For this purpose, deep learning is particularly well suited because each HSI measurement produces a large volume of high-dimensional data. Indeed, deep learning has been explored in detail for HSI, as reviewed by Jia et al.,[34] and for medical applications specifically.[35,36] For brain tumor resection, the technique is very promising,[37] and several studies have used support vector machines, random forest models, and simple convolutional neural networks (CNNs) to segment and classify tissues *in vivo*.[38–40] Other approaches include majority voting-based fusions of k-nearest neighbors (KNNs), hierarchical k-means clustering, and dimensionality reduction techniques such as principal component analysis or t-distributed stochastic neighbor embedding.[41,42] These papers used 61 images from 34 patients with a resulting median macro F1-score of 70% in detecting tumors. Rinesh et al. used KNNs and multilayer perceptrons (MLPs).[43] The HELICoiD (Hyperspectral Imaging Cancer Detection) dataset,[44] which consists of 36 data cubes from 22 patients, has been widely used. For instance, Manni et al. achieved 80% accuracy in classifying tumor, healthy tissue, and blood vessels using a CNN,[45] and Hao et al. combined different deep learning architectures in a multi-step pipeline to reach 96% accuracy in glioblastoma identification.[46] Other methods used pathological slides,[47] with most mentioned based on small datasets.[48]

Given the small datasets, many of these papers have not yet had sufficiently good results to be clinically useful and likely do not generalize very well. This is partly due to the cost of labeling many hyperspectral images. As a result, modern architectures for medical image segmentation, such as U-Net,[49] V-Net,[50] or graph neural networks,[51] have seen little use. Autoencoders[52] or generative adversarial networks[53] can use unsupervised learning for certain tasks to avoid the labeling problem but require large volumes of data. Jia et al. describe some approaches to overcome the lack of data in HSI,[54] and self- or unsupervised approaches have been used in general HSI,[55–57] but not in neurosurgery. In addition, these papers all represent end-to-end attempts to take a raw data cube containing high-grade glioma and output a segmentation. This approach is unlikely to generalize well to other devices, hospitals, or tumor types. Instead, a more fine-grained method may generalize
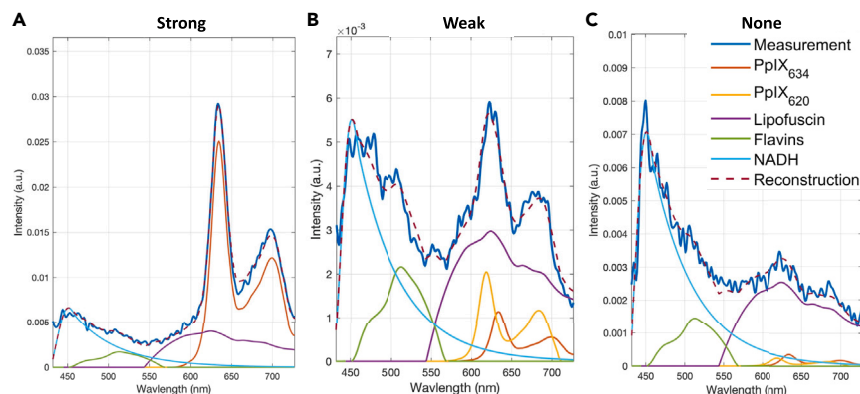
**Figure 2. Example unmixing of three spectra with different PpIX content**

(A–C) Sample unmixing of three spectra with strong (A), weak (B), and very weak (C) PpIX content. The blue line is the measured spectrum, while the purple dashed line is the fit. The other spectra are the endmembers, scaled according to their abundance and summed to create the fitted spectrum. Unmixing enables recovery of PpIX abundance despite autofluorescence.

better, in which the core steps of the process are individually optimized and rooted in the physics of the system. These steps include image acquisition, correction, unmixing, and interpretation of endmember abundances. The surrounding elements of device-specific processing can be kept separate. This separation also enables more flexible use of the results. For example, endmember abundances may be used to identify tumor tissue, classify the tumor type, or provide information about biomarkers such as isocitrate dehydrogenase (IDH) mutation, which is clinically highly relevant.[17]

As described before, classical methods for unmixing have some limitations. Therefore, research has explored deep-learning-based unmixing. Zhang et al. successfully applied CNNs to this task to obtain endmember abundances on four open-source agricultural HSI datasets.[58] No similarly large dataset is available for brain surgery. Wang et al. used CNNs to obtain slightly better performance than non-negative matrix factorization on simulated and real geological HSI data.[59] Others have used fully connected MLPs,[60] CNNs,[61] and auto-associative neural networks[62] to unmix spectra without prior knowledge of the endmembers. However, these are not as effective when the endmember spectra are known, as in our case. An attractive solution called the endmember-guided unmixing network used autoencoders in a Siamese configuration to enforce certain relevant constraints, with good results.[63] A review on-deep learning-based unmixing by Bhatt and Joshi shows that existing work is relatively minimal and preliminary.[64] Much of the research does not use *a priori* known endmember spectra, and to the authors' knowledge, none focuses on attenuation correction, neurosurgery, or HSI for fluorescence imaging.

This paper, therefore, describes a method of deep-learning-based correction and unmixing of HSI data cubes for fluorescence-guided resection of brain tumors. This improves on classical methods, can fit into any HSI pipeline in brain surgery, and gives generalizability and flexibility in the use of the endmember abundances. This is facilitated by the first use, to the authors' knowledge, of modern architectures, including deep autoencoders and residual networks[65] in HSI for brain tumor surgery. It is also the first use of a large and broadly diverse dataset for deep learning in HSI for neurosurgery, including 184 patients and 891 fluorescence HSI data cubes from 12 tumor types, all four World Health Organization grades, with IDH mutant and wild-type samples, and labeled solid tumor, infiltrating zone, and reactive brain ("healthy") tissue. The models are optimized using phantoms and pig brain homogenate (PBH) data with known PpIX concentration. Due to the design's physical underpinning, we show not only better quantitative results on these distributions but also improvements in generalizing to human data.

## RESULTS

### Phantom and PBH results

Figure 3 shows the true and predicted PpIX concentration in PBH data using attenuation correction and hyperspectral unmixing network (ACU-Net) in contrast to the former approach, dual-band attenuation, and partial least-squares (PLS) regression.[66] For PLS, all methods are evaluated using the same cross-validation data splits as described for ACU-Net training. The ACU-Net result has lower variance, indicating that the attenuation correction is effective. Additionally, the unmixed PpIX abundances are linear with the known abundances. Thus, the unmixing is also effective. In fact, the coefficient of determination for the PBH data was 0.97 using ACU-Net, compared to 0.82 with the benchmark method. The R value was similarly strongly improved for phantom data, as shown in Table 1.

This shows that the supervised deep learning method can outperform classical methods. However, the semi-supervised ACU-SA method, too, shows a marked improvement in performance compared to the benchmark, with R values comparable to the supervised model. All the results, R values, and root-mean-square error (RMSE in μg/mL) for phantom and PBH data with the four methods are shown in Table 1.

In addition to PpIX quantification metrics, we have evaluated the runtime of each of the methods to validate whether the developed deep learning approach could be a potential step toward a real-time intraoperative technique. We observed, as shown in Table 2, that the mean runtime per pixel for the ACU-Net is greater than two times faster than the previous benchmark method.
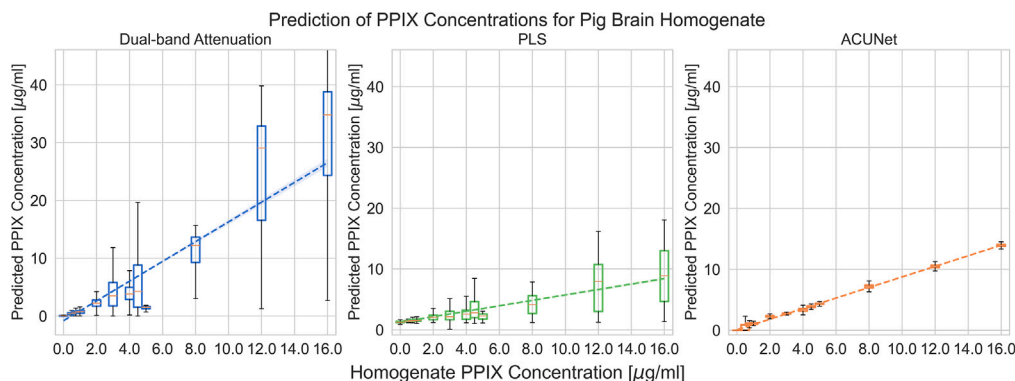
**Figure 3. Linearity and variance of ACU-Net normalization and unmixing compared to previous methods**

Due to heterogeneous scattering and absorption, there is large variation in the measurements for a given known concentration. With ACU-Net, we see greatly improved linearity at a much lower variance and, consequently, a higher coefficient of determination than with the classical method and partial least-squares. These plots used the PBH data. The boxes are the interquartile range, with a line at the median, and the whiskers indicate the minimum and maximum.

### Extension to human data

Though the human data endmember abundances were not known, and no R or RMSE values could be computed, we nevertheless tested the methods on the human data to compare how well they generalize. For these tests, the models were trained on PBH data only, and the mean squared error (MSE) of the spectral reconstruction was measured on the PBH and human data. Good reconstruction does not guarantee accurate underlying endmember abundances, but it does provide some comparison of generalization. As shown in Table 3, ACU-Net and ACU-SA both generalize better to human data than the naive MLP does.

Though the phantom and PBH results are promising, the critical question is whether the same results hold true in human data. While we currently do not have the true endmember abundances and thus cannot assess the performance quantitatively, we can observe that the average MSE reconstruction error of the ACU-Net is comparable to the benchmark method and much better than both the PLS and MLP. Note that the NNLS unmixing minimizes the sum of squared errors, so it is not possible to outperform it in this metric. It shows, however, that the model outputs are reasonable and close to optimal and that it generalizes better to human data than other existing methods.

In addition, the ability of the model to differentiate between healthy and tumor tissue is essential. Producing false-positive PpIX abundance readings, i.e., non-zero computed PpIX abundances where the actual abundance is zero, can be detrimental as they may cause erroneous resection of healthy brain tissue. Therefore, we measured the false-positive rate of the different methods on reactively altered human brain tissue, which should contain little to no PpIX. For these tests, the models were trained on the PBH data and tested on human data. The results in Table 4 show that the ACU-Net architecture outperforms existing methods.

### Qualitative results

Furthermore, differences in the output PpIX concentration maps are observed. In many cases where strong spots of specular reflection caused anomalous results in the dual-band normalization,[16] the ACU-SA can remove the artifacts. This may be because the white light spectra in these cases were sometimes saturated, so the dual-band normalization would not sufficiently compensate, while a deep learning approach can better cope. In addition, previous papers have noted the difficulty of calibrating the unmixing output due to the nonlinear nature of PpIX fluorescence and the presence of more than one fluorescing state with different peak wavelengths.[4,16,27] These factors lead, with the previous method, to unexpectedly large output PpIX concentrations in many cases. However, with the ACU-SA, the values appear far more reasonable, adhering more to expected values with less extreme variation. These factors are illustrated in Figure 4 and suggest that the deep learning approaches may have several benefits over classical methods for processing human data.

**Table 1. Comparison of proposed end-to-end learning-based normalization and unmixing compared to the benchmark dual-band normalization followed by non-negative least squares unmixing**

|  | Dual-band | ACU-Net | ACU-SA | PLS | MLP |
|---|---|---|---|---|---|
| Phantom data | R = 0.93 | R = **0.997** | R = 0.98 | R = 0.93 | R = 0.998 |
|  | RMSE = 3.77 | RMSE = **0.19** | RMSE = 0.51 | RMSE = 0.35 | RMSE = 1.31 |
| Pig brain homogenate | R = 0.82 | R = **0.99** | R = 0.91 | R = 0.67 | R = 0.92 |
|  | RMSE = 4.17 | RMSE = **0.33** | RMSE = 0.81 | RMSE = 2.10 | RMSE = 1.94 |

PLS and MLP approaches are also compared. The coefficient of determination between known and computed PpIX concentration is used for consistency with previous normalization work.[61,67] For human data, no labels are available, so the reconstruction's MSE (ReMSE) is used.

**Table 2. Comparison of runtime for each of the methods (mean $\pm$ standard deviation for a full 21,000-pixel test dataset, and per pixel)**

| Runtime | Benchmark | ACU-Net | PLS |
|---|---|---|---|
| Total test set | 7720 $\pm$ 1,240 ms | 3400 $\pm$ 303 ms | 447 $\pm$ 45.3 ms |
| Mean per pixel | 367.6 $\pm$ 59.0 μs | 161.9 $\pm$ 14.4 μs | 21.3 $\pm$ 2.2 μs |

All differences are significant ($p < 0.05$).

For ACU-Net, although we do not explicitly train to achieve a normalized fluorescence emission spectrum, we observe that the reconstructions do converge to a reasonable spectrum for samples of both phantom and PBH given the same PpIX concentrations. This is shown in Figure 5.

## DISCUSSION

The results show that both supervised and semi-supervised learning outperform classical methods for correcting and unmixing hyperspectral brain tumor data. The performance of the semi-supervised method is promising for the field, as it shows that improved performance may be achieved without labeled datasets. Instead, data such as our human measurements can be used without ground truth abundance values. In this way, such models could be trained with large volumes of data and may generalize well to new human measurements. Additionally, it is shown that the ACU-Net method generalizes better to human data when trained on PBH data than existing classical or learning-based methods and that it achieves lower false-positive rates. Further work is required to continue improving the performance and to show quantitatively that it is effective on human data. This may involve chemical or histopathological assessment of samples co-registered to the HSI measurements, allowing for comparison of known absolute PpIX concentrations. The dataset should also be expanded to include non-tumor tissue to decrease false-positive endmember abundances. Additionally, enhancing the dataset with ground truth labels for the concentrations of both states of PpIX and the other fluorophores would better constrain the outputs of the relative abundance model outputs. Currently, only PpIX$_{634}$ labels are available in phantom and PBH data.

For better generalizability and interpretability, it is best to separate the normalization and unmixing steps or at least have an intermediate state, which is the normalized spectra. Then, for example, the normalization network could be trained on phantom data with concentration labels and then attached to an unmixing network, which was trained unsupervised on human data. In this way, the whole model would generalize better since the unmixing cannot be trained on phantom data, which contain different endmember spectra than human brain, and the normalization is best trained with phantoms of constant, known concentration. This is achieved to a degree in this study but requires further investigation. Although ACU-Net and ACU-SA both achieved similarly high R value, we observed cases where the intermediate predicted normalized spectrum did not resemble a real measured spectrum. This indicates that the domain of the HU function the deep learning models learn is too large. Future work should find methods to constrain the shape of the predicted normalized spectrum more strongly to prevent the ACU-SA architecture from functioning as an end-to-end model and defeating the purpose of having a distinct normalization module. A related challenge is that the normalized spectrum is not known *a priori*. This is why both ACU-Net and ACU-SA rely on either an indirect or latent representation during training, which is not guaranteed to converge to true physical normalized spectrum. If phantoms are not sufficiently homogenous, the assumption that a common normalized spectrum exists is tenuous.

This study on the use of deep learning for analysis of hyperspectral images in fluorescence-guided neurosurgery invites several avenues of future research. These include integrating increasingly sophisticated models emerging from deep learning research, adding further constraints to enhance modeling accuracy, and enriching the available datasets to bolster the effectiveness of models. For example, to enable supervised learning on human data, mass spectrometry could be used to determine ground truth labels. It is also likely that spatial interactions between adjacent pixels in the hyperspectral images, which are currently not accounted for in our models, exist. Notably, relevant studies, including those using deep learning models for unmixing and otherwise analyzing hyperspectral images, have demonstrated improved performance when examining larger image regions instead of individual pixels. Therefore, adopting a model similar to ACU-SA but using a 2D CNN to account for spatial information can enhance correctional capabilities. This would likely improve the spatial smoothness of the abundance overlay plots and better handle localized artifacts such as bright reflections, as shown in Figure 4.

Another promising direction for future research is the integration of product and quotient relations into deep learning models. Previous studies[22,68,69] have successfully utilized scaling factors that multiply or divide the measured fluorescence emission spectra for normalization. However, standard deep neural networks (DNNs) are better at capturing additive and non-linear relationships rather than direct multiplicative or divisive interactions. Incorporating multiplication or division operations directly or explicitly transforming them to log space into the

**Table 3. Reconstruction MSE of the models trained on only PBH**

| | ACU-Net | ACU-SA | MLP |
|---|---|---|---|
| PBH | 5.16e−5 | 5.48e−5 | 5.02e−5 |
| Human data | 5.95e−4 | 2.77e−4 | 7.87e−4 |

The dual-band method is excluded because it uses non-negative least squares so the MSE is minimal and there is no concept of generalization. ACU-SA and ACU-Net both generalize better to human data than a naive MLP.

**Table 4. False-positive rate in human brain tissue—i.e., the percentage of spectra with zero expected PpIX abundance for which the method computed a non-zero value**

| Method | Dual-Band | MLP | ACU-Net |
|---|---|---|---|
| *False-positive rate* | 13.1% | 12.9% | 8.39% |

model's architecture could enable a DNN to represent these simpler analytical relations more efficiently, thus reducing the likelihood of over-fitting and potentially offering a more accurate model of reality. However, it is essential to exercise caution regarding non-differentiability when incorporating these operations, as they can pose challenges in the gradient-based optimization process typically used in training DNNs.

### Limitations of the study

As outlined in the Discussion, there are several limitations of the current study that warrant future research. The primary limitation is that there were no ground truth labels for the human data, so performance in humans had to be assessed indirectly. Furthermore, the data are *ex vivo* and the imaging device is slow, so intraoperative use of the models *in vivo* will require more evaluation and potentially adjustment of the models for use with snapshot hyperspectral devices. The models themselves did not utilize spatial information, which would likely improve performance. Additionally, there were some cases where the intermediate predicted normalized spectrum did not resemble a real measured spectrum, so further improvement of the models would be beneficial.

### Conclusion

This paper has introduced two deep learning architectures that outperform prior methods for attenuation correction and unmixing of hyperspectral images in fluorescence-guided brain tumor surgery. The architectures explicitly enforce adherence to physical models of the system and condition on prior knowledge of the present endmember spectra, thus retaining some of the reliability and explainability of classical methods. Furthermore, the second introduced architecture can be trained in a semi-supervised manner, which allows the use of unlabeled human data and encourages better generalizability. The developed methods greatly improve the efficacy of the spectral correction and subsequent unmixing, decreasing unwanted variance and increasing the linearity of the estimated endmember abundances with respect to the expected abundances. They also decrease false-positive PpIX measurements and generalize better to human data than existing methods. These models will thus enable more accurate classification of brain tumors and tumor margins for intraoperative guidance in future work.

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and data should be directed to and will be fulfilled by the lead contact, Dr. Eric Suero Molina (Eric.Suero@ukmuenster.de).

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

- The human data cannot be shared for privacy reasons, but phantom and PBH data may be shared upon reasonable request to the lead contact.
- The code for the described deep learning models is available on the repository linked in the key resources table.
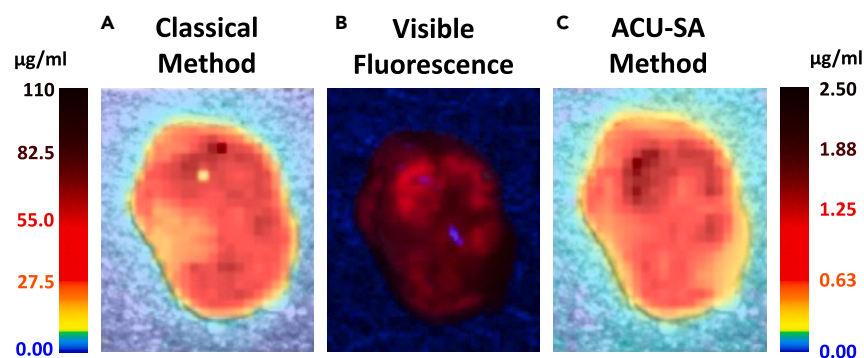


**Figure 4. PpIX concentration map showing qualitative benefits of the proposed method**
PpIX concentration computed across a brain tumor sample using the classical method (A) and the ACU-SA (C). The deep-learning-based method shows a far more reasonable concentration range and better handles bright specular reflections in the top center of the sample. The visible fluorescence (RGB image, B) shows very similar patterns to the unmixing results.
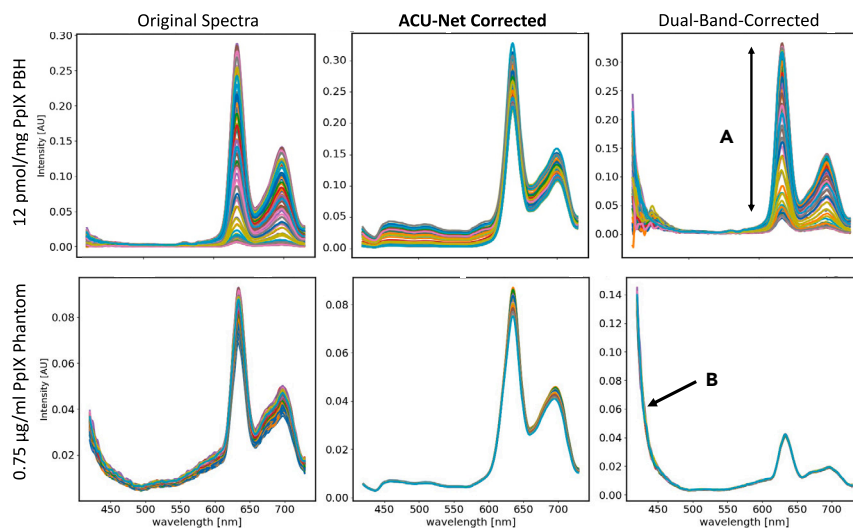
**Figure 5. Corrected fluorescence emission spectra computed for both PpIX phantoms (top) and pig brain homogenate (bottom) using ACU-Net (middle) and the benchmark dual-band attenuation method (right)**

(A and B) The deep-learning-based correction shows lower variance given the same concentration (A) and better corrects for the blue light tail near the excitation wavelengths (B) than the classical dual-band method.

- Further information about the human data is included in Table S1. Any additional information required to reanalyze the data reported in this paper may be made available from the lead contact upon request.

## AUTHOR CONTRIBUTIONS

Conception and design, D.B., A.X., J.G., and E.S.M.; acquisition of data, E.S.M.; statistical analysis and interpretation, D.B., A.X., J.G., B.L., and E.S.M.; drafting the article, D.B., A.X., and J.G.; critically revising the article, all authors; technical support, E.S.M.; study supervision, E.S.M.

## DECLARATION OF INTERESTS

E.S.M. received research support from Carl Zeiss Meditec AG. W.S. has received speaker and consultant fees from SBI ALA Pharma, medac, Carl Zeiss Meditec AG, NXDC, and research support from Zeiss.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - *Ex vivo* animal material
  - Human participants
- METHOD DETAILS
  - Architecture
  - Dataset
  - Experiments
- QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111273.

## REFERENCES

1. Stepp, H., and Stummer, W. (2018). 5-ALA in the management of malignant glioma. Lasers Surg. Med. *50*, 399–419. https://doi.org/10.1002/LSM.22933.

2. Stummer, W., Pichlmeier, U., Meinel, T., Wiestler, O.D., Zanella, F., and Reulen, H.J.; ALA-Glioma Study Group (2006). Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. Lancet Oncol. *7*, 392–401. https://doi.org/10.1016/S1470-2045(06)70665-9.

3. Hadjipanayis, C.G., and Stummer, W. (2019). 5-ALA and FDA approval for glioma surgery. J. Neuro Oncol. *141*, 479–486. https://doi.org/10.1007/S11060-019-03098-Y.

4. Suero Molina, E., Black, D., Walke, A., Azemi, G., D'Alessandro, F., König, S., and Stummer, W. (2023). Unraveling the blue shift in porphyrin fluorescence in glioma: The 620 nm peak and its potential significance in tumor biology. Front. Neurosci. *17*, 1261679. https://doi.org/10.3389/fnins.2023.1261679.

5. Schupper, A.J., Rao, M., Mohammadi, N., Baron, R., Lee, J.Y.K., Acerbi, F., and Hadjipanayis, C.G. (2021). Fluorescence-Guided Surgery: A Review on Timing and Use in Brain Tumor Surgery. Front. Neurol. *12*, 682151. https://doi.org/10.3389/fneur.2021.682151.

6. Kaneko, S., Suero Molina, E., Sporns, P., Schipmann, S., Black, D., and Stummer, W. (2022). Fluorescence real-time kinetics of protoporphyrin IX after 5-ALA administration in low-grade glioma. J. Neurosurg. *136*, 9–15. https://doi.org/10.3171/2020.10.JNS202881.

7. Suero Molina, E., Black, D., Kaneko, S., Müther, M., and Stummer, W. (2022). Double dose of 5-aminolevulinic acid and its effect on protoporphyrin IX accumulation in low-grade glioma. J. Neurosurg. *137*, 943–952. https://doi.org/10.3171/2021.12.JNS211724.

8. Suero Molina, E., Kaneko, S., Black, D., and Stummer, W. (2021). 5-Aminolevulinic acid-induced porphyrin contents in various brain tumors: implications regarding imaging device design and their validation. Neurosurgery *89*, 1132–1140.

9. Ma, J., Sun, D.W., Pu, H., Cheng, J.H., and Wei, Q. (2019). Advanced Techniques for Hyperspectral Imaging in the Food Industry: Principles and Recent Applications. Annu. Rev. Food Sci. Technol. *10*, 197–220. https://doi.org/10.1146/ANNUREV-FOOD-032818-121155/CITE/REFWORKS.

10. Dong, X., Jakobi, M., Wang, S., Köhler, M.H., Zhang, X., and Koch, A.W. (2019). A review of hyperspectral imaging for nanoscale materials research. Appl. Spectrosc. Rev. *54*, 285–305. https://doi.org/10.1080/05704928.2018.1463235.

11. Lu, B., Dao, P., Liu, J., He, Y., and Shang, J. (2020). Recent advances of hyperspectral imaging technology and applications in agriculture. Remote Sens. (Basel) *12*, 2659.

12. Hege, E.K., O'Connell, D., Johnson, W., Basty, S., and Dereniak, E.L. (2004). Hyperspectral imaging for astronomy and space surveillance. Imaging Spectrometry IX *5159*, 380–391. https://doi.org/10.1117/12.506426.

13. Jarry, G., Henry, F., and Kaiser, R. (2000). Anisotropy and multiple scattering in thick mammalian tissues. JOSA A *17*, 149–153.

14. Iordache, M.-D., Bioucas-Dias, J.M., and Plaza, A. (2011). Sparse unmixing of hyperspectral data. IEEE Trans. Geosci. Remote Sens. *49*, 2014–2039.

15. Black, D., Kaneko, S., Walke, A., König, S., Stummer, W., and Suero Molina, E. (2021). Characterization of autofluorescence and quantitative protoporphyrin IX biomarkers for optical spectroscopy-guided glioma surgery. Sci. Rep. *11*, 20009–20012. https://doi.org/10.1038/s41598-021-99228-6.

16. Walke, A., Black, D., Valdes, P.A., Stummer, W., König, S., and Suero-Molina, E. (2023). Challenges in, and recommendations for, hyperspectral imaging in ex vivo malignant glioma biopsy measurements. Sci. Rep. *13*, 3829. https://doi.org/10.1038/S41598-023-30680-2.

17. Black, D., Byrne, D., Walke, A., Liu, S., Di Ieva, A., Kaneko, S., Stummer, W., Salcudean, T., Suero Molina, E., and med Eric Suero Molina, P. (2023). Towards Machine Learning-based Quantitative Hyperspectral Image Guidance for Brain Tumor Resection. Preprint at arXiv. https://doi.org/10.1038/s43856-024-00562-3.

18. Leclerc, P., Ray, C., Mahieu-Williame, L., Alston, L., Frindel, C., Brevet, P.F., Meyronet, D., Guyotat, J., Montcel, B., and Rousseau, D. (2020). Machine learning-based prediction of glioma margin from 5-ALA induced PpIX fluorescence spectroscopy. Sci. Rep. *10*, 1462. https://doi.org/10.1038/s41598-020-58299-7.

19. Martinez, B., Leon, R., Fabelo, H., Ortega, S., Piñeiro, J.F., Szolna, A., Hernandez, M., Espino, C., O'shanahan, A.J., Carrera, D., et al. (2019). Most Relevant Spectral Bands Identification for Brain Cancer Detection Using Hyperspectral Imaging. Sensors *19*, 5481. https://doi.org/10.3390/S19245481.

20. Baig, N., Fabelo, H., Ortega, S., Callico, G.M., Alirezaie, J., and Umapathy, K. (2021). Empirical Mode Decomposition Based Hyperspectral Data Analysis for Brain Tumor Classification. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), pp. 2274–2277. https://doi.org/10.1109/EMBC46164.2021.9629676.

21. Giannantonio, T., Alperovich, A., Semeraro, P., Atzori, M., Zhang, X., Hauger, C., Freytag, A., Luthman, S., Vandebriel, R., Jayapala, M., et al. (2023). Intra-operative brain tumor detection with deep learning-optimized hyperspectral imaging.. Preprint at arXiv, 80–98. https://doi.org/10.1117/12.2646999.

22. Valdés, P.A., Leblond, F., Kim, A., Wilson, B.C., Paulsen, K.D., and Roberts, D.W. (2012). A spectrally constrained dual-band normalization technique for protoporphyrin IX quantification in fluorescence-guided surgery. Opt. Lett. *37*, 1817–1819.

23. Valdés, P.A., Leblond, F., Jacobs, V.L., Wilson, B.C., Paulsen, K.D., and Roberts, D.W. (2012). Quantitative, spectrally-resolved intraoperative fluorescence imaging. Sci. Rep. *2*, 798.

24. Yoon, J., Grigoroiu, A., and Bohndiek, S.E. (2020). A background correction method to compensate illumination variation in hyperspectral imaging. PLoS One *15*, e0229502.

25. Bradley, R.S., and Thorniley, M.S. (2006). A review of attenuation correction techniques for tissue fluorescence. J. R. Soc. Interface *3*, 1–13. https://doi.org/10.1098/RSIF.2005.0066.

26. Yoon, S.-C., and Park, B. (2015). Hyperspectral image processing methods. Hyperspectral Imaging Technology in Food and Agriculture, 81–101.

27. Alston, L., Mahieu-Williame, L., Hebert, M., Kantapareddy, P., Meyronet, D., Rousseau, D., Guyotat, J., and Montcel, B. (2019). Spectral complexity of 5-ALA induced PpIX fluorescence in guided surgery: a clinical study towards the discrimination of healthy tissue and margin boundaries in high and low grade gliomas. Biomed. Opt Express *10*, 2478–2492.

28. Fürtjes, G., Reinecke, D., von Spreckelsen, N., Meißner, A.K., Rueß, D., Timmer, M., Freudiger, C., Ion-Margineanu, A., Khalid, F., Watrinet, K., et al. (2023). Intraoperative microscopic autofluorescence detection and characterization in brain tumors using stimulated Raman histology and two-photon fluorescence. Front. Oncol. *13*, 1146031. https://doi.org/10.3389/FONC.2023.1146031/BIBTEX.

29. Black, D., Liquet, B., Kaneko, S., Di Ieva, A., Stummer, W., and Molina, E.S. (2024). A Spectral Library and Method for Sparse Unmixing of Hyperspectral Images in Fluorescence Guided Resection of Brain Tumors. Biomed. Opt Express *15*, 4406–4424.

30. Geladi, P., and Kowalski, B.R. (1986). Partial Least-Squares Regression - a Tutorial. Anal. Chim. Acta *185*, 1–17. https://doi.org/10.1016/0003-2670(86)80028-9.

31. Bro, R., and DeJong, S. (1997). A fast non-negativity-constrained least squares algorithm. J. Chemom. *11*, 393–401. https://doi.org/10.1002/(Sici)1099-128x(199709/10)11:5<393::Aid-Cem483>3.3.Co;2-C.

32. Wang, R., Lemus, A.A., Henneberry, C.M., Ying, Y., Feng, Y., and Valm, A.M. (2023). Unmixing biological fluorescence image data with sparse and low-rank Poisson regression. Bioinformatics *39*, btad159.

33. Coates, P.B. (1972). Photomultiplier noise statistics. J. Phys. D Appl. Phys. *5*, 915–930.

34. Jia, S., Jiang, S., Lin, Z., Li, N., Xu, M., and Yu, S. (2021). A survey: Deep learning for hyperspectral image classification with few labeled samples. Neurocomputing *448*, 179–204. https://doi.org/10.1016/j.neucom.2021.03.035.

35. Khan, U., Paheding, S., Elkin, C.P., and Devabhaktuni, V.K. (2021). Trends in Deep Learning for Medical Hyperspectral Image Analysis. IEEE Access *9*, 79534–79548. https://doi.org/10.1109/ACCESS.2021.3068392.

36. Cui, R., Yu, H., Xu, T., Xing, X., Cao, X., Yan, K., and Chen, J. (2022). Deep Learning in Medical Hyperspectral Images: A Review. Sensors *22*, 9790. https://doi.org/10.3390/S22249790.

37. Ebner, M., Nabavi, E., Shapey, J., Xie, Y., Liebmann, F., Spirig, J.M., Hoch, A., Farshad, M., Saeed, S.R., Bradford, R., et al. (2021). Intraoperative hyperspectral label-free imaging: from system design to first-in-patient translation. J Phys D Appl Phys *54*, 294003. https://doi.org/10.1088/1361-6463/ABFBF6.

38. Ruiz, L., Martin, A., Urbanos, G., Villanueva, M., Sancho, J., Rosa, G., Villa, M., Chavarrias, M., Perez, A., Juarez, E., et al. (2020). Multiclass Brain Tumor Classification Using Hyperspectral Imaging and Supervised Machine Learning. In 2020 35th Conference on Design of Circuits and Integrated Systems, DCIS 2020. https://doi.org/10.1109/DCIS51330.2020.9268650.

39. Urbanos, G., Martín, A., Vázquez, G., Villanueva, M., Villa, M., Jimenez-Roldan, L., Chavarrías, M., Lagares, A., Juárez, E., and Sanz, C. (2021). Supervised Machine Learning

Methods and Hyperspectral Imaging Techniques Jointly Applied for Brain Cancer Classification. Sensors *21*, 3827. https://doi.org/10.3390/S21113827.

40. Fabelo, H., Halicek, M., Ortega, S., Shahedi, M., Szolna, A., Piñeiro, J.F., Sosa, C., O'Shanahan, A.J., Bisshopp, S., Espino, C., et al. (2019). Deep Learning-Based Framework for In Vivo Identification of Glioblastoma Tumor using Hyperspectral Images of Human Brain. Sensors *19*, 920. https://doi.org/10.3390/S19040920.

41. Fabelo, H., Ortega, S., Ravi, D., Kiran, B.R., Sosa, C., Bulters, D., Callicó, G.M., Bulstrode, H., Szolna, A., Piñeiro, J.F., et al. (2018). Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations. PLoS One *13*, e0193721. https://doi.org/10.1371/JOURNAL.PONE.0193721.

42. Leon, R., Fabelo, H., Ortega, S., Cruz-Guerrero, I.A., Campos-Delgado, D.U., Szolna, A., Piñeiro, J.F., Espino, C., O'Shanahan, A.J., Hernandez, M., et al. (2023). Hyperspectral imaging benchmark based on machine learning for intraoperative brain tumour detection. npj Precis. Oncol. *7*, 119–217. https://doi.org/10.1038/s41698-023-00475-9.

43. Rinesh, S., Maheswari, K., Arthi, B., Sherubha, P., Vijay, A., Sridhar, S., Rajendran, T., and Waji, Y.A. (2022). Investigations on Brain Tumor Classification Using Hybrid Machine Learning Algorithms. J. Healthc. Eng. *2022*, 2761847. https://doi.org/10.1155/2022/2761847.

44. Fabelo, H., Ortega, S., Szolna, A., Bulters, D., Piñeiro, J.F., Kabwama, S., JO'Shanahan, A., Bulstrode, H., Bisshopp, S., Kiran, B.R., et al. (2019). In-vivo hyperspectral human brain image database for brain cancer detection. IEEE Access *7*, 39098–39116.

45. Manni, F., van der Sommen, F., Fabelo, H., Zinger, S., Shan, C., Edström, E., Elmi-Terander, A., Ortega, S., Marrero Callicó, G., and de With, P.H.N. (2020). Hyperspectral Imaging for Glioblastoma Surgery: Improving Tumor Identification Using a Deep Spectral-Spatial Approach. Sensors *20*, 6955. https://doi.org/10.3390/S20236955.

46. Hao, Q., Pei, Y., Zhou, R., Sun, B., Sun, J., Li, S., and Kang, X. (2021). Fusing Multiple Deep Models for in Vivo Human Brain Hyperspectral Image Classification to Identify Glioblastoma Tumor. IEEE Trans. Instrum. Meas. *70*, 1–14. https://doi.org/10.1109/TIM.2021.3117634.

47. Callicó, G.M., Fabelo, H., Camacho, R., de la Luz Plaza, M., Callicó, G.M., and Sarmiento, R. (2018). Detecting brain tumor in pathological slides using hyperspectral imaging. Biomed. Opt Express *9*, 818–831. https://doi.org/10.1364/BOE.9.000818.

48. Puustinen, S., Vrzáková, H., Hyttinen, J., Rauramaa, T., Fält, P., Hauta-Kasari, M., Bednarik, R., Koivisto, T., Rantala, S., von und zu Fraunberg, M., et al. (2023). Hyperspectral Imaging in Brain Tumor Surgery—Evidence of Machine Learning-Based Performance. World Neurosurg. *175*, e614–e635. https://doi.org/10.1016/J.WNEU.2023.03.149.

49. Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (Springer), pp. 234–241.

50. Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). V-net: Fully convolutional networks for volumetric medical image segmentation. In 2016 fourth international conference on 3D vision (3DV) (Ieee), pp. 565–571.

51. Li, Y., Zhang, Y., Cui, W., Lei, B., Kuang, X., and Zhang, T. (2022). Dual encoder-based dynamic-channel graph convolutional network with edge enhancement for retinal vessel segmentation. IEEE Trans. Med. Imaging *41*, 1975–1989.

52. Zhou, Q., Wang, Q., Bao, Y., Kong, L., Jin, X., and Ou, W. (2022). Laednet: A lightweight attention encoder–decoder network for ultrasound medical image segmentation. Comput. Electr. Eng. *99*, 107777.

53. Xun, S., Li, D., Zhu, H., Chen, M., Wang, J., Li, J., Chen, M., Wu, B., Zhang, H., Chai, X., et al. (2022). Generative adversarial networks in medical image segmentation: A review. Comput. Biol. Med. *140*, 105063.

54. Jia, S., Jiang, S., Lin, Z., Li, N., Xu, M., and Yu, S. (2021). A survey: Deep learning for hyperspectral image classification with few labeled samples. Neurocomputing *448*, 179–204. https://doi.org/10.1016/J.NEUCOM.2021.03.035.

55. Gao, K., Liu, B., Yu, X., and Yu, A. (2022). Unsupervised Meta Learning With Multiview Constraints for Hyperspectral Image Small Sample set Classification. IEEE Trans. Image Process. *31*, 3449–3462. https://doi.org/10.1109/TIP.2022.3169689.

56. Li, Y., Wu, R., Tan, Q., Yang, Z., and Huang, H. (2023). Masked Spectral Bands Modeling With Shifted Windows: An Excellent Self-Supervised Learner for Classification of Medical Hyperspectral Images. IEEE Signal Process. Lett. *30*, 543–547. https://doi.org/10.1109/LSP.2023.3273506.

57. Dong, Y., Shi, W., Du, B., Hu, X., and Zhang, L. (2022). Asymmetric Weighted Logistic Metric Learning for Hyperspectral Target Detection. IEEE Trans. Cybern. *52*, 11093–11106. https://doi.org/10.1109/TCYB.2021.3070909.

58. Zhang, X., Lin, T., Xu, J., Luo, X., and Ying, Y. (2019). DeepSpectra: An end-to-end deep learning approach for quantitative spectral analysis. Anal. Chim. Acta *1058*, 48–57. https://doi.org/10.1016/J.ACA.2019.01.002.

59. Wang, C.J., Li, H., and Tang, Y.Y. (2019). Hyperspectral Unmixing Using Deep Learning. In International Conference on Wavelet Analysis and Pattern Recognition (IEEE), pp. 1–7. https://doi.org/10.1109/ICWAPR48189.2019.8946465.

60. Deshpande, V.S., and Bhatt, J.S. (2021). A practical approach for hyperspectral unmixing using deep learning. Geosci. Rem. Sens. Lett. IEEE *19*, 1–5.

61. Zhang, X., Sun, Y., Zhang, J., Wu, P., and Jiao, L. (2018). Hyperspectral unmixing via deep convolutional neural networks. IEEE Geosci. Remote Sens. Lett. *15*, 1755–1759.

62. Licciardi, G.A., and Del Frate, F. (2011). Pixel unmixing in hyperspectral data by means of neural networks. IEEE Trans. Geosci. Remote Sens. *49*, 4163–4172. https://doi.org/10.1109/TGRS.2011.2160950.

63. Hong, D., Gao, L., Yao, J., Yokoya, N., Chanussot, J., Heiden, U., and Zhang, B. (2022). Endmember-guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing. IEEE Trans. Neural Netw. Learn. Syst. *33*, 6518–6531.

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

65. He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016-December (IEEE), pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

66. Geladi, P., and Kowalski, B.R. (1986). Partial least-squares regression: a tutorial. Anal. Chim. Acta *185*, 1–17. https://doi.org/10.1016/0003-2670(86)80028-9.

67. Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., and Girschick, R. (2019). Detectron2. https://github.com/facebookresearch/detectron2.

68. Hull, E., Ediger, M., Unione, A., Deemer, E., Stroman, M., and Baynes, J. (2004). Noninvasive, optical detection of diabetes: model studies with porcine skin. Opt Express *12*, 4496–4510. https://doi.org/10.1364/OPEX.12.004496.

69. Finlay, J.C., Conover, D.L., Hull, E.L., and Foster, T.H. (2001). Porphyrin Bleaching and PDT-induced Spectral Changes are Irradiance Dependent in ALA-sensitized Normal Rat Skin In Vivo. Photochem. Photobiol. *73*, 54–63. https://doi.org/10.1562/0031-8655(2001)0730054PBAPIS2.0.CO2.

70. He, K.M., Zhang, X.Y., Ren, S.Q., and Sun, J. (2016). Deep Residual Learning for Image Recognition. In 2016 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr), pp. 770–778. https://doi.org/10.1109/Cvpr.2016.90.

71. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. In 2016 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr), pp. 2818–2826. https://doi.org/10.1109/Cvpr.2016.308.

72. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature *521*, 436–444. https://doi.org/10.1038/nature14539.

73. Kendall, A., Gal, Y., and Cipolla, R. (2018). Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

74. Palsson, B., Sigurdsson, J., Sveinsson, J.R., and Ulfarsson, M.O. (2018). Hyperspectral Unmixing Using a Neural Network Autoencoder. IEEE Access *6*, 25646–25656. https://doi.org/10.1109/ACCESS.2018.2818280.

75. Anichini, G., Leiloglou, M., Hu, Z., O'Neill, K., and Elson, D. (2024). Hyperspectral and multispectral imaging in neurosurgery: a systematic literature review and meta-analysis. Eur. J. Surg. Oncol. *108293*, 108293. https://doi.org/10.1016/J.EJSO.2024.108293.

# STAR★METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| **Biological samples** | | |
| | Samples removed from patients undergoing brain tumor surgery with fluorescence guidance at the University Hospital Muenster | |
| **Software and algorithms** | | |
| | https://github.com/dgblack/acunet_glioma | |

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Since the presented models require a mix of labeled and unlabeled data, three datasets were used in this paper: (1) brain tissue phantoms were created using known concentrations of PpIX, (2) PBH was spiked with known concentrations of PpIX, and (3) human brain tumor tissue was extracted during surgery and imaged *ex vivo*. All samples were measured on the same HSI device at the University Hospital of Münster, described below.

For phantoms, PpIX was mixed with Intralipid 20% (Fresenius Kabi GmbH, Bad Homburg, Germany) and red dye (McCormick, Baltimore, USA) in dimethyl sulfoxide (DMSO; Merck KGaA, Darmstadt, Germany) solvent to simulate the scattering and absorption, respectively, in human tissue, as described by Valdes et al..[22,23] The PpIX concentrations were (0.0, 0.2, 0.6, 1.25, 2.5 μg/mL). By varying the other components, the following optical properties were achieved: absorption at 405 nm: $\mu_{a,\,405\,nm} = 18, 42, 60\ cm^{-1}$; reduced scattering at 635 nm: $\mu'_{s,\,635\,nm} = 8.7, 11.6, 14.5\ cm^{-1}$. More details are found in previous work.[16]

### Ex vivo animal material

For the PBH, pig brain was obtained from a local butcher and separated into anatomical sections of cerebrum, cerebellum, hypothalamus, and brain stem/spinal cord. The tissue was washed with distilled water, cut into 10 × 10 × 10 mm pieces, and homogenized using a blender (VDI 12, VWR International, Hannover, Germany). The pH was controlled using 0.5 M tris(hydroxymethyl)aminomethane (Tris-base, Serva, Heidelberg, Germany) buffer and hydrochloric acid (HCl, Honeywell Riedel–de Haen, Seelze, Germany). For each sample, 200 to 600 mg of the homogenates were spiked with PpIX (Enzo Life Sciences GmbH, Lörrach, Germany) stock solution (300 pmol/μL in DMSO) to the desired concentrations (0.0, 0.5, 0.75, 1.0, 2.0, 3.0 and 4.0 pmol/mg) and homogenized using a vortex mixer. The PBH samples were placed in a Petri dish, making samples of about 4 × 4 × 2 mm. Approval for experiments with pig brains was given by the Health and Veterinary Office Münster (Reg.-No. 05 515 1052 21). More details about the PBH are available in previous work.[16]

### Human participants

The human data used in this study was measured over six years (2018–2023) at the University Hospital Münster, Münster, Germany. Patients undergoing surgery for various brain tumors were given a standard dose of 20 mg/kg of 5-ALA (Gliolan, medac, Wedel, Germany) orally 4 h before induction of anesthesia. All procedures performed in these studies followed the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. All experiments and clinical data analysis were approved by the local Ethics Committee (2015-632-f-S and 2020-644-f-S), and informed consent was obtained from all patients.

Tissue resected by the surgeons was immediately taken to the hyperspectral imaging (HSI) device and imaged *ex vivo* before being given to pathology. Each tissue sample measurement produced one data cube which on average contained approximately 623 spectra. In total, data cubes were measured for 891 biopsies from 184 patients, resulting in 555666 human brain tumor spectra. The tumor types are shown below.

| Category | # of Data Cubes | # of Patients |
|---|---|---|
| *Tissue Type* | *632* | *130* |
| Pilocytic Astrocytoma | 5 | 2 |
| Diffuse Astrocytoma | 60 | 17 |

*(Continued on next page)*

**Continued**

| Category | # of Data Cubes | # of Patients |
|---|---|---|
| Anaplastic Astrocytoma | 51 | 10 |
| Glioblastoma | 415 | 77 |
| Grade II Oligodendroglioma | 24 | 5 |
| Ganglioglioma | 4 | 2 |
| Medulloblastoma | 6 | 2 |
| Anaplastic Ependymoma | 8 | 2 |
| Anaplastic Oligodendroglioma | 4 | 1 |
| Meningioma | 37 | 8 |
| Metastasis | 6 | 2 |
| Radiation Necrosis | 20 | 4 |
| *Margins (Gliomas)* | *288* | *67* |
| Reactively altered brain tissue | 100 | 22 |
| Infiltrating zone | 57 | 18 |
| Solid tumor | 131 | 27 |
| *WHO Grade (Gliomas)* | *571* | *119* |
| Grade I | 9 | 3 |
| Grade II | 84 | 20 |
| Grade III | 57 | 15 |
| Grade IV | 421 | 81 |
| *IDH Classification* | *411* | *76* |
| Mutant | 126 | 26 |
| Wildtype | 285 | 50 |

Of the 184 patients, 56.7% identified as male and 43.3% female. The ages ranged from 1 to 82, with a mean of 51.6 and median of 55. No meaningful difference in any of the endmember abundances was found as a function of age or sex.

## METHOD DETAILS

### Architecture

Two neural network (NN) architectures were developed and tested in Python: a supervised model called ACU-Net and a semi-supervised autoencoder model called ACU-SA, inspired by EGU-Net.[63] Given the data's characteristics, we employ a 1D deep Convolutional Neural Network (1D CNN) architecture since neighboring wavelengths of the fluorescence and white-light spectra exhibit more correlation than those farther apart. This spatial and spectral correlation aligns well with the inductive bias inherent in CNNs. Additionally, we leverage residual connections, which allow for bypasses of certain layers[70] and have been demonstrated to be a robust heuristic choice that improves the quality of learned features.[71]

For both models, the input data is $X \in \mathbb{R}^{m \times n \times 2}$ where $n$ is the number of spectra and $m$ is the number of wavelength samples in each spectrum. We use the fluorescence emission spectrum $\Phi_{Fluo} \in \mathbb{R}^m$, which is captured while exciting the region with light at $\lambda = 405$ nm, and the white light reflectance spectrum $\Phi_{Ref} \in \mathbb{R}^m$, which is captured while illuminating the region with broadband white light as explained in the Introduction. The two spectra are stacked to form a two-channel input spectrum, which utilizes the locality bias of the CNN. Let $K$ be the number of known endmember spectra. The matrix whose columns are the endmember spectra is $B = \begin{bmatrix} \Phi_{spec,1} & \cdots & \Phi_{spec,k} \end{bmatrix} \in \mathbb{R}^{m \times K}$.

The HSI attenuation correction aims to correct the fluorescence emission spectra so that those originating from samples with equal fluorophore concentration have equal magnitudes irrespective of local optical or geometric properties. In other words, the goal becomes to minimize the variation between spectra of equal fluorophore content. Suppose there is an ideal corrected spectrum, $\Phi_c$, which is the pure emission of the fluorophores with all effects corrected for. Then, the correction seeks to minimize the variance between the predicted fluorescence spectra $\widehat{\Phi}_{fluo,i}$, and $\Phi_c, i \in [1,n]$. Thus, we use the mean squared error (MSE $= \frac{1}{n}\sum_{i=1}^{n} \left\| \widehat{\Phi}_{fluo,i} - \Phi_c \right\|^2$) for the proposed models when predicting the true fluorescence emission spectra. Note $MSE(x) = Bias(x)^2 + Var(x)$[72], so minimizing this objective function does indeed minimize the variance between the predicted and the true normalized spectra. For models in which the abundances are output rather than reconstructed spectra, i.e., the output is the $c_k$ from Equation 1 rather than the $\Phi_{fluo}$, MSE is also used.

## ACU-Net

The Attenuation Correction and Unmixing Network (ACU-Net) is a 1D CNN with four residual blocks, each containing 2–3 same-convolutions, each followed by a small max pooling layer to reduce the dimensions of the feature maps. Between each residual block, there is a convolution layer that approximately doubles the number of feature channels. A kernel size of 5 is used in the early layers and 3 in the later ones. The output of the convolutional layers is inputted to three fully connected layers. The architecture is shown in Figure 6. The white light and fluorescence emission spectra are stacked, so convolutions are performed together, as described above.
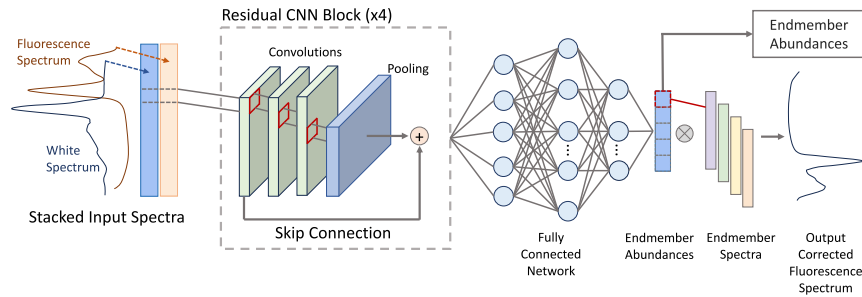


**Spectrally informed attenuation correction and hyperspectral unmixing network architecture**
The inputs are the fluorescence and white-light spectra (orange and black, respectively, on the far left).

The goal of the ACU-Net is to learn the mapping $f : \mathbb{R}^{m \times 2} \to \mathbb{R}^{K}$ from the raw measured spectrum to the absolute endmember abundance vector, $z$, which includes $PpIX_{620}$, $PpIX_{634}$, and three primary autofluorescence sources: lipofuscin, NADH, and flavins.[15] Other autofluorescence may be present,[28] but these 5 spectra have been shown to fit well.[15] This mapping is shown in Equation 2. The ground truth absolute PpIX concentration, denoted $c_{PPIX}$, is known for the phantoms and is known on average for the PBH, as described by Walke et al.[16]

$$z = f(\Phi_{Fluo}, \Phi_{Ref}), z \in \mathbb{R}^{K} \qquad \text{(Equation 2)}$$

Ground truth abundances are not, however, known for human data. Thus, a second loss - the reconstruction loss - is also considered with the aim of better generalization to human data. Let the relative abundance vector be $\widehat{z} = \frac{z}{\|z\|_2}$. We define the normalized reconstructed spectrum $\widehat{\Phi}_{Fluo} = \sum_{k=1}^{K} \widehat{z}_k \Phi_{spec,k} = B\widehat{z}$, which should be as close as possible to the true corrected spectrum, $\Phi_c$ described above. The normalization is important to avoid bias toward strong PpIX spectra which have much larger magnitude than weak ones. The mapping $f$ thus also aims to minimize $MSE(\widehat{\Phi}_{fluo,i} - \Phi_c)$. There is no known ground truth spectrum $\Phi_c$. Instead, ACU-Net uses the non-corrected $\Phi_{Fluo}$, hypothesizing that by training on a large and diverse dataset, $\widehat{\Phi}_{fluo}$ will converge toward an average representation that best characterizes $\Phi_c$. Additionally, it is essential to note that the learned $\widehat{\Phi}_{fluo,i}$ will not fit as precisely as methods employing least squares (LS), which are mathematically optimal and tend to overfit. Instead, by utilizing a deep neural network (DNN), we aim to more effectively learn the corrected fluorescence spectrum $\Phi_C$, and the abundances underlying the noisy measurement.

Using a rectified linear unit (ReLU) activation function at the output of the final layer enforces the non-negativity constraint on the relative abundance values. Finally, we use a weighted loss to train the model to minimize both the error in predicted concentration and the reconstruction error. Since the two objectives are of different scales and it is unknown how the structure of our architecture may affect the learning, the loss weights are also parameterized by considering the homoscedastic uncertainty of each task as outlined by Kendal et al.[73] Denoting $\sigma_C$ and $\sigma_{rec}$ as the learned parameters for weighing the concentration prediction and spectrum reconstruction components of the architecture, we write the total loss function for one measured spectrum in Equation 3.

$$L = \frac{1}{2\sigma_C^2}(z_1 - c_{PPIX})^2 + \frac{1}{2\sigma_{rec}^2}\|Bz - \Phi_{fluo}\|_2^2 + \log(\sigma_C \sigma_{rec})$$

$$z = f(\Phi_{Fluo}, \Phi_{Ref}) \qquad \text{(Equation 3)}$$

## ACU-SA

The challenge with ACU-Net is that it requires ground-truth abundance labels, which are only available for phantom data. Therefore, we also propose a semi-supervised model. Attenuation Correction and Unmixing by a Spectrally-informed Autoencoder (ACU-SA) is similar to the EGU-Net,[63] using an endmember-guided semi-supervised approach to the unmixing process. ACU-SA consists of two main components: one for hyperspectral unmixing (HU) and one explicitly for normalization. The HU portion consists of a Siamese autoencoder architecture, as shown in Figure 7, outlined in green. The objective of this portion is to learn a mapping : $\mathbb{R}^m \to \mathbb{R}^K$, from the normalized fluorescence spectrum to the absolute endmember abundances, like ACU-Net. However, unlike ACU-Net, this portion takes the attenuation-corrected fluorescence emission spectrum as an input rather than the stacked raw spectra and, through its autoencoder structure, unmixes and reconstructs it.

The HU component has the same architecture as the ACU-Net. Then, ACU-SA also includes a standalone CNN normalization model (blue outline in Figure 7) whose objective is to learn the mapping $g : \mathbb{R}^{m \times 2} \to \mathbb{R}^{m \times 1}$, from the two captured spectra to an intermediate representation, which we train to be the normalized/corrected fluorescence spectrum. Together, the normalization model takes the stacked white and fluorescence spectra, performs the attenuation correction, and feeds into the HU autoencoder network, which unmixes it into the absolute endmember abundances. Our normalization model is a shallow 1D-CNN with four convolutional layers and no residual blocks.
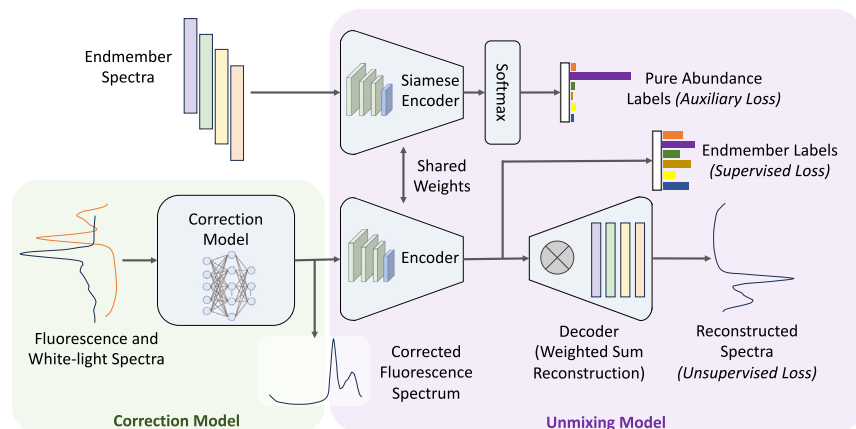


**Endmember-guided normalization (blue outline) and unmixing (green outline) network for semi-supervised learning through an autoencoder architecture**

The endmember embeddings can be used for a supervised loss, while the autoencoder reconstruction is used for semi-supervised training. A second encoder with identical parameters is used with the pure endmember spectra as input, to condition the network on the known endmember spectra.

For supervised learning, the output embeddings from the encoder can be compared to known abundances using the MSE. Otherwise, the decoder reconstructs the spectrum from the abundance values so it can be compared to the input spectrum to obtain an unsupervised reconstruction loss. As with ACU-Net, the decoder uses the output embeddings as weights in the linear combination from Equation 1. Thus, the decoder has fixed parameter weights to ensure the encoders embeddings represent the real endmember abundances.

A twin encoder with shared weights to the HU encoder is used with a SoftMax output and evaluated with a cross-entropy loss. The pure endmember spectra are input to this network, and the output should ideally be a one-hot vector. For example, if the second endmember is input, the unmixing should output zero for all the endmembers except the second, which should be one. In this way, the independence of the endmember spectra is enforced, and we ensure that the output embeddings each correspond to only one endmember. This conditions the network on our *a priori* knowledge of the endmember spectra and has been shown to be effective in deep neural networks for HU.[63]

ACU-SA is trained in two stages. First, the HU network is trained to learn $f$ for the PBH and homogenate datasets. We use a small NN as opposed to other linear and nonlinear HU methods such as least squares and non-negative matrix factorization because it is fully differentiable and easily be incorporated with the other components in ACU-SA. There is also evidence that DNN autoencoders are more robust to environmental noise for HU.[63,74] Since this stage is fully self-supervised, we can augment the training data with synthetic data composed by creating random linear combinations of the known endmember spectra plus noise to help the HU module learn unmixing more effectively, and we can use unlabeled human data. The loss function used for training the HU is given in Equation 4, where $\hat{e}_k \in \mathbb{R}^K$ is all zeros with a 1 in the $k^{th}$ element.

$$L_{HU} = \frac{1}{2K\sigma_{EG}^2} \sum_{k=1}^{K} CE\big(f(\Phi_{spec,i})\big) + \frac{1}{2\sigma_{rec}^2}\|Bz - \widehat{\Phi}_{fluo}\|^2 + \log(\sigma_{EG}\sigma_{rec})$$

$$z = f(\widehat{\Phi}_{fluo})CE(\phi)$$

$$= \log \frac{e^{\phi^T \hat{e}_k}}{\sum_{j \in \{1,\dots,K\}} e^{\phi^T \hat{e}_j}} \qquad \text{(Equation 4)}$$

Here $\sigma_{EG}$ and $\sigma_{rec}$ are again learned loss weightings as used in ACU-Net. For the second stage of training, the weights of the HU module are frozen, and the normalization module is attached. Then, given a much smaller amount of data labeled with their PpIX concentrations, the full network can be trained, optimizing only the weights for the normalization module. The loss function for this stage is shown in Equation 5, where $[x]_i$ represents the $i^{th}$ element of vector $x$.

$$L = \left( [f(g(\Phi_{fluo}, \Phi_{ref}))]_1 - c_{PpIX} \right)^2 \qquad \text{(Equation 5)}$$

The models are physics-informed because they take advantage of the spatial and spectral correlation in the measurements and are optimized with respect to abundances of known fluorescence emission spectra of the predominant fluorophores in brain tissue. Additionally, compared to other works using DNN models which directly perform semantic segmentation of tissue, our model outputs a prediction for a definite and physical quantity. Furthermore, our approach splits correction and unmixing into two modules which can be trained or modified individually, and conditions the unmixing autoencoder to correspond directly with the known endmembers by utilizing a Siamese network and decoding through an explicit weighted sum of the endmember spectra.
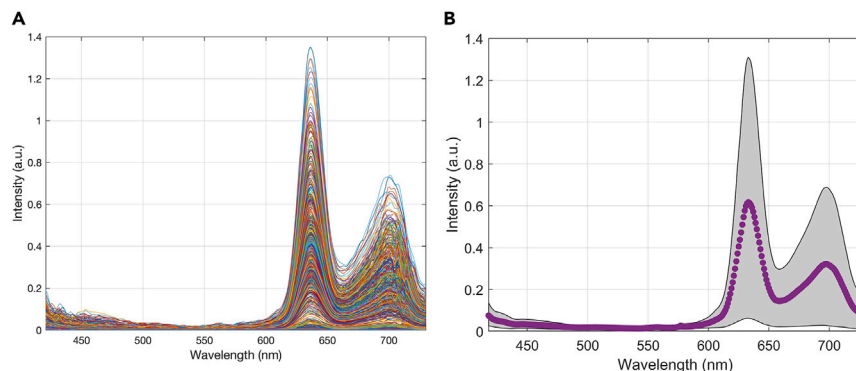
### Dataset

Samples of each type (human, PBH, phantom) are shown in Figure 8. All the samples from these sources were imaged using an HSI device previously described, and some were used in prior research into 5-ALA dosage and timing, tissue type classification, and optimization-based unmixing.[6–8,15–17] The sample was illuminated with white light to capture the white light spectra, blue light from a 405 nm LED for the fluorescence spectra, and not at all for dark spectra, which were used to remove the dark noise of the camera sensor. The reflected and emitted light was captured with a ZEISS Opmi Pico microscope (Carl Zeiss Meditec AG, Oberkochen, Germany) and passed through several low and high-pass filters to remove, for example, the brightly reflected blue excitation light. The light then passed through a liquid crystal tunable filter (Meadowlark Optics, Longmont, CO, USA) to a scientific metal oxide semiconductor (sCMOS) camera (PCO.Edge, Excelitas Technologies, Waltham, MA, USA). Data cubes were captured by sweeping the filter through the visible range from 421 to 730 nm in 3 nm steps and capturing a 2048 x 2048-pixel grayscale image at every sampling wavelength. Each image had a 500 ms exposure time to ensure good signal-to-noise ratio even from faint fluorescence. Additionally, 10 x 10 regions of pixels were averaged to reduce noise. The microscope focus was such that each region was 210 × 210 μm in size.
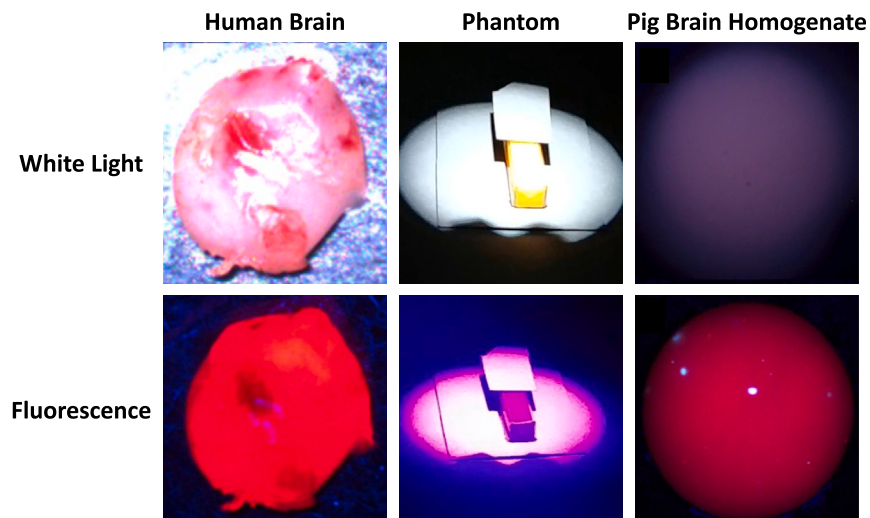


**RGB images of typical samples of human brain, phantom, and pig brain homogenate, each under white light and blue light (fluorescence) illumination**
The PBH images are used under CC BY 4.0 license from ref. [17]

Once captured, each data cube contained the sample of interest surrounded by background of the slide. Extracting the spectra from only the sample by manual segmentation is tedious, so classical computer vision techniques of edge and blob detection and morphological opening were used to detect the sample automatically. This was later augmented using a Detectron 2 model trained on our images.[67] Within these selected areas, regions of 10 x 10 pixels were averaged to increase the signal-to-noise ratio, and as many non-overlapping regions as possible were extracted from the biopsy to ensure independent data samples. The spectra were then corrected for the filter transmission curves and wavelength-dependent sensitivity of the camera. Approximately 500–1000 spectra were measured from each biopsy.

In total, data cubes were measured for 891 biopsies from 184 patients, resulting in 555666 human brain tumor spectra. The human data is shown in Figure 9. The phantom data consisted of 9277 spectra, and the PBH samples were large and constituted 198816 spectra.

**1,000 typical human fluorescence spectra were randomly sampled from the dataset of 555,666 total spectra**

These show clear PpIX content and vary widely in magnitude. (A) shows the spectra themselves while (B) shows the mean spectrum's sample points with the range of one standard deviation in gray.

### Experiments

Various tests were performed to determine the performance of the models on the dataset. As a baseline method or benchmark, the classical attenuation correction and unmixing procedures described in the Introduction were used, including dual-band normalization and nonnegative least squares unmixing. This is currently the most commonly used method in the field. Partial least-squares (PLS) regression[66] and a multi-layer perceptron (MLP) model were also used for comparison. PLS is representative of other PCA-based methods commonly employed in many recent studies.[18,75] The naive deep learning approach was an MLP with an input layer size of 610; the input fluorescence and white light spectra were horizontally stacked. The hidden layer sizes were 8, 5, and 8, with an output size of 310. The model was trained the same way as ACU-Net: the output at the size-5 hidden layer was optimized to the concentration labels, while the 310-dimensional final output was used to compute reconstruction loss. All other optimizer parameters, such as the batch size, learning rate scheduler, etc., were the same as for the ACU-Net.

To evaluate the performance, we used not only the MSE of the reconstructed spectra or calculated abundance vectors, but primarily the correlation coefficient (R) between the measured and ground truth PpIX concentrations. This should ideally be linear, so an R as close to 1 as possible is desired. In this way, the method can be calibrated with a single scaling factor. We also evaluated the runtime of each method since speed is important for real-time intraoperative imaging. These tests were carried out on a laptop with an Nvidia GeForce GTX 1050 GPU and an Intel Core i7-8550U CPU by running a Python timeit library test for a test set of 20 hyperspectral pixels and averaging over 12 runs.

To test the fully supervised ACU-Net, it was necessary to use the phantoms and PBH data, which had ground truth labels. It was possible to train the ACU-SA on human data. However, assessing its performance was difficult without known abundances, and thus comparing methods was impossible. Therefore, the ACU-SA was also trained on the phantom and PBH data for quantitative evaluation before using the human data for a qualitative analysis. For training the ACU-SA, each dataset was split approximately 85% and 15% into training and testing sets, respectively. The split was performed by sample of pig brain or vial of phantom rather than by pixel to avoid bias. All results presented below are on test data unseen during training. The models were trained using the AdamW optimizer with an adaptive learning rate that decreases on training loss plateau. No hyperparameter tuning was done.

### QUANTIFICATION AND STATISTICAL ANALYSIS

For statistical significance tests of the difference between two distributions, we used the two-sample Kolmogorov-Smirnov test (kstest2.m in MATLAB) because the data is continuous, and the test does not make assumptions about its distribution. A $p$ value of less than 0.05 was considered statistically significant. Where a value is given in the paper as $x \pm y$, x is the mean and y is the standard deviation. For all implementation and testing of the deep learning models, Python was used with various packages including NumPy, SciPy, Scikit-image, PyTorch, and TimeIt.