Application of Transfer Learning for Rapid Calibration of Spatially-resolved Diffuse Reflectance Probes for Extraction of Tissue Optical Properties

Md Nafiz Hannan,[a] Timothy M. Baran[b,c,d]
[a] University of Rochester, Department of Physics and Astronomy, Rochester, New York, United States
[b] University of Rochester Medical Center, Department of Imaging Sciences, Rochester, New York, United States
[c] University of Rochester, Department of Biomedical Engineering, Rochester, New York, United States
[d] University of Rochester, The Institute of Optics, Rochester, New York, United States

Abstract

Significance: Treatment planning for light-based therapies including photodynamic therapy requires tissue optical property knowledge. These are recoverable with spatially-resolved diffuse reflectance spectroscopy (DRS), but requires precise source-detector separation (SDS) determination and time-consuming simulations.

Aim: An artificial neural network (ANN) to map from DRS at short SDS to optical properties was created. This trained ANN was adapted to fiber-optic probes with varying SDS using transfer learning.

Approach: An ANN mapping from measurements to Monte Carlo simulation to optical properties was created with one fiber-optic probe. A second probe with different SDS was used for transfer learning algorithm creation. Data from a third were used to test this algorithm.

Results: The initial ANN recovered absorber concentration with RMSE=0.29 µM (7.5% mean error) and μ_s' at 665 nm ($\mu_{s,665}'$) with RMSE=0.77 cm^{-1} (2.5% mean error). For probe-2, transfer learning significantly improved absorber concentration (0.38 vs. 1.67 µM, p=0.0005) and $\mu_{s,665}'$ (0.71 vs. 1.8 cm^{-1}, p=0.0005) recovery. A third probe also showed improved absorber (0.7 vs. 4.1 µM, p<0.0001) and $\mu_{s,665}'$ (1.68 vs. 2.08 cm^{-1}, p=0.2) recovery.

Conclusions: A data-driven approach to optical property extraction can be used to rapidly calibrate new fiber-optic probes with varying SDS, with as few as three calibration spectra.

Keywords: Diffuse reflectance spectroscopy, machine learning, neural network, transfer learning, Monte Carlo simulation

1 Introduction

Photodynamic therapy (PDT), which produces an antimicrobial effect through the photochemical production of reactive oxygen species, is a promising treatment modality for eradication of infectious disease[1]. We are particularly interested in the use of PDT to treat deep tissue abscesses, which are localized collections of infected, purulent fluid surrounded by a fibrous pseudo-capsule. We have shown that PDT is safe and feasible for this purpose in a Phase 1 clinical trial[2], and that PDT is efficacious against bacteria cultured from human abscesses[3]. Effective PDT requires sufficient doses of both the photosensitizer and treatment light. As the light dose is dependent upon the optical properties of the target tissue, namely absorption and scattering, accurate measurement of tissue optical properties is required for designing efficacious PDT treatment plans. To that end, we have constructed and validated a spatially-resolved diffuse reflectance spectroscopy system that is capable of extracting tissue optical properties using a fiber-optic probe placed in contact with the abscess wall[4]. Critically, the fiber-optic probe used to deliver light and collect diffusely reflected light was required to fit through the drainage catheter used for standard of care abscess drainage, which has an inner diameter of only 2 mm. This spectroscopy system was utilized during our recently completed Phase 1 clinical trial[2] to quantify human abscess wall optical properties [5]. Substantial inter-patient variability in the abscess wall optical properties was found[5]. We have shown that patient-specific treatment planning incorporating knowledge of these optical properties greatly improves delivery of an efficacious fluence rate to the abscess wall[6, 7]. Measurement of individual patient optical properties is therefore of high importance for the clinical success of PDT for this application.

To extract optical properties from diffuse reflectance spectroscopy, different approaches have been used, broadly categorized into two groups: analytical and numerical. Analytical approximations to the radiative transport equation such as the diffusion[8, 9] and P_3 [10] approximations have been utilized by many research groups. However, these approximations place limitations on source-detector separation, requiring detectors to be multiple transport mean free paths from the source. As described above, our primary clinical target requires that a fiber optic probe with multiple source detectors separations (SDS) must fit through a catheter with a diameter of 2 mm. As a consequence, these approximations are not valid for the probes we aim to deploy clinically, and more broadly for spatially-resolved diffuse reflectance at short SDS. Many research groups have therefore employed numerical approaches, such as the Monte Carlo lookup table (MCLUT) inverse model for optical property recovery[4, 11]. This approach is more flexible in terms of SDS range and is also valid for a wide range of optical properties. However, the usage of MC requires accurate SDS measurements for model creation, and generation of an extensive MCLUT which remains resource and time-intensive despite graphics processing unit (GPU) acceleration. Moreover, a distinct MC library must be generated for every probe employed. This is undesirable for cases where similar but distinct probes need be deployed, such as for single-use devices.

With advancements in computational power and machine learning (ML) algorithms, numerous research groups have turned to fully data-driven ML approaches for recovering optical properties from diffuse reflectance spectroscopy (DRS). Farrell *et al* pioneered the use of an artificial neural network (ANN) trained with the diffusion approximation to extract optical properties from simulated experimental data[8]. Recognizing the limitations of the diffusion approximation,

Kienle *et al* subsequently employed an ANN trained with MC simulations[12]. More recently, Nguyen *et al* demonstrated that deep learning models showed lower error and faster runtime compared to a MCLUT inverse model[13]. Lan *et al* developed an ANN model that achieved comparable accuracy to the traditional Monte Carlo-based inverse model while offering improved speed and flexibility[14].

However, these prior studies were confined to the performance of ANN-based optical property recovery for a single device. Data obtained from analytical solutions [8] and/or MC libraries [12, 14] for a specific probe in combination with phantom data collected by that probe were used to train these ANN models. None of these studies showed extension of ANNs to the case of unseen probes with variable SDS. As the ANNs created by these groups were probe-specific, they are not suitable for clinical adoption or more widespread application where rapid calibration of a large number of new devices is required. An ANN created for one specific device cannot be expected to perform well for other probes having dissimilar SDS, as the DRS signal will vary in unknown ways. To circumvent this with current methods, either the probe SDS would need to be very tightly controlled or a separate ANN would need to be created for each probe. Both of these solutions would impede clinical implementation by increasing costs due to tight tolerancing and prolonging the time required for calibration.

To facilitate our envisioned multi-center clinical trials and eventual widespread clinical adaptation, we propose a Transfer Learning (TL) approach for rapid probe-to-probe calibration. TL is a widely employed approach in the ML field, wherein models trained for one specific task are partially retrained to address a related task that has limited available data. We hypothesize that by using a feature-extraction based TL algorithm, an ANN created for one probe can be adapted to similar probes via a small dataset consisting of only a few calibration spectra collected by the target probe. This will enable the creation of an initial ANN for a single probe, which can be easily adapted to target probes without the need for the time and resource-intensive steps mentioned earlier, thus rendering large-scale clinical adoption more feasible.

This study followed a three-step workflow. First, an initial artificial neural network (ANN) was created and validated using data captured with a single fiber optic probe (probe-1). This initial ANN served as a pre-trained model for subsequent transfer learning steps. Second, we utilized data from a second fiber optic probe with slight differences in source-detector separation (probe-2) to create and validate a TL algorithm, which included selecting and optimizing the TL algorithm, TL dataset size, and associated hyperparameters. Finally, this TL algorithm was applied to data from a third fiber optic probe (probe-3) in a “real-world” scenario to independently assess its performance and applicability.

2 Methods

2.1 *Spectroscopy System*

The fiber optic probe-based diffuse reflectance system utilized in this study has been extensively described in our earlier publication [4]. Briefly, it is a continuous wave spectroscopy system in which broadband light from a tungsten halogen lamp (HL-2000-HP-FHSA, Ocean Optics, Inc.,

Largo, FL) is emitted from a source fiber into the sample. Eight detector fibers, which are placed at different source-detector separations (SDS) from the source fiber on the face of the probe, collect the reflected light for detection by a spectrometer (QE Pro, Ocean Optics). An encrypted, password-protected laptop computer is used to operate this system via a custom interface created in LabVIEW (National Instruments, Austin, TX).

2.2 *Probe Characterization*

Three custom-made fiber optic probes were used, each having an outer diameter of 2 mm. These probes were manufactured by Pioneer Optics Company (Bloomfield, CT), following the same design specifications. These specifications involved positioning the fiber faces at the distal end of the probe in a "plus" pattern, with a desired distance of 300 µm between adjacent fiber centers. While the design intended for each of these probes to possess the same SDS, the delivered probes had varying SDS values compared to each other. At the short separations used in these probes, precise knowledge of SDS is important for accurate recovery of tissue optical properties.

In order to determine the actual SDS for each probe, a stereomicroscope (SMZ1500, Nikon Instruments, Melville, NY) was used. The distal end of each probe was imaged both with and without a US Air Force resolution target (USAF 1951 1X, Edmund Optics, Inc., Barrington, NJ) in the same frame. The air force target was used as a distance reference. The centers of each fiber were identified and SDS for each fiber were determined. An image of the distal face of probe-1 showing the position of the source and detector fibers is presented in Figure 1. The SDS for all 8 detectors fibers for each probe are given in Table 1.

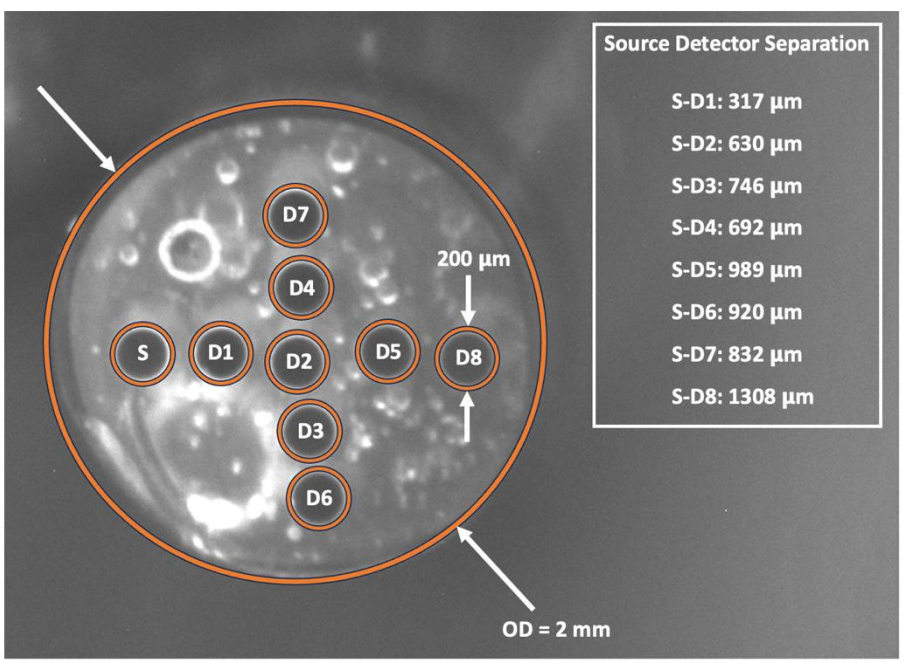

Figure 1: Image of the distal face of probe-1 showing the 200 µm diameter fibers. The source fiber (S) and detector fibers (D1-D8) are denoted using orange circles. The outside diameter is abbreviated as "OD."

Table 1: Source detector separation in µm for the 8 detector fibers of probes 1, 2 and 3

Probe No.	Detector Fiber 1	Detector Fiber 2	Detector Fiber 3	Detector Fiber 4	Detector Fiber 5	Detector Fiber 6	Detector Fiber 7	Detector Fiber 8
Probe 1	317	630	746	692	989	920	832	1308
Probe 2	315	648	738	761	985	914	911	1342
Probe 3	325	736	804	829	1082	951	990	1424

2.3 *Artificial Neural Networks*

2.3.1 *Creation and Validation of Initial ANN (ANN_1)*

An artificial neural network (ANN) approach was employed to recover optical properties from diffuse reflectance spectra (DRS). An initial ANN (ANN_1) was trained to map from the experimentally collected DRS for probe-1 to corresponding optical properties (absorption and reduced scattering spectra). It had 8 input nodes, one for each SDS, and 2 output nodes, one for each optical property (absorption and reduced scattering). It was created by combining two separate ANN ensembles, illustrated in Figure 2. The first ANN ensemble ($ANN_{EXP1\text{-}MC1}$) takes in experimental diffuse reflectance intensities at the 8 detector fibers as inputs, with outputs corresponding to Monte Carlo simulated reflectance intensities for these 8 fibers. The second ANN ensemble ($ANN_{MC1\text{-}OP}$) takes in these 8 simulated reflectance intensities and outputs corresponding optical properties. To get the final output of the ensemble models, predictions of individual ANNs were averaged. ANNs used in this research were created using the Keras API [15].

2.3.1.1 *Creation and Validation of $ANN_{EXP1\text{-}MC1}$ Network*

The $ANN_{EXP1\text{-}MC1}$ network was an ensemble of 10 individual ANNs as shown in Figure 2. Each of these ANNs were created from the same dataset ($TV_{EXP1\text{-}MC1}$ described in section 2.5.1.1) but with different train-validation splits and weight initializations. Given the relatively limited number of data points available for training (8848 samples), we opted for an ensemble approach to construct this network, which enhanced its stability and robustness. The input signals (experimental reflectance intensity) were not scaled. However, as the outputs (simulated reflectance intensity) of different fibers were of vastly different magnitudes, they were rescaled to [0,1] range using min-max normalization.

To train each of these ANNs, the dataset was randomly divided into two categories: Training (80%) and Validation (20%). Each individual ANN had 5 hidden layers (30 neurons in each layer) and an output layer consisting of 8 neurons resulting in 4238 parameters. The hidden layers used rectified linear unit (ReLU) activation, while the output layer employed a linear activation function. The optimal number of hidden layers and number of neurons in these layers was found using KerasTuner[15], the details of which are provided in the supplemental material. Training was continued for up to 1000 epochs. An early stopping callback was implemented to halt training if there was no improvement in the validation loss for 50 epochs, preventing overfitting. The model that demonstrated best performance on the validation set during training was restored and saved. The Adam optimizer with a mean squared error (MSE) cost function was used with a learning rate of 0.0005.

2.3.1.2 *Creation and Validation of $ANN_{MC1\text{-}OP}$ Network*

The optimal $ANN_{MC1\text{-}OP}$ network was an ensemble of 5 ANNs, each with 6 hidden layers depicted as depicted in Figure 2. The details of the optimization are provided in the supplemental material. The $TV_{MC1\text{-}OP1}$ dataset described in section 2.5.2 was used to train this network. Both the input and output data were scaled to [0,1] range. The dataset was randomly divided into training (80%) and test set (20%). An automated algorithm called Deep jointly informed neural networks (DJINN)[16] was used for automatic selection of the best ANN structure and hyperparameters without the need for time-consuming manual searching. DJINN automatically determined the near-optimal hyperparameters for our dataset, which were 250 epochs, batch size of 4916, and a learning rate of 0.007. By automating hyperparameter tuning and network structure selection, DJINN enabled us to obtain a high-performance model more efficiently. DJINN employed the Adam optimizer to minimize the mean squared error loss function (MSE). Each hidden layer utilized the rectified linear unit (ReLU) activation function, while linear activation was applied to the output layers.

Finally, $ANN_{EXP1\text{-}MC1}$ and $ANN_{MC1\text{-}OP}$ were integrated end-to-end to form the initial ANN (ANN_1) as illustrated in Figure 2. To independently assess its performance, we subjected ANN_1 to evaluation using a distinct test dataset ($E_{EXP1\text{-}OP1}$) that was not included in training or validation. The performance of ANN_1 was evaluated based on its ability to accurately predict both absorption (μ_a) and reduced scattering (μ_s') coefficients. Predicted absorption spectra were fitted with the known chromophore basis function to retrieve absorber concentration. Predicted reduced scattering coefficient was fitted with a power-law relationship of the form $\mu_s' = a\left({}^{\lambda}/_{\lambda_0}\right)^{-b}$, where a and b are fitting coefficients, λ corresponds to wavelength in nm, and λ_0 is a normalization wavelength. This ANN_1 served as the basis for the transfer learning technique discussed in the next section.

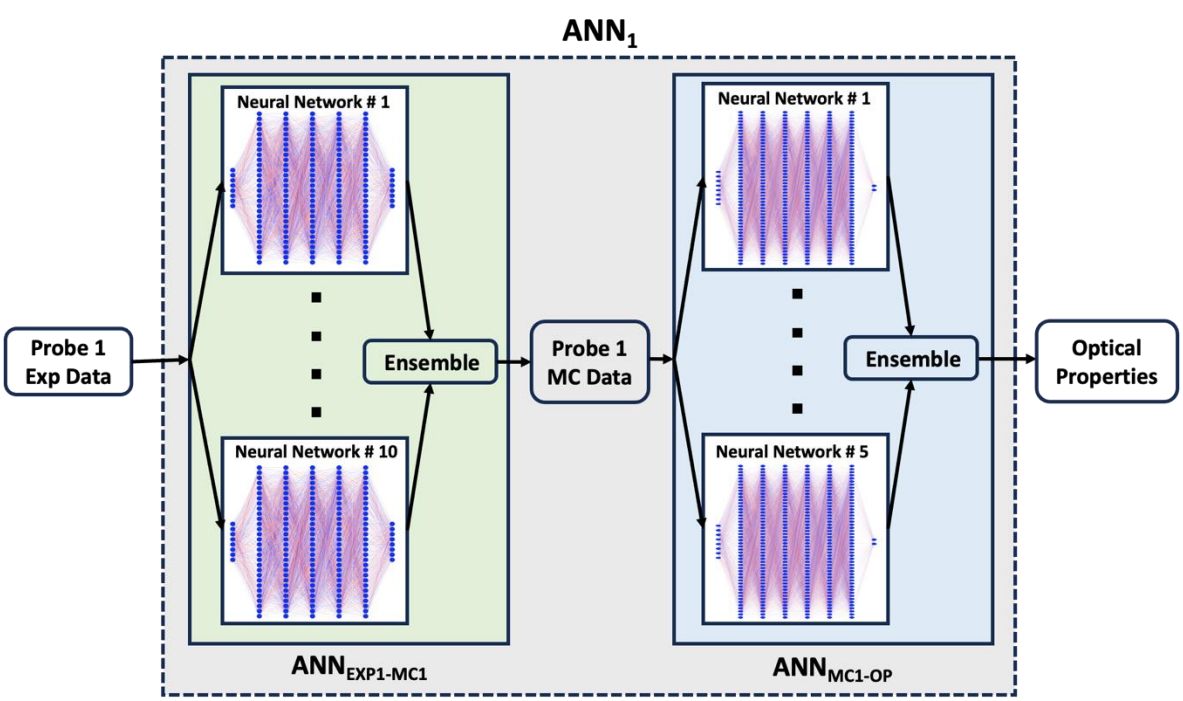

Figure 2: Initial ANN (ANN_1) for optical property extraction, which was created by combining $ANN_{EXP1\text{-}MC1}$ ensemble and $ANN_{MC1\text{-}OP}$ ensemble.

2.4 *Transfer Learning*

In this study, a feature extraction technique was used for transfer learning (TL). For this technique, a pre-trained model's intermediate layers are used as feature extractors to capture relevant patterns and representations from data, followed by training additional task-specific layers on top to adapt the model for a specific task. To achieve this, only the output layer or the top few layers of the pre-trained model are made trainable while all other layers remain frozen. By training these unfrozen layers using a small transfer learning dataset, the pre-trained model is adapted to the target task. This approach is most suitable when the target dataset is similar to the data the model was pre-trained on.

Here, the $ANN_{EXP1\text{-}MC1}$ network of ANN_1 was used as the pre-trained model. To accomplish probe-to-probe transfer learning, only the output layer of the $ANN_{EXP1\text{-}MC1}$ ensemble was modified. The output layer of each individual ANN of the $ANN_{EXP1\text{-}MC1}$ ensemble was retrained using a small TL dataset collected by the target probe to create an ANN_{TL} ensemble for the target probe as shown in Figure 3. The inputs for this TL dataset were experimental reflectance signals collected from selected TL phantoms with known optical properties described in section 2.5.1, and the outputs were simulated reflectance for probe-1 obtained from the probe 1 MC library (MC_1) described in section 2.5.2 for these known optical properties. Hence, the ANN_{TL} ensemble essentially learned the mapping from the target probe's experimental reflectance signal to the corresponding simulated reflectance signal for probe-1. This allowed us to reuse $ANN_{MC1\text{-}OP}$, which eliminated the need for time-consuming generation of the MC library and subsequent ANN creation for the target probe. ANN_{TL} was joined with $ANN_{MC1\text{-}OP}$ described earlier to create an $ANN_{target(probe)}$ that took the target probe's experimental reflectance intensities as inputs and outputs corresponding optical properties.

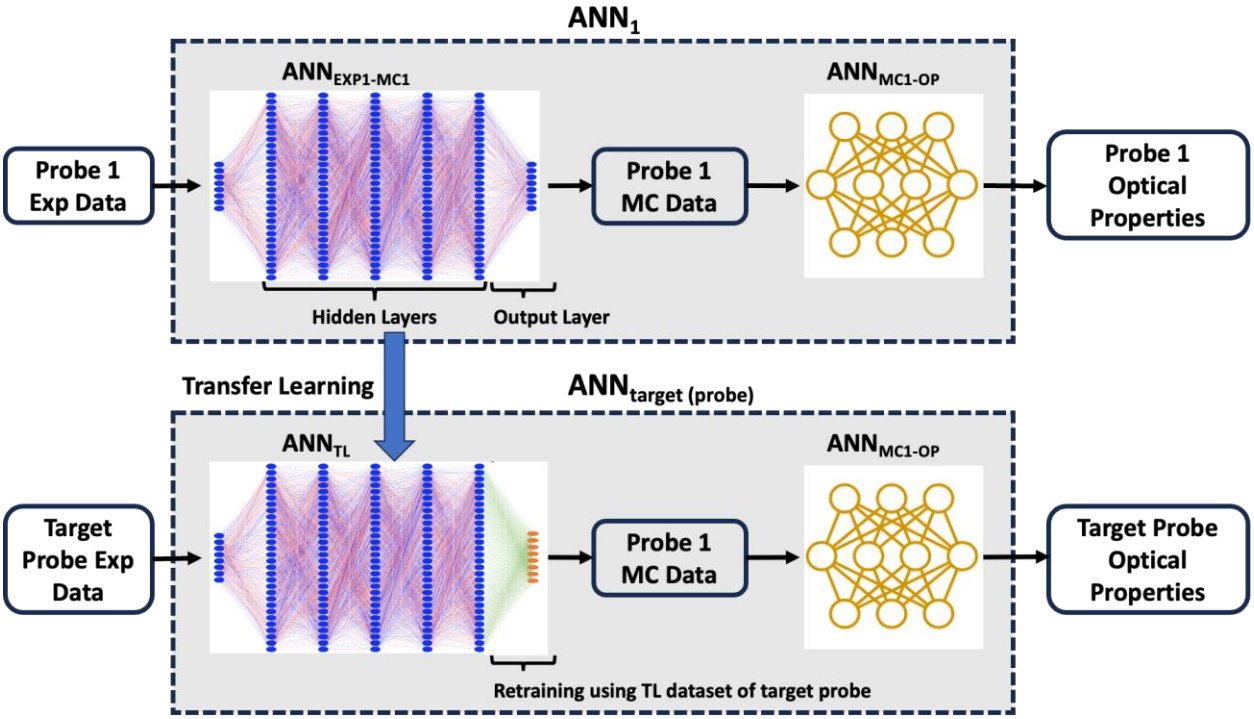

Figure 3: Transfer Learning to create $ANN_{target\ (probe)}$ from ANN_1 by retraining the output layer of $ANN_{EXP1\text{-}MC1}$. Although the TL algorithm was applied to all ANN of the $ANN_{EXP1\text{-}MC1}$ ensemble, only one ANN is shown here for clarity.

Successful transfer learning relies on an effective transfer learning strategy, which includes the size and diversity of the transfer learning dataset, and optimal selection of hyperparameters (e.g., learning rate, epochs, batch size etc.). In order to assess our TL approach, we selected the performance of ANN_1 on the test dataset ($E_{EXP1\text{-}OP1}$) as the benchmark. The effective validation of our TL model ANN_{target} was determined by its capability to attain performance levels on par with those of ANN_1. Root mean squared error (RMSE) and Mean absolute percent error (MAPE) were used as the error metrics.

In our clinical context, it is impractical to collect a large number of calibration spectra, even though having a larger number of calibration spectra with diverse conditions in the TL dataset could be beneficial. Hence, to understand the effect of TL dataset size and its diversity on the TL performance, our TL technique was applied to the pre-trained model ($ANN_{EXP1\text{-}MC1}$) with varying sized probe-2 TL datasets. Different subsets of the $TL_{EXP2\text{-}MC1}$ dataset were used to create different TL models and their performance were evaluated on a test dataset ($E_{EXP2\text{-}MC1}$). This enabled us to determine which spectra should be incorporated into the TL dataset to ensure its effectiveness. In total, the $TL_{EXP2\text{-}MC1}$ dataset consisted of 30 spectra. At first, we used the full $TL_{EXP2\text{-}MC1}$ dataset to create the TL model. Subsequently, we reduced the size of the $TL_{EXP2\text{-}MC1}$ dataset so that it only included phantoms with the lowest ($\mu_a = 0.05\ cm^{-1}$) and highest absorption ($\mu_a = 1\ cm^{-1}$) coefficient at 665 nm for each scattering condition, which resulted in 10 spectra. These spectra held particular importance for effective TL, as they spanned the outermost boundaries of the dataset. Subsequently, we used different smaller combinations of these 10 spectra for TL, down to a single spectrum.

During TL training, we randomly partitioned the TL dataset into two distinct sets: a training subset (80%) and a validation subset (20%). This validation dataset was used to prevent our TL model from overfitting to the TL dataset by using an early stopping callback (patience = 10)[15]. This low patience early stopping callback, coupled with a low learning rate (10^{-4}), prevented the TL model from experiencing "catastrophic forgetting"[17] which occurs when a network forgets previously learned information when adapting to new tasks, ensuring the preservation of knowledge learned from initial probe dataset. Although the maximum number of epochs was set to 500, the training process was terminated well before reaching this limit due to the early stopping callback.

The knowledge learned from this step allowed us to create a TL algorithm that included the required number and conditions of TL spectra, and optimal hyperparameter settings for successful probe-to-probe TL. To validate the generalizability and robustness of this TL algorithm, the exact algorithm was applied to a new probe (probe-3). The TL model was created using the $TL_{EXP3\text{-}MC1}$ dataset and performance of the resulting TL model was evaluated on an independent test dataset ($E_{EXP3\text{-}OP3}$).

2.5 *Datasets*

2.5.1 *Phantom Experiments and Associated Datasets*

Three stages of phantom data were collected using three different probes. In all phantoms, methylene blue (Akorn, Inc., Lake Forest, IL) was used as the absorber and Intralipid 20%

(Fresenius Kabi AG, Bad Homburg, Germany) was used as the scatterer. Appropriate amounts of MB and Intralipid were mixed with 300 mL of distilled water (DI) in a black spray-painted container (Rust-Oleum Flat Black, Vernon Hills, IL) to create the phantoms described in subsections 2.5.1.1, 2.5.1.2 and 2.5.1.3. Concentration ranges of each component were selected to encompass the optical property range pertinent to our clinical application[4, 5]. Separate MB and Intralipid stocks were used for data collection for each separate probe, with each stock characterized as described below.

In order to calculate the necessary volumes of Intralipid and MB to add to DI water to achieve the desired optical properties, absorption (μ_a) spectra of MB, scattering (μ_s) spectra of Intralipid, and scattering anisotropy coefficients (g) for Intralipid were required. The first two were measured using a commercial spectrophotometer (Cary 50 Varian, Santa Clara, CA). To determine the scattering anisotropy of each Intralipid stock, we created 30 phantoms with μ_s = 25, 37.5, 62.5, 75 and 125 cm^{-1} at 665 nm, each of which had μ_a = 0.05, 0.1, 0.25, 0.5, 0.75, and 1.0 cm^{-1} at 665 nm from each stock. These phantoms were distinct from those used for training and/or validation. An iterative fitting algorithm was used to determine *g* by minimizing the difference between experimental signal and corresponding simulated signal obtained from the MC library described in section 2.5.2. The calculated *g* values were similar to literature values[18, 19] and showed small inter-stock variation. This small variation was relevant as it considerably affected the quality of fitting of the algorithm. This was expected because of the relatively small SDS present in our probes, making them considerably more sensitive to variations in scattering anisotropy. Hence, it was important for us to determine the value of *g* for each Intralipid stock rather than relying on the fixed values from the literature.

Our experimental phantom data collection process is briefly described here, the details of which can be found in our earlier work[4]. We first allowed the source lamp to reach a steady state over approximately 30 minutes. Calibration measurements were taken using a 3-inch calibration sphere to account for fiber throughput and lamp power variation. The probe was then placed in contact with the liquid phantom while continuously stirring. Light was emitted into the phantom using the source fiber, and diffuse reflectance spectra were collected through each of the detector fibers. For both the calibration measurements and phantom measurements, corresponding dark measurements were taken and subtracted. Then, the dark corrected spectra were normalized by the corresponding dark corrected calibration data for each detection fiber to obtain the final DRS. Each spectrum consisted of 316 wavelengths (500-740 nm).

Sub-sections 2.5.1.1, 2.5.1.2, and 2.5.1.3 detail phantom datasets and their optical properties where TV, TL, and E denote training and validation, transfer learning, and testing datasets, respectively. Figure 4 visually represents these datasets with blue, green, and orange denoting training, transfer learning, and test sets, respectively. The subscript associated each dataset is divided into two parts, with the first part indicating the input and the second part indicating the output of each dataset. EXP, MC, and OP respectively denote experimental reflectance intensity, simulated reflectance intensity, and optical properties. The numerical value in the subscript specifies the probe or MC library utilized in generating the dataset.

2.5.1.1 *Probe-1 Dataset*

Phantom data collected using probe-1 were used to create the two following datasets (see Figure 4):

- $TV_{EXP1\text{-}MC1}$ consisted of phantoms with μ_s = 25, 37.5, 62.5, 75 and 125 cm^{-1} at 665 nm, each of which had μ_a = 0.05, 0.1, 0.25, 0.5, 0.75, and 1.0 cm^{-1} at 665 nm (30 phantoms)
- $E_{EXP1\text{-}OP1}$ included phantoms with μ_s = 50 and 100 cm^{-1} at 665 nm, each of which had μ_a = 0.05, 0.1, 0.25, 0.5, 0.75, and 1.0 cm^{-1} at 665 nm (12 phantoms)

2.5.1.2 *Probe-2 Dataset*

Phantom data collected using probe-2 were used to create the two following datasets (see Figure 4):

- $TL_{EXP2\text{-}MC1}$ consisted of phantoms with μ_s = 25, 37.5, 62.5, 75 and 125 cm^{-1} at 665 nm, each of which had μ_a = 0.05, 0.1, 0.25, 0.5, 0.75, and 1.0 cm^{-1} at 665 nm (30 phantoms)
- $E_{EXP2\text{-}MC1}$ and $E_{EXP2\text{-}OP2}$ included phantoms with μ_s = 50 and 100 cm^{-1} at 665 nm, each of which had μ_a = 0.05, 0.1, 0.25, 0.5, 0.75, and 1.0 cm^{-1} at 665 nm (12 phantoms)

2.5.1.3 *Probe-3 Dataset*

Phantom data were collected using probe-3 to create the following two datasets (see Figure 4):

- $TL_{EXP3\text{-}MC1}$ comprised of the phantoms which had μ_s = 25 cm^{-1} and μ_a = 1 cm^{-1}, μ_s = 62.5 cm^{-1} and μ_a = 0.05 cm^{-1}, and μ_s = 125 cm^{-1} and μ_a = 1 cm^{-1}. All coefficients are given for 665 nm. (3 phantoms)
- $E_{EXP3\text{-}OP3}$ consisted of phantoms with μ_s = 50, 75 and 100 cm^{-1} at 665 nm, each of which had μ_a = 0.05, 0.1, 0.25, 0.5, 0.75, and 1.0 cm^{-1} at 665 nm (18 phantoms)

<table>
<tr><td>Scattering at 665 nm (μ_s) [cm^{-1}]</td><td colspan="3">25</td><td colspan="3">37.5</td><td colspan="3">50</td><td colspan="3">62.5</td><td colspan="3">75</td><td colspan="3">100</td><td colspan="3">125</td></tr>
<tr><td>Absorption at 665 nm (μ_a) [cm^{-1}]</td><td>0.05</td><td>…</td><td>1.00</td><td>0.05</td><td>…</td><td>1.00</td><td>0.05</td><td>…</td><td>1.00</td><td>0.05</td><td>…</td><td>1.00</td><td>0.05</td><td>…</td><td>1.00</td><td>0.05</td><td>…</td><td>1.00</td><td>0.05</td><td>…</td><td>1.00</td></tr>
<tr><td>Probe 1</td><td colspan="6">$TV_{EXP1\text{-}MC1}$</td><td colspan="3">$E_{EXP1\text{-}OP1}$</td><td colspan="6">$TV_{EXP1\text{-}MC1}$</td><td colspan="3">$E_{EXP1\text{-}OP1}$</td><td colspan="6">$TV_{EXP1\text{-}MC1}$</td></tr>
<tr><td>Probe 2</td><td colspan="6">$TL_{EXP2\text{-}MC1}$</td><td colspan="3">$E_{EXP2\text{-}OP2}$</td><td colspan="6">$TL_{EXP2\text{-}MC1}$</td><td colspan="3">$E_{EXP2\text{-}OP2}$</td><td colspan="6">$TL_{EXP2\text{-}MC1}$</td></tr>
<tr><td>Probe 3</td><td></td><td colspan="2">$TL_{EXP3\text{-}MC1}$</td><td colspan="3"></td><td colspan="3">$E_{EXP3\text{-}OP3}$</td><td colspan="2">$TL_{EXP3\text{-}MC1}$</td><td colspan="4"></td><td colspan="6">$E_{EXP3\text{-}OP3}$</td><td colspan="3"></td><td colspan="2">$TL_{EXP3\text{-}MC1}$</td></tr>
</table>

Figure 4: Data for different probes and their corresponding optical property ranges. The blue, green, and orange colors correspond to training, transfer learning, and test datasets, respectively.

2.5.2 *Monte Carlo Dataset*

A Monte Carlo (MC) model for probe-1 was constructed for optical property retrieval from diffuse reflectance spectra. A detailed account is of this is available in our previous publication[4]. In summary, we generated the MC library (MC_1) by conducting graphics processing unit (GPU)

accelerated MC simulations across a wide range of parameter combinations, encompassing values for absorption coefficient (μ_a) from 0.0001 to 25 cm^{-1} and scattering coefficient (μ_s) from 1 to 250 cm^{-1}. These simulations were carried out using a refractive index of n=1.37 and employed the Henyey-Greenstein phase function with a scattering anisotropy factor (g) of 0.7, resulting in reduced scattering coefficients (μ_s') from 0.3 to 75 cm^{-1} for the MC library. The simulations were executed using 10^8 photon packets. The computations were performed on a Quadro RTX6000 GPU equipped with 24 GB of GPU memory (NVIDIA Corporation, Santa Clara, CA), which took about 7 days to run all optical property combinations. This MC library was used to create a $TV_{MC1-OP1}$ dataset where the inputs were simulated reflectance intensity at 8 fibers and outputs were corresponding absorption coefficients (μ_a) and reduced scattering coefficients (μ_s').

2.6 *Statistical Analysis*

Measured values are summarized as mean ± standard deviation. Paired comparisons were performed using the Wilcoxon matched-pairs signed rank test. Comparisons between the number of spectra in the transfer learning dataset used the Kruskal-Wallis test, with Dunn's test for multiple comparisons. GraphPad Prism (v10, GraphPad Software, Inc., Boston, MA) was used for all statistical analysis.

3 Results

3.1 *Performance of Initial Neural Network (ANN_1)*

3.1.1 *Training and Performance of $ANN_{EXP1-MC1}$ Ensemble*

Figure 5 demonstrates the progression of both training and validation losses (MSE) against the number of epochs for a representative ANN (ANN #1) selected from the $ANN_{EXP1-MC1}$ ensemble. The plot illustrates the convergence of the ANN as training proceeded, with training and validation losses closely aligning. Early stopping callback halted the training at 544 epochs.

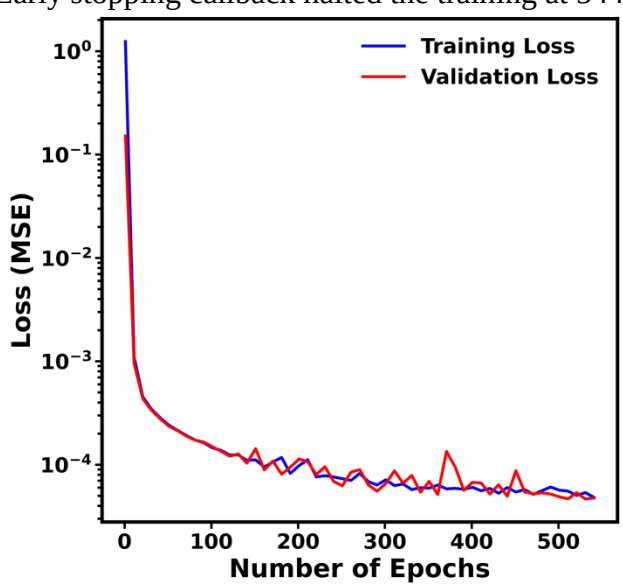

Figure 5: Loss (MSE) vs epochs for ANN #1 of the $ANN_{EXP1-MC1}$ ensemble. Blue and red colors correspond to training and validation loss, respectively.

The performance of the $ANN_{EXP1\text{-}MC1}$ ensemble was evaluated on the entire $TV_{EXP1\text{-}MC1}$ dataset to ensure it properly captured the relationship between experimental and simulated reflectance signals across all detector fibers. Figure 6 demonstrates that the ensemble accurately learned the relationship between input and outputs. The RMSE and MAPE values were small across all fibers, as detailed in Table 2. The ensemble's performance was evaluated on the entire $TV_{EXP1\text{-}MC1}$ dataset.

Table 2: RMSE and MAPE for predicted simulated reflectance signal across 8 detector fibers

Metric	Detector Fiber 1	Detector Fiber 2	Detector Fiber 3	Detector Fiber 4	Detector Fiber 5	Detector Fiber 6	Detector Fiber 7	Detector Fiber 8
RMSE	5.34 $x10^{11}$	2.18 $x10^{11}$	1.80 $x10^{11}$	1.99 $x10^{11}$	1.36 $x10^{11}$	1.45 $x10^{11}$	1.59 $x10^{11}$	1.05 $x10^{11}$
MAPE	0.59%	0.536%	0.55%	0.55%	0.62%	0.59%	0.57%	0.85%

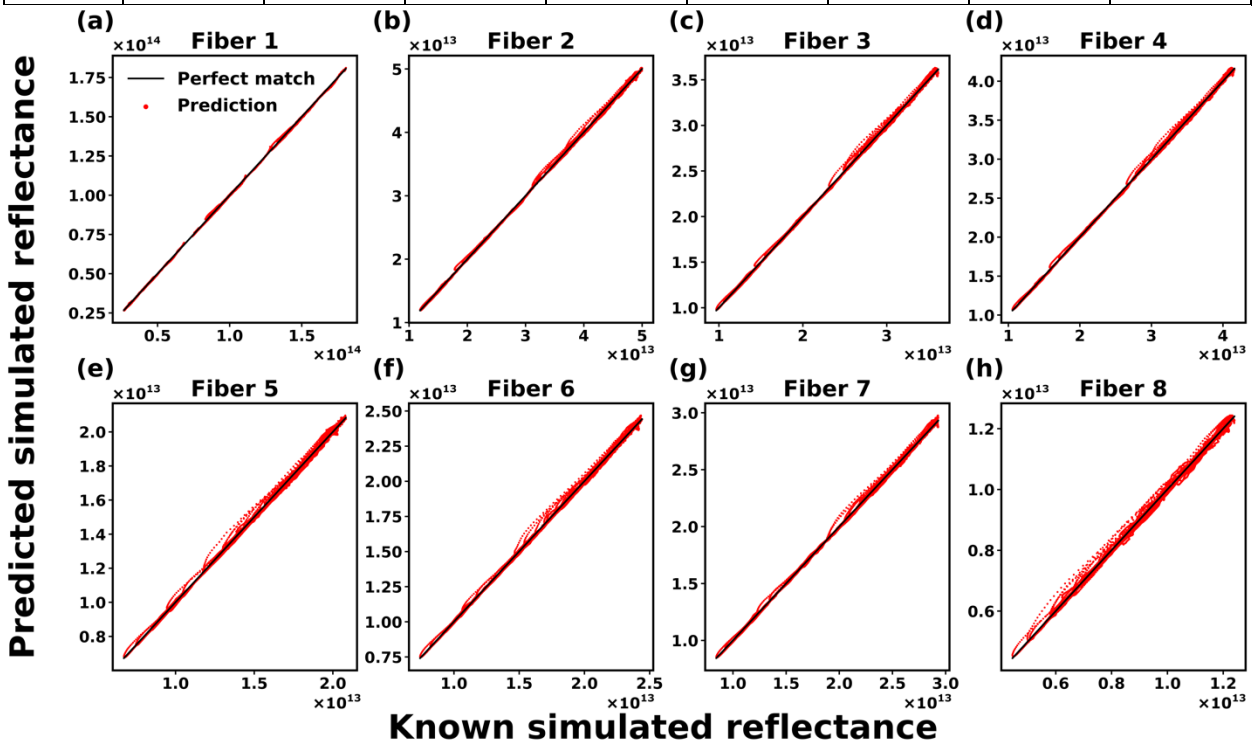

Figure 6: Performance of $ANN_{EXP1\text{-}MC1}$ for all eight detector fibers are shown in sub-figures (a) to (h), respectively. Red dots indicate predictions and solid black lines denote perfect prediction. The units for simulated reflectance signals are detected photon weight.

3.1.2 *Training and Performance of $ANN_{MC1\text{-}OP}$ Ensemble*

Figure 7 shows the training loss (MSE) vs number of epochs for the $ANN_{MC1\text{-}OP}$ ensemble, demonstrating training convergence.

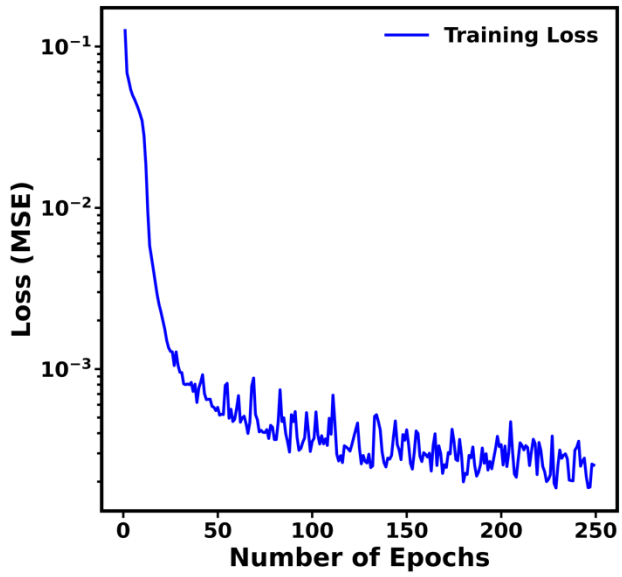

Figure 7: Loss (MSE) vs epoch for of $ANN_{MC1\text{-}OP}$ ensemble. Blue color corresponds to training loss.

The performance of the $ANN_{MC1\text{-}OP}$ ensemble on its test set is depicted in Figure 8. Absorption coefficient (μ_a) was predicted with a small RMSE = 0.087 cm^{-1} (MAPE = 290%). This elevated MAPE can be attributed to discrepancies arising from very small absorption coefficients ($\mu_a < 0.01$ cm^{-1}). For $\mu_a > 0.01$ cm^{-1}, RMSE was 0.091 cm^{-1} (MAPE = 6.4%). Reduced scattering coefficient (μ_s') was predicted with a small RMSE = 0.48 cm^{-1} (MAPE = 1.1%). These low RMSE values indicate that $ANN_{MC1\text{-}OP}$ successfully learned the relationship between simulated reflectance intensity and corresponding optical properties.

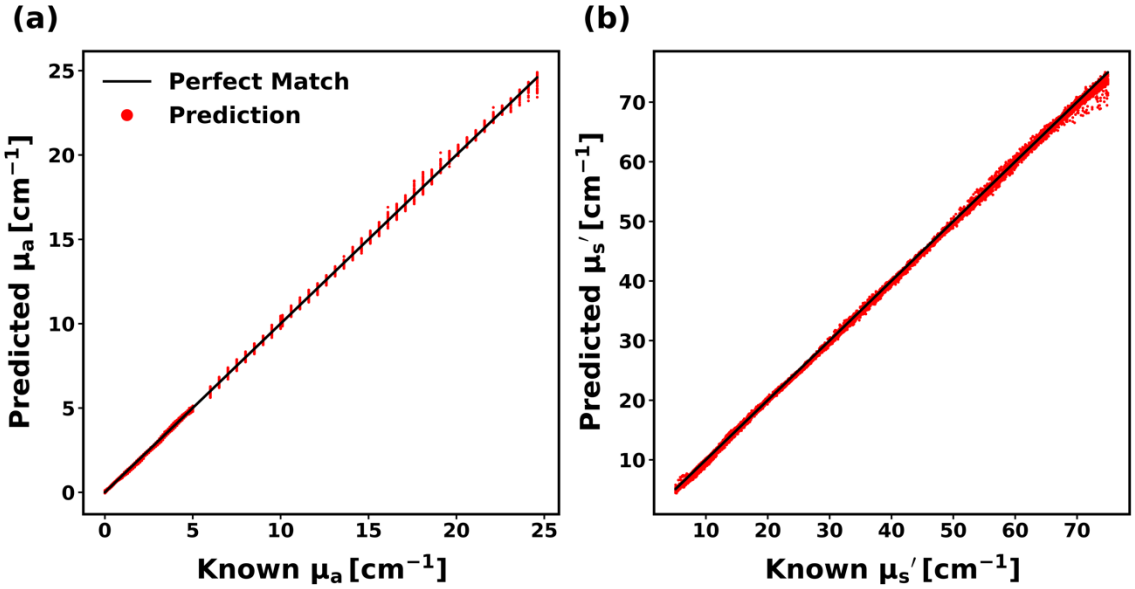

Figure 8: Predicted (a) absorption coefficients (μ_a) and (b) reduced scattering coefficients (μ_s'). Red dots represent predictions and solid black lines denote perfect prediction.

3.1.3 *Performance of the Initial ANN (ANN_1)*

The performance of the initial ANN (ANN_1) on the test set ($E_{EXP1\text{-}OP1}$) is presented in Figure 9. The absorption and reduced scattering spectra predictions from ANN_1 for a representative phantom (with μ_s = 100 cm^{-1} at 665 nm and μ_a values of 0.05, 0.1, 0.25, 0.5, 0.75, and 1.0 cm^{-1} at 665 nm) from the test dataset are illustrated in Figures 9(a) and (b). The predicted absorption and reduced scattering spectra were fitted using the procedure mentioned in section 2.3.1.2, resulting in highly accurate fitting.

Figure 9(c) illustrates that ANN_1 accurately predicted the MB concentration of the phantoms in the test dataset with a small RMSE = 0.29 μM (MAPE = 7.46%). It also performed well in predicting the reduced scattering coefficient (μ_s') across all wavelengths, as evidenced by a low RMSE = 1.08 cm^{-1} (MAPE = 3.6%). The reduced scattering coefficient at 665 nm ($\mu_{s,665}'$) was predicted with RMSE = 0.77 cm^{-1} (MAPE = 2.54%) which is shown in Figure 9(d).

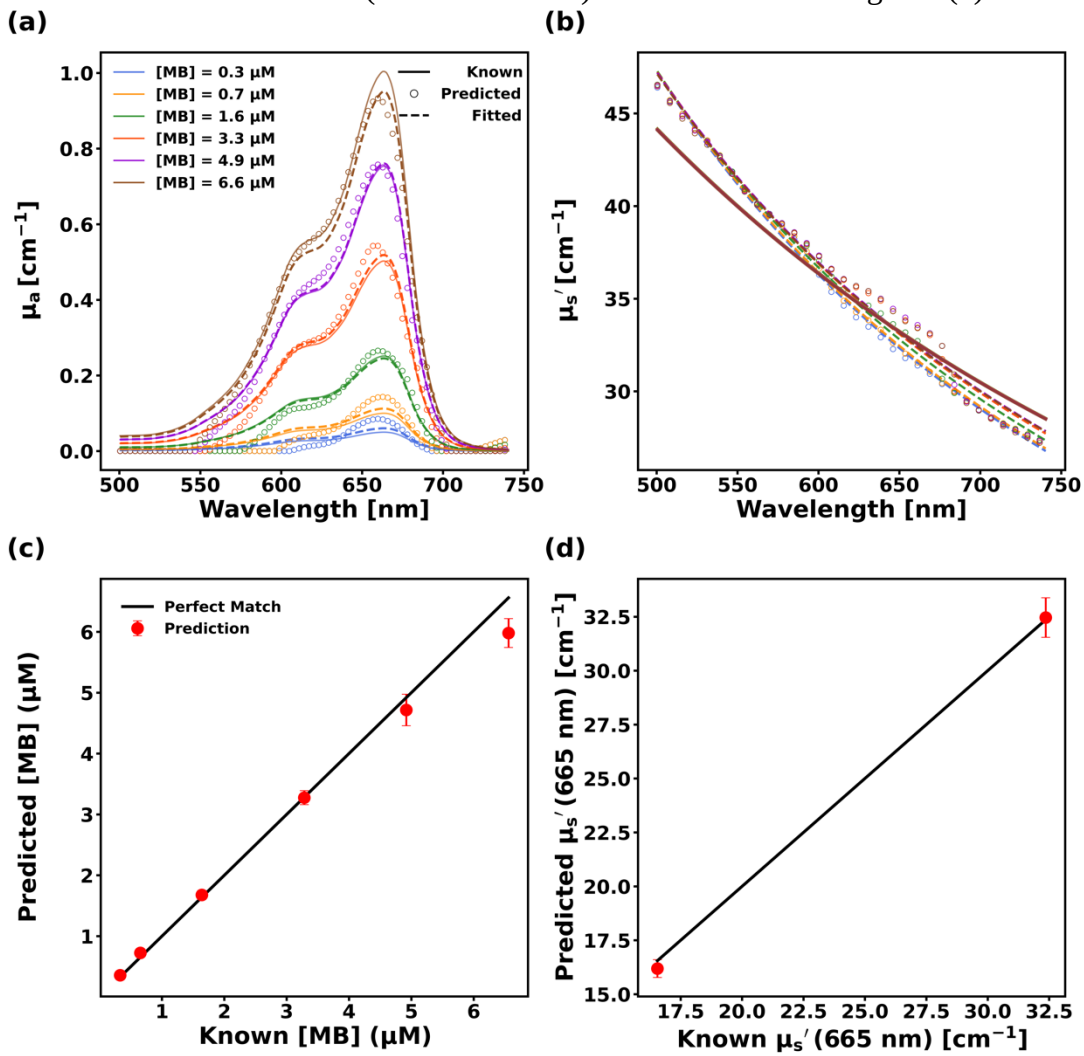

Figure 9: Predicted (a) absorption spectra and (b) reduced scattering spectra for a representative phantom. Solid lines, open circles, and dashed lines correspond to known, predicted, and fitted spectra respectively, while colors indicate MB concentration. Predicted (c) MB concentration and (d) μ_s' at 665 nm across phantoms. Predictions are denoted by red dots and solid black lines denote perfect prediction. Symbols indicate mean recovered values across all phantom measurements at that value, while error bars correspond to standard deviation across phantoms with identical optical properties. Error bars are included in all cases but are not always visible.

3.2 *Transfer Learning*

3.2.1 *Performance of Initial ANN (ANN_1) on Probe-2 data without Transfer Learning*

To access the performance of ANN_1 for data collected with another probe, we evaluated its performance on the $E_{EXP2\text{-}OP2}$ dataset. Figures 10(a) and (b) illustrate the absorption and reduced scattering spectra predictions of ANN_1 for a representative phantom from $E_{EXP2\text{-}OP2}$. ANN_1 predicted much higher absorption than known absorption for this phantom as evident from Figure 10(a). The MB concentration and fitted reduced scattering spectra were obtained by the same methods described in sub-section 3.1.3.

The failure of ANN_1 to accurately predict MB concentration from probe-2 data can be seen in Figure 10(c), which shows the performance of ANN_1 for the entire $E_{EXP2\text{-}OP2}$ dataset. The MB concentration of phantoms was recovered with a large RMSE = 1.67 μM (MAPE = 154.84%). It also performed worse in predicting the reduced scattering coefficient (μ_s') across all wavelengths as evidenced by comparably higher RMSE = 2.2 cm^{-1} (MAPE = 4.85%). The reduced scattering coefficient at 665 nm ($\mu_{s,665}'$) was predicted with RMSE = 1.76 cm^{-1} (MAPE = 5.14%), as shown in Figure 10(d). These results indicate that employing ANN_1, which was trained using probe-1 data, for optical property extraction from data collected by probe-2 results in large errors in predictions, particularly for absorption.

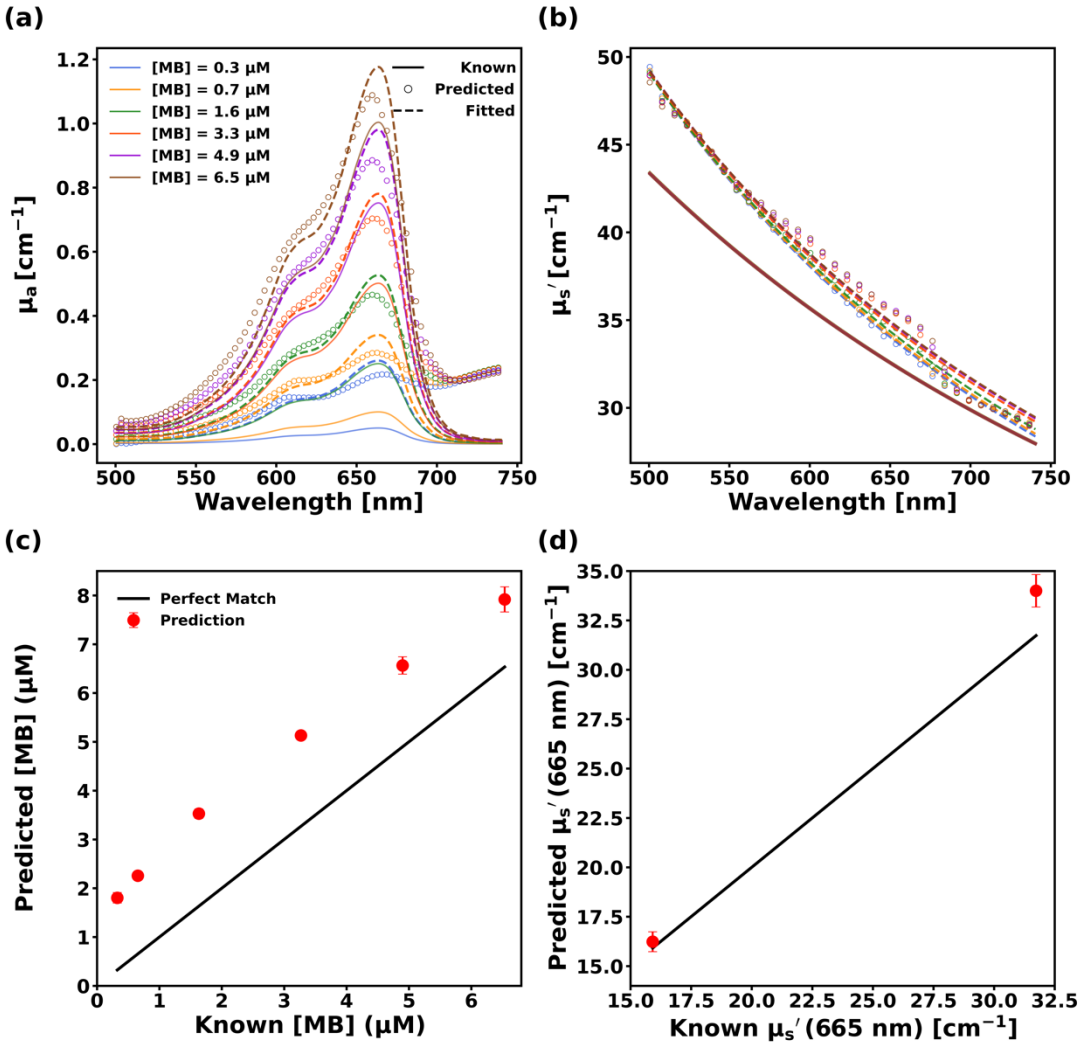

Figure 10: Predicted (a) absorption spectra and (b) reduced scattering spectra for a representative phantom measured with probe-2 using the neural network trained on probe-1 data (ANN_1) show poor quality extraction of optical properties. Solid lines, open circles, and dashed lines correspond to known, predicted, and fitted spectra respectively while colors indicate MB concentration. Predicted (c) MB concentration and (d) μ_s' at 665 nm across phantoms also show poor performance, where predictions are denoted by red dots and solid black lines denote perfect prediction. Error bars are included in all cases but are not always visible.

3.2.2 *Feature Extraction*

From the results of sub-section 3.2.1, it is evident that usage of the initial ANN (ANN_1) for probe-2 data leads to highly inaccurate prediction of optical properties. To adapt ANN_1 to probe-2, we employed a feature extraction TL technique, the results of which are provided in sub-sections 3.2.2.1, 3.2.2.2, and 3.2.2.3.

3.2.2.1 *Number of Training Spectra vs Performance*

Figure 11 presents the prediction error of different TL models created from different-sized TL training datasets on the $E_{EXP2\text{-}MC1}$ dataset. The prediction errors of TL training datasets with the same number of spectra but varying known optical properties were used to find the mean and standard deviation of error for that particular number of spectra. Figure 11 elucidates that using 10 spectra for TL, which includes spectra with the lowest (μ_a = 0.05 cm^{-1}) and highest absorption (μ_a = 1 cm^{-1}) coefficient at 665 nm for each scattering condition, leads to identical error compared to that of using all 30 spectra in the $TL_{EXP2\text{-}MC1}$ dataset (0.021 vs. 0.021). We found that as we incorporate smaller and smaller subsets of these 10 spectra in the TL dataset (from 5 spectra to 1 spectrum), the performance of the resultant TL model worsens, indicated by the higher MSE shown in Figure 11. The usage of a smaller number of spectra also leads to a higher variance in the error compared to the usage of a higher number of spectra.

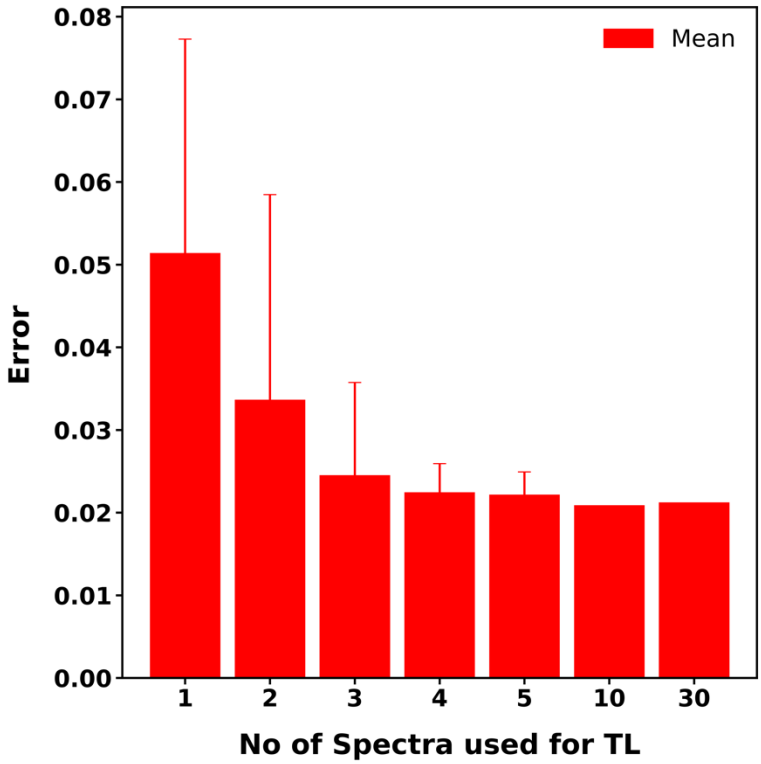

Figure 11: Prediction error for ANN_{TL} models created from TL datasets with varying number of training spectra. The X-axis denotes the number of spectra in the TL dataset, and the Y-axis denotes the prediction error. Solid bars represent mean values across all combinations of the specified number of spectra, with error bars corresponding to standard deviation.

3.2.2.2 *Selection of Spectra and Creation of TL Algorithm*

We chose to utilize 3 training spectra in our TL algorithm, as we can see that utilizing 3 spectra for TL results in comparable error to utilizing all 30 spectra (0.024 vs. 0.021). Utilization of only 2 training spectra resulted in an increase in error (0.034 vs. 0.024). To select the optimal spectra to be incorporated into the TL dataset, we considered the strengths and weaknesses of ANNs. ANNs work accurately and reliably only within the range of the data they were trained on. Beyond

this range, performance can degrade quite rapidly[20]. Hence, to maximize the range of validity of our TL model, we decided to integrate two spectra from the extremes of the TL dataset ($TL_{EXP2-MC1}$). Specifically, we included the spectrum with the lowest scattering and highest absorption (μ_s = 25 cm^{-1} and μ_a = 1 cm^{-1} at 665 nm), as well as the spectrum with the highest scattering and absorption (μ_s = 125 cm^{-1} and μ_a = 1 cm^{-1} at 665 nm). Moreover, we also included an intermediate spectrum (μ_s = 62.5 cm^{-1} and μ_a = 0.05 cm^{-1} at 665 nm) to improve interpolation. In this step, we also determined the optimal hyper-parameters for TL (e.g., learning rate, epochs, etc.). Based on these, a TL algorithm was created for probe-to-probe TL which will allow us to create a $ANN_{target(probe)}$ for the target probe optical property recovery.

The algorithm was:

1) Collection of 3 spectra from 3 known phantoms using the target probe to create the TL dataset. The phantoms are:
    a. μ_s = 25 cm^{-1} and μ_a = 1 cm^{-1} at 665 nm
    b. μ_s = 62.5 cm^{-1} and μ_a = 0.05 cm^{-1} at 665 nm
    c. μ_s = 125 cm^{-1} and μ_a = 1 cm^{-1} at 665 nm
2) Usage of the TL dataset to adapt $ANN_{EXP1-MC1}$ to create ANN_{TL} adapted for target probe
3) Integration of ANN_{TL} with ANN_{MC1-OP} to create $ANN_{target(probe)}$ which is capable of recovering optical properties from DRS collected by the target probe.

3.2.2.3 *Performance of $ANN_{target(2)}$ on Probe-2 Data*

This algorithm was applied to ANN_1 to create $ANN_{target(2)}$ for probe-2 and its performance was evaluated on $E_{EXP2-OP2}$. Figures 12(a) and (b) illustrate the absorption and reduced scattering spectra predictions of $ANN_{target(2)}$ for a representative phantom from $E_{EXP2-OP2}$. We observe much better agreement between the known and predicted spectra for both absorption and reduced scattering compared to the incorrect insertion of probe-2 data into ANN_1.

The ability of $ANN_{target(2)}$ to accurately predict MB concentration can be seen in Figure 12(c), which shows the performance of ANN_{target} for the entire $E_{EXP2-OP2}$ dataset. The MB concentration of phantoms was recovered with a small RMSE = 0.38 μM (MAPE = 24.82%). It also showed superior performance in predicting the reduced scattering coefficient (μ_s') across all wavelengths as evidenced by low RMSE = 0.88 cm^{-1} (MAPE = 3.68%). The reduced scattering coefficient at 665 nm ($\mu_{s,665}'$) was predicted with RMSE = 0.71 cm^{-1} (MAPE = 3.01%) which is shown in Figure 12(d).

Overall, $ANN_{target(2)}$ exhibited significantly improved performance on the $E_{EXP2-OP2}$ dataset in comparison to ANN_1 both in predicting MB concentration ($p = 0.0005$) and reduced scattering coefficient at 665 nm ($p = 0.0005$).

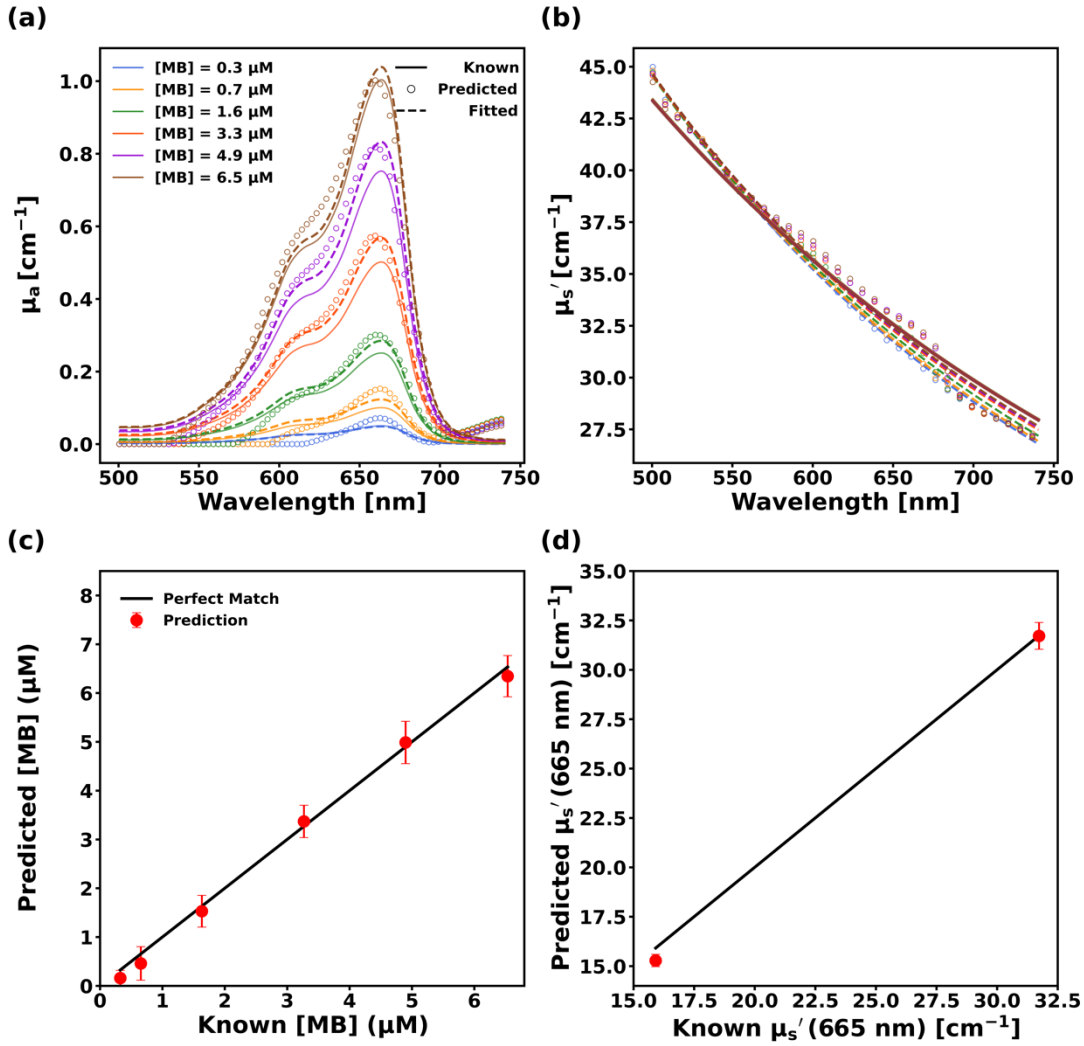

Figure 12: Predicted (a) absorption spectra and (b) reduced scattering spectra for a representative phantom measured with probe-2 after transfer learning. Solid lines, open circles, and dashed lines correspond to known, predicted, and fitted spectra respectively while colors indicate MB concentration. Predicted (c) MB concentration and (d) μ_s' at 665 nm across phantoms, where predictions are denoted by red dots and solid black lines denote perfect prediction. Error bars are included in all cases but are not always visible.

3.2.3 *Application of TL Algorithm to Probe-3*

The performance of ANN_1 on probe-3 data ($E_{EXP3\text{-}OP3}$) is shown in Figures 13(a) and (b) as the pre-TL prediction. The MB concentration prediction error was huge, with a RMSE = 4.1 μM (MAPE = 361.08%). For phantoms with $\mu_a > 0.1$ cm^{-1}, MB concentration was predicted with RMSE = 4.06 μM (MAPE = 124.51%). Prediction of the reduced scattering coefficient (μ_s') across all wavelengths was also poor, with RMSE = 2.47 cm^{-1} (MAPE = 5.68%). The reduced scattering coefficient at 665 nm ($\mu_{s,665}'$) was predicted with an RMSE = 2.08 cm^{-1} (MAPE = 5.45%) as shown in Figure 13(b).

The TL algorithm described in Section 3.2.2.2 was applied to ANN_1 to create $ANN_{target(3)}$ for probe-3 and its performance was also evaluated on $E_{EXP3\text{-}OP3}$, which are shown in Figures 13(a) and (b) as the post-TL prediction. The MB concentration of phantoms was recovered with a RMSE

= 0.73 μM (MAPE = 61.89%), as depicted in Figure 13 (a). This high MAPE can be attributed to mismatch in prediction of low-absorption phantoms for which slight mismatch in prediction causes huge MAPE values. For phantoms with $\mu_a > 0.1$ cm^{-1}, MB concentration was predicted with RMSE = 0.72 μM (MAPE = 22.8%). It was also able to predict reduced scattering coefficient (μ_s') across all wavelengths with a low RMSE = 1.67 cm^{-1} (MAPE = 4.11%). The reduced scattering coefficient at 665 nm ($\mu_{s,665}'$) was predicted with RMSE = 1.68 cm^{-1} (MAPE = 5.43%) which is shown in Figure 13(b).

$ANN_{target(3)}$ exhibited significantly improved performance on the $E_{EXP3\text{-}OP3}$ dataset in comparison to ANN_1 for MB concentration prediction ($p < 0.0001$), while prediction of the reduced scattering coefficient at 665 nm was not significantly improved ($p = 0.2$).

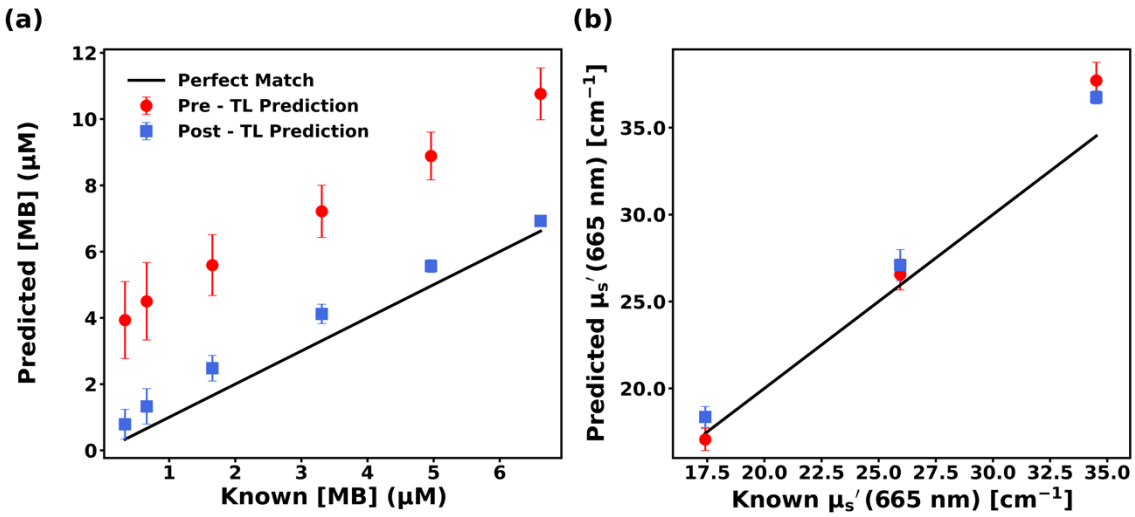

Figure 13: Predicted (a) MB concentration and (b) μ_s' at 665 nm for phantoms measured with probe-3 for both before (pre-TL) and after (post-TL) the application of transfer learning. Pre-TL and post-TL predictions are denoted by red dots and blue squares, respectively, while solid black lines denote perfect prediction. Error bars are included in all cases but are not always visible.

4 Discussion

In this study, we created and validated an ANN that can accurately predict optical properties from DRS collected by a specific fiber optic probe. Then, we investigated utilization of this ANN for optical property prediction from DRS collected by other probes having dissimilar SDS. We found that doing so led to large errors in prediction. To solve this issue, we utilized a feature-extraction based transfer learning algorithm. We showed that this algorithm can reliably adapt an ANN created for one probe to a target probe for accurate optical property recovery by the target probe. This probe-to-probe calibration can be achieved by using a small transfer learning dataset consisting of only three spectra collected by the target probe.

Other groups have used a similar ANN approach for optical property recovery from DRS. Ivančič *et al* predicted μ_a with RMSE = 0.09 cm^{-1} (relative RMSE = 2.6%) and μ_s' with RMSE = 0.26 cm^{-1} (relative RMSE = 1.6%) at SDS ranging from 220-1200 μm[21]. However, their performance evaluation was limited to a single low scattering and high absorption phantom. Furthermore, they employed separate ANNs for predicting each optical parameter which resulted in four distinct ANNs, whereas we used a single ANN for predicting both absorption and reduced scattering

spectra. Chong *et al* demonstrated recovery of μ_a and μ_s' with mean relative error ranging from 1.62-5.92% and 0.8-1.6%, respectively, with their baseline neural network and its physics-guided variants, but only for a simulation dataset[20]. An *et al* developed ANNs for their spectroscopy system, which predicted μ_a with mean absolute error ranging from 0.13-0.17 cm^{-1} and μ_s' ranging from 1.43-1.71 cm^{-1} [22]. However, their ANN prediction was limited to only four wavelengths compared to our broad-spectrum prediction ranging from 500-740 nm. Lan *et al* predicted μ_a with Euclidean distance = 0.25 and μ_s' with Euclidean distance = 4.14, which is somewhat difficult to interpret[14]. With our initial ANN (ANN_1), we achieved comparable performance to these previous studies. However, these prior publications only investigated the performance of ANN-based optical property recovery for a single instance of their respective fiber optic probes. The effects of SDS mismatch between probes on the recovered optical properties was not previously explored. This is vitally important for applications requiring multiple copies of a designed fiber optic probe, or single use devices.

We have shown that ANNs for optical property extraction are probe-specific. In other words, an ANN created for one probe can't be used for another probe if their SDS are considerably different. This is particularly true for the short SDS required by our clinical fiber optic probe. Precise knowledge of SDS is therefore required for traditional methods, such as Monte Carlo lookup tables[4, 11] or analytical approximations[10], as well as generation of Monte Carlo models for ANNs trained on simulation data. Hence, each time we intend to extract optical properties using a new probe, we are required to perform time-consuming probe characterization, probe-specific MC library creation, and phantom data collection, as well as ANN creation and validation to ensure satisfactory performance with this new probe. These steps require a substantial amount of effort, technical expertise, and time (7-14 days). These factors can severely limit the clinical implementation of an ANN-based spectroscopy system for our envisioned multi-center clinical trials or to transform this system into a widely-used product on a large scale[23].

As we have demonstrated, our transfer learning algorithm offers a possible solution to these problems. It eliminates the need for performing the aforementioned lengthy and time-consuming probe-specific MC library and ANN creation for each new probe, reducing the time and expertise required. Even without code optimization, our probe-to-probe calibration can be done rapidly within 10 minutes, so it can be easily integrated into a clinical workflow. We envision that this would utilize well-characterized solid phantoms for transfer learning, which have been shown to be highly reproducible[24]. Furthermore, this transfer learning algorithm could potentially reduce the cost of probe manufacturing. As we have shown that our algorithm allows for accurate recovery of data from probes with relatively large differences in SDS from those specified, this could allow for the utilization of less expensive probes with looser tolerances on fiber positioning. This could be extended to the utilization of sterile disposable probes for straightforward integration into the clinical workflow.

We observed significantly better prediction performance after the application of the transfer learning algorithm for both probes 2 and 3. For probe-2, errors in both MB concentration and scattering coefficient prediction were reduced up to 77.2% and 60%, respectively. For probe 3, errors in both MB concentration and scattering coefficient prediction were reduced up to 82.2% and 32.4%, respectively. These findings indicate that the transfer learning algorithm enhanced the

prediction accuracy of both probes 2 and 3, demonstrating the generalizability of our proposed TL algorithm.

We have shown that the magnitude of prediction error depends on the SDS mismatch between the initial probe and the target probe. This was apparent when data from probe-2 or 3 were used as inputs to the initial ANN (ANN_1) without transfer learning. For probe-2 data, the ANN_1 predicted MB concentration with RMSE = 1.67 μM (MAPE = 154.84%) whereas for probe 3 data, it predicted MB concentration with RMSE = 4.1 μM (MAPE = 361.08%) which is 2.45 times higher. Between probes 1 and 2, the SDS mismatch had RMSE = 39.69 μm (MAPE = 3.46%) whereas between probes 1 and 3, the SDS mismatch had RMSE = 100.91 μm (MAPE = 10.94%) which is 2.54 times higher. It is evident that a higher mismatch in SDS leads to higher prediction errors. Due to this greater mismatch between fiber positioning between probes 1 and 3 compared to that of probes 1 and 2, even after the application of transfer learning the MB concentration recovery had somewhat higher RMSE = 0.73 μM (MAPE = 61.89%) for probe 3. This implies that there may be a maximum SDS difference that can be tolerated by this transfer learning approach. Unfortunately, we do not have a sufficient selection of probes to test this directly, but this will be an area of future study.

Although our proposed transfer learning algorithm has great potential in facilitating multi-center trials and ultimately widespread adaption of probe-based optical property recovery using ANN, it is not without limitations. We employed only three probes in this study which prevents us from creating a general guideline involving tolerance in fiber positioning and its effect on optical property prediction. The differences among the probes were substantial but not overwhelmingly large. In future research, we are planning to investigate these issues more thoroughly by employing a larger number of probes with varying SDS. One weakness of the ANN approach is the inability to perform accurately and reliably outside the training range[25]. Although the ANNs used in this study were trained on datasets that covered the range pertinent to our clinical application, there is a possibility that these ANNs will encounter out-of-domain data in a clinical scenario. We plan to utilize physics-guided neural networks for optical recovery in our future endeavors as these networks have shown superior out-of-domain performance for different tasks[20]. The reliability of our TL algorithm may be compromised by uncertainties in the knowledge of the optical properties for the phantoms used for transfer learning. This could be mitigated by the usage of solid phantoms with identical optical properties, including regular validation of these phantoms with another well-validated spectroscopy system.

Disclosures

The authors have no conflicts of interest to disclose.

Code, Data, and Materials

Data are available upon reasonable request from the authors.

Acknowledgments

This work was funded by grant EB029921 from the National Institutes of Health.

References:

1. Hamblin, M.R. and T. Hasan, *Photodynamic therapy: A new antimicrobial approach to infectious disease?* Photochem Photobiol Sci, 2004. **3**: p. 436-450.
2. Baran, T.M., et al., *Photodynamic therapy is a safe and feasible adjunct to percutaneous drainage of deep tissue abscesses: Results of a first in humans Phase 1 clinical trial.* medRxiv, 2023. **2023.10.16.23297086**.
3. Haidaris, C.G., et al., *Effective photodynamic therapy against microbial populations in human deep tissue abscess aspirates.* Lasers Surg Med, 2013. **45**: p. 509-516.
4. Bridger, K.G., J.R. Roccabruna, and T.M. Baran, *Optical property recovery with spatially-resolved diffuse reflectance at short source-detector separations using a compact fiber-optic probe.* Biomed Opt Express, 2021. **12**(12): p. 7388-7404.
5. Hannan, M.N., A.K. Sharma, and T.M. Baran, *First in human measurements of abscess cavity optical properties and methylene blue uptake prior to photodynamic therapy by in vivo diffuse reflectance spectroscopy.* medRvix, 2023. **2023.10.16.23297088**.
6. Li, Z., et al., *Effects of patient-specific treatment planning on eligibility for photodynamic therapy of deep tissue abscess cavities: retrospective Monte Carlo simulation study.* Journal of Biomedical Optics, 2022. **27**(8): p. 083007.
7. Li, Z., et al., *Treatment planning for photodynamic therapy of abscess cavities using patient-specific optical properties measured prior to illumination.* medRxiv, 2023. **2023.10.23.23297420**.
8. Farrell, T.J., B.C. Wilson, and M.S. Patterson, *The use of a neural network to determine tissue optical properties from spatially resolved diffuse reflectance measurements.* Phys Med Biol, 1992. **37**(12): p. 2281-6.
9. Farrell, T.J., M.S. Patterson, and B. Wilson, *A diffusion theory model of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties in vivo.* Medical Physics, 1992. **19**(4): p. 879-888.
10. Hull, E.L. and T.H. Foster, *Steady-state reflectance spectroscopy in the P_3 approximation.* J Opt Soc Am A, 2001. **18**(3): p. 584-599.
11. Hennessy, R., et al., *Monte Carlo lookup table-based inverse model for extracting optical properties from tissue-simulating phantoms using diffuse reflectance spectroscopy.* J Biomed Opt, 2013. **18**(3): p. 037003.
12. Kienle, A., et al., *Spatially resolved absolute diffuse reflectance measurements for noninvasive determination of the optical scattering and absorption coefficients of biological tissue.* Appl Opt, 1996. **35**(13): p. 2304-14.
13. Nguyen, M., et al., *Machine learning to extract physiological parameters from multispectral diffuse reflectance spectroscopy.* Journal of Biomedical Optics, 2021. **26**(5): p. 052912.
14. Lan, Q., R.G. McClarren, and K. Vishwanath, *Neural network-based inverse model for diffuse reflectance spectroscopy.* Biomed Opt Express, 2023. **14**(9): p. 4725-4738.
15. Chollet, François, and Others. *Keras*. 2015; Available from: https://keras.io.
16. Humbird, K.D., J.L. Peterson, and R.G. McClarren, *Deep Neural Network Initialization With Decision Trees.* IEEE Trans Neural Netw Learn Syst, 2019. **30**(5): p. 1286-1295.

17. Goodfellow, I.J., et al., *An Empirical Investigation of Catastrophic Forgetting in Gradient-Based Neural Networks.* arXiv, 2013. **1312.6211**.
18. van Staveren, H.J., et al., *Light scattering in lntralipid-10% in the wavelength range of 400–1100 nm.* Applied Optics, 1991. **30**(31): p. 4507-4514.
19. Michels, R., F. Foschum, and A. Kienle, *Optical properties of fat emulsions.* Opt Express, 2008. **16**(8): p. 5907-25.
20. Chong, K.C. and M. Pramanik, *Physics-guided neural network for tissue optical properties estimation.* Biomed Opt Express, 2023. **14**(6): p. 2576-2590.
21. Ivancic, M., et al., *Efficient estimation of subdiffusive optical parameters in real time from spatially resolved reflectance by artificial neural networks.* Opt Lett, 2018. **43**(12): p. 2901-2904.
22. An, J., et al., *Neural network-based optimization of sub-diffuse reflectance spectroscopy for improved parameter prediction and efficient data collection.* J Biophotonics, 2023. **16**(5): p. e202200375.
23. Wilson, B.C., M. Jermyn, and F. Leblond, *Challenges and opportunities in clinical translation of biomedical optical spectroscopy and imaging.* J Biomed Opt, 2018. **23**(3): p. 1-13.
24. Zhao, F., et al., *Reproducibility of identical solid phantoms.* J Biomed Opt, 2022. **27**(7).
25. Kelly, C.J., et al., *Key challenges for delivering clinical impact with artificial intelligence.* BMC Med, 2019. **17**(1): p. 195.