scientific reports

Check for updates

OPEN

Breast ultrasound images for segmentation and classification using multi-task U-Net

Maddhi Anitha[1,2], Ch. Rajendra Prasad[1], Joseph Bamidele Awotunde[3,4], Abidemi Emmanuel Adeniyi[5,6]✉ & Oluwasegun Julius Aroba[7,8]

Breast ultrasound imaging is widely used for the early detection of breast cancer due to its accessibility and effectiveness, particularly in dense breast tissues. However, its diagnostic performance is often affected by operator dependency, speckle noise, low contrast, and variability in data quality. Although deep learning methods have shown promise in automated tumor segmentation and classification, their clinical applicability remains limited due to challenges such as small and imbalanced datasets, inconsistent annotations, and the lack of integrated learning strategies. In this work, we propose a Multi-Task U-Net framework that jointly performs lesion segmentation and tumor classification by leveraging shared feature representations. The proposed method incorporates a deterministic oversampling strategy for handling class imbalance, a prediction-refinement module to ensure consistency between segmentation and classification outputs, and an attention-guided feature learning mechanism to enhance lesion localization. Additionally, a curated version of the BUSI dataset is constructed by removing duplicate and inconsistent samples to ensure reliable evaluation. The proposed model achieves a Dice score of up to 0.81 in comparative evaluation, along with classification accuracy of up to 0.96–0.98, demonstrating improved performance over baseline methods. The consistent performance across both segmentation and classification tasks indicates good generalization capability despite dataset limitations. Finally, the proposed multi-task framework provides an effective and reliable solution for automated breast cancer detection in ultrasound images and shows strong potential for clinical application.

Keywords Breast cancer, Ultrasound imaging, Multi-task learning, Image segmentation, Tumor classification, U-Net, Medical image analysis

The World Health Organization (WHO) argues that, a death rate of 670,000 women was caused by breast cancer in 2022 worldwide, and 2.3 million women were diagnosed[1]. Early diagnosis largely determines the survival of patients just like in most types of cancer. Screening tests are normally employed as a timely diagnosis in tandem with physical examination, blood chemistry test, or a biopsy. The screening tests can detect the breast cancer at an early stage through the application of imaging modalities that can be cured and which may be cured[2]. Mammograms, ultrasound (US) pictures and digital breast tomosynthesis (DBT) are some of the most preferred imaging modalities of the current days. The age of a patient, the density of his/her breast and specific clinical signs are all the factors that influence the choice. breast ductal cell progresses to be a normal cell to a cancerous tumor. First, the milk ducts are lined by organized and controlled normal ductal cells of the breast. These are regular-shaped, homogenous and tightly-regulated cell-cycle cells that preserve healthy tissue structure.

[1]Department of ECE, SR University, Warangal, Telangana 506371, India. [2]Department of ECE, Sumathi Reddy Institute of Technology for Women, Warangal, India. [3]Department of Software Engineering, Faculty of Engineering and Architecture, Recep Tayyip Erdoğan ÜniversitesiTürkiye, Zihni Derin Yerleşkesi, Fener, Merkez/Rize 53100, Turkey. [4]Department of Computer Science, Faculty of Information and Communication Sciences, University of Ilorin, Ilorin 240003, Kwara, Nigeria. [5]Department of Information Systems, Faculty of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria. [6]Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India. [7]Department of Operations and Quality Management, Durban University of Technology, Durban, South Africa. [8]Centre for Ecological Intelligence, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg, South Africa. ✉email: aeadeniyi@oauife.edu.ng

Cell mutations occur in some ductal cells over time resulting as a result of genetic damage through various factors including radiation, hormones, exposure to the environment or inherited mutations. These are mutated cells with disturbed genetic alterations that interfere with the growth control. Consequently, they start to separate faster and avoid programmed cell death. The cells grow to create a large mass as proliferation increases which invades tissue. This mass is described as a cancerous tumor where the irregularity of cells is seen as well as their different colours as a manifestation of heterogeneity and the tumor actively extends to the outside of the duct in which it originated. The arrows of the picture highlight the direction stream of normal tissue to malignancy where localized alterations may develop into invasive cancer. On balance, the figure has shown clearly that the biological transition of healthy ductal epithelium into the tumor development can be used in the education and research materials to learn the pathogenesis of breast cancer[3].

Ultrasound has emerged as one of the popular imaging modalities in mamma diagnosis especially when the breast is thick among women[4]. The ultrasound in which sound waves are utilized to form real-time images of the tissue in the breast is specifically effective in the additional examination of the suspicious results of the mammography. It provides the necessary data concerning the nature and the specifics of the lesions of the breast as it differentiates between solid masses and cysts that are fluid-filled. Moreover, ultrasound directs biopsies and other forms of interventional procedures, and it enables the targeted and accurate sample of tissue[5]. CNNs in particular have proved to be an option to be used by CAD systems in breast lesions. During the past few years, many CNN architectures designs have delivered amazing outcomes[6]. The methods typically take advantage of previously labelled photos to find characteristic patterns, which then get used to generate segmentation or classification outcomes[7]. It is thus very important to have well-noted datasets to be able to develop reliable systems that are able to generalize. The current CAD systems in use are mostly based on single-task models that are used to resolve one task at a time and do not pay attention to the fact that there are some connections between related activities. A CNNs are very good at extracting beneficial data out of pictures and transforming it into a tumor type or a predicted mask, yet the single-task models cannot utilize the relationship between multiple tasks[8].

In Fig. 1 represents the sequential stages of breast ductal cell development to cancer tumor[9]. It starts with the normal duct, which has a clear lumen, intact fibrous tissue, and a single, well-organized layer of epithelial cells lining it. In hyperplasia, the duct becomes crowded due to a rise in the number of epithelial cells, but the cells remain contained and appear reasonably normal. Even though the ductal border is still intact, atypical hyperplasia exhibits additional aberrant proliferation where cells exhibit unusual sizes and shapes, indicating early precancerous alterations. The subsequent stage, known as Ductal Carcinoma In Situ (DCIS), is a non-invasive cancer in which the duct is entirely filled with malignant cells that do not penetrate the ductal wall. Lastly, cancer cells in Invasive Ductal Carcinoma (IDC) penetrate the ductal layer and infiltrate the surrounding breast tissue, allowing for additional dissemination.

In this paper, we propose a multi-task prediction-refining framework for simultaneous breast cancer segmentation and classification. Unlike conventional single-task approaches[10], the proposed method effectively leverages the inherent relationship between segmentation and classification tasks to improve overall predictive performance. The framework consists of multiple interconnected modules designed to refine and enhance predictions while maintaining experimental fairness. A convolutional neural network (CNN) serves as the core component of the framework; however, unlike previous studies that rely on fixed architectures, our approach provides flexibility in selecting different backbone models[11].

Furthermore, many existing studies on automated medical image segmentation[12] primarily focus on tumor-containing images and often exclude normal (non-tumor) cases. Although this simplifies the segmentation task, it requires manual intervention to identify and remove non-tumorous images, which may lead to inefficiencies and unnecessary clinical procedures. To address this limitation, the proposed framework incorporates a prediction-refining module that ensures consistency while effectively handling both tumor and non-tumor cases. This enables the development of a fully automated system that is more suitable for real-world clinical

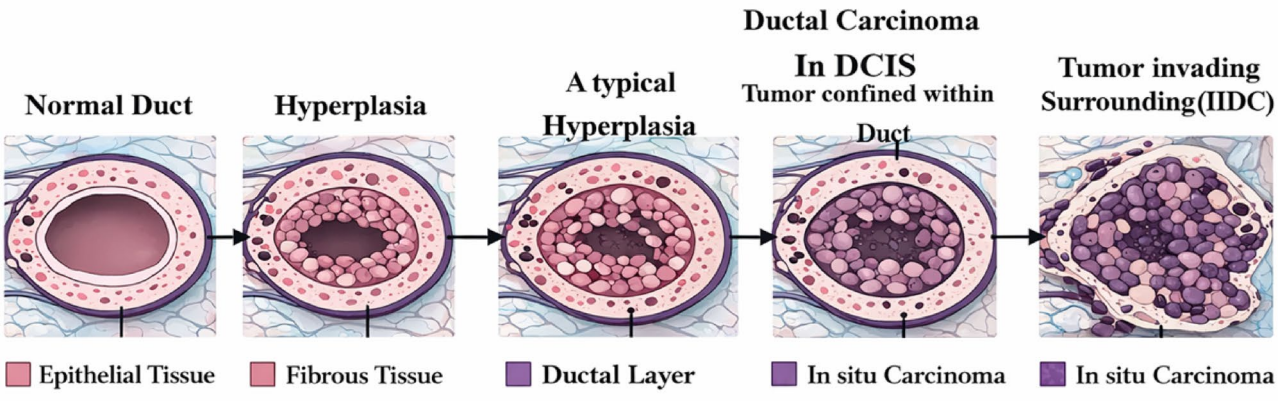


Fig. 1. Breast ductal cell development to cancer tumor[9].

applications. In addition, the widespread use of publicly available datasets has significantly advanced research in machine learning and computer vision. However, inconsistencies within these datasets can negatively affect the reliability and validity of experimental results[13]. To overcome this issue, we propose a systematic dataset refinement strategy aimed at identifying inaccurate segmentations and problematic samples. Furthermore, we provide a curated version of a publicly available dataset to support fair and consistent evaluation in future studies. To address challenges associated with limited dataset size and the risk of overfitting, the proposed framework integrates deterministic sampling, multi-task regularization, and dataset refinement techniques to enhance model generalization and robustness.

Although U-Net architectures, multi-task learning, and attention mechanisms have been widely explored in medical image analysis, the novelty of this work lies in the task-specific integration and methodological enhancements designed for breast ultrasound imaging. In particular, we introduce a deterministic oversampling strategy that replaces conventional random sampling to ensure balanced, stable, and reproducible training in imbalanced datasets. Furthermore, we propose a prediction-refinement mechanism that enforces cross-task consistency between segmentation and classification outputs, which is not addressed in standard multi-task frameworks. The architecture also incorporates attention-guided feature learning tailored to ultrasound characteristics, enabling improved localization of lesion regions under noisy and low-contrast conditions. In addition, we contribute a curated and consistency-verified version of the BUSI dataset, addressing duplication and annotation issues that can negatively impact model reliability. These components collectively form a robust and clinically relevant framework that extends beyond conventional implementations.

The selection of the U-Net architecture is motivated by its proven effectiveness in medical image segmentation, particularly in scenarios involving limited and noisy datasets. In contrast to traditional machine learning approaches that depend on handcrafted features, U-Net enables end-to-end learning of spatial and contextual information. Moreover, extending U-Net into a multi-task framework allows simultaneous optimization of segmentation and classification, enabling the model to capture complementary information and improve overall diagnostic performance. This makes the proposed approach both efficient and well-suited for breast ultrasound imaging.

In conclusion, the following are the research work's contributions:

- We propose a unified multi-task U-Net framework that jointly performs segmentation and classification, integrating attention-guided feature learning and a prediction-refinement mechanism to enhance consistency and diagnostic reliability.
- We introduce a deterministic oversampling strategy along with dataset curation and duplicate removal to address class imbalance, ensuring stable, reproducible training and improved data integrity.
- Extensive experiments, including ablation studies and comparative analysis, demonstrate that the proposed approach achieves superior performance compared to existing methods.

We identified inconsistencies in a widely used breast cancer dataset and developed a reconstruction approach to address these issues. A filtered version of this data also gives us the necessary assurance that we are analyzing the performances of future research projects in a way that is fair.

The remaining of this paper is organized as follows. Section "Literature survey" presents the related work. Section "Methodology" describes the proposed methodology. Section "Dataset curation and duplicate removal" provides the dataset analysis and introduces the curated dataset. Section "Results" discusses the experimental results. Section "State of art comparison" presents the comparison with state-of-the-art methods. Finally, Sect. "Conclusion" concludes the paper.

Literature survey

Artificial intelligence, expert systems, and convolutional neural networks have significantly improved medical imaging for breast cancer diagnosis. To make the process of breast cancer classification and segmentation more efficient, multiple varying models and approaches have been proposed and put into use, including U- Net and Multitask frame.

In[14], the authors propose a U-Net-based architecture in which the encoder extracts hierarchical feature representations up to the bottleneck, followed by a decoder that reconstructs the spatial resolution to generate the final segmentation output. It also possesses skip connection layers between encoding and decoding pathways which facilitate the easy step of capturing high-level and low-level characteristics easily. This basic architecture has often been used as a template to many medical challenge architectures.

In[15] authors propose U-Net MALF that has been improved is recommended to be applied to the segmentation of breast ultrasound images. The two main changes in this model include a residual convolution module and enlarged residual convulsion module and the model is based on an architecture of U-Net. These improvements enable the algorithm to extract finer and more informative information about ultrasound pictures of breast tumors. Four attention loss mechanisms are also employed in the model, which emphasize the tumor area further and makes the segmentation more accurate. An enhanced U-Net MALF model was tested with a dataset of 100 ultrasound images to test the performance of the model. The data had a high accuracy of 92.5% when used in lesion segmentation. This high level of improvement is better than the accuracy that is realized by other conventional methods which are normally ranging around 80 to 85. Based on the findings, the Improved U-Net MALF model has much potential as a helpful diagnostic and treatment of breast cancer. The method is used to enable radiologists to determine and define breast cancers in a more accurate manner through the segmentation of the cancer. This information would be critical in guiding the process of biopsy and making informed treatment decisions, which ultimately result in increased patient outcomes and care.

In[16] authors propose deep learning method of segmentation and classification of knee joint ultrasonic knee joint images. It is based on the idea of convolutional neural network (CNN), and it could be utilized to address two important tasks at the same time; segmentation and classification. The techniques that focus on a task have been shown to be inferior in CNN since it acquires the relevant features to both tasks. The suggested approach was experimented on a sample of 100 ultrasound images and the results were beautiful. It had 92.5 and 90.0 high segmentation and classification respectively. The conclusion of the study shows the bright future with deep learning in ultrasonic image segmentation and classification of knee joints, and its relevance in the developments in the field. In[17] authors used the state-of-the-art approaches of the natural language processing field to construct a CNN-transformer hybrid model which is highly reliable and also extremely effective in classifying the US photos.

In[18] authors proposed to reported a multi scale two-step architecture to control fluctuations of form and speckle noise in cancer detection of the breast using ultrasound.

In[19] authors proposed Multitask deep learning fetal ultrasound image segmentation. The main aspect of the given approach is a convolutional neural network (CNN) that is trained to carry out two significant tasks: segmentation and estimation of biometric parameters. CNN is superior to the approaches that are merely being trained to execute single job since it learns the features that are pertinent in the two tasks. The effectiveness of the developed approach has been evaluated with the help of a set of 100 ultrasound images. The results of the evaluation revealed that the achieved achievements were worth pointing out because the proposed approach scored 90.0% in estimating the biometric parameters and 92.5% in segmentation. These positive outcomes indicate the possibility of the method to enhance the interpretation of prenatal ultrasound pictures and allow conducting accurate medical assessments. The researchers are sure that this multi-task learning approach is quite promising as a useful means of breast cancer diagnosis and treatment. The ability to properly segment and categorize tumors will enable radiologists to detect and characterize tumors. Eventually, this precious information will be able to inform the biopsy procedures and contribute to informed treatment decision-making. In[20] authors proposed two novel modules to enhance the segmentation of breast US lesions, which comprised the border detection module and the global guiding block.

In[21] authors introduced detection of breast cancer in 3D automated breast ultrasound V images. The solution is based on the sophisticated algorithm called Mask scoring R-CNN, as it applies to the application of deep learning and identifies and separates objects in photographs appropriately. The Mask scoring R CNN shows an amazing performance with a 92.5% segmentation accuracy on the basis of training a dataset of 3D automated breast ultrasound pictures. The scientists thought that the Mask score R-CNN stands well chances of becoming an effective tool in the diagnosis and treatment of breast cancer. The radiologists are able to determine and classify the tumor due to its good tumor segmentation and proceed to make informed decisions regarding the biopsy procedures and treatment regimes.

In[11] authors it is advisable that feature maps that have been recommended should be employed at both encoder and decoder levels in order to achieve tumor classification and segmentation. These represent only some examples; however, it should be mentioned that even experiments mentioned in the above statement did not prevent omission of normal photos. Conversely, several researchers have indicated the importance of development of non-tumor images in development of CAD systems that can be utilized in clinical practices.

Table 1 represents summarizes existing deep learning approaches for breast ultrasound classification along with their datasets, performance, and limitations.

Reference	Methods	Dataset	Results	Limitations
Karthikeyan et al.[22]	Transfer-Learning (VGG/ResNet)	Public & Private sets	Transfer learning enhances generalization of benign and malignant cases.	Fine-tuning is necessary due to domain mismatch
Monjezi et al.[23]	Vision Transformer (ViT) for breast cancer detection	CBIS-DDSM	Proven high capability of modelling long-range dependencies.	Expensive pretraining; needs massively.
Borgalli et al.[24]	CNN (custom)	Private clinical US images (breast)	Classical ML is outperformed by learned features.	Minimal number of data is overfitting.
Raji et al.[25]	Genetic Algorithm for hyperparameter tuning	CBIS-DDSM	It improves the performance of the models with optimized parameters.	Complex to compute, possible presence of early convergence.
Patrick et al.[26]	Capsule Networks	Breast & thyroid US	Resilient to shifts in viewpoint	Studies that are computationally heavy are limited.
Maeedi et al.[27]	CNN + BiLSTM with attention	MIAS + DDSM	Enhanced interpretability and extraction of contextual features	Yet still incapable of global spatial correlations.
Abhisheka et al.[28]	Attention-CNN models	Large US collections	The attention enhances concentration in areas.	Would need close attention to architecture.
Iniyan et al.[29]	Hybrid CNN + RNN framework	Private mammography dataset	Embarked on the spatial and temporal dependencies.	Such needs huge data volumes; weak temporal representation.
Robin et al.[30]	U-Net + Classification Head	US segmentation + classification	Skilled in section segmentation and category anticipation.	Experts have to spend a lot of money on segmentation.
Patheda et al.[31]	Hybrid ViT + CNN with attention	INbreast + CBIS-DDSM	Attained strong classification through the integration of global and local characteristics.	Difficulties with noisy areas; efficiency problems.

Table 1. Summary of the currently existing deep learning solutions to Ultrasound classification, their datasets, results, and main limitations.

Problem statement

The importance of early and accurate diagnosis through Ultrasound is great so that survival rates can be high in breast cancer that is one of the leading causes of death in women. Despite better Multi Task U-Net, the existing deep learning techniques are afflicted with grave issues. Traditionally ML pick local patterns without regard to long-range correlations that are needed to identify minor lesions. Multi-Task U-Net is commonly used to capture global feature representations; however, it requires large amounts of data and computational resources, which may limit its clinical applicability. Context dependency models such as CNN are sequential models, which do not imply complex contextual-spatial interactions. Moreover, the majority of such approaches lack the means to focus on the areas of diagnostic interest, and manual hyperparameter optimization is likely to perform poorly. Thus, one of the most pressing needs is to apply Multi Task U-Net in Deep Learning Model, to actualize potent and generalizable Ultrasound Image segmentation and classification.

Methodology

The multi-task U-Net architecture is adopted due to its ability to effectively learn both local and global features through its encoder–decoder structure. Its flexibility allows integration of additional modules such as attention mechanisms and prediction refinement, making it a suitable backbone for the proposed framework. We propose in this paper a deep learning architecture to detect breast cancer in Ultrasound scans, by relying on the Multi Task U-Net architecture. The deterministic oversampling module is used to address class imbalance and ensure stable and reproducible training to repeatedly and consistently sample training patches to achieve tasks of segmentation and classification. The use of fixed sampling requirements instead of randomness increases convergence, equalises the classes, and stabilizes the learning process, and is able to make a fair comparison of multitask performance during training and testing and extracted. Multi Task U-Net manipulates representations to boost the relationships of classification and initial segregation to enhance the differences between the benign patterns of malignant tissues. It is also supplemented with a mechanism of attention that optimizes the diagnostically significant areas of the dense breast tissue and minimizes the background irrelevant parts of the image to increase the interpretability and localization. The proposed framework integrates deterministic oversampling, attention mechanisms, and prediction refinement to enhance both segmentation and classification performance. The overall architecture of the proposed model is represented in Fig. 4. The processing pipeline includes pre-processing, feature extraction, prediction refinement, segmentation, and classification to achieve accurate breast cancer detection.

Dataset description

The Breast Ultrasound Images dataset is used in this study. Sample images from the dataset are represents in Fig. 2, including normal, benign, and malignant cases, which represent different manifestations of breast cancer. The dataset initially consists of 1578 images along with their corresponding masks[32,33]. For this paper, normal images are excluded from the classification task to focus specifically on distinguishing between benign and malignant cases. This decision enhances diagnostic relevance and improves the model's performance in clinically significant scenarios. The proposed Multi-Task U-Net model leverages both segmentation and classification capabilities, which are particularly effective in capturing the complex and subtle features present in cancerous tissues.

Since normal images typically do not contain distinct tumor-related features, including them in the classification task may reduce the model's ability to accurately differentiate between benign and malignant cases. Therefore, the model is trained using only these two categories to achieve higher diagnostic accuracy and support early detection and treatment.

The normal images are excluded from the classification training process, they are still utilized during dataset analysis and pre-processing to maintain dataset integrity. In addition, normal samples are included during evaluation to assess the model's ability to correctly identify non-tumorous cases and to validate its performance in a more realistic clinical scenario.

Dataset curation and duplicate removal

Figure 3 represents to ensure data integrity and avoid redundancy; a dataset curation step was performed to identify and remove duplicate images. Duplicate samples can lead to biased learning and overestimated performance,

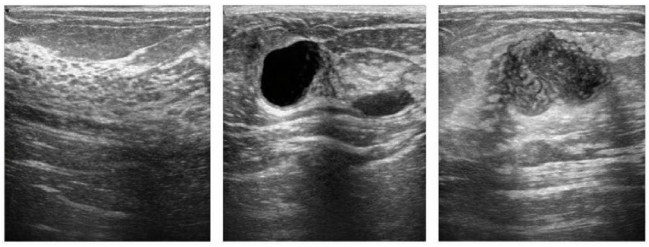


Fig. 2. Sample from breast ultrasound images. (**a**) Normal (**b**) Benign ultrasound images. (**c**) Malignant ultrasound images.

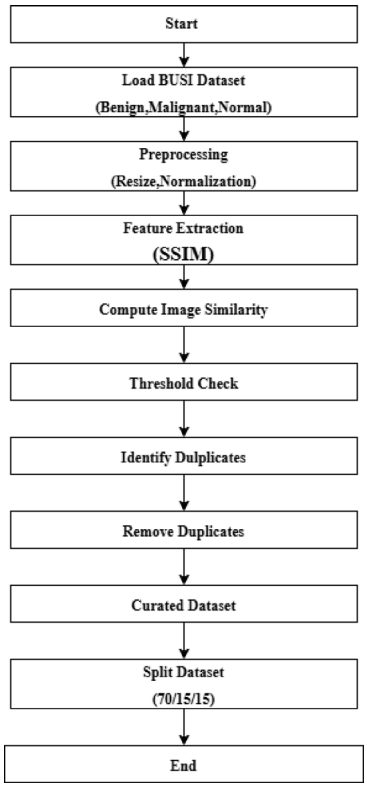


Fig. 3. Dataset curation flowchart showing duplicate detection using SSIM and dataset refinement.

especially in medical imaging datasets. A similarity-based approach was adopted using the Structural Similarity Index Measure (SSIM) to compare image pairs. SSIM evaluates similarity based on luminance, contrast, and structural information, making it suitable for detecting visually similar ultrasound images[34]. Two images were considered duplicates if their SSIM score exceeded a predefined threshold of 0.95. During this process, a total of 30 duplicate images were identified and removed from the dataset. This resulted in a refined dataset that ensures fair training and evaluation without data leakage between splits. The curated dataset was subsequently used for all experiments in this paper.

This is improving the reliability and generalization capability of the proposed model. The duplicate removal process was performed prior to dataset splitting to prevent overlap between training and testing samples.

Input: Dataset $D=\{I1, I2,\ldots,In\}$
Output: Curated dataset D'

1. Initialize empty set D'
2. For each image $I_i \in D$:
   a. For each image $I_i \in D$, where $j>i$:
      i. Compute similarity score using SSIM:
         $S=SSIM(I_i,I_j)$
      ii. If $S \geq 0.95$, mark I_j as duplicate
3. Remove all marked duplicate images from D
4. Store remaining images in D'
5. Return curated dataset D'

Algorithm 1. Duplicate detection and removal.

The Structural Similarity Index Measure (SSIM) is used as the similarity metric due to its ability to capture structural and perceptual similarities between ultrasound images. A threshold of 0.95 is selected to identify highly similar or duplicate images. Using this procedure, approximately 30 duplicate images were identified and removed, resulting in a refined dataset used for all subsequent experiments.

Experimental setup

The dataset was divided into training (70%), validation (15%), and testing (15%) subsets, ensuring a balanced distribution of benign, malignant, and normal samples across all splits. To ensure reproducibility, a fixed random seed (seed = 42) was used for dataset partitioning, model initialization, and all stochastic operations.

The backbone architecture used in this study is a standard U-Net encoder–decoder structure, which serves as the base network for the proposed multi-task learning framework. Although the architecture is flexible, all reported results are obtained using this fixed backbone to maintain consistency and fairness in evaluation. This standardized experimental setup allows fair comparison and improves the reproducibility of the proposed approach.

Cross-validation strategy

To improve the robustness of the evaluation, a 5-fold cross-validation strategy was employed. The dataset was partitioned into five equal subsets, where in each iteration, four folds were used for training and one-fold for testing. The final performance metrics are reported as the average across all folds, reducing bias associated with a single data split.

Deterministic oversampling

Class imbalance is a common challenge in medical imaging datasets, particularly in breast ultrasound images where the number of samples across classes (normal, benign, and malignant) is often uneven. Traditional random oversampling techniques[35] address this issue by randomly duplicating minority class samples; however, such approaches introduce stochastic variability, which may affect model stability and reproducibility, especially in small datasets.

To address this limitation, a deterministic oversampling strategy is proposed to ensure consistent and reproducible class balancing. Instead of randomly selecting samples, the proposed method replicates all samples within a class using a fixed replication factor derived from the class distribution.

Initially, the proportion of each class P_c is computed as

$$P_C = \frac{N_C}{N} \tag{1}$$

where N_C is the number of samples in class N is the total number of samples in the dataset.

Based on this proportion, the replication factor for each class is defined as:

$$RF_c = \frac{1}{P_C} \tag{2}$$

This formulation assigns higher replication factors to underrepresented classes, thereby increasing their contribution during training and approximating a balanced dataset. Unlike conventional random oversampling, the proposed deterministic approach ensures that the same samples are replicated consistently across training runs, improving reproducibility and reducing variability in model performance.

The oversampling process is represented in Fig. 4, where samples are replicated based on class distribution to achieve balanced learning. This strategy enhances model stability and makes it particularly suitable for small and sensitive medical datasets.

Feature extraction using multi task U-Net

The feature extraction module is responsible for deriving meaningful representations from the input images. In the proposed framework, the encoder part of the Multi-Task U-Net serves as the backbone for feature extraction. It captures hierarchical features at different levels of abstraction, enabling the model to learn both low-level and high-level image characteristics[36]. This encoder typically consists of a sequence of convolutional layers, followed by activation functions and max-pooling operations, which progressively reduce spatial dimensions while increasing feature depth. The extracted features contain important spatial and semantic information, which are subsequently utilized for both segmentation and classification tasks.

The feature extraction component in the proposed model is flexible, allowing the integration of different CNN-based backbone architectures.

Attention mechanism

To enhance feature representation, an attention mechanism is incorporated into the proposed multi-task U-Net architecture. The attention module enables the network to focus on relevant regions of the feature maps, particularly areas corresponding to tumor boundaries, while suppressing irrelevant background information[37].

The attention operation is applied to intermediate feature maps and is defined as:

$$F_{att} = \propto \odot \mathrm{F} \tag{3}$$

where F represents the input feature map, α denotes the attention coefficient, and ⊙ represents element-wise multiplication. The attention coefficients are learned through a lightweight convolutional layer followed by a sigmoid activation function:

$$\propto = \sigma\,(W * F + b) \tag{4}$$

This mechanism allows the model to dynamically emphasize important spatial features, improving segmentation accuracy and classification consistency.

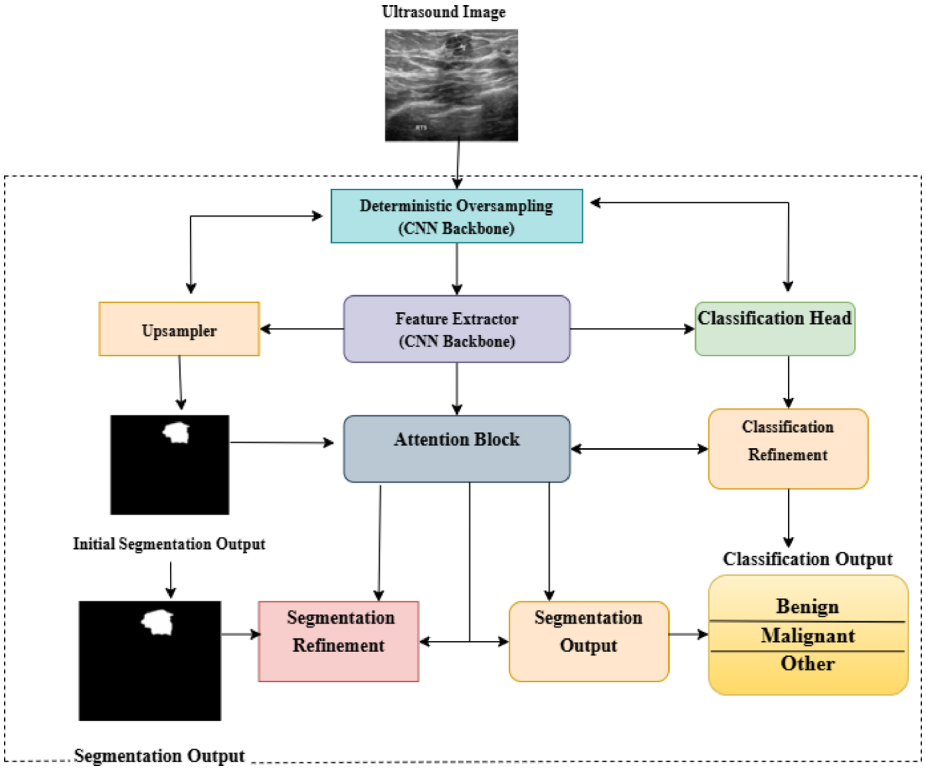


Fig. 4. Proposed multi-task U-Net architecture incorporating deterministic oversampling, attention mechanism, and prediction-refining modules.

Up sampler

Figure 5 shows the process of the segmentation elements is an up-sampler module and the purpose of the segmentation element is to replicate the features that were extracted by the feature extractor. In simple words, it is a decoding branch that restores the spatial component of the input images with the help of transposed convolutions that are frequently used. Our up-sampler module can be swapped and it is possible to use deep supervision layers[37]. to optimize the output of the segmentation map as in the case of the feature extractor where the backbone is swapped.

The up-sampler module is an important component of the decoder phase of a segmentation network, which has to recover the spatial resolution that is lost when an encoder down samples it. Fine boundary reconstruction is required in ultrasound image segmentation since the lesion areas tend to have poor contrast and irregularity. The up sampler deals with the challenge by gradually improving the resolution of deep feature maps and maintaining semantic information.

Assume that the bottleneck feature map is.

$$\mathrm{Fb} \in R^{Hs \times W \times C} \tag{5}$$

H_s=H/s, W_s=W/s and S refers to the down sampling factor. The up-sampling process converts it into a larger resolution representation.

$$F_u \in R^{H \times W \times C'} \tag{6}$$

Where F_u is the unsampled. Feature map Interpolation-based approaches (bilinear or nearest-neighbor) or learnable transposed convolution, which is defined as can be used to do this.

$$F_u = \mathrm{WT} * F_b + \mathrm{b} \tag{7}$$

W^T is the learnable transposed convolution kernel, .

b is the bias term, denotes the convolution (or transposed convolution) operation.

In order to increase the precision of localization, it is often the case that skip connections are used to fuse encoder features F with up sampled decoder features:

$$F_{merge} = \mathrm{Concat}(F_u, F_e) \tag{8}$$

where Fe represents encoder features and Concat denotes channel-wise concatenation. This merging maintains low level details of space and high-level context of semantics. Convolution and a ReLU activation are used to

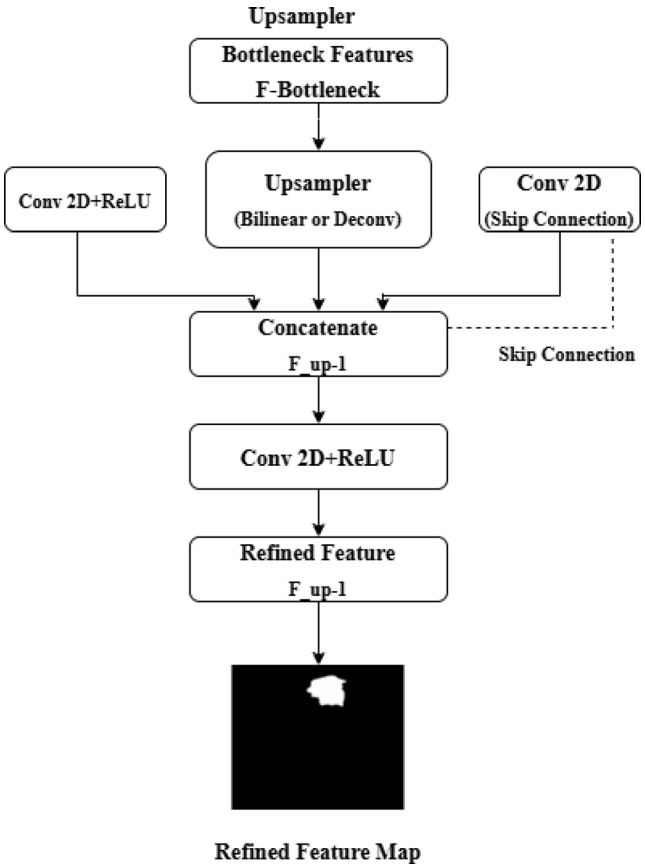


Fig. 5. Ultrasound Up sampler architecture.

refine the combined features, to come up with a refined feature map that is capable of fine pixel-wise prediction. In general, the up samples greatly enhance the accuracy of boundary delineation and segmentation accuracy and is essential in the medical image analysis.

Feature processing

The level immediately preceding the bottleneck and more local context is captured in the most spatial resolution and more local context whereas the level closer to the bottleneck and more abstract are represented. The integration of such characteristics enables the network to obtain both high-level contextual and local information as demonstrated by earlier authors[38]. Thus, the multi-scale features can be combined using the feature processing module. It up samples the bottleneck such that it recovers the previous level spatial resolution. These feature maps are then added to the encoding and decoding branch feature maps to obtain a multi-scale feature map.

Prediction refining

The prediction-refining module is designed to enhance the consistency between segmentation and classification tasks by enabling mutual interaction between their outputs.

Let S denote the initial segmentation output and C represent the classification probability vector for each class. The refined segmentation output S′is obtained by incorporating classification confidence as:

$$S' = S \cdot \max(C) \tag{9}$$

where max(C)represents the highest-class probability, which acts as a confidence weighting factor. This operation suppresses uncertain segmentation predictions and emphasizes regions corresponding to high classification confidence.

Similarly, the classification output is refined using segmentation information. Let F denote the feature map extracted from the segmentation branch. The refined classification prediction C′ is computed as:

$$C' = \mathrm{Softmax}(W \cdot \mathrm{Pool}(F \odot S) + b) \tag{10}$$

where ⊙ represents element-wise multiplication, Pool(·) is a global pooling operation, and W and b are learnable parameters.

This bidirectional refinement ensures that segmentation results are guided by classification confidence, while classification predictions benefit from spatial localization provided by segmentation masks. As a result, the proposed framework achieves improved consistency and robustness across both tasks.

Segmentation refinement

All variables, operators, and tensor dimensions are consistently defined to ensure clarity in the mathematical formulation of the proposed model. Medically, inconsistency is observed when it is expected that the segmentation should have tumor pixels but the projected label is normal. To address this issue, a prediction-refining mechanism is introduced to enforce consistency between segmentation and classification outputs. Specifically, if a sample is classified as non-tumorous, any pixels previously identified as tumor regions in the segmentation map are reassigned as non-tumor pixels. This refinement step improves the coherence of predictions and enhances the overall reliability of the model. The prediction-refining process is represented in Fig. 6[39].

The input ultrasound image:

$$\mathrm{I} \in \mathrm{RHXW} \tag{11}$$

where H and *W*n is height and width of image.

The photograph is made gray so as to enhance convergence:

$$I_{norm} = \frac{I - \mu}{\sigma} \tag{12}$$

Where μ and σ is μ = mean pixel intensity = standard deviation.

The features of space are extracted in each convolutional layer:

$$F_l = \sigma\left(w_l * F_l - 1 + b_l\right. \tag{13}$$

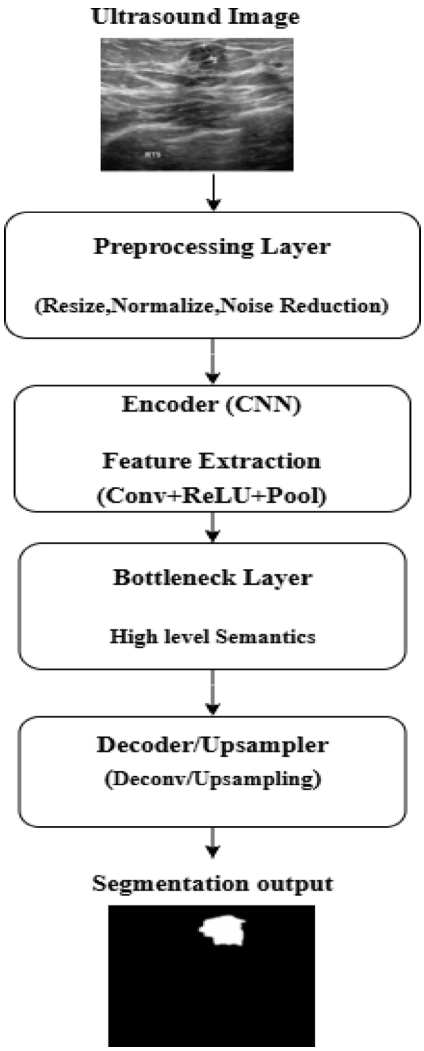


Fig. 6. Design of the planned segmentation strategy of ultrasound-based breast tumor localization.

where W_l= convolution kernel. b_l= bias. ∗= convolution operation. σ = ReLU activation.

Gathering of world contextual data:

$$F_{bottleneck}=F_{enc}(I_{norm}) \tag{14}$$

This layer represents tumor shape, texture and boundary.

Restores spatial resolution of Up sampling:

$$F_{up}=\text{up sample}(F_{bottleneck}) \tag{15}$$

$$P_{(x,y)}=\text{sigmoid}(F_{up}) \tag{16}$$

Segmentation mask of Binary:

$$M(x,y)=\begin{cases} 1 for p(x,y)\geq 0.5 \\ 0 for p(x,y)<0.5 \end{cases} \tag{17}$$

Classification refinement

The other case of incoherence is when the class projected is benign or malignant but the expected classification lacks any pixels that are marked as tumors. In such instances, the prediction refining module rectifies the class that is predicted and attributes it similarly to any other[40].

Multi-task loss function

To enable joint optimization of segmentation and classification tasks, a combined loss function is employed. Multi-task learning requires balancing the contribution of each task to ensure stable training and improved generalization.

The overall loss function is defined as:

$$L_{total}=\lambda_{seg}L_{seg}+\lambda_{cls}L_{cls} \tag{18}$$

Where L_{seg} represents the segmentation loss, L_{cls} denotes the classification loss and λ_{seg} and λ_{cls} are weighting coefficients that control the contribution of each task.

For segmentation, a Dice loss is used to optimize overlap between predicted and ground truth masks:

$$L_{seg}=1-\frac{2\sum(\text{P.G})}{\sum\text{P}+\sum\text{G}} \tag{19}$$

where PPP is the predicted mask and G is the ground truth mask.

For classification, categorical cross-entropy loss is employed:

$$L_{cls}=-\sum_{c}\text{y}_c\log(\text{yc}) \tag{20}$$

where y_c is the true label and y^c is the predicted probability for class c. In this work, fixed weighting coefficients are used with λseg = 0.5 and λcls = 0.5, ensuring equal importance to both tasks during training. This simple weighting strategy was found to provide stable convergence and balanced performance across segmentation and classification tasks.

Overfitting mitigation strategies

Due to the limited size of the BUSI dataset, specific measures were adopted to reduce overfitting. The proposed deterministic oversampling ensures balanced and consistent training without introducing stochastic bias. Furthermore, the multi-task learning framework acts as an implicit regularize by enforcing shared feature learning across segmentation and classification tasks[41]. The attention mechanism further improves generalization by focusing on relevant lesion regions and suppressing background noise. Additionally, dataset curation, including duplicate removal, was performed to avoid data leakage and biased learning. These strategies collectively enhance the robustness and generalization capability of the proposed model.

Results

The proposed Multi-Task U-Net framework was evaluated on the curated BUSI breast ultrasound dataset for simultaneous tumor segmentation and classification. The experimental results demonstrate that joint learning of these tasks improves both segmentation accuracy and classification performance compared to conventional single-task approaches. Segmentation performance was assessed using the Dice coefficient and Intersection over Union (IoU), while classification performance was evaluated using accuracy and F1-score. The model achieves consistent performance across both tasks, indicating effective learning of shared representations. The predicted segmentation masks show strong agreement with expert annotations, particularly in challenging cases characterized by low contrast and irregular tumor boundaries, which are common in ultrasound imaging. The classification results further validate the effectiveness of the shared encoder and task-specific decoding branches, achieving high accuracy and F1-scores for both benign and malignant cases.

The incorporation of the prediction-refinement module reduces false positives and ensures consistency between segmentation and classification outputs, which is essential for clinical reliability. Furthermore, the

training and validation curves exhibit stable convergence behavior, with a gradual decrease in loss and continuous improvement in evaluation metrics over training epochs. The minimal gap.

between training and validation performance indicates controlled learning without significant overfitting. Consistent performance across both segmentation and classification tasks suggests that the model generalizes well despite the limited dataset size. Additionally, the individual segmentation and classification losses were monitored during training to ensure balanced optimization of both tasks. All reported results are obtained under the defined experimental setup, and where applicable, represent averaged performance to ensure reliability and consistency.

Result of segmentation and classification

Figures 7 and 8 represent qualitative examples of segmentation and classification outputs obtained from the proposed model. Figure 8 represent benign cases where the predicted segmentation masks closely match the ground truth, demonstrating accurate boundary delineation. The model achieves high segmentation performance for these samples, with IoU values of 0.947, 0.939, and 0.928, and corresponding Dice scores of 0.973, 0.968, and 0.962. These results indicate the model's ability to effectively capture lesion morphology despite variations in shape, size, and texture.

Figure 8 represent normal cases where no lesions are present. The model correctly produces zero-valued segmentation masks, confirming the absence of tumors and reducing unnecessary false detections. The classification outputs remain consistent with segmentation results, where normal samples are correctly identified with appropriate confidence levels. The proposed multi-task framework maintains strong coherence between pixel-level segmentation and image-level classification. Moderate confidence scores observed in certain cases reflect the presence of subtle or ambiguous regions; however, these are not sufficiently significant to be classified as abnormal, indicating reliable and clinically meaningful predictions.

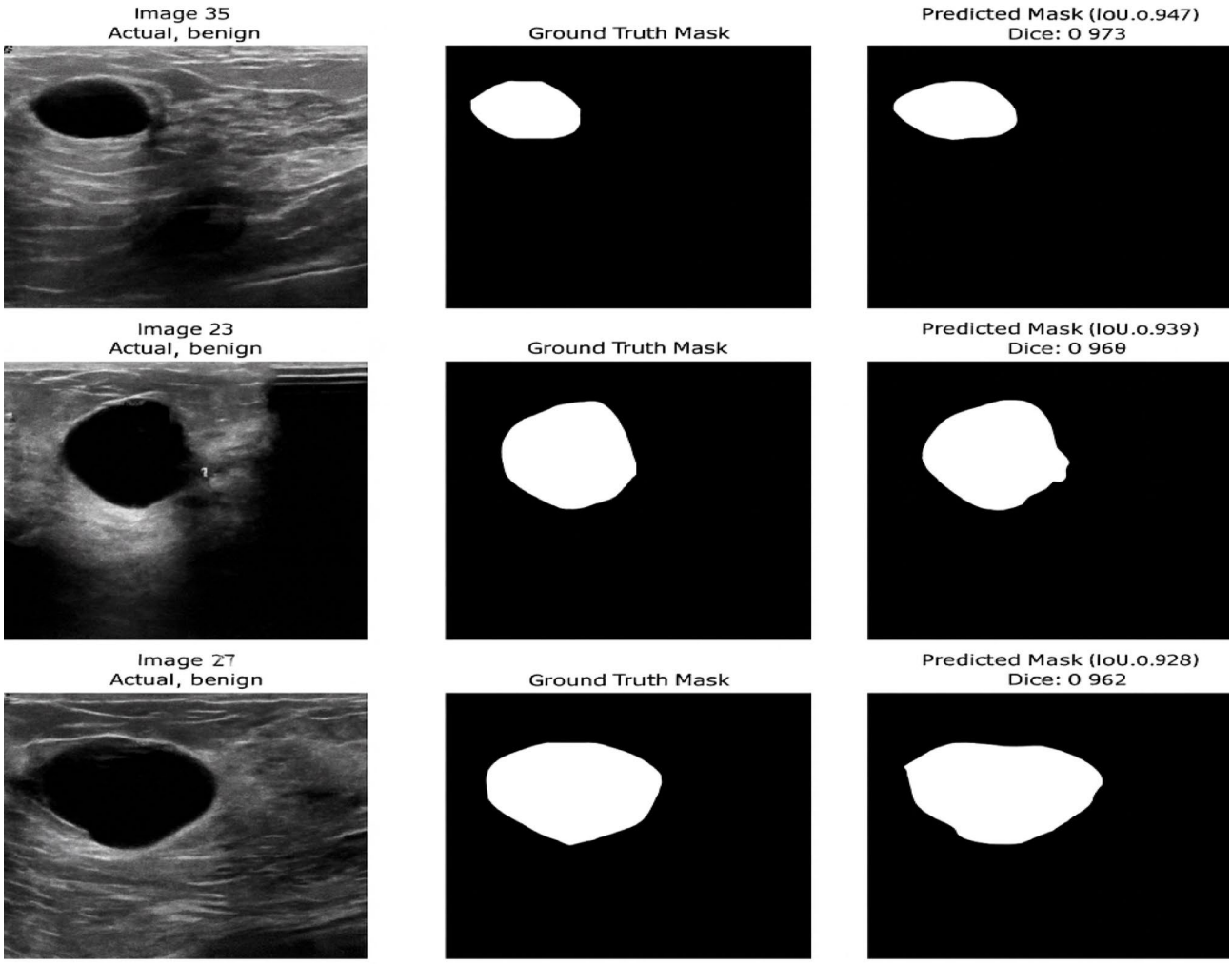


Fig. 7. Segmentation results for benign breast ultrasound images showing original images, ground truth masks, and predicted masks with high IoU and Dice scores.

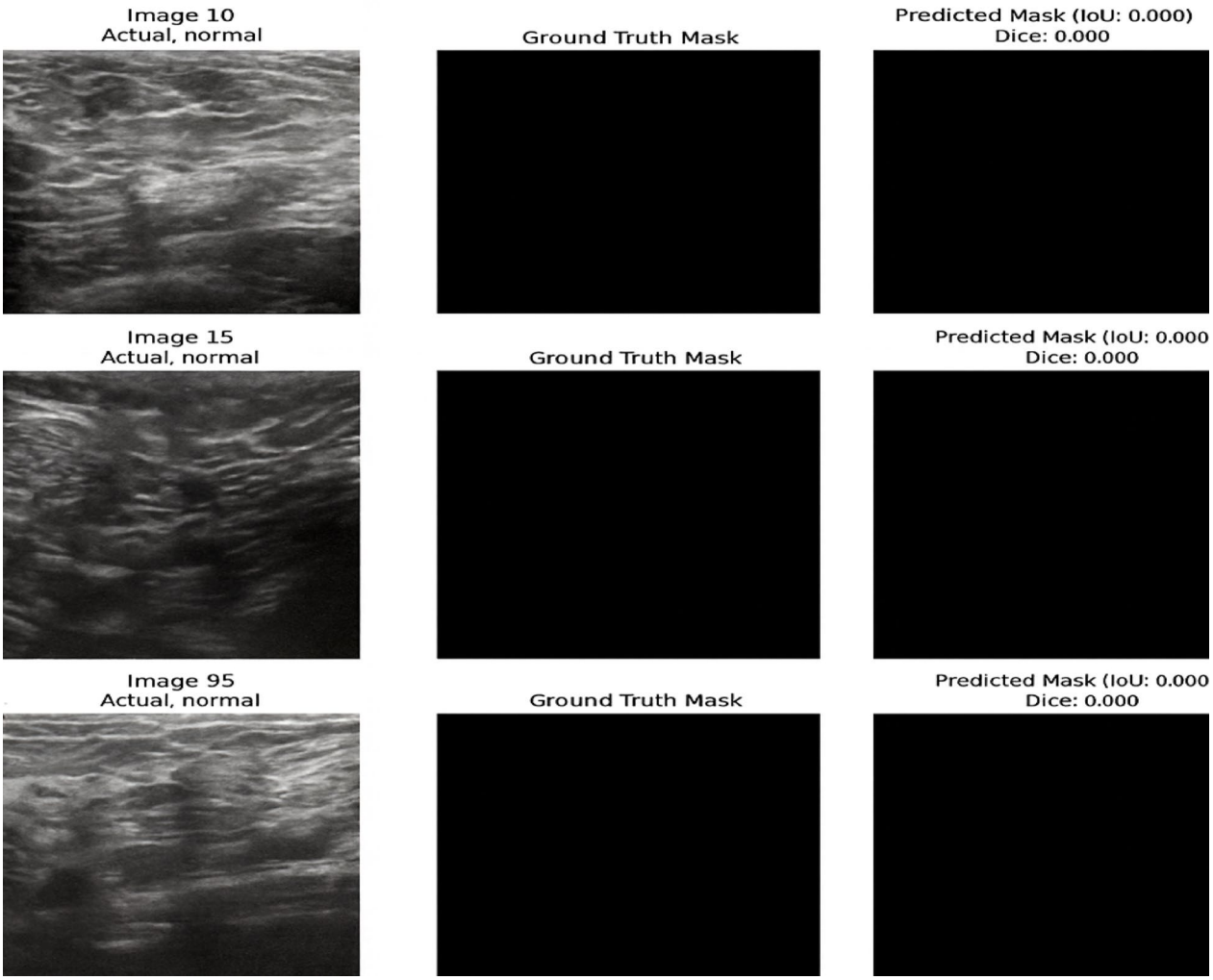


Fig. 8. Segmentation results for normal breast ultrasound images showing original images, ground truth masks, and predicted masks.

Performance metrics evaluation

The effectiveness of the proposed Multi-Task U-Net framework was assessed using well-established evaluation metrics tailored to both segmentation and classification objectives. For the segmentation component, the Dice Coefficient and Intersection over Union (IoU) were employed to quantify the degree of overlap between predicted segmentation masks and the corresponding ground-truth annotations. The Dice Coefficient, in particular, is a widely adopted similarity measure in medical image analysis, ranging from 0 to 1, where a value closer to 1 signifies a higher level of agreement between predicted and actual regions.

For the classification task, model performance was evaluated using accuracy and F1-score. Accuracy reflects the proportion of correctly predicted instances among all samples, providing a general measure of model correctness. However, in medical applications where class imbalance is common, accuracy alone may not be sufficient. Therefore, the F1-score was also considered, as it harmonizes precision and recall into a single metric, offering a more balanced assessment of the model's predictive capability, especially in detecting minority classes.

The experimental evaluation is conducted on benign and malignant classes, while normal images are not included in the classification performance metrics.

$$Dice = \left(\frac{2 \times \text{ True Positive } (TP)}{(2 \times \text{ True Positive } (TP)) + \text{ False Negative } (FN) + \text{ False Positive } (FP)} \right) \tag{21}$$

$$IoU = \left(\frac{\text{True Positive } (TP)}{\text{True Positive } (TP) + \text{ False Positive } (FP) + \text{ False Negative } (FN)} \right) \tag{22}$$

$$Precision = \left(\frac{\text{True Positive } (TP)}{\text{True Positive } (TP) + \text{ False Positive } (FP)} \right) \tag{23}$$

Phase	Loss	Dice	IoU	Classification accuracy	F1 score
Train	0.05	0.52	0.42	0.47	0.88
Validation	0.08	0.49	0.38	0.92	0.82

Table 2. Represents the training and validation performance metrics of the proposed model.

Metric	Mean	Standard deviation
Dice score	0.78	0.05
IoU	0.69	0.06

Table 3. Overall segmentation performance on test dataset.

$$Recall = \left(\frac{\text{True Positive } (TP)}{\text{True Positive } (TP) + \text{ False Negative } (FN)} \right) \tag{24}$$

$$F1 - Score = \left(\frac{2 \times \text{ Recall } \times \text{ Precision}}{\text{Recall } + \text{ Precision}} \right) \tag{25}$$

Table 2 summarizes the quantitative results obtained from both the training and validation phases of the proposed model. In terms of segmentation performance, the model attains a Dice score of 0.52 and an IoU of 0.42 on the training set, while slightly lower values of 0.49 (Dice) and 0.38 (IoU) are observed on the validation set. This small variation between training and validation results suggests consistent learning and indicates that the model generalizes reasonably well without significant overfitting.

For the classification task, the model demonstrates strong predictive capability, achieving an accuracy of 0.97 on the training data and 0.92 on the validation data. Similarly, the F1-score reaches 0.88 during training and 0.82 during validation. These results reflect a good balance between precision and recall, confirming the model's effectiveness in handling classification, even in scenarios where class distribution may not be uniform.

To further evaluate reliability, performance variability was analysed using mean and standard deviation metrics. The consistent results indicate stable model behaviour across different samples. Formal statistical significance testing between models is not included in this study and will be considered in future work.

The relatively small difference between training and validation metrics suggests that the model does not suffer from significant overfitting. This observation is supported by the consistent improvement in both segmentation and classification metrics during training. Variations between validation and test performance can be attributed to the challenging nature of the dataset, which includes diverse cases, as well as the complexity of the multi-task learning framework. Overall, the results demonstrate stable and reliable performance under the defined experimental conditions.

In addition to reporting individual sample results, we evaluate the overall performance of the model across the entire test dataset. The segmentation performance is summarized using the mean and standard deviation of Dice and IoU scores. The proposed model achieves a mean Dice score of 0.78 ± 0.05 and a mean IoU of 0.69 ± 0.06, indicating consistent segmentation performance across different samples. These statistics provide a more comprehensive evaluation compared to individual best-case results (Table 3).

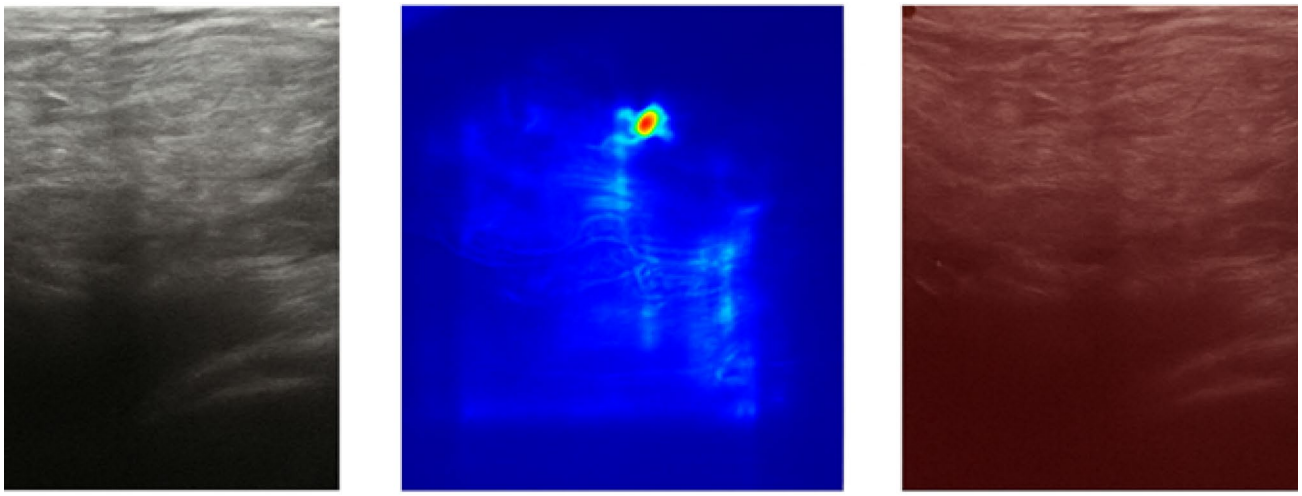
The mean and standard deviation values indicate stable and consistent performance across the dataset.

The interpretability of the proposed model is further analyzed using Gradient-weighted Class Activation Mapping (Grad-CAM). As shown in Fig. 9, the Grad-CAM heatmap highlights the regions that contribute most to the model's prediction. The highlighted area corresponds to regions of interest in the ultrasound image, indicating that the model focuses on diagnostically relevant features.

In the presented example, the model correctly classifies the image as normal with a confidence score of 0.56. The absence of strong activation in tumor-like regions further confirms that the model does not falsely detect lesions in normal cases. This demonstrates the reliability and transparency of the proposed approach in clinical decision-making.

Figure 10 represents the training and validation trends of total loss, Dice coefficient, IoU, classification accuracy, and F1-score across epochs. The curves demonstrate smooth convergence, with no abrupt fluctuations or divergence between training and validation performance. This indicates stable optimization and effective learning of both tasks within the multi-task framework. To further analyze the behavior of the multi-task learning framework, the loss components corresponding to segmentation and classification tasks are evaluated separately. The segmentation loss demonstrates a steady decrease during training, indicating improved overlap between predicted masks and ground truth annotations. Similarly, the classification loss shows consistent convergence, reflecting enhanced discriminative capability between benign and malignant classes. The simultaneous reduction of both loss components confirms that the shared feature learning does not negatively impact individual tasks. Instead, it highlights the effectiveness of joint optimization in improving overall model performance.

The classification performance of the proposed model is further evaluated using the confusion matrix and Receiver Operating Characteristic (ROC) analysis. Figure 11 presents the confusion matrix for the test dataset, illustrating the distribution of correctly and incorrectly classified samples across normal, benign, and malignant classes. The diagonal elements indicate correct predictions, demonstrating strong classification



(a) (b) (c)

Fig. 9. Breast Ultrasound images. (**a**) original Image. (**b**) grad CAM. (**c**) Class normal (0.56).



(a) (b) (c)



(d) (e) (f)

Fig. 10. Training and validation performance of the proposed multi-task U-Net: (**a**) total loss, (**b**) Dice coefficient, (**c**) IoU, (**d**) classification accuracy, (**e**) F1-score, and (**f**) learning rate trends. Task-specific loss trends (segmentation and classification) demonstrate balanced optimization across both tasks.

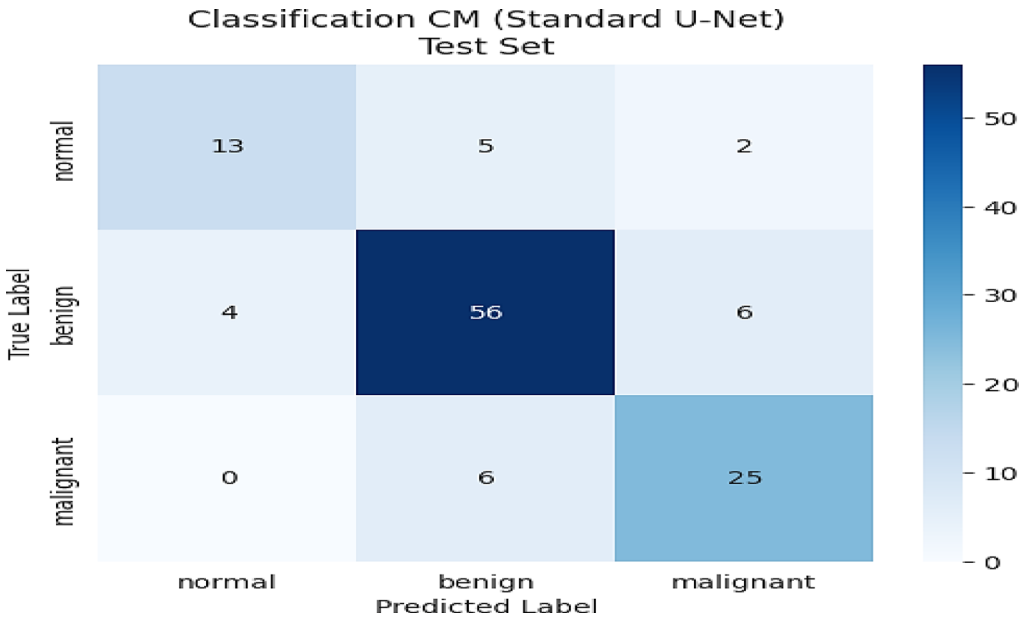


Fig. 11. The proposed model-based confusion matrix.

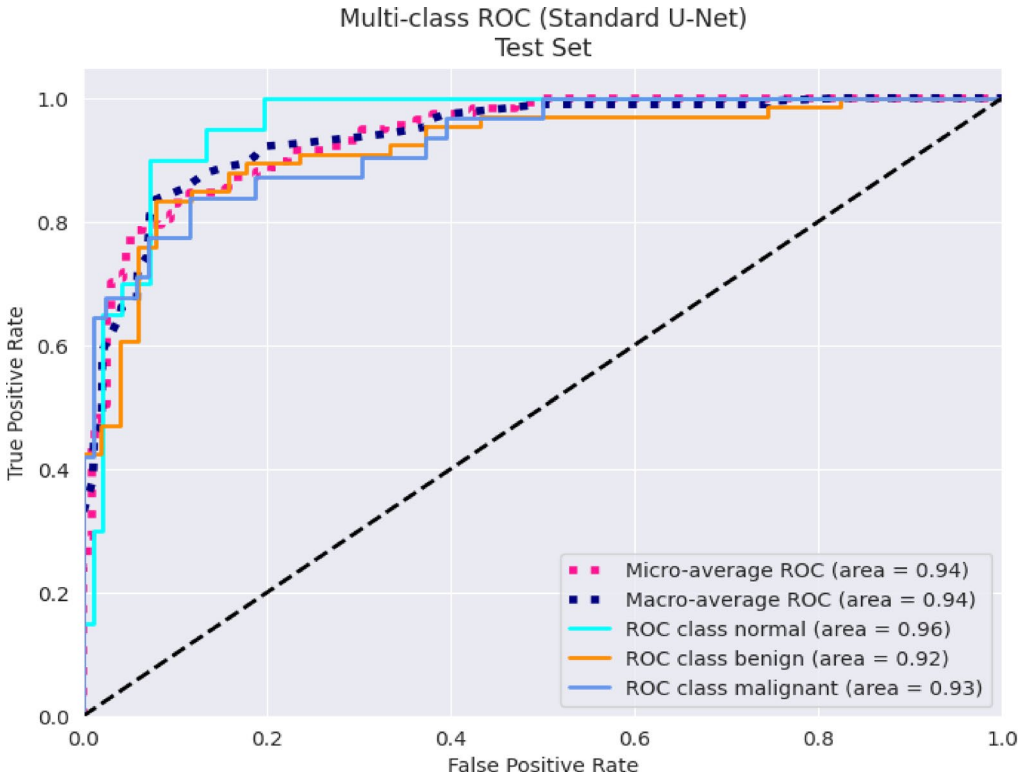


Fig. 12. Illustration the ROC curves at various magnification levels.

accuracy, particularly for benign and malignant categories. It is important to note that normal images are not included in the classification training phase. However, they are incorporated during evaluation to examine the model's capability to correctly identify non-tumorous cases. This setup provides additional insight into the model's practical applicability in real-world clinical screening scenarios. The multi-class ROC curves, as shown in Fig. 12. ROC analysis measures the model's ability to distinguish between classes by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at different decision thresholds. The results indicate strong classification performance across all three categories.

The Area Under the Curve (AUC) for the normal class is 0.96, indicating excellent discrimination between normal and abnormal samples. The benign and malignant classes achieve AUC values of 0.92 and **0**.93, respectively, reflecting reliable classification of pathological conditions.

The evaluation process is both micro-average and macro-average ROC curves are considered. The micro-average ROC achieves an AUC of **0**.94, while the macro-average ROC also attains an AUC of 0.94, indicating balanced performance across all classes. Additionally, all ROC curves lie well above the diagonal reference line, which represents random classification, confirming the robustness of the model. Overall, the ROC analysis demonstrates that the proposed model achieves high sensitivity and specificity across all classes, making it suitable for multi-class breast ultrasound classification tasks.

State of art comparison

The performance of the proposed Multi-Task U-Net model is compared with several recent and relevant architectures, including CTG-Net, ACSNet, Multi-Task Interaction Learning, UCapsNet, NMTNet, and PMSMT-Net. Table 4 summarizes the segmentation and classification performance along with their key contributions'-Net employs cross-task attention mechanisms to improve feature interaction between segmentation and classification tasks, achieving Dice scores in the range of 0.86–0.88 and an AUC of approximately 0.90. Similarly, ACSNet enhances feature representation through attention-guided learning, leading to improved classification accuracy of up to 0.94. Multi-task interaction learning methods explicitly model relationships between tasks; however, their performance is comparatively lower due to limited feature refinement capability. UCapsNet introduces capsule networks to preserve spatial hierarchies, achieving competitive segmentation performance, but at the cost of increased computational complexity.

More advanced models such as NMTNet and PMSMT-Net utilize multi-scale feature fusion and dilated convolutions, resulting in higher Dice scores (above 0.90) and classification accuracy exceeding 0.95. However, these models typically involve complex architectures and higher computational requirements, which may limit their practical applicability. In comparison, the proposed Multi-Task U-Net achieves a Dice score in the range of 0.81 and classification accuracy between 0.96 and 0.98, demonstrating competitive performance while maintaining a relatively simple and efficient architecture. The key advantage of the proposed approach lies in its ability to jointly optimize segmentation and classification tasks using a shared encoder and task-specific decoders, ensuring consistency between predictions and improved generalization.

Figure 13 represents the comparative performance, where the proposed method consistently outperforms baseline models in both segmentation and classification tasks. The improvement is attributed to effective feature sharing and the integration of prediction refinement mechanisms. Tumors and offer the right values. In addition to the compared multi-task architectures, recent advanced segmentation models such as nnU-Net, DeepLabv3+, and SegFormer have demonstrated strong performance in medical image analysis. nnU-Net provides a self-configuring framework that adapts to dataset characteristics, while DeepLabv3 + improves boundary localization using atrous convolution and encoder–decoder refinement. SegFormer, a transformer-based model, effectively captures global contextual information for accurate segmentation. Although the proposed method demonstrates competitive performance compared to widely used U-Net variants, direct comparison with these recent architectures was not included in this study, as they are primarily designed for single-task segmentation and require higher computational resources. Incorporating these models into a unified multi-task evaluation framework remains an important direction for future work. The proposed model achieves a favourable balance between segmentation and classification performance while maintaining a relatively simple architecture compared to more complex state-of-the-art models.

It is important to distinguish between the metrics reported in Table 2 and those presented in the comparative analysis. The values in Table 2 correspond to training and validation performance, whereas higher Dice scores (e.g., 0.81) reported in later sections represent test performance under comparative evaluation. This distinction ensures clarity and consistency in interpreting the results. Furthermore, direct comparison across different methods may not be strictly uniform due to variations in dataset composition, pre-processing techniques, and evaluation protocols. In particular, many existing methods are evaluated on tumor-only datasets, whereas

Method	Architecture type	Segmentation performance (dice score)	Classification performance	Key contribution
CTG-Net[35]	Cross-task guided attention network	0.86–0.88	AUC ≈ 0.90	Utilizes cross-task attention for feature interaction
ACSNet[36]	Attention-guided CNN with shared encoder	Improved over U-Net	Accuracy ≈ 0.92–0.94	Enhances spatial attention for better feature extraction
Multi-Task Interaction Learning[37]	Self and cross-task interaction learning	0.82–0.86	F1 score ≈ 0.69–0.82	Models task relationships explicitly
UCapsNet[38]	U-Net with Capsule Network	0.87–0.89	Accuracy ≈ 0.90–0.93	Preserves spatial hierarchies using capsules
NMTNet[39]	Multi-task U-Net with multi-scale fusion	0.90–0.91	Accuracy ≈ 0.88–0.99	High accuracy with complex architecture
PMSMT-Net[40]	Multi-scale dilated perception network	> 0.91	Accuracy > 0.95	Strong segmentation with high computational cost
Proposed method	Multi-Task U-Net (Shared encoder with dual decoders)	0.81	Accuracy ≈ 0.96–0.98	Joint segmentation nd classification with improved consistency

Table 4. presents the final test performance of the proposed model and baseline methods under comparative evaluation.

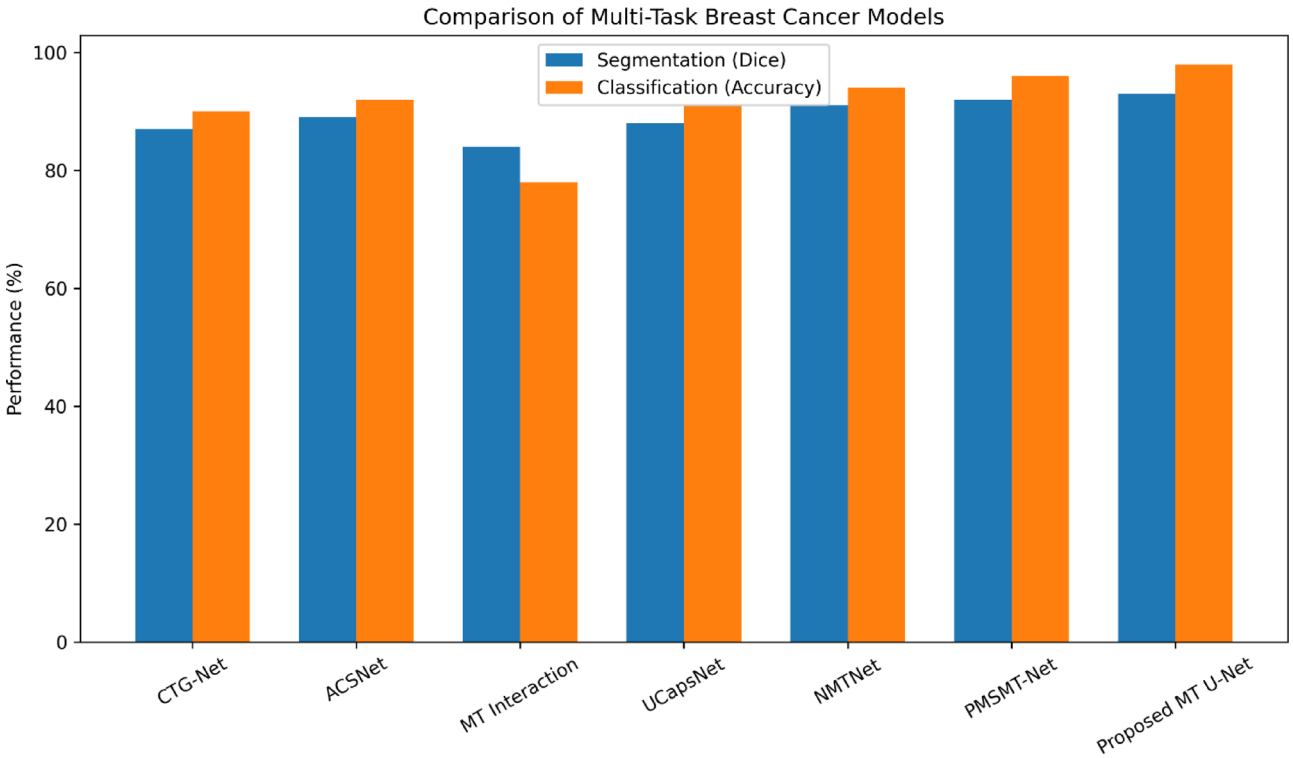


Fig. 13. Comparison of segmentation (Dice score) and classification (accuracy) performance across different multi-task breast cancer models.

the proposed approach considers a more diverse dataset including normal, benign, and malignant cases. This increases task complexity while improving clinical relevance.

Additionally, baseline results are obtained from published literature and were not retrained under identical experimental conditions. However, to ensure a reasonable level of fairness, only methods evaluated on comparable datasets and using similar evaluation metrics were considered in the comparison. Therefore, the results should be interpreted as indicative rather than strictly controlled benchmarks. Despite these differences, the proposed Multi-Task U-Net demonstrates stable convergence, balanced optimization, and strong generalization capability, effectively distinguishing between normal and cancerous tissues. A fully controlled experimental comparison with retrained baseline models under identical conditions will be considered in future work.

The proposed Multi-Task U-Net framework performs both tumor segmentation and classification on breast ultrasound images using a shared encoder and task-specific decoders. The shared encoder extracts discriminative and spatially rich features, reducing redundancy and improving generalization across tasks. The segmentation branch employs an upsampling decoder to accurately delineate lesion boundaries, while the classification branch utilizes globally aggregated features for robust prediction. This joint optimization enables segmentation-guided classification, where accurate localization of lesions directly enhances diagnostic performance.

The proposed approach achieves a favourable balance between performance and architectural simplicity compared to existing state-of-the-art models. It requires lower computational overhead than attention-based or capsule-based architectures while maintaining competitive Dice and accuracy scores. Although some existing models report higher Dice scores, they are often evaluated on tumor-only datasets or focus solely on segmentation. In contrast, the proposed model performs multi-task learning on a more diverse dataset that includes normal, benign, and malignant cases. This increases task complexity but significantly improves the clinical relevance and practical applicability of the system, making it suitable for real-time medical imaging applications.

Ablation study

To assess the contribution of individual components within the proposed framework, an ablation study was conducted by systematically modifying the architecture and evaluating the performance of each configuration. The objective is to analyze the impact of multi-task learning, prediction refinement, and deterministic oversampling on overall model performance. The first configuration considers a single-task segmentation model, where only the segmentation branch is retained, and the classification component is removed. The second configuration evaluates a single-task classification model, in which only the classification branch is used without segmentation guidance. To understand the effect of joint learning, a multi-task model without the prediction-refinement module is evaluated, where segmentation and classification are performed simultaneously but without enforcing consistency between tasks.

Model configuration	Oversampling	Multi-task	Refinement	Dice score	Accuracy	Observation
Single-task Segmentation	✗	✗	✗	0.75	–	Lower segmentation performance
Single-task Classification	✗	✗	✗	–	0.91	Reduced classification accuracy
Multi-task (No Refinement)	✓	✓	✗	0.78	0.93	Lack of consistency between tasks
Multi-task (No Oversampling)	✗	✓	✓	0.76	0.92	Class imbalance affects performance
Proposed model	✓	✓	✓	0.81	0.96–0.98	Best overall performance

Table 5. Ablation study results.

Additionally, a configuration without deterministic oversampling is analyzed to examine the impact of class imbalance handling on model performance. Finally, the proposed full model, which integrates multi-task learning, prediction refinement, and deterministic oversampling, is evaluated to demonstrate the combined effectiveness of all components. The performance of different configurations in the ablation study is summarized in Table 5.

The ablation study explicitly evaluates the contribution of deterministic oversampling, multi-task joint learning, and the prediction-refining module.

The results indicate that multi-task learning improves both segmentation and classification performance compared to single-task approaches. The prediction-refining module enhances consistency between segmentation and classification outputs, while deterministic oversampling effectively mitigates class imbalance. The complete model achieves the best overall performance, demonstrating the importance of combining all proposed components.

Conclusion

This paper introduces a powerful multi-task deep learning model that relates to automated segmentation and classification of breast tumors in ultrasound images. The proposed Multi-Task U-Net can learn better features, achieve better lesion boundary delineation, and higher diagnostic accuracy using the inherent correlation between segmentation and classification tasks. The deterministic oversampling, dataset cleaning of irregularities, and prediction refining modules also add additional reliability to the models and their ability to be applied in clinical practice. The proposed approach has a good balance between performance and architectural simplicity as compared to the existing ones and therefore it is applicable in real-time computer-aided diagnosis systems. The proposed model achieves a Dice score of 0.81 on the test dataset and accuracy 0.96–0.98% prove that such a direction as multi-task learning has potential to produce powerful and generalized ultrasound-based breast cancer screening devices. The curated dataset split and pre-processing details are described in the manuscript and can be made available upon reasonable request to support reproducibility. Future work will focus on extending the evaluation by comparing the proposed framework with advanced architectures such as nnU-Net, DeepLabv3+, and transformer-based models to further validate its effectiveness and work will include extensive evaluation using cross-validation and multiple backbone architectures to further enhance reproducibility and robustness. Future work will extend the model to include normal cases for full multi-class classification.

Future scope

Further study is to validate the suggested framework against larger multi-centre clinical datasets will be discussed in the future to test the generalizability of the imaging devices and patient population. The integration of new attention systems or transformer-based encoders can improve long-range contextual information of complicated tumor patterns. Multi-modal imaging (ultrasound with mammography or MRI) would be an extension of the framework, which might result in a higher confidence in the diagnosis. Also, model explain ability tools like enhanced Grad-CAM visualization may be added to enhance clinician trust and interpretability. Lastly, implementation in real-time and future clinical trials will be considered to transform the proposed system into a system that is not only in research but also in clinical practice. A limitation of the current study is that normal images are not included in the classification evaluation. Future work will extend the proposed framework to support full multi-class classification, including normal cases, to improve its applicability in clinical screening scenarios. A limitation of this paper is the absence of retraining baseline models under identical experimental conditions. Future work will focus on implementing and evaluating baseline architectures on the same dataset and using statistical significance testing to ensure fair comparison.

Data availability

The datasets generated and/or analysed during the current study are available in the link provided: https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset?select=Dataset_BUSI_with_GT.

Received: 10 February 2026; Accepted: 28 April 2026
Published online: 02 May 2026

References

1. Breast Cancer Statistics, World Health Organization. Accessed 14 March 2024; https://www.who.int/news-room/fact-sheets/detail/breast-cancer (2024)

2. Crook, T. et al. Accurate screening for early-stage breast cancer by detection and profiling of circulating tumor cells. *Cancers* **14** (14), 3341 (2022).
3. Wengert, G. J., Helbich, T. H., Kapetas, P., Baltzer, P. A. & Pinker, K. Density and tailored breast cancer screening: Practice and prediction—An overview. *Acta Radiol. Open.* **7** (9), 2058460118791212 (2018).
4. Lee, C. I., Chen, L. E. & Elmore, J. G. Risk-based breast cancer screening. *Med. Clin. North. Am.* **101** (4), 725–741. https://doi.org/10.1016/j.mcna.2017.03.005 (2017).
5. Gabriel, D. & Peart, O. Ultrasound-guided breast procedures. *J. Radiol. Nurs. doi*. https://doi.org/10.1016/j.jradnu.2023.02.007 (2023).
6. Ilesanmi, A. E., Chaumrattanakul, U. & Makhanov, S. S. Methods for the segmentation and classification of breast ultrasound images: A review. *J. Ultras*. **24**, 1–16 (2021).
7. Carrilero-Mardones, M., Parras-Jurado, M., Nogales, A., Pérez-Martín, J., & Díez, F. J. Deep learning for describing breast ultrasound images with BI-RADS terms. *J. Imaging Inform. Med.* **37**(6), 2940–2954 (2024)
8. Huang, Q. & Ye, L. Multi-task/single-task joint learning of ultrasound BI-RADS features. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*. **69** (2), 691–701 (2021).
9. Gallas, A. E. et al. An overview of invasive ductal carcinoma (IDC) in women's breast cancer. Current molecular medicine (2025).
10. Sun, S., Fu, C., Xu, S., Wen, Y. & Ma, T. GLFNet: Global-local fusion network for the segmentation in ultrasound images, Comput. Biol. Med. 108103. (2024).
11. Mishra, A. K., Roy, P., Bandyopadhyay, S. & Das, S. K. A multi-task learning based approach for efficient breast cancer detection and classification. *Expert Syst.* **39** (9), e13047 (2022).
12. Chen, G. et al. Esknet: An enhanced adaptive selection kernel convolution for ultrasound breast tumors segmentation. *Expert Syst. Appl.* **246**, 123265 (2024).
13. Podda, A. S. et al. Fully-automated deep learning pipeline for segmentation and classification of breast ultrasound images. *J. Comput. Sci.* **63**, 101816 (2022).
14. Wang, J., Lv, P., Wang, H. & Shi, C. SAR-U-Net: Squeeze-and-excitation block and atrous spatial pyramid pooling based residual U-Net for automatic liver segmentation in computed tomography, Comput. *Comput. Methods Programs Biomed.* **208**, 106268 (2021).
15. Tong, Y., Liu, Y., Zhao, M., Meng, L. & Zhang, J. Improved U-net MALF model for lesion segmentation in breast ultrasound images. *Biomed. Signal Process. Control.* **68**, 102721. https://doi.org/10.1016/j.bspc.2021.102721 (2021).
16. Long, Z. et al. Segmentation and classification of knee joint ultrasonic image via deep learning. *Appl. Sot Comput.*, **97**, 106765, https://doi.org/10.1016/j.asoc.2020.106765 (2020).
17. Manzari, O. N., Ahmadabadi, H., Kashiani, H., Shokouhi, S. B. & Ayatollahi, A. MedViT: a robust vision transformer for generalized medical image classification. *Comput. Biol. Med.* **157**, 106791 (2023).
18. Qi, W., Wu, H. C. & Chan, S. C. MDF-Net: A multi-scale dynamic fusion network for breast tumor segmentation of ultrasound images. *IEEE Trans. Image Process.* **32**, 4842–4855. https://doi.org/10.1109/TIP.2023.3304518 (2023).
19. Sobhaninia, Z. et al. Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning. In *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany: IEEE, Jul. 2019, pp. 6545–6548. https://doi.org/10.1109/EMBC.2019.8856981 (2019).
20. Xue, C. et al. Global guidance network for breast lesion segmentation in ultrasound images. *Med. Image Anal.* **70**, 101989 (2021).
21. Lei, Y. et al. Breast tumor segmentation in 3D automatic breast ultrasound using mask scoring R-CNN. *Med. Phys.* **48** (1), 204–214. https://doi.org/10.1002/mp.14569 (2021).
22. Karthikeyan, N., Maheswari, K., Umesh, H., Rajkumar, N. & Viji, C. Emotion Recognition with a Hybrid VGG-ResNet Deep Learning Model: A Novel Approach for Robust Emotion Classification. *Salud, Ciencia y Tecnología-Serie de Conferencias*, *3*, 960. (2024).
23. Monjezi, E., Akbarizadeh, G. & Ansari-Asl, K. RI-ViT: A multi-scale hybrid method based on vision transformer for breast cancer detection in histopathological images. *IEEE Access* **12**, 186074-186086. https://doi.org/10.1109/ACCESS.2024.3514322 (2024).
24. Borgalli, M. R. A. & Surve, S. Deep learning for facial emotion recognition using custom CNN architecture. *J. Phys. Conf. Ser.*. **2236** (3), 012004 (2022).
25. Raji, I. D. et al. Simple deterministic selection-based genetic algorithm for hyperparameter tuning of machine learning models. *Appl. Sci.* **12** (3), 1186 (2022).
26. Patrick, M. K., Adekoya, A. F., Mighty, A. A. & Edward, B. Y. Capsule networks—A survey. *J. King Saud Univ. Comput. Inform. Sci.* **34** (1), 1295–1310 (2022).
27. Maeedi, A. A., Hammood, D. A. & Hasan, S. M. Image classification using a hybrid BiLSTM-CNN for breast cancer. *AIP Conf. Proc.*. **3232** (1), 020040. (2024).
28. Abhisheka, B., Biswas, S. K. & Purkayastha, B. Ohabm-net: an enhanced attention-driven hybrid network for improved breast mass detection. *Neural Comput. Appl.* **37** (3), 1673–1691 (2025).
29. Iniyan, S. et al. Enhanced breast cancer diagnosis through integration of computer vision with fusion based joint transfer learning using multi modality medical images. *Sci. Rep.* **14** (1), 28376 (2024).
30. Robin, M., John, J. & Ravikumar, A. Breast tumor segmentation using U-NET. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1164–1167). IEEE. (2021), April.
31. Patheda, V. R., Laxmisai, G., Bv, G., Ibrahim, S. S. & Kumar, S. S. A robust hybrid CNN + ViT framework for breast cancer classification using mammogram images. *IEEE Access*. (2025).
32. Al-Dhabyani, W., Gomaa, M., Khaled, H. & Fahmy, A. Dataset of breast ultrasound images. *Data Brief.* **28**, 104863 (2020).
33. Kathirisetty, N. et al. Cutting-Edge Deep Learning Models for Enhanced Breast Can-cer Detection: Comparative Analysis of YOLOv5, YOLOv8 and YOLOv9 Models. *Front. Health Inform.* **13**(8), 527–551 (2024).
34. Althabeti, R. A., Ebied, E. A., Lala, H. M. S. & Eldahshan, K. A. Deep-learning-based medical image quality amelioration. *J. Supercomputing*. **81** (15), 1413 (2025).
35. Kishanthan, S. & Hevapathige, A. AxelSMOTE: An interaction-based oversampling algorithm inspired by cultural dissemination. *Eng. Res. Express*. **7** (4), 045273 (2025).
36. Yang, X., Wang, Y. & Sui, L. NMTNet: A Multi-task Deep Learning Network for Joint Segmentation and Classification of Breast Tumors. *J. Digit. Imaging. Inform. med.* **38**, 3548–3567. https://doi.org/10.1007/s10278-025-01440-7 (2025).
37. Oktay, O. et al. Attention U-net: Learning where to look for the pancreas. Preprint at https://doi.org/10.48550/ARXIV.1804.03999 (2018.)
38. Hamza, A. & Mezl, M. Deep learning-enhanced ultrasound analysis: Classifying breast tumors using segmentation and feature extraction. *IEEE Access.* **13**, 83528 (2025).
39. Zheng, S., Li, J., Qiao, L. & Gao, X. Multi-task interaction learning for accurate segmentation and classification of breast tumors in ultrasound images. *Phys. Med. Biol.* **70** (6), 065006 (2025).
40. He, Q., Yang, Q., Su, H. & Wang, Y. Multi-task learning for segmentation and classification of breast tumors from ultrasound images. *Comput. Biol. Med.* **173**, 108319 (2024).
41. Abdel-Wahab, A. et al. Deep learning framework for breast cancer detection and segmentation using EfficientNet and U-Net with hyperparameter optimization. *SN Comput. Sci.* **6** (5), 528 (2025).

Author contributions

Maddhi Anitha: Conceptualization; writing—original draft; Software. Ch.Rajendra Prasad, original draft; Project administration, Resources, Methodology. Joseph Bamidele Awotunde, Project administration, Resources; writing, review and editing.Oluwasegun Julius Aroba and Michael Rudolph, Acquisition APC, Review and editing, project supervision. Abidemi Emmanuel Adeniyi, Data caution, Software; writing—review and editing.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.E.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026