

Two-stage classification strategy for breast cancer diagnosis using ultrasound-guided diffuse optical tomography and deep learning

Menghao Zhang,^{a,†} Shuying Li,^{b,†} Minghao Xue,^b and Quing Zhu^{a,b,c,*}

^aWashington University in St. Louis, Department of Electrical and Systems Engineering, St. Louis, Missouri, United States

^bWashington University in St. Louis, Department of Biomedical Engineering, St. Louis, Missouri, United States

^cWashington University School of Medicine, Department of Radiology, St. Louis, Missouri, United States

ABSTRACT. **Significance:** Ultrasound (US)-guided diffuse optical tomography (DOT) has demonstrated great potential for breast cancer diagnosis in which real-time or near real-time diagnosis with high accuracy is desired.

Aim: We aim to use US-guided DOT to achieve an automated, fast, and accurate classification of breast lesions.

Approach: We propose a two-stage classification strategy with deep learning. In the first stage, US images and histograms created from DOT perturbation measurements are combined to predict benign lesions. Then the non-benign suspicious lesions are passed through to the second stage, which combine US image features, DOT histogram features, and 3D DOT reconstructed images for final diagnosis.

Results: The first stage alone identified 73.0% of benign cases without image reconstruction. In distinguishing between benign and malignant breast lesions in patient data, the two-stage classification approach achieved an area under the receiver operating characteristic curve of 0.946, outperforming the diagnoses of all single-modality models and of a single-stage classification model that combines all US images, DOT histogram, and imaging features.

Conclusions: The proposed two-stage classification strategy achieves better classification accuracy than single-modality-only models and a single-stage classification model that combines all features. It can potentially distinguish breast cancers from benign lesions in near real-time.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JBO.28.8.086002](https://doi.org/10.1117/1.JBO.28.8.086002)]

Keywords: diffuse optical tomography; ultrasound; breast cancer diagnosis; deep learning; classifier

Paper 230056GRR received Mar. 8, 2023; revised Jul. 29, 2023; accepted Aug. 2, 2023; published Aug. 26, 2023.

1 Introduction

Breast cancer is the most common cancer among women in the United States, with around 2.9 million estimated new cases in 2022.¹ X-ray mammography is the major modality for both screening and diagnosis;^{2,3} however, it has limited sensitivity when used in dense breasts.⁴ For average-risk patients with dense breasts, breast ultrasound (US) has been widely adopted as both a diagnostic tool and supplementary screening tool.^{5,6} Magnetic resonance imaging is also a

*Address all correspondence to Quing Zhu, zhu.q@wustl.edu

†Authors contributed equally to this study.

supplementary tool for screening dense breasts of high-risk women,⁷ but its high cost makes it accessible to only a small patient population. Despite improvements in imaging tools, a low positive predictive value for biopsy recommendation has remained a problem. In 2017, the Breast Cancer Surveillance Consortium reported only a 27.5% positive predictive value for biopsy recommendation.²

In the last two decades, diffuse optical tomography (DOT), which utilizes diffused near-infrared (NIR) light to map tissue optical properties, has been widely explored for non-invasive breast cancer diagnosis.^{8–22} US-guided DOT, utilizing co-registered US images to provide lesion size and depth information for improved DOT reconstruction accuracy, has shown success in breast cancer diagnosis and treatment monitoring.^{23–28} A recent study reported that the adjunctive use of US-guided DOT can reduce benign biopsies by 23.5%.²⁹

Recently, machine learning (ML) has been applied to DOT in bulk tissue optical property estimation,^{30,31} image reconstruction,^{32–39} and breast cancer diagnosis.^{40–43} Using simulated breast lesions, Di Sciacca et al.⁴² applied logistic regression, support vector machines (SVMs), and a fully connected network to reconstruct optical properties and achieved a best accuracy of 78%. Xu et al.⁴³ applied a convolutional neural network (CNN) to 1260 2D optical tomographic images sliced from 3D images collected from 63 women with dense breasts and achieved a 0.95 area under the receiver operating characteristic (ROC) curve (AUC). Inspired by recent multi-model fusion techniques in medical imaging,^{44–46} Zhang et al. combined US features extracted by a modified VGG-11 network with DOT images to form a new neural network for classification, and it achieved an AUC of 0.931 in distinguishing between benign and malignant breast lesions.⁴⁰ However, DOT's relatively slow data processing and image reconstruction speeds have hindered real-time diagnosis. To address this problem, our group has developed a two-stage strategy.⁴¹ The first stage uses breast imaging reporting and data system (BI-RADS) readings from radiologists together with features from DOT frequency domain perturbation measurement data represented in two-dimensional histograms (DOT histograms). These data are combined by a random forest classifier to screen out ~60% of benign lesions without the need of reconstruction. The second stage then utilizes reconstructed DOT images and an SVM classifier to improve the overall diagnostic accuracy. However, this strategy relies on radiologists' readings, which can be unavailable or have limited accuracy in remote or low-resource settings. Additionally, it relies on limited hand-crafted features from DOT measurement data and DOT images, which may not capture all important features.

In this study, we further improve the two-stage classification strategy with deep learning and we automate it to achieve faster and more accurate classification for breast lesions without BI-RADS readings or hand-crafted features. In the first stage, we train two CNN-based classifiers using US images and DOT histograms. The predictions of the CNNs are averaged to single out benign lesions. The rest of the non-benign suspicious lesions pass through the second stage, which combines US image features, DOT histograms, and DOT reconstructed images for diagnosis of this group. This two-stage classification strategy combines the diagnoses of first and second stage classifiers and has achieved the excellent diagnostic performance by filtering out 70% of the benign lesions in near real-time without DOT image reconstruction.

2 Methods

2.1 Data

A total of 254 patients (169 with benign lesions and 85 with malignant lesions) were studied to evaluate the proposed diagnostic strategy. In the benign group, the maximum dimensions of the lesions, measured by US, ranged from 0.31 to 5.30 cm, with a mean of 1.48 cm, and the depths ranged from 0.65 to 3.20 cm, with a mean of 1.50 cm. In the malignant group, the maximum dimensions of the lesions ranged from 0.43 to 5.52 cm, with a mean of 2.05 cm, and the depths ranged from 0.70 to 3.00 cm, with a mean of 1.53 cm. DOT patient data and co-registered US images were acquired with our US-guided DOT systems at four wavelengths in the NIR range.^{24,47} The clinical study was approved by the local Institutional Review Boards and was compliant with the Health Insurance Portability and Accountability Act. Informed consent was obtained from each patient, and data from patients were de-identified. Because of our limited dataset, the results could be influenced by the training-testing splits. Thus, we chose to employ a bootstrapping approach to

Table 1 Simulated lesion and background tissue optical properties in the pre-training dataset.

Target radius (cm)	Target center depth (cm)	Target μ_a (cm ⁻¹)	Target μ_s' (cm ⁻¹)	Background tissue μ_a (cm ⁻¹)	Background tissue μ_s' (cm ⁻¹)
0.375 ~ 1.5	0.7 ~ 2.7	0.06 ~ 0.30	4 ~ 8	FEM: 0.01 ~ 0.06	FEM: 4 ~ 8
MC: 20 digital breast phantoms with fat fractions 20% ~ 80% ^{48,49}					

evaluate the performance of our models by shuffling the dataset 50 times, providing a more robust model performance than using a small testing dataset. To use a balanced dataset, in each of the 50 runs, 120 lesions (60 benign and 60 malignant) were selected as the training set and the other 50 (25 benign and 25 malignant) were used as the testing set. The hyperparameters were tuned based on the averaged performance of the 50 runs. In addition, using the finite element method (FEM) and Monte Carlo (MC) simulations, we generated a total of 880 sets of DOT measurements to pretrain the DOT models before fine-tuning the models with patient data. For FEM simulation, we used COMSOL software to generate the forward measurements. One hemisphere was used to simulate breast tissue, and the other layer was used to simulate the chest wall.³⁷ The VICTRE digital breast phantoms were incorporated with Monte Carlo simulation to simulate heterogeneous breast tissue.^{48,49} The digital phantoms were uniformly compressed numerically in the z -direction to a depth of 5 cm to simulate the compression that occurred when using our handheld reflection-mode DOT probe.⁴⁸ Details of the simulated lesion and background tissue optical properties are in Table 1. For both simulation settings, the probe locations were moved slightly around the center locations to mimic a clinical scenario in which the probe is moved around the lesion location to obtain multiple sets of measurements from a single lesion. Then all measurements obtained from a single lesion were merged into one DOT histogram.

2.1.1 US images

Figure 1 shows co-registered US images of a benign lesion and a malignant lesion. The benign lesion has a clear boundary, whereas the malignant one has an irregular and ill-defined boundary. These features were used by the CNN to classify different types of breast lesions. To utilize CNN weights pre-trained with the ImageNet dataset,⁵⁰ we cropped the lesion area of each US image to a region of interest (ROI) and resized it to $3 \times 224 \times 224$ pixels before inputting it to the CNN. Here, 3 is the number of channels and 224×224 is the height \times width of each channel. After resizing, the pixel sizes ranged from 0.044 to 0.123 mm. Finally, the images (ROIs) were normalized based on the mean and standard deviation of the ImageNet dataset.

2.1.2 DOT frequency domain data (DOT histogram)

We normalized the measured frequency-domain diffuse reflectance obtained from the lesion side breast, U_l , to the contralateral normal breast, U_r , to produce the “perturbation”

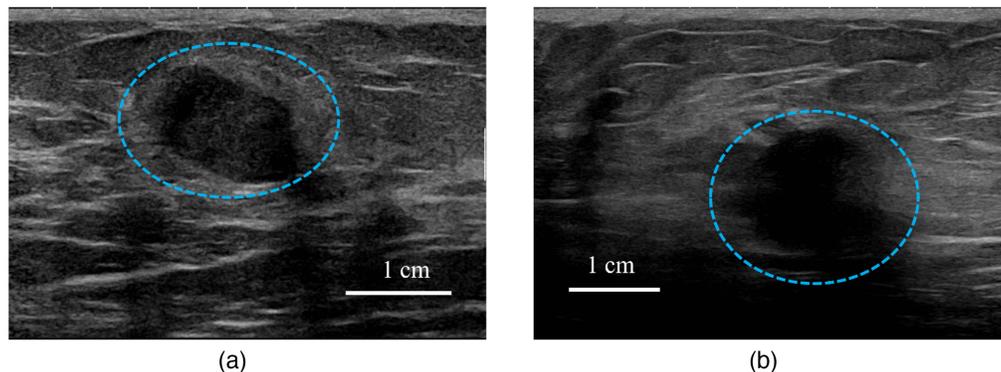


Fig. 1 Examples of co-registered US images, with lesions marked by blue circles. (a) A benign lesion and (b) a malignant lesion.

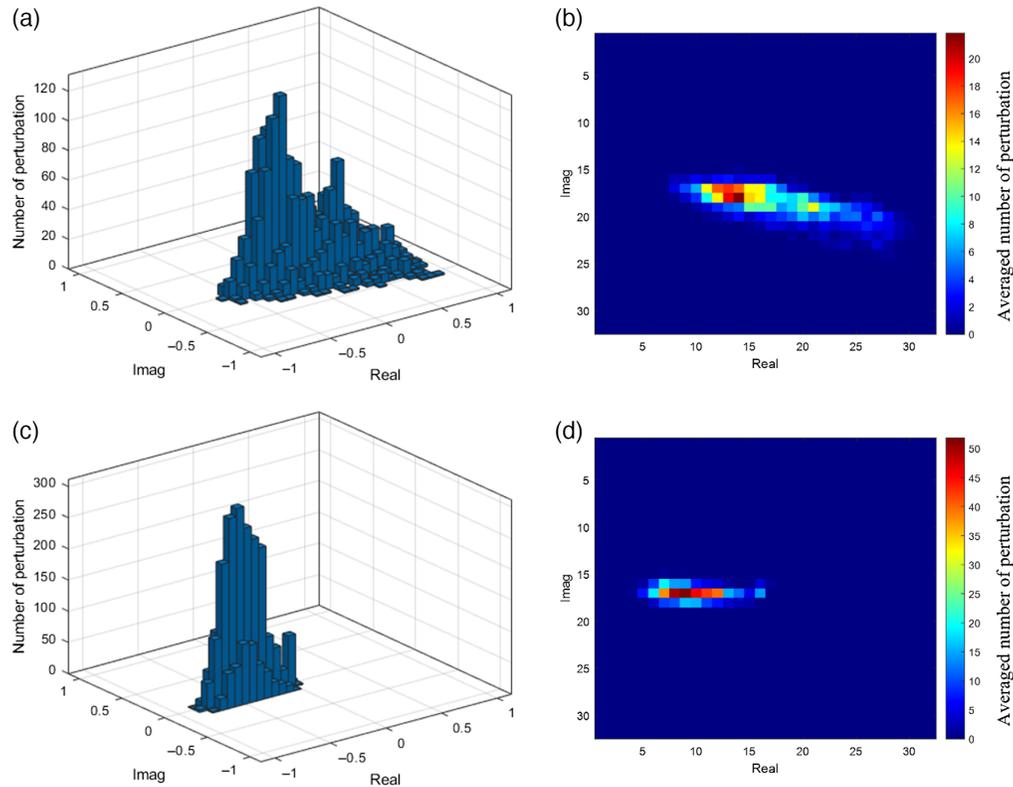


Fig. 2 Examples of DOT histograms and 2D representations. (a) A benign case and (c) its 2D representation after averaging. (b) A malignant case and (d) its 2D representation after averaging.

$$\text{pert} = \frac{U_l - U_r}{U_r} = \frac{U_l}{U_r} - 1 = \text{real}(\text{pert}) + j \times \text{imag}(\text{pert}).$$

We used a two-dimensional representation of the DOT perturbation measurements, with the x -axis being the real part of the perturbation and the y -axis being the imaginary part. For each lesion, the real and imaginary perturbation parts were used together to obtain a bivariate histogram with 32×32 bins, as shown in Fig. 2. For a benign lesion, which typically has lower absorption, the difference between the lesion side and reference side measurements is typically small, which means U_l and U_r are similar and their ratio is close to 1, so the points are more distributed around the origin. For a malignant lesion, typically with higher absorption, U_l is smaller than U_r , so the perturbation is skewed toward -1 in the negative real axis. This histogram representation can handle different source-detector geometries. For each patient, all DOT perturbations from multi-spectral measurements and repeated lesion measurements were combined into one 2D histogram, and the total number of points in one 2D histogram is the number of sources \times number of detectors \times number of wavelengths \times number of repeated lesion measurements. The histograms were then averaged based on the number of lesion files used, the number of wavelengths, and the number of detectors in the DOT system.

2.1.3 DOT reconstructed images

To reconstruct the unknown changes in the target absorption coefficients $\delta\mu_a$ compared to the reference side, we used the Born approximation to linearize the inverse problem:⁵¹

$$f(x) = \arg \min_{\delta\mu_a} \left(\|\text{pert} - W\delta\mu_a\|^2 + \frac{\lambda}{2} \|\delta\mu_a\|^2 \right),$$

where W is a sensitivity matrix calculated using the absorption and reduced scattering coefficients of the reference side and λ is a regularization parameter. The conjugate gradient algorithm and a dual mesh scheme using co-registered US guidance were employed to solve the inverse

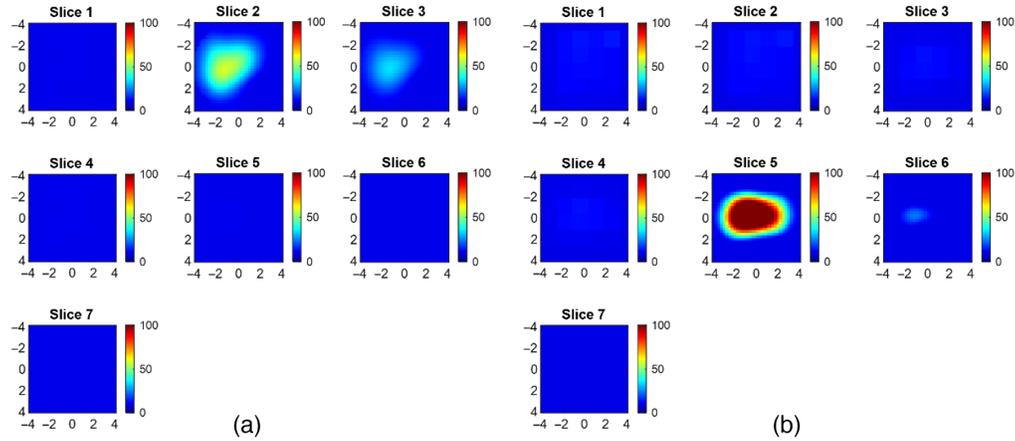


Fig. 3 Examples of DOT reconstructed tHb images. Color bar indicates tHb values in μM . (a) A benign lesion and (b) a malignant lesion.

problem. After reconstructing the absorption maps, the total hemoglobin (tHb) distribution was calculated using all four wavelengths. Figure 3 shows reconstructed images of a benign and a malignant lesion, the same cases as in Fig. 2. The benign lesion has a relatively lower tHb concentration than the malignant lesion. Besides, the malignant lesion shows a “light shadowing” effect, because the tHb in the topmost layer is greater than the tHb of the underlying layers.⁵² The resulting high absorption by the cancer in the topmost layer significantly reduces the number of photons reaching the deeper layers.

2.2 First Stage

In the first stage, two individual CNNs were trained with US images only and DOT histograms only, as shown in Figs. 4(a) and 4(b). Because we had a clear set of benign or malignant labels, supervised learning was used to train and evaluate all the CNN models.

2.2.1 US only CNN

We used a modified VGG-11 neural network⁵³ to extract features from co-registered US images, as shown in Fig. 4(a). The model was pre-trained with the ImageNet database. After the first 26 layers of the VGG-11 and a maxpooling layer, the output $512 \times 7 \times 7$ matrix was zero-padded to a $512 \times 8 \times 8$ matrix. Then another convolutional layer with a 3×3 kernel size and batch normalization was added to generate a $64 \times 8 \times 8$ feature matrix. Following feature extraction, a binary classification model (benign versus malignant) was obtained using three fully connected layers. The weights of the first nine layers were kept unchanged from the pre-trained VGG-11 using ImageNet and then the model was fine-tuned using open-source US images^{54,55} and US images acquired with our US-guided DOT system. The fine-tuning was performed for 5 epochs with a learning rate of 0.001, followed by another 5 epochs with a learning rate of $1e^{-4}$, using stochastic gradient descent with a momentum of 0.9 as the optimizer. After training, the predicted probability of malignancy was extracted and combined with DOT histogram predictions for the first stage classification. The features before the fully connected layers were also extracted to be combined with DOT histogram features and DOT image features for the second stage classification.

2.2.2 DOT histogram feature extraction CNN

The DOT histogram feature extraction CNN is shown in Fig. 4(b). The inputs are the 2D DOT histograms, measuring $1 \times 32 \times 32$ pixels. After two convolutional layers with a 3×3 kernel size, followed by batch normalization and 2×2 maxpooling, the input histogram goes through another convolutional layer with a 3×3 kernel size and two fully connected layers. The output is the final prediction of the lesion type. We pre-trained the model with simulated DOT histograms

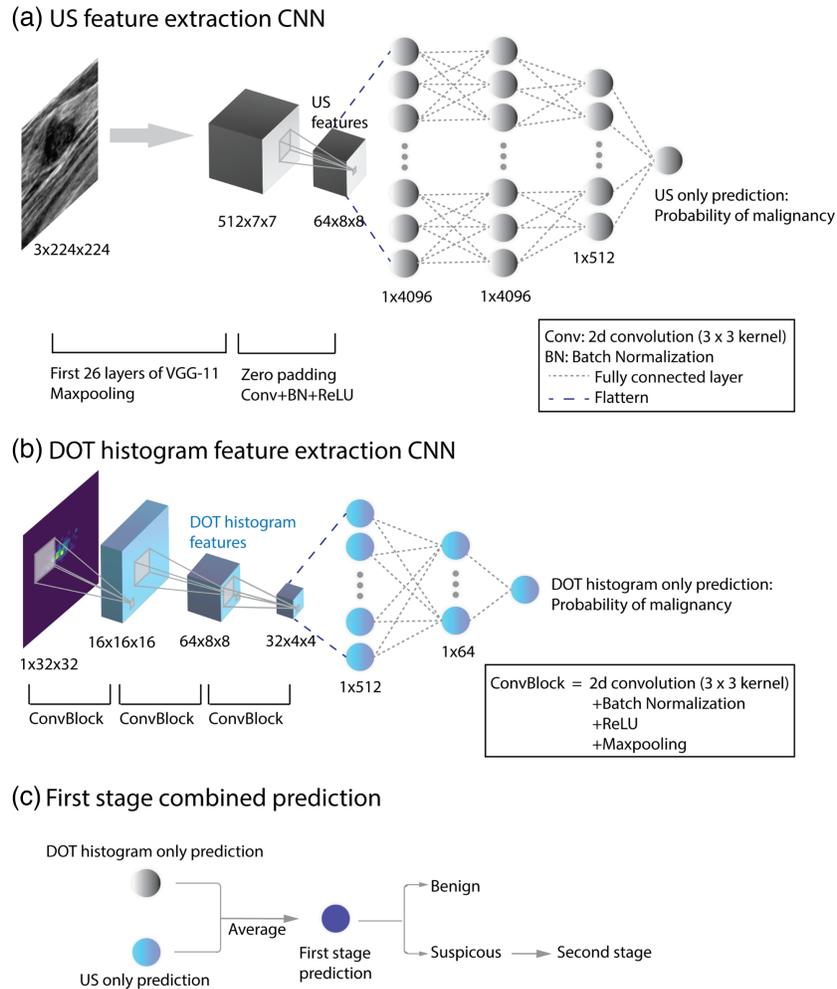


Fig. 4 (a)–(c) First stage of the two-stage classification strategy.

for 20 epochs with a learning rate of $1e^{-4}$ and then fine-tuned it with patient data for 20 epochs with a learning rate of $1e^{-4}$. The Adam optimizer and the cosine annealing learning rate scheduler were used. After the model was trained, the predicted probability of benign versus suspicious was extracted and combined with US only predictions for first stage classification. After two convolutional layers, the features were extracted to be combined with US and DOT images in the second stage for identifying suspicious lesions.

2.2.3 First stage combined prediction

In the first stage, as shown in Fig. 4(c), the probabilities of malignancy predicted from the US-only CNN and the DOT histogram-only CNN were averaged to generate the prediction of benign versus suspicious for near real-time diagnosis.

2.3 Second Stage

2.3.1 DOT image feature extraction CNN

A similar CNN, as shown in Fig. 5(a), was built to extract DOT images. The inputs are the reconstructed tHb maps, which are $3 \times 32 \times 32$ matrixes, and the output is the probability of the lesion being malignant. The three convolutional layers have a convolutional kernel of 3×3 and are followed by batch normalization, and two fully connected layers are applied to perform classification. The DOT image feature extraction CNN was trained on the simulation data with a learning rate of 0.001 for 60 epochs and then fine-tuned with patient data with a

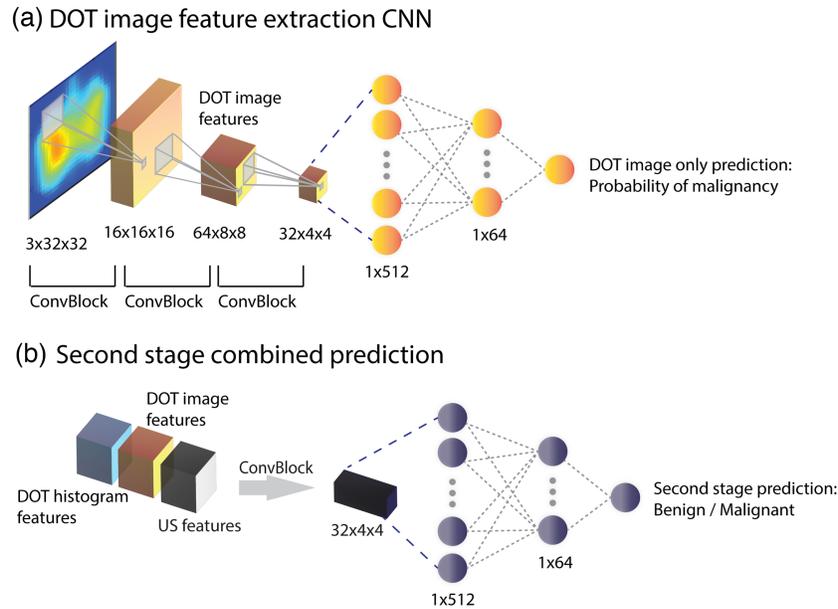


Fig. 5 (a) and (b) Second stage of the two-stage classification strategy.

learning rate of 0.001 for 10 epochs. The features after two convolutional layers were extracted to be combined with DOT histogram features and US features for the second-stage classification.

2.3.2 Two-stage Classification Strategy

As shown in Fig. 5(b), in the second stage, DOT image features, DOT histogram features, and US features are concatenated to generate $192 \times 8 \times 8$ input matrices, followed by a convolutional layer with a 3×3 kernel size and two fully connected layers to output the final prediction of the lesion type for the suspicious group from the first stage. In the two-stage strategy, to recover the optical properties of the suspicious lesions, DOT reconstruction is performed for the suspicious cases identified by the first stage classifier. Then, the DOT image features are extracted using the DOT image feature extraction CNN. Next, to obtain the final classification, the DOT histograms, DOT image features, and US features are concatenated and fed into the second stage classifier. The final predictions for all cases thus comprise the benign predictions from the first stage and the predictions of suspicious cases from the second stage.

The two-stage strategy is shown in Fig. 6. Once the data acquisition is complete, the co-registered US images and the perturbation measurements are generated and put into the first stage classifier, which combines the predictions from the US only CNN and DOT histogram only CNN. To lower the false negative rate while maintaining a high true negative rate, the prediction threshold is set at a value lower than 0.5, the value commonly used in binary classification. The threshold for the first stage classification model was determined by a grid search of all thresholds lower than 0.5, with a step size of 0.01. Based on the grid search results, we set a threshold value of 0.12, which minimized the false negative rate while maintaining high accuracy in predicting benign cases. Lesions with predicted probabilities lower than the threshold are classified as benign lesions, and the rest are classified as suspicious lesions.

3 Experimental Results

The [Supplementary Material](#) provides the number of trainable parameters for each model as well as examples of training and testing losses.

3.1 First Stage Classification Results

On average, the first stage identified 18.2 of 25 cases as benign, so the benign accuracy was 73.0%. However, an average of 0.7 malignant cases were falsely classified as benign, resulting in a false negative rate of 3.0%. Among the 50 testing set cases, an average of 30.1 cases were

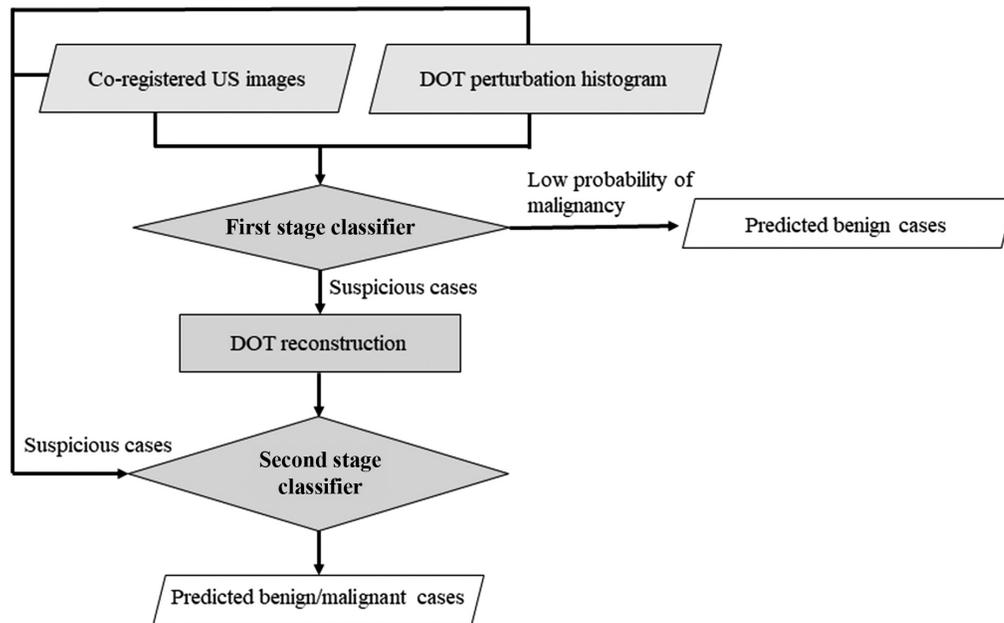


Fig. 6 Two-stage classification workflow.

classified into the suspicious group and fed into the second stage classifier. These results are better than those from US-only models, which can filter out only 68.1% benign cases, with a false negative rate of 9.8%, or histogram-only models, which can filter out 54.6% benign cases, with a false negative rate of 9.4%.

3.2 Two-Stage Classification Strategy Results

After the first stage classifier excluded part of the benign cases, the suspicious cases above the threshold, with a higher probability of malignancy, were advanced into the second stage classifier. For those cases, DOT images were reconstructed using the conjugate gradient descent method with regularization, and the reconstructed image features were extracted using the DOT image feature extraction CNN. Next, the reconstructed DOT image features, DOT histogram features, and US features for the suspicious group were put into the second stage classifier to obtain the final classification. The averaged ROCs from 50 runs for the two-stage classification strategy, along with the single-modality models, are shown in Fig. 7. The first stage classifier predicts only two groups: benign and suspicious cases. In the second stage, the suspicious cases are further classified as either benign or malignant. Biopsy confirmed the true positive and false positive predictions from the second stage, and the biopsy-confirmed true negative and false negative predictions from both stages were used to generate the two-stage combined ROC curve. The confusion matrix used to generate the ROC curve is shown in Table 2, where a 1 or 2 denotes results from the first or second stage. TM, FM, FB, and TB respectively denote true malignant, false malignant, false benign, and true benign. As shown in Fig. 7, the two-stage classification strategy provides a more accurate classification (AUC = 0.946, 95% CI: 0.939 to 0.954) than any of the single-modality based classifiers.

We also evaluated the performance of a single-stage model that combines features extracted from DOT histograms, DOT images, and US images. For a fair comparison, the evaluation is performed for all testing cases, not limited to only the suspicious cases classified by the first stage model. This model achieved an AUC of 0.933 (95% CI: 0.924 to 0.941), which is comparable with that of the two-stage classification strategy; however, the latter can achieve near real time diagnosis for 70% of the benign lesions.

3.3 Comparison with Traditional Feature-Based Machine Learning Approaches

As a comparison, we manually extracted features from DOT histograms and DOT images following Ref. 28 and used two radiologists' BI-RADS readings. Then we used random forest or SVM classifiers instead of deep learning classifiers to perform the classification, with the results shown in Table 3.

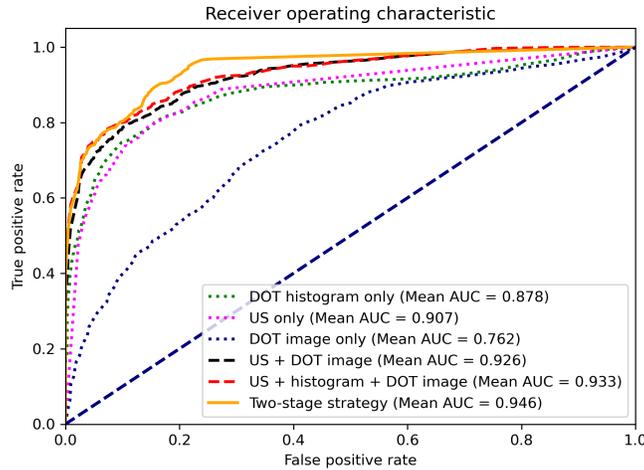


Fig. 7 ROC curves for testing data.

Table 2 Confusion matrix of two-stage combined results, where a 1 or 2 denotes results from the first or second stage. TM, FM, FB, and TB denote true malignant, false malignant, false benign, and true benign, respectively.

Predicted	Biopsy	
	Malignant	Benign
Malignant	TM2	FM2
Benign	FB2 + FB1	TB2 + TB1

Table 3 Comparison of deep-learning based classification with traditional feature-based classification.

Traditional ML input	Traditional ML method	Traditional ML Performance	Deep learning counterpart input	Deep learning counterpart performance
12 DOT histogram features	Random forest classifier	AUC = 0.715	DOT histograms	AUC = 0.878
First stage: 12 DOT histogram features and BI-RADS from 2 radiologists	Random forest classifier	61.8% of benign cases were singled out, with a 6.6% false negative rate	First stage: DOT histograms and US images	73.0% of benign cases were singled out, with a 3.0% false negative rate
2 DOT functional features: total hemoglobin and oxy hemoglobin	SVM classifier	AUC = 0.744	DOT reconstructed total hemoglobin maps	AUC = 0.762
DOT histogram features, BI-RADS, and DOT functional features	Random forest classifier	AUC = 0.931	DOT histograms, US images, and DOT reconstructed total hemoglobin maps	AUC = 0.933
Two-stage classification strategy	First stage: DOT histogram features and BI-RADS Second stage: DOT histogram features, BI-RADS, and functional features	AUC = 0.925	Two-stage classification strategy	AUC = 0.946

These findings suggest that deep learning methods outperform traditional ML methods when using DOT histograms and DOT images only. Additionally, deep learning models combining DOT with US features achieve results better than or comparable to the results of traditional ML models utilizing radiologists' BI-RADS readings.

4 Conclusion

In this paper, we proposed a two-stage strategy using deep-learning models to classify whether a breast lesion is malignant or benign. In the first stage, DOT histogram features and US features are extracted using two feature extraction CNN models. Next, to single out those benign cases with a low probability of malignancy, the averaged prediction scores from the two models are calculated as the first stage result, which yields predictions of benign or suspicious. This strategy can identify 73.0% of benign cases in the first stage without the need for imaging reconstruction. In the second stage, the DOT image features, extracted by another feature extraction CNN model, are combined with the DOT histograms and US features and input into the second-stage classifier to obtain the final diagnostic classification. This strategy achieved near real-time diagnosis for a significant portion of breast lesions, without the need for image reconstruction in the first stage, and the second stage ensures the overall prediction accuracy. Combining the first stage's high sensitivity and the second stage's high specificity, the two-step classification strategy provides better classification accuracy, with an AUC of 0.946, than either single-modality-only models or a single-stage classification model that combines all features.

The proposed two-stage classification strategy has shown promising results on clinical patient data, with high robustness over 50 random training-testing splits. However, there are still limitations to this approach. First, the first-stage model has a false negative rate of 3.0%, even with a threshold set much lower than 0.5. This performance could be improved by incorporating radiologists' BI-RADS scores, which have extremely high sensitivity, albeit low specificity. Second, the second-stage model involves DOT image reconstruction, which requires manual data preprocessing, such as reference selection and outlier removal. This limitation can be mitigated by applying an automatic data preprocessing strategy to fully automate the entire two-stage classification strategy. Additionally, we utilized traditional iterative methods for DOT reconstruction to accommodate data collected from various DOT systems, and these methods are usually slower than deep learning reconstruction. As more data are accumulated from a specific DOT system, deep learning reconstruction³⁷ could be used to achieve real-time diagnosis for all cases. Furthermore, with more data in the future, it will be possible to reserve an additional testing set that has never been exposed to the model during training and thus the robustness of the model could be further evaluated.

Disclosures

The authors declare that there are no conflicts of interest related to this article.

Code, Data, and Materials Availability

Associated code has been uploaded to https://github.com/OpticalUltrasoundImaging/DOT_two_stage. Data are available from the corresponding author upon reasonable request.

Acknowledgments

The authors acknowledge the funding support from U.S. National Cancer Institute (Grant No. R01CA228047). We also thank Professor James Ballard for reviewing and editing the manuscript.

References

1. American Cancer Society, "Cancer Statistics Center," <http://cancerstatisticscenter.cancer.org>.
2. B. L. Sprague et al., "National performance benchmarks for modern diagnostic digital mammography: update from the Breast Cancer Surveillance Consortium," *Radiology* **283**(1), 59–69 (2017).
3. H. J. Schünemann et al., "Breast cancer screening and diagnosis: a synopsis of the european breast guidelines," *Ann. Intern. Med.* **172**(1), 46–56 (2020).

4. P. A. Carney et al., "Erratum: Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography (Annals of Internal Medicine (2003) 138 (168-175))," *Ann. Intern. Med.* **138**(9), 771 (2003).
5. A. S. Tagliafico et al., "Adjunct screening with tomosynthesis or ultrasound in women with mammography-negative dense breasts: interim report of a prospective comparative trial," *J. Clin. Oncol.* **34**(16), 1882–1888 (2016).
6. W. A. Berg et al., "Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer," *JAMA* **299**(18), 2151–2163 (2008).
7. M. F. Bakker et al., "Supplemental MRI screening for women with extremely dense breast tissue," *N. Engl. J. Med.* **381**(22), 2091–2102 (2019).
8. B. Chance et al., "Breast cancer detection based on incremental biochemical and physiological properties of breast cancers: a six-year, two-site study," *Acad. Radiol.* **12**(8), 925–933 (2005).
9. Q. Zhu and S. Poplack, "A review of optical breast imaging: multi-modality systems for breast cancer diagnosis," *Eur. J. Radiol.* **129**, 109067 (2020).
10. V. Heffer Pera et al., "Near-infrared imaging of the human breast: complementing hemoglobin concentration maps with oxygenation images," *J. Biomed. Opt.* **9**(6), 1152 (2004).
11. R. Choe et al., "Diffuse optical tomography of breast cancer during neoadjuvant chemotherapy: a case study with comparison to MRI," *Med. Phys.* **32**(4), 1128–1139 (2005).
12. B. J. Tromberg et al., "Diffuse optics in breast cancer: detecting tumors in pre-menopausal women and monitoring neoadjuvant chemotherapy," *Imaging Breast Cancer* **7**, 279–285 (2005).
13. D. R. Leff et al., "Diffuse optical imaging of the healthy and diseased breast: a systematic review," *Breast Cancer Res. Treat.* **108**(1), 9–22 (2008).
14. E. Y. Chae et al., "Development of digital breast tomosynthesis and diffuse optical tomography fusion imaging for breast cancer detection," *Sci. Rep.* **10**(1), 13127 (2020).
15. M. A. Mastanduno et al., "MR-guided near-infrared spectral tomography increases diagnostic performance of breast MRI," *Clin. Cancer Res.* **21**(17), 3906–3912 (2015).
16. Q. Fang et al., "Combined optical and x-ray tomosynthesis breast imaging," *Radiology* **258**(1), 89–97 (2011).
17. P. Taroni et al., "Clinical trial of time-resolved scanning optical mammography at 4 wavelengths between 683 and 975 nm," *J. Biomed. Opt.* **9**(3), 464 (2004).
18. W. Zhi et al., "Solid breast lesions: clinical experience with US-guided diffuse optical tomography combined with conventional US," *Radiology* **265**(2), 371–378 (2012).
19. V. Ntziachristos et al., "Concurrent MRI and diffuse optical tomography of breast after indocyanine green enhancement," *Proc. Natl. Acad. Sci. U. S. A.* **97**(6), 2767–2772 (2000).
20. B. Brooksby et al., "Combining near-infrared tomography and magnetic resonance imaging to study in vivo breast tissue: implementation of a Laplacian-type regularization to incorporate magnetic resonance structure," *J. Biomed. Opt.* **10**(5), 051504 (2005).
21. R. Choe et al., "Differentiation of benign and malignant breast tumors by in-vivo three-dimensional parallel-plate diffuse optical tomography," *J. Biomed. Opt.* **14**(2), 024020 (2009).
22. M. A. Mastanduno et al., "Sensitivity of MRI-guided near-infrared spectroscopy clinical breast exam data and its impact on diagnostic performance," *Biomed. Opt. Express* **5**(9), 3103 (2014).
23. Q. Zhu et al., "Ultrasound-guided optical tomographic imaging of malignant and benign breast lesions: initial clinical results of 19 cases," *Neoplasia* **5**(5), 379–388 (2003).
24. Q. Zhu et al., "Assessment of functional differences in malignant and benign breast lesions and improvement of diagnostic accuracy by using US-guided diffuse optical tomography in conjunction with conventional US," *Radiology* **280**(2), 387–397 (2016).
25. Q. Zhu, N. Chen, and S. H. Kurtzman, "Imaging tumor angiogenesis by use of combined near-infrared diffusive light and ultrasound," *Opt. Lett.* **28**(5), 337 (2003).
26. Q. Zhu, S. Tannenbaum, and S. H. Kurtzman, "Optical tomography with ultrasound localization for breast cancer diagnosis and treatment monitoring," *Surg. Oncol. Clin. N. Am.* **16**(2), 307–321 (2007).
27. Q. Zhu et al., "Benign versus malignant breast masses: optical differentiation with US-guided optical imaging reconstruction," *Radiology* **237**(1), 57–66 (2005).
28. Q. Zhu et al., "Breast cancer: assessing response to neoadjuvant chemotherapy by using US-guided near-infrared tomography," *Radiology* **266**(2), 433–442 (2013).
29. S. P. Poplack et al., "Prospective assessment of adjunctive ultrasound-guided diffuse optical tomography in women undergoing breast biopsy: impact on BI-RADS assessments," *Eur. J. Radiol.* **145**(May), 110029 (2021).
30. S. Sabir et al., "Convolutional neural network-based approach to estimate bulk optical properties in diffuse optical tomography," *Appl. Opt.* **59**(5), 1461 (2020).
31. M. Zhang et al., "Deep learning-based method to accurately estimate breast tissue optical properties in the presence of the chest wall," *J. Biomed. Opt.* **26**(10), 106004 (2021).

32. J. Feng et al., “Back-propagation neural network-based reconstruction algorithm for diffuse optical tomography,” *J. Biomed. Opt.* **24**(5), 051407 (2018).
33. J. Feng et al., “Deep-learning based image reconstruction for MRI-guided near-infrared spectral tomography,” *Optica* **9**(3), 264 (2022).
34. N. Murad, M. Pan, and Y. F. Hsu, “Reconstruction and localization of tumors in breast optical imaging via convolution neural network based on batch normalization layers,” *IEEE Access* **10**, 57850–57864 (2022).
35. H. B. Yedder et al., “Multitask deep learning reconstruction and localization of lesions in limited angle diffuse optical tomography,” *IEEE Trans. Med. Imaging* **41**, 515–530 (2021).
36. J. Yoo et al., “Deep learning diffuse optical tomography,” *IEEE Trans. Med. Imaging* **39**(4), 877–887 (2020).
37. Y. Zou et al., “Machine learning model with physical constraints for diffuse optical tomography,” *Biomed. Opt. Express* **12**(9), 5720 (2021).
38. B. Deng et al., “FDU-net: deep learning-based three-dimensional diffuse optical image reconstruction,” *IEEE Trans Med Imaging* **42**, 2439–2450 (2023).
39. N. Murad, M.-C. Pan, and Y.-F. Hsu, “Periodic-net: an end-to-end data driven framework for diffuse optical imaging of breast cancer from noisy boundary data,” *J. Biomed. Opt.* **28**(2), 026001 (2023).
40. M. Zhang et al., “A fusion deep learning approach combining diffuse optical tomography and ultrasound for improving breast cancer classification,” *Biomed. Opt. Express* **14**(4), 1636–1646 (2023).
41. K. M. S. Uddin et al., “Optimal breast cancer diagnostic strategy using combined ultrasound and diffuse optical tomography,” *Biomed. Opt. Express* **11**(5), 2722 (2020).
42. G. Di Sciacca et al., “Evaluation of a pipeline for simulation, reconstruction, and classification in ultrasound-aided diffuse optical tomography of breast tumors,” *J. Biomed. Opt.* **27**(03), 036003 (2022).
43. Q. Xu, X. Wang, and H. Jiang, “Convolutional neural network for breast cancer diagnosis using diffuse optical tomography,” *Vis. Comput. Ind. Biomed. Art* **2**(1), 1–6 (2019).
44. S. C. Huang et al., “Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection,” *Sci. Rep.* **10**(1), 22147 (2020).
45. S. C. Huang et al., “Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines,” *NPJ Digital Med.* **3**(1), 136 (2020).
46. N. Nasir et al., “Multi-modal image classification of COVID-19 cases using computed tomography and X-rays scans,” *Intell. Syst. Appl.* **17**(Nov. 2022), 200160 (2023).
47. H. Vavadi et al., “Compact ultrasound-guided diffuse optical tomography system for breast cancer imaging,” *J. Biomed. Opt.* **24**(2), 021203 (2018).
48. S. Li, M. Zhang, and Q. Zhu, “Ultrasound segmentation-guided edge artifact reduction in diffuse optical tomography using connected component analysis,” *Biomed. Opt. Express* **12**(8), 5320 (2021).
49. A. Badano et al., “Evaluation of digital breast tomosynthesis as replacement of full-field digital mammography using an in Silico imaging trial,” *JAMA Network Open* **1**(7), e185474 (2018).
50. J. Deng et al., “ImageNet: a large-scale hierarchical image database,” *2009 IEEE Conf. Comput. Vis. Pattern Recognit.*, 248–255, IEEE (2009).
51. K. M. S. Uddin et al., “Two step imaging reconstruction using truncated pseudoinverse as a preliminary estimate in ultrasound guided diffuse optical tomography,” *Biomed. Opt. Express* **8**(12), 5437 (2017).
52. M. Zhang et al., “Target depth-regularized reconstruction in diffuse optical tomography using ultrasound segmentation as prior information,” *Biomed. Opt. Express* **11**(6), 3331 (2020).
53. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv: 1409.1556, 2015, <https://arxiv.org/abs/1409.1556>.
54. W. Al-Dhabyani et al., “Dataset of breast ultrasound images,” *Data Br.* **28**, 104863 (2020).
55. P. S. Rodrigues, “Breast ultrasound image,” Mendeley Data, 2017, <https://data.mendeley.com/datasets/wmy84gzngw/1>.

Menghao Zhang received his doctoral degree from the Department of Electrical and System Engineering at Washington University in St. Louis in 2023. He received his bachelor’s degree from the University of Electronic Science and Technology of China in 2015. Throughout his studies, he has worked on enhancing breast cancer diagnosis through ultrasound-guided diffuse optical tomography.

Shuying Li completed her doctoral degree from the Department of Biomedical Engineering at Washington University in St. Louis in 2023. She also holds a bachelor’s degree from Zhejiang University and a master’s degree from the University of Michigan. Her research is primarily focused on utilizing optical imaging techniques such as diffuse optical tomography, optical coherence tomography, and spatial frequency domain imaging for cancer diagnosis. She is currently a postdoctoral candidate at Boston University.

Minghao Xue earned his bachelor's degree from Sun Yat-sen University in China in 2020 and commenced his doctoral studies in biomedical engineering at Washington University in Saint Louis in 2021. His research primarily revolves around the application of deep learning and ultrasound-guided diffuse optical tomography (US-guided DOT). His current focus involves the automated study of US-guided DOT clinical translation.

Qing Zhu joined Washington University in St. Louis in the Department of Biomedical Engineering in July 2016. Previously, she was a professor of electrical and computer engineering and biomedical engineering at the University of Connecticut. Her research interests are focused on multimodality photoacoustic, ultrasound, optical coherence tomography, and structured light imaging techniques for cancer detection and treatment assessment and prediction.