

# Deep Orthogonal Multi-Wavelength Fusion for Tomogram-Free Diagnosis in Diffuse Optical Imaging

Hanene Ben Yedder<sup>a,\*</sup>, Ben Cardoen<sup>a</sup>, Majid Shokoufi<sup>b</sup>, Farid Golnaraghi<sup>b</sup>, Ghassan Hamarneh<sup>a</sup>

<sup>a</sup>Medical Image Analysis Lab, School of Computing Science, Simon Fraser University, BC Canada V5A 1S6

<sup>b</sup>School of Mechatronic Systems Engineering, Simon Fraser University, BC Canada V5A 1S6

---

## ARTICLE INFO

---

*Article history:*

---

*Keywords:* Deep learning, Diffuse Diffuse optical tomography, Breast cancer, Lesion, Diagnosis, Reconstruction, Direct prediction, multitask learning, Multi-wavelength, Real time, Handheld probe.

---

## ABSTRACT

---

Novel portable diffuse optical tomography (DOT) devices for breast cancer lesions hold great promise for non-invasive, non-ionizing breast cancer screening. Critical to this capability is not just the identification of lesions but rather the complex problem of discriminating between malignant and benign lesions. To accurately reconstruct the highly heterogeneous tissue of a cancer lesion in healthy breast tissue using DOT, multiple wavelengths can be leveraged to maximize signal penetration while minimizing sensitivity to noise. However, these wavelength responses can overlap, capture common information, and correlate, potentially confounding reconstruction and downstream end tasks. We show that an orthogonal fusion loss regularizes multi-wavelength DOT leading to improved reconstruction and accuracy of end-to-end discrimination of malignant versus benign lesions. We further show that our raw-to-task model significantly reduces computational complexity without sacrificing accuracy, making it ideal for real-time throughput, desired in medical settings where handheld devices have severely restricted power budgets. Furthermore, our results indicate that image reconstruction is not necessary for unbiased classification of lesions with a balanced accuracy of 77% and 66% on the synthetic dataset and clinical dataset, respectively, using the raw-to-task model. Code is available at <https://github.com/sfu-mial/FuseNet>

---

## 1. Introduction

Breast cancer is the most frequently diagnosed cancer among women [1]. Pre-screening is usually carried out using self-breast examinations, which can suffer from high false-positive rates, or clinical breast examinations [2]. Although breast lumps are often benign, such as lipoma, cyst, or hamartoma, lesion malignancies may appear with a non-palpable sign; hence, regular screenings are critical [3]. While mammography is the most commonly used screening tool today, it has potential cumulative health risks due to its reliance on ionizing radiation and low sensitivity in patients with thick breast tissue [4]. Furthermore, the acquisition device's complexity and size limit patient screening throughput [5].

Imaging modalities based on near-infrared light are emerging as tools for biomedical diagnosis, given the non-ionizing nature of infrared light as well as their ability to penetrate a few centimeters into human structures, such as the skull, brain, and breast [6]. The recent progress of optical sensors makes

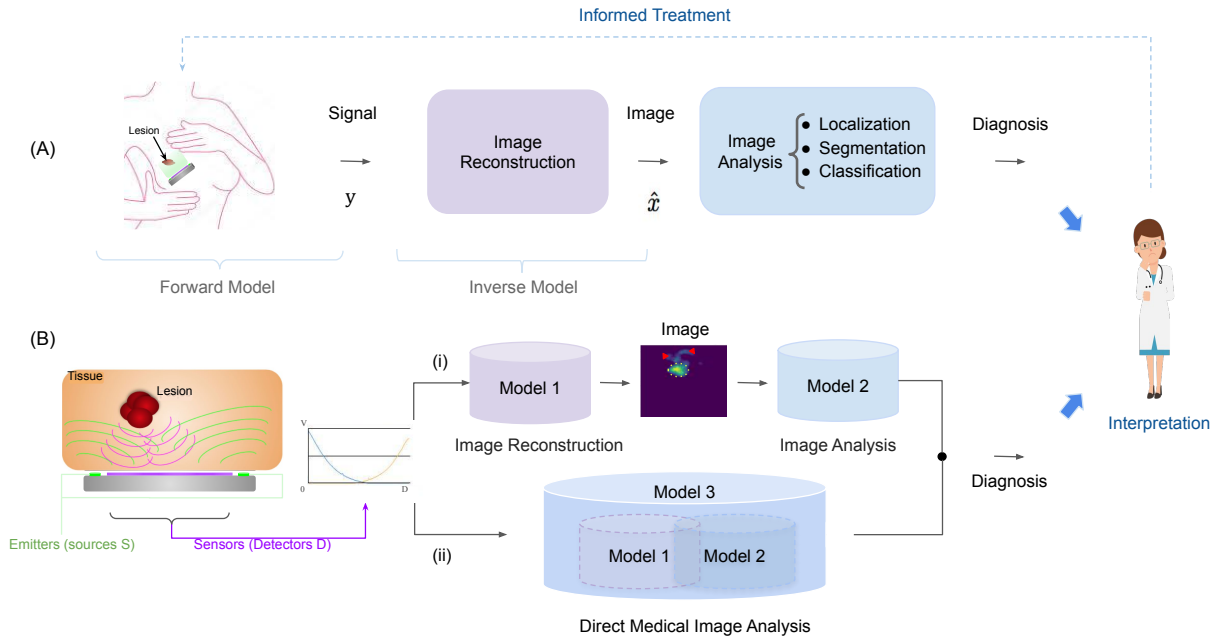
optical-based modalities increasingly attractive. Diffusion optical tomography (DOT) uses near-infrared light to image soft tissues, offering several advantages in terms of safety, costs, portability, and sensitivity to functional changes [7]. This technique has shown great potential in investigating functional brain imaging [8, 9] and breast cancer screening [10, 11]. Fig. 1-A shows a typical breast screening workflow in the medical setting.

DOT measures the distribution of tissue optical properties as a function of absorption and scattering coefficients. These properties are closely correlated to physiological markers and allow indirect quantitative assessment of tissue malignancy [7, 12]. Indeed, marked variations between healthy and tumor tissue are observed in terms of optical properties and chromophore components (e.g., oxy/deoxy hemoglobin and collagen) [13]. In particular, normal breasts and lesions can be separated in terms of optical coefficients at several wavelengths [14, 12, 15].

These properties make DOT a potentially promising tool in the pre-screening of patients in a clinical setting, saving them from unnecessary exposure to more precise but potentially harmful ionizing modalities such as CT. In such a setting,

---

\*Correspondence to be addressed to {hbenyed, hamarneh}@sfu.ca



**Fig. 1. Typical breast cancer screening workflow.** (A) Images reconstructed by an inverse model, from signals collected by the acquisition hardware, are analyzed for assessment, diagnosis, and treatment prognosis. (B) Screening pipelines can be divided into two paradigms: (i) Accurate reconstruction followed by image based classification. (ii) A direct prediction model omits image reconstruction to focus solely on the ultimate task and can help overcome sub-task errors, e.g., reconstruction induced false positives, marked by red triangles in this scenario, in paradigm (i).

there is a clear need for both low latency, i.e., method inference speed, preferably real-time, and accurate reconstruction and classification.

A DOT scanner is comprised of an array of emitters and receivers, using low-powered LEDs or lasers to measure the optical transmission [16] or reflection [10] of light beamed into the tissue at various locations on the tissue surface. While an optimized probe design enables reduced hardware complexity and better portability, it increases the complexity of the reconstruction task, mainly when the number of sources is limited [17].

### 1.1. DOT Reconstruction Algorithms

DOT image reconstruction is an ill-posed inverse problem, subject to artifacts [18]. Reconstruction quality and depth sensitivity are inversely proportional to the distance between source and detector and noise level, and strongly depend on the reconstruction method [19]. In addition, the highly heterogeneous nature of malignant cancerous tissue further complicates the reconstruction task [20]. A portable design with a limited power budget, significantly reducing the number of available sensor-detector pairs and the available computational envelope for reconstruction, complicates matters further by increasing the ill-posedness of the reconstruction problem. We refer the interested reader to Aspri *et al.* [21] for a more in-depth mathematical modeling and review.

While difference imaging approaches rely on a reference measurement to recover the change in the tissue's optical properties, e.g., a reference tissue, a phantom, or the previous condition like a rest stage in brain DOT, absolute imaging approaches

use a single set of measurements to reconstruct optical coefficients. In this manuscript, we focus on absolute imaging.

Traditional image reconstruction techniques commonly rely on non-linear methods minimizing an objective function, iteratively until convergence, e.g., gradient and Newton-type methods [37]. Based on an initial homogeneous tissue optical properties estimate, the difference between the measured signal and the modelled data is used to iteratively update the estimate until achieving convergence within acceptable limits with the measured data. Regularization terms are leveraged to ensure convergence by restricting the space of all possible solutions to only a subset of physically plausible ones. A comprehensive review is presented in [19].

Even though non-linear methods follow directly from the underlying mathematical problem formulation, in practice they have a high computational cost as each iteration needs to be optimized independently at reconstruction time, prohibiting real-time reconstruction. Furthermore, the reconstruction accuracy is easily compromised as the number of sources and detectors is reduced, and reconstruction of complex shapes can become challenging [18]. To address these shortcomings, researchers have explored deep learning (DL) as an alternative approach [38, 39]. A deep learning model for DOT reconstruction is typically trained in a supervised setting on *in silico* or phantom training data pairs. By incorporating complex and diversified data samples, the model can selectively enrich its feature space to improve performance on real-world data. Recent studies, e.g., [25, 17, 29, 26], have shown image reconstruction and classification are faster and more accurate when deep learning algorithms are used instead of conventional reconstruc-

**Table 1. Regularization approaches in state-of-the-art DL-reconstruction methods. D: design approach; FF: feed-forward, I: iterative unrolled based model; M: multi-modal/wavelength; S/P: in silico/phantom data; and C: clinical patient data.**

	Ref	D	M	S/P	C	Approach to mitigate ill-posedness
Sabir et al., Ben Yedder et al.(2018), Nizam et al.(2022)	[22, 23, 24]	FF	×	✓	×	CNN learns the nonlinear end-to-end mapping
Ben Yedder et al.(2019)	[25]	FF	×	✓	×	Promote appearance similarity
Mozumder et al.(2021), Mozumder et al.(2022)	[26, 27]	I	×	✓	×	Augment Gauss-Newton algorithm with deep learning
Nizam et al.(2023)	[28]	FF	✓	✓	×	Micro-CT structural prior
Yoo et al.	[29]	FF	×	✓	×	Model based on Lippmann-Schwinger equation
Fang et al.	[30]	FF	×	✓	×	Reflection model as sum of features from different depths
Zhao et al.	[31]	I	×	✓	×	Data-driven unrolled network promoting appearance similarity
Zou et al.(2021), Li et al., Zou et al.(2022)	[32, 33, 34]	FF	✓	✓	✓	Multi-modal representation learning (US+DOT)
Ben Yedder et al.(2021)	[17]	FF	×	✓	✓	Deep spatial-wise attention network
Ren et al.	[35]	FF	×	✓		Adaptive sampling based neural fields
Murad et al.	[36]	FF	×	✓		Dual signal and image encoder network
Ours	–	FF	✓	✓	✓	Orthogonal multi-wavelength representation learning

tion methods. One advantage deep learning-based algorithms have over classical reconstruction methods is that they can exploit implicitly learned feature encodings from the DOT sensor data, whereas classical reconstruction algorithms can exploit only priors encoded by human designers [38]. Recent advances tackle the problem of ill-posedness in a variety of ways. We summarize the closely related approaches in a tabulated overview (Table 1), and refer the interested reader to [39, 40] for a more in-depth review.

### 1.2. Multi-wavelength DOT

The primary motivation for multi-wavelength DOT is to exploit the different but complementary responses of chromophores, found in different breast tissue components, to multi-wavelength excitation, given that chromophores absorb photons at different rates at distinct wavelengths [41]. A multi-spectral image can be obtained using several LEDs or lasers of multiple wavelengths as illumination. The different LEDs can be used consecutively to capture an image per wavelength or combined as one multi-spectral image [42].

The captured multi-wavelength data can provide more spatial and contextual information, enabling more robust and accurate identification and discrimination of disease-correlated biological anomalies [12]. This wavelength sensitivity is leveraged to analyze optical spectra and reconstruct images of the exposed tissue for diagnostic purposes, given that recovered chromophore concentration changes can convey information about functional brain vascular events and the characterization and monitoring of breast lesions [43]. The choice of wavelength in DOT involves a trade-off between improved separation of optical properties and the limitations associated with decreased penetration and reduced signal-to-noise ratio [44, 45]. The specific requirements of a given imaging task will dictate the optimal balance between these factors.

Utilizing multi-wavelength data for improving DOT image reconstruction and diagnosis has been an active field of research, illustrating that the accuracy of the optical coefficient can be improved using measurements with multiple wavelengths [46, 47, 42, 48]. Recent studies, summarized in a tabulated overview (Table 2) have shown that using measurements with multiple wavelengths can improve the recovery of optical coefficients and provide higher SNR and lower error [49, 42]. Improvement, however, depends on wavelengths

selection scheme and utilized instrument given the specific noise impact [48]. This finding is supported by Zimek et al. [50], who reported that adding dimensions can harm discriminative potential if those dimensions do not improve the signal-to-noise ratio.

Augmenting DOT with ultrasound is finding recent adoption as well, an example of multi-modal fusion [51, 52, 53], [54]. The aforementioned art is based on conventional reconstruction algorithms. To the best of our knowledge, no deep learning-based method has explored the merit of exploiting multiple wavelengths in DOT-reconstruction and diagnosis.

### 1.3. Multi-wavelength as Data Fusion

Data fusion models mimic higher cognitive abstraction in the human brain by synthesizing information from multiple sources for improved decision-making. While data fusion is non-trivial, the resulting contribution of multiple data sources or multi-modal data can significantly improve the performance of deep learning models [58, 59]. The underlying motivation for collecting multi-modal data is to learn the optimal joint representation from rich and complementary features of the same object or scene. In the context of combining multiple information sources to learn more powerful representations, the terms ‘early’ and ‘late’ fusion are commonly used [60]. Early fusion refers to concatenating input data from multiple sources in separate channels before presenting it as input to the network. In contrast, late fusion involves processing each input data individually and aggregating their output. Mid-fusion restricts cross-data flow to later layers of the network, allowing early layers to specialize in learning and extracting data-specific patterns [61].

Attention mechanisms have been shown to be suitable for the fusion of features that usually suffer from confounding issues such as conflicting or canceling information, correlation, and noise. Attention provides an approach to learn to select informative subsets of the data and the relationship between data streams, before fusing them into a single comprehensive representation [59, 62]. A multi-head attention architecture, transformer based models have recently gained increased adoption [63, 61]. However, the high computational cost and complexity, scaling adversely with input sequence length, remain a significant challenge, especially given the real-time requirement.

**Table 2. Multi-wavelength DOT for Image Reconstruction and Diagnosis. Di: dimension; Freq: frequency range; D: dataset, S: simulation, P: phantom, C: clinical; and MF:Multi-Frequency.**

Ref	Leveraging different frequency schemes	DI	Freq (MHz)	D
Chen et al. [49]	Improve joint optical coefficients recons.	2D	100-250	S
Shifa et al. [42]	Enhance fused MF image quality	2D	100-1000	S
Liu et al. [55]	Compensate physiological and noise interference in recons.	3D	361-382	P
Kazanci et al. [56]	Frequency shifting for reduced recons. ill-posedness	2D	$100+5*i, \forall i \in \{0, \dots, 100\}$	S
Perkins et al. [57]	Minimize the effect of phase data and improve contrast	2D	{78, 141, 203}	P
Applegate et al. [48]	Evaluate the impact of frequency selection	2D	50-500	S
Taroni et al. [14]	Discrimination between malignant and benign tissue	2D	283-472	C

Self-supervised learning (SSL) based on a joint embedding architecture, driven by the maximization of the information content of the network branches' embedding, opened the door to the application of joint-embedding SSL to multi-modal signals [64]. The idea is to produce independent embedding variables, removing confounding effects such as partial correlation and avoiding modal collapse between data streams by encouraging architecture diversity between branches, using loss based normalized cross-correlation matrix [65] or explicit variance-preservation term for each embedding [64].

Imposing orthogonal constraints in linear and convolutional neural network layers can act as a form of regularization that can help improve task performance and be beneficial for the network's generalization [66, 67]. Orthogonality in feature space was proposed to encourage intra-class compactness and inter-class separation of the deep features, and has shown improvement in classification tasks [68]. Multi-modal orthogonalization has been used to force uni-modal embeddings to provide independent and complementary information to the fused prediction [69, 70]. Another advantage is that an orthogonal encoding can enforce the learning of a more sparse correlation-free representation. The resulting smaller encoding can reduce architecture dimensions, and serve as an implicit regularization.

#### 1.4. Towards Direct Medical Image Analysis in DOT

Traditional computational pipelines in biomedical imaging involve solving tasks sequentially (Fig. 1-B.i, e.g., segmentation followed by classification or detection). Although each task is usually solved separately, the useful clinical information extracted by the second task highly depends on the first task's results. While a 'joint' or multi-stage model where different tasks are lumped together, for example, image reconstruction then classifying diagnosis, can benefit from feature sharing and joint parameters tuning for both tasks, significant computational resources are required to optimize sub-tasks that may not necessarily lead to end-task improvements. In contrast, in the direct medical image analysis [71] (DMIA) paradigm, end task results are directly inferred from raw/original data (e.g., raw sensors or whole image/volume) as illustrated in Fig. 1-B.ii. Therefore, the model can focus solely on the end task, reclaiming some of the computational resources for improved results while requiring fewer resources. For instance, Wu et al. [72] trained a neural network for joint reconstruction and lung nodule detection from raw acquisitions and showed performance

improvement compared to a two-stage approach. Hussain et al. [73] had shown that a segmentation-free kidney volume estimation can help overcome segmentation errors and limitations and reduce the false-positive area estimates. In a similar perspective, Taghanaki et al. [74] investigated a segmentation-free tumor volume and activity estimation in PET images. More recently, Abhishek et al. [75] illustrated that, in the context of cancerous skin lesions, predicting the management decisions directly can be a simpler problem to address than predicting the diagnosis followed by management decisions, as one action can be prescribed to multiple subsets of disease classes.

#### 1.5. Contributions

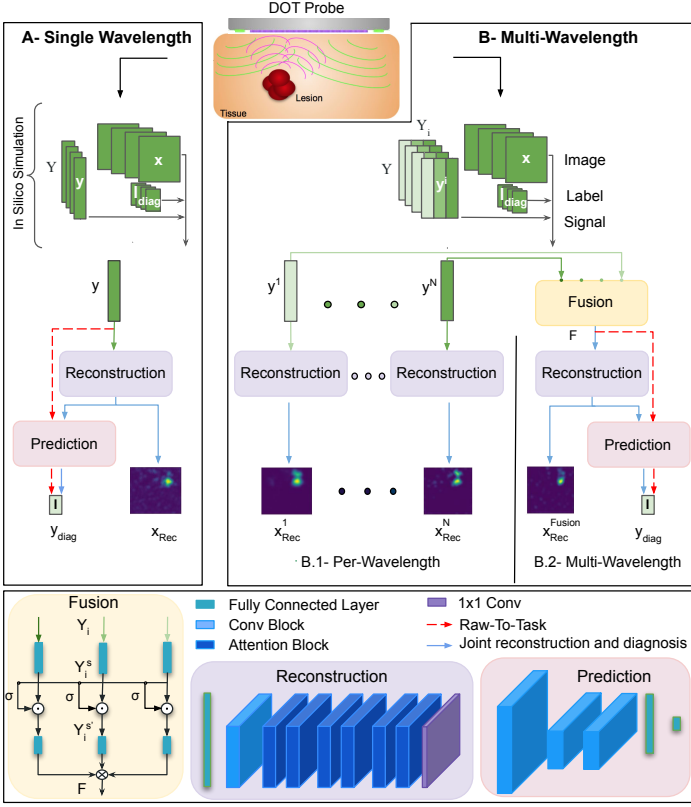
We make the following contributions in this paper:

(i) We investigate the benefit of multi-wavelength data on the quality of DOT reconstruction and breast lesion diagnosis. Previously, many works have addressed the multi-wavelength reconstruction problem or diagnosis, albeit using conventional methods. Despite the importance of multi-wavelength acquisition for chromophore reconstruction, no deep learning framework has investigated multi-wavelength fusion nor joint reconstruction and diagnosis to date. Here, we present a novel approach designed to recover the optical properties of breast tissue from multi-wavelength data with a deep orthogonal fusion model followed by a diagnosis.

(ii) To the best of our knowledge, this is the first deep learning-based method that investigates the merits of tackling the diagnosis prediction task from raw sensor data directly without image reconstruction in DOT (direct prediction<sup>1</sup>). Results with and without reconstruction are contrasted using a modular pipeline, highlighting the potential of the raw-to-task model for improved accuracy, while reducing computational complexity.

(iii) We extend a fusion network [62] by training models using an orthogonalization loss function [68] to maximize the independent contribution of each wavelength data and emphasize their collective strength, with improved predictive performance compared to a single wavelength model.

<sup>1</sup>While our direct prediction contribution omits the 'tomogram' part of DOT, and thus works directly on near-infrared (NIR) sensor data, our fusion contribution applies both to tomogram reconstruction as well as tomogram-free reconstruction. Thus, we continue to use DOT throughout the paper instead of NIR.



**Fig. 2. Architecture overview of the proposed DOT image reconstruction and diagnosis method. (A) single-wavelength and (B) multi-wavelength signals ( $y$ ) along with corresponding ground truth diagnosis labels ( $l_{diag}$ ) and images ( $x$ ) are used to train the model. In the single-wavelength variant of our method (A),  $y$ , is used as input to the image reconstruction, then the resulting image is used for diagnosis prediction. In multi-wavelength, note the two variants: (B.1) per-wavelength reconstruction and (B.2) multi-spectral reconstruction and diagnosis. For both single and multi-wavelength, the red dashed lines depict the raw-to-task flow, where the image reconstruction is skipped and the diagnosis is predicted directly from  $y$ . The bottom panel shows the details of the multi-wavelength fusion, reconstruction, and prediction modules.**

Section 2, introduces our proposed model for multi-wavelength DOT fusion and defines the two prediction pipelines (raw-to-task and joint reconstruction and diagnosis). Physics-based computational simulation and real patient datasets are detailed in Section 3.1.1. In silico performance results are presented in Section 3.2 and results on real-world data in Section 3.3. We conclude the paper by discussing insights and limitations on interpretability, speed, and adaptive dynamic treatment in Section 4.

## 2. Methodology

Solving the inverse problem in DOT recovers the spatial distribution of a tissue's optical properties  $x \in \mathbb{R}^{W \times H}$  based on the measured boundary data  $y^i \in \mathbb{R}^{S \times D \times N}$ , from  $S$  sources (emitters) with  $D$  sensors (detectors) at different wavelengths  $i \in \{1, N\}$ . The learned inverse function  $\mathcal{F}^{-1}(\cdot)$  maps the raw measurements  $y$  to an image estimate  $\hat{x}$  while remaining faithful to the underlying physics constraints. Learning the inverse

function  $\mathcal{F}^{-1}(\cdot)$  is carried out by solving:

$$\theta^* = \operatorname{argmin}_{\theta} \mathcal{L}(\mathcal{F}^{-1}(y^i; \theta); x) + \lambda \mathcal{R}(\mathcal{F}^{-1}(y^i; \theta)), \quad (1)$$

where  $\mathcal{L}$  and  $\mathcal{R}$  are the network loss function and the regularization,  $\theta$  are the optimized network weights that parameterize  $\mathcal{F}^{-1}$ . The reconstruction of an image based on the fusion of all raw signals from diverse wavelengths is considered as well by using the fusion network described in Section 2.1. While reconstructing an accurate 2D/3D image/volume from collected measurements has been the mainstream task in DOT, in a clinical setting, the ultimate purpose is not necessarily obtaining the image itself but rather making an informed clinical diagnosis or management decision, such as lesion detection and classification into predefined classes. To compare the impact of omitting the reconstruction and directly predicting the end task, we implemented two architectures: The first, FuseNet, reflects classical approaches, i.e., a classification module is appended to the output of the reconstruction layer to make a prediction, where the result of the multi-spectral reconstruction is used to supervise the classification task (Section 2.2). Whereas the second, Raw-to-Task, uses the same classification module to make a prediction based on the fused raw data directly, i.e., no reconstruction is considered in between. The ultimate goal is to study the ability of deep learning to provide superior prediction based on the raw signal only while reducing model complexity and computational cost (Section 2.3).

### 2.1. Fusion Network

Given multi-wavelength raw data paired with known diagnosis outcomes, the objective is to learn a robust multi-wavelength representation in a supervised learning setting. While many fusion strategies have been proposed in computer vision, natural language processing, and multimodal biomedical data, strategies for fusing data in multi-wavelength DOT data remain unexplored in deep learning-based approaches. Inspired by recent methods for multimodal data fusion [62, 69], we adopt a similar attention-based mechanism to control the expressiveness of features from each input wavelength before constructing the multi-wavelength embedding, while uniquely feeding the raw data directly with no further pre-processing. Let  $Y \in \mathbb{R}^{S \times D \times N \times M}$  be a training mini-batch including  $M$  tissue samples, each collected using  $N$  wavelengths such that  $Y = [Y_1, Y_2, \dots, Y_N]$  where for each wavelength  $i$ ,  $Y_i = [y_1^i, \dots, y_M^i]$  includes data for  $M$  samples. When  $N > 1$ , input measurements from each wavelength are combined using the fusion branch (Fusion, Fig. 2). To reduce the impact of noisy input features and compress the size of the feature space, each  $Y_i$  is first passed through a fully connected layer of length  $F_1$ , with ReLU activation, outputting  $Y_i^s \in \mathbb{R}^{F_1 \times M}$ , followed by an attention mechanism that scores the relevance of each feature in  $Y_i$ . We define wavelengths  $\ni$  as the set  $\{j\}$  such that  $j \in \{1, \dots, N\} \setminus \{i\}$ , i.e., for wavelengths other than  $i$ . A linear transformation  $W_A$  of wavelengths  $Y_{\ni}$ , that would score the relative importance of each feature in  $i$ , is learned.  $W_A$  is a learned weight matrix parameters for feature gating. The attention weights vector  $a_i$  is then applied to  $Y_i^s$ , an element-wise product of scores and features, to form the

attention-weighted embedding  $Y_i^{s'} \in \mathbb{R}^{l \times 1 \times M}$ :

$$Y_i^{s'} = a_i * Y_i^s = \sigma(W_A * [Y_\ominus]) * Y_i^s. \quad (2)$$

Finally, attention-weighted embeddings are passed through a fully connected layer of length  $F_2$ , with ReLU activation, then combined through a Kronecker product between all wavelength embeddings to capture possible interactions. Each vector is appended by 1 to capture partial interactions between wavelengths [62]. The final fused embedding is then defined as:

$$F = \begin{bmatrix} 1 \\ Y_1^{s'} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_2^{s'} \end{bmatrix} \otimes \dots \otimes \begin{bmatrix} 1 \\ Y_N^{s'} \end{bmatrix}. \quad (3)$$

$F \in \mathbb{R}^{l' \times l' \times l' \times M}$ , for  $N = 3$  and  $l' = F_2 + 1$ , is a  $N$ -dimensional hypercube of all wavelength interactions.

**Table 3. Network Layers Description for Joint Multi-wavelength Reconstruction and Diagnosis**

Branch	Description	Details
Reconstruction Branch	Fully Connected Layer	$128 \times 128$
	Convolutional Layer	32 channels, filter size 3, ReLU
	Residual Attention Blocks	4 blocks, 32 channels, filter size 3, ReLU
	Squeeze and Excite Modules	Attention map: Convolutional layer with a $1 \times 1$ kernel, Sigmoid activation, Multiply input tensor element-wise with the attention map
Prediction Branch	Convolutional Layers	2 FC layers with max pooling
	Classification Layers	2 FC final layers for diagnosis prediction
Fusion Branch	Gated Multimodal Units	FC layers with ReLU, Gating mechanisms (multi-spectral branch)
Input	Raw Data	$Y \in \mathbb{R}^{S \times D \times N}$
Output	Per-Wavelength Reconstructed Image	$x_{Rec}^i \forall i \in \{1, \dots, N\}$
	Multi-Spectral Reconstructed Image	$x_{Rec}^{Fusion}$
	Diagnosis Prediction	$y_{diag}$

## 2.2. Joint Multi-wavelength Reconstruction and Diagnosis

The task is to recover tissue optical properties and diagnosis outcome given raw signal data. While a single wavelength model SW-JRD (Fig. 2-A), used as a baseline, relies on a single wavelength measurement to reconstruct spatially distributed optical coefficients and predict diagnosis, a multi-wavelength model (FuseNet) relies on a joint representation from multiple wavelength measurements (Fig. 2-B). The multi-wavelength model, including a network with multiple branches as shown in (Fig. 2-B), inputs  $N$  measurements of the same scanned tissue at  $N$  wavelengths. A multi-spectral image that combines all wavelength measurements, using the fusion branch encoding (Fig. 2-B.2), is reconstructed and passed to a classification module for diagnosis prediction. Furthermore, a per-wavelength image is reconstructed using each modulated wavelength signal. As depicted in (Fig. 2-B), the FuseNet model outputs are  $x_{Rec}^i \forall i \in \{1, \dots, N\}$ ,  $x_{Rec}^{Fusion}$ , and  $y_{diag}$  which denote the per-wavelength reconstructed image (Fig. 2-B.1), the multi-spectral reconstructed image, and the predicted diagnosis label (Fig. 2-B.2), respectively.

Using multiple inputs, per wavelength network reconstruction branches (Fig. 2-B) learn independent representations, where features derived from each input measurement ( $Y_i$ ) are only useful for the corresponding output  $x_{Rec}^i$ . Furthermore, given the differences in initialization, the branches can converge to disconnected modes in weight space, thereby behaving

as independently trained neural networks. Empirically, we observe that they converge to distinct optima. For this multi-task reconstruction and prediction model, we extend the multi-task framework [17] and train a model to simultaneously reconstruct a per-wavelength image, localize the lesion, and predict the diagnosis.

The reconstruction branch (Fig. 2) implements the design detailed in the multi-task framework [17] with a fully connected layer,  $128 \times 128$ , followed by a convolutional layer and 4 residual attention blocks with 32 channels, filters of size of 3 and ReLU activation, to produce the final reconstruction image. While the first and last layers are shallow feature extractors, the attention blocks extract hierarchical attention-aware features with modules of the form: two convolutions followed by squeeze and excite modules. This deep spatial-wise attention network attends to the most important features by reweighting features according to their interdependencies in feature space and filtering noisy ones. In contrast to the difference approach, which uses a reference measurement of healthy tissue to compute the contrasted inverse image, our reconstruction module is designed to learn the mapping directly, from the measured data to the desired output, without the need for any prior knowledge or references that can bias the search space. Furthermore, obtaining such a reference measurement from a homogeneous background in a clinical setting, such as a breast cancer screening, is not trivial; hence, we consider absolute imaging, where the network learns the inverse mapping between sensor measurements and the image domain directly.

The prediction branch (Fig. 2) includes 2 convolutional layers with max pooling and two final classification layers. Raw data from different wavelengths are passed to the reconstruction branch except for the multi-spectral subnetwork, where raw data from different wavelengths are first fused via the fusion branch. The fused features are passed to the reconstruction branch, which outputs a multi-spectral image followed by a classification layer to output the final classification prediction. Table 3 details the architecture for the joint multi-wavelength reconstruction and diagnosis serving as a foundational framework for other ablation architectures, which encompass subsets of its components.

The multi-task loss ( $L_{MULTI}$ ) encompasses all three tasks: reconstruction, lesion localization, and diagnosis as a sum of losses for each task is defined as follows:

$$L_{MULTI} = L_{REC} + L_{DIAG} \quad (4)$$

where  $L_{REC}$  and  $L_{DIAG}$  denote the reconstruction loss and the diagnosis losses, respectively.

### 2.2.1. Reconstruction loss

We adopt the reconstruction loss defined by Ben Yedder et al. [17]. The mean square error loss  $L_{MSE}$  combined with the location loss  $L_{LOC}$  guide the image reconstruction and lesion localization of the network as per (5).  $L_{MSE}$  recovers the pixel-wise representation of the image.

$$\begin{aligned} L_{REC} &= L_{MSE} + \beta L_{LOC}, \\ L_{LOC} &= \|DT(\mathcal{F}^{-1}(y_i, \theta), x) - DT(x)\|, \end{aligned} \quad (5)$$

**Table 4. Summary of variants of our method architectures input and output details\*. N: number of wavelengths; S: sources; D: detectors; H: height; W: width.**

	Input	Output	
		Direct prediction	Joint reconstruction and diagnosis
Single-Wavelength	$Y_1 \in \mathbb{R}^{S \times D}$	$y_{diag} \in \mathbb{R}$	$x_{Rec} \in \mathbb{R}^{W \times H}$
		$y_{diag} \in \mathbb{R}$	$y_{diag} \in \mathbb{R}$
		<b>SW-DP</b>	<b>SW-JRD</b>
Multi-Wavelength	$Y \in \mathbb{R}^{S \times D \times N}$	$y_{diag} \in \mathbb{R}$	$x_{Rec}^j \in \mathbb{R}^{W \times H} \quad \forall i \in \{1, \dots, N\}$
			$x_{Rec}^{Fusion} \in \mathbb{R}^{W \times H}$
			$y_{diag} \in \mathbb{R}$
		<b>Raw-to-Task</b>	<b>FuseNet</b>

where DT denotes the distance transform and computes the Euclidean distance between the image pixel location and the lesion boundaries,  $\theta$  denotes the parameters of the multi-task model, and  $\beta \in [0, 1]$  is a hyper-parameter controlling the contribution of  $L_{LOC}$ .

### 2.2.2. Diagnosis loss

The diagnosis loss,  $L_{DIAG}$ , is a weighted sum of the categorical cross entropy loss  $L_{CE}$ , and the orthogonal projection loss  $L_{OPL}$ :

$$L_{DIAG} = L_{CE} + \gamma L_{OPL}, \quad (6)$$

$$\begin{aligned} \text{where: } L_{CE} &= L_{CE}(x, l_{diag} | \Theta) \\ &= - \sum_{j=1}^{n_{diag}} l_{diag,j} \cdot \log(\phi(x | \Theta)_j), \end{aligned} \quad (7)$$

$$\begin{aligned} L_{OPL} &= (1 - s) + |d| \\ s &= \sum_{\substack{i,j \in B \\ y_i=y_j}} \langle \mathbf{f}_i, \mathbf{f}_j \rangle, \quad d = \sum_{\substack{i,k \in B \\ y_i \neq y_k}} \langle \mathbf{f}_i, \mathbf{f}_k \rangle, \end{aligned}$$

$n_{diag}$ ,  $l_{diag}$  denote the number of classes in the diagnosis prediction tasks and ground truth label, respectively.  $\phi(x|\Theta)_j$  denotes the predicted probability for the  $j^{th}$  class by the model parameterized by  $\Theta$ .  $\gamma \in [0, 1]$  is a hyper-parameter balancing the contribution of the  $L_{OPL}$ .  $|x|$  is the absolute value operator,  $\langle x, y \rangle$  the cosine similarity operator applied on two vectors, and  $B$  denotes the mini-batch size.

The orthogonal projection loss  $L_{OPL}$ , as defined in [68], is used to maximize separability between classes by enforcing class-wise orthogonality in the intermediate feature space and simultaneously ensuring inter-class orthogonality (d term) and intra-class clustering ((1-s) term) within a mini-batch.

### 2.3. Direct Prediction: Raw to Task Model

The ultimate aim of DOT-based screening is the early identification and classification of breast cancer lesions. Therefore, we investigate if focusing exclusively on the end task, at the cost of omitting the reconstruction of a 2D image, can perform better or worse compared to classification with the intermediate reconstruction. Without the need to reconstruct a 2D image, the architecture and computational complexity reduce significantly,

**Table 5. Optical coefficients distributions on the In Silico Dataset for wavelengths in 690-850 nm spectrum [76]**

		Healthy tissue	Benign	Malignant
Absorption $\nu_a(cm^{-1})$	690	$0.042 \pm 0.013$	$0.08 \pm 0.04$	$0.110 \pm 0.066$
	750	$0.046 \pm 0.024$		$0.100 \pm 0.060$
	800	$0.052 \pm 0.015$		$0.118 \pm 0.096$
	850	$0.032 \pm 0.005$		$0.124 \pm 0.089$
Scattering $\nu_s(cm^{-1})$	690	$12.9 \pm 2.3$	$19.4 \pm 8.4$	$13.5 \pm 4.7$
	750	$8.70 \pm 2.2$		$11.6 \pm 3.9$
	800	$10.5 \pm 1.2$		$12.2 \pm 1.7$
	850	$8.40 \pm 0.4$		$9.10 \pm 1.9$

leading to a reduction in power consumption and data computation latency. The classification module is used to make predictions based on the fused raw data, where combined features, extracted from different wavelengths using the fusion branch (Section 2.1), are passed to a convolutional layer for the prediction task and a final classification layer with the associated loss (Fig. 2-dashed lines). The diagnosis loss function,  $L_{DIAG}$ , is used to train the model given the raw input measurement where:

$$\begin{aligned} L_{CE} &= L_{CE}(y_i, \dots, y_N, l_{diag} | \Theta) \\ &= - \sum_{j=1}^{n_{diag}} l_{diag,j} \cdot \log(\phi(y | \Theta)_j), \end{aligned} \quad (8)$$

$y_i$  denotes the  $i^{th}$  measurement of the raw data and  $\phi(y^{(i)}|\Theta)_j$  denotes the predicted probability for the  $j^{th}$  class given an input  $y^{(i)}$  by the model parameterized by  $\Theta$ . The orthogonal projection loss  $L_{OPL}$  (7) is used to maximize separability between classes in the feature space.

FuseNet, Raw-to-Task, SW-JRD and SW-DP models are trained separately while using the same modules: fusion, reconstruction, and prediction modules. Table 4 summarizes different models input and output details.

### 2.4. Transfer Learning Network

In medical imaging settings, transfer learning [77] can be used to bridge the gap between simulated and clinical data by transferring knowledge learned from simulated data to improve the performance of models on clinical data [78]. This is particularly important in medical imaging and relatively new imaging devices such as DOB probes, where obtaining large quantities of annotated clinical data can be challenging and expensive. Similar to Ben Yedder et al. [17], we use transfer learning to render an in silico trained network applicable to real world data and reduce the disparities between real-world acquisition  $y^p$  and in silico simulated data  $y^s$ . A multi-layer perceptron (MLP) network is used to tackle the domain shift by minimizes the transfer learning loss  $\mathcal{L}_{TL}$  over  $N_p$  sets of real data measurements obtained using a phantom solution and their correspond-

ing tissue-equivalent simulated data:

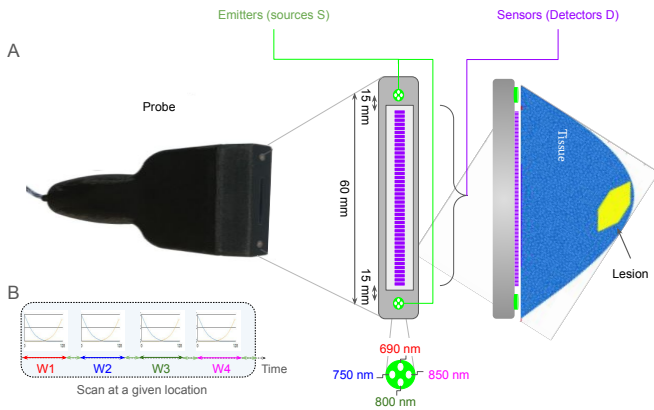
$$\theta^* = \underset{\theta}{\operatorname{argmin}} \mathcal{L}_{\text{TL}}(\theta)$$

where

$$\begin{aligned} \mathcal{L}_{\text{TL}}(\theta) = & \sum_{i=1}^{N_p} \|\phi(y_i^p; \theta) - y_i^s\| \\ & + \alpha \sum_{i=1}^{N_p} \sum_{j=1}^{D-w+1} \|\phi(y_{[j-w \ j+w]}^p; \theta) - y_{[j-w \ j+w]}^s\| \end{aligned} \quad (9)$$

$w$  is the size of the sliding window,  $D$  is the number of detectors,  $\alpha$  is a hyper-parameter that is used to control the contribution of the windowed mean absolute error loss. At inference time, the final reconstructed and diagnosis results are computed as:

$$\hat{x}^* = \mathcal{F}^{-1}(\phi(y^p)). \quad (10)$$



**Fig. 3.** (A) Aligning the spatial distribution of simulated sources and detectors with the physical layout of the probe (left). A sample synthetic mesh is also shown (right). (B) The light sources can be turned on and off individually or sequentially, with a period of 83ms, sufficient for functional imaging of biological tissue.

### 3. Results

We present results on both in-silico and clinical data. Results were obtained by training the model on the in-silico data. A transfer learning network, adapted from [17] and trained on a phantom dataset, bridges the distributions shift that is unavoidable when switching between in silico and real world data. A Gaussian noise was added to the signal, mimicking real world signal fluctuation, to improve model robustness to sensor noise and mimic the real-world drift of device characteristics on different clinics in between calibrations. This noise model depicts the highly variable noise to each individual detector as caused by sensor noise and interference of refracting light. Consistent with previous work [17, 29, 79], we set  $\sigma = 10\%$  of the maximum sensor value. Besides the simulated noise, the probe accounts for ambient light, the predominant noise source, as well by capturing a frame without any active emitters and then subtracting it from the actual data measurement, taken during

**Table 6. Summary of clinical data**

	Tumor Position	Tumor size (cm)	Tumor Type
Patient 1	Left Breast	1 × 0.8 × 0.7	BI-RADS 7
	Right Breast	2.5 × 0.8 × 0.8	
Patient 2	Left Breast	1.1 × 0.8 × 0.7	Benign
	Right Breast	2.2 × 1.7 × 1.7	
Patient 3	Left Breast	1 × 1 × 1	Non-invasive ductal
Patient 4	Left Breast	2.5 × 1.7 × 3.5	BI-RADS 5
Patient 5	Right Breast	2.4	BI-RADS 4
Patient 6	Right Breast	2.3 × 2.2 × 1.5	BI-RADS 4
Patient 7	Left Breast	1.7 × 1.4 × 1.2	BI-RADS 5
Patient 8	Left Breast	1.6 × 0.8 × 0.8	BI-RADS 5
Patient 9	Right Breast	2.2 × 2.1 × 2.3	Invasive ductal

clinical tests, before feeding it into our model. Performance evaluation captures image reconstruction quality, diagnosis accuracy, and speed. The next section provides details.

#### 3.1. Experimental Design

##### 3.1.1. Dataset

We simulate light propagation into tissue at different light wavelengths, 690, 750, 800, and 850 nm, illuminating the tissue sequentially, using the physics-based Toast++ software [80]. Probe geometry, with two LED sources and a row of 128 detectors placed in the same straight line, as illustrated in Fig. 1-B and Fig. 3, was configured to reflect real physical DOT probe geometry [10], used clinically, in terms of the number and geometry of sources and detectors and the used wavelengths. We collect training samples from synthesized tissues with known optical properties and labels. Lesions are modeled as tissue with perturbed optical coefficients embedded in an otherwise homogeneous diffusive medium. A set of 2D images with various lesion sizes, shapes, and positions discretized into finite element nodes (triangular meshes) is synthesized. In order to mimic real breast tissue optical parameters, we base the optical properties on realistic optical coefficient values that reflect characteristics of real breast lesions, such as microcalcifications and masses, [76, 14] as summarized in Table 5. A total of 4000 sample data pairs (256-D × 4 vectors, 2D image, label) are used to train and test our method, with 1750 benign samples and 1750 malignant samples, respectively, and 500 healthy (i.e., lesion free). The optical distributions are randomly selected using values from Table 5. Each sample includes the collected measurement vectors, one for each wavelength, the ground truth image, and the diagnosis label. The training dataset size is chosen as a compromise between training time and in-silico performance. We focused on the diversity of simulated scenarios while also being mindful of computational resources. While we note that in silico and phantom data can result in very large datasets [81], we focused on the diversity of simulated scenarios and adopted a dataset with various lesion numbers, sizes, and depths to emulate realistic conditions where the optical properties of the anomaly and surrounding tissues were taken from available in vivo breast tissue experimental data.

Our recently developed hand-held breast scanner (DOB-probe) [10, 82] was used to collect real patient data to test our method. The probe, as illustrated in Fig. 3, is composed of two symmetrical multi-wavelength light-emitting diodes (eLED) sources, illuminating the tissue symmetrically and delivering



near-infrared light to a body surface at different points, and featuring 128 co-linear equi-spaced detectors (a row) that measure the back-scattered light from the tissue and emitted from the boundary. Each eLED source consists of four different near-infrared (NIR) wavelengths: 690, 750, 800, and 850 nm. The penetration of photons within these wavelengths is optimized due to the minimal absorption of breast fat and water at the selected wavelengths. These encapsulated light-emitting diodes serve as pencil beam sources, emitting light into the scattering medium, and function as multi-wavelength pointed illumination sources within the tissue. Note that the wavelengths share variable overlap in the spectrum [14], motivating further the need for orthogonal encoding. To train the transfer learning module breast-mimetic phantoms, with known inhomogeneity locations, and DOB-probe were used to collect measurements [17].

Following the ethics and institutional review board approval protocol <sup>2</sup>, clinical data were collected from 9 participants diagnosed with breast tumors [83]. In a normal clinical pre-screening exam, a breast is usually divided into four quadrants, and different measurements are collected on each quadrant. Given that the used probe is in clinical studies [10, 42, 17], patients with known cancer localization are considered, and sweeps over the lesion location and the opposite healthy breast are collected. This step was essential to proving that the technology we introduced works well with human tissue. For each patient, height, weight, age, and gender, as well as details of the subjects' breast cancer, briefly summarized in Table 6, were recorded. The clinical dataset, reflecting the variety seen in clinical practice, was designed to acquire sufficient variety of sample from individuals presenting with benign tissue, masses, and microcalcifications, while adhering to resource and ethical constraints, we refer the interested reader to [83] for further details. Patients were placed in a supine position, and scans at multiple points over the lesion location and healthy breast were collected. On average, four different measurements (scans) were taken on each breast. Even though no reconstruction ground truth is available for real-world data, it is invaluable to detect robustness and real-world performance, with partial ground truth known from other modalities on the same patients. The precise location, size, and type of the tumor lesion were determined via mammography, ultrasound, or biopsy. Note that these details were only used for model performance evaluation and metrics calculation, while raw sensor signal only was inputted to the model. Another advantage of our direct prediction approach is that the absence of pixel-wise ground truth is less problematic compared to reconstruction based classification, as only the diagnosis label is required.

### 3.1.2. Implementation

Models were implemented in the Keras TensorFlow framework and trained for 100 epochs on an NVIDIA Titan X GPU. By optimizing the model's performance on the validation set, we set all hyper-parameters as follows: batch size to 16, learn-

Table 7. Hyperparameters summary.

Hyperparameter	Value
Framework	Keras TensorFlow
Training Epochs	100
GPU	NVIDIA Titan X
Batch Size	16
Learning Rate	$10^{-4}$
Optimizer	Adam
Initialization	Xavier
Early Stopping	10 epochs
Data Split	80/10/10
$\beta$ (Eq.5)	0.2
$\alpha$ (Eq.9)	0.5
$D$ (Eq. 9)	4
$\gamma$ (Eq. 6)	0.5
$F_1$	32
$F_2$	16

ing rate to  $10^{-4}$ , optimizer set to Adam, and initialization to Xavier. Early stopping was used if the validation loss had not improved within 10 epochs. To ensure a balanced and representative in silico training process, we employ an 80/10/10 data split strategy. This results in 3200 samples, per frequency, for training, 400 samples for validation, and another 400 samples for testing. For each class, the data split maintains the same proportion across the training, validation, and test sets. This approach ensures that our model is trained on a diverse range of examples while still allowing for robust evaluation on unseen data. Hyper-parameters  $\beta$  (5),  $\alpha$  and  $D$  (9), and  $\gamma$  (6) were set to 0.2, 0.5, 4 and 0.5, respectively. The fully connected units for the fusion branch were set to 32 and 16 for  $F_1$  and  $F_2$ , respectively. Table.7 provides a comprehensive summary of the model hyperparameters.

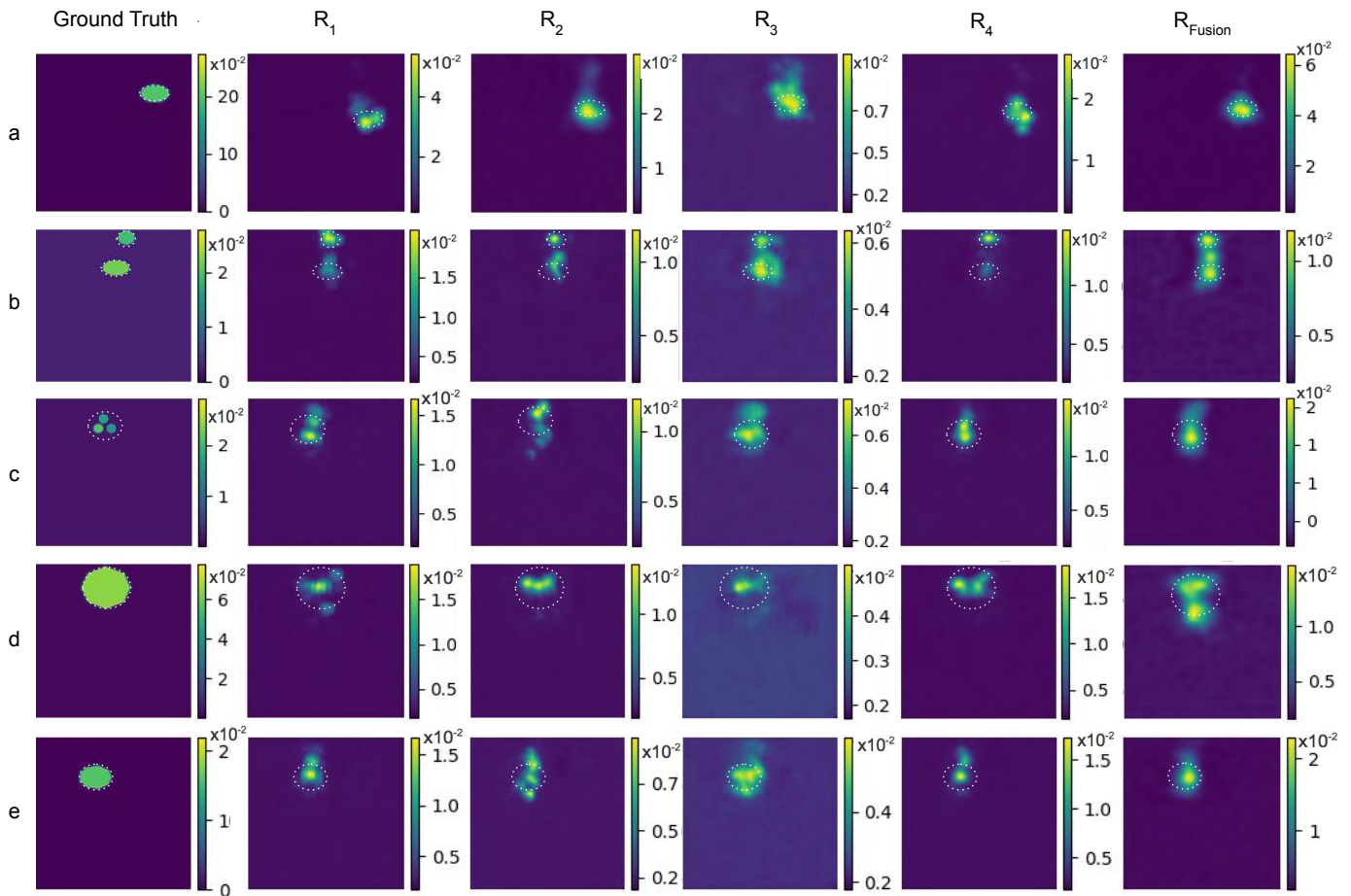
### 3.1.3. Evaluation metrics

To quantify the models' robustness, we look at (i) lesion localization error (Loc. Error); (ii) peak signal-to-noise ratio (PSNR); (iii) structural similarity index (SSIM); and (iv) Fuzzy Jaccard for reconstruction quantification, while the balanced accuracy (BA), F1 score (F1), precision P, recall R, Matthews correlation coefficient (MCC), and confusion matrix are reported for the classification task quantification.

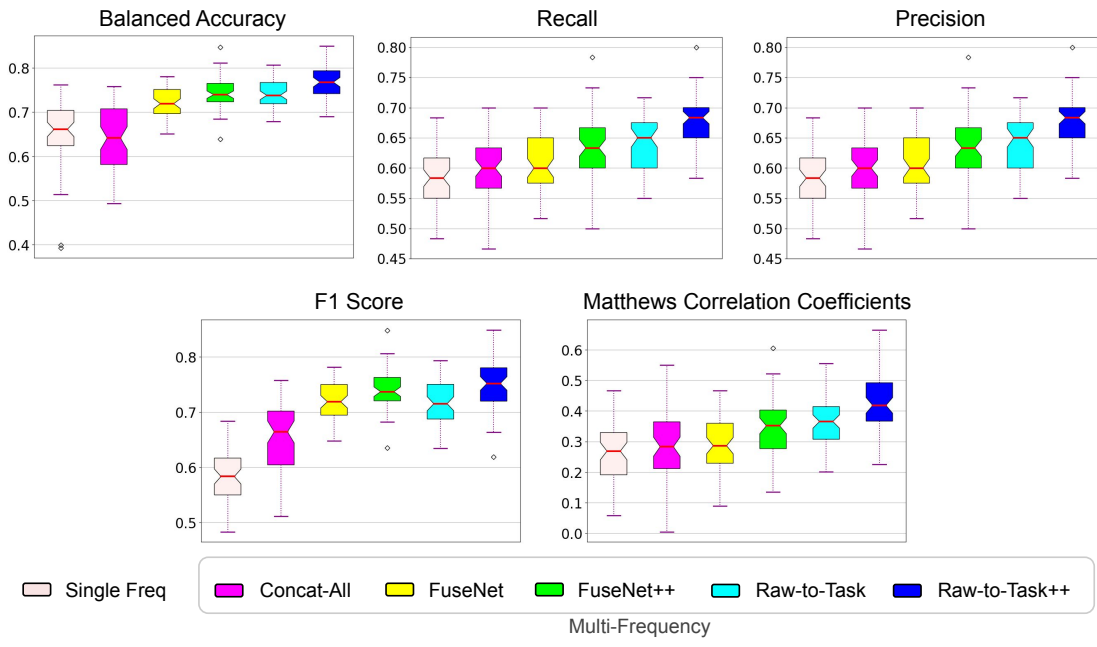
$$\begin{aligned}
 BA &= \frac{1}{2} \left( \frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \\
 P &= \frac{TP}{(TP + FP)}, \\
 R &= \frac{TP}{(TP + FN)}, \\
 F1 &= 2 \frac{P * R}{P + R} \\
 MCC &= \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}
 \end{aligned} \tag{11}$$

True positive (TP) is the number of correctly predicted samples as positive, while false positive (FP) is the number of wrongly predicted samples as positive. False negative (FN) is

<sup>2</sup>Farid Golnaraghi, Study Approval Feb 08-2017, Fraser Health, Certificate of Ethical Approval for Harmonized Minimal Risk Clinical Study, FHREB 2014-065



**Fig. 4. Qualitative reconstruction performance of absorption coefficients using the FuseNet++ on in silico samples with varying ground truth lesion sizes, locations, and numbers. Our multi-spectral results ( $R_{Fusion}$ ) show an overall superiority in terms of generally improved background/foreground contrast and a better differentiation between lesion sizes and lesion localization compared to per-wavelength reconstruction results ( $R_1$  to  $R_4$ ) at wavelengths 690, 750, 800, and 850 nm, respectively.**



**Fig. 5. Quantitative diagnosis performance of different models when one vs multi-wavelength are used. Overall results show improved prediction performances in multi-wavelength models. Note the significant improvement when FuseNet is used compared to a simple concatenation (Concat-All). Results using the FuseNet++ enforce the benefit of feature space orthogonality. Raw-to-task++, in which all network capacity is dedicated to the end task, shows an overall performance gain.**

**Table 8. Quantitative results on in silico test dataset.** Losses are defined in Section 2-B; Loc.error: lesion localization error; PSNR: peak signal-to-noise ratio; SSIM: structural similarity index; BA: balanced accuracy; F1: F1 score. †: value not supported by method, ‡: image reconstruction skipped.

	Loss			Loc. Error (pixel, ↓)	PSNR (dB, ↑)	SSIM (↑)	Fuzzy Jaccard (↑)	Runtime (ms, ↓)	BA ↑	F1 ↑
	$L_{REC}$	$L_{CE}$	$L_{OPL}$							
Single-Wavelength	✓	✓	†	17.7 ± 21.9	19.1 ± 4.8	0.80 ± 0.05	0.60 ± 0.17	23	0.65	0.65
Concat-All	✓	✓	†	20.4 ± 18.4	19.6 ± 6.2	0.73 ± 0.17	0.61 ± 0.18	28	0.63	0.65
FuseNet	✓	✓	-	17.6 ± 23.3	20.2 ± 4.1	0.88 ± 0.05	0.62 ± 0.19	31	0.72	0.72
FuseNet++	✓	✓	✓	<b>15.7 ± 12.7</b>	<b>21.2 ± 4.4</b>	<b>0.89 ± 0.03</b>	<b>0.64 ± 0.18</b>	32	0.74	0.74
Raw-to-Task	†	✓	-			‡		<b>15</b>	0.74	0.72
Raw-to-Task++	†	✓	✓			‡		<b>15</b>	<b>0.77</b>	<b>0.75</b>

the number of wrongly predicted samples as negative, while true negative (TN) is the number of correctly predicted negative class samples over the number of classes in the prediction tasks. Recall quantifies the number of positive class samples properly identified by the model, while precision measures the number of correct positive predictions made by the model. BA, used when quantifying performance on imbalanced data, measures the average accuracy obtained from all classes. MCC measures the quality of multi-class classifications and is informative in cases of skewed class distributions.

For the computational cost at inference, we quantify the forward pass of the model, measured in ms per example.

To evaluate the performance of our models, we contrast the results when using one wavelength with many wavelengths in the FuseNet and the Raw-to-task model. We present results on in-silico data and clinical data.

### 3.2. Results on Synthetic Data

Trained on the in silico data and tested on a separate test set of 240 images, we compare the reconstruction and prediction performance of our FuseNet and the prediction performance with the Raw-to-Task counterpart.

#### 3.2.1. Joint reconstruction and diagnosis

Figure 4, illustrates reconstruction results on selected in silico samples with different lesion sizes, numbers, locations, and depths. In order to offer clinicians more details, results based on each wavelength separately ( $R_i$ ) as well as results that use all wavelengths are shown, with the latter showing more consistent performance. The joint model successfully exploits the presence of the different wavelengths and generally shows an improved background/foreground contrast. For example, the difference in signature for 3 small but proximate lesions is marked in different wavelength results ( $R_1$  to  $R_4$ ) (row c), while a more accurately reconstructed sphere size is provided by the fusion result  $R_{\text{Fusion}}$  in row (d). Detecting heterogeneity in lesions is critical for correct treatment estimation given that it is a proxy indicator of evolutionary pressure in the lesion, selecting for more resistant cancer sub-populations. Table 8 presents the quantitative results of the ablation study, where the contribution of different losses and modular choices of the architecture to model performance are quantified. Rows 1 to 4 highlight the benefit of using multi-wavelength fusion on the reconstruction task. A naive multiple wavelengths concatenation will not necessarily improve results, which agrees with the findings reported by [48], illustrating the impact of adding noisy

dimensions on performances. Nonetheless, we see improved results for FuseNet. When fusion branch and  $L_{OPL}$  are used jointly (FuseNet++), the features contribution from each wavelength is maximized in contrast to simple features concatenation (Concat-All) at the price of a minimal computational increase (only 9%).

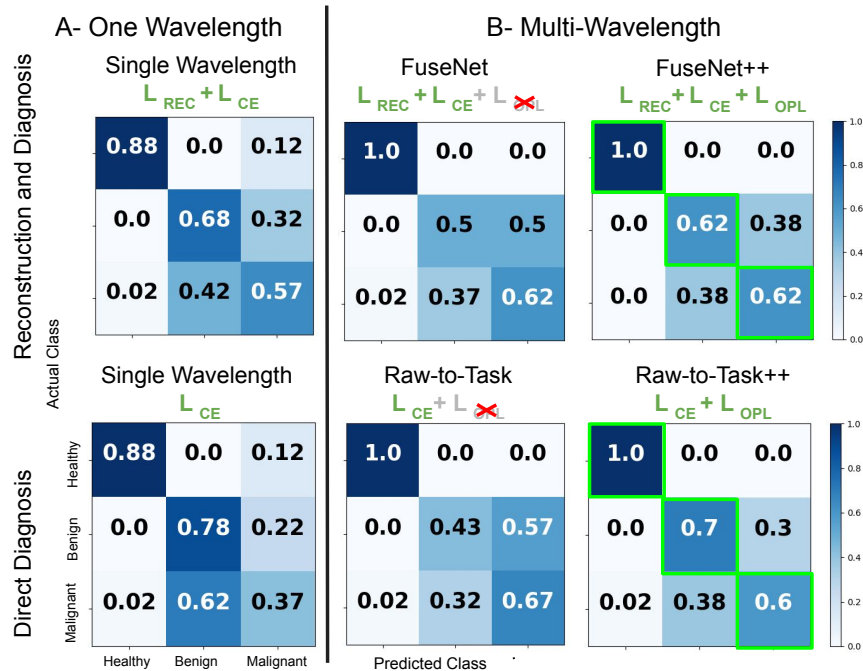
Prediction performance highlighted in Table 8 and Fig. 5 show an overall improvement when more input wavelengths are available, with a boost in performance when FuseNet and FuseNet++ are used. Confusion matrices (Fig. 6-A,B) show a clear discrimination between healthy and lesion features when more data, in the form of more wavelengths, is available. Further, improved benign and malignant discrimination is observed when feature orthogonality is leveraged (Fig. 6-B) as well as a reduction in healthy false negative.

#### 3.2.2. Direct prediction

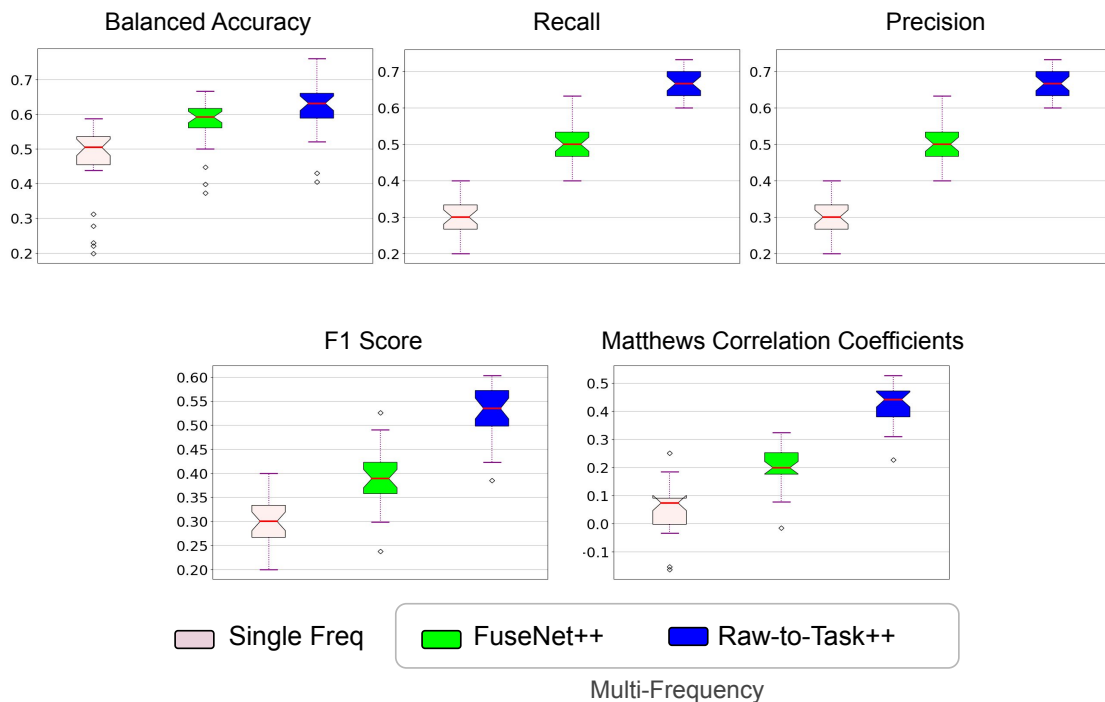
In Figure 6, similarly to the joint model, the direct prediction model results using a single wavelength input (SW-DP) (Fig. 6-A) are contrasted with raw-to-task prediction results using multiple wavelengths as input (Fig. 6-B). A clear discrimination between features is apparent when more data, in the form of multiple wavelengths, is available, especially when discriminating between healthy and lesion; the primary application in DOT-based screening deployments. Raw-to-task model significantly reduces computational complexity (Table 8-Runtime), enabling lower latency and higher throughput in real medical settings. Next, we tested the contribution of individual loss function terms and architecture component on overall diagnosis performance. Figure 5 shows the diagnosis performance on the test set for the best value of  $\gamma$  and highlights the benefits of the feature orthogonality constraint in breast cancer diagnosis, where tumoral and non-tumoral breast lesion differentiation is challenging. Contrasting FuseNet++ and Raw-to-task++ (Fig. 5-6) illustrates performance gain when all network capacity is dedicated to the end task rather than intermediate ones.

### 3.3. Results on Clinical Data

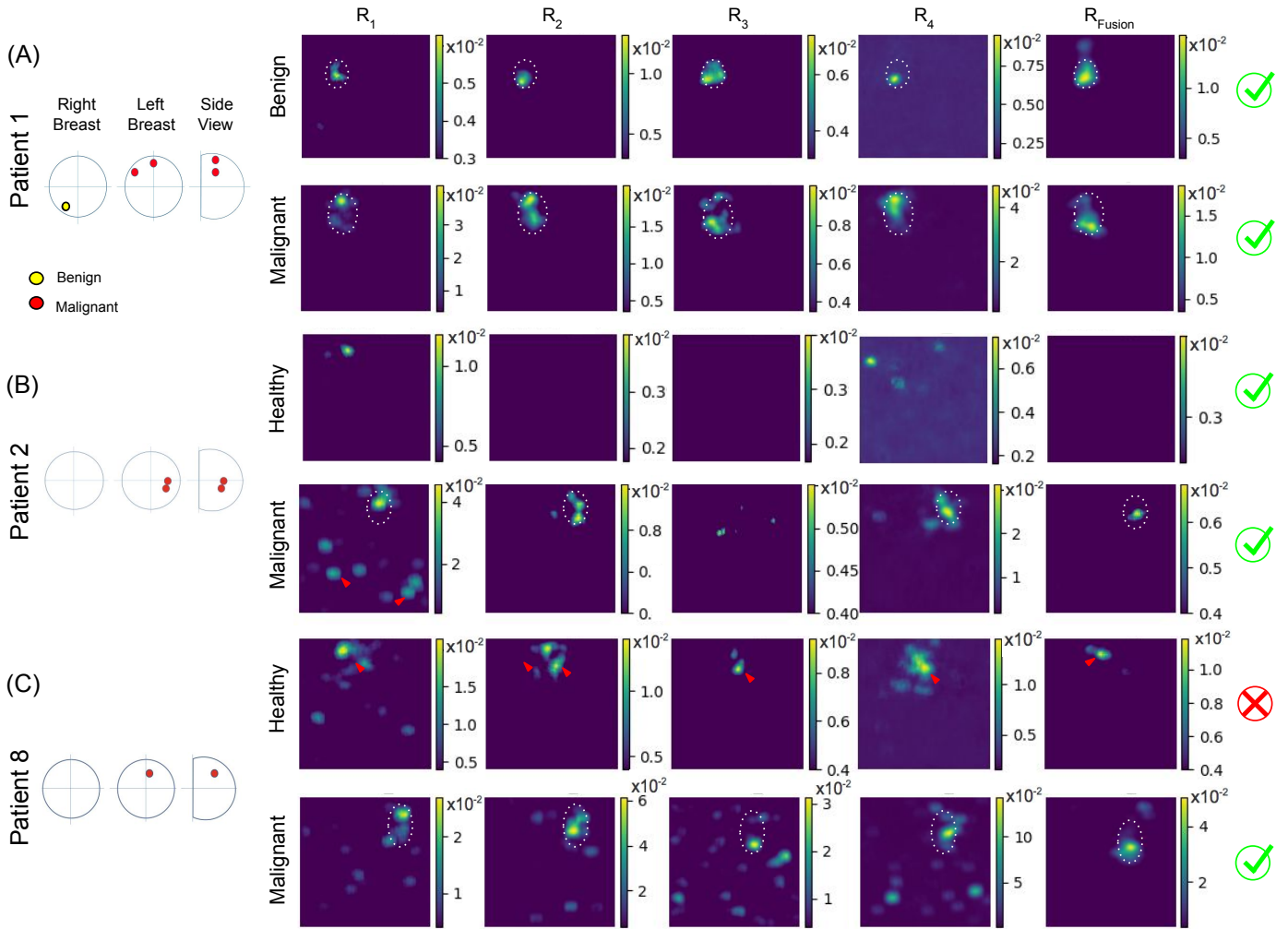
We present qualitative reconstruction performance in Fig. 8 (and Appendix Figure A. 1 for all patients results), and quantitative performance, Fig. 7, Fig. 9, and Table 9, on breast scans from patients diagnosed with breast tumors. Results are computed on all patients, with lesion details summarized in Table 6. The probe is placed close to the likely location of each identified lesion, and a set of scans are made. The opposite healthy breast,



**Fig. 6.** Diagnosis prediction confusion matrices when (A) one vs (B) multi-wavelength inputs are used. Note the improvement in accuracy of unbiased lesion classification (benign, malign) vs healthy when multiple wavelengths are used, as illustrated by the higher values along the diagonal. Results of FuseNet++ highlight the benefit of encouraging orthogonality in enhancing benign vs malignant separability while reducing healthy false negative. Raw-to-task++ further improves separability at the expense of minimal false negative (2%).



**Fig. 7.** Quantitative diagnosis performance when one vs multi-wavelength are used on clinical dataset. Overall results show improved prediction performances in multi-wavelength models compared to single wavelength and indicate that image reconstruction is not necessary for unbiased classification and can even lead to biased results. Note the marked improvement when Raw-to-task++ is used compared to FuseNet++.



**Fig. 8. Qualitative reconstruction results in clinical patients with benign and malignant tumors. Approximate lesion sizes and locations were obtained with joint modalities (details in Table 6). Note, in (A-C), the ability of FuseNet++ to reconstruct lesions. In (B), the robustness of orthogonal fusion to noise ( $R_{Fusion}$ ) compared to ( $R_1$  to  $R_4$ ) (healthy row) is highlighted. (C) highlights a false positive reconstruction case, marked with red triangles, that remain less critical than false negative cases where a tumor is missed. Note the noisy reconstruction in  $R_1$  to  $R_4$ , suggesting a quite noisy input signal. Comprehensive reconstruction results for all patients can be found in Appendix file A.**

for each patient, is scanned as a contrastive reference. Weak labels were attributed to each set of scans regardless of the probe’s closeness to the tumor localization. The distribution of healthy, benign, and malignant lesions in the clinical data is documented in Table 9 (number of scans column). As a partial ground truth, patients underwent mammography and/or Ultrasound scans to obtain estimated lesion dimensions and biopsies to confirm tumor type. While lesions are accurately reconstructed in most cases, as shown in Fig. 8, with clear foreground and background discrimination in  $R_{\text{Fusion}}$  as well as  $R_1$  to  $R_4$ , healthy cases, capturing only background readings, highlight a better robustness of orthogonal fusion,  $R_{\text{Fusion}}$ , to noise.

Figure 7 reports quantitative prediction performance on single vs. multi-wavelength data and highlights the overall improved performance when more wavelengths are used. Note the biased classification results when image reconstruction supervises the prediction task, FuseNet++, compared to direct prediction from raw data, Raw-to-task++.

The confusion matrix, Fig. 9, shows improved discrimination between healthy and lesion features with the raw-to-task model. If we consider that a key feature of the reconstruction based classification is the interpretable angle of such results, we note that the raw-to-task model has the added advantage, in addition to improved performances, that it omits potentially confounding explanations, where reconstruction artifacts can mislead experts. Indeed, in recent work on explainable artificial intelligence, such confounding explainers were identified as a roadblock [84]. Table 9 reports raw-to-task model diagnosis performances on each data class to highlight the clinical data imbalance compared to a balanced training data scheme.

In Figure 8 and Appendix Figure A. 1, we illustrate some failure cases at the limit of detection capability, with false positives (Fig. 8-C and Appendix Figure A. 1-G,H, marked with red triangles) and false negatives (Appendix Figure A. 1-I). For screening purposes, false negatives are more critical; false positives would eventually be resolved by follow-up diagnosis. Note that the discriminatory power of the detection is limited by tumor depth, shape, and noise level. It may require several scans over breast tissue in order to be captured. The failure cases here are from a single scan measurement only, not aggregated.

Although the transfer learning network, trained using phantom data, bridges to some extent the disparity between in silico (training) data and real-world data, its performance on clinical data reveals that it can still be misled by significant real-world variations, such as differences in illumination and noise levels. Additionally, since each tumor is unique, tumor heterogeneity can result in distinct acquisition signatures that may not be presented in the training data. These failure cases highlight the need for more clinical data (patient data) to better train the transfer learning module. Current results present a proof of concept, where validation on larger and more diverse datasets is still required.

### 3.4. Effect of Lesion Localization on Accuracy

We quantify the effect of lesion location on lesion detection accuracy in Figure 10, where we classify whether a lesion is

**Table 9. Quantitative results on clinical dataset using Raw-to-Task++**

	Precision	Recall	F1-score	Number of scans
Healthy	0.71	0.65	0.68	32
Benign	0.11	0.5	0.18	2
Malignant	0.78	0.68	0.73	44
Weighted-Avg	0.73	0.66	0.69	78

present or not. The penetration depth into breast tissue is approximately half the distance between the source and detectors [85],  $\sim 2.5$  cm for our DOT probe. Our results confirm the expected reduction in lesion detection accuracy as the lesions decrease in size or increase in depth.

## 4. Discussion

### 4.1. Breast Lesion Characterization

The detection and characterization of breast lesions, such as microcalcifications and masses, play a crucial role in breast cancer screening and diagnosis. Microcalcifications, identified as tiny calcium deposits on mammograms, can serve as important indicators of potential abnormalities in the breast tissue. While microcalcifications may suggest benign conditions, they can also be associated with malignant changes, highlighting the need for careful evaluation and follow-up. On the other hand, breast masses can manifest in various forms, including solid, fluid-filled, or a combination of both. These masses may represent a wide spectrum of pathologies, ranging from benign tumors to malignant lesions. Given the diverse nature of breast masses, accurate characterization is essential for appropriate clinical management and treatment planning. In our study, we utilized absorption coefficients in the in silico data to simulate a broad range of absorption and scattering characteristics reflective of real-world conditions. This approach aims to provide a comprehensive understanding of the tissue properties and enhance the accuracy of our analysis. Our clinical dataset comprises patients presenting with microcalcifications, masses, and benign tissue, reflecting the heterogeneity observed in clinical practice. However, it is important to note that our primary objective was not to differentiate between subtypes of malignant tissues. Instead, our focus was on assessing the method’s capability to discriminate between healthy, malignant, and benign tissue exclusively. By doing so, we aim to guide clinicians in identifying cases of malignancy and directing patients towards appropriate follow-up modalities for further evaluation and management.

### 4.2. Model Performance vs Interpretability

In order to be an effective tool in clinical settings, a clinician’s trust is essential. A combination of good performance, as quantified by accuracy and other metrics, and an interpretable model increases trust. Neither deep learning based reconstruction nor classical iterative algorithms provide a path from pixel to sensor value in a way that a clinician can easily understand. While a reconstructed image may seem to increase interpretability, it is typically not created in an interpretable way

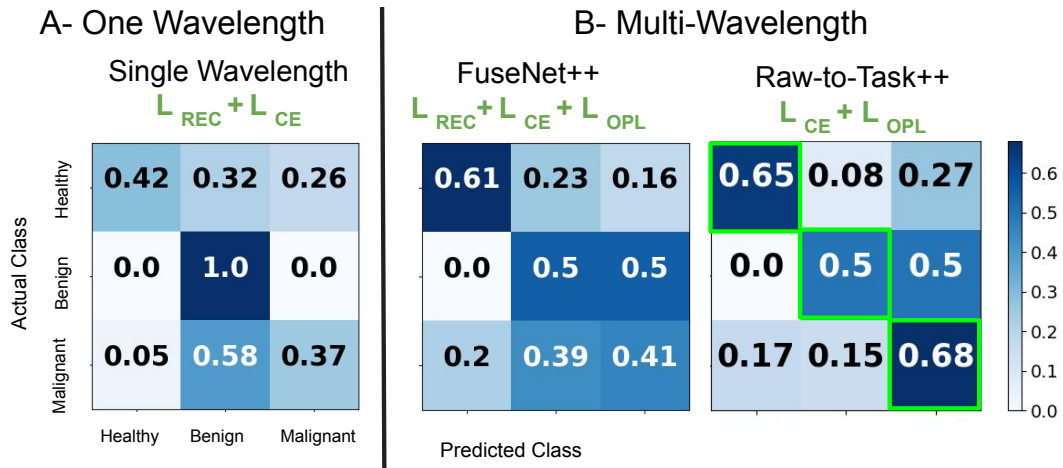


Fig. 9. Clinical data diagnosis prediction confusion matrices when (A) one vs (B) multi-wavelength inputs are used. Note the improvement in accuracy of lesion classification in the Raw-to-Task model, despite the imbalance in the data.

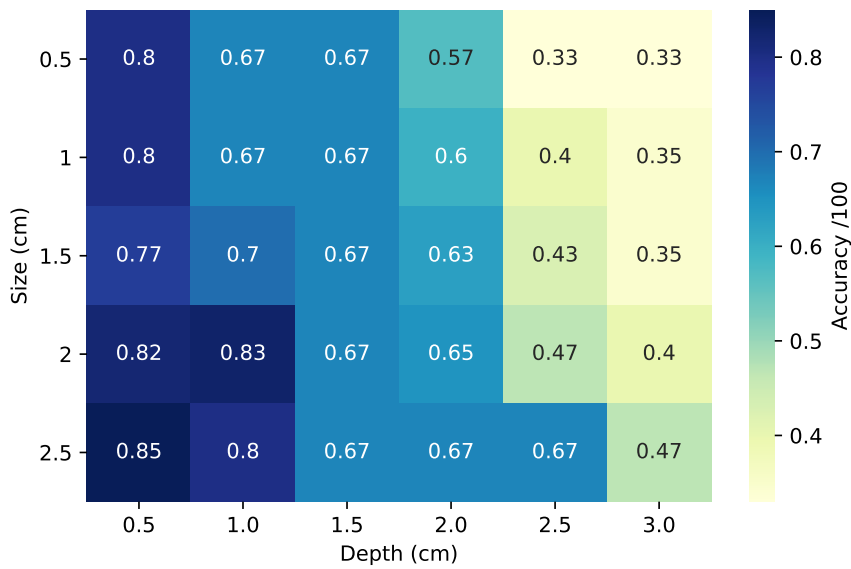


Fig. 10. Effect of lesion depth and radius on model prediction accuracy. Note how the more superficial (closer to the skin surface) and larger lesions are more accurately detected.

and is not necessarily causally related to the classification decision. Omitting the reconstructed image, while increasing performance, would not therefore reduce the trust a clinician has in our direct to task contribution.

To assess the performance of the algorithms, we employed various image quality assessment metrics such as Loc. Error, PSNR, Fuzzy Jaccard, and SSIM, as well as evaluated runtime. Beyond using standard metrics, we emphasized the critical importance of clinical outcomes by focusing on domain-specific metrics: The evaluation of our methods considered patient diagnosis outcomes, which remains the most valuable measure in the medical imaging domain.

#### 4.3. Future Directions and Clinical Implications

Cancer treatment regimens, especially for treatment-resistant lesions, are shifting towards adaptive or dynamic treatment models, such as the recent game theory-driven treatment of resistant prostate tumor patients [86] and treatment decisions guided neoadjuvant chemotherapy [87]. However, these require accurate, unbiased, and specialized task-specific models. Our raw-to-task approach can be extended to develop models specializing in multiple tasks, not just diagnostics. Examples are prediction of lesion type, progression, localization, and tumor heterogeneity, all the way to successful treatment regimens ahead of time, paving the way for adaptive personalized medicine and disease management [75].

We demonstrated the direct improvements, achieved by incorporating the multi-wavelength fusion block and the direct-to-classification approach. Future work may involve integrating these two contributions into other baselines.

A key focus of this work was to leverage orthogonality in mitigating confounding factors induced by multi-wavelength fusion. However, as noted as early as 1936 by [88], orthogonal representations need not be informative, and thus, in a deep learning setting can also lead to orthogonal or independent encodings that are less or uninformative, as we encountered in our own experiments. The heterogeneity of lesions, especially malignant ones, ensures that no two malignant lesions will likely be the same, thus driving the need for diagnostic capability that focuses on identifying the diverse lesion types, not necessarily the reconstructed image.

Our raw-to-task approach offers potential benefits for dynamic treatment models in cancer care, enabling adaptive personalized medicine and disease management. However, it is essential to acknowledge limitations and challenges, including the need for accurate and specialized task-specific models, and the potential sparsity of real-world patient data. Future research directions may explore additional factors, such as lesion type prediction and treatment regimen optimization, to further enhance clinical decision-making in breast cancer diagnosis and management.

## 5. Conclusion

We introduce deep learning-based multi-wavelength orthogonal fusion for diffuse optical tomography with end-to-end classification of malignancy of breast lesions. Orthogonal fusion of

multi-wavelength improved both image reconstruction quality and accuracy of tumoral and non-tumoral breast lesions' discrimination. In addition, we show that raw-to-task learning can improve classification without requiring reconstruction in a real-time setting making it suitable for real-time throughput, crucial for medical settings with constrained power budgets. Our findings suggest that image reconstruction is unnecessary for unbiased lesion classification. The multi-wavelength fusion increased accuracy by 9% and 16% on the synthetic dataset and clinical dataset, respectively, using FuseNet++. Additionally, direct classification further enhanced accuracy by an additional 3% and 8% on the synthetic dataset and clinical dataset, respectively, using the raw-to-task approach.

## Acknowledgments

This study was conducted through the Fraser Health: ethical approval reference number FHREB 2014-065 – PI Dr. Farid Golnaraghi, and supported in part by grants from Natural Sciences and Engineering Research Council of Canada (NSERC) and Simon Fraser University graduate fellowships and graduate dean's entrance scholarship. The authors are grateful for the reviewers' feedback, which greatly improved the manuscript. This research was enabled in part by support provided by Westgrid ([www.westgrid.ca](http://www.westgrid.ca)) and the Digital Research Alliance of Canada ([alliancecan.ca](http://alliancecan.ca)).

## References

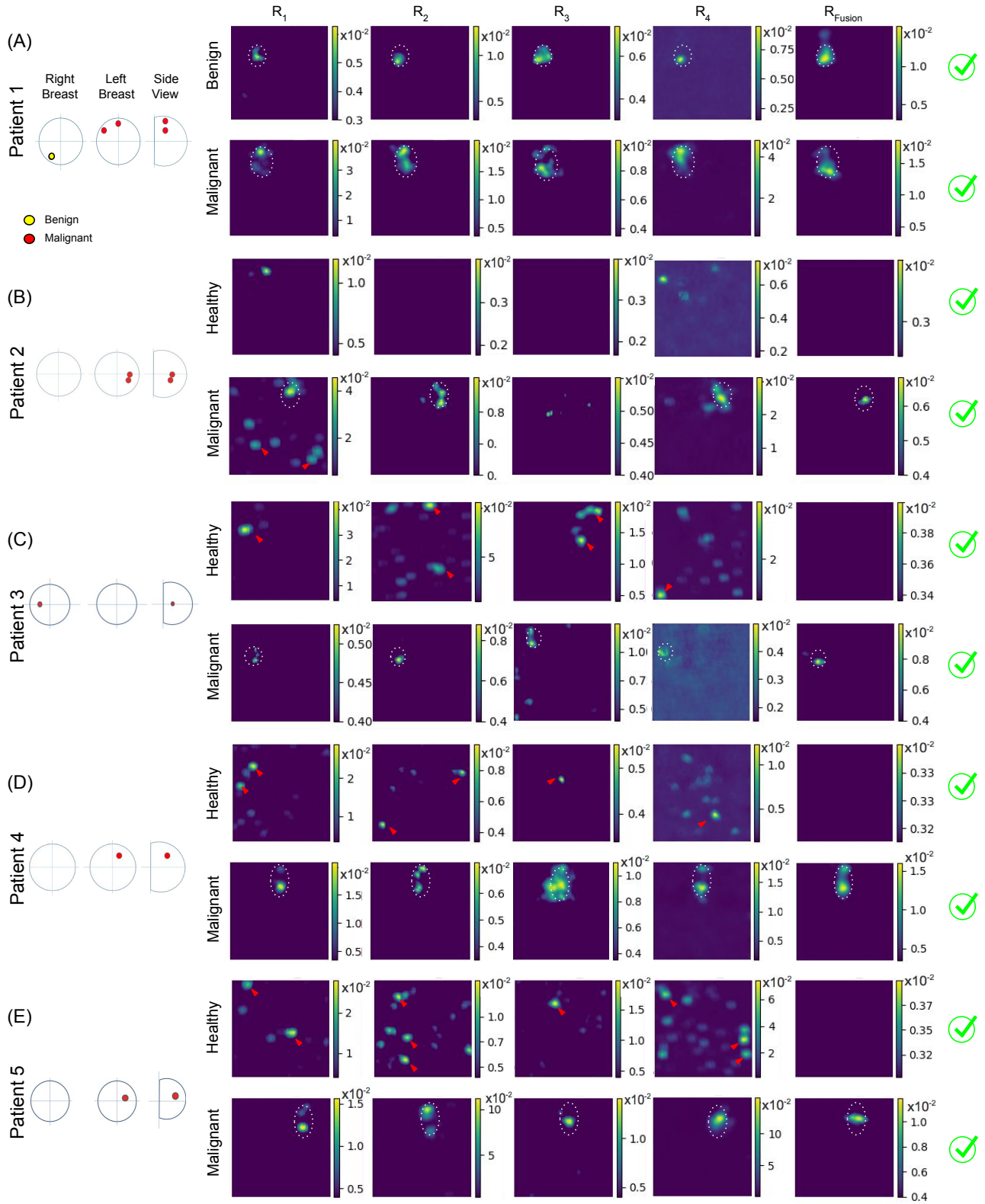
- [1] M. Akram, M. Iqbal, M. Daniyal, A. U. Khan, Awareness and current knowledge of breast cancer, *Biological research* 50 (2017) 1–23.
- [2] C. Coleman, Early detection and screening for breast cancer, in: *Seminars in oncology nursing*, Vol. 33, Elsevier, 2017, pp. 141–155.
- [3] A. G. Waks, E. P. Winer, Breast cancer treatment: a review, *Jama* 321 (3) (2019) 288–300.
- [4] M. K. Shetty, Screening for breast cancer with mammography: current status and an overview, *Indian journal of surgical oncology* 1 (2010) 218–223.
- [5] S. Iranmakani, T. Mortezaadeh, F. Sajadian, M. F. Ghaziani, A. Ghafari, D. Khezerloo, et al., A review of various modalities in breast imaging: technical aspects and clinical outcomes, *Egyptian Journal of Radiology and Nuclear Medicine* 51 (1) (2020) 1–22.
- [6] A. Dalla Mora, D. Contini, S. Arridge, F. Martelli, A. Tosi, G. Boso, et al., Towards next-generation time-domain diffuse optics for extreme depth penetration and sensitivity, *Biomedical optics express* 6 (5) (2015) 1749–1760.
- [7] M. Applegate, R. Istfan, S. Spink, A. Tank, D. Roblyer, Recent advances in high speed diffuse optical imaging in biomedicine, *APL Photonics* 5 (4) (2020) 040802.
- [8] C. Chen, F. Tian, H. Liu, J. Huang, Diffuse optical tomography enhanced by clustered sparsity for functional brain imaging, *IEEE transactions on medical imaging* 33 (12) (2014) 2323–2331.
- [9] H. Zhao, E. M. othersand Frijia, E. V. Rosas, L. Collins-Jones, G. Smith, R. Nixon-Hill, et al., Design and validation of a mechanically flexible and ultra-lightweight high-density diffuse optical tomography system for functional neuroimaging of newborns, *Neurophotonics* 8 (1) (2021) 015011.
- [10] M. Shokoufi, F. Golnaraghi, Handheld diffuse optical breast scanner probe for cross-sectional imaging of breast tissue, *Journal of Innovative Optical Health Sciences* 12 (02) (2019) 1950008.
- [11] M. L. Altoe, A. Marone, H. K. Kim, K. Kalinsky, D. L. Hershman, A. H. Hielscher, et al., Diffuse optical tomography of the breast: a potential modifiable biomarker of breast cancer risk with neoadjuvant chemotherapy, *Biomedical Optics Express* 10 (8) (2019) 4305–4315.

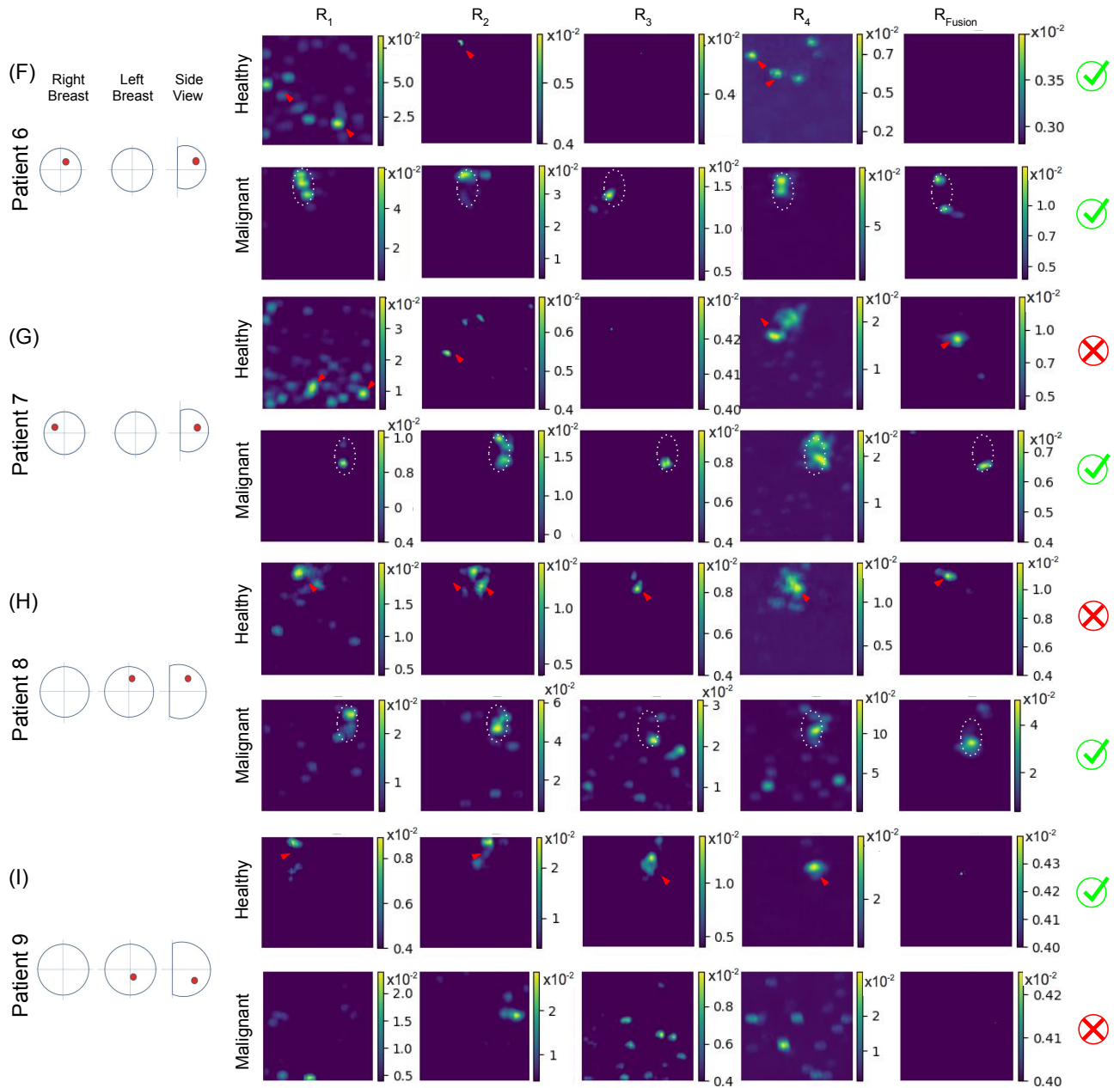


- [12] S. Mahdy, O. Hamdy, M. A. Hassan, M. A. Eldosoky, A modified source-detector configuration for the discrimination between normal and diseased human breast based on the continuous-wave diffuse optical imaging approach: a simulation study, *Lasers in Medical Science* (2021) 1–10.
- [13] T. Durduran, R. Choe, W. B. Baker, A. G. Yodh, Diffuse optics for tissue monitoring and tomography, *Reports on progress in physics* 73 (7) (2010) 076701.
- [14] P. Taroni, A. M. Paganoni, F. Ieva, A. Pifferi, G. Quarto, F. Abbate, et al., Non-invasive optical estimate of tissue composition to differentiate malignant from benign breast lesions: A pilot study, *Scientific reports* 7 (1) (2017) 1–11.
- [15] M. Momtahan, S. Momtahan, R. Remaseshan, F. Golnaraghi, Early detection of breast cancer using diffuse optical probe and ensemble learning method, in: *2023 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO)*, IEEE, 2023, pp. 139–142.
- [16] F. S. Azar, K. Lee, A. Khamene, R. Choe, A. Corlu, S. D. Konecky, et al., Standardized platform for coregistration of nonconcurrent diffuse optical and magnetic resonance breast images obtained in different geometries, *Journal of biomedical optics* 12 (5) (2007) 051902–051902.
- [17] H. Ben Yedder, B. Cardoen, M. Shokoufi, F. Golnaraghi, G. Hamarneh, Multitask deep learning reconstruction and localization of lesions in limited angle diffuse optical tomography, *IEEE Transactions on Medical Imaging* 41 (3) (2021) 515–530.
- [18] Y. Hoshi, Y. Yamada, Overview of diffuse optical tomography and its clinical applications, *Journal of Biomedical Optics* 21 (9) (2016) 091312.
- [19] S. R. Arridge, J. C. Schotland, Optical tomography: forward and inverse problems, *Inverse problems* 25 (12) (2009) 123010.
- [20] F. Beca, K. Polyak, Intratumor heterogeneity in breast cancer, in: *Novel Biomarkers in the Continuum of Breast Cancer*, Springer, 2016, pp. 169–189.
- [21] A. Aspri, A. Benfenati, P. Causin, C. Cavaterra, G. Naldi, Mathematical and numerical challenges in diffuse optical tomography inverse problems, *Discrete and Continuous Dynamical Systems-S* (2023) 0–0.
- [22] S. Sabir, S. Cho, Y. Kim, R. Pua, D. Heo, K. H. Kim, et al., Convolutional neural network-based approach to estimate bulk optical properties in diffuse optical tomography, *Applied Optics* 59 (5) (2020) 1461–1470.
- [23] H. Ben Yedder, A. BenTaieb, M. Shokoufi, A. Zahiremami, F. Golnaraghi, G. Hamarneh, Deep learning based image reconstruction for diffuse optical tomography, in: *International Workshop on Machine Learning for Medical Image Reconstruction*, Springer, 2018, pp. 112–119.
- [24] N. I. Nizam, M. Ochoa, J. T. Smith, X. Intes, Wide-field diffuse optical tomography using deep learning, in: *Optical Tomography and Spectroscopy*, Optica Publishing Group, 2022, pp. OW4D–7.
- [25] H. Ben Yedder, M. Shokoufi, B. Cardoen, F. Golnaraghi, G. Hamarneh, Limited-angle diffuse optical tomography image reconstruction using deep learning, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 66–74.
- [26] M. Mozumder, A. Hauptmann, I. Nissilä, S. R. Arridge, T. Tarvainen, A model-based iterative learning approach for diffuse optical tomography, *IEEE Transactions on Medical Imaging* 41 (5) (2021) 1289–1299.
- [27] M. Mozumder, A. Hauptmann, S. R. Arridge, T. Tarvainen, Diffuse optical tomography utilizing model-based learning, in: *Optics and the Brain*, Optica Publishing Group, 2022, pp. JTU3A–10.
- [28] N. I. Nizam, M. Ochoa, J. T. Smith, X. Intes, Deep learning-based fusion of widefield diffuse optical tomography and micro-ct structural priors for accurate 3d reconstructions, *Biomedical Optics Express* 14 (3) (2023) 1041–1053.
- [29] J. Yoo, S. Sabir, D. Heo, K. H. Kim, A. Wahab, Y. Choi, et al., Deep learning diffuse optical tomography, *IEEE Transactions on medical imaging* 39 (4) (2019) 877–887.
- [30] X. Fang, C. Gao, Y. Li, T. Li, Solving heterogenous region for diffuse optical tomography with a convolutional forward calculation model and the inverse neural network, in: *Advanced Optical Imaging Technologies III*, Vol. 11549, SPIE, 2020, pp. 50–60.
- [31] Y. Zhao, A. Raghuram, F. Wang, S. H. Kim, A. H. Hielscher, J. T. Robinson, et al., Unrolled-dot: an interpretable deep network for diffuse optical tomography, *Journal of Biomedical Optics* 28 (3) (2023) 036002.
- [32] Y. Zou, Y. Zeng, S. Li, Q. Zhu, Machine learning model with physical constraints for diffuse optical tomography, *Biomedical Optics Express* 12 (9) (2021) 5720–5735.
- [33] S. Li, M. Zhang, M. Xue, Q. Zhu, Difference imaging from single measurements in diffuse optical tomography: a deep learning approach, *Journal of Biomedical Optics* 27 (8) (2022) 086003.
- [34] Y. Zou, Y. Zeng, S. Li, Q. Zhu, Unsupervised machine learning model for dot reconstruction, in: *Optical Tomography and Spectroscopy of Tissue XIV*, Vol. 11639, SPIE, 2021, pp. 23–36.
- [35] W. Ren, S. Shen, L. Li, S. Gao, Y. Wang, L. Gu, S. Li, X. Zhu, J. Jiang, J. Yu, High-resolution tomographic reconstruction of optical absorbance through scattering media using neural fields, *arXiv preprint arXiv:2304.01682* (2023).
- [36] N. Murad, M.-C. Pan, Y.-F. Hsu, Optimizing diffuse optical imaging for breast tissues with a dual-encoder neural network to preserve small structural information and fine features, *Journal of Medical Imaging* 10 (6) (2023) 066003–066003.
- [37] L. Zhang, G. Zhang, Brief review on learning-based methods for optical tomography, *Journal of Innovative Optical Health Sciences* 12 (06) (2019) 1930011.
- [38] H. Ben Yedder, B. Cardoen, G. Hamarneh, Deep learning for biomedical image reconstruction: A survey, *Artificial Intelligence Review* (2020) 1–33.
- [39] G. M. Balasubramaniam, B. Wiesel, N. Biton, R. Kumar, J. Kupferman, S. Arnon, Tutorial on the use of deep learning in diffuse optical tomography, *Electronics* 11 (3) (2022) 305.
- [40] J. T. Smith, M. Ochoa, D. Faulkner, G. Haskins, X. Intes, Deep learning in macroscopic diffuse optical imaging, *Journal of Biomedical Optics* 27 (2) (2022) 020901–020901.
- [41] A. Godavarty, S. Rodriguez, Y.-J. Jung, S. Gonzalez, Optical imaging for breast cancer prescreening, *Breast Cancer: Targets and Therapy* 7 (2015) 193.
- [42] N. Z. Shifa, M. M. R. Sayem, M. A. Islam, Improved image reconstruction using multi frequency data for diffuse optical tomography, in: *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, IEEE, 2021, pp. 264–268.
- [43] M. Doulgarakis, A. T. Eggebrecht, H. Dehghani, High-density functional diffuse optical tomography based on frequency-domain measurements improves image quality and spatial resolution, *Neurophotonics* 6 (3) (2019) 035007.
- [44] M. B. Unlu, O. Birgul, R. Shafiqi, G. Gulsen, O. Nalcioglu, Diffuse optical tomographic reconstruction using multifrequency data, *Journal of Biomedical Optics* 11 (5) (2006) 054008.
- [45] H. K. Kim, U. J. Netz, J. Beuthan, A. H. Hielscher, Optimal source-modulation frequencies for transport-theory-based optical tomography of small-tissue volumes, *Optics express* 16 (22) (2008) 18082–18101.
- [46] X. Intes, B. Chance, Multi-frequency diffuse optical tomography, *Journal of Modern Optics* 52 (15) (2005) 2139–2159.
- [47] V. Mudeng, W. Nisa, S. S. Suprpto, Computational image reconstruction for multi-frequency diffuse optical tomography, *Journal of King Saud University-Computer and Information Sciences* (2021).
- [48] M. B. Applegate, C. A. Gómez, D. Roblyer, Frequency selection in frequency domain diffuse optical spectroscopy, in: *Optical Tomography and Spectroscopy of Tissue XIV*, Vol. 11639, International Society for Optics and Photonics, 2021, p. 116390N.
- [49] C. Chen, V. C. Kavuri, X. Wang, R. Li, H. Liu, J. Huang, Multi-frequency diffuse optical tomography for cancer detection, in: *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2015, pp. 67–70.
- [50] A. Zimek, E. Schubert, H.-P. Kriegel, A survey on unsupervised outlier detection in high-dimensional numerical data, *Statistical Analysis and Data Mining: The ASA Data Science Journal* 5 (5) (2012) 363–387.
- [51] M. Althobaiti, H. Vavadi, Q. Zhu, Diffuse optical tomography reconstruction method using ultrasound images as prior for regularization matrix, *Journal of biomedical optics* 22 (2) (2017) 026002–026002.
- [52] A. Pifferi, A. Dalla Mora, L. Di Sieno, E. Ferocino, A. Tosi, E. Conca, et al., Solus: an innovative multimodal imaging system to improve breast cancer diagnosis through diffuse optics and ultrasounds, in: *Optical Tomography and Spectroscopy of Tissue XIV*, Vol. 11639, International Society for Optics and Photonics, 2021, p. 116390C.
- [53] G. Di Sciaccia, G. Maffei, A. Farina, A. Dalla Mora, A. Pifferi, P. Taroni, et al., Evaluation of a pipeline for simulation, reconstruction, and classification in ultrasound-aided diffuse optical tomography of breast tumors, *Journal of biomedical optics* 27 (3) (2022) 036003.
- [54] M. Zhang, M. Xue, S. Li, Y. Zou, Q. Zhu, Fusion deep learning approach combining diffuse optical tomography and ultrasound for improv-

- ing breast cancer classification, *Biomedical Optics Express* 14 (4) (2023) 1636–1646.
- [55] D. Liu, Y. Zhang, L. Bai, P. Zhang, F. Gao, Combining two-layer semi-three-dimensional reconstruction and multi-wavelength image fusion for functional diffuse optical tomography, *IEEE Transactions on Computational Imaging* 7 (2021) 1055–1068.
- [56] H. O. Kazanci, O. Oral, Frequency shifting model for diffuse optical tomography, *Optical and Quantum Electronics* 53 (11) (2021) 1–6.
- [57] G. A. Perkins, A. T. Eggebrecht, H. Dehghani, Multi-modulated frequency domain high density diffuse optical tomography, *Biomedical Optics Express* 13 (10) (2022) 5275–5294.
- [58] Y. Zhang, D. Sidibé, O. Morel, F. Mériaudeau, Deep multimodal fusion for semantic image segmentation: A survey, *Image and Vision Computing* 105 (2021) 104042.
- [59] W. Guo, J. Wang, S. Wang, Deep multimodal representation learning: A survey, *IEEE Access* 7 (2019) 63373–63394.
- [60] G. Patel, J. Dolz, Weakly supervised segmentation with cross-modality equivariant constraints., *Medical Image Analysis* (2022) 102374.
- [61] A. Nagrani, S. Yang, A. Arnab, A. Jansen, C. Schmid, C. Sun, Attention bottlenecks for multimodal fusion, *Advances in Neural Information Processing Systems* 34 (2021).
- [62] R. J. Chen, M. Y. Lu, J. Wang, D. F. Williamson, S. J. Rodig, N. I. Lindeman, et al., Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis, *IEEE Transactions on Medical Imaging* (2020).
- [63] A. Bozic, P. Palafox, J. Thies, A. Dai, M. Nießner, Transformerfusion: Monocular rgb scene reconstruction using transformers, *Advances in Neural Information Processing Systems* 34 (2021).
- [64] A. Bardes, J. Ponce, Y. LeCun, VICReg: Variance-invariance-covariance regularization for self-supervised learning, *arXiv preprint arXiv:2105.04906* (2021).
- [65] J. Zbontar, L. Jing, I. Misra, Y. LeCun, S. Deny, Barlow twins: Self-supervised learning via redundancy reduction, in: *International Conference on Machine Learning*, PMLR, 2021, pp. 12310–12320.
- [66] N. Bansal, X. Chen, Z. Wang, Can we gain more from orthogonality regularizations in training deep networks?, *Advances in Neural Information Processing Systems* 31 (2018).
- [67] L. Huang, X. Liu, B. Lang, A. W. Yu, Y. Wang, B. Li, Orthogonal weight normalization: Solution to optimization over multiple dependent stiefel manifolds in deep neural networks, in: *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [68] K. Ranasinghe, M. Naseer, M. Hayat, S. Khan, F. S. Khan, Orthogonal projection loss, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12333–12343.
- [69] N. Braman, J. W. Gordon, E. T. Goossens, C. Willis, M. C. Stumpe, J. Venkataraman, Deep orthogonal fusion: Multimodal prognostic biomarker discovery integrating radiology, pathology, genomic, and clinical data, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2021, pp. 667–677.
- [70] Z. Chen, Y. Liu, Y. Zhang, Q. Li, A. D. N. Initiative, et al., Orthogonal latent space learning with feature weighting and graph learning for multimodal alzheimer’s disease diagnosis, *Medical Image Analysis* 84 (2023) 102698.
- [71] X. Zhen, S. Li, Towards direct medical image analysis without segmentation, *arXiv preprint arXiv:1510.06375* (2015).
- [72] D. Wu, K. Kim, B. Dong, Q. Li, End-to-end abnormality detection in medical imaging, *OpenReview* (2018).
- [73] M. A. Hussain, G. Hamarneh, T. W. O’Connell, M. F. Mohammed, R. Abugharbieh, Segmentation-free estimation of kidney volumes in ct with dual regression forests, in: *International Workshop on Machine Learning in Medical Imaging*, Springer, 2016, pp. 156–163.
- [74] S. A. Taghanaki, N. Duggan, H. Ma, X. Hou, A. Celler, F. Benard, et al., Segmentation-free direct tumor volume and metabolic activity estimation from pet scans, *Computerized Medical Imaging and Graphics* 63 (2018) 52–66.
- [75] K. Abhishek, J. Kawahara, G. Hamarneh, Predicting the clinical management of skin lesions using deep learning, *Scientific reports* 11 (1) (2021) 1–14.
- [76] A. B. Kononov, E. A. Genina, A. N. Bashkatov, Diffuse optical mammotomography: state-of-the-art and prospects, *Journal of Biomedical Photonics & Engineering* 2 (2) (2016) 020202–1.
- [77] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He, A comprehensive survey on transfer learning, *Proceedings of the IEEE* 109 (1) (2020) 43–76.
- [78] X. Yu, J. Wang, Q.-Q. Hong, R. Teku, S.-H. Wang, Y.-D. Zhang, Transfer learning for medical images analyses: A survey, *Neurocomputing* 489 (2022) 230–254.
- [79] J. Feng, Q. Sun, Z. Li, Z. Sun, K. Jia, Back-propagation neural network-based reconstruction algorithm for diffuse optical tomography, *Journal of biomedical optics* 24 (5) (2018) 051407.
- [80] M. Schweiger, S. R. Arridge, The toast++ software suite for forward and inverse modeling in optical tomography, *Journal of biomedical optics* 19 (4) (2014) 040801.
- [81] W. Shao, Y. Du, Microwave imaging by deep learning network: Feasibility and training method, *IEEE transactions on antennas and propagation* 68 (7) (2020) 5626–5635.
- [82] F. Golnaraghi, M. Shokoufi, Handheld probe and system for imaging human tissue, eP3570743A1; EP3570743A4; US11457818B2; US2019328233A1; US2022409058A1; WO2018132908A (Oct. 31 2019).
- [83] M. Shokoufi, Z. Haeri, B. S. Kharb, F. Golnaraghi, Novel handheld diffuse optical spectroscopy probe for breast cancer assessment: Clinical study, *Journal of Biomedical Science* 6 (2017). URL <https://api.semanticscholar.org/CorpusID:4654450>
- [84] M. Jacobs, M. F. Pradier, T. H. McCoy, R. H. Perlis, F. Doshi-Velez, K. Z. Gajos, How machine-learning recommendations influence clinician treatment selections: the example of antidepressant selection, *Translational psychiatry* 11 (1) (2021) 1–9.
- [85] C.-W. Sun, Biophotonics for tissue oxygenation analysis, in: *Biophotonics for Medical Applications*, Elsevier, 2015, pp. 301–320.
- [86] J. West, Y. Ma, P. K. Newton, Capitalizing on competition: An evolutionary model of competitive release in metastatic castration resistant prostate cancer treatment, *Journal of Theoretical Biology* 455 (2018) 249–260.
- [87] S. Momtahan, M. Momtahan, R. Ramaseshan, F. Golnaraghi, An optical sensory system for assessment of residual cancer burden in breast cancer patients undergoing neoadjuvant chemotherapy, *Sensors* 23 (12) (2023) 5761.
- [88] R. A. Fisher, The use of multiple measurements in taxonomic problems, *Annals of eugenics* 7 (2) (1936) 179–188.

**Appendix A**  
**Reconstruction Results Across Clinical Patients**





**Fig. 1.** Qualitative reconstruction results for all clinical patients with both benign and malignant tumors are presented in this study. In panel A, reconstruction from a benign case is highlighted. In panel (B-F), the robustness of orthogonal fusion to noise ( $R_{Fusion}$ ) is notable when compared to ( $R_1$  to  $R_4$ ) in the healthy row. This highlights the superior noise resistance of the orthogonal fusion approach. (G,I) Examples of reconstruction failure cases. (G,H) show some false positive reconstruction cases. However for screening false positives are less critical, as follow-up exam can reveal them, than false negatives where a tumor is missed. It is worth noting the noisy reconstruction observed in  $R_1$  to  $R_4$ , evidencing a relatively noisy input signal.