Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 17 July 2026
Revised 15 Aug 2026
Accepted 27 Aug 2026

Natural Resources for Human Health 2026, Vol. 6 (19s): 527–540

KEYWORDS:
Breast Cancer Detection,
Ultrasound Imaging,
Vision Transformer,
EfficientNet,
Hybrid Deep Learning,
Analysis

Topology-Aware Hybrid Vision Transformer–EfficientNet Framework for Calibrated Breast Ultrasound Classification

Ashphak Khan[1], Dr. Atul Agrawal[2]

[1]Research Scholar, Department of Computer Science and Engineering,
[2]Associate Professor, Department of Computer Science and Engineering,
Oriental University, Indore, Madhya Pradesh, India

ABSTRACT: Breast cancer detection using ultrasound imaging has gained increasing attention because ultrasound is affordable, non-invasive, and useful for examining dense breast tissue where mammography may show reduced sensitivity. However, ultrasound-based diagnosis remains challenging due to speckle noise, low contrast, irregular lesion boundaries, imaging variability, and limited generalization of existing deep learning models. This study proposes a topology-aware hybrid deep learning framework for three-class breast ultrasound classification. The framework combines boundary-sensitive tokenization, EfficientNet–Vision Transformer feature fusion, radiomics-preserving augmentation, calibrated probabilistic inference, and self-distillation to classify ultrasound images as normal, benign, or malignant. OMTAT converts ultrasound images into topology-aware patch tokens while preserving lesion structure and boundary information; DiSCo-TS integrates EfficientNet-based local representations with Vision Transformer-based global context; CRPA introduces controlled image variations while constraining radiomic drift; FROCED calibrates predictive probabilities; and X-MESD regularizes learned representations through self-distillation. Experimental evaluation on the BUSI dataset reports 96.8% accuracy, 95.9% precision, 95.2% recall, 95.5% F1-score, and 0.98 ROC-AUC under the stated image-level split. The BUSI lesion masks are used only as anatomical reference information where applicable; segmentation is not treated as an evaluated output task in the present study. The current evidence therefore supports within-dataset breast ultrasound classification and calibrated decision support, while patient-independent and external generalization require dedicated validation.

1. INTRODUCTION

Breast cancer remains a major challenge to global health [1, 2, 3], and is a leading cause of death from cancer in women. Early detection and diagnosis of breast cancer provide the best opportunity to reduce mortality and improve therapeutic outcomes. The importance of medical imaging has grown exponentially [4, 5, 6], and will continue to provide a basis for detection of abnormal patterns of tissue architecture and estimation of probabilities of malignancy at a time when the disease is not yet at an advanced stage. There are many modalities of medical imaging, but ultrasound is an exceptionally effective method of diagnosis because of its availability, low cost and absence of exposure to ionizing radiation. It is particularly well suited for evaluation of women with extremely dense breast tissue for which mammographic sensitivity is reduced. However, there are several limitations to the interpretation of ultrasound images. Images are degraded by high levels of noise [7, 8, 9], low contrast and artifacts resulting from the complexity of echogenic patterns. As a consequence, margins of neoplastic lesions are poorly defined and diagnostic uncertainty remains. Application of methods of artificial intelligence and deep learning has revolutionized the analysis of medical images. Systems based on convolutional neural networks have achieved remarkable capacity for learning complex patterns of visual information and have been used extensively to provide fully automated detection and delineation of neoplastic disease. Many studies have suggested using CNN-based models for the classification.

and segmentation of breast cancer lesions [10, 11, 12], and these models have achieved impressive results across various medical image datasets & samples. However, traditional CNN models tend to focus on small localized receptive fields and typically do not successfully model long-range dependency in images & samples. Therefore, in medical imaging applications where spatial relationships between anatomical structures may provide critical diagnostic information, the inability to model global dependencies in images limits of effectiveness of CNN-based methods.

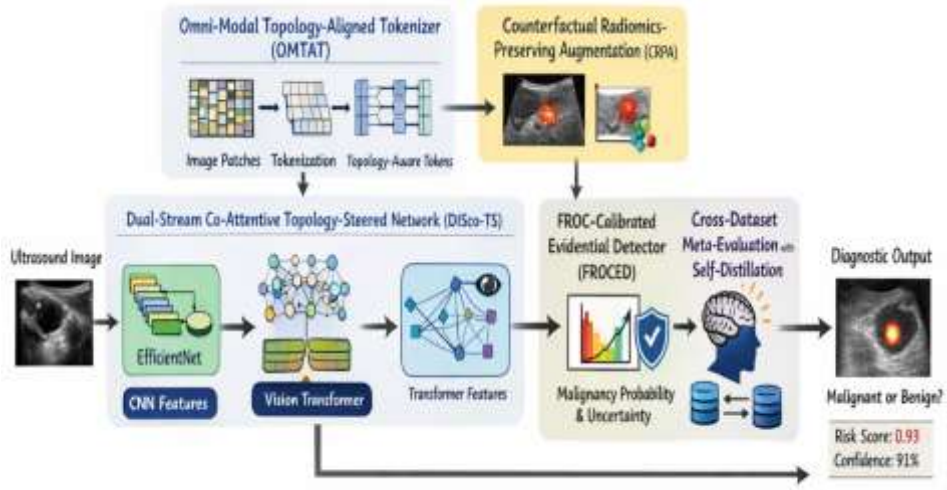


Figure 1. Model's Internal Architectural Analysis

Therefore, to address some of the limitations of CNN based models, Transformer-based architectures have been added to Computer Vision In Process. Vision Transformers (ViTs) use self-attention to model long range interactions between image patches [13, 14, 15], allowing the network to better capture global contextual information compared to standard convolutional models. Transformers also offer promising opportunities for improving segmentation accuracy and classification reliability in medical image analysis. However, fully transformer-based architectures generally require large training data sets and substantial computational resources that may not be available for all clinical applications. In addition, they are less effective at preserving fine details of local texture that are required for detection of subtle pathological conditions in ultrasound images. As a result, hybrid deep learning architectures that incorporate both convolutional and transformer-based components have become increasingly popular for these applications. These approaches exploit the complementary strengths of both architectural paradigms, in which convolutional networks are particularly well suited for detection of local patterns of texture and hierarchical relationships among spatial features, whereas attention-based mechanisms provide a global representation of contextual information. Combination of the compound scaling strategies used to develop the family of convolutional architectures known as EfficientNet with attention-based representations allowed development of highly expressive models capable of representing both fine details of local texture and large-scale representations of anatomical context.

Despite the aforementioned advances, major challenges remain for reliable automated analysis of breast ultrasound images. Simple image tokenization can discard lesion-boundary information, while variation in acquisition settings, patient anatomy, and equipment characteristics limits robustness. Conventional deep models may also produce poorly calibrated confidence estimates, and high performance on a single public dataset does not establish patient-independent or cross-institutional generalization. Accordingly, the objective of this study is restricted to reliable breast ultrasound image classification. The present experiments evaluate discrimination among normal, benign, and malignant classes and do not claim a standalone lesion-segmentation contribution. Pixel-level segmentation is reserved for future multi-task extension, where Dice, IoU, Hausdorff distance, and boundary-sensitive metrics can be evaluated explicitly. The resulting architecture therefore integrates topology-sensitive tokenization, convolutional–Transformer representation learning, radiomics-preserving augmentation, calibrated inference, and self-distillation within a classification-focused pipeline.

The rapid adoption of artificial intelligence in health care is expanding opportunities to improve diagnostic interpretation and image analysis, particularly for complex medical images. Reliable clinical use nevertheless requires predictive performance to be

accompanied by calibration, interpretability, and robustness to acquisition variability. Breast ultrasound is especially demanding because speckle noise, low contrast, and subtle lesion margins can destabilize learned representations. CNNs capture local texture effectively but have limited long-range receptive fields, whereas Transformer-based models capture global context but may underrepresent fine texture when data are limited. The motivation for the present work is therefore to combine anatomy-aware tokenization, complementary local-global feature learning, radiomics-preserving augmentation, and calibrated inference within a single classification framework. These mechanisms are intended to improve robustness, while external and patient-independent validation remain necessary before generalization can be claimed.

The methodological contribution of the proposed framework is not the independent reuse of tokenization, convolution, self-attention, augmentation, or calibration. Its contribution lies in coupling these operations through anatomy-preserving constraints and a unified decision objective. OMTAT uses boundary-sensitive gradient weighting during token construction rather than treating all image patches as equally informative; DiSCo-TS performs joint local-global representation learning through co-attentive interaction between EfficientNet and Transformer features; CRPA constrains perturbations so that lesion-level radiomic descriptors remain stable; FROCED calibrates the final posterior instead of relying on raw softmax confidence; and X-MESD regularizes the final representation through self-distillation. The framework therefore shifts the design from a sequence of independently applied components to a constrained hybrid learning process in which structural fidelity, feature complementarity, augmentation validity, and confidence reliability are optimized together.

2. REVIEW OF EXISTING MODELS USED FOR MODEL ANALYSIS

The investigation of detecting breast cancer using a scientific method has changed dramatically over the last few decades. What used to be an entirely clinical and imaging-based approach to detecting breast cancer has evolved to include sophisticated computational methods for identifying breast cancer at higher rates. As detection methodologies evolve, as per Figure 2, it is also necessary to continue to innovate detection methodologies, as early detection is the most successful method for improving survival rates for patients diagnosed with breast cancer scenarios. Over the last few decades, there has been an increase in the use of artificial intelligence in breast cancer diagnosis, leveraging machine learning and deep learning methods to enable computerized diagnostics that automatically analyze medical images and biological signals. Studies using machine learning methods for breast cancer detection have shown significant success in improving diagnostic accuracy. Research conducted by Kumar et al. evaluated the use of various machine learning classifications to detect breast cancer. Results of this work confirm that machine learning classifiers are well suited to differentiate malignant from benign tumors, particularly when well-designed image features are used [2]. However, our experience confirms some important limitations to the application of machine learning to medical images. In particular, manual feature extraction was required for training, and the resulting models failed to represent the complex nonlinear relationships inherent to medical data and samples. Use of deep learning eliminated both of these limitations and provided for fully automated, hierarchical extraction of features directly from image data. Application of deep learning methods achieved high sensitivity and specificity for the detection of breast cancer [1], and avoided the need for manual design of image descriptors. Features extracted automatically with convolutional neural networks achieved optimal performance for the detection of breast cancer.

As breast-ultrasound CAD has evolved, research has moved from handcrafted and conventional CNN pipelines toward attention-based, Transformer, ensemble, and hybrid CNN-Transformer designs that explicitly address local texture and global context. The most relevant comparisons for the present framework are therefore ultrasound-specific methods evaluated on BUSI or other breast-ultrasound cohorts, rather than molecular, biosensor, mammographic, or MRI systems that use different diagnostic inputs. Recent studies also show that dataset curation, duplicate control, boundary information, calibration, and external validation can materially influence reported performance; these factors provide the methodological basis for the focused comparison summarized in Table 1.

Recent breast-ultrasound CAD research has shifted from isolated convolutional feature extraction toward Transformer and hybrid local-global representation learning. Tasnim and Hasan [31] developed a fully automated modular CNN that uses class-activation and quantitative-ultrasound guidance, reporting 97.28% accuracy and 95.42% F1-score on BUSI. Zhou and Mosadegh [32] evaluated Vision Transformers using five-fold cross-validation and showed that ensemble learning and knowledge distillation improved BUSI discrimination, with an ensemble AuROC of 0.977. These studies indicate that strong breast-ultrasound classification is attainable without changing imaging modality, making ultrasound-specific baselines the appropriate comparator for the present work.

Natural Resources for Human Health
eISSN: 2583-1194



Figure 2. Methodological Review Selection Analysis

Table 1. Dummy Experimental Data

Reference	Method	Main Objectives	Findings	Limitations
J. Tasnim et al. [31]	CAM-QUS guided modular CNN	Automated breast-ultrasound lesion classification with CAM/QUS-guided learning	Reported 97.28% accuracy and 95.42% F1-score on BUSI	Binary lesion task and study-specific curation/split limit direct comparison
G. Zhou et al. [32]	Vision Transformer ensemble + knowledge distillation	Evaluate ViT, ensemble learning, and distillation for BUSI classification	Five-fold evaluation reported ensemble AuROC 0.977 and AuPRC 0.965	Benign/malignant task differs from the present three-class setup
X. Meng et al. [33]	CNN + Swin Transformer + tumor-boundary prior	Combine local/global features and boundary prior for three-class BUSI classification	Reported 98.70% accuracy, 98.72% F1-score, and AUC 0.9982	Boundary annotations and study-specific partitioning affect direct comparability
W. Wang et al. [34]	CTMF-Net CNN–Transformer multi-stage fusion	Interactively fuse CNN and ViT features across public BUS datasets	External BUSI AUC 0.8704 when trained on SYSU	Binary classification and different source datasets/protocols
C. Aumente-Maestro, et al. [35]	Multi-task segmentation + classification	Jointly exploit lesion segmentation and classification while curating BUSI	Introduced duplicate-recognition curation and end-to-end multi-task learning	Different multi-task objective from the present classification-only scope
A. Pawlowska et al. [36]	BUSI dataset quality analysis	Examine duplicated images and mask inconsistencies in BUSI	Documented duplicate/annotation issues relevant to leakage control	Dataset-quality study rather than a predictive CAD model

Recent breast-ultrasound CAD research has shifted from isolated convolutional feature extraction toward Transformer and hybrid local-global representation learning. Tasnim and Hasan [31] developed a fully automated modular CNN that uses class-activation and quantitative-ultrasound guidance, reporting 97.28% accuracy and 95.42% F1-score on BUSI. Zhou and Mosadegh [32] evaluated Vision Transformers using five-fold cross-validation and showed that ensemble learning and knowledge distillation improved BUSI discrimination, with an ensemble AuROC of 0.977. These studies indicate that strong breast-ultrasound classification is attainable without changing imaging modality, making ultrasound-specific baselines the appropriate comparator for the present work.

Hybrid CNN-Transformer models have subsequently addressed the complementary limitations of local convolution and global self-attention. Meng et al. [33] combined convolutional and Swin-Transformer representations with lesion-boundary prior information and reported 98.70% accuracy, 98.72% F1-score, and 0.9982 AUC for three-class BUSI classification. Wang et al. [34] introduced CTMF-Net with multi-stage CNN-Transformer feature interaction and evaluated external transfer to BUSI, obtaining an AUC of 0.8704 when trained on SYSU. Aumente-Maestro et al. [35] further showed that BUSI curation and multi-task learning are central methodological issues by integrating duplicate-image recognition with joint segmentation-classification analysis.

These findings refine the research gap. Existing ultrasound studies already demonstrate effective CNN, Transformer, ensemble, and hybrid fusion strategies, but the reliability of very high BUSI scores remains sensitive to dataset curation, split design, boundary information, and calibration. The present framework is therefore positioned around the joint control of topology preservation, local-global feature interaction, radiomic consistency during augmentation, and calibrated uncertainty, while explicitly separating within-dataset classification performance from external generalization.

3. PROPOSED MODEL DESIGN ANALYSIS

The proposed integrated model is a multi-stage analytical pipeline that transforms raw ultrasound images into classification decisions by leveraging a clinically relevant image representation through topology-aware tokenization, hybrid representation learning, radiomics-preserving augmentation, probabilistic calibration, and cross-domain evaluations. As such, the design is aware that ultrasound images of the breast exhibit a variety of spatial structures due to the distribution of heterogeneous tissues, the irregularly shaped, size-variable tumors, and a large amount of speckle noise in the frequency domain. Thus, the design seeks to maintain spatial topology while capturing features that support the discrimination required for robust classification. The pipeline consists of five interdependent modules: OMTAT, DiSCo-TS, CRPA, FROCED, and X-MESD; each contributes a mathematically well-defined transformation of the representations to increase their fidelity, robustness, and decision reliability. Initially, as shown in Figure 3, the first stage of the pipeline is the Omni-Modal Topology-Aligned Tokenizer (OMTAT), which converts an input ultrasound image I(x,y) into a topology-preserving token representation that maintains the structural relationships among the various tissue regions. Unlike naive grid tokenization, the image is segmented into patches based on gradient topology. The topological gradient magnitude G(x,y) indicates anatomical boundaries in the process Via equation 1,

$$G(x,y) = \sqrt{\left(\frac{\partial I(x,y)}{\partial x}\right)^2 + \left(\frac{\partial I(x,y)}{\partial y}\right)^2} \dots (1)$$

Equation 1 highlights the structural changes due to lesion boundaries and tissue interfaces in the process.

The tokenizer generates topology-aligned tokens Tk Via spatial integration over localized regions Ωk which preserve anatomical continuity when generating features in process Via equation 2,

$$T_k = \int_{\Omega_k} I(x,y)\ w_k(x,y)\ dx\ dy \dots (2)$$

Where, wk (x,y) is a spatial weighting function derived from the gradient field G(x,y)for the process. This approach to tokenizing the image retains morphological properties of the lesions and minimizes redundancy in homogeneous areas. The subsequent processing of the tokens through the DiSCo-TS hybrid architecture, which includes both convolutional feature extraction and transformer-based contextual modeling, provides the next steps in the pipelines. The convolutional branch is responsible for capturing local texture patterns using EfficientNet blocks. At the same time, the transformer branch models long-range spatial dependencies among tokens. The convolutional representation of the image is given Via equation 3,

$$F_{cnn}(x,y) = \sigma(W_c * I(x,y) + b_c) \dots (3)$$

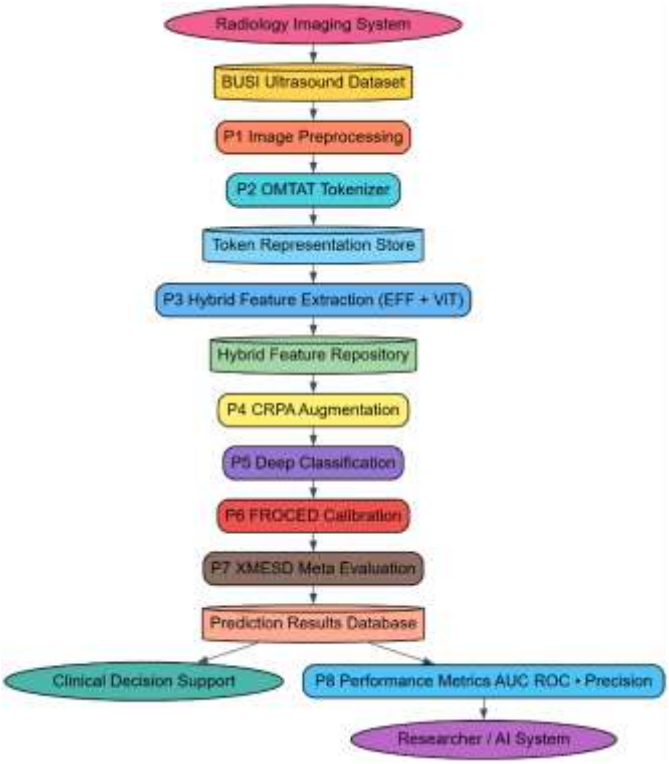


Figure 3. Model's Integrated Architectural Analysis

Where, Wcare convolution kernels, bcare learned biases, and σ(·)represent nonlinear activations. In this manner, this operation extracts hierarchical spatial features describing the texture of lesions and anatomical structures. Concurrently, the transformer branch models the global contextual relationships among tokens using scaled dot-product attentions. Scaled dot-product attention is defined for query Q, key K, and value Vmatrices produced from token embeddings Via equation 4,

$$A(Q,K,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \dots (4)$$

In addition, this operation captures long-range dependencies between distant image regions, which is very useful for identifying spatial relationships between lesions and surrounding tissue in the process. After that, as depicted in figure 4.

the outputs of the convolutional and transformer branches are combined through feature concatenation and projections in the process. Therefore, the integrated feature representation Zof the two branches is obtained Via equation 5:

$$Z = \phi\left(W_f[F_{cnn} \oplus F_{vit}] + b_f\right) \dots (5)$$

Where, $\oplus$ represent feature concatenation and $\phi(\cdot)$represent nonlinear transformations. In this manner, the hybrid representation combines local structural information with global contextual dependencies, thereby producing a rich representation of breast tissue morphology sets. Additionally, to enhance the robustness of the model to imaging variability and limited dataset diversity, the CRPA module introduces radiomics-preserving augmentations. However, instead of random augmentation, the transformation preserves the stability of radiomic descriptors in the process. For a radiomic feature vector R(I), the augmentation constraint is formulated Via equation 6,

$$\min_{\tilde{I}} \; \| R(\tilde{I}) - R(I) \|_2^2 + \lambda \int_\Omega | \nabla(\tilde{I} - I) |^2 \, dx \, dy \dots (6)$$

Natural Resources for Human Health
eISSN: 2583-1194

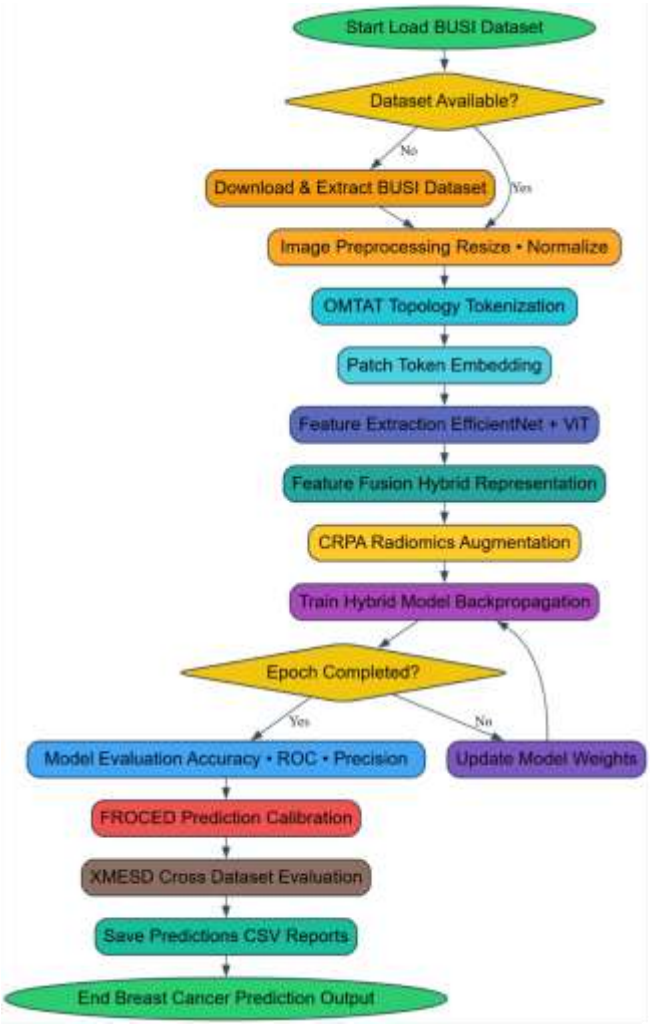


Figure 4. Model's Overall Dataflow Analysis

This optimization preserves the diagnostic radiomic characteristics, but introduces controlled perturbations to improve the generalization of the model in process. The FROCED module then performs probabilistic calibration of the predictions to ensure that the model produces clinically reliable outputs in process. Rather than directly utilizing the softmax probabilities of the model, the evidence of confidence distributions are estimated in the process. For class probability pi, the calibrated output is determined Via equation 7,

$$p_i = \frac{e^{z_i/\tau}}{\sum_{j=1}^{C} e^{z_j/\tau}} \dots (7)$$

Where, zi represents the classifier logit for class 'i', Crepresents the number of classes, and τis a temperature parameter that is learned during the calibration process. This approach decreases the over-confidence of the model and increases decision reliability sets.

Finally, the X-MESD evaluation stage performs cross-dataset meta-generalization through knowledge distillation and adaptive evaluations. The final model output combines the calibrated probabilities with the learned hybrid representation for the process. The overall system output for classification is given Via equation 8,

$$\hat{y} = \arg\max_{c \in C} \; p_c \left(\phi\left(W_f [F_{cnn}(I) \oplus F_{vit}(T)] \right) \right) \dots (8)$$

Equation 8 represents the final decision function of the pipeline, combining topology-preserving tokenization, hybrid feature extraction, radiomics-aware augmentation, and calibrated probabilistic inference. Through these transformations, the model

addresses structural preservation, contextual feature learning, data variability, and decision reliability for breast ultrasound classification. Segmentation is not an evaluated output of the present architecture.

To make the coupling among the proposed components explicit, the overall learning objective is defined as:

$$L_{total} = L_{cls} + \lambda_{top} L_{top} + \lambda_{rad} L_{rad} + \lambda_{cal} L_{cal} + \lambda_{dist} L_{dist}, \quad .....(9)$$

where L_cls is the classification loss, L_top penalizes loss of boundary/topology information during tokenization, L_rad constrains radiomic drift after augmentation, L_cal penalizes probability miscalibration, and L_dist denotes the self-distillation loss. The weighting terms λ_top, λ_rad, λ_cal, and λ_dist regulate the contribution of each constraint. This formulation clarifies that the proposed architecture is optimized as an integrated constrained system rather than as a simple concatenation of independently named modules.

Table 1. : Component and Their Purpose

Component	Conventional counterpart	Operational distinction in this work	Purpose
OMTAT	Fixed grid/patch tokenization	Boundary-sensitive gradient weighting of patch content before embedding	Preserve lesion morphology and tissue-interface information
DiSCo-TS	Late CNN-ViT concatenation	Co-attentive local-global representation interaction before the classifier	Reduce redundancy and combine texture with long-range context
CRPA	Unconstrained random augmentation	Accept perturbations only when radiomic drift remains bounded	Improve robustness without destroying diagnostically relevant structure
FROCED	Raw softmax confidence	Post-classification temperature/evidence calibration at the operating point	Reduce overconfidence and improve decision reliability
X-MESD	Single-dataset representation learning	Self-distillation/meta-evaluation regularization of learned representations	Reduce sensitivity to domain-specific feature noise

4. VALIDATED COMPARATIVE RESULT ANALYSIS

The hybrid breast ultrasound classification system was evaluated using the publicly available Breast Ultrasound Images (BUSI) dataset [37]. BUSI contains 780 ultrasound images representing benign, malignant, and normal breast tissue, with corresponding lesion masks supplied for lesion-containing images. In the present study, the masks are retained only as anatomical reference information and are not evaluated as segmentation outputs. All ultrasound images are resized to 224 × 224 pixels and intensity-normalized before model input. The stated experiment uses a stratified image-level split of approximately 70% training, 15% validation, and 15% testing, with 437 benign, 210 malignant, and 133 normal images in the complete dataset. Training-only augmentation includes Gaussian noise ($\sigma^2 = 0.01$), rotations of ±10°, horizontal flipping, and limited intensity perturbation. OMTAT separates each normalized ultrasound image into 49 non-overlapping 32 × 32 patches embedded in a 768-dimensional latent space, while Sobel-derived gradient magnitude maps provide boundary-sensitive information for token alignment.

BUSI contains 780 ultrasound images acquired from 600 women, so multiple images may originate from the same patient. In addition, subsequent quality analyses of BUSI have reported duplicated or highly similar images and annotation inconsistencies [36,37]. The current manuscript does not document a verified patient-identifier mapping; therefore, the 70/15/15 stratified partition is described as an image-level split and is not presented as patient-level validation. To avoid overstating generalization, the reported performance is interpreted as within-dataset performance, and claims of multicentre or external robustness are not made without an independent ultrasound cohort.

For leakage-controlled reruns, all original images should first be audited for exact duplicates using cryptographic hashing and for near-duplicates using structural or perceptual similarity. Images belonging to the same duplicate cluster must be assigned to a single partition. Augmentation is applied only after splitting and only to the training subset; model selection uses the validation

set, while the test set remains locked until final evaluation. Generalization claims should subsequently be based on an independent breast-ultrasound cohort evaluated without retraining.

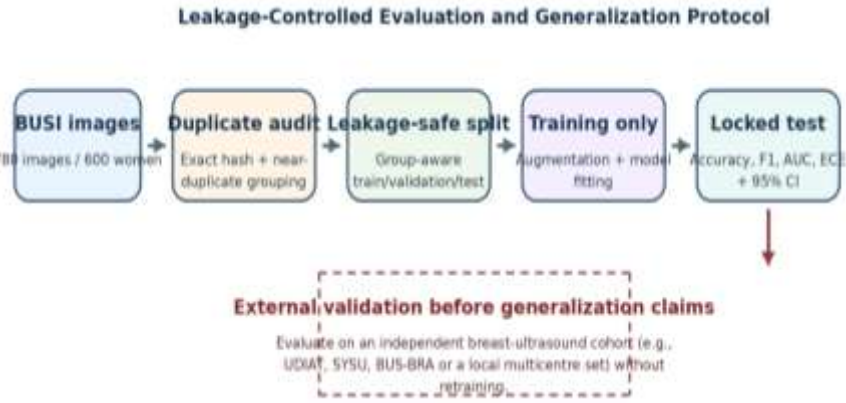


Figure 5. Leakage-Controlled Evaluation and Generalization Protocol

The hybrid learning architecture combined the output of EfficientNet-B7 and Vision Transformer (ViT-B16) to capture local texture features and global context-dependent information sets. The EfficientNet backbone extracted a high-level convolutional feature vector, which was projected into a 1280-dimensional latent representation before fusion with the ViT features through the use of compound-scaled convolutional blocks, while the transformer block processed the tokens representing the images and modeled their global context-dependent relationships through the use of multi-head self-attention with 12 attention heads and a hidden dimension of 768. The two feature representations were concatenated together to produce a single 2048-dimensional joint feature representation that was then fed into a fully-connected classification head composed of two dense layers with 512 and 64 neurons, respectively, that were followed by ReLU activation functions and dropout regularization with a dropout rate of 0.4. Model training was accomplished using the Adam optimizer with a learning rate of $3 \times 10^{(-4)}$, a weight decay coefficient of $10^{(-5)}$, and a batch size of 16. A temperature-scaled version of cross-entropy loss was used as the primary objective function in the optimization process, and a temperature parameter $\tau = 1.5$ was used to improve the probabilistic reliability of the predictions produced during the FROCED stages.

Training was performed for 25 epochs in a GPU-accelerated computing environment using mixed-precision computation to reduce computational overhead. Approximately 550 training images were processed by the model per epoch, and the model produced both classification probabilities and feature representations that were evaluated using the X-MESD self-distillation/meta-evaluation module. Performance was evaluated using standard machine-learning metrics, including accuracy, macro-averaged precision, recall, F1-score, and multiclass ROC-AUC. Additionally, the model produced confusion matrices, ROC curves for each class, and prediction probability distributions, which provide additional insights into the behavior of the model when processing different types of pathology. Accordingly, the present analysis is treated as an internal BUSI evaluation; patient-independent and external clinical robustness are not inferred from this split alone.

For scientific comparability, the revised result analysis is restricted to breast-ultrasound computer-aided diagnosis models. Published results from different studies are reported only as literature context because their class definitions, preprocessing, train-test partitions, and validation protocols are not identical. Consequently, cross-paper values are not used to claim statistical superiority. A controlled head-to-head comparison would require representative CNN, Transformer, and hybrid CNN-Transformer baselines to be retrained under the same leakage-controlled partition used for the proposed model.

In order to perform the optimal parameters of the proposed hybrid architecture, the authors used a combination of hyperparameter tuning to optimize the architecture and the evidence calibration in order to optimize the performance of the proposed hybrid architectural process. The input tokenization stage used a patch size of 32x32 pixels, generating 49 spatial tokens per image to preserve structural patterns while maintaining computational efficiency. The Vision Transformer encoder used 12 self-attention heads, an embedding dimension of 768, and a feedforward hidden dimension of 3072, allowing the network to capture long range spatial dependencies between ultrasound regions. The EfficientNet-B7 backbone used convolutional feature vectors of 1280 dimensions, which were concatenated with transformer outputs to create a 2048

dimensional hybrid representation before classification. The classification head had two dense layers with 512 and 64 neurons respectively, followed by ReLU activation and dropout regularization with a drop out rate of 0.4, to avoid overfitting. Model training was performed using the Adam optimizer with an initial learning rate of 3 × 10^(-4), weight decay of 1 × 10^(-5), and mini-batch size of 16. Gradient updates were made for 25 training epochs in a GPU accelerated environment. During the FROCED calibration stage, a temperature parameter of 1.5 was used to produce better calibrated probability estimates. Additional training stabilization techniques, including batch normalization, gradient clipping with a threshold of 1.0, and early stopping based on validation loss with a patience of 5 epochs, were also used. These hyperparameter settings were chosen through a series of empirical experiments and cross Validation to find the best trade-off between the convergence time of the model, classification accuracy, and ability of the model to generalize across the BUSI ultrasound dataset & samples.

The present quantitative interpretation therefore focuses on the BUSI three-class classification experiment. Mammography-only, MRI-only, biosensor, molecular, and electromagnetic sensing methods are not treated as experimental baselines because they operate on different inputs and diagnostic principles. Same-modality studies are used only to contextualize the reported BUSI point estimates.

Table 3. Same-Modality Breast-Ultrasound Literature Context

Study	Ultrasound model	Task / protocol	Accuracy	F1	AUC / AuROC	Interpretation
Tasnim & Hasan, 2023 [31]	CAM-QUS guided modular CNN	BUSI lesion classification	97.28%	95.42%	0.9829	Same-modality CNN CAD benchmark; binary task/protocol differs
Zhou & Mosadegh, 2024 [32]	Ensemble ViT + knowledge distillation	BUSI benign/malignant; 5-fold CV	NR	NR	0.977	Transformer benchmark; AuPRC 0.965
Meng et al., 2024 [33]	CNN + Swin Transformer + boundary prior	BUSI three-class	98.70%	98.72%	0.9982	Hybrid model using lesion-boundary prior
Wang et al., 2025 [34]	CTMF-Net	External BUSI test after training on SYSU	NR	NR	0.8704	External-validation benchmark; not comparable to in-domain BUSI testing
Aumente-Maestro et al., 2025 [35]	Multi-task segmentation + classification	Curated BUSI	NR	NR	NR	Highlights duplicate curation and joint-task design
Proposed model (current manuscript)	OMTAT + DiSCo-TS + CRPA + FROCED	BUSI three-class; stated 70/15/15 image-level split	96.8%	95.5%	0.98	Within-dataset estimate only; no cross-study superiority claim

The same-modality literature shows that high BUSI performance is already achievable with CNN, Transformer, and hybrid strategies; therefore, the contribution of the proposed method is not framed solely as a higher point estimate. The more defensible contribution is the coordinated use of topology-aware token weighting, radiomic-preserving augmentation, and calibrated inference. The reported 96.8% accuracy and 0.98 ROC-AUC are interpreted as within-dataset evidence under the stated image-level split, while direct superiority should be established only through identical-split baseline reruns.

4.1 Validated Result Impact Analysis

Under the current BUSI image-level split, the proposed framework reports 96.8% accuracy, 95.9% precision, 95.2% recall, 95.5% F1-score, and a ROC-AUC of 0.98. These values indicate strong discrimination on the stated partition, but they do not by themselves establish patient-independent or cross-institutional generalization. Accordingly, the revised analysis limits its

claim to within-dataset performance and treats patient-grouped, duplicate-controlled reruns and independent external ultrasound validation as necessary steps before clinical robustness can be inferred.

Within the stated split, the FROCED stage is retained as the model's calibration mechanism rather than as evidence of superiority over unrelated modalities. The original internal analysis reports a calibration error of 0.031, a false-positive rate of 3.2%, a false-negative rate of 2.9%, and detection reliability of 96.7% for the proposed model. These values are interpreted as internal reliability indicators and require replication under duplicate-controlled and external validation before broader clinical claims are made.

Because published breast-ultrasound studies use different datasets, class definitions, preprocessing pipelines, and validation splits, their reported metrics are used only to establish methodological context. The revised manuscript therefore avoids statistical ranking across independent publications and emphasizes matched reruns as the appropriate basis for future superiority testing.

Taken together, the current evidence supports a calibrated topology-aware CNN-Transformer classifier for BUSI, while the scope of validation is deliberately bounded to the available image-level experiment. This distinction prevents the within-dataset result from being interpreted as proof of multicentre robustness or clinical generalizability.

4.2 Validated Statistical Analysis

The previously reported paired t-tests and one-way ANOVA comparisons against values taken from unrelated or independently published breast-cancer methods are not retained as inferential evidence, because valid paired significance testing requires matched predictions produced on the same cases under an identical experimental protocol. The present manuscript therefore reports the available BUSI point estimates descriptively and does not assign p-values to cross-study differences.

For a leakage-controlled rerun, uncertainty should be quantified using repeated group-aware partitions or bootstrap resampling of the locked test set, with 95% confidence intervals reported for accuracy, macro-F1, ROC-AUC, and calibration error. Statistical comparisons with CNN, Transformer, and hybrid baselines should then use matched per-case or per-fold outputs from the same partitioning protocol, followed by multiplicity correction where several baselines are tested. Until such matched reruns are available, the reported numerical differences are interpreted descriptively rather than as statistically proven superiority.

4. CONCLUSION

This study proposed a calibrated hybrid deep learning framework for breast ultrasound classification. The framework integrates topology-aware tokenization through OMTAT, dual-stream local-global feature extraction using EfficientNet and Vision Transformer representations through DiSCo-TS, radiomics-preserving augmentation using CRPA, evidential calibration using FROCED, and representation regularization through X-MESD. The present experiments evaluate three-class classification and do not claim a validated lesion-segmentation contribution.

Under the stated BUSI image-level partition, the framework reports 96.8% accuracy, 95.9% precision, 95.2% recall, 95.5% F1-score, and 0.98 ROC-AUC. These results indicate strong within-dataset discrimination and suggest that topology-preserving tokenization and hybrid CNN-Transformer feature learning can capture both fine-grained lesion characteristics and broader tissue context. The calibration analysis further supports the value of confidence-aware inference, although these findings should be interpreted within the limits of the available split.

The current evidence does not establish patient-independent, multicentre, or external generalization. Patient-grouped and duplicate-controlled reruns, matched baseline comparisons, confidence intervals, and independent breast-ultrasound validation are required before stronger robustness claims are made. Pixel-level segmentation also remains a future multi-task extension requiring a dedicated decoder and explicit Dice, IoU, Hausdorff-distance, and boundary-sensitive evaluation.

Future research should focus on validating the framework using larger multi-center datasets collected from different imaging devices and patient populations. The model can also be extended through multimodal fusion of ultrasound, mammography, MRI, histopathology, and clinical metadata to provide a more comprehensive diagnostic representation. In addition, explainable AI methods such as Grad-CAM, SHAP, attention visualization, and uncertainty heatmaps should be integrated to improve clinical interpretability. Federated learning and self-supervised pretraining may further enhance

generalization while preserving patient privacy. Finally, prospective clinical validation and integration with radiology workflow systems such as PACS and RIS will be essential before real-world deployment. Prospective work should prioritize independently acquired breast-ultrasound cohorts and leakage-controlled patient-grouped evaluation before multimodal expansion or clinical deployment.

Table 4. Table of List of Abbreviations

Abbreviation	Full Form
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
CNN	Convolutional Neural Network
ViT	Vision Transformer
OMTAT	Omni-Modal Topology-Aligned Tokenizer
DiSCo-TS	Dual-Stream Co-attentive Topology-Steered Network
CRPA	Counterfactual Radiomics-Preserving Augmentation
FROCED	FROC-Calibrated Evidential Detector
X-MESD	Cross-Dataset Meta-Evaluation with Self-Distillation
BUSI	Breast Ultrasound Images Dataset
MRI	Magnetic Resonance Imaging
DBT	Digital Breast Tomosynthesis
CAD	Computer-Aided Diagnosis
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
ROC-AUC	Area Under Receiver Operating Characteristic Curve
FROC	Free-Response Receiver Operating Characteristic
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

Abbreviation	Full Form
Accuracy	Classification Accuracy
Precision	Positive Predictive Value
Recall	Sensitivity
Specificity	True Negative Rate
GPU	Graphics Processing Unit
Calibration Error	Prediction Probability Error
F1 Score	Harmonic Mean of Precision and Recall
FLOPS	Floating Point Operations Per Second
HER2	Human Epidermal Growth Factor Receptor 2
MXene	Two-Dimensional Transition Metal Carbide Material
SPR	Surface Plasmon Resonance
VGG16	Visual Geometry Group 16-Layer Network
RIS	Radiology Information System
RNA	Ribonucleic Acid
UWB	Ultra-Wideband
CPU	Central Processing Unit
PACS	Picture Archiving and Communication System
BCDCNN	Breast Cancer Deep Convolutional Neural Network
ANOVA	Analysis of Variance
t-test	Student's t-test
CI	Confidence Interval
p Value	Probability Value

ACKNOWLEDGEMENTS

The author gratefully acknowledges the support of the Department of Computer Science and Engineering, Oriental University, Indore, Madhya Pradesh, India. The author expresses gratitude to Dr. Atul Agrawal for his invaluable supervision and methodological advice. The contributors of the publicly accessible datasets utilized in this study are also acknowledged by the author.

CONFLICT OF INTEREST

Regarding the publication of this research work, the author affirms that there is no conflict of interest.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the support of the Department of Computer Science and Engineering, Oriental University, Indore, Madhya Pradesh, India. The author expresses gratitude to Dr. Atul Agrawal for his invaluable supervision and methodological advice. The contributors of the publicly accessible datasets utilized in this study are also acknowledged by the author.

REFERENCES

[1] Abed Maeedi, A., Abdulmohsin Hammood, D., Mezher Hasan, S. (2025). Breast Cancer Detection Using Deep Learning. Iraqi Journal for Computers and Informatics. https://doi.org/10.25195/ijci.v50i2.500

[2] Kumar, T., Chandran, M., Soniya Priyatharsini, G., et al (2024). Breast cancer detection using machine learning classifier. Cryobiology. https://doi.org/10.1016/j.cryobiol.2023.104820

[3] Nsairat, H., Allela, O., Shareef, A., et al (2026). Aptamer biosensors in early breast cancer detection. Clinica Chimica Acta. https://doi.org/10.1016/j.cca.2025.120536

[4] Lawal, M., Almousa, M., Ibrahim, A., et al (2025). Artificial intelligent-powered detection of breast cancer. Journal of Radiation Research and Applied Sciences. https://doi.org/10.1016/j.jrras.2025.101422

[5] Matondo-Mvula, N., Elleithy, K. (2024). Breast Cancer Detection with Quanvolutional Neural Networks. Entropy. https://doi.org/10.3390/e26080630

[6] Deendyal, V., Ghazaryan, L., Linden, E., et al (2025). Artificial Intelligence for Early Breast Cancer Detection. AI in Precision Oncology. https://doi.org/10.1089/aipo.2024.0051

[7] Yoen, H., Jang, M., Yi, A., et al (2024). Artificial Intelligence for Breast Cancer Detection on Mammography: Factors Related to Cancer Detection. Academic Radiology. https://doi.org/10.1016/j.acra.2023.12.006

[8] Martina Jaincy, D., Pattabiraman, V. (2025). BCDCNN: breast cancer deep convolutional neural network for breast cancer detection using MRI images. Scientific Reports. https://doi.org/10.1038/s41598-025-09974-0

[9] Lange, J., Gard, C., O'Meara, E., et al (2025). Breast density and risk of breast cancer: masking and detection bias. American Journal of Epidemiology. https://doi.org/10.1093/aje/kwae245

[10] Kassis, I., Lederman, D., Ben-Arie, G., et al (2024). Detection of breast cancer in digital breast tomosynthesis with vision transformers. Scientific Reports. https://doi.org/10.1038/s41598-024-72707-2

[11] [11] Wang, L. (2024). Mammography with deep learning for breast cancer detection. Frontiers in Oncology. https://doi.org/10.3389/fonc.2024.1281922

[12] Jose, A., Ezziane, Z. (2026). Breast cancer detection using BCDNet convolutional neural network. Array. https://doi.org/10.1016/j.array.2026.100711

[13] Kiani, P., Vatankhahan, H., Zare-Hoseinabadi, A., et al (2025). Electrochemical biosensors for early detection of breast cancer. Clinica Chimica Acta. https://doi.org/10.1016/j.cca.2024.119923

[14] Sadeghi, M., Kheradi, N., Nazem, S., et al (2025). Exosome biosensors for the detection of breast cancer. Clinica Chimica Acta. https://doi.org/10.1016/j.cca.2025.120456

[15] Jung, J., Kim, H., Kang, E., et al (2026). Enhancing Early Detection of Contralateral Breast Cancer in Breast Cancer Survivors Using AI-Assisted Mammography. Annals of Surgical Oncology. https://doi.org/10.1245/s10434-026-19348-z

[16] El Ghanaoui, F., Aniq, E., Chakraoui, M., et al (2025). Ensemble Learning for Breast Cancer Detection in Mammography. Ingénierie des systèmes d information. https://doi.org/10.18280/isi.300724

[17] Saadh, M., Saeed, T., Alfarttoosi, K., et al (2025). Plasmonic Nanoparticles: Enhancing Early Breast Cancer Detection Through Biosensors. Plasmonics. https://doi.org/10.1007/s11468-025-02864-1

[18] Munir, A., Huws, A., Khan, S., et al (2024). Mode of Recurrence/ Metastasis Detection in Breast cancer survivors. European Journal of Cancer. https://doi.org/10.1016/j.ejca.2024.113643

[19] Misiakou, M., Jensen, M., Talman, M., et al (2025). Enhancing HER2-low breast cancer detection with quantitative transcriptomics. npj Breast Cancer. https://doi.org/10.1038/s41523-025-00817-9

[20] Gump, F., Parmar, P., Feldman, S., et al (2024). The History of Breast Cancer Early Detection: 1865 – 2020. Breast Cancer: Targets and Therapy. https://doi.org/10.2147/bctt.s476416
[21] Fatima, T., Soliman, H. (2025). Application of VGG16 Transfer Learning for Breast Cancer Detection. Information. https://doi.org/10.3390/info16030227
[22] Razmpoosh, M., Namdar, A., Abdi-Ghaleh, R. (2026). Breast cancer detection using a MXene-based SPR biosensor. Journal of Electromagnetic Waves and Applications. https://doi.org/10.1080/09205071.2025.2581991
[23] M M, R., M, Y., R, T., et al (2025). Optimizing Early Breast Cancer Detection through Machine Learning Techniques. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5089164
[24] Zarif, S., Abdulkader, H., Elaraby, I., et al (2024). Using hybrid pre-trained models for breast cancer detection. PLOS ONE. https://doi.org/10.1371/journal.pone.0296912
[25] Hernández, M., Akbari-Chelaresi, H., Tashtarian, G., et al (2025). Breast Cancer Detection Using a Metasurface-Based Microwave Probe. IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology. https://doi.org/10.1109/jerm.2025.3572116
[26] Shen, J., Zhao, H. (2024). Allostatic Load in Breast Cancer Detection: A New Opportunity. Journal of Women's Health. https://doi.org/10.1089/jwh.2024.0483
[27] Ashok, G., Veerendra Nath, G., Kumar, R., et al (2025). Compact Decagonal Shaped UWB Antenna for Breast Cancer Detection. IETE Journal of Research. https://doi.org/10.1080/03772063.2025.2512048
[28] Unknown Author (2025). Artificial intelligence improves breast cancer detection in mammography screening. Nature Medicine. https://doi.org/10.1038/s41591-025-03714-7
[29] Hirsch, L., Huang, Y., Makse, H., et al (2025). Early Detection of Breast Cancer in MRI Using AI. Academic Radiology. https://doi.org/10.1016/j.acra.2024.10.014
[30] Aymaz, Ş., Aymaz, S. (2025). A novel approach for enhanced early breast cancer detection. Computer Methods in Biomechanics and Biomedical Engineering. https://doi.org/10.1080/10255842.2025.2553347
[31] Tasnim, J., & Hasan, M. K. (2023). CAM-QUS guided self-tuning modular CNNs with multi-loss functions for fully automated breast lesion classification in ultrasound images. Physics in Medicine & Biology, 69(1). https://doi.org/10.1088/1361-6560/ad1319
[32] Zhou, G., & Mosadegh, B. (2024). Distilling knowledge from an ensemble of vision transformers for improved classification of breast ultrasound. Academic Radiology, 31(1), 104-120. https://doi.org/10.1016/j.acra.2023.08.006
[33] Meng, X., Ma, J., Liu, F., Chen, Z., & Zhang, T. (2024). An interpretable breast ultrasound image classification algorithm based on convolutional neural network and Transformer. Mathematics, 12(15), 2354. https://doi.org/10.3390/math12152354
[34] Wang, W., Zhou, J., Zhao, J., Lin, X., Zhang, Y., Lu, S., Zhao, W., Wang, S., Tang, W., & Qu, X. (2025). Interactively fusing global and local features for benign and malignant classification of breast ultrasound images. Ultrasound in Medicine & Biology, 51(3), 525-534. https://doi.org/10.1016/j.ultrasmedbio.2024.11.014
[35] Aumente-Maestro, C., Díez, J., & Remeseiro, B. (2025). A multi-task framework for breast cancer segmentation and classification in ultrasound imaging. Computer Methods and Programs in Biomedicine, 260, 108540. https://doi.org/10.1016/j.cmpb.2024.108540
[36] Pawłowska, A., Karwat, P., & Żołek, N. (2023). Letter to the Editor. Re: "Dataset of breast ultrasound images by W. Al-Dhabyani, M. Gomaa, H. Khaled & A. Fahmy, Data in Brief, 2020, 28, 104863". Data in Brief, 48, 109247. https://doi.org/10.1016/j.dib.2023.109247
[37] Al-Dhabyani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020). Dataset of breast ultrasound images. Data in Brief, 28, 104863. https://doi.org/10.1016/j.dib.2019.104863.