


Multi-class classification of breast tissue using optical coherence tomography and attenuation imaging combined via deep learning: supplement

KEN Y. FOO,^{1,2,*}  KYLE NEWMAN,^{1,2} QI FANG,^{1,2} PEIJUN GONG,^{1,2} HINA M. ISMAIL,^{1,2} DEVINA D. LAKHIANI,^{1,2} RENATE ZILKENS,^{1,3} BENJAMIN F. DESSAUVAGIE,^{4,5} BRUCE LATHAM,^{5,6} CHRISTOBEL M. SAUNDERS,^{3,7,8,9} LIXIN CHIN,^{1,2} AND BRENDAN F. KENNEDY^{1,2,10}

¹*BRITelab, Harry Perkins Institute of Medical Research, QEII Medical Centre, Nedlands, and Centre for Medical Research, The University of Western Australia, Perth, WA 6009, Australia*

²*Department of Electrical, Electronic & Computer Engineering, School of Engineering, The University of Western Australia, Perth, WA 6009, Australia*

³*Division of Surgery, Medical School, The University of Western Australia, Perth, WA 6009, Australia*

⁴*Division of Pathology and Laboratory Medicine, Medical School, The University of Western Australia, Perth, WA 6009, Australia*

⁵*PathWest, Fiona Stanley Hospital, Murdoch, WA 6150, Australia*

⁶*School of Medicine, The University of Notre Dame, Fremantle, WA 6160, Australia*

⁷*Breast Centre, Fiona Stanley Hospital, Murdoch, WA 6150, Australia*

⁸*Breast Clinic, Royal Perth Hospital, Perth, WA 6000, Australia*

⁹*Department of Surgery, Melbourne Medical School, The University of Melbourne, Parkville, VIC 3010, Australia*

¹⁰*Australian Research Council Centre for Personalised Therapeutics Technologies, Perth, WA 6000, Australia*

*ken.foo@research.uwa.edu.au

This supplement published with Optica Publishing Group on 12 May 2022 by The Authors under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) in the format provided by the authors and unedited. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Supplement DOI: <https://doi.org/10.6084/m9.figshare.19656768>

Parent Article DOI: <https://doi.org/10.1364/BOE.455110>

**Multi-class classification of breast
tissue using optical coherence
tomography and attenuation imaging
combined via deep learning:
supplemental document**

1. CROSS-VALIDATION DATASETS

Table S1 provides the details of the datasets for each cross-validation group.

Table S1. Summary of the datasets used to train, validate, and test the CNN for each cross-validation group. Percentages in brackets indicate the percentage of each set in each group corresponding to each tissue class.

Cross-validation group	Set	Class			Total
		Adipose tissue	Benign dense tissue	Malignant tissue	
1	Training	2,170 (52.5%)	841 (20.3%)	1,122 (27.1%)	4,133
	Validation	542 (52.5%)	210 (20.3%)	281 (27.2%)	1,033
	Test	336 (52.7%)	93 (14.6%)	209 (32.8%)	638
2	Training	2,178 (51.6%)	828 (19.6%)	1,214 (28.8%)	4,220
	Validation	544 (51.6%)	207 (19.6%)	303 (28.7%)	1,054
	Test	326 (61.5%)	109 (20.6%)	95 (17.9%)	530
3	Training	2,157 (52.1%)	862 (20.8%)	1,121 (27.1%)	4,140
	Validation	539 (52.1%)	215 (20.8%)	280 (27.1%)	1,034
	Test	352 (55.9%)	67 (10.6%)	211 (33.5%)	630
4	Training	2,235 (54.2%)	788 (19.1%)	1,102 (26.7%)	4,125
	Validation	559 (54.2%)	197 (19.1%)	275 (26.7%)	1,031
	Test	254 (39.2%)	159 (24.5%)	235 (36.3%)	648
5	Training	2,227 (52.7%)	846 (20.0%)	1,155 (27.3%)	4,228
	Validation	557 (52.6%)	212 (20.0%)	289 (27.3%)	1,058
	Test	264 (51.0%)	86 (16.6%)	168 (32.4%)	518
6	Training	2,239 (52.4%)	854 (20.0%)	1,180 (27.6%)	4,273
	Validation	560 (52.4%)	213 (19.9%)	295 (27.6%)	1,068
	Test	249 (53.8%)	77 (16.6%)	137 (29.6%)	463
7	Training	2,167 (52.0%)	827 (19.8%)	1,174 (28.2%)	4,168
	Validation	542 (52.0%)	207 (19.8%)	294 (28.2%)	1,043
	Test	339 (57.2%)	110 (18.5%)	144 (24.3%)	593
8	Training	2,140 (51.2%)	843 (20.2%)	1,193 (28.6%)	4,176
	Validation	535 (51.2%)	211 (20.2%)	298 (28.5%)	1,044
	Test	373 (63.9%)	90 (15.4%)	121 (20.7%)	584
9	Training	2,156 (51.9%)	834 (20.1%)	1,163 (28.0%)	4,153
	Validation	539 (51.9%)	208 (20.0%)	291 (28.0%)	1,038
	Test	353 (57.6%)	102 (16.6%)	158 (25.8%)	613
10	Training	2,277 (54.6%)	714 (17.1%)	1,182 (28.3%)	4,173
	Validation	569 (54.5%)	179 (17.1%)	296 (28.4%)	1,044
	Test	202 (34.4%)	251 (42.8%)	134 (22.8%)	587
Total		3,048 (52.5%)	1,144 (19.7%)	1,612 (27.8%)	5,804

2. RESNET-18 ARCHITECTURE

Figure S1 presents the architecture of the ResNet-18 CNN used in this study. This network contains 18 layers (represented by the blocks), which includes eight “building blocks” that each contain two 3×3 convolutional layers and a shortcut connection.

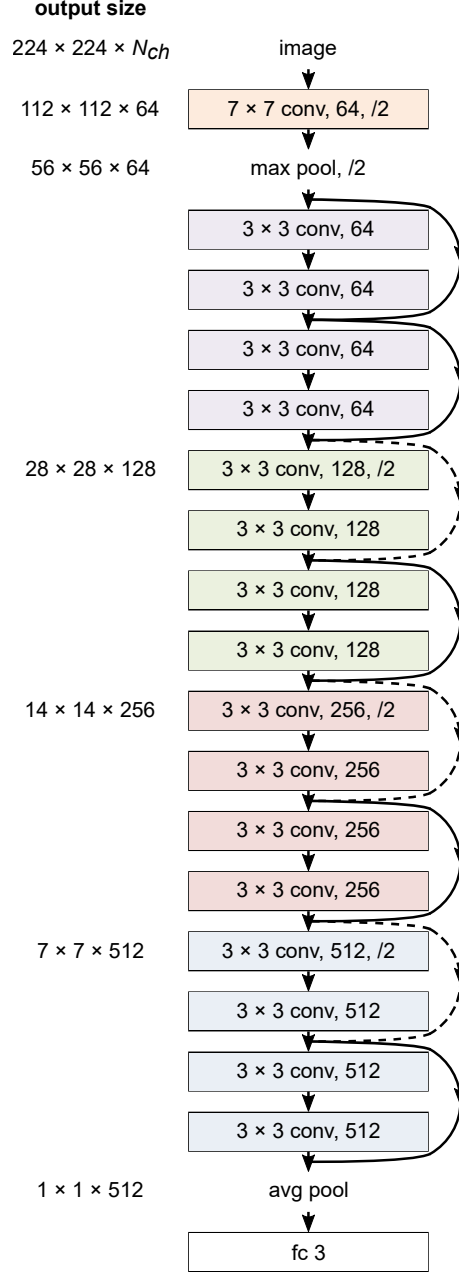


Fig. S1. ResNet-18 architecture. Each block represents one layer. Colored blocks represent convolutional layers, with the text indicating the kernel size, the number of filters, and, if applicable, the downsampling factor (achieved using a stride of 2). The white block represents a fully connected layer, with the text indicating the number of outputs. Shortcut connections are indicated by the curved arrow lines; dashed lines indicate where 1×1 convolutions with a stride of 2 are used to match dimensions. The output sizes from layers that change the output size are indicated on the left, where N_{ch} represents the number of channels in the input image ($N_{ch} = 1$ for the OCT network and the attenuation network, and $N_{ch} = 2$ for the combined network).

3. ATTENUATION NETWORK

This Section presents a comparison of the classification performance of the attenuation network compared to the OCT network and the combined network. The architecture and training methodology for the attenuation network is identical to that of the OCT network (described in Section 2.2 of the main manuscript), except attenuation sub-images only are used for training, validation, and testing. Table S2 shows the confusion matrix and classification performance of the attenuation network, and Tables S3–S5 provide p -values for the two-sided Wilcoxon signed-rank test comparing the OCT network and the attenuation network, the attenuation network and the combined network, and the OCT network and the combined network, respectively.

Table S2. Confusion matrix and classification performance of the attenuation network on the test sets. The confusion matrix entries are the sum of the confusion matrices from each cross-validation group and the percentage of the true class. The sensitivity, specificity, PPV, NPV, and accuracy for each class, and the total accuracy and MCC across all classes, are the mean \pm standard deviation across all cross-validation groups.

		Attenuation network		
Histopathology		Adipose tissue	Benign dense tissue	Malignant tissue
Sub-images		3,048	1,144	1,612
CNN prediction	Adipose tissue	2,922 (95.9%)	97 (8.5%)	68 (4.2%)
	Benign dense tissue	76 (2.5%)	690 (60.3%)	283 (17.6%)
	Malignant tissue	50 (1.6%)	357 (31.2%)	1,261 (78.2%)
Sensitivity (%)		95.9 \pm 2.2	60.0 \pm 16.5	79.0 \pm 10.2
Specificity (%)		93.8 \pm 4.0	92.4 \pm 5.2	90.3 \pm 7.0
PPV (%)		94.5 \pm 3.5	67.0 \pm 12.2	76.7 \pm 13.0
NPV (%)		95.3 \pm 2.5	90.3 \pm 7.8	91.9 \pm 4.7
Accuracy (%)		95.0 \pm 1.8	86.0 \pm 5.3	87.1 \pm 5.4
		Attenuation network		
Total accuracy (%)		84.1 \pm 5.5		
MCC		0.741 \pm 0.064		

From Table S3, we see a statistically significant difference in performance between the OCT network and the attenuation network in sensitivity, NPV, and accuracy in adipose tissue, as well as in accuracy in malignant tissue. Comparing Table S2 to Table 3, we see that while the attenuation network accuracy in malignant tissue is higher than that of the OCT network by 1.6%, the OCT network outperforms the attenuation network across all adipose tissue metrics by $\sim 1\%$. This may be because the algorithm for computing the attenuation coefficient involves fitting a curve to the OCT signal with depth, however, the OCT signal in adipose tissue consists of regions of low signal (from the lipid-filled interior of adipose cells) interrupted by sharp peaks (from the cell boundaries), which makes curve fitting challenging and may cause erroneous results. Therefore, while attenuation imaging may provide more contrast in dense tissue, OCT may provide more contrast in adipose tissue. As a result, between the OCT network and the attenuation network, neither network may be considered dominant across all metrics.

Table S3. p -values for the two-sided Wilcoxon signed-rank test applied to the performance metrics of the OCT and attenuation networks. The null hypothesis is that the median of the differences $X_i - Y_i$ is zero, against the alternative that it is not zero, where X_i is the performance on Group i for the OCT network, and Y_i is the performance on the same group for the attenuation network. Asterisks indicate a statistically significant difference ($\alpha = 0.05$).

Histopathology	$X = \text{OCT network}, Y = \text{Attenuation network}$		
	Adipose tissue	Benign dense tissue	Malignant tissue
Sensitivity	0.0195*	0.8457	0.1055
Specificity	0.1309	0.6953	0.3074
PPV	0.1055	0.8457	0.4316
NPV	0.0098*	0.6953	0.0645
Accuracy	0.0059*	0.3223	0.0273*
$X = \text{OCT network}, Y = \text{Attenuation network}$			
Total accuracy	0.3750		
MCC	0.6953		

From Table S4, we see a statistically significant difference in performance between the attenuation network and the combined network in accuracy in adipose tissue, as well as sensitivity in benign dense tissue. Comparing Table S2 to Table 3, we see that the combined network outperforms the attenuation network by 1.2% for accuracy in adipose tissue, and 8.0% for sensitivity in benign dense tissue. Similarly, as shown in Table S5 and discussed in Section 3, the combined network outperforms the OCT network across several metrics in benign dense tissue and malignant tissue, as well as total accuracy and MCC across all classes. Therefore, using OCT and attenuation imaging together in the combined network yields better results than using either OCT or attenuation imaging individually in the OCT network or the attenuation network, respectively.

Table S4. p -values for the two-sided Wilcoxon signed-rank test applied to the performance metrics of the attenuation and combined networks. The null hypothesis is that the median of the differences $X_i - Y_i$ is zero, against the alternative that it is not zero, where X_i is the performance on Group i for the attenuation network, and Y_i is the performance on the same group for the combined network. Asterisks indicate a statistically significant difference ($\alpha = 0.05$).

$X = \text{Attenuation network}, Y = \text{Combined network}$			
Histopathology	Adipose tissue	Benign dense tissue	Malignant tissue
Sensitivity	0.0662	0.0371*	0.4316
Specificity	0.0645	0.4922	0.3223
PPV	0.0645	0.4922	0.1309
NPV	0.1309	0.0840	0.3750
Accuracy	0.0098*	0.3223	0.4726
$X = \text{Attenuation network}, Y = \text{Combined network}$			
Total accuracy		0.1934	
MCC		0.0840	

Table S5. p -values for the two-sided Wilcoxon signed-rank test applied to the performance metrics of the OCT and combined networks. The null hypothesis is that the median of the differences $X_i - Y_i$ is zero, against the alternative that it is not zero, where X_i is the performance on Group i for the OCT network, and Y_i is the performance on the same group for the combined network. Asterisks indicate a statistically significant difference ($\alpha = 0.05$).

$X = \text{OCT network}, Y = \text{Combined network}$			
Histopathology	Adipose tissue	Benign dense tissue	Malignant tissue
Sensitivity	0.8762	0.0107*	0.5566
Specificity	0.3583	0.6953	0.0273*
PPV	0.4142	0.1934	0.0137*
NPV	0.9187	0.0059*	0.3750
Accuracy	0.6831	0.0039*	0.0020*
$X = \text{OCT network}, Y = \text{Combined network}$			
Total accuracy		0.0039*	
MCC		0.0273*	