

1 **No Sonographer, no Radiologist: An AI-enabled Comprehensive**
2 **Breast Ultrasound Diagnostic System for Low-Resource**
3 **Settings**

4 **Authors:** Emilio J. Ochoa¹, Luis C. Revilla², Stefano E. Romero, ²Giancarlo A.
5 Guarnizo², Thomas J. Marini^{3,5}, Kevin J. Parker⁴, Yu Zhao^{3,5}, Galen Brennan^{3,5}, Jonah
6 Kan⁶, Steven Meng⁷, Ann Dozier⁸, Anna Weiss^{9,10}, Benjamin Castaneda¹

7 ¹*Department of Biomedical Engineering, University of Rochester, Rochester, New York, USA;* ²
8 *Departamento de Ingeniería, Pontificia Universidad Católica del Perú, Lima, Perú;* ³*Department of*
9 *Imaging Sciences, University of Rochester Medical Center, Rochester, New York, USA;* ⁴*Department*
10 *of Electrical Engineering, University of Rochester, New York, USA;* ⁵*University of Rochester Medical*
11 *Center, Rochester, New York, USA;* ⁶*Department of Internal Medicine, Mayo Clinic, Rochester,*
12 *Minnesota, USA;* ⁷*Department of Radiology, UT Southwestern Medical Center, Dallas, Texas, USA;* ⁸
13 *Department of Public Health, University of Rochester Medica Center, Rochester, New York, USA;* ⁹
14 *Division of Surgical Oncology, Department of Surgery, University of Rochester Medical Center,*
15 *Rochester, New York, USA;* ¹⁰*Wilmot Cancer Center, University of Rochester Medical Center,*
16 *Rochester, New York, USA.*

17 **Keywords:** VSI, breast ultrasound, breast cancer research, artificial intelligence, telemedicine.

18 **Corresponding Author:** Emilio J. Ochoa. Department of Biomedical Engineering, University
19 of Rochester, Rochester, New York, USA Email: eochoaal@ur.rochester.edu

20
21 **Abstract**

22 Background
23 Breast cancer is the most common non-skin related malignancy and the leading cause of cancer death
24 in women. Mammography remains the gold standard for early detection; however, its accessibility is
25 limited in low-resource settings due to cost and technical complexity. Ultrasound (US) is a viable

26 alternative, but its implementation is hindered by the scarcity of trained radiologists and sonographers.
27 Volume sweep imaging (VSI) has addressed the issue of US acquisition by enabling non-specialists to
28 perform standardized scans. However, these still require expert interpretation, limiting their impact. To
29 overcome this barrier, we propose a fully automated Breast VSI (VSI-B) system integrating artificial
30 intelligence (AI) for segmentation and classification of breast lesions, aiming to provide an accessible
31 diagnostic tool for low-resource environments.

32 **Methods**

33 This study developed an AI-driven diagnostic system for VSI-B, combining a segmentation model
34 (Attention U-Net 3D) with a classification model for lesion detection. A total of 98 patients with
35 palpable breast lumps were included in the study. The dataset consisted of 392 VSI-B US videos and
36 2,100 classified frames. A new method was implemented to enhance mass identification by selecting
37 key frames for analysis. A majority voting algorithm was used to optimize lesion classification. The
38 system's performance was assessed based on sensitivity, specificity, and accuracy.

39 **Results**

40 Following a detection step that achieved 100% sensitivity and 93.6% specificity for cancer and no
41 cancer patients, as well as 95.0% sensitivity and 63.0% specificity for mass and no mass patients, the
42 DenseNet classification model reached 87% accuracy, 100% sensitivity, and 83% specificity. A
43 majority voting algorithm optimized classification, yielding an AUC of 0.91.

44 **Conclusions**

45 This study highlights the potential of an AI-enabled VSI-B system as a reliable diagnostic tool in low-
46 resource settings. By integrating multi-modal segmentation and classification, the system automates
47 breast lesion detection and stratification, reducing reliance on radiologists. The results suggest that this
48 approach could enhance early breast cancer diagnosis and guide clinical decision-making, particularly
49 in underserved regions.

50

51 **Introduction**

52 Approximately 2.1 million women are diagnosed with breast cancer worldwide each year, and
53 approximately 685,000 deaths occur annually (1). Breast cancer incidence and mortality are expected

54 to increase, and it is projected to reach more than 3.2 million new cases by 2030(2), with at least 60%
55 of deaths in low- and middle-income countries (LMICs)(3).

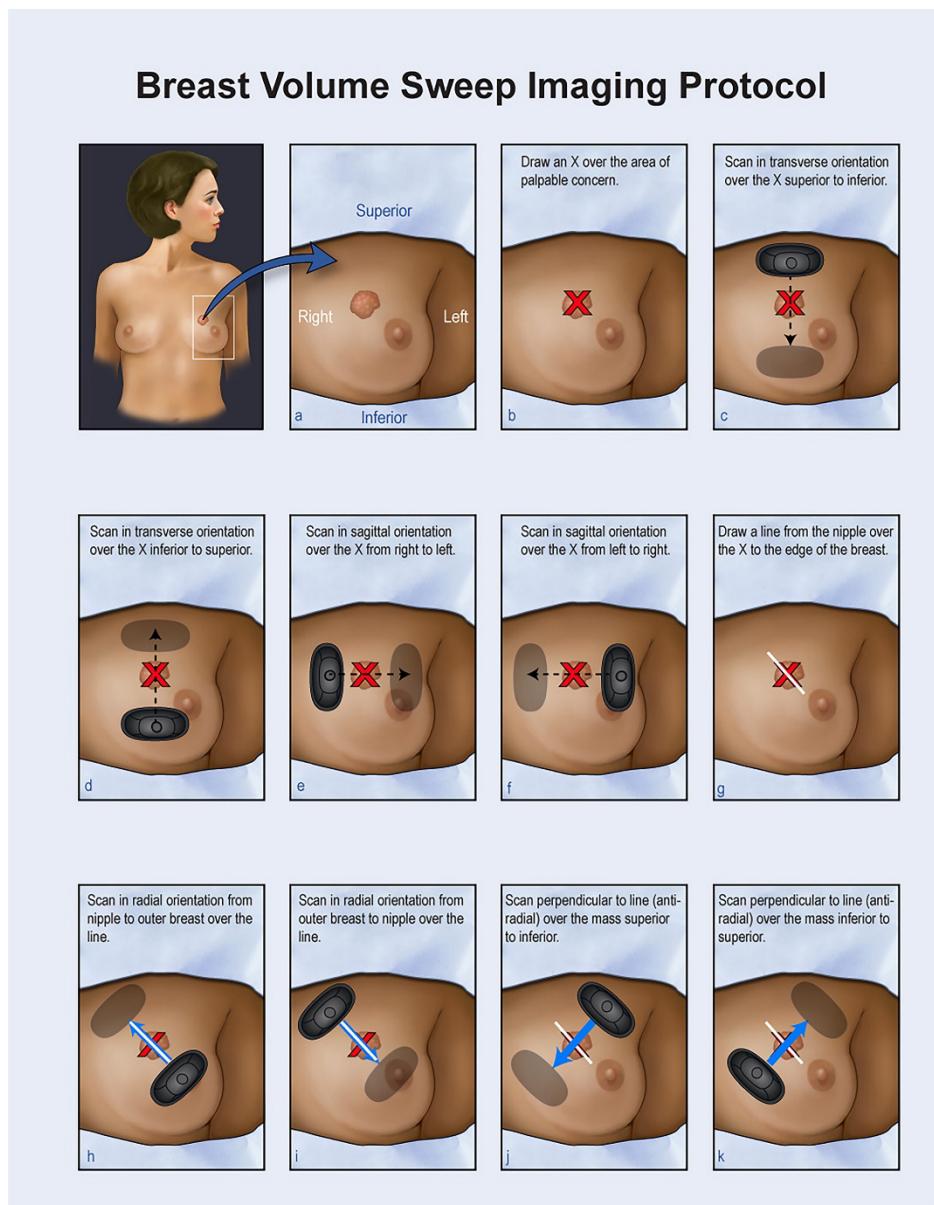
56 Early diagnosis by mammography is often associated with the term screening in high-income
57 countries, based on the assumption that mammography will always be required (4). Unfortunately, use
58 of mammography equipment in LMICs is limited by the shortage of human and economic resources,
59 especially in rural zones. Mexico has previously reported that less than 10% of patients are diagnosed
60 with early stage breast cancer whereas 47% are diagnosed with most advanced stages (III and IV) (5). In
61 a recent study of breast cancer diagnosis delays in women in Peru, 93% of the subjects self-diagnosed
62 their own breast cancer, with an average time of 407 days from the onset of symptoms to the start of
63 treatment (6). Given the severity of this public health problem, novel solutions are urgently needed to
64 improve the accessibility of medical imaging in LMICs, especially in rural regions. Early detection
65 drastically improves outcomes: when breast cancer is diagnosed at Stage I—defined as a tumor 2 cm or
66 smaller with no lymph node involvement (Stage IA) or a similarly small tumor with microscopic nodal
67 metastases (Stage IB), the five-year breast cancer-specific survival rate is 98 %–100 %. By contrast,
68 Stage III disease—characterized by locally advanced tumors with significant regional lymph node
69 spread or invasion of the chest wall or skin, but without distant metastases—carries a five-year survival
70 of only 66 %–98 % (7).

71 Point-of-care ultrasound (US) is a portable, noninvasive, and cost-effective medical imaging
72 modality with the potential to improve diagnosis accessibility. However, deployment of this modality
73 requires trained specialists for both acquisition and diagnosis with years of experience, increasing the
74 risk of late diagnosis and, therefore, of severe or fatal complications (8). To standardize and reduce the
75 training for the acquisition process; Volume sweep imaging (VSI) is a streamlined and time-efficient
76 asynchronous standardized protocol to enable the acquisition of US videos by inexperienced personnel
77 and has been validated for obstetric, lung, right upper quadrant, thyroid and breast scanning (8-20).

78 In the breast VSI (VSI-B) protocol, the operator, after a 2-hour training, sweeps the US probe
79 over the target region to obtain a complete volumetric acquisition. The video clips are then sent to a
80 specialist for interpretation and diagnosis. These sweeps are standardized and based exclusively on
81 external body landmarks. Figure 1 illustrates the VSI-B protocol, outlining the steps required to obtain

82 a volumetric acquisition of the breast using an US probe. High levels of agreement between the VSI-B
83 and the standard of care for mass visualization have been reported (Cohen's Kappa 0.95; CI: 0.89-1).
84 However, there is still no automation process to interpret the images; VSI-B still relies on specialists
85 which are experiencing broad shortages, resulting in an increased patient load for diagnosis and an
86 immediate result even more challenging (8).

87



88

89 *Figure 1.* Illustration of the standardized VSI-B protocol for breast mass evaluation. The process
90 involves **8 US sweeps** over the area of palpable concern. (a) The area of interest is identified. (b) An
91 'X' is marked over the palpable mass. (c-f) Transverse and sagittal sweeps are performed in superior-

92 to-inferior, inferior-to-superior, right-to-left, and left-to-right directions. (g) A line is drawn from the
93 nipple to the marked area. (h-i) Radial sweeps are performed along this line, from nipple to outer
94 breast and back. (j-k) Anti-radial sweeps are done perpendicular to the line, covering superior-to-
95 inferior and inferior-to-superior directions. This systematic approach enhances mass visualization (8).

96 Artificial intelligence (AI) may be the answer to this problem. AI has demonstrated potential to
97 improve the accuracy of breast cancer diagnosis by US, resulting in high sensitivity (85%; 95% CI:
98 70%-94%) and specificity (73%; 95% CI: 56%-85%) values using 2D US datasets (21).

99 The transition to incorporate AI models within medical equipment begins with systems or
100 software to aid in the detection and characterization of lesions or tissue abnormalities (22). A recent
101 study evaluated the accuracy of Samsung's S-Detect model, a proprietary system requiring images from
102 Samsung devices, for the classification of breast lesions using one US image per patient from the VSI-
103 B dataset, selected by a physician. The results demonstrated a high level of agreement between the
104 model's classifications of cancer, cysts, fibroadenomas, and lipomas, compared to the standard expert
105 report ($\kappa = 0.73$). Furthermore, the model achieved a sensitivity of 100% in detecting the 20 malignant
106 cases, along with a specificity of 86% (14).

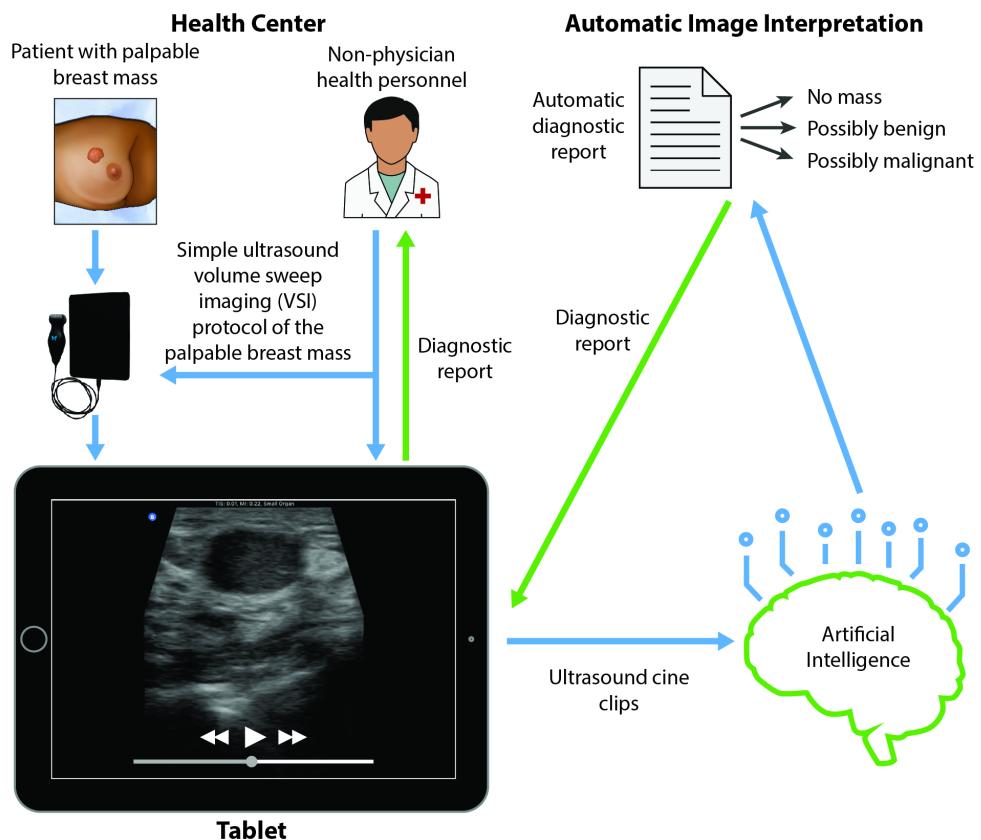
107 Other state-of-the-art architectures have been proposed, such as the Attention U-Net and the
108 Sharp attention U-Net(23). Attention U-Net 3D developed important improvements over the classical
109 model, such as the soft attention mechanism. These architectures have already been applied and tested
110 in US imaging with the breast US images (BUSI) dataset (24). However, it is important to note that the
111 BUSI dataset is not specifically designed for low-resource settings; it consists only of segmented frames
112 rather than video sequences, unlike the VSI-B protocol, which is better suited for such environments.
113 Models for segmentation using the VSI protocol have been proposed in both obstetric (15, 25, 26) and
114 breast (27, 28) protocols. Systems that integrate both detection and classification of breast nodules have
115 not yet been developed in low-resource areas.

116 The present research aimed to develop an automated system to support the segmentation,
117 detection and classification of palpable breast lumps for the early diagnosis of breast cancer. Our system
118 was designed with portable, inexpensive, VSI-B US data for conditions where a radiologist or

119 sonographer is not available, so that a provisional recommendation can be determined where no other
120 local option exists.

121

122 **Material and Methods**

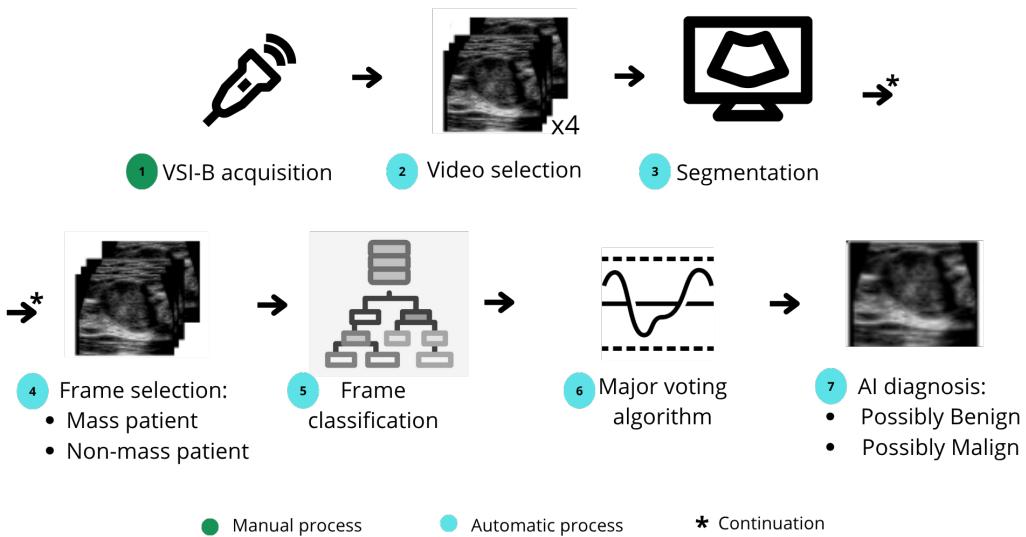


123

124 *Figure 2. Illustrated workflow of an automated breast lesion assessment system. A non-physician*
125 *health worker performs a simple VSI-B protocol on a patient with a palpable breast mass, capturing*
126 *US cine clips on a tablet. These clips are then uploaded to an artificial intelligence (AI) system for*
127 *automatic image interpretation, producing a diagnostic report that categorizes the lesion as “no*
128 *mass,” “possibly benign,” or “possibly malignant”. Blue arrows represent the technical flow of*
129 *information: ultrasound acquisition, image transfer, and automated interpretation. Green arrows*
130 *represent the diagnostic feedback loop, showing how the final report is transmitted back to the health*
131 *worker to guide clinical decision-making.*

132

133 System description



134

135 *Figure 3. Proposed technical system. Step 1 involves a direct intervention from the healthcare*
 136 *personnel, while steps 2-8 are an automated process.*

137

138 The proposed system for this project is shown in Figure 2. A more detailed technical description is
 139 listed here in nine steps (Figure 3). Step one involves direct intervention from healthcare personnel.
 140 Steps two to eight are an automated process:

- 141 1. Clinical evaluation, referral of the patient to medical imaging and acquisition of VSI-B
 142 protocol.
- 143 2. Preprocessing of the data for segmentation.
- 144 3. Segmentation with Attention U-Net 3D, followed.
- 145 4. Post-processing with a Gaussian mask. A discrimination algorithm defines whether it is a “non-
 146 mass” patient or a “mass” patient to pass to the binary classification stage.
- 147 5. Selection of four videos and five representative frames for each video.
- 148 6. Frame classification of lesions is conducted: “Possibly benign” or “Possibly malignant”.
- 149 7. Threshold algorithm decision.
- 150 8. Binary results are delivered to support the classification for the physician's analysis.

151

152 These steps will be described in further detail below, along with the one-time initialization of the data
153 and training of the model.

154

155 1) VSI-B dataset acquisition and ground truth

156 US data was collected in a previous work by the University of Rochester, New York(8); using
157 the VSI protocol applied to breast as shown in Figure 1. The VSI-B protocol was carried out by
158 healthcare workers in clinics after a two-hour training. A hand-held US probe IQ+ (Butterfly Network,
159 MA, USA) with a small organs pre-setting was used. This probe was connected to a tablet for
160 acquisition. Results were compared with standard-of-care (SOC) US images from a Logiq e10 scanner
161 (General Electric, MA, USA) or an Epiq 7G scanner (Phillips, Amsterdam) (8).

162 The sample set consisted of 160 patients with a total of 170 palpable breast lumps. For our
163 evaluation, we included all patients with a cyst, cancer, fibroadenoma, or normal exam. As this is a
164 preliminary study of the potential of AI, we excluded any discrepant cases between VSI and standard
165 of care, any other category of lesions, and cases where the lesion was unsegmentable. Pathology was
166 determined by biopsy whenever possible. However, in cases where biopsy was not performed, BI-
167 RADS 2 or 3 were considered benign. Patients categorized as "non-mass" in the ground truth did not
168 present any masses on US imaging. Patients categorized with a "mass" had either a cyst, cancer, or
169 fibroadenoma. For patients with a detected mass, "non-cancer" indicates that the mass was visible on
170 US, but SOC determined that it was not malignant (non-mass, fibroadenoma, cyst). Finally, cases with
171 malignancy were "cancer."

172 For patients with detected masses, four US videos were selected per patient, one of each type
173 of sweep (transverse, sagittal, radial, and antiradial). Sweeps were chosen between each pair based on
174 the sweep visualizing the largest portion of the mass. If both sweeps showed equal amounts of the mass,
175 the first sweep was chosen. These videos were analyzed and labeled by a trained radiologist using the
176 "Video Labeler App" tool from MATLAB R2021 (MathWorks, Natick, Massachusetts, USA),
177 generating masks to delineate tumor shapes.

178 **Table 1** provides a detailed breakdown of patients with confirmed diagnoses of cancer or non-cancer,
179 including those with cysts and fibroadenomas, resulting in a total of 98 subjects.

180

181 Due to the variable length of the videos, ground truth masks were generated only on the areas
182 where the tumor was located (Figure 4). The size of the video frames with their labeled masks were
183 originally 1696×1080 pixels. Frames with ground truth annotations assigned were interpolated in the
184 intermediate frames since the doctor segmented the ground truth by skipping one frame(29, 30). Black
185 masks were created as background in regions without labels. These steps produced masks for the entire
186 video, allowing for a comparison between the medical ground truth and the predicted outcomes. For
187 non-mass patients, entirely black labels were generated for all the videos.

188 2) Preparation of videos for segmentation

189 To preprocess the segmentation dataset, we selected the four labeled sweeps per patient. To keep the
190 “no-mass” cohort balanced, we likewise chose four sweeps for each of those patients and applied black
191 masks to them, yielding a total of 392 videos (98 patients). In the 2D domain, 1696×1080 pixels were
192 cropped to 912×912 to only consider the US images and not the Butterfly IQ+ interface. After this, a
193 resize to 128×128 was proposed due to computational limitations. The video length was resized to 128
194 frames to obtain 3D volumes, ensuring an average resize between all the videos (average length of the
195 videos: 222 frames). Contrast limited adaptive histogram equalization (CLAHE) was applied as a
196 preprocessing step. A K-fold cross-validation ($k=10$) was performed as a method of training and
197 validation to maximize data utilization while obtaining stable performance estimates across folds in a
198 relatively small video dataset. To evaluate the model's performance, two metrics were utilized: the Dice
199 coefficient and accuracy, as specified in Equations 1 and 2, respectively. The Dice coefficient, as shown
200 in Equation 1, quantifies the similarity between two sets A and B , where A represents the predicted
201 segmentation mask and B represents the ground truth segmentation mask. Accuracy, on the other hand,
202 refers to the proportion of total predictions that the model correctly classifies as either mass or non-
203 mass patients.

204

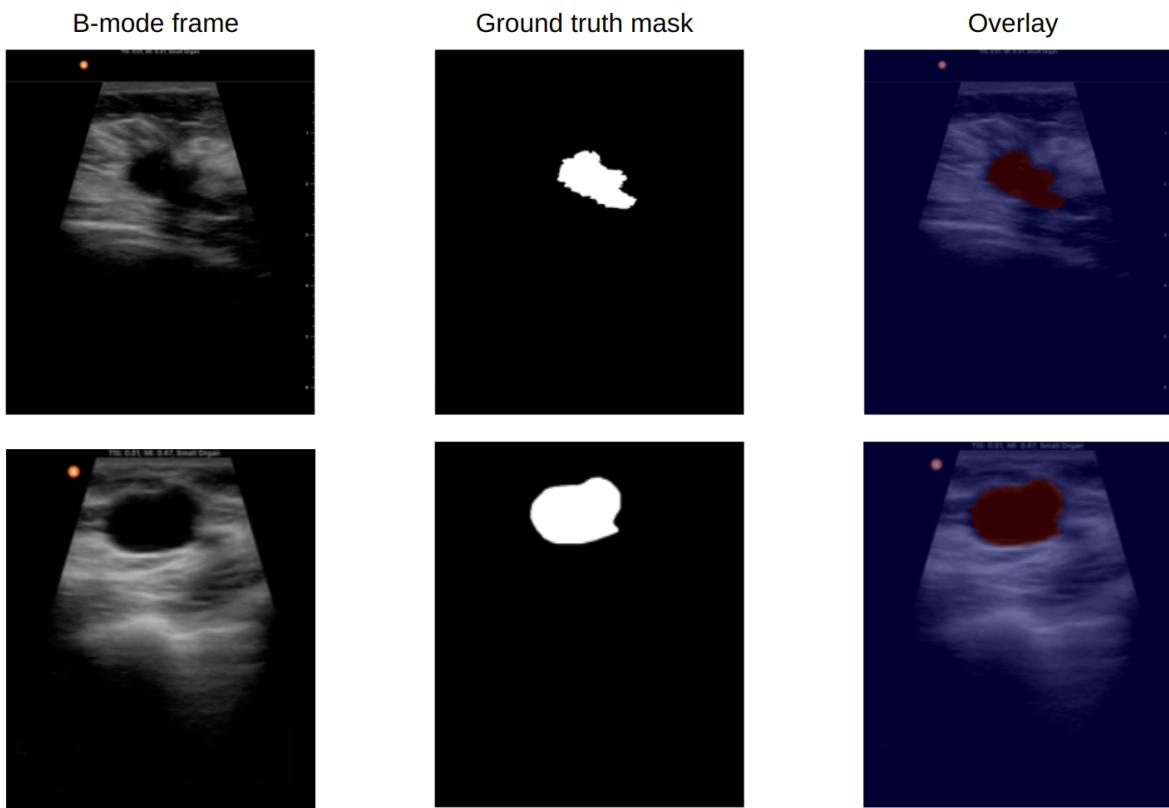
205
$$Dice = \frac{2 \|A \cap B\|}{\|A\| + \|B\|} \quad (\text{Eq. 1})$$

206

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}$$

208 (Eq. 2)

209



210

211 *Figure 4. B-mode image, physician-segmented mask, and overlay in two different patients. Top:*
 212 *segmented tumor represents a malignant case in the sixth sweep (radial). Bottom: benign tumor in the*
 213 *eighth sweep (perpendicular).*

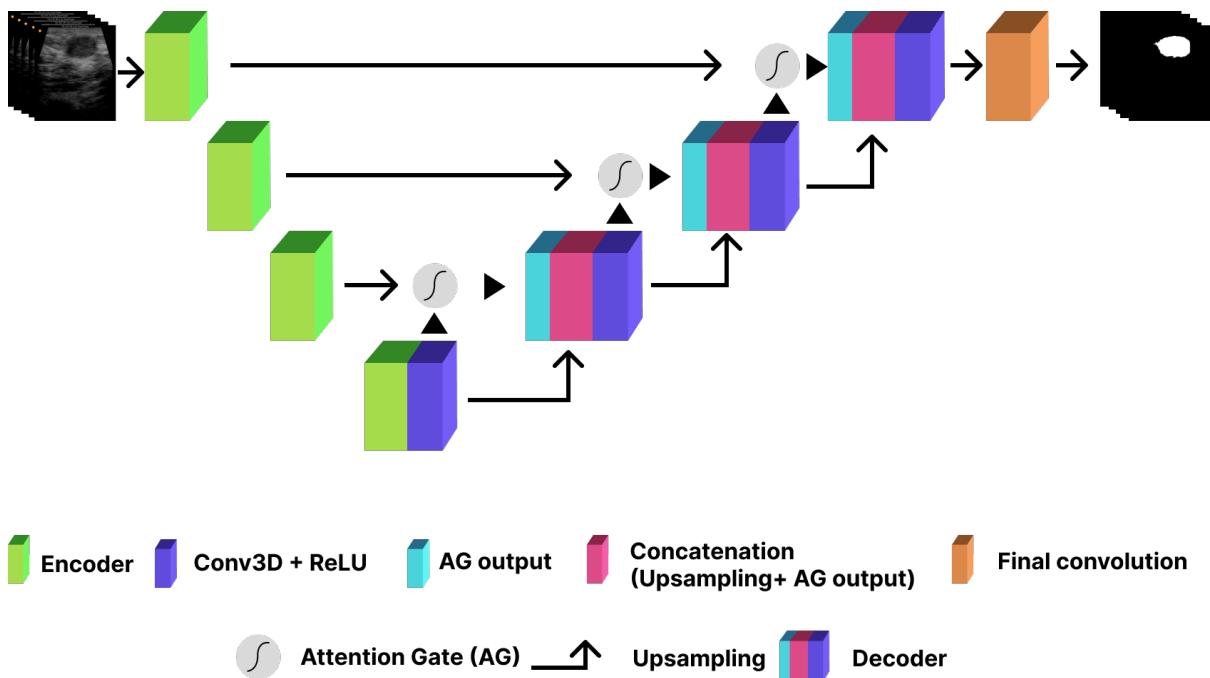
214

215 3) Segmentation

216 Tumor segmentation was performed by Attention U-Net 3D (Figure 5), an architecture tested
 217 on the BraTS 2019 dataset (Multimodal Brain Tumor Segmentation Challenge dataset) (31). As a 3D
 218 attention map was obtained, average pooling for channel correlation was performed in parallel. Finally,
 219 skip-connections were fused to reduce sparsity and singularity, also improving generic learning and
 220 segmentation prediction. Attention U-Net 3D hyperparameters included: 90 epochs, a batch size of 1,
 221 Binary Cross Entropy Dice loss with 0.5 weight, and a net depth of 4 layers (16 to 128 filters; 256 filters

222 in the bridge). Training was processed on a 3080 GPU (NVIDIA, Santa Clara, California, USA) using
223 PyTorch Lightning 2.0 (32) and Python 3.9.12.

224



225

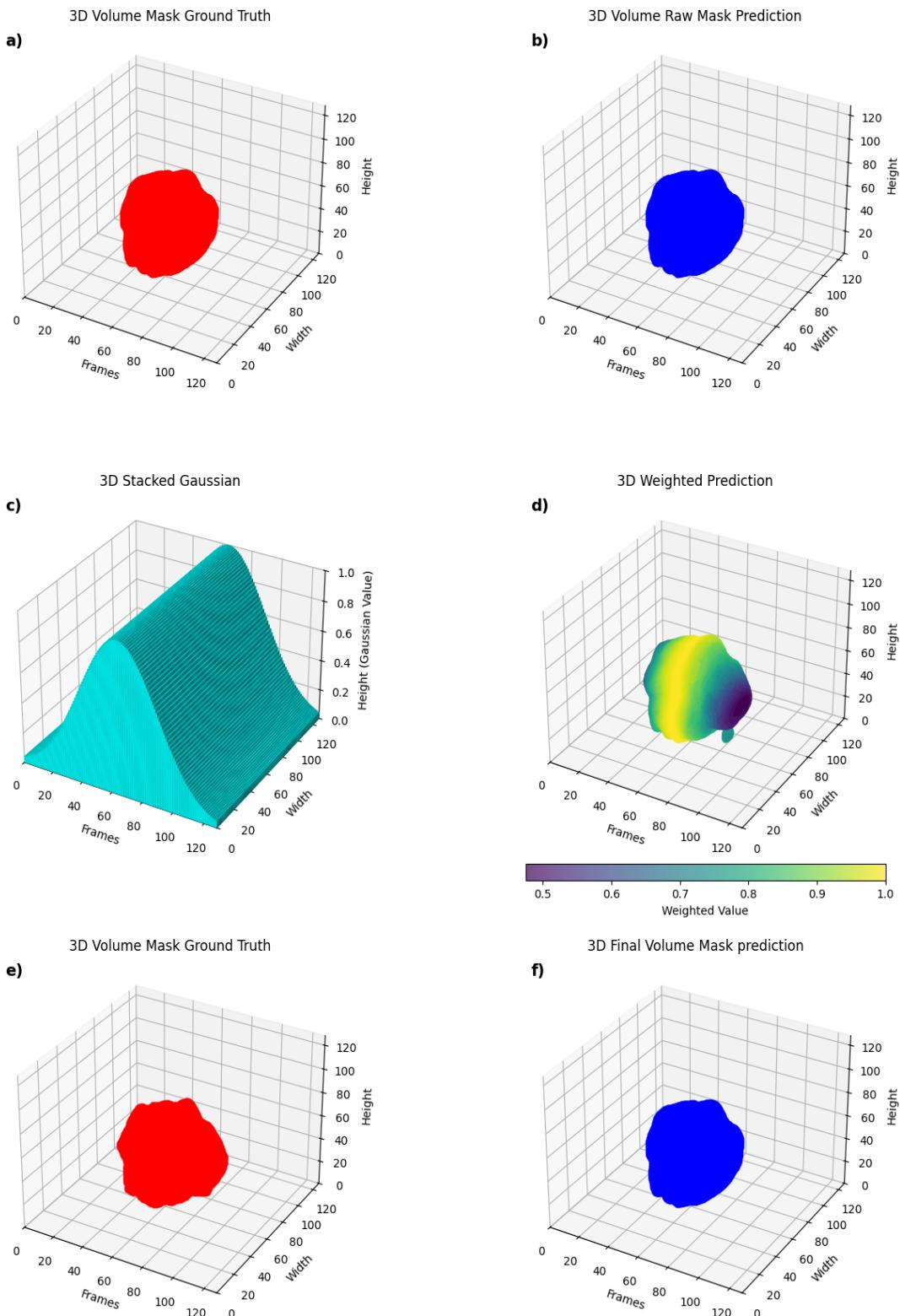
226 *Figure 5.* Diagram depicting the architecture of the 3D Attention U-Net model, designed for
227 volumetric medical image segmentation. The model processes 3D US volumes as input, applying
228 multiple encoding and decoding blocks with attention mechanisms to enhance the segmentation of
229 relevant structures. The encoding path captures spatial and contextual information through
230 convolutional and downsampling operations, while the decoding path progressively reconstructs the
231 segmented volume using upsampling and concatenation. Attention gates are integrated to suppress
232 irrelevant regions and focus on salient anatomical features, improving the precision of segmentation.

233

234 *4) Mask postprocessing*

235 To improve the tumor detection process using US imaging, the analysis began with a plot of
236 the normalized mask area versus frame, where 100% indicated that a full mask covered the entire image.
237 This plot was independent by volume: if a frame contained two separate masks, they were treated as
238 distinct volumes and represented by two individual area measurements. This approach, illustrated in

239 Figure 6, enabled a precise analysis of each potential tumor site as an independent entity, ensuring that
240 the data was not conflated.



241

242 *Figure 6. Mask postprocessing with a 3D stacked Gaussian filter. Panels (a) and (e) show the*
243 *original 3D volume ground truth mask for comparison (same figure). Panel (b) presents the raw 3D*

volume mask prediction obtained directly from the model, which exhibits initial irregularities. Panel (c) demonstrates the 3D stacked Gaussian filter applied to the mask, where the Gaussian values are distributed across the volume, as indicated by the smooth surface. Panel (d) visualizes the weighted prediction after applying the Gaussian filter, highlighting the refined probabilistic values with a colormap that scales from 0.5 to 1.0. Finally, panel (f) displays the postprocessed 3D volume mask prediction.

250

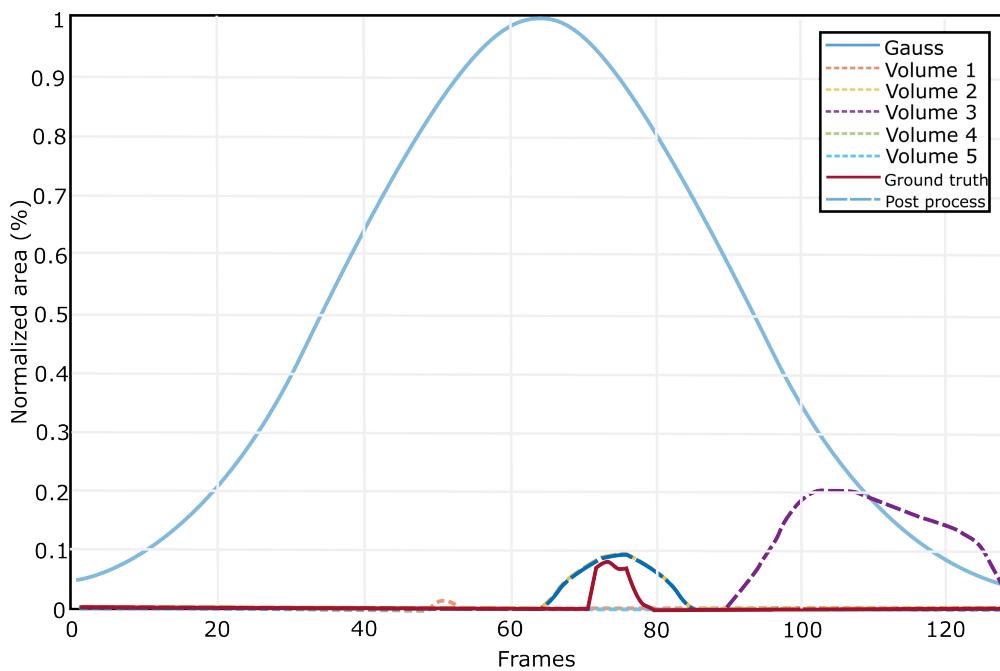
To enhance the accuracy of predictions, a 3D Gaussian window was applied to this plot. Figure 6c shows this Gaussian voxel and its outputs, which played a vital role in smoothing the data outputs from our Attention U-Net 3D model, reducing temporal and spatial noise and emphasizing areas with the highest segmentation probability, assigning greater weight to the central frames—where the likelihood of correctly identifying a tumor was higher—while reducing the influence of peripheral frames, which were more prone to noise. By multiplying the Gaussian voxel with the segmented mask, noise was effectively minimized, allowing for the extraction of more relevant information.

After applying the Gaussian filter, every data point below a 0.01% normalized area was used to eliminate small, irrelevant areas considered as noise, further refining the data. The threshold value of 0.01% was tuned based on a range from 0.005% to 0.04%; the highest detection accuracy was achieved at 0.01%. This threshold was particularly important because it corresponded to a tumor diameter of approximately 0.2 cm, a size generally not considered clinically significant for breast cancer. This approach allowed the focus to remain on more substantial and potentially hazardous tumor indications, ensuring that the analysis was both precise and clinically relevant without being overly conservative.

Finally, after thresholding, a volume that exhibits the highest peak area in the plot was selected. Figure 7 shows this pivotal selection to identify the most significant volumetric data in a 2D plot, representing the most probable tumor presences. The protocol is explicitly designed for assessment around a single clinically palpable lesion per acquisition. When clinical suspicion indicates multiple palpable lesions, the protocol specifies separate VSI acquisitions for each lesion. Therefore, our current model assumes the presence of a single lesion per VSI scan and is not intended for segmentation and separation of multiple simultaneous lesions.

272 To address the issue of false segmentations in non-mass patients and refine the classification of
273 patients as either mass or non-mass, the post-processing discrimination algorithm in Figure 8 was
274 implemented. This discrimination algorithm evaluates the number of masks generated per video for
275 each patient and decides which patient will be detected as a non-mass or mass patient based on
276 predefined criteria. Initially, for each patient, the presence of segmented masks is counted across the
277 videos. The algorithm evaluates the presence of segmentation masks across four predictions per patient.
278 If two or more predictions are classified as "empty" (no mass detected), the patient is classified as a
279 "non-mass" patient. Conversely, if fewer than two predictions are empty, the patient is classified as
280 having a "presence of mass" and proceeds to the classifier for further analysis. This threshold was
281 selected after validation across different thresholds to balance sensitivity and specificity.

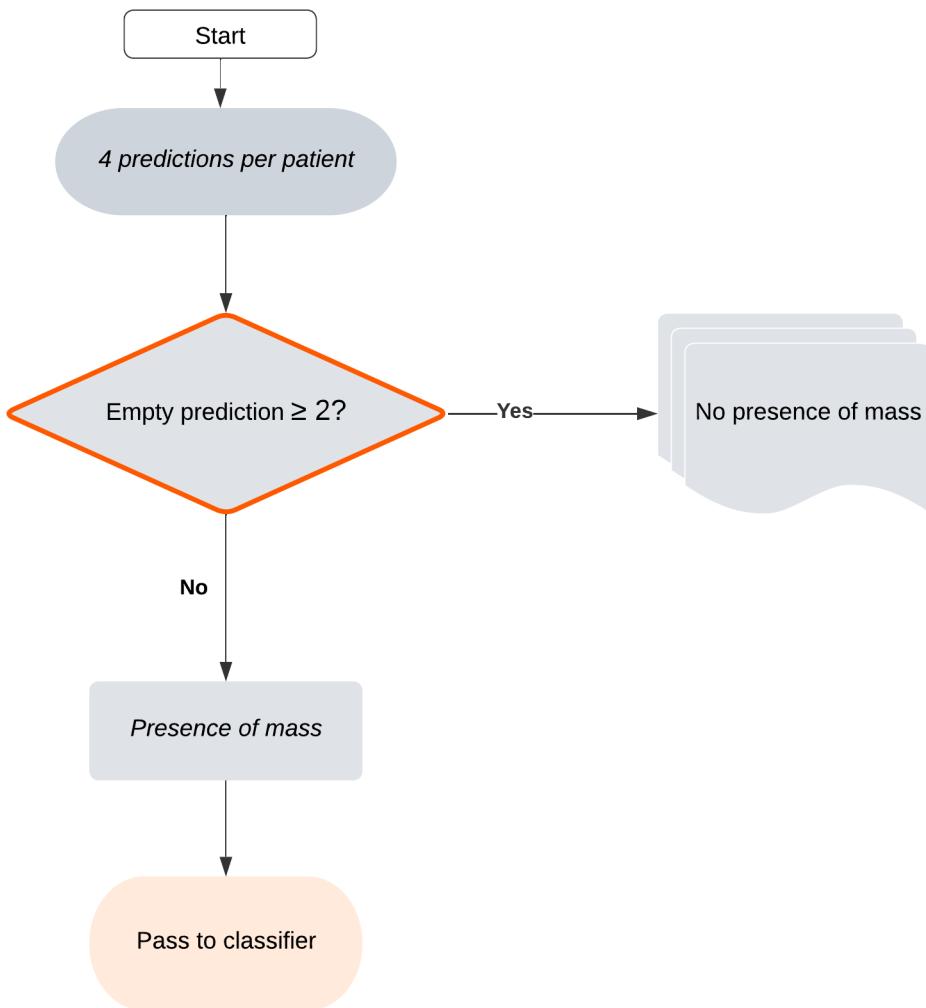
282



283

284 *Figure 7. Normalized mask area vs. frame after post-processing in a 2D representation. While five*
285 *segmented volumes were originally detected, the post-processing step correctly identifies the ground-*
286 *truth volume as the only valid mask. Volumes 1, 2, and 4 are removed using a 0.01% threshold,*
287 *Volume 3 is discarded based on the Gaussian probability plot, and Volume 5 remains as the final*
288 *desired mask.*

289



290

291 *Figure 8. Discrimination algorithm of segmentations before classifying breast US frames. Patients*
 292 *who had two or more videos with empty masks were considered as non-mass patients and discarded*
 293 *for the classification stage.*

294

295 5) Frames selection for classification

296 All selected frames maintained the original resolution for classification. Since classification
 297 was a 2D task, similarity between the ground truth of the same patient was addressed through a data
 298 cleaning process. For this purpose, 5 frames per video per patient were initially selected (all selected
 299 frames had been labeled ground-truth): the frame with the largest tumor area, the two previous frames,
 300 and the two subsequent frames. Previous and subsequent frames were randomly selected with the

301 condition of exceeding 80% of the tumor area size of the frames. The target assignment for these frames
302 was based on histopathology results, with cancer frames considered malignant and all other frames
303 considered benign. Final data for training consisted of 2100 frames. The resize of 224×224 pixels was
304 used as it is a standard input dimension for many convolutional neural networks (CNNs) and has been
305 empirically determined to offer a balance between computational efficiency and sufficient detail
306 retention for effective feature extraction in image classification tasks. Leave-one-out cross-validation
307 (LOOCV) was performed as the training dataset was split into different folds (one per patient), resulting
308 in 60 folds as it is more suitable for limited sample sizes and ensures complete independence between
309 training and testing data by leaving out all data from one patient at a time.

310 6) Frame classification

311 Training was carried out on five CNN architectures, commonly used in deep learning (DL) for
312 image recognition tasks, particularly in medical imaging: MobileNet, DenseNet, ResNet, Vgg16, and
313 Inception (33-37). MobileNet utilizes depthwise separable convolutions to significantly reduce the
314 number of parameters, making it efficient for mobile and embedded vision applications. DenseNet
315 connects each layer to every other layer in a feed-forward fashion, which improves gradient flow and
316 encourages feature reuse. ResNet employs residual learning by introducing shortcut connections that
317 bypass one or more layers, allowing for the training of very deep networks. Vgg16, known for its
318 simplicity, uses 16 layers with small 3×3 convolutions to increase depth and feature extraction
319 capabilities. Inception, also known as GoogLeNet, incorporates multi-scale convolutions within the
320 same layer to capture various spatial patterns and reduce computational cost. These architectures have
321 been effectively utilized in various medical imaging applications, such as tumor detection, organ
322 segmentation, and disease classification, due to their ability to accurately extract and interpret complex
323 features from medical images (38). A comparative analysis of accuracy, sensitivity, and specificity was
324 made to select the best model for the final evaluation strategy.

325 7) Major voting algorithm

326 To make a final decision on the classification of all tested frames per patient, a majority voting
327 algorithm was designed. Different decision thresholds (20-50%) were used to assess the feasibility of
328 the proposed pipeline for VSI-B application. This major voting algorithm determined the classification

329 as malignant if a specific number of frames, exceeding the threshold, were classified as malignant. Prior
330 to this, an analysis of various cases was conducted using only the best-performing classification model
331 (as determined by accuracy, sensitivity, and specificity metrics) to establish the optimal decision
332 threshold, ensuring the reliability and accuracy of the classification method. A major voting algorithm
333 was evaluated by measuring the validation metrics for each threshold and using the ROC-AUC (receiver
334 operating characteristic area under the curve) metric. Binary labels were added to each patient:
335 “Possibly Benign”, “Possibly Malignant”.

336

337 **Results**

338 From the original 170 cases, n= 58 miscellaneous lesions were removed from the current
339 analysis. Additionally, n=4 cases where VSI-B did not identify a mass seen on standard of care, n=2
340 cases of DCIS without a sonographic mass, and n=8 cases of fibroadenomas or malignancy where there
341 was difficulty in segmenting a mass lesion were removed from analysis. After exclusions, there were
342 98 total subjects, n=38 cases no mass and n=60 mass lesions. **Table 1** provides a detailed breakdown
343 of patients with confirmed diagnoses of cancer or non-cancer, including those with cysts and
344 fibroadenomas.

345 Table 1. Characteristics of the sample studied (N=98).

346

Variable	Category	N (%)
Age		40 ± 14*
Sex		
	Female	98 (100%)
Race		
	African American	20 (20.0%)
	Hispanic	5 (5.0%)
	White	73 (75.0%)
BMI		29 ± 7*
Time since apparition		595 ± 1325*
Breast with palpable mass	Left	50 (51.0%)
	Right	48 (49.0%)

Pathology	No mass	38 (39.0%)
	Benign	47 (48.0%)
	Malign	13 (13.0%)
Pain	Yes	44 (45.0%)

*Median (IQR)

347

348

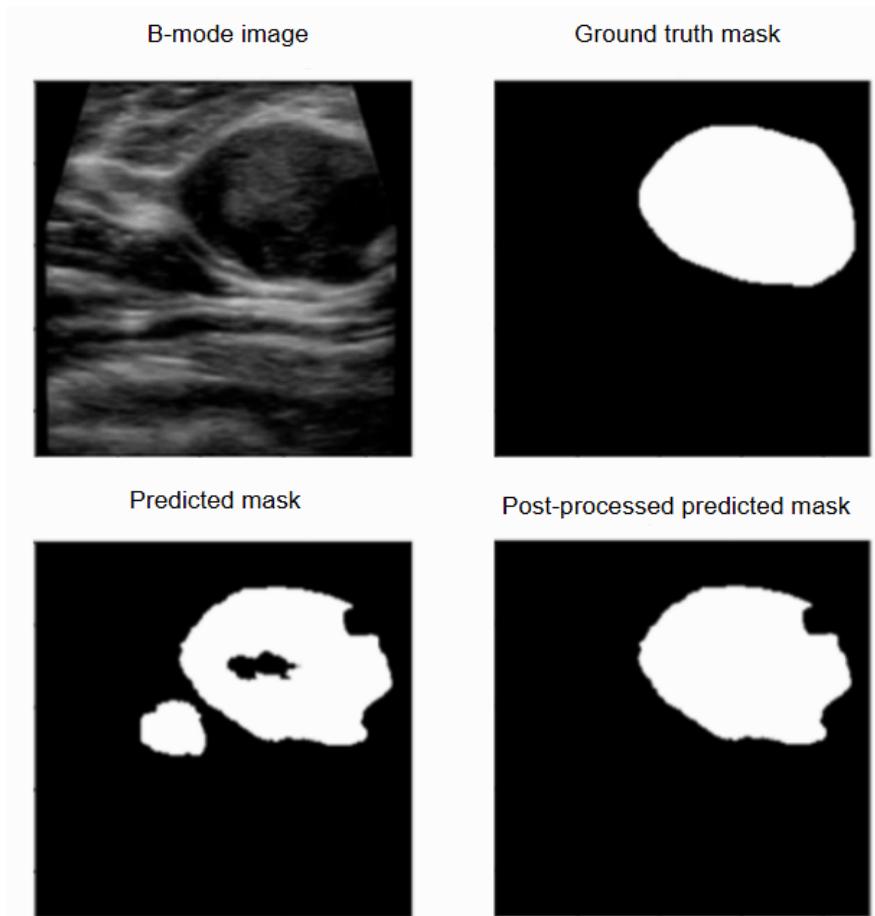
349 *I. Segmentation*

350 Table 2 presents the results for the segmentation model. The Dice score improved from 53.7% (95%
 351 CI: 47.9–61.4%) to 62.9% (95% CI: 55.3–70.5%) after the 1D Gaussian window application with a
 352 threshold of 0.01%. Notably, the variance across folds was 0.71%. A paired t-test confirmed the
 353 statistical significance of this improvement ($p = 0.0216$).

354 Table 2: Dice coefficient comparison for segmentation results using the post-processing technique.

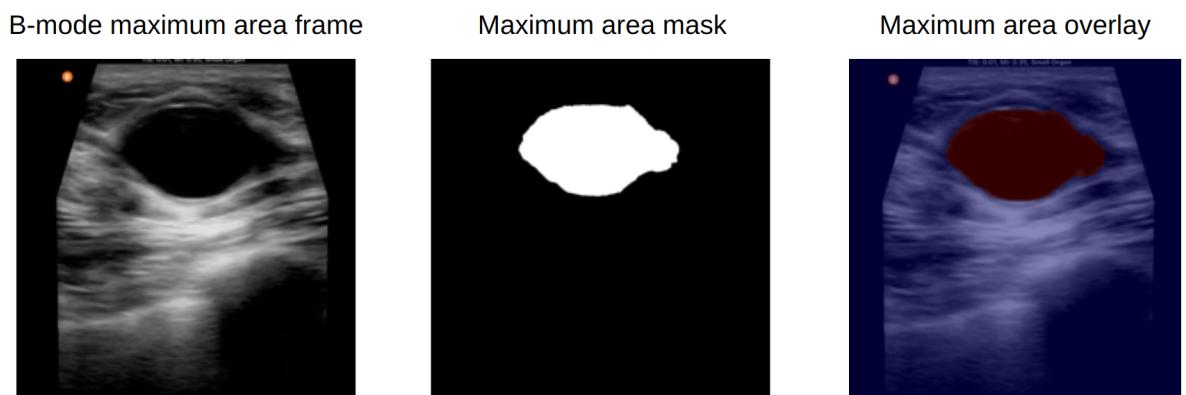
Dice Coefficient	Dice (%) \pm SD
Without post-processing	53.7 \pm 39.1
With post-processing	62.9 \pm 38.5

355



356

357 *Figure 9. Post-processing pipeline. Top-left: Obtained frame. Top-right: Ground truth mask. Bottom-*
 358 *left: Prediction generated by neural network. Bottom-right: Prediction post-processed*
 359



360

361 *Figure 10. Cropped Frame acquired with Butterfly IQ+ with the maximum area segmented. Left:*
 362 *Original US image of a benign patient. Center: Mask predicted by Attention-U-Net 3D. Right:*
 363 *Overlay of the patient's US frame with larger area of the predicted tumor mask.*

364 Before classification, a clear improvement was observed with the post-processing in figure 9.
365 Discrimination algorithm gave not only the possible mass/non-mass patient, but in case of a mass
366 patients the maximum area mask to be analyzed (Figure 10) with 4 more frames per video. The best
367 result was obtained with the threshold of 2 or more unmasked videos to be considered non-mass
368 patients. Similarly, the existence of 2 or more videos with masks to categorize patients with mass
369 improved the evaluation metrics. Assessment in detecting breast lumps at patient level reached 95.0%
370 sensitivity and 63.0% specificity. Considering only mass patients; Cancer and non-cancer detection had
371 100% sensitivity and 93.6% specificity (Table 3)

372

373 Table 3: Model's specificity and sensitivity metrics at patient level for breast detection on mass
374 patients versus non-mass patients, and cancer patients versus non-cancer patients. In the case of mass
375 detection, the specificity of 63% might seem relatively low; however, this does not pose a critical issue
376 within the diagnostic workflow. Since patients identified as "having a mass" proceed to the cancer
377 classifier, all misclassified cases at this stage have been ultimately re-categorized as benign, which is
378 beneficial as it minimizes the risk of missing malignant cases.

379

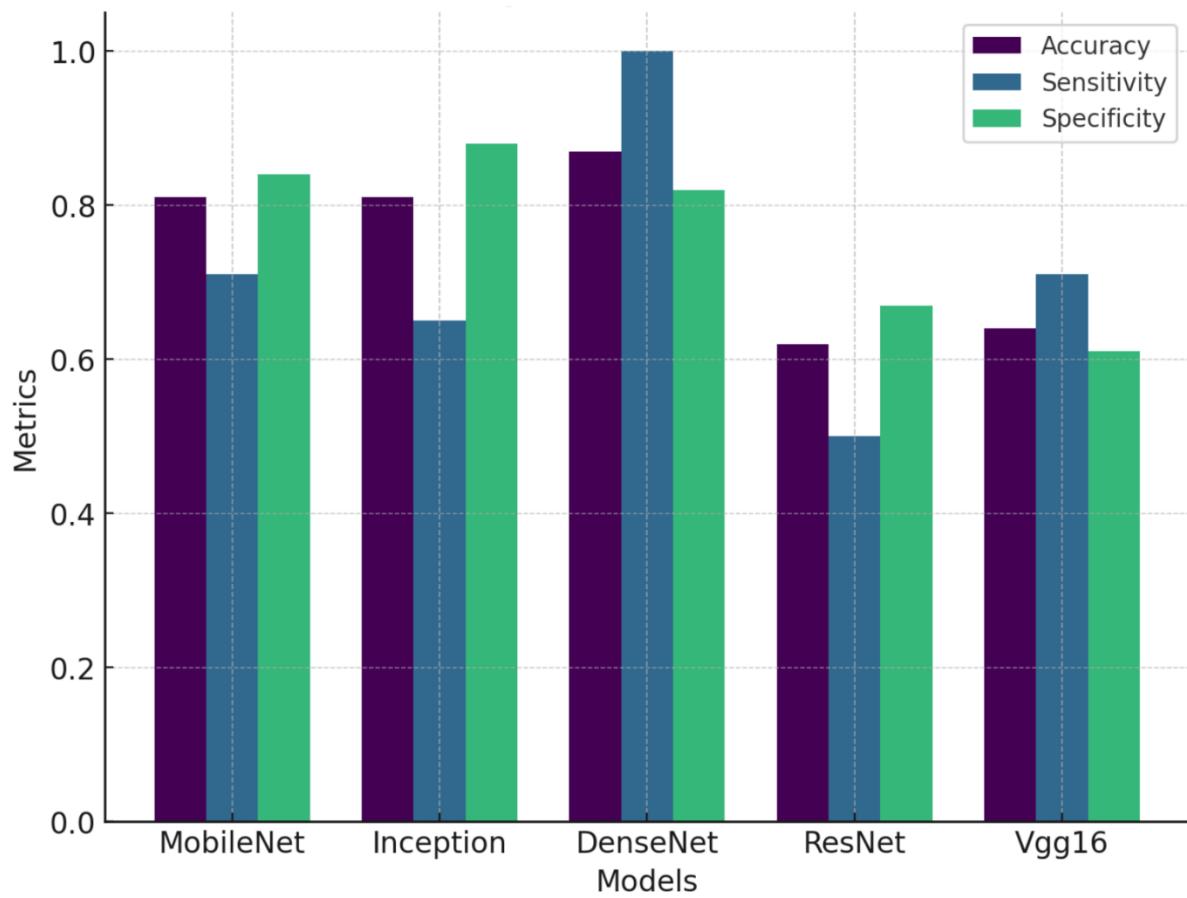
Detection	Specificity (%)	Sensitivity (%)
Mass/ non-mass	63.0	95.0
Cancer/ no cancer	93.6	100

380

381 2. Classification

382 Classification models showed high sensitivity, accuracy, and specificity. Figure 11 shows that
383 among the models selected, DenseNet reached 87.2% accuracy (95% CI: 76.5–93.5%), 100%
384 sensitivity (95% CI: 78.5–100%), and 81.8% specificity (95% CI: 68.4–90.4%). Furthermore, the
385 patient-level variance of correct classification across LOOCV folds was 11.3%. This DenseNet-based
386 model had the highest overall metrics except for specificity, where Inception reached 87.9% specificity
387 (95% CI: 75.4–94.5%). DenseNet, MobileNet, and Inception were the only architectures surpassing
388 80% accuracy and specificity. ResNet had the lowest performance with 61.7% accuracy (95% CI: 49.1–

389 72.9%) and 50.0% sensitivity (95% CI: 26.8–73.2%), while Vgg16 had the lowest specificity, 60.6%
390 (95% CI: 46.2–73.4%).

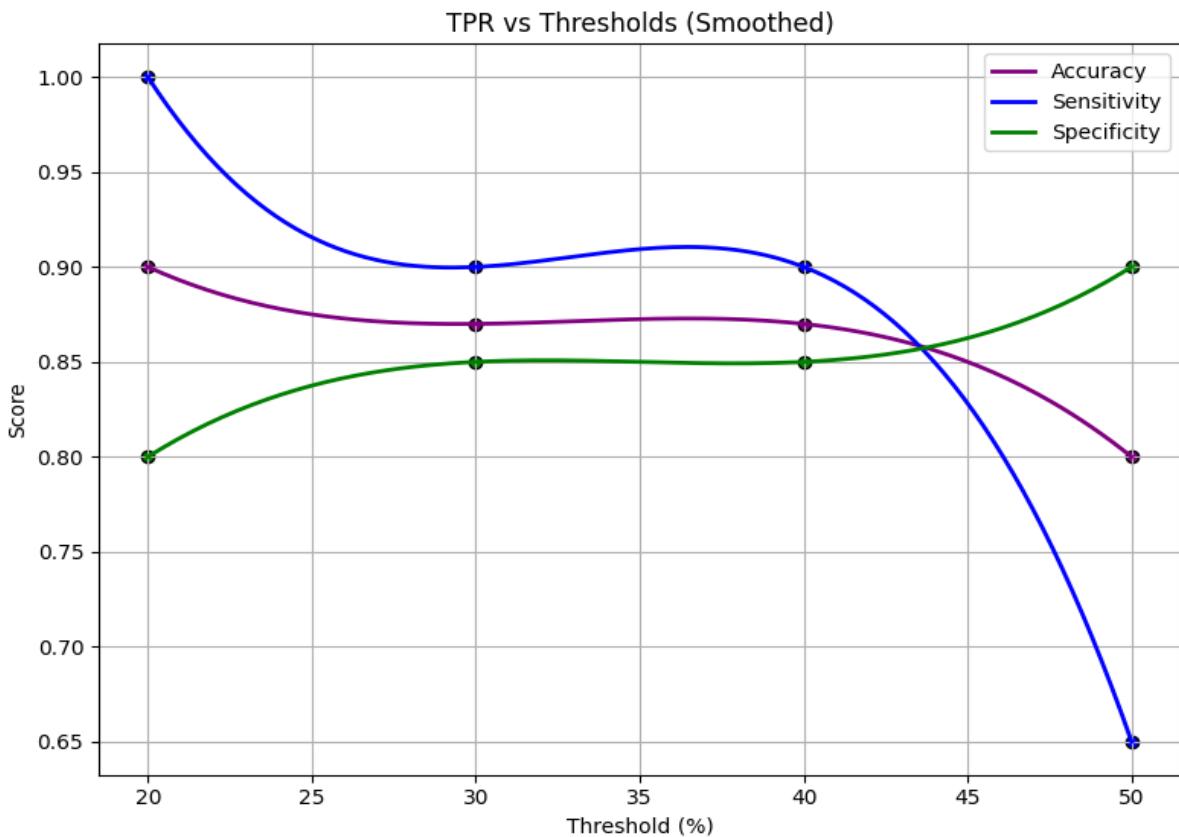


391

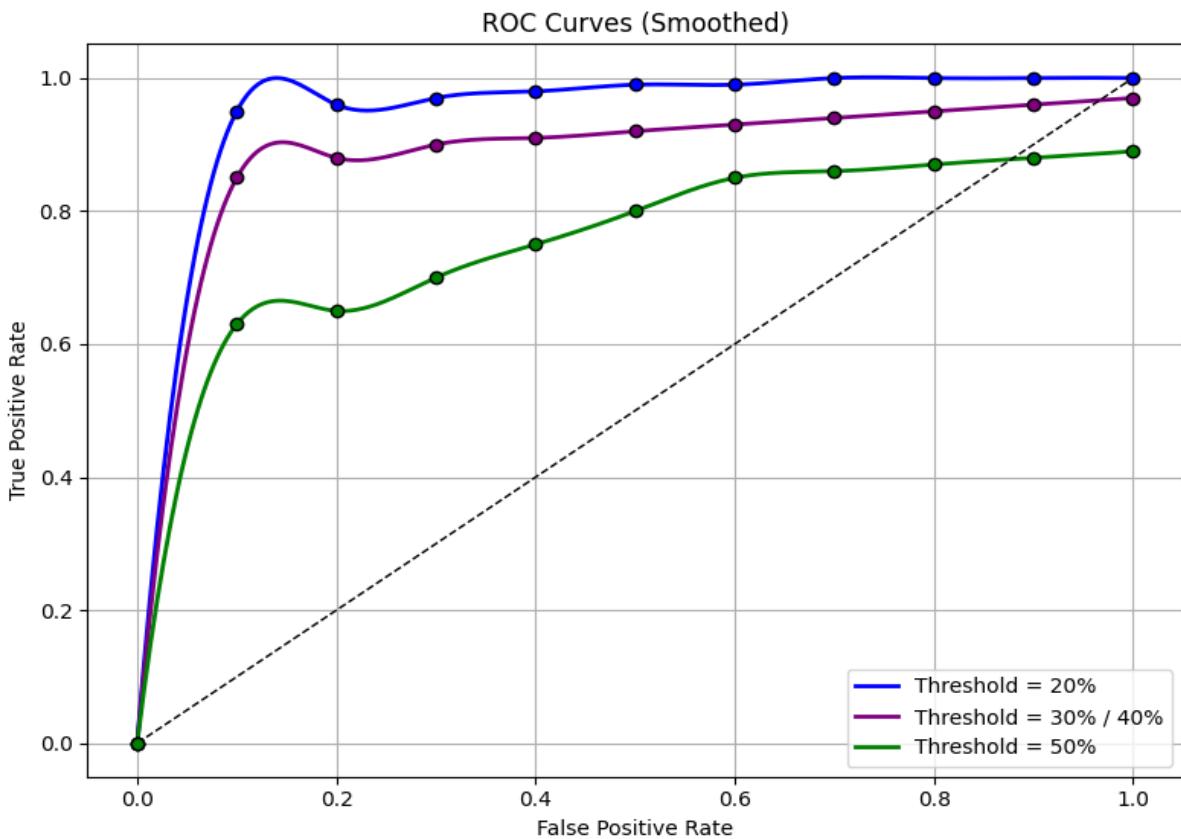
392 *Figure 11. Classification metrics for U-Net segmentation, comparing the 5 architectures, with a*
393 *threshold of 20%.*

394 3. Major voting algorithm

395 Test patients from the LOOCV were evaluated with DenseNet architecture to assess the major
396 voting algorithm for classification across different thresholds, as Figure 12 and Figure 13 depict. The
397 threshold of 20% had the highest metrics. In comparison, the rest of thresholds tested (30%-50%), had
398 lower results. The ROC-AUC of threshold of 20% was 0.91. These results indicated a superior
399 performance of the DenseNet architecture at these thresholds.



400
401 *Figure 12. DenseNet classification metrics across different thresholds. Threshold of 20% had the*
402 *highest scores in accuracy and sensitivity.*
403



404

405 *Figure 13. ROC curves of the DenseNet across different thresholds. The best result is found with the*
 406 *threshold of 20% with an AUC of 0.91.*

407 **Discussion**

408 This work explored the potential of an automated system for breast lump diagnosis using VSI-
 409 B and AI. The integration of this fully automated system can serve as a valuable decision-support tool
 410 by providing clear patient management recommendations based on detected findings. For non-mass
 411 cases, the system suggests no findings of concern. When a benign mass is detected, it could recommend
 412 a follow-up in six months. In cases where a potential cancerous nodule is identified, the model does not
 413 directly diagnose cancer but rather advises referral to a hospital for further evaluation. This structured
 414 recommendation system helps complete the missing care pathway in LMICs by offering a streamlined
 415 approach to patient management, eliminating the need for acquisition and manual image selection by a
 416 specialist. By integrating pathology-based classifications, primary care assessments, and imaging
 417 findings, the model enhances diagnostic workflows, ensuring that patients receive timely and
 418 appropriate care. These findings suggest that combining 3D segmentation with 2D classification on

419 VSI-B images may provide a robust and reproducible approach for breast cancer screening in resource-
420 limited settings.

421 Prior work with the VSI-B protocol, include the use of a WATUNet, has demonstrated
422 promising performance by applying a 2D breast cancer segmentation approach to US frames (28).
423 Results with WATUNet serves as an important benchmark for our research, offering a well-established
424 reference point for VSI-based segmentation accuracy. Our previous study also assessed breast tumor
425 segmentation comparison using VSI-B US protocol previously, where using a 2D multi-input attention
426 U-Net achieved 72.45% of Dice coefficient (27). However, both studies proposed 2D methods, that
427 often overlook the temporal and volumetric information present in clinical US videos. Using videos for
428 the segmentation task provides the incorporation of temporal information from multiple frames. This
429 temporal attention approach leverages the greater volume of data in videos compared to single images,
430 potentially improving segmentation accuracy.

431 The proposed approach was fully automated, from segmentation to classification, streamlining
432 the diagnostic process and reducing the need for manual intervention. In contrast, S-Detect relies on a
433 physician to identify a single “representative” frame. Restricting the analysis to a single frame may omit
434 variations that appear elsewhere in the clip, potentially skewing diagnostic accuracy (14). Although
435 achieving relevant results in segmentation metrics, previous works such as obtained with WATUNet
436 and Ochoa et. al. focused solely on segmenting the videos and did not evaluate classification at the
437 patient level (27, 28). Performance metrics, particularly sensitivity and specificity, are critical in
438 assessing the clinical utility of diagnostic models. Our approach attains a sensitivity of 100% and a
439 specificity of 83%—including 100% specificity for non-mass cases—indicating strong reliability in
440 correctly identifying malignant lesions and minimizing false positives, even in a diverse patient
441 population. S-Detect with VSI-B reported a sensitivity of 100% and a specificity of 86%, metrics that
442 are like those achieved in this study. Achieving similar performance levels to a clinical benchmark is
443 especially significant in healthcare and telemedicine, as it suggests that our fully automated, video-
444 based pipeline could offer a viable alternative for remote or resource-limited settings.

445 This seemingly lower non-mass detection specificity is due to the inherent nature of the VSI-B
446 protocol, which originally consider that all patients have a palpable breast lump, but this not necessarily

447 means a sonographic correlation. Consequently, the initial dataset was predominantly composed of
448 patients with a mass (either benign or malignant). The inclusion of palpable lumps without sonographic
449 correlation (non-mass cases in the sweep) in the training dataset is a novel aspect of this study, allowing
450 for the development of a more comprehensive diagnostic system that accounts for all three possibilities:
451 the presence of a benign mass, a malignant mass, or no mass at all. Increasing the number on non-mass
452 cases would improve the specificity value for this category.

453 In terms of classification with VSI-B protocol, an exploratory 2D classification approach was
454 proposed using S-detect software, but the methodology focused on only a single frame per patient (14).
455 However, this selection of the frame was physician-dependent, leading to a manual process rather than
456 an automated one. While this suits the process for this already trained system, it potentially omitted
457 important temporal and volumetric details contained in the full US clip. Besides this, there was not a
458 validation of the S-Detect system on non-mass cases acquired with VSI-B protocol, leaving a gap in
459 fully capturing the spectrum of breast abnormalities (14). Recognizing this shortfall, we explicitly
460 included non-mass patients, aiming to offer a more comprehensive approach that addresses real-world
461 clinical variability.

462 Generalizability is a key requirement for AI in healthcare, which is why the present study
463 employed a patient-wise cross-validation method based on our previous study(27), which ensured that
464 the model is validated on unseen patient data, thereby reducing the risk of overfitting, and enhancing
465 generalizability. On the other hand, WATUNet used random frames for train-test splitting, a simpler
466 method that might reduce generalizability, and needed manual revision to prevent this issue and
467 overfitting (28). S-Detect system has an already trained model, so this methodology did not apply for
468 that project (14).

469 Successful deployment of the complete pipeline requires careful consideration of deployment
470 logistics, such as training and hardware requirements. The VSI-B acquisition protocol can be mastered
471 in under two hours, consistently yielding high inter operator agreement among non-specialist
472 personnel(8). More complex protocols were successfully tested in low-resource regions such as
473 obstetric and right upper quadrant, showing that is possible to deploy breast VSI training in a limited
474 setting. To enable image capture, storage, transmission and processing in resource limited settings, we

475 used a telemedicine device called the medical box, which features a simplified user interface and relies
476 on CPU mode processing(39). On this platform the end-to-end workflow completes in approximately
477 four minutes per patient, demonstrating that the pipeline can be deployed cost effectively in low
478 resource settings without specialized computational resources.

479 While high-end cart-based ultrasound systems generally deliver superior resolution, advanced
480 imaging modalities (e.g., spectral Doppler, elastography) and greater depth penetration, enhancing
481 AI-driven accuracy, POCUS handheld devices like the Butterfly iQ+ trade some of these capabilities
482 for portability and affordability(40). Field evaluations have also noted that POCUS devices tend to have
483 reduced imaging performance due to processing power requirements and occasional overheating during
484 continuous use (40, 41); without affecting the clinical diagnosis. Despite these limitations, the
485 Butterfly iQ+ remains a viable solution in low-resource settings, and our medical box platform's
486 ultrasound-agnostic design enables seamless adoption with high-end machines in better-resourced
487 environments without modification of the end-to-end workflow.

488 In comparison with Automated Breast Ultrasound Systems (ABUS), our proposed VSI-B
489 approach presents key advantages specifically tailored for low-resource settings. Recent clinical studies
490 have reported sensitivity and specificity for ABUS of approximately 83% and 91%, respectively,
491 underscoring its diagnostic efficacy as a complementary tool for breast cancer detection(42). However,
492 ABUS systems entail high costs (equipment prices can reach approximately \$300,000) and demand
493 specialized infrastructure with stable electricity supply, complicating their deployment in rural or
494 resource-limited areas(43).

495 Regarding future steps our study presents several limitations. First, the dataset was collected in
496 a single center with a specific acquisition protocol and equipment, which may limit the
497 representativeness of broader clinical populations. Second, while the model was validated on an
498 independent test set, its generalizability to untrained operators or different ultrasound devices remains
499 untested, which could impact real-world applicability. Third, the risk of automation errors, especially
500 false negatives in low-quality images or atypical cases, necessitates that the AI system be used as an
501 assistive tool rather than a standalone diagnostic method. Prospective multi-center studies and external
502 validations are needed to address these limitations before clinical deployment. Increasing the volume

503 of patients in the dataset is likely to improve the specificity and reliability of the AI, especially in
504 difficult cases. In this study, we focus on cysts, fibroadenomas, and cancers but other types of mass
505 lesions such as abscess, fat necrosis, skin lesions, lipomas and multiple lesions per patient would also
506 benefit from analysis with AI. At the time of this study, we did not have enough of these other types of
507 lesions to perform meaningful analysis. Thus, it is important to recognize that our current study will
508 have some limitations in terms of generalizability that should be addressed in future work. Also, a larger
509 number of ablation studies are necessary to calibrate and optimize the process of applying a threshold
510 on the Gaussian mask, as it may prevent the detection and segmentation of masses smaller than 0.001%
511 of the image area. Key frame selection and 3D model comparison can be proposed. While redundancies
512 in the data, such as segmenting only 4 of the 8 clips, may contribute to precision by reducing noise and
513 focusing on key frames, this can also lead to longer processing times. It is crucial to assess whether the
514 additional frames are enhancing the model's accuracy or simply increasing computational load without
515 significant gains in precision. Balancing redundancy with time efficiency will be key in refining the
516 VSI-B-based system.

517

518 Conclusion

519 VSI-B provides a potentially low-cost, portable, and non-invasive method for breast imaging,
520 which can be crucial in areas with limited access to medical imaging. The integration of AI with VSI-
521 B could enhance the reach of VSI-B by automating the detection and classification process, thus
522 reducing the dependency on highly trained radiologists. This model, with its high sensitivity and
523 specificity for malignancy, supports early detection and accurate diagnosis of breast cancer, improving
524 patient outcomes through timely and effective treatment. The combination of VSI-B and AI in this study
525 demonstrates a promising approach to addressing the challenges of breast cancer diagnosis.

526

527 Availability of data and materials statements

528 The code used for this study is publicly available on GitHub at
529 <https://github.com/castanedalab/Breast-AI-model>. The database used for the analysis in this study was
530 provided by the University of Rochester and contains proprietary information. Due to the sensitive

531 nature of the data and confidentiality agreements, the raw data cannot be made publicly available.
532 However, anonymized and aggregated data that support the findings of this study is available from the
533 corresponding author on reasonable request. Requests for data access should be directed to the
534 corresponding author and will be reviewed to ensure compliance with confidentiality and intellectual
535 property agreements

536 **Funding Declaration**

537 This work was funded by the Dirección de Fomento de la Investigación from Pontificia
538 Universidad Católica del Perú, through grant CAP 2023-F-0020, project PI1094

539

540 **Author contributions statement**

541 EJO, LCR, SER, and GAG were involved in the AI development, study design, data analysis and
542 manuscript writing. TJM and KJP were involved in the AI development, study design, data acquisition,
543 data analysis, and manuscript writing. YZ, GB, JK, and SM were involved the data acquisition, image
544 segmentation, data analysis, and manuscript writing. AD and AW were involved in the study design,
545 data interpretation, and manuscript production. BC oversaw the entire study and was involved in data
546 analysis, AI development, and manuscript production.

547

548 **Conflict of interest**

549 The authors have no conflict of interest

550 **Ethics and use of human participants statement**

551 This study was approved by the University of Rochester Research Subjects and Review Board
552 (Study 00005262). Informed consent was obtained from every participant. The acquisition
553 process adhered to all relevant guidelines and regulations established by the University of Rochester
554 and was conducted in accordance with the principles outlined in the Declaration of Helsinki. To ensure
555 compliance with HIPAA guidelines, all video clips underwent a thorough anonymization process.

556

557

558 **References**

- 559 1. Francies FZ, Hull R, Khanyile R, Dlamini Z. Breast cancer in low-middle income countries:
560 abnormality in splicing and lack of targeted treatment options. Am J Cancer Res. 2020;10(5):1568-91.
- 561 2. Ginsburg O, Rositch AF, Conteh L, Mutebi M, Paskett ED, Subramanian S. Breast Cancer
562 Disparities Among Women in Low- and Middle-Income Countries. Current Breast Cancer Reports.
563 2018;10(3):179-86.
- 564 3. Torre LA, Islami F, Siegel RL, Ward EM, Jemal A. Global Cancer in Women: Burden and
565 Trends. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for
566 Cancer Research, cosponsored by the American Society of Preventive Oncology. 2017;26(4):444-57.
- 567 4. Duggan C, Dvaladze A, Rositch AF, Ginsburg O, Yip C-H, Horton S, et al. The Breast Health
568 Global Initiative 2018 Global Summit on Improving Breast Healthcare Through Resource-Stratified
569 Phased Implementation: Methods and overview. Cancer. 2020;126(S10):2339-52.
- 570 5. Unger-Saldaña K, Peláez-Ballestas I, Infante-Castañeda C. Development and validation of a
571 questionnaire to assess delay in treatment for breast cancer. BMC Cancer. 2012;12(1):626.
- 572 6. Romanoff A, Constant TH, Johnson KM, Guadiamos MC, Vega AMB, Zunt J, et al.
573 Association of Previous Clinical Breast Examination With Reduced Delays and Earlier-Stage Breast
574 Cancer Diagnosis Among Women in Peru. JAMA Oncol. 2017;3(11):1563-7.
- 575 7. Weiss A, Chavez-MacGregor M, Lichtensztajn DY, Yi M, Tadros A, Hortobagyi GN, et al.
576 Validation study of the American Joint Committee on Cancer eighth edition prognostic stage compared
577 with the anatomic stage in breast cancer. JAMA oncology. 2018;4(2):203-9.
- 578 8. Marini TJ, Castaneda B, Iyer R, Baran TM, Nemer O, Dozier AM, et al. Breast Ultrasound
579 Volume Sweep Imaging: A New Horizon in Expanding Imaging Access for Breast Cancer Detection. J
580 Ultrasound Med. 2023;42(4):817-32.
- 581 9. Erlick M, Marini T, Drennan K, Dozier A, Castaneda B, Baran T, et al. Assessment of a Brief
582 Standardized Obstetric Ultrasound Training Program for Individuals Without Prior Ultrasound
583 Experience. Ultrasound quarterly. 2022.

- 584 10. Toscano M, Marini T, Lennon C, Erlick M, Silva H, Crofton K, et al. Diagnosis of Pregnancy
585 Complications Using Blind Ultrasound Sweeps Performed by Individuals Without Prior Formal
586 Ultrasound Training. *Obstetrics & Gynecology*. 2023;141(5).
- 587 11. Marini TJ, Castaneda B, Baran T, O'Connor TP, Garra B, Tamayo L, et al. Lung Ultrasound
588 Volume Sweep Imaging for Pneumonia Detection in Rural Areas: Piloting Training in Rural Peru.
589 *Journal of Clinical Imaging Science*. 9:35.
- 590 12. Marini TJ, Weis JM, Baran TM, Kan J, Meng S, Yeo A, et al. Lung ultrasound volume sweep
591 imaging for respiratory illness: a new horizon in expanding imaging access. *BMJ open respiratory*
592 *research*. 2021;8(1).
- 593 13. Marini TJ, Oppenheimer DC, Baran TM, Rubens DJ, Toscano M, Drennan K, et al. New
594 Ultrasound Telediagnostic System for Low-Resource Areas. *Journal of Ultrasound in Medicine*.
595 2021;40(3):583-95.
- 596 14. Marini TJ, Castaneda B, Parker K, Baran TM, Romero S, Iyer R, et al. No sonographer, no
597 radiologist: Assessing accuracy of artificial intelligence on breast ultrasound volume sweep imaging
598 scans. *PLOS Digital Health*. 2022;1(11):e0000148.
- 599 15. Arroyo J, Marini TJ, Saavedra AC, Toscano M, Baran TM, Drennan K, et al. No sonographer,
600 no radiologist: New system for automatic prenatal detection of fetal biometry, fetal presentation, and
601 placental location. *PloS one*. 2022;17(2):e0262107.
- 602 16. Marini TJ, Castaneda B, Satheesh M, Zhao YT, Reátegui-Rivera CM, Sifuentes W, et al.
603 Sustainable volume sweep imaging lung teleultrasound in Peru: Public health perspectives from a new
604 frontier in expanding access to imaging. *Frontiers in health services*. 2023;3:1002208.
- 605 17. Toscano M, Marini TJ, Drennan K, Baran TM, Kan J, Garra B, et al. Testing telediagnostic
606 obstetric ultrasound in Peru: a new horizon in expanding access to prenatal ultrasound. *BMC Pregnancy*
607 and *Childbirth*. 2021;21(1):328.
- 608 18. Marini TJ, Oppenheimer DC, Baran TM, Rubens DJ, Dozier A, Garra B, et al. Testing
609 telediagnostic right upper quadrant abdominal ultrasound in Peru: A new horizon in expanding access
610 to imaging in rural and underserved areas. *PloS one*. 2021;16(8):e0255919.

- 611 19. Marini TJ, Weiss SL, Gupta A, Zhao YT, Baran TM, Garra B, et al. Testing telediagnostic
612 thyroid ultrasound in Peru: a new horizon in expanding access to imaging in rural and underserved
613 areas. *Journal of Endocrinological Investigation*. 2021.
- 614 20. Marini TJ, Kaproth-Joslin K, Ambrosini R, Baran TM, Dozier AM, Zhao YT, et al. Volume
615 sweep imaging lung teleultrasound for detection of COVID-19 in Peru: a multicentre pilot study.
616 2022;12(10):e061332.
- 617 21. Zhou LQ, Wu XL, Huang SY, Wu GG, Ye HR, Wei Q, et al. Lymph Node Metastasis Prediction
618 from Primary Breast Cancer US Images Using Deep Learning. *Radiology*. 2020;294(1):19-28.
- 619 22. Browne JL, Pascual M, Perez J, Salazar S, Valero B, Rodriguez I, et al. AI: Can It Make a
620 Difference to the Predictive Value of Ultrasound Breast Biopsy? *Diagnostics (Basel)*. 2023;13(4).
- 621 23. Khaledyan D, Marini TJ, M. Baran T, O'Connell A, Parker K. Enhancing breast ultrasound
622 segmentation through fine-tuning and optimization techniques: Sharp attention UNet. *PloS one*.
623 2023;18(12):e0289195.
- 624 24. Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. *Data
625 Brief*. 2020;28:104863.
- 626 25. Arroyo J, Saavedra A, Tamayo L, Egoavil M, Ramos B, Castaneda B. Automatic fetal
627 presentation diagnosis from ultrasound images for rural zones: head location as an indicator for fetal
628 presentation: SPIE; 2021.
- 629 26. Saavedra AC, Arroyo J, Tamayo L, Egoavil M, Ramos B, Castaneda B, editors. Automatic
630 ultrasound assessment of placenta previa during the third trimester for rural areas. 2020 IEEE
631 International Ultrasonics Symposium (IUS); 2020 7-11 Sept. 2020.
- 632 27. Ochoa EJ, Romero SE, Marini TJ, O'Connell A, Brennan G, Kan J, et al., editors. A comparison
633 between Deep Learning architectures for the assessment of breast tumor segmentation using VSI
634 ultrasound protocol. 2024 46th Annual International Conference of the IEEE Engineering in Medicine
635 and Biology Society (EMBC); 2024: IEEE.
- 636 28. Khaledyan D, Marini TJ, O'Connell A, Meng S, Kan J, Brennan G, et al. WATUNet: a deep
637 neural network for segmentation of volumetric sweep imaging ultrasound. *Machine Learning: Science
638 and Technology*. 2024;5(1):015042.

- 639 29. Schenk A, Praise G, Peitgen H-O, editors. Efficient Semiautomatic Segmentation of
640 3D Objects in Medical Images2000; Berlin, Heidelberg: Springer Berlin Heidelberg.
- 641 30. Raya SP, Udupa JK. Shape-based interpolation of multidimensional objects. IEEE Transactions
642 on Medical Imaging. 1990;9(1):32-42.
- 643 31. Islam M, Vibashan VS, Jose VJM, Wijethilake N, Utkarsh U, Ren H, editors. Brain Tumor
644 Segmentation and Survival Prediction Using 3D Attention UNet2020; Cham: Springer International
645 Publishing.
- 646 32. Falcon W. The PyTorch Lightning Team, 2023. Pytorch lightning.
- 647 33. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition.
648 Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
- 649 34. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, editors. Densely connected convolutional
650 networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.
- 651 35. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al., editors. Going deeper with
652 convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition; 2015.
- 653 36. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets:
654 Efficient convolutional neural networks for mobile vision applications. arXiv preprint
655 arXiv:170404861. 2017.
- 656 37. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image
657 recognition. arXiv preprint arXiv:14091556. 2014.
- 658 38. Tang Y-X, Tang Y-B, Peng Y, Yan K, Bagheri M, Redd BA, et al. Automated abnormality
659 classification of chest radiographs using deep convolutional neural networks. NPJ digital medicine.
660 2020;3(1):70.
- 661 39. Marini TJ, Oppenheimer DC, Baran TM, Rubens DJ, Toscano M, Drennan K, et al. New
662 Ultrasound Telediagnostic System for Low-Resource Areas: Pilot Results From Peru. Journal of
663 Ultrasound in Medicine. 2021;40(3):583-95.
- 664 40. Salimi N, Gonzalez-Fiol A, Yanez ND, Fardelmann KL, Harmon E, Kohari K, et al. Ultrasound
665 image quality comparison between a handheld ultrasound transducer and mid-range ultrasound
666 machine. POCUS journal. 2022;7(1):154.

667 41. Burleson SL, Swanson JF, Shufflebarger EF, Wallace DW, Heimann MA, Crosby JC, et al.
668 Evaluation of a novel handheld point-of-care ultrasound device in an African emergency department.
669 The Ultrasound Journal. 2020;12(1):53.

670 42. Dan Q, Zheng T, Liu L, Sun D, Chen Y. Ultrasound for breast cancer screening in resource-
671 limited settings: current practice and future directions. Cancers. 2023;15(7):2112.

672 43. Company GE. Commercial Reimbursement and Utilization for Invenia ABUS (Automated
673 Breast Ultrasound). General Electric Company; 2019.

674 Appendix A

675 Table A. Number of cases and final classification by type of diagnosis. Non-mass and mass patients
676 were determined through a discrimination algorithm after segmentation. Non-cancer and cancer were
677 discriminated at the binary classification stage.

678

Diagnosis	Non-mass	Non-cancer	Cancer
N	38	47	13
Correctly classified	24	39	13
Incorrectly classified	14*	8	0

*All non-mass patients incorrectly discriminated as mass patients in the segmentation stage were
classified as non-cancer at the binary classification stage

679

680