SPLEEN-US: Open Annotated Spleen Ultrasound Dataset for Precision Neuromodulation

Table 1: Existing Spleen Ultrasound Datasets (* our proposed dataset)

Name	No. of subjects	No. of images + masks	2D/cine	Metadata	Access
Spleenex	450	450	2D	n/a	restricted
US sim & seg (CycleGAN)	11	61 926 synthetic	2D	n/a	public
DeepSPV	0	150 synthetic	2D	n/a	public
Roboflow US	208	208	2D	n/a	public
SPLEEN-US *	**2000**	**cine loops + masks**	**2D cine**	**yes**	**public**

Figure 1: Proposed acquisition, segmentation, and proofreading workflow. We will recruit 2000 subjects without medical conditions from a diverse student population and acquire abdominal ultrasound focused on the spleen. Automated 2D cine-loop segmentation will be followed by expert proofreading using our streamlined QC software to iteratively improve our automatic algorithm. For each subject, we will release the ultrasound images, segmentation masks, and metadata (age, sex, race/ethnicity, BMI, socioeconomic indicators).

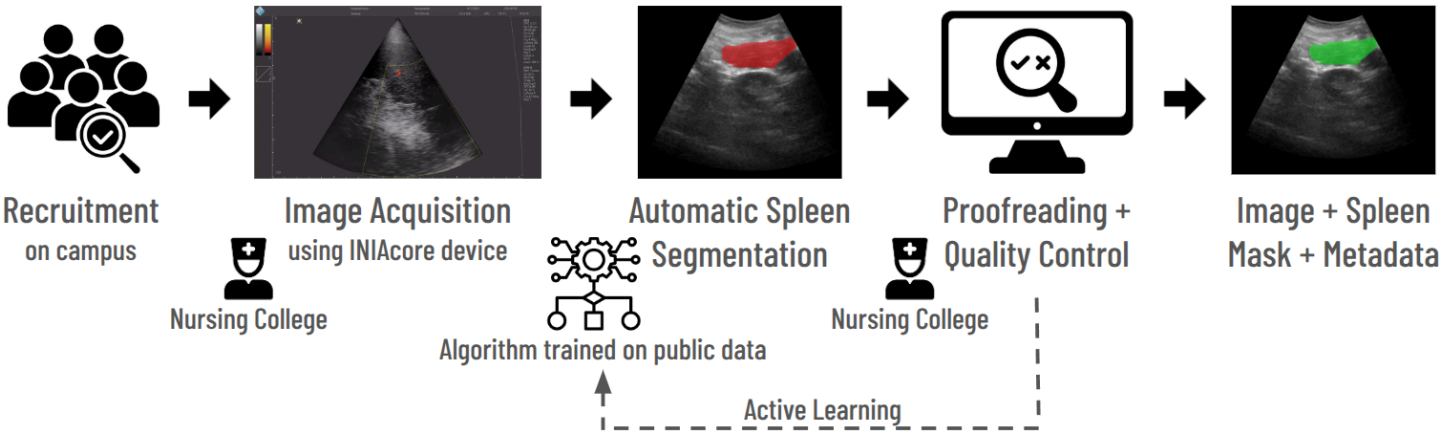


Figure 2: Preliminary feasibility study. We trained a UNet-based classifier on the Roboflow US public dataset and tested on previously unseen samples of the same dataset (left). While not perfect, the classifier performed relatively well (Test Dice 0.91). However, without retraining, the classifier fails on the US sim & seg (CycleGAN) dataset indicating a lack of generalization.

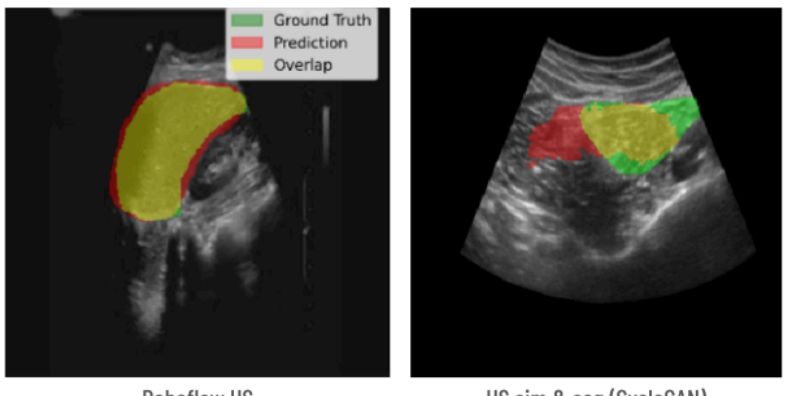
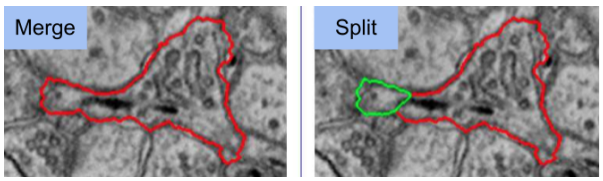


Figure 3: Related work. The PI team has significant experience in biomedical image segmentation algorithms across modalities, scientific visualization, and streamlined quality control software for proofreading.

Our single-click carotid plaque segmentation is 2.89x faster than manual [JCCT 2023].



Our guided neuron proofreading is 7.5x faster than interactive error correction [CVPR 2018].