Deep Learning–Based Ultrasound Lesion Segmentation and Subtype Classification

Applicant: Varsha Venkatapathy, University of North Carolina at Chapel Hill
Department: Joint Department of Biomedical Engineering

Abstract

Despite its widespread use, current clinical and research workflows for breast ultrasound (US)—a radiation-free, bedside imaging modality—remain hampered by several critical limitations. First, lesion delineation is still largely manual, leading to inconsistent region-of-interest selection and high inter-operator variability. Second, many machine-learning (ML) approaches either ignore precise tumor localization—working on full frames and diluting diagnostic signal—or depend on slow, handcrafted radiomic feature extraction that cannot scale. Third, few systems incorporate attention mechanisms, which human experts use instinctively to focus on subtle textural or morphological cues such as spiculations or acoustic shadowing. Finally, real-time deployment across multiple ultrasound platforms is virtually nonexistent, restricting these tools to offline research. Our project directly targets each shortcoming by (1) automating lesion segmentation with an attention-enhanced U-Net—eliminating manual tracing and standardizing region selection; (2) integrating squeeze-and-excitation and self-attention modules into a ResNet classifier to amplify critical sub-visual features without costly handcrafted features; and (3) packaging the end-to-end pipeline into a lightweight inference engine validated on three commercial scanners, guaranteeing subtype predictions in under 50 milliseconds per image for seamless point-of-care use.

Intellectual Merit

Breast ultrasound (US) —a safe, real-time imaging modality that uses sound waves instead of ionizing radiation—remains highly operator-dependent and shows significant variability between readers. Recent studies, however, demonstrate the strong potential of machine learning (ML)—algorithms that learn patterns directly from data—to improve diagnostic consistency on US images. For example, Ferre et al. (2023) extracted 249 radiomic features (quantitative measurements of texture and shape) from B-mode (brightness-mode) scans of 88 patients and trained a logistic regression (LR) model to distinguish triple-negative from non-triple-negative tumors (area under the receiver-operating-characteristic curve, or AUC = 0.824) and HER2-positive vs HER2-negative lesions (AUC = 0.778). Pacal (2022) applied a Vision Transformer—a deep neural network architecture that uses self-attention to weight the importance of different image regions—to 780 images from the public BUSI dataset, achieving 88.6 % overall accuracy and an F1-score (the harmonic mean of precision and recall) of 88.7 %. Michael et al. (2022) used a LightGBM gradient-boosting machine—an ensemble of decision trees tuned via Bayesian optimization (Tree-structured Parzen Estimator)—on a proprietary cohort of approximately 912 cases, reporting 99.86 % accuracy and 99.6 % recall (sensitivity). Finally, Becker et al. (2018) trained a generic convolutional neural network (CNN)—a class of ML models specialized for image tasks—on 632 mixed benign and malignant lesions, achieving AUC 0.84 with an ultra-fast inference time (the time to process one image) of just 0.019 seconds. While these approaches establish ML's feasibility on ultrasound, they each leave gaps in full automation, precise tumor localization, attention-driven feature extraction, and real-time clinical deployment. To fill these gaps, our integrated pipeline (1) employs an attention-enhanced U-Net—an encoder–decoder CNN with spatial and channel attention modules—to generate high-fidelity lesion masks without manual tracing, (2) uses an attention-augmented ResNet, a convolutional neural network architecture with "residual" shortcut connections that ease the training of very deep models, further enhanced by squeeze-and-excitation and self-attention layers to focus on critical textural and morphological cues for accurate subtype classification, and (3) packages the entire workflow into a lightweight inference engine validated on three different clinical ultrasound systems, delivering segmentation and classification in under 50 milliseconds per image.

Specific Aims

Aim 1: We will train a residual U-Net—a convolutional encoder–decoder network with "skip" connections that preserve spatial detail—augmented by spatial and channel attention modules on 100 expert-annotated B-mode ultrasound volumes. Our objective is to achieve a Dice coefficient ≥ 0.80, a measure of overlap defined as twice the shared area divided by the sum of the predicted and true mask areas, and an Intersection-over-Union (IoU) ≥ 0.75, the ratio of overlap area to the union of predicted and true mask areas, ensuring precise, fully automated delineation of lesion boundaries without any manual tracing.

Aim 2: Building on the segmentation masks from Aim 1, we will develop a ResNet (Residual Network)—a deep CNN architecture with shortcut connections to ease training—enhanced by squeeze-and-excitation blocks (which adaptively reweight feature channels) and self-attention layers to focus on critical textural and morphological cues. We will train and validate this classifier on an external cohort of 500 cases, targeting an area under the receiver operating characteristic curve (AUC) ≥ 0.85 for molecular subtype discrimination, and then package the entire end-to-end pipeline into a lightweight application optimized to run inference in under 50 milliseconds per image on three different clinical ultrasound platforms

Preliminary Data

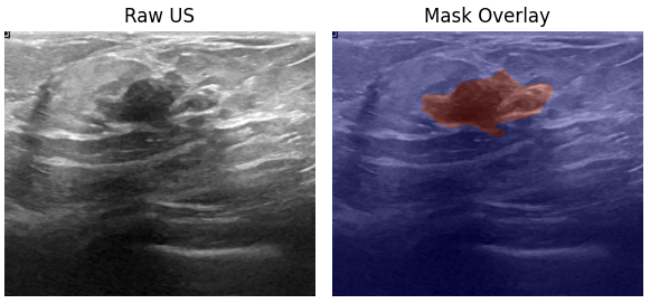

Fig. 1 shows a raw B-mode ultrasound frame alongside its expert-annotated lesion mask. Our proof-of-concept pipeline crops each image to that mask and classifies it with a lightweight ResNet (N = 130 test cases), yielding 65 % accuracy, AUC = 0.55 (vs. 0.50 chance) and 0.11 s/image inference. This confirms end-to-end feasibility—segmentation followed by focused classification—but the modest AUC establishes a baseline: by incorporating spatial and channel attention in Aim 1 and Aim 2, we anticipate boosting AUC to ≥ 0.85 for robust subtype discrimination.

Research Plan/Timeline

Over the first six months, we will curate a multi-institutional dataset of 500 expert-annotated B-mode ultrasound volumes and train a baseline U-Net model for lesion segmentation. In months 7–12, we will integrate spatial and channel attention modules into the U-Net and validate its performance on held-out data, targeting a Dice coefficient of ≥ 0.80 and an Intersection-over-Union (IoU) of ≥ 0.75 for precise lesion delineation. During months 13–18, we will develop a ResNet classifier enhanced by squeeze-and-excitation blocks and self-attention layers, optimizing molecular subtype prediction through five-fold cross-validation on 500 labeled cases. Finally, in months 19–24, we will package the full segmentation-plus-classification pipeline into a lightweight graphical user interface (GUI) or web application, conduct cross-scanner validation to ensure classification accuracy ≥ 0.80 and inference latency ≤ 50 ms per image, and release all code, models, and documentation as open-source tools.

Broader Impacts

This open-source pipeline will enhance ultrasound diagnosis in resource-limited settings—reducing unnecessary biopsies and accelerating treatment—while fostering equitable rural healthcare and laying the groundwork for future AI diagnostics. It will train the next generation of biomedical imaging researchers by mentoring two undergraduates (one from an underrepresented group), hosting workshops and a “Women in Medical AI” seminar, and co-leading summer internships for students from minority-serving institutions, with all materials shared via webinars and the project website. PI approval, retrospective datasets, GPU compute resources, and clinical scanners are already secured.

Reference

[1] Ferre R, Elst J, Senthilnathan S, et al. Breast Dis. 2023;42:59–66; [2] Pacal I. J Inst Sci Technol. 2022;12(4):1917–1927; [3] Michael E, Ma H, Li H, Qi S. Biomed Res Int. 2022;2022:8482022; [4] Becker AS, Mueller M, Stoffel E, et al. Br J Radiol. 2018;91(1083):20170576.