

Table 1. Supplemental: Dataset and Active Learning strategy testing details

Dataset	Training Size	Test Size	Step Size	Max Size	Strategy	Repetitions (Supervised)	Repetitions (Semi-supervised)
US Dataset	849	199	50	750	Random	20	20
					RiggedDraw ($\alpha = 5$)	20	19
					RiggedDraw ($\alpha = 25$)	20	20
					RiggedDraw ($\alpha = 50$)	19	19
					LeastConfidence	19	19
					KMeans	20	20
					WAAL	N/A	19
Pneumonia MNIST	850	624	50	750	Random	19	20
					RiggedDraw ($\alpha = 5$)	19	19
					RiggedDraw ($\alpha = 25$)	20	19
					RiggedDraw ($\alpha = 50$)	19	20
					LeastConfidence	20	20
					KMeans	20	19
					WAAL	N/A	19
BreaKHis	850	364	50	750	Random	19	17
					RiggedDraw ($\alpha = 5$)	19	18
					RiggedDraw ($\alpha = 25$)	19	21
					RiggedDraw ($\alpha = 50$)	19	19
					LeastConfidence	20	18
					KMeans	18	17
					WAAL	N/A	16

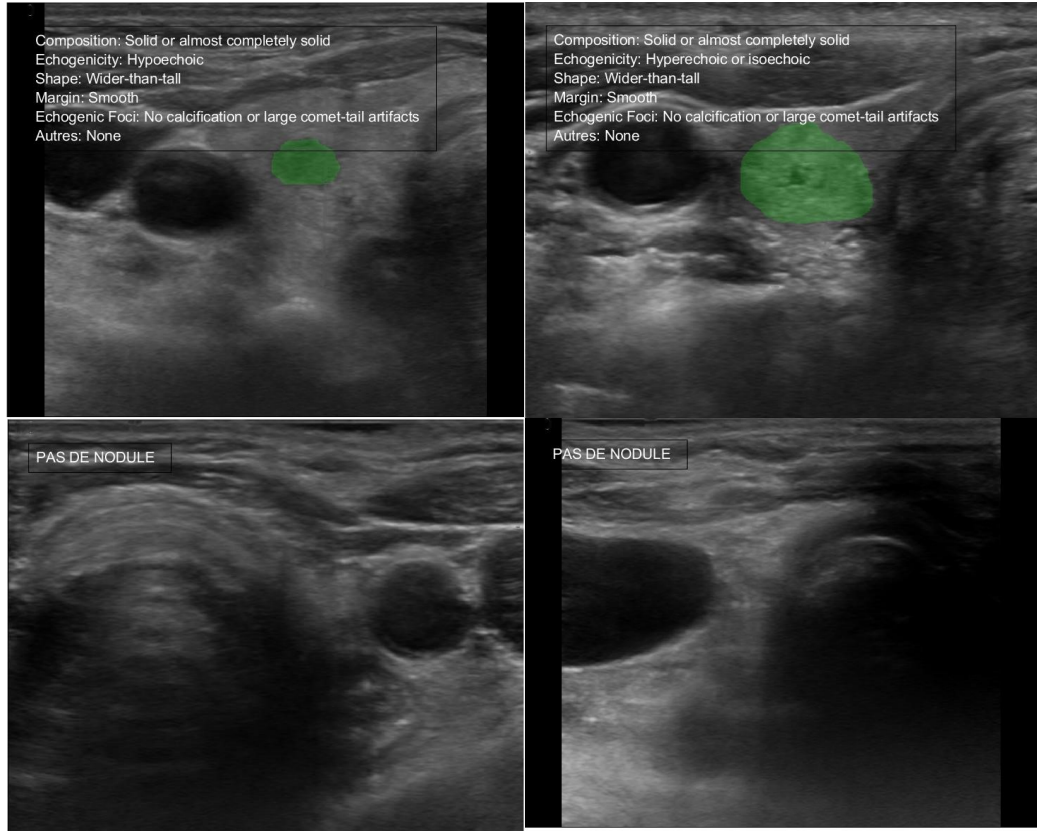


Fig. 1. Supplemental: Sample annotated ultrasound images from the dataset, with and without nodules.

Table 2. Supplemental: Supervised learning AUBC values for different Rigged Draw weights, with * indicating p-values < 0.05 when compared to random selection.

Dataset	Measure	$\alpha = 5$	$\alpha = 25$	$\alpha = 50$
US Dataset	Mean	0.639	0.639	0.642
	Median	0.639	0.641	0.640
	STD	0.012	0.012	0.008
Pneumonia MNIST	Mean	0.916	0.919	0.917
	Median	0.915	0.920	0.918
	STD	0.007	0.004	0.005
BreaKHis	Mean	0.788	0.787	0.791*
	Median	0.784	0.793	0.793
	STD	0.020	0.022	0.017

Table 3. Supplemental: Semi-supervised learning AUBC values for different Rigged Draw weights, with * indicating p-values < 0.05 when compared to random selection.

Dataset	Measure	$\alpha = 5$	$\alpha = 25$	$\alpha = 50$
US Dataset	Mean	0.754*	0.754*	0.751*
	Median	0.755	0.755	0.752
	STD	0.008	0.007	0.012
Pneumonia MNIST	Mean	0.923	0.924*	0.924
	Median	0.923	0.923	0.924
	STD	0.009	0.006	0.006
BreaKHis	Mean	0.799	0.805	0.804
	Median	0.798	0.802	0.805
	STD	0.019	0.014	0.018