Supplemental Information for Learning a Clinically-Relevant Concept Bottleneck for Lesion Detection in Breast Ultrasound

Table 1. Additional characteristics of the study sample. *All image-level counts reflect the number of images collected, not the resulting number of images after dual-view images were split. BUS = breast ultrasound. Dx = diagnosis. BI-RADS feature lesion counts are reported as dichotomized for model training and evaluation, not as defined by the ACR [1].

Dataset Characteristic, Unit	Train	Validation	Test
Women, N	693	101	200
Mean age at BUS, years (SD)	63.6 (12.8)	62.9 (11.7)	63.3 (13.9)
Mean age at Dx, years (SD)	63.7 (12.5)	62.9 (12.2)	63.0 (13.8)
Images, N*	6,260	910	1,684
Images with benign findings, N (%)	4,587 (73.3)	661 (72.6)	1,307 (77.6)
Images with malig. findings, N (%)	1,673 (26.7)	249 (27.4)	377 (22.4)
Images on PHILIPS system, N (%)	2,881 (46.0)	489 (53.7)	1,113 (66.1)
Images on SIEMENS system, N (%)	2,949 (47.1)	381 (41.9)	539 (32.0)
Images on ATL system, N (%)	430 (6.9)	40 (4.4)	32 (1.9)
BI-RADS 1/2/3 images, N (%)	3,914 (62.5)	566 (62.2)	1,159 (68.8)
BI-RADS 4 images, N (%)	1,410 (22.5)	235 (25.8)	357 (21.2)
BI-RADS 5/6 images, N (%)	638 (10.2)	92 (10.1)	137 (8.1)
BI-RADS 0/Unk. images, N (%)	298 (4.8)	17 (1.9)	31 (1.8)
Lesion Views, N	4,203	573	872
Oval lesions, N (%)	2,801 (66.6)	368 (64.2)	573 (65.7)
Irregular/round lesions, N (%)	1,402 (33.4)	205 (35.8)	299 (34.3)
Parallel lesions, N (%)	3,560 (84.7)	449 (78.4)	752 (86.2)
Not parallel lesions, N (%)	643 (15.3)	124 (21.6)	120 (13.8)
Circumscribed lesions, N (%)	2,937 (69.9)	401 (70.0)	598 (68.6)
Not circumscribed lesions, N (%)	1,266 (30.1)	172 (30.0)	274 (31.4)
Anechoic lesions, N (%)	1,259 (30.0)	168 (29.3)	339 (38.9)
Not anechoic lesions, N (%)	2,944 (70.0)	405 (70.7)	533 (61.1)
Lesions w/o posterior feats., N (%)	2,988 (71.1)	383 (66.8)	560 (64.2)
Lesions w/posterior feats., N (%)	1,215 (28.9)	190 (33.2)	312 (35.8)

Table 2. Hyperparameter search space and chosen hyperparameter values for all model training stages and variations. For each stage, the hyperparameters were tuned over 25 trials using Optuna's TPESampler [2]. Frozen stage = stage at which the ResNet-101 FPN is frozen during training. FC = fully-connected. In Stage 2, the number of filters corresponds to the number in each convolutional layer in a residual block.

Training Stage	Hyperparameter	Search Space	Value
Concept Bottleneck Model	frozen stage	{1, 2, 3, 4, 5}	4
Stage 1:	# box head conv. layers	{1, 2, 3, 4, 5}	3
Lesion Detection	# box head FC layers	{1, 2, 3}	4
	# mask head conv. layers	{1, 2, 3, 4, 5}	5
	momentum	Uniform [0.1, 0.9]	0.9
Stage 2:	# filters in 1st layer	{512, 256, 128, 64}	512
Concept Classification	# filters in 2nd layer	{256, 128, 64, 32}	64
	base learning rate	LogUniform [1e-7, 1e-1]	0.093
	momentum	Uniform [0.1, 0.9]	0.8
Stage 3a:	FC layer width	{2048, 1024, 512, 256, 128, 64}	512
Cancer Classification	base learning rate	LogUniform [1e-7, 1e-1]	4×10^{-4}
(w/o side channel)	intermediate sigmoid	{True, False}	False
	momentum	Uniform [0.1, 0.9]	0.5
Stage 3b:	base learning rate	LogUniform [1e-7, 1e-1]	6×10^{-6}
(w/ side channel)	momentum	Uniform [0.1, 0.9]	0.8
Baseline Model	base learning rate	LogUniform [1e-7, 1e-1]	7×10^{-2}
	momentum	Uniform [0.1, 0.9]	0.1

References

1. CJ, D.O., et al., ACR BI-RADS ® Atlas, Breast Imaging Reporting and Data System. 2013, Reston, VA: American College of Radiology.
2. Akiba, T., et al. Optuna: A Next-generation Hyperparameter Optimization Framework. INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING. ACM.