scientific reports

OPEN

Learned spectral decoloring enables photoacoustic oximetry

Janek Gröhl[1,2✉], Thomas Kirchner[3], Tim J. Adler[1,4], Lina Hacker[7], Niklas Holzwarth[1,5], Adrián Hernández-Aguilera[6], Mildred A. Herrera[6], Edgar Santos[6], Sarah E. Bohndiek[7,8] & Lena Maier-Hein[1,2✉]

The ability of photoacoustic imaging to measure functional tissue properties, such as blood oxygenation sO_2, enables a wide variety of possible applications. sO_2 can be computed from the ratio of oxyhemoglobin HbO_2 and deoxyhemoglobin Hb, which can be distuinguished by multispectral photoacoustic imaging due to their distinct wavelength-dependent absorption. However, current methods for estimating sO_2 yield inaccurate results in realistic settings, due to the unknown and wavelength-dependent influence of the light fluence on the signal. In this work, we propose *learned spectral decoloring* to enable blood oxygenation measurements to be inferred from multispectral photoacoustic imaging. The method computes sO_2 pixel-wise, directly from initial pressure spectra $S_{p_0}(\lambda, x)$, which represent initial pressure values at a fixed spatial location x over all recorded wavelengths λ. The method is compared to linear unmixing approaches, as well as pO_2 and blood gas analysis reference measurements. Experimental results suggest that the proposed method is able to obtain sO_2 estimates from multispectral photoacoustic measurements in silico, in vitro, and in vivo.

Tissue blood oxygen saturation sO_2 is an indicator of the health status of a patient[1] and can be used for numerous intra-operative applications. Furthermore, characteristic changes in local sO_2 are associated with some of the hallmarks of cancer[2]. State-of-the-art methods to obtain this value are limited as they are either invasive (e.g. arterial blood gas analysis[3]), lack practicability and accuracy (e.g. blood oxygen level-dependent magnetic resonance imaging[4] or functional near infrared spectroscopy[5]), or only yield a rough global estimate from peripheral vasculature (e.g. pulse-oximetry[6]). None of these techniques yield real-time, spatially-resolved sO_2 estimates. Photoacoustic imaging (PAI) promises to mitigate many of these disadvantages, because it is non-invasive and provides real-time measurements of spatially-resolved sO_2.

Many methods that aim to achieve sO_2 estimation using PAI have been proposed to date[7], including model-based inversion techniques[8–12] and data-driven approaches[13–19]. Due to limitations in terms of applicability, repeatability, or ease-of-use, none of the advanced methods are routinely used in the field of multispectral PAI. Instead, the most commonly applied technique for estimating sO_2 is linear unmixing (LU)[20,21], which assumes a linear combination of relevant chromophores (in the instance of sO_2: oxyhemoglobin HbO_2 and deoxyhemoglobin Hb) to the signal. The core assumption for LU algorithms is that the signal intensities of the reconstructed PA image S, which is an approximation of the underlying initial pressure distribution p_0, are only proportional to the optical absorption coefficients μ_a of the chromophore distribution $S(\lambda) \approx p_0(\lambda) \propto \mu_a(\lambda)$. This assumption does not hold in practise, because the reconstructed image is also proportional to both the Grüneisen parameter Γ, and the light fluence ϕ: $S(\lambda) \approx p_0(\lambda) \propto \mu_a(\lambda) \cdot \Gamma \cdot \phi(\lambda)$.

While Γ is not assumed to be wavelength-dependent, the fluence $\phi(\lambda)$ is dependent on the optical tissue properties and as such the wavelength λ. Due to this interdependency, $\phi(\lambda)$ has a non-trivial and non-linear influence on the recorded multispectral signal $S(\lambda)$. As the optical absorption and scattering coefficients change with wavelength, so does the fluence, which leads to changes in PA signal. This effect is generally referred to as *spectral corruption*[8] or *spectral coloring*[22]. Even small absorption coefficients in the background medium can lead to coloring effects, depending on the depth in the medium. Methods focusing on quantitative photoacoustic imaging have been proposed to solve the ill-posed inverse problem of estimating the optical absorption properties of tissue from the initial pressure distribution[13,16,22–24]. However, while showing great theoretical promise,

[1]Computer Assisted Medical Interventions, German Cancer Research Center, Heidelberg, Germany. [2]Medical Faculty, Heidelberg University, Heidelberg, Germany. [3]Institute of Applied Physics, Biomedical Photonics, Bern University, Bern, Switzerland. [4]Faculty of Mathematics and Computer Science, Heidelberg University, Heidelberg, Germany. [5]Faculty of Physics and Astronomy, Heidelberg University, Heidelberg, Germany. [6]Department of Neurosurgery, Heidelberg University Hospital, Heidelberg, Germany. [7]Department of Physics, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, UK. [8]Cancer Research UK Cambridge Institute, University of Cambridge, Robinson Way, Cambridge CB2 0RE, UK. ✉email: j.groehl@dkfz.de; l.maier-hein@dkfz.de
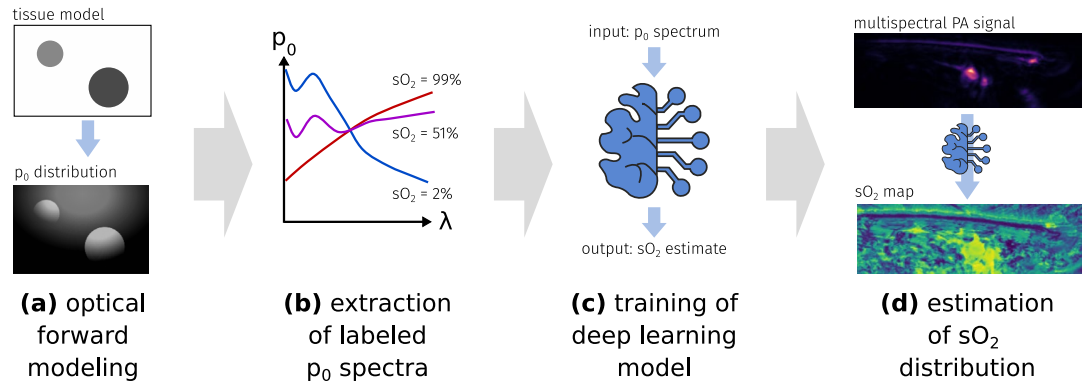
Figure 1. Overview of the methodology: first, numerous p_0 spectra are extracted from optical forward simulations (**a**). These pixel-wise spectra are retrieved from the multi-wavelength simulation by evaluating the p_0 intensity at a fixed pixel location as a function of wavelength (**b**). The data is then used to train a deep learning algorithm (**c**), which afterwards is able to estimate sO_2 values on data that comprises the same wavelengths (**d**).

these methods have not been proven to work for in vivo measurements of complex media. One reason for this is that there is a *domain gap* between simulated and real photoacoustic data. Therefore, accurately calculating blood oxygenation values from multispectral photoacoustic signals in a clinical context remains challenging[25].

Our work introduces a data-driven approach for tackling the problem of sO_2 estimation by introducing the method of *learned spectral decoloring* (LSD). LSD is based on the assumption that the domain gap can be addressed by performing training on initial pressure data by using normalized pixel-wise p_0 spectra. If this assumption holds, it should be possible to train the algorithm using only data from PA optical forward simulations. Using a digital twin of the PA device and a synthetic representation of the target tissue, the LSD algorithm learns how spectral coloring can affect the p_0 spectra at different spatial locations in the tissue. In several experiments, we train the algorithm on in silico data and apply it to various in silico, in vitro, and in vivo data sets. According to the results, sO_2 estimates from spectrally colored data are generally feasible with LSD, which shows promising advantages compared to LU. For example, LSD consistently exhibits a higher dynamic range than sO_2 estimates obtained using LU techniques and enables orders of magnitude faster sO_2 estimation.

Methods

This section will first introduce the principle of the *learned spectral decoloring* (LSD) method as depicted in Fig. 1, then will present the data sets, and finally will specify the implementation details.

Concept overview. Assuming that the Grüneisen parameter Γ is constant, p_0 can be expressed with the following equation: $p_0(\lambda) \propto \mu_a(\lambda) \cdot \phi(\lambda)$. Prior work by Tzoumas et al.[8] has shown that it is theoretically feasible to compute pixel-wise sO_2 values based on p_0 spectra by assuming linear mixture models for both μ_a and ϕ. Here, the optical absorption coefficient can be expressed by the weighted sum of the constituting chromophore spectra c: $\mu_a = \sum_k^N a_k \cdot c_k$ and the fluence can be expressed as a weighted sum of eigenspectra e: $\phi = \sum_k^M b_k \cdot e_k$. Both c_k and e_k are known *a priori*. It is assumed that the extraction of p_0 with sufficient wavelengths enables a reliable sO_2 estimation from the spectrum, as p_0 can only be explained by a limited amount of plausible μ_a and ϕ combinations. However, it is challenging to constrain a conventional minimization algorithm to yield a unique solution when used on handheld PA device geometries. In recent work of Olefir et al.[26] it was shown that deep learning algorithms can help to mitigate these issues. One of the core assumptions of *learned spectral decoloring* (LSD) is that the optimal set of constraints for the inversion can be learned from the wavelength-dependent changes of p_0 using data-driven approaches. A key challenge in implementing such a method is the lack of labeled ground truth data. This is addressed by using simulated in silico training data to train a neural network that approximates a function f_{LSD} which maps the initial pressure spectra S_{p_0} to corresponding blood oxygenation saturation sO_2 values:

$$f_{LSD} : S_{p_0} = \begin{pmatrix} p_{0_{\lambda_1}} \\ \dots \\ p_{0_{\lambda_n}} \end{pmatrix} \in \mathbb{R}^n \to sO_2 \in \mathbb{R}, \tag{1}$$

where n is the number of recorded wavelengths. f_{LSD} is a neural network that is trained to compensate for different levels of spectral coloring and that learns a mapping strategy in which many differently colored p_0 spectra correspond to the same sO_2 value. Due to an inherent lack of ground truth sO_2 values for experimental p_0 measurements, the method is trained on simulated data sets that can be optimized for the specific applications and wavelengths. Many samples of differently colored spectra are obtained from the same in silico sample by extracting single-pixel spectra from multiple spatial locations. For example, the influence of the gradual absorption of energy by water is expressed to a greater extent for deep samples, as there has been a greater amount of
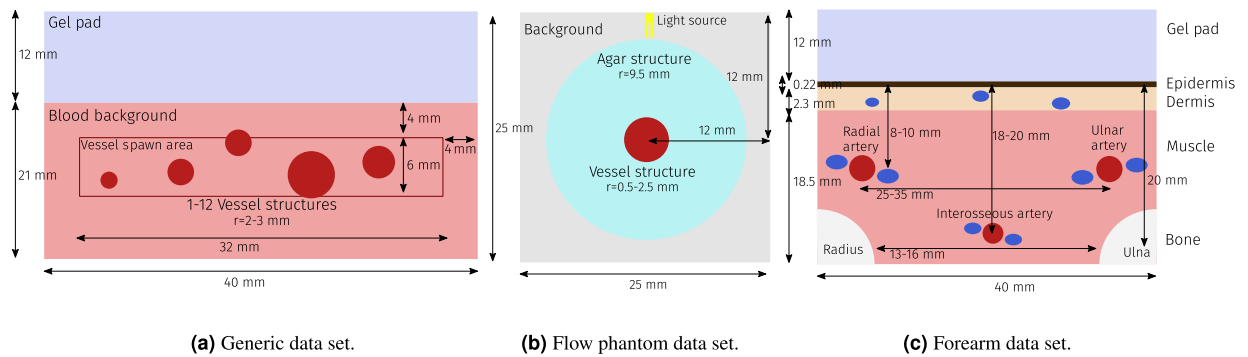
Figure 2. Schematic representation of the three in silico data sets. (**a**) shows the generic data set, (**b**) depicts the in silico flow phantom data set, and (**c**) visualizes the structures of the forearm data set. The vascular structures are simulated as tubes. The dark blue structures specifically correspond to veins.(**a, c**) were simulated with a digital twin of a custom PAI device based on the DiPhAs imaging system (Gröhl et al.[28]), whereas (**b**) was simulated with a pencil beam as the illumination source.

interaction between the light and the chromophore. This leads to different p_0 spectra for the same set of optical parameters. A visual representation of the spectra extraction from simulated p_0 data is shown in Fig. 1.

During training, the algorithm is given tuples $\left(S_{p_0}, sO_2\right)$, with $S_{p_0} \in \mathbb{R}^n$ and $sO_2 \in \mathbb{R}$. Each spectrum S_{p_0} is normalized such that all vector components sum to one ($\sum_{i=1}^{n} S_{p_{0_{\lambda_i}}} = 1$).

The amplitude information of the recorded spectra is crucial for quantitative photoacoustic imaging which aims to obtain absolute concentrations of chromophores in tissue. However, by sacrificing the amplitude information, we also eliminate the need to calibrate the in silico training data to the acquisition device and the specific target domain. This drawback is tolerable, as we are only interested in the relative ratios of HbO_2 and Hb. Furthermore, it constitutes an important step towards bridging the *domain gap* and thereby enables the LSD method to be applied to real data with unknown calibration of the PA measurement device. Real data needs to be reconstructed from raw time series pressure data in order to form a spatial image of signal amplitudes. A reconstruction algorithm can introduce artifacts to the image, for example, due to limited view geometries[27]. However, in this work, the biases introduced by the reconstruction are assumed to be independent of the wavelength, allowing the algorithm to be directly trained on p_0 data without the need to simulate the acoustic forward model as well.

Data sets. Several data sets were used in this work for training and validation. These validation data comprised (a) an in silico data set, (b) in vitro data set from a blood flow phantom, and in vivo data sets from (c) an open porcine brain and (d) forearms of healthy human volunteers. The purpose of the in silico data sets was to evaluate the sO_2 estimation accuracy that could potentially be achieved by the LSD method; the in vitro data set was used to investigate if the method is capable of recovering the entire value range when the sO_2 is chemically decreased from 100% to 0% in a controlled manner and the in vivo data sets were used to demonstrate that the method is capable of obtaining plausible values in real data derived from complex media.

(a) **Synthetic data.** A total of three simulated data sets were generated to train the models and test the method in different scenarios (Fig. 2). All of these had different a priori assumptions for the underlying optical tissue properties.

1. The *generic data set* (Fig. 2a). This data set contained a generic tissue representation without skin-specific chromophores such as melanin and was used for estimations of the open brain data. It contained randomly distributed vessel structures with 100% blood volume fraction, a homogeneous background medium with 0.5% blood volume fraction, and a scattering coefficient of 10 cm^{-1}. All structures were initialized with the same random blood oxygenation levels that were drawn from a uniform distribution from 0% to 100% oxygenation. The phantoms were simulated with 26 wavelengths, equidistant from 700 nm to 950 nm, using a multi-threaded adaptation of the Monte Carlo framework mcxyz[29] with 10^7 photons for each simulation.

2. The *flow phantom data set* (Fig. 2b). This data set was designed to resemble the geometric setup of the oxygenation flow phantom as presented by Gehrung et al.[30]. Following the specifications of the agar phantom, the structure was assumed to have a reduced scattering coefficient μ_s' of 5 cm^{-1}. For increased variability, we randomly changed the water content between 50 and 100%. The agar structure contains a single tubular structure containing blood with a haemoglobin concentration of 150g/L and an oxygenation level uniformly randomized between 0 and 100%. In order to reduce the influence of discretization artifacts and to increase the effective number of samples, the radius of the tube was uniformly randomized between 0.5 and 2.5 mm for each simulation. The phantoms were simulated with the same wavelengths used for data acquisition: {660, 664, 680, 684, 694, 700, 708, 715, 730, 735, 760, 770, 775, 779, 800, 850, 950} nm. The MCX simulation framework[31,32] was used to simulate this data set due to its fast computational speed with 10^7 photons for each simulation.

Tissue type	Blood volume [%]	Oxygenation [%]	Melanin [%]	Water [%]
Gel Pad	0	–	0	0
Epidermis	0	–	2.2 ± 1[34]	0
Dermis	1	80 ± 10	0	58[35]
Muscle	1	80 ± 10	0	68[35]
Vessel	100	0–100	0	0
Artery	100	0.95 ± 5[36]	0	0
Vein	100	70 ± 10[37]	0	0
Bone	0	–	0	19 ± 1[38]

Table 1. Assumed property ranges and chromophore abundances for the different tissue types. For each instance of a forearm phantom was created, random values for ranges X–Y were drawn from a uniform distribution and values for ranges $X \pm Y$ were drawn from a Gaussian normal distribution. The given values refer to the volume fraction of the chromophore. Here, whole blood was assumed to have a hemoglobin concentration of 150 g/L[33].

3. The *forearm model data set* (Fig. 2c). The forearm data set was designed to mimic all structures in the human forearm, including the chromophore melanin that is present in the epidermis. The synthetic forearm phantoms were also simulated with 26 wavelengths, equidistant from 700 to 950 nm, using the MCX simulation framework with 10^7 photons for each simulation. The optical properties of the different structures were modeled as reviewed by Jacques[33] and a constant anisotropy of $g = 0.9$ was assumed. Table 1 shows the assumed physiological ranges for different parameters in the respective structures.

For all simulated data sets, spectra were extracted from a region of interest (ROI), defined as vessel structures where the signal at the isosbestic point of 800 nm was higher than a noise equivalent threshold (determined by calculating the pixel-wise contrast-to-noise ratio (CNR) and setting a threshold of CNR ≥ 2). This was done because hardware limitations in overall sensitivity and acoustic frequency responses make it impossible to extract any meaningful information hidden deep in a blood vessel in in vivo images.

(b) **Blood flow phantom data.** This data set consisted of three measurements of a blood flow phantom setup, including reference blood oxygenation measurements provided by partial oxygen pressure (pO₂) needle probes. It contains measurements of two human blood samples and a rat blood sample. A diagram of the measurement setup and a detailed description of the data acquisition process can be found in Hacker et al.[39]. The data was measured at the University of Cambridge using an *MSOT inVision 256-TF* imaging system (iThera Medical GmbH, Munich, Germany). The blood samples were first chemically oxygenated and then chemically deoxygenated during the measurement process, theoretically going from 100% blood oxygenation to 0% blood oxygenation over the measurement time, with continuous reference measurements being taken by pO₂ needle probes. These pO₂ measurements were translated into sO₂ estimates using the Severinghaus equation[40,41]. For the evaluation of the method on this data set, ten consecutive frames of the same wavelength were averaged to account for laser pulse energy fluctuations. The tube structure was automatically segmented by only taking pixels in which the signal at 800 nm was greater than 2×10^4 MSOT signal units into account. This threshold was chosen to yield a good fit of the vessel structure for each of the data sets. This step was necessary, as the tubular structure was subject to slight movements over the imaging duration, and as such, it was not feasile using a constant manually-segmented ROI for all of the images.

(c) **Porcine brain data.** This data set consists of a multispectral image series that was taken from a previous animal experiment in which a porcine brain was imaged during open brain surgery[42]. The used images were acquired as part of a pilot study to see the hemodynamic responses of the brain during spreading depolarization. The specific data used for this study corresponded to a baseline measurement done to establish the capability of PAI to distinguish different hemodynamic states which were induced by using different levels of respiratory oxygen. During the entire length of the imaging procedure, 38 minutes, the animal was supplied with different levels of respiratory oxygen (rO₂) mixed in the mechanical ventilation air flow, to induce changes in the hemodynamics of the brain. Specifically, the rO₂ was set to these values during the experiment: 35% from minute 0 to 5 (baseline), 21% from minute 5 to 10 (normoxia), 0% from minute 10 to 16 (anoxia), started recovery with 21% from minute 16 to 21 (normoxia), and 100% from minute 21 to 26 (hyperoxia). Finally, the rO₂ was again set to the baseline (35%) for the remainder of the experiment. Towards the end of each interval, we took arterial blood gas measurements as a reference. In the original study, we used three month old female German Landrace swines weighing 30–35 kg with a sample size of N = 3. In this evaluation, we used the experiment that had the most reliable blood gas measurements as assessed by a clinician and analysed the blood oxygenation over time using linear unmixing and learned spectral decoloring on the photoacoustic measurements. The images were recorded at the same wavelengths as in the training data set (700–950 nm at intervals of 10 nm). They were normalized by the recorded laser
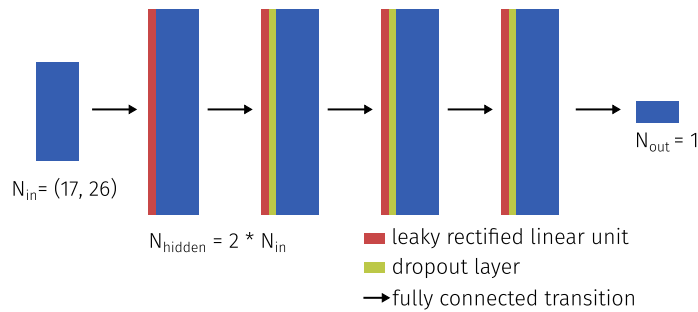
Figure 3. Visualization of the network architecture used for this work. The hidden layer has a size of twice the input layer. The blue color represents a layer of the network, the black arrows correspond to a fully connected transition, where every neuron of the previous layer is connected to the next. Red and green represent leaky rectified linear units and dropout layers, respectively.

energy and reconstructed with the delay-and-sum algorithm using a hamming window and were recorded using a custom PAI device based on the DiPhAs imaging system (Kirchner et al.[43]). Further details on the experimental design and the general image acquisition can be found in previous publications of the experiments[42,44].

(d) **Human forearm data.** This data set consists of multispectral images of the left and right forearms of three healthy human volunteers. The images were taken at three distinct positions a distance of approximately 2, 4, and 6 cm from the radiocarpal joint, leading to a total of 18 image sequences. The person operating the handheld PAI device attempted to capture either the *arteria radialis* or the *arteria ulnaris* in the imaging plane. These vessels could be identified by their pulsating motion which is induced by the heartbeat. The data sets were recorded using the MSOT Acuity Echo PAI device (iThera Medical GmbH, Munich, Germany). The images were taken at a wavelength range from 700–950 nm in intervals of 10 nm for at least 30 s. The ten subsequent multispectral sequences with the least amount of movement were averaged to account for laser intensity fluctuations and to increase the robustness against motion artifacts. Finally, the images were scaled to a size of 256×128 pixels. The vessel structures were manually annotated using the corresponding PA and US images. For the final segmentation masks only pixels were considered that had a CNR ≥ 2 and vessels containing less than 20 pixels were excluded to ensure a relevant sample size for the sO_2 estimates.

Deep learning models. For each of the training data sets, a separate fully connected feed-forward neural network was trained to account for domain-specific differences. The feed-forward architecture was chosen because we decided to use a well-understood baseline method for this initial study and a single-pixel approach does not warrant the use of convolutional neural networks. The number of input features was set to the number of wavelengths in the respective multispectral sequence (26 for the forearm and generic data set and 17 for the flow phantom data set), the model contained four hidden layers and the size of these hidden layers was set to be twice the size of the input vector (see Fig. 3).

The models were trained for 100 epochs, where one epoch contained 500 batches of size 10^4. The initial learning rate was set to 10^{-2} and was updated every two epochs to $new_{lr} = 10^{-2} \times 0.9^{(epoch/2)}$. After each leaky rectified linear unit, we applied a dropout of 20% to prevent the network from overfitting. Tracking of the validation losses over the number of epochs showed that the validation loss did not significantly decrease after as little as ten epochs using this training scheme. For each synthetic data set, a model was trained on 75% of the available data, 5% of the data were used for validation and the final 20% of the data was used as a held-out test set. The reported hyperparameters were optimized based on the performance of the method on the validation set. The test set was only evaluated once, at the end of the training process, to obtain the final results.

Linear unmixing. As the LU technique constitutes the state-of-the-art method in functional parameter estimation for photoacoustic imaging, it was used as a reference method to compare the proposed LSD method to. It was performed using literature absorption spectra of pure Hb and HbO_2 as reviewed by Jacques[33]. The unmixing method was implemented in Python 3.7., using the *minimize* function of the *scipy* python package that implements the SLSQP (Sequential Least SQares Programming) algorithm for finding the best fit. The unmixing was done exclusively for Hb and HbO_2, using initial values of 0.5.

Ethical approval. The healthy human volunteer experiments were carried out in accordance with relevant guidelines and regulations and were approved by the ethics committee of the medical faculty of Heidelberg University under reference number S-451/2020. The study is registered with the German Clinical Trials Register under reference number DRKS00023205. All porcine experiments were carried out in accordance with relevant guidelines and regulations, including the ARRIVE guidelines, and protocols were approved by the institutional
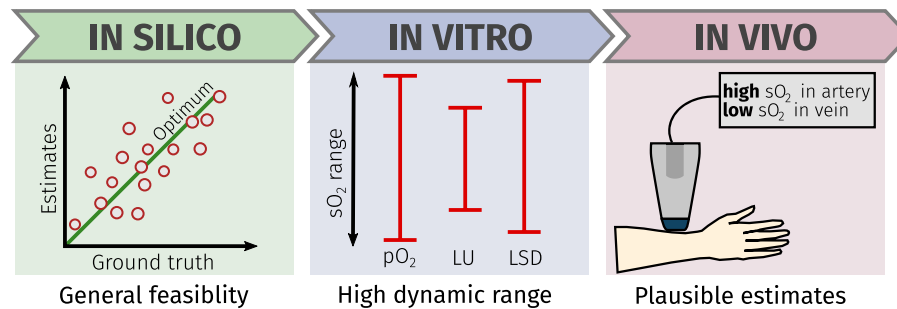
Figure 4. Stylized summary of the key findings of the experiments. The in silico experiments demonstrated the general feasibility of the LSD method, the in vitro experiments revealed the large dynamic range of the LSD estimates, and the in vivo experiments showed that the method yields more plausible estimates than LU even in complex situations.
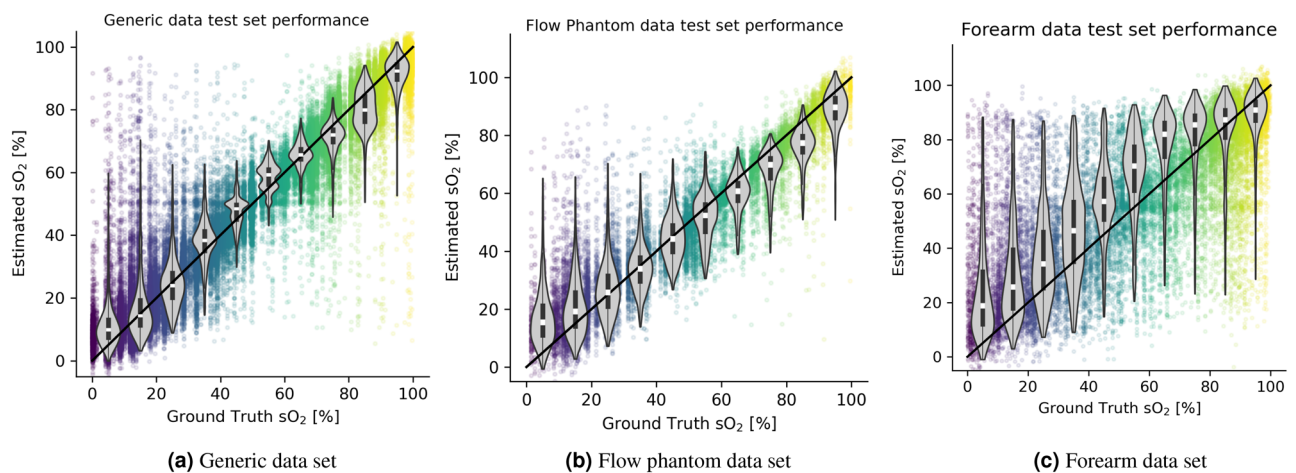

(a) Generic data set   (b) Flow phantom data set   (c) Forearm data set

Figure 5. In silico estimation results for the generic data set (**a**), the flow phantom data set (**b**), and the forearm data set (**c**). The scatter plot is colored with the ground truth oxygenation value. The violin plots show the estimated sO_2 for the ground truth sO_2 intervals in increments of 10%. As such, in addition to the scatter plot, there is one violin plot for all ground truth sO_2 values in equidistant steps of 10%.

animal care and use committee in Karlsruhe, Baden-Wuerttemberg, Germany (Protocol No. 35-9185.81/G-174/16).

Informed consent. Informed consent was obtained from all volunteers.

Results
This section presents the results of the LSD method on the simulated in silico data sets, on the in vitro flow phantom data sets, and on the in vivo porcine brain and human forearm data. A separate LSD regressor was trained for each of the respective in silico training data sets. The key findings of the experiments are summarized in Fig. 4.

In silico results. Figure 5 illustrates the performance of the respective deep learning model when tasked with predicting sO_2 values for the test set from the generic data set, the flow phantom data set, and the forearm data set. Here, the relative sO_2 estimation error is reported, which was calculated using the equation $e_{sO_2} = |sO_2^{EST} - sO_2^{GT}|/sO_2^{GT}$, with sO_2^{EST} being the estimated oxygen saturation and sO_2^{GT} being the ground truth oxygen saturation.

The median relative sO_2 estimation error for the model trained and tested on the generic tissue model data set was 6.1%, with an interquartile range (IQR) of (2.4%, 18.7%). On the flow phantom data set, the LSD method achieved a median relative estimation error of 9.9%, with an IQR of (3.6%, 28.5%). The largest error produced by the LSD method was found in the in silico forearm data set with a relative median quantification error of 15.0%, and an IQR of (5.3%, 45.4%).

The median *absolute* sO_2 quantification error on the test sets was well below 10 percentage points for all data sets. The model that was trained and tested on the forearm data set achieved 7.9 percentage points median absolute sO_2 estimation error with an IQR of (3.5, 17.1) percentage points.
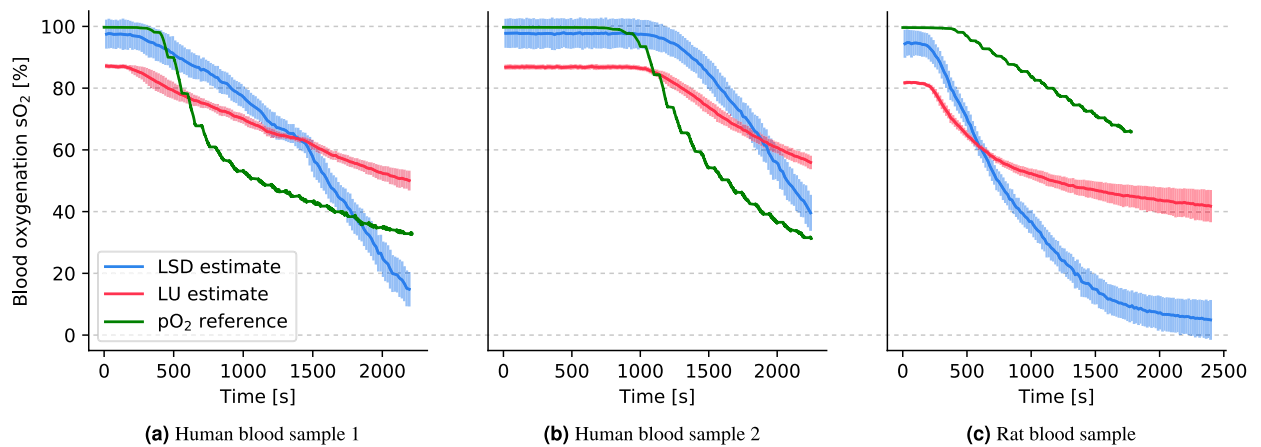
Figure 6. The mean oxygenation estimation results from three different measurement methods, shown over time on three different blood samples: (1) LSD in blue, (2) LU in red, and (3) pO$_2$ reference measurement in green. The standard deviation of the estimations within the ROI for LU and LSD unmixing is shown around the mean estimate in the corresponding color. The graphs are shown for human blood samples (**a**) and (**b**) and for a rat blood sample (**c**).

In vitro results. For a comparative analysis of the flow phantom data, three techniques of sO$_2$ estimation are shown in Fig. 6: spectral unmixing using LU, spectral unmixing using the proposed LSD approach, and pO$_2$ probe reference measurements. The mean and standard deviation of the estimates for both the LU and LSD approach are shown on the graphs.

All three of these measurements showed a monotonous decrease in the blood oxygenation level over the time frame of the experiment. The pO$_2$ reference measurements yielded a value range from $\approx 100\%$ to 5% blood oxygenation. In contrast, LU exhibited a dynamic range of $\approx 85\%$ to 40%, whereas LSD results exhibited a dynamic range of $\approx 95\%$ to 5% on all three data sets. It should be noted that the pO$_2$ references could not be perfectly aligned due to constraints of the equipment and procedure. The temporal calibration of the human blood sample (b) was manually adjusted by 1000 s, as the curves did not seem to match initially. Furthermore, the pO$_2$ reference was stopped prematurely in the rat blood data set.

In addition, Fig. 7 shows the depth-dependent behavior of the unmixing result for two time-points in the measurement. In both cases one can clearly see, that the rim-core differences in the sO$_2$ estimates are more pronounced in LU when compared to LSD. The LSD estimates are more homogeneous throughout the imaging cross section.

In vivo results. In the porcine brain image series, LSD was observed to increase the dynamic range of the predictions while maintaining the same tendency (high values were mapped to higher values and low values were mapped to lower values). Median oxygenation values were computed and tracked over time in a manually placed ROI in Fig. 8. In addition to the comparison of LSD and LU, arterial blood gas analysis measurements were taken and are shown in the figure as well.

Figure 9 shows the results of the method on the in vivo forearm data set on two example images. For completeness, the results on all 18 image slices are available in the Supplemental Material S1. The LSD estimates were obtained from a deep learning model trained on the synthetic forearm data set. The results were compared to data analyzed with LU. On the left image, the arterial vessel structures are estimated to have a blood oxygenation levels of 64.7% with LU and 89.6% with LSD (yellow) and 68.2% with LU and 91.6% with LSD (red). On the right image, LSD estimates a blood oxygenation level of 81.9% for the artery (red), while LU yields 60.4%. The mean spectra of these ROIs are plotted in the figure.

Discussion

This paper introduces the machine learning-based method *LSD*, which is able to account for the spectral coloring effects in multispectral photoacoustic imaging when estimating the blood oxygen saturation of tissue. Other recently published deep learning-based methods have only reported results on simulated data[14–17], or preliminary results on simple phantom setups[18,19]. This shortcoming in the field may be attributed to the *domain gap* between simulated data and real measurements. LSD is based on the assumption that the consideration of spatial relations in PA measurements in simulations amplifies the *domain gap* between simulation and reality. Therefore, it is trained on single-pixel p$_0$ spectra at different spatial locations within tissue. We demonstrated that this method is able to predict plausible blood oxygenation levels in vivo and has distinct advantages compared to linear spectral unmixing as shown in both in vitro and in vivo data sets.

In vivo measurements of the forearms of healthy human volunteers suggest that LSD is capable of yielding physiologically plausible sO$_2$ measurements. In line with the expected literature values for physiological value ranges of blood oxygenation, highly oxygen saturated blood is systematically estimated to have higher sO$_2$ with LSD than LU, while poorly oxygenated blood is systematically estimated to have lower sO$_2$ with LSD than LU,
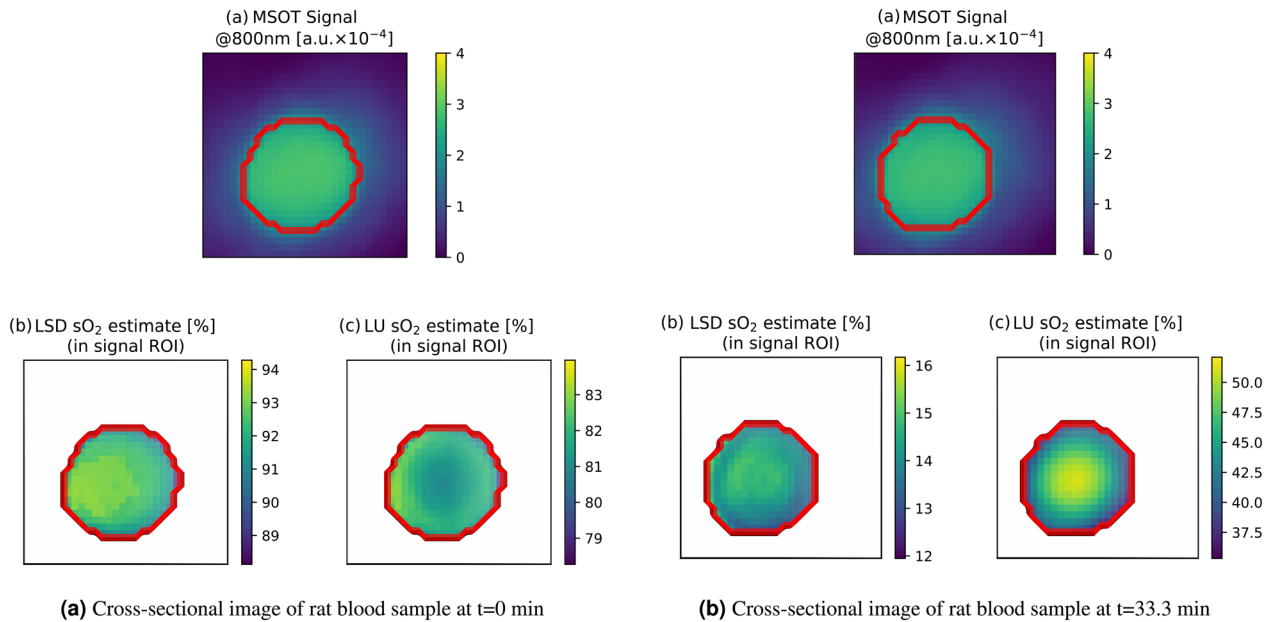
(a) Cross-sectional image of rat blood sample at t=0 min   **(b)** Cross-sectional image of rat blood sample at t=33.3 min

Figure 7. Visualization of a cross-sectional view through the flow phantom rat blood data. The left plot (**a**) shows the spatial distribution of sO$_2$ estimates in the beginning of the experiment (t = 0 min) and the right plot (**b**) towards the end of the experiment (t = 33.3 min). The rim-core differences in the sO$_2$ estimates are more pronounced in LU when compared to LSD.
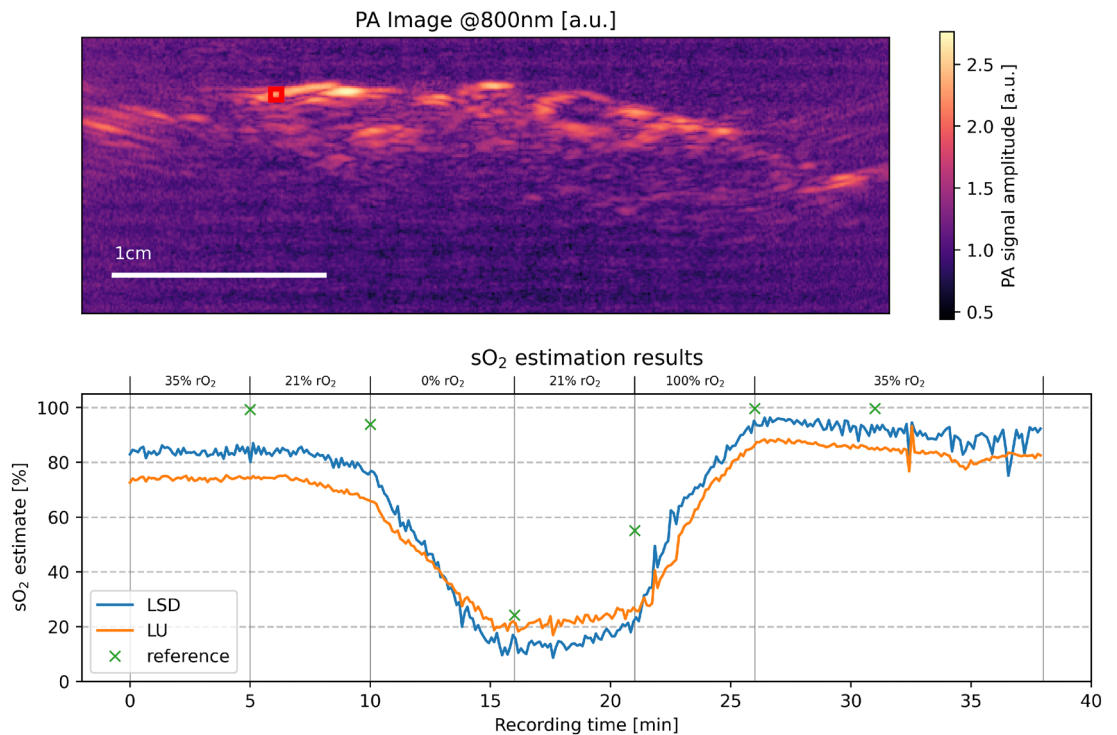
Figure 8. The results of LSD in vivo on an open porcine brain with a deep learning model trained on the generic tissue data set. The LSD results are compared to the LU results. The red rectangle shows an ROI which the LSD (blue) and LU (orange) results were computed on. The green crosses mark the time points and values of reference arterial blood gas analysis (BGA) measurements.

bringing the estimations closer to the anticipated ground truth. This conclusion is supported by the performance of the LSD method on in vitro flow phantom data. Here, LSD exhibits a higher dynamic range of the sO$_2$ estimates than LU while showing a monotonous decrease in sO$_2$ that could be confirmed with pO$_2$ reference measurements. While the steady decrease of the sO$_2$ measurements is apparent for all three approaches, the pO$_2$ measurements decrease at different rates compared to both LSD and LU. One reason for this mismatch might be
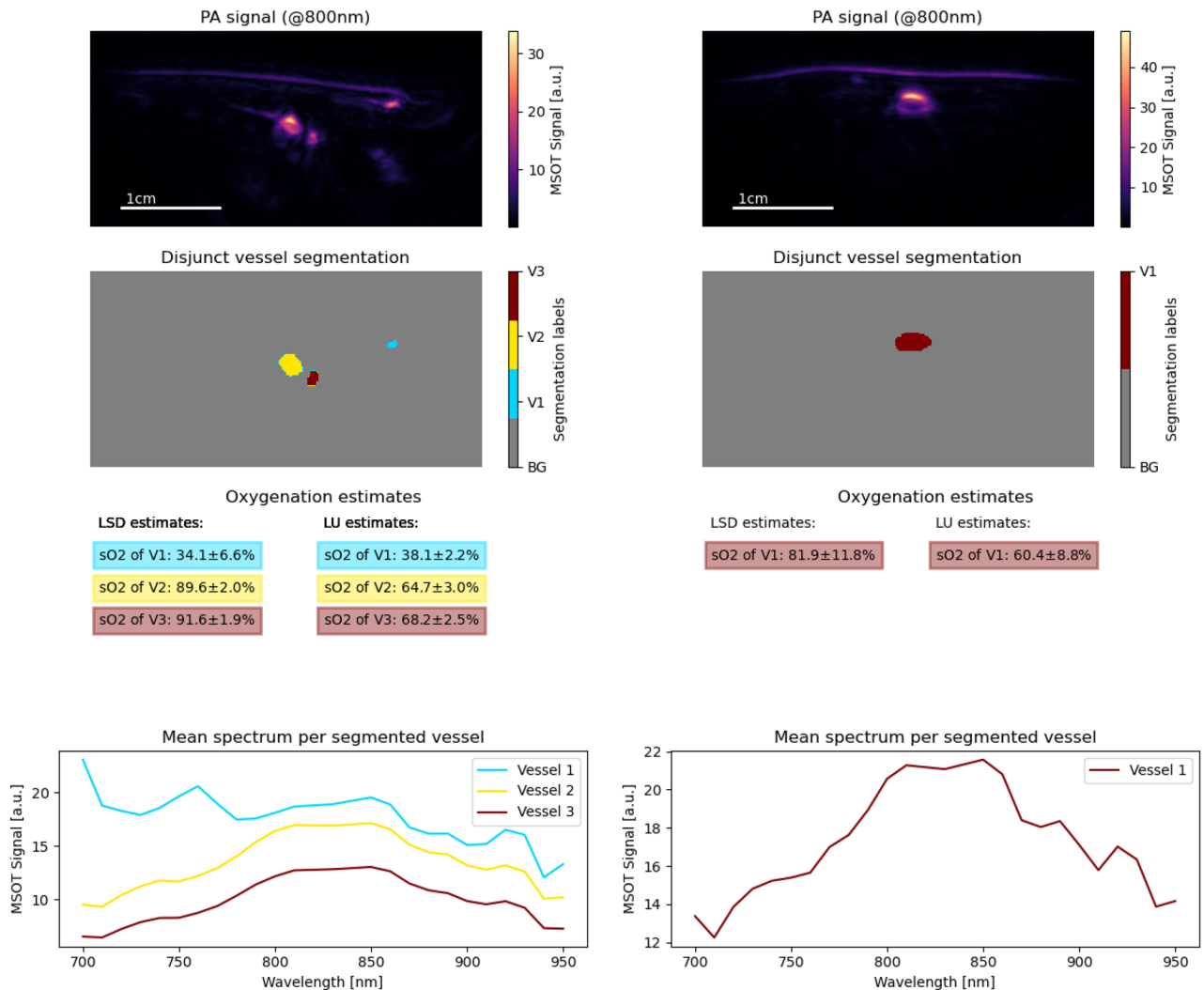
Figure 9. The results of LSD in vivo on the forearms of two healthy human volunteers. The LSD model was trained on the in silico forearm data set. The results were compared to the results for data analyzed with LU. The top row shows the PA signal at 800 nm, the second row shows the segmentation masks of the imaged vessels, the third row compares the LU and the LSD sO$_2$ estimates, and the last row shows the mean MSOT signal spectrum for each of these vessels.

the empirical nature of the Severinghaus equation, which was used for translating the measurements into sO$_2$ values, along with an unknown pH value and a temperature mismatch. Moreover, as the pO$_2$ probe could not be placed inside the PAI system, PA and pO$_2$ measurements were made at different positions along the blood tubing leading to further misalignments. In future work, this might be remedied by adding blood gas analyzer measurements as a reference, by using a more physiological blood gas oxygenation process instead of chemical oxygenation, or by creating a ground truth reference using a phantom setup in which, a priori, the exact absolute concentrations of the mixed chromophores are known. Nevertheless, the LSD approach shows great promise on this data set, through the expanded dynamic range of the response and by significantly reducing the rim-core effect that is a sign of spectral coloring in depth. An undisputed advantage of machine learning methods is their fast inference speed after training. For a 256 × 128 pixels image with 26 wavelengths, the inference time was measured to be as low as 211 ± 9 ms using a central processing unit (CPU) (AMD Ryzen 5 1600 Six-Core Processor) and 2.4 ± 0.4 ms using a graphics processing unit (GPU) (Nvidia GTX 1080ti, 11 GB). As such, the real-time capability of this method would mostly be impeded by the hardware constraints and not by the method. For example, slow sO$_2$ estimation times can be expected when using many wavelengths and averaging steps on a system with a low pulse repetition rate.

The LSD method is capable of performing estimations in realistic settings, despite being trained on simulated data, demonstrating the ability of the approach to bridge the domain gap between real and simulated images. In fact, to our knowledge, no prior work has successfully applied convolutional neural network-based approaches to sO$_2$ estimation on entire images of initial pressure in vivo [45]. On the other hand, since single-pixel spectra can be highly ambiguous when considering a too large solution space, thus we strike a careful balance between generalizability and applicability of the method by designing a suitable data set for a specific application. In

addition to this trade-off, another severe limitation of the method is that the inversion model is wavelength-dependent, which means that a trained model can only estimate sO_2 values for a specific set of pre-determined wavelengths. Furthermore, while the variation in the target structures included in the data sets increased, so did the level of inaccuracy of the in silico sO_2 inversion results. Only in the forearm data set the background and the vessel structures were allowed to take different oxygenation values. This can increase the difficulty for inversion because a source of ambiguity is added to the problem. In order to obtain more accurate results in these cases, techniques for spatial regularization could be employed in future to counteract the ambiguity of the inversion problem. We trained the LSD models directly on p_0 spectra and thus assume that the acoustic forward process is independent of the incident wavelength and has no major influence on the spectral behavior. Considering the impact that reconstruction algorithm-specific artifacts can have on the spectra it may be beneficial to incorporate acoustic forward modelling and image reconstruction into the simulation pipeline in the future.

In summary, LSD provides encouraging advantages over conventional LU techniques for sO_2 estimation: the expanded dynamic range of the response, the significantly reduction of rim-core effects that are a sign of spectral coloring in depth, as well as fast inference times. Thus, the method comprises a promising framework for more accurate estimations of PAI biomarkers in the future.

Data availability

References

1. Farrohknia, N. *et al.* Emergency department triage scales and their components: A systematic review of the scientific evidence. *Scand. J. Trauma Resuscitation Emerg. Med.* **19**, 42. https://doi.org/10.1186/1757-7241-19-42 (2011).
2. Brown, E., Brunker, J. & Bohndiek, S. E. Photoacoustic imaging as a tool to probe the tumour microenvironment. *Dis. Models Mech.* **12**, dmm039636 (2019).
3. Kelly, A.-M. Can venous blood gas analysis replace arterial in emergency medical care. *Emerg. Med. Austral.* **22**, 493–498. https://doi.org/10.1111/j.1742-6723.2010.01344.x (2010).
4. Haacke, E. M., Lai, S., Yablonskiy, D. A. & Lin, W. In vivo validation of the bold mechanism: A review of signal changes in gradient echo functional mri in the presence of flow. *Int. J. Imaging Syst. Technol.* **6**, 153–163. https://doi.org/10.1002/ima.1850060204 (1995).
5. Tak, S. & Ye, J. C. Statistical analysis of FNIRS data: A comprehensive review. *Neuroimage* **85**, 72–91. https://doi.org/10.1016/j.neuroimage.2013.06.016 (2014).
6. Severinghaus, J. W. & Honda, Y. History of blood gas analysis. vii. Pulse oximetry. *J. Clin. Monit.* **3**, 135–138. https://doi.org/10.1007/BF00858362 (1987).
7. Li, M., Tang, Y. & Yao, J. Photoacoustic tomography of blood oxygenation: A mini review. *Photoacoustics* **10**, 65–73. https://doi.org/10.1016/j.pacs.2018.05.001 (2018).
8. Tzoumas, S. *et al.* Eigenspectra optoacoustic tomography achieves quantitative blood oxygenation imaging deep in tissues. *Nat. Commun.* **7**, 12121. https://doi.org/10.1038/ncomms12121 (2016).
9. Perekatova, V., Subochev, P., Kirillin, M. Y. & Turchin, I. Fluence compensated optoacoustic measurements of blood oxygen saturation in vivo at two optimal wavelengths. In *Photons Plus Ultrasound: Imaging and Sensing 2017*, vol. 10064, 100645K. https://doi.org/10.1117/12.2250851 (International Society for Optics and Photonics) (2017).
10. Tzoumas, S., Deliolanis, N. C., Morscher, S. & Ntziachristos, V. Unmixing molecular agents from absorbing tissue in multispectral optoacoustic tomography. *IEEE Trans. Med. Imaging* **33**, 48–60. https://doi.org/10.1109/TMI.2013.2279994 (2013).
11. Ding, L. *et al.* Non-negative constrained inversion approaches for unmixing chromophores in multispectral optoacoustic tomography. In *Photons Plus Ultrasound: Imaging and Sensing 2017*, vol. 10064, 100641B. https://doi.org/10.1117/12.2252807 (International Society for Optics and Photonics) (2017).
12. Glatz, J., Deliolanis, N. C., Buehler, A., Razansky, D. & Ntziachristos, V. Blind source unmixing in multi-spectral optoacoustic tomography. *Opt. Express.* **19**, 3175–3184. https://doi.org/10.1364/OE.19.003175 (2011).
13. Kirchner, T., Gröhl, J. & Maier-Hein, L. Context encoding enables machine learning-based quantitative photoacoustics. *J. Biomed. Opt.* **23**, 056008. https://doi.org/10.1117/1.JBO.23.5.056008 (2018).
14. Hoffer-Hawlik, K. & Luke, G. P. *absO2luteU-Net: Tissue Oxygenation Calculation Using Photoacoustic Imaging and Convolutional Neural Networks*. Master's thesis, Dartmouth College (2019).
15. Luke, G. P., Hoffer-Hawlik, K., Van Namen, A. C. & Shang, R. O-net: A convolutional neural network for quantitative photoacoustic image segmentation and oximetry. arXiv:1911.01935 **arXiv preprint** (2019).
16. Cai, C., Deng, K., Ma, C. & Luo, J. End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging. *Opt. Lett.* **43**, 2752–2755. https://doi.org/10.1186/1757-7241-19-421 (2018).
17. Yang, C., Lan, H., Zhong, H. & Gao, F. Quantitative photoacoustic blood oxygenation imaging using deep residual and recurrent neural network. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 741–744. https://doi.org/10.1109/ISBI.2019.8759438 IEEE (2019).
18. Durairaj, D. A. *et al.* Unsupervised deep learning approach for photoacoustic spectral unmixing. In *Photons Plus Ultrasound: Imaging and Sensing 2020*, vol. 11240, 112403H. https://doi.org/10.1117/12.2546964 (International Society for Optics and Photonics) (2020).
19. Chen, T. *et al.* A deep learning method based on u-net for quantitative photoacoustic imaging. In *Photons Plus Ultrasound: Imaging and Sensing 2020*, vol. 11240, 112403V (International Society for Optics and Photonics) (2020).
20. Agrawal, S. *et al.* Functional, molecular and structural imaging using led-based photoacoustic and ultrasound imaging system. In *Photons Plus Ultrasound: Imaging and Sensing 2020*, vol. 11240, 112405A. https://doi.org/10.1117/12.2547048 (International Society for Optics and Photonics) (2020).
21. Roll, W. *et al.* Multispectral optoacoustic tomography of benign and malignant thyroid disorders: A pilot study. *J. Nucl. Med.* **60**, 1461–1466. https://doi.org/10.1186/1757-7241-19-422 (2019).
22. Cox, B. T., Laufer, J. G., Beard, P. C. & Arridge, S. R. Quantitative spectroscopic photoacoustic imaging: A review. *J. Biomed. Opt.* **17**, 061202. https://doi.org/10.1117/1.JBO.17.6.061202 (2012).
23. Tarvainen, T., Cox, B. T., Kaipio, J. & Arridge, S. R. Reconstructing absorption and scattering distributions in quantitative photoacoustic tomography. *Inverse Prob.* **28**, 084009 (2012).

24. Buchmann, J., Kaplan, B., Powell, S., Prohaska, S. & Laufer, J. Quantitative pa tomography of high resolution 3-d images: Experimental validation in a tissue phantom. *Photoacoustics* **100157**, 20 (2020).
25. Hochuli, R., An, L., Beard, P. C. & Cox, B. T. Estimating blood oxygenation from photoacoustic images: Can a simple linear spectroscopic inversion ever work?. *J. Biomed. Opt.* **24**, 121914. https://doi.org/10.1117/1.JBO.24.12.121914 (2019).
26. Olefir, I. *et al.* Deep learning based spectral unmixing for optoacoustic imaging of tissue oxygen saturation. *IEEE Trans. Med. Imaging* **20**, 20 (2020).
27. Paltauf, G., Nuster, R., Haltmeier, M. & Burgholzer, P. Experimental evaluation of reconstruction algorithms for limited view photoacoustic tomography with line detectors. *Inverse Prob.* **23**, S81 (2007).
28. Gröhl, J., Kirchner, T., Adler, T. & Maier-Hein, L. Confidence estimation for machine learning-based quantitative photoacoustics. *J. Imaging* **4**, 147. https://doi.org/10.3390/jimaging4120147 (2018).
29. Jacques, S. L. Coupling 3d monte carlo light transport in optically heterogeneous tissues to photoacoustic signal generation. *Photoacoustics* **2**, 137–142. https://doi.org/10.1186/1757-7241-19-42 6 (2014).
30. Gehrung, M., Bohndiek, S. E. & Brunker, J. Development of a blood oxygenation phantom for photoacoustic tomography combined with online pO2 detection and flow spectrometry. *J. Biomed. Opt.* **24**, 1–11. https://doi.org/10.1117/1.JBO.24.12.121908 (2019).
31. Fang, Q. & Boas, D. A. Monte carlo simulation of photon migration in 3d turbid media accelerated by graphics processing units. *Opt. Express* **17**, 20178–20190. https://doi.org/10.1186/1757-7241-19-42 8 (2009).
32. Yu, L., Nina-Paravecino, F., Kaeli, D. R. & Fang, Q. Scalable and massively parallel monte carlo photon transport simulations for heterogeneous computing platforms. *J. Biomed. Opt.* **23**, 010504. https://doi.org/10.1117/1.JBO.23.1.010504 (2018).
33. Jacques, S. L. Optical properties of biological tissues: A review. *Phys. Med. Biol.* **58**, R37. https://doi.org/10.1088/0031-9155/58/11/R37 (2013).
34. Alaluf, S. *et al.* Ethnic variation in melanin content and composition in photoexposed and photoprotected human skin. *Pigment Cell Res.* **15**, 112–118. https://doi.org/10.1111/j.1742-6723.2010.01344.x 1 (2002).
35. Forbes, R. *et al.* The composition of the adult human body as determined by chemical analysis. *J. Biol. Chem.* **203**, 359–366 (1953).
36. Merrick, E. B. & Hayes, T. J. Continuous, non-invasive measurements of arterial blood oxygen levels. *Hewlett-packard J.* **28**, 2–9 (1976).
37. Molnar, Z. & Nemeth, M. Monitoring of tissue oxygenation: An everyday clinical challenge. *Front. Med.* **4**, 247. https://doi.org/10.3389/fmed.2017.00247 (2018).
38. Timmins, P. & Wall, J. Bone water. *Calcif. Tissue Res.* **23**, 1–5. https://doi.org/10.1007/BF02012759 (1977).
39. Hacker, L., Brunker, J., Smith, E. S., Quiros-Gonzalez, I. & Bohndiek, S. E. Photoacoustics resolves species-specific differences in hemoglobin concentration and oxygenation. *J. Biomed. Opt.* **25**, 095002 (2020).
40. Severinghaus, J. W. Simple, accurate equations for human blood o2 dissociation computations. *J. Appl. Physiol.* **46**, 599–602. https://doi.org/10.1111/j.1742-6723.2010.01344.x 3 (1979).
41. Collins, J.-A., Rudenski, A., Gibson, J., Howard, L. & O'Driscoll, R. Relating oxygen partial pressure, saturation and content: The haemoglobin-oxygen dissociation curve. *Breathe* **11**, 194–201. https://doi.org/10.1183/20734735.001415 (2015).
42. Kirchner, T. *et al.* Photoacoustics can image spreading depolarization deep in gyrencephalic brain. *Sci. Rep.* **9**, 8661. https://doi.org/10.1038/s41598-019-44935-4 (2019).
43. Kirchner, T. *et al.* An open-source software platform for translational photoacoustic research and its application to motion-corrected blood oxygenation estimation. arXiv:1901.09781**arXiv preprint** (2019).
44. Kirchner, T. *et al.* Photoacoustic monitoring of blood oxygenation during neurosurgical interventions. In *Photons Plus Ultrasound: Imaging and Sensing 2019*, vol. 10878, 108780C (International Society for Optics and Photonics) (2019).
45. Gröhl, J., Schellenberg, M., Dreher, K. & Maier-Hein, L. Deep learning for biomedical photoacoustic imaging: A review. arXiv:2011.02744**arXiv preprint** (2020).

Acknowledgements

Author contributions

Conceptualization: J.G., T.K., T.J.A., and L.M.-H.; data acquisition J.G., T.K., T.J.A., A.H.-A., M.A.H., E.S., and L.H.: data curation, J.G., N.H., T.K., and L.H.: formal analysis, J.G.: funding acquisition, L.M.-H.: investigation, J.G., T.K., T.J.A., L.H., and N.H.: methodology, J.G., T.K., and L.M.-H.: project administration, L.M.-H.: software, J.G.: supervision, E.S., S.E.B., and L.M.-H.: validation, J.G.: visualization, J.G.: method naming, T.K.: writing–original draft, J.G.; and writing–review and editing, J.G., T.K., T.J.A., N.H., M.A.H., A.H.-A., L.H., S.E.B., and L.M.-H.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-83405-8.

Correspondence and requests for materials should be addressed to J.G. or L.M.-H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.