Breast Cancer Classification and Segmentation Using Deep Learning on Ultrasound Images

Doha Saad Dajam[1], Ayman Qahmash[2]
Department of Data Science-College of Computer Science, King Khalid University, Abha, Saudi Arabia[1]
Department of Computer Science-College of Computer Science, King Khalid University, Abha, Saudi Arabia[2]

***Abstract*—Breast cancer continues to pose a major health challenge for women worldwide, highlighting the critical role of accurate and early detection methods in improving patient outcomes. Ultrasound imaging, a commonly used and non-invasive method, is especially useful for identifying tissue irregularities in younger women or individuals with dense breast tissue. However, accurate interpretation of ultrasound images is challenging due to variability in human analysis and limitations in existing deep learning models, which often struggle with small, imbalanced datasets and lack generalizability compared to models trained on natural images. To tackle these challenges, we introduce a dual deep learning framework that combines image classification and tumor segmentation using breast ultrasound images. The classification component evaluates four models (Custom CNN, VGG16, InceptionV3, and MobileNet) while the segmentation module employs a MobileNet-optimized U-Net architecture for precise boundary localization. We validate our approach using the publicly available BUSI dataset, achieving a 98% classification accuracy with MobileNet and a Dice coefficient of 0.8959 for segmentation, indicating high model reliability and spatial agreement. Our method demonstrates a robust, efficient solution to automate breast cancer detection and localization, with potential to support radiologists in early and accurate diagnosis.**

***Keywords*—Breast cancer; Convolutional Neural Networks (CNNs); tumor segmentation; MobileNet; dice coefficient; BUSI Dataset**

I. Introduction

Breast cancer stands as the most prevalent cancer and the leading contributor to cancer-related mortality among women around the world, posing a major threat to their health and quality of life [1]. Early detection of the disease is key to improving the effectiveness of treatment, reducing mortality rates, and enhancing the quality of life of patients [2]. Though traditional diagnostic techniques such as mammography and MRI are valuable, they also have their own drawbacks like the risk of false positives, invasive biopsy, and patient discomfort during the procedure. It is under these circumstances that ultrasound imaging has arrived as a helpful, non-surgical alternative, particularly for women younger than 40 years and women with dense breasts, since it has the ability to distinguish between fluid-filled cysts and solid tumors [3].

Recent advancements in artificial intelligence (AI) and deep learning have commenced the transformation of medical image analysis, providing novel solutions to surpass the constraints of human interpretation. Particularly in automated image identification applications, deep learning (DL) techniques, particularly convolutional neural networks (CNNs), have shown impressive results. These techniques can learn complex feature representations directly from imaging data, enabling high accuracy in detecting and classifying abnormalities in medical images. For breast cancer imaging, researchers have developed AI systems that approach expert radiologist-level performance in identifying malignancies on both mammograms and ultrasound scans. Compared to traditional computer-aided diagnosis using hand-crafted features, CNN-based approaches automatically extract optimal features and have proven more robust across varying image qualities. By leveraging large datasets and powerful GPUs, deep learning models can be trained to recognize subtle patterns indicative of cancer that might elude the human eye [4]. This has facilitated the development of automated diagnostic algorithms for breast ultrasound, aimed at improving accuracy, consistency, and efficiency in radiological practice. Furthermore, modern DL techniques like transfer learning (pretraining on massive general-image datasets and fine-tuning on medical images) help mitigate data scarcity issues, and ensemble models combine multiple network predictions to boost performance. Segmentation networks have also advanced, enabling precise localization of tumors within images. These developments collectively enhance the capabilities of breast imaging diagnostics beyond what conventional methods can achieve.

Despite the promise of deep learning in medical imaging, significant challenges remain. One major hurdle is the limited size of many curated medical image datasets. Acquiring and labeling medical images is time-consuming, costly, and often constrained by privacy concerns. In breast ultrasound imaging, publicly available datasets have only hundreds of images, far smaller than the thousands or millions of images often used to train robust CNNs in general computer vision. Training deep networks on such small datasets risks overfitting, where the model learns spurious details specific to the training set rather than general patterns of disease. This may cause a poor performance on new patients or images from different hospitals. Additionally, medical images can vary widely in quality and characteristics: ultrasound scans, for example, differ based on the machine manufacturer, technician technique, and patient body habitus. A model that performs well on one clinic's ultrasound data might not generalize to another's if these differences are not accounted for. This domain shift and limited diversity in training data make generalization a core challenge. Traditional transfer learning from natural image datasets only partially addresses this, since features learned from photographs may not capture the nuances

of ultrasound textures or artifacts. Researchers have begun exploring strategies like multi-stage transfer learning – first pretraining on a similar medical imaging task before fine-tuning on the target task – and data augmentation techniques to expand dataset variability. Ensuring that deep learning models are robust, generalizable, and not overly sensitive to training data peculiarities is an active area of research [5].

Our research provides important contributions by presenting a deep learning-enabled advanced system that is most applicable to the precise segmentation and classification of breast ultrasound images. With the help of extensive preprocessing, better augmentation strategies, hyperparameter tuning, and more recent model architectures, such as custom CNNs and U-Net segmentation models, our solution aims to enhance diagnostic accuracy, reduce false positive and negative rates, and assist radiologists in making correct and timely decisions. Furthermore, the paper offers a comparative study of different deep learning techniques and determines the optimal approaches to create an efficient, generalized, and powerful diagnostic tool, thereby helping to fight one of the world's largest killers of women.

The following sections are outlined as follows: Section II reviews related work on deep learning techniques for ultrasound image classification and segmentation. Section III outlines the proposed method using CNN and U-Net models. Section IV presents the results, evaluating model performance and discusses the results obtained. Section V concludes with key takeaways and future work suggestions.

II. Related Work

In this section, the literature on the application of DL methods in breast cancer identification using ultrasound images is reviewed. Transfer learning, ensemble methods, and segmentation models are just a few of the techniques that have been adopted to increase classification and segmentation accuracy. While progress has been made, issues regarding dataset limitations and model generalization still need to be addressed, demonstrating an area that requires continued exploration.

Hijab et al. developed a deep learning method to classify ultrasound images of malignant breast tumours using transfer learning [6]. They adopted a strategy that encompassed training a deep CNN on a dataset of 1,000 ultrasound images (500 benign and 500 malignant cases). They investigated three models: a baseline CNN model trained from scratch, VGG16-based transfer learning model, and a fine-tuned VGG16 model. As indicated by the results presented, the fine-tuned model achieved the best performance (0.97), followed by the transfer learning model achieving a performance of 0.94 and, finally, a performance value of 0.82 in the baseline model. It demonstrated the effectiveness of fine-tuning pre-trained models on medical imaging datasets to improve classification accuracy and overcome issues associated with limited training data and overfitting.

Ayana et al. proposed a multistage transfer learning (MSTL) method for classifying breast cancer in ultrasound images, leveraging both natural and medical image datasets [7]. It follows the steps using an ImageNet pre-trained model, then transfer-learn on the cancer cell microscopic images, and then transfer-learn on ultrasound images to classify them as "malignant" or "benign". The method attained a test accuracy of 99% on the "Mendeley" dataset and 98.7% on the "MT-Small-Dataset", representing a significant improvement in classification accuracy. In contrast, the study showed that integration of cancer cell line images as an intermediary step in MSTL was superior to CTL approaches, suggesting that transfer learning based on better deep learning is indeed possible for early breast cancer diagnosis.

Islam et al. introduced an Ensemble Deep CNN (EDCNN) model for detecting and classifying breast cancer using ultrasound images [8]. This model combined features from both MobileNet and Xception architectures, resulting in significant performance gains over various transfer learning models and the Vision Transformer. Moreover, authors utilized U-Net for image segmentation that provided accurate identification and extraction of tumor areas along with Grad-CAM to enhance the transparency of model's decision-making. The EDCNN model outperformed other popular models in the dataset by achieving 87.82% and 85.69% accuracy in the two datasets respectively. This study was proved to be a potential tool for clinical applications, since the advanced deep learning techniques coupled with image segmentation will make it possible for higher diagnostic accuracy and could aid in early detection of breast cancer.

Kim et al. introduced a weakly-supervised deep learning algorithm for diagnosing breast cancer using ultrasound images, with the goal of reducing the effort and potential bias associated with manual region-of-interest (ROI) annotation [9]. The model was trained on a dataset of 1,000 unannotated ultrasound images, evenly split between benign and malignant cases, and was evaluated on both internal and external datasets. The results demonstrated that the diagnostic performance of the weakly-supervised model was on par with fully-supervised approaches, achieving area under the curve (AUC) values between 0.86 and 0.96.

Uysal and Köse carried out research geared to enhance breast cancer detection through ultrasound images and deep learning-based classification models [10]. Using a dataset of 780 ultrasound images that was divided into training and validation sets, they employed three models: VGG16, ResNet50, and ResNeXt50. The dataset consisted of benign, malignant and normal classes. The images were preprocessed and augmented with center crop, normalization, and random data augmentation. ResNeXt50 has the highest obtained accuracy of 85.83% among cases tested. This study also reflected the power of artificial intelligence, especially deep learning in automating the diagnostic process in medicine, overcoming the subjectivity that is a hallmark of the human decision-making process and accelerating the time it takes to analyze and diagnose a sample.

Wei et al. introduced a multi-feature fusion multi-task network that tackles classification and segmentation simultaneously on breast ultrasound images [11]. Their framework, enhanced with attention modules to better exploit shared features, was tested on the BUSI dataset and a large ultrasound video dataset. It achieved around 95% accuracy on

BUSI and significantly improved segmentation quality, and about 87% accuracy on a more challenging external ultrasound video set (MIBUS). This demonstrates that carefully designing multi-task architectures can yield high performance on both tasks, addressing the limitations of models that excel only in either classification or localization.

Aumente-Maestro et al. similarly developed an end-to-end multi-task *CNN* for concurrent tumor segmentation and classification [12]. A key contribution of their work was an in-depth curation of the BUSI dataset – removing duplicated or inconsistent images – to create a cleaner training set of 450 ultrasound images spanning benign, malignant, and normal classes. Using this refined dataset, their joint model yielded approximately 15% higher Dice and accuracy than training separate models, ultimately reaching about 79–80% classification accuracy and markedly improved mask quality (Dice ≈0.75). These results underscore the benefit of multi-task learning, as the shared representations improved both the identification of tumor presence and the delineation of tumor boundaries.

Madhu et al. took a two-step approach by first segmenting and then classifying tumors [13]. They presented UCapsNet, which combines an enhanced U-Net for tumor segmentation with a Capsule Network for classifying the segmented tumor region. Evaluated on the BUSI dataset, this method achieved near-perfect results – after segmenting the lesion, the capsule-based classifier attained 99.22% accuracy in distinguishing malignant from benign tumors (with 99.52% sensitivity). The extremely high performance suggests that precise segmentation coupled with an advanced classifier that preserves spatial feature hierarchies can dramatically improve diagnostic accuracy, albeit on a relatively small dataset.

Shilaskar et al. focused on a straightforward but effective pipeline using separate models for each task [14]. They employed VGG-16 for classifying ultrasound images and U-Net for segmenting tumors within those images. Using the standard BUSI dataset of 780 images (with ground-truth masks), their system reached 90% classification accuracy and about 98% segmentation accuracy in detecting tumor regions. This dual-model approach illustrates a practical way to integrate classification and segmentation: the CNN provides a probability of malignancy while the U-Net yields the tumor contour, together providing a more comprehensive output to assist radiologists.

Table I summarizes and compares the related works mentioned previously, showing the models they used, the dataset, and their accuracies.

TABLE I. COMPARISON OF RELATED WORK

Authors	Title	Year	Model	Dataset	Accuracy
Hijab et al [6]	"Breast Cancer Classification in Ultrasound Images using Transfer Learning"	2019	CNN Pre-trained VGG16 Fine-tuned VGG16	1300 ultrasound images, augmented to 21,600.	CNN: 79%
Ayana et al [7]	"A Novel Multistage Transfer Learning for Ultrasound Breast Cancer Image Classification"	2022	MSTL: EfficientNetB2, InceptionV3, and ResNet50.	Cancer cell (20,400), Mendeley (200), MT-Small (400).	Mendeley: 99% MT-Small: 98.7%.
Islam et al. [8]	"Enhancing breast cancer segmentation and classification: An Ensemble Deep Convolutional Neural Network and U-net approach on ultrasound images"	2024	EDCNN	Dataset 1 (BUSI): 780 ultrasound images (normal, benign, malignant) Dataset 2 (UDAIT): 163 ultrasound images (110 benign, 53 malignant)	Dataset 1: 87.82% Dataset 2: 85.69%
Kim et al [9]	"Weakly-supervised deep learning for ultrasound diagnosis of breast cancer"	2021	VGG16, ResNet34 GoogLeNet.	1400 ultrasound images from 971 patients.	Not specified
Uysal and Köse [10]	"Classification of Breast Cancer Ultrasound Images with Deep Learning-Based Models"	2022	ResNet50 ResNeXt50 VGG16	780 ultrasound images (benign, malignant, normal) from 600 patients (Kaggle, 400×400 px).	ResNet50: 85.4% ResNeXt50: 85.83% VGG16: 81.11%
Wei et al. [11]	"A Novel Deep Learning Model for Breast Tumor Ultrasound Image Classification with Lesion Region Perception"	2024	MFFMT (ResNet18 & ResNet50 backbones; multi-task)	BUSI: 780 images (benign, malignant, normal); MIBUS: 25,272 frames from 188 videos (benign vs malignant)	BUSI: ~95%; MIBUS: ~87%
Aumente-Maestro et al. [12]	"A multi-task framework for breast cancer segmentation and classification in ultrasound imaging"	2025	Multi-task CNN (UNet++ or nnU-Net backbone)	BUSI: 780 images (3 classes), curated to 450 images (duplicate removed)	≈80%
Madhu et al. [13]	"UCapsNet: A Two-Stage DL Model Using U-Net and Capsule Network for Breast Cancer Segmentation and Classification in US Imaging"	2024	UCapsNet (U-Net + Capsule Network)	BUSI: 780 ultrasound images (with tumor masks)	99.22%
Shilaskar et al. [14]	"Classification and Segmentation of Breast Tumor Ultrasound Images using VGG-16 and U-Net"	2025	VGG16 + U-Net (dual-model pipeline)	BUSI: 780 images (normal, benign, malignant)	90%

Deep-learning studies on breast-ultrasound still tend to excel at one task while overlooking others. Hijab et al. fine-tuned VGG16 on 1,300 images and reported 97 % classification accuracy, but the model provided no lesion outlines, limiting clinical usefulness [6]. Ayana et al. pushed transfer learning further with a multistage strategy: after a second pre-training step on cancer-cell microscopy, their ResNet50 reached 99 % accuracy on the Mendeley set and 98.7

% on MT-Small again for classification alone [7]. More recently, Islam et al. combined MobileNet and Xception (EDCNN) yet achieved only ≈88 % accuracy on BUSI and ≈86 % on UDAIT, while Uysal & Köse's experiments with VGG16/ResNet derivatives topped out at ≈86 % [8] [10]. Segmentation has been even less explored: most papers either omit quantitative mask metrics or rely on separate U-Net pipelines. An exception is Kim et al., who introduced weakly-supervised CNNs that dispense with ROI annotation; their networks attained AUC 0.92–0.96 internally and 0.86–0.90 externally, and localized virtually all malignant masses, but still treated classification and localization as loosely coupled outputs [9].

More recent approaches have begun integrating both tasks. Wei et al. proposed a multi-feature fusion multi-task (MFFMT) network that achieved ≈95% accuracy on the BUSI dataset and boosted segmentation Dice scores significantly compared to single-task models [11]. Aumente-Maestro et al. developed a curated BUSI subset and used a multi-task CNN to jointly segment and classify, improving performance on both fronts and achieving ≈80% classification accuracy with Dice ≈0.75 [12]. Madhu et al. introduced a two-stage UCapsNet model, segmenting with U-Net and classifying with a Capsule Network, resulting in a remarkably high 99.22% classification accuracy [13]. Finally, Shilaskar et al. adopted a dual-model pipeline with VGG16 for classification and U-Net for segmentation, attaining 90% and 98% accuracy respectively, showing that even a modular approach can provide comprehensive outputs [14].

Building on these insights, our paper will integrate both diagnosis and delineation in a single lightweight pipeline. A MobileNet-based classifier will share features with a streamlined U-Net decoder, allowing real-time prediction of class probabilities *and* pixel-accurate tumor contours. By favouring depth-wise separable convolutions over heavyweight backbones, the system will run on mid-tier GPUs or edge devices, overcoming the deployment barriers faced by VGG16- or ResNet101-centric solutions. In addition, we will validate on the public BUSI set and follow Kim et al.'s lead in limiting manual annotations, thereby ensuring transparency, reproducibility, and less curation overhead.

Through this integrated, resource-efficient design we aim to supply radiologists with both a diagnostic label and a precise lesion contour in real time, thereby bridging the gap between algorithmic performance reported in prior studies and the practical demands of everyday clinical workflows.

III. METHODOLOGY

Our proposed approach implements a two-part deep learning system for breast ultrasound analysis: one part focuses on image classification, and the other on tumor segmentation. Fig. 1 presents an overview of the system architecture. In the classification module, we employ a CNN-based model to identify each ultrasound image as malignant tumor, benign tumor, or normal tissue. Rather than relying on a single network, we perform a comparative evaluation of several convolutional neural network architectures to determine the most effective model for this task. In particular, we explore transfer learning with established models (VGG16, MobileNet, and InceptionV3) as well as a custom CNN trained from scratch. By using transfer learning, the models benefit from feature representations learned on large-scale image datasets, which is advantageous given the limited size of medical image data. The classification network takes a preprocessed ultrasound image as input and outputs class probabilities for the three categories, ultimately assigning the image to the class with highest probability via a Softmax layer.

In parallel, the segmentation module is designed to delineate the breast tumor within the ultrasound image. For this purpose, we adopt a U-Net–based architecture owing to its proven effectiveness in biomedical image segmentation. To tailor U-Net for our needs, we integrate a MobileNet encoder as the contracting path of the U-Net. This MobileNet-optimized U-Net uses the efficient MobileNet convolutional blocks to extract high-level features while downsampling the image, and then a symmetrical expanding path (decoder) to produce a binary mask highlighting the tumor region. Skip connections between the encoder and decoder ensure that fine-grained spatial details are preserved in the final segmentation output. The result is a pixel-wise segmentation map where each pixel is classified as either tumor or background tissue. By training this model on ultrasound images with corresponding tumor masks, it learns to accurately localize lesions.

Overall, the methodology can be summarized as a dual pipeline: the input ultrasound image passes through the classification CNN to yield a diagnosis (normal/ benign/malignant) and simultaneously through the U-Net segmentation network to yield a highlighted tumor region (if a tumor is present). These outputs can be combined to provide a radiologist with both a diagnostic prediction and a visual overlay of the tumor contours on the ultrasound image. We implemented the framework using the BUSI dataset for both training and evaluation. Extensive preprocessing (e.g., normalization, data augmentation) was applied to improve model generalizability. Hyperparameters for each model were tuned empirically to optimize performance. In the following, we detail key components of our methodology, including the convolutional building blocks and specific network architectures used for classification (MobileNet, VGG16) and segmentation (U-Net).

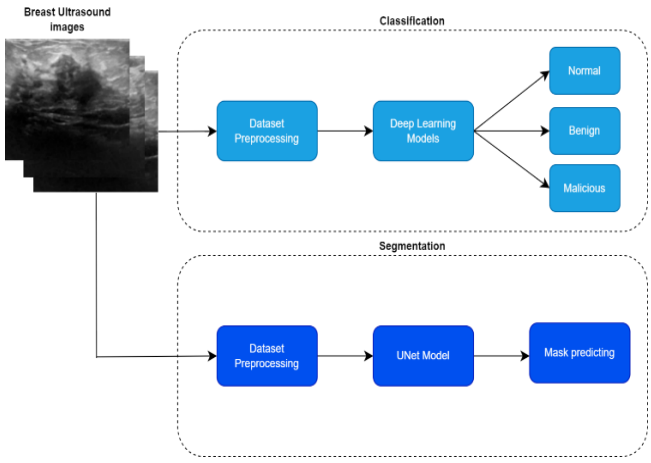

Fig. 1. The proposed structure of the system consists of two parts: classification and segmentation.

A. Classification Models

The classification models are used to classify ultrasound breast images into three categories: "malignant tumor", "benign tumor", and "normal breast". Four different models were used to perform the classification process, and therefore conducted a comprehensive study to determine the best and most appropriate classification model. The four models are Custom CNN, VGG16, InceptionV3, and MobileNet. Fig. 2 represents the classification approach used in the research. The approach is the same for all four classifiers, only the model differs. This approach begins by reading the data, processing it, training the selected model, and finally evaluating the model.

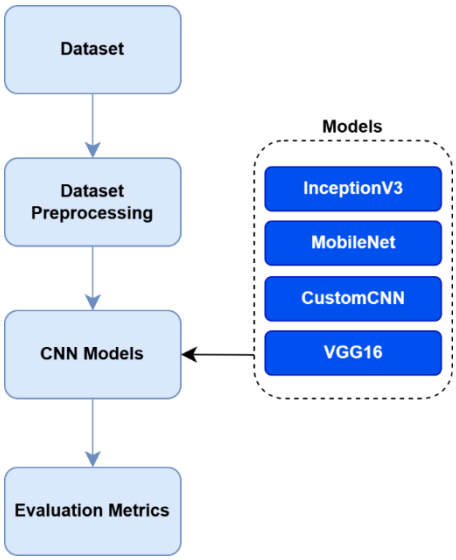

Fig. 2. Our Classification approach.

1) Custom CNN: Every convolutional neural network architecture comprises several key layers that hierarchically process input data. convolutional and pooling layers, the network incorporates:

a) Convolution layer: The convolutional layer (CL) is used to extract local features from the input images using a set of Trainable filters. Each convolution filter has a small spatial extent, like a 3*3 filter, and a depth equal to the number of input channels. As the filter slides across the input's width and height, it computes dot products between the filter weights and the underlying image patch, producing a feature map that highlights the locations of specific features within the input. The quantity of filters in a CL defines the total number of output feature maps (channels) generated by that layer. Enhancing the filter count can improve the model's ability to capture intricate features, though it also escalates computational demands, necessitating a balanced approach based on the task requirements.

b) Pooling layer: Pooling layers (down-sampling layers) are positioned within CLs to gradually decrease the spatial dimensions of feature maps, maintaining the most essential information. Each pooling operation considers a localized area (e.g., 2×2 window) of the input feature map and computes a single summary statistic for that region. These regions typically do not overlap, so pooling effectively partitions the feature map into disjoint segments. Typical pooling functions include average pooling that determines the region's mean value, and max pooling that selects the peak value inside a region. In the custom CNN model developed for this paper, max pooling is employed. Max pooling selects the highest activation in each region as the representative output, thereby capturing the most salient features and reducing data size for subsequent layers.

c) Fully Connected (Dense) layers: These layers act as the classifier component, transforming extracted features into class predictions. Each neuron connects to all activations from the previous layer, enabling high-level feature integration.

d) Activation functions: Non-linear transformations are critical for learning complex patterns. We employ:

- ReLU (Rectified Linear Unit): Applied after each convolutional and dense layer (except output), defined as $f(x) = max(0, x)$. This activation provides computational efficiency while alleviating vanishing gradients.
- SoftMax: The final layer utilizes SoftMax activation produce normalized class probabilities for our three tumor categories (normal, benign, malignant).

$$\text{Softmax}(x) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}} \quad (1)$$

Our custom CNN architecture employs three convolutional blocks (32-64-128 filters) with max pooling for hierarchical feature extraction from ultrasound images. The network transitions to dense layers 512→256 units with dropout regularization before final SoftMax classification.

2) MobileNet architecture: MobileNet is a compact CNN architecture optimized for efficient performance on devices with constrained computational power [15]. The hallmark of MobileNet is its use of depthwise separable convolutions as the primary building block. A depthwise separable convolution is a factorized form of the standard convolution that drastically reduces the number of parameters and multiplications required. It breaks the convolution into two stages: first, a depthwise convolution where a single filter is applied independently to each input channel (slice) of the feature map, and second, a pointwise convolution (1×1 convolution) that combines the outputs of the depthwise step across channels. In a traditional convolution layer, if we have N input channels and M output channels with a k×k filter, we would use k×k×N×M parameters. In contrast, a depthwise separable convolution uses only k×k×N parameters for the depthwise stage plus 1×1×N×M for the pointwise stage, leading to a significant reduction in total computations. In fact, this factorization can reduce the computational cost by about 8 to 9 times compared to a standard convolution of equivalent dimensions, while preserving a large portion of the representational power. This efficiency makes MobileNet particularly attractive for tasks like ours, where we aim to deploy complex models without incurring prohibitive computation.

Despite its light weight, MobileNet maintains strong performance through its clever design. The architecture consists of a sequence of layers that intermix depthwise separable convolution blocks with additional operations such as batch normalization and non-linear activations. In MobileNet's original formulation (often referred to as MobileNet V1), the network begins with a single ordinary convolution layer, and thereafter every convolution is depthwise separable. Each such block typically includes: a depthwise convolution (per-channel spatial filtering), a Batch Normalization layer (to stabilize learning by normalizing activations), a nonlinear activation (ReLU), then a 1×1 pointwise convolution to integrate features, followed again by BatchNorm and ReLU. By repeating these blocks with varying numbers of filters, MobileNet builds up a deep network. An occasional stride-2 convolution is used in some blocks to perform downsampling (instead of using separate pooling layers), reducing feature map size while increasing depth. Overall, the baseline MobileNet architecture comprises 28 layers when counting depthwise and pointwise convolutions separately. After the convolutional feature extraction layers, MobileNet includes an average pooling layer that aggregates the spatial information (producing a 1×1 representation per channel), followed by a final fully connected (dense) layer or a 1×1 convolution that produces the class scores, and a closing Softmax activation to output class probabilities. In our implementation for breast image classification, we initialize MobileNet with weights pre-trained on ImageNet (to leverage learned general features) and then fine-tune it on the ultrasound dataset. MobileNet's efficiency does not come at the cost of accuracy in our experiments – in fact, its performance was superior to the heavier models for this task (as discussed later). The combination of computational thrift and discriminative power makes MobileNet well-suited as the core of the classification module in our dual framework.

3) Visual Geometry Group 16 (VGG16): VGG16 is a deep CNN architecture that is widely recognized for its simple and uniform design, which has made it a common benchmark in image classification research [16]. The name "VGG16" refers to the model developed by the "Visual Geometry Group" (VGG) at Oxford, with 16 layers of weights (13 convolutional layers and 3 fully-connected layers). VGG16 was originally introduced for the "ImageNet Large Scale Visual Recognition Challenge" and demonstrated that a deep network with small filters could achieve excellent accuracy. Although it is a relatively large model in terms of parameters, its straightforward architecture provides a useful comparison for more modern networks. The input to VGG16 is typically a fixed-size image of 224 × 224 pixels (with 3 color channels), so we resize our grayscale ultrasound images accordingly by duplicating the single channel or adapting the first layer to single-channel input. The core of VGG16 is organized into five convolutional blocks. Each block consists of multiple convolutional layers using very small 3 × 3 kernels (with stride 1 and same-padding so that spatial dimensions are preserved) followed by a 2×2 max pooling layer that halves the spatial resolution. For example, the first block might have two conv layers of 64 filters each, then a max pool; the next block conv layers of 128 filters, then pool; and so on, typically doubling the number of filters after each pooling. This design means that as we go deeper, feature maps become smaller in spatial size but richer in depth (channels), enabling the network to learn hierarchical features at multiple scales. Using 3×3 filters throughout (instead of larger kernels) was a key design choice: stacking two 3×3 conv layers has an effective receptive field of 5×5 but with fewer parameters and more non-linearities than a single 5×5 layer, which improves learning. The repeated pattern of conv → conv → pool in VGG16 yields a very uniform architecture that is easier to implement and tune.

After the final convolutional block, VGG16 transitions to the classification head of the network. The feature maps output by the conv stack are flattened into a single vector (or alternatively, global average pooling could be used, but in the standard VGG16 they do a flatten). This is followed by three fully-connected layers. The first two dense layers in VGG16 each have 4096 neurons, which are quite large and contribute significantly to the parameter count of the model. These act as high-level feature combiners, where the network can learn complex non-linear combinations of the convolutional features. After these two layers, a smaller fully-connected layer produces the final outputs. In the original ImageNet model, this third dense layer has 1000 units (one for each class in ImageNet), but in our case we adjust it to have 3 output units corresponding to the classes (normal, benign, malignant). Each fully-connected layer is followed by a ReLU activation function, and the first two have dropout regularization in the original architecture to prevent overfitting. The network concludes with a Softmax layer (built into the last dense layer in many implementations) that outputs class probabilities summing to 1. Throughout the network – from convolutional layers to the dense layers – ReLU activations are used, introducing the non-linear capabilities needed for the network to learn complicated patterns. In our use of VGG16 via transfer learning, we leverage the pre-trained convolutional layers as a fixed feature extractor or fine-tune them on the ultrasound dataset (experimenting with both strategies). The appeal of VGG16 in our study is its proven performance and simplicity: it often serves as a baseline model, and by comparing it to newer architectures like MobileNet, we can quantify the improvements gained by modern designs. While VGG16 is computationally heavier, it provides a useful reference for how a conventional deep CNN performs on breast ultrasound classification. The insights from VGG16's performance also guided some of our model tuning, such as the importance of data augmentation to combat overfitting given the model's large capacity.

4) InceptionV3: InceptionV3 is a CNN architecture designed for high computational efficiency and performance, addressing some limitations of simply making networks deeper [17]. While very deep networks (such as early VGG-style models) achieved impressive results, they often incurred extremely high computational costs and were prone to overfitting, especially when training data was limited. InceptionV3 builds upon the Inception series of architectures

by using clever factorization of convolutional kernels and other techniques to reduce computation while maintaining representational strength. For example, a large convolution (e.g., 5×5) may be factorized into two smaller convolutions (e.g., two 3×3 convolutions or a 1×N followed by N×1 convolution), which lowers computational load. Additionally, InceptionV3 incorporates aggressive regularization methods (such as label smoothing and dropout) to combat overfitting.

The design of InceptionV3 is guided by four key principles that balance network depth and width for optimal efficiency:

- Avoiding representational bottlenecks: The architecture is structured to prevent early layers from severely restricting the information flow (e.g., by not making any layer too narrow in terms of feature maps).
- Processing at higher dimensions when feasible: InceptionV3 maintains relatively high-dimensional feature representations internally, as higher dimensional spaces can make it easier for the network to disentangle complex information (provided the computation is manageable).
- Using low-dimensional embeddings for spatial aggregation: The network employs 1×1 convolutions (bottleneck layers) to reduce dimensionality before expensive operations. These low-dimensional embeddings allow for combining spatial information (e.g., in pooling or in larger convolutions) without significant loss of representational capacity.
- Balancing width and depth: Instead of only increasing the depth (number of layers), InceptionV3 also increases the width (number of parallel paths or filters) of the network in a judicious way. Expanding the network in both directions (width and depth) simultaneously yields better performance for a given computational budget than merely going deeper.

By adhering to these principles, InceptionV3 achieves strong performance on image recognition tasks with a more efficient use of parameters and computations compared to earlier very-deep models.

B. Segmentation

Image segmentation involves dividing a digital image into several segments or regions, each representing a meaningful component of the scene. The primary aim is to simplify or transform the image representation to make it more useful for analysis, which is crucial for locating objects and boundaries (e.g., tumors in medical images). Over the years, a variety of segmentation algorithms have been developed in computer vision, ranging from early classical methods to more advanced techniques. Early approaches include thresholding (separating regions based on intensity thresholds), region growing (iteratively merging pixels or regions that satisfy homogeneity criteria), K-means clustering (grouping pixels into *K* clusters based on feature similarity), and watershed algorithms (treating the image as a topographic surface and finding catchment basins) [18] [19] [20]. More advanced traditional techniques involve active contours (snakes), graph cuts, and sparsity-based methods, each bringing improvements in capturing object shapes or incorporating prior knowledge into the segmentation process [19] [21].

In our segmentation module, we utilize a U-Net to learn the mapping from ultrasound images to binary masks of tumor vs. background [22]. Given the limited number of training images, U-Net's efficiency and reliance on augmented data are well-suited to our problem. We enhanced the basic U-Net by using a pretrained MobileNet encoder (as mentioned in the methodology overview) to initialize the contracting path with robust feature extractors. The decoder was kept relatively standard, with up-convolution layers and concatenation of encoder features via skip connections. We trained the network using a combination of binary cross-entropy and Dice loss (a common practice to handle class imbalance in segmentation and directly optimize for overlap with ground truth). The output of the segmentation network is a probability map which we threshold to obtain the final binary mask of the tumor region. By leveraging U-Net, our system achieves accurate delineation of breast tumors, effectively separating them from healthy tissue in the ultrasound images. This segmentation is valuable on its own – for example, to estimate tumor size or visualize shape – and it also complements the classification result. The combination of a class prediction with a segmented tumor outline can give radiologists greater confidence in the AI's output, as it provides both an answer and an explanation (highlighting *where* the model sees a tumor). U-Net's proven capability to yield high accuracy on limited data is a major reason it excels in our application, helping to overcome the dataset size challenge and producing reliable segmentations that generalize well to new ultrasound scans [23][24].

1) UNet Components:

a) Encoder: The encoder uses a pre-trained MobileNet backbone to extract hierarchical features from the input image. Instead of traditional pooling layers, it relies on strided convolutions to progressively reduce spatial dimensions while increasing feature depth. The five encoder layers (conv1_relu, conv_pw_3_relu, conv_pw_5_relu, conv_pw_11_relu, conv_pw_13_relu) downscale the image from 256×256 to 8×8, capturing high-level semantic information at different scales.

b) Decoder: The decoder is not fully symmetric to the encoder but follows a U-Net structure. It uses transposed convolutions (Conv2DTranspose) to upsample feature maps, doubling their resolution at each step. Instead of simple skip connections, the decoder concatenates upsampled features with resized encoder outputs (via 1×1 convolutions for channel alignment). This helps recover spatial details while maintaining learned features.

c) Skip connections: At each decoder stage, feature maps from the corresponding encoder layer are resized and concatenated with the upsampled decoder features. These connections bridge the semantic gap between high-resolution encoder features (e.g., conv1_relu at 128×128) and low-resolution decoder features, improving localization accuracy.

d) Final layer: The decoder's last step upsamples to the original input size (256×256) and applies a 1×1 convolution

with sigmoid activation to produce a binary segmentation mask. Unlike traditional U-Net, our model uses 32 filters in the final upsampling before reducing to a single-channel output, balancing detail preservation and computational efficiency. Fig. 3 represents the proposed UNet Model.

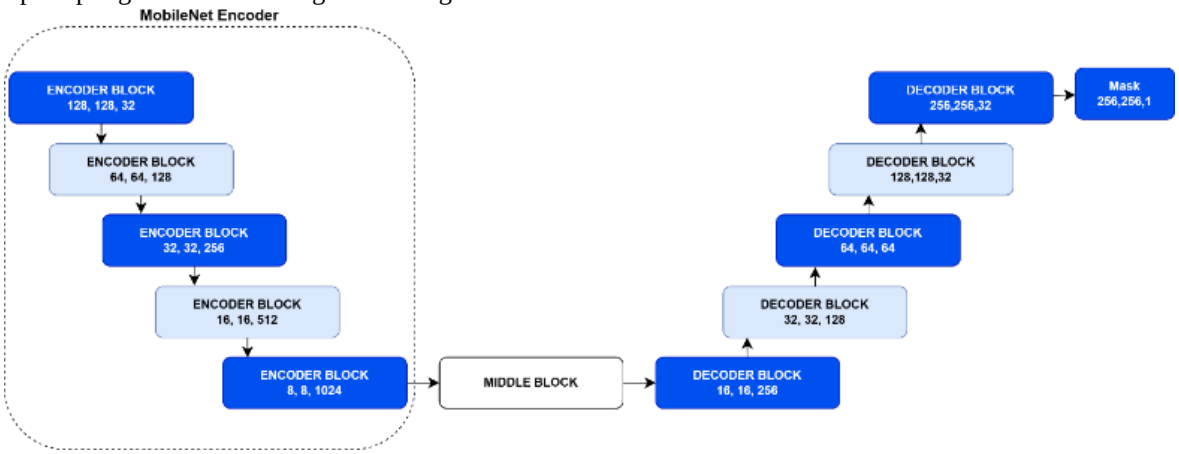

Fig. 3. Our Proposed UNet model.

C. Evaluation Metrics

1) Classification metrics: To evaluate classification performance, we use four standard metrics (precision, recall, F1-score, and accuracy) each with its formal definition. In (2), (3), and (5), TP is the true positive, FP is the false positive, TN is the true negative, and FN is the false negative.

a) Precision: Precision evaluates how accurate the positive predictions are.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

A higher precision indicates that the model has a low false-positive rate, i.e., when it predicts a lesion is malignant (positive), it is often correct.

b) Recall: Recall, also known as sensitivity, assesses a model's ability to correctly identify all actual positive cases.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

A higher recall means the model misses few positive instances (low false-negative rate), correctly detecting most tumors that are present.

c) F1-score: The F1-score—computed as the harmonic mean of precision and recall—provides one balanced measure of how accurately a model predicts the positive class. A high F1-score signals that the model achieves strong precision and recall simultaneously, meaning it identifies positives well while keeping both types of errors low. It is calculated using the formula:

$$F1-\text{score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

d) Accuracy: Accuracy represents the overall correctness of the model and is defined as the proportion of all predictions that are correct. Formally:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

This metric gives the fraction of images (of any class) that are classified correctly. While accuracy is useful, it can be misleading in imbalanced datasets, which is why the above precision, recall, and F1 metrics are also reported for a more complete evaluation.

2) Segmentation metric

a) Dice coefficient: Evaluates pixel-wise agreement between predicted and ground truth masks:

$$Dice = 2 * \frac{[X \cap Y]}{[X]+[Y]} \quad (6)$$

Where X is the predicted mask and Y is the ground truth. Values range from 0 (no overlap) to 1 (perfect match).

IV. Results

This part will introduce and evaluate the effectiveness of the deep learning framework that we proposed for classification and for segmentation of the ultrasound images of breast. The experimental setup evaluates four different classification models—Custom CNN, VGG16, MobileNet, and InceptionV3-and one U-Net-based segmentation model. Performance will be assessed by the multi-evaluation metrics mentioned previously.

A. Dataset

We conducted our experiments using the "Dataset of Breast Ultrasound Images" (BUSI), which is the first breast-ultrasound collection released for open use [25]. The BUSI dataset contains a total of 780 ultrasound images collected from 600 female patients, with ages ranging from 25 to 75 years. Each image is a grayscale breast ultrasound scan (with an average resolution of roughly 500×500 pixels) that has been labeled by expert radiologists into one of three categories:

- Normal: healthy breast tissue with no evident tumors (typically the scans of volunteers or the contralateral healthy breast).
- Benign: presence of a non-cancerous tumor or lesion (e.g. fibroadenomas or cysts).

- Malignant: presence of a cancerous tumor.

Ultrasound imaging is commonly employed for early detection of breast cancer, especially in younger women and individuals with dense breast tissue, as it can distiguish between solid tumors and fluid-filled cysts. The BUSI dataset provides a diverse set of examples for these classes, supporting both the training and evaluation of classification and segmentation models in this study. Fig 4 shows a few representative ultrasound images from the dataset as examples of each class.

The class distribution in the BUSI dataset is somewhat imbalanced, reflecting real-world prevalence in a clinical setting. Out of the 780 images, 133 are normal, 437 are benign, and 210 are malignant. Thus, benign cases form the largest group, which is expected since many screened abnormalities turn out to be benign, while malignant cases are fewer. This imbalance was taken into account during model training and evaluation by using appropriate metrics (like macro-averaged F1-score) and techniques (like class-balanced batch sampling and data augmentation) to ensure the models perform well across all categories. The class distribution of the dataset by category (Benign, Malignant, and Normal) is shown in Fig. 5.

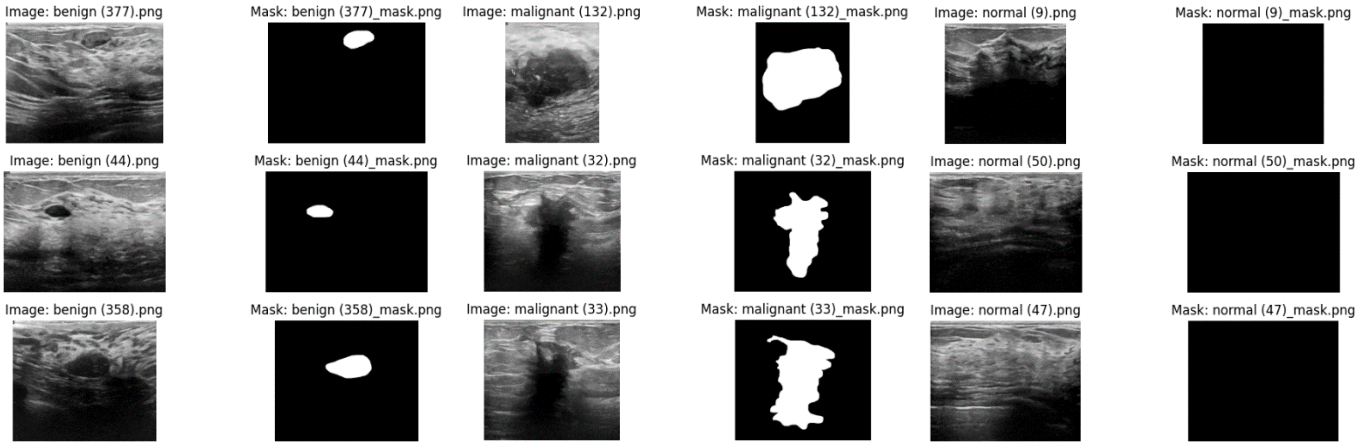

Fig. 4. Random samples from dataset.

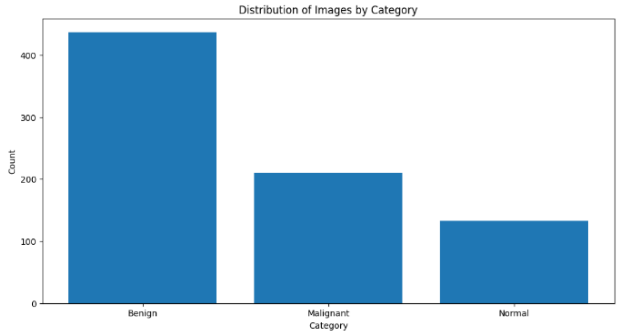

Fig. 5. Class distribution of dataset.

B. Dataset Preprocessing

1) Data normalization: Before feeding the images into the neural network models, we applied data normalization to standardize the input scale. Normalization is the process of rescaling numeric data from different ranges into a common scale, typically between 0 and 1 (or sometimes -1 and 1). This step is important because features (in this case, pixel intensity values) can have vastly different scales, and if left unnormalized, those with larger magnitudes could unduly influence the model's learning process. This ensures that all attributes share a consistent scale. For the normalization process, we apply the equation below, which will generate a new range from 0 to 1.

$$New\ Pixel\ Value = \frac{Old\ Pixel\ Value}{255} \quad (7)$$

2) Data augmentation: To further improve the model's generalizability and address the limited size of the dataset, we employed data augmentation techniques during training [26]. Data augmentation artificially expands the training set by creating modified versions of the original images, thereby providing the model with a more varied set of examples to learn from. In our case, each original ultrasound image was subjected to random transformations to generate new, plausible images. These transformations included small rotations (up to a few degrees), shifts in the horizontal or vertical direction (translating the image by a fraction of its width or height), slight shearing, adjustments to brightness (making the image lighter or darker), zooming in/out, and horizontal flipping. By applying these perturbations, the model is exposed to different scenarios of how a tumor might appear in an ultrasound, which reduces the chance of overfitting to the original training images.

The augmented dataset is both larger and more diverse, which leads to more robust learning. Models trained with augmentation tend to perform better on unseen data because they have learned to handle variations in image orientation, position, scale, illumination, and other conditions. In summary, data augmentation improves the generalization of the deep learning models, ultimately enhancing their accuracy and reliability when deployed on new ultrasound scans. The data augmentation parameters are summarized in Table II.

TABLE II. DATA AUGMENTATION PARAMETERS

Parameter	Value / Range	Description
Rotation Range	5°	Maximum rotation angle in degrees
Width Shift Range	0.1 (10% of width)	Horizontal translation range
Height Shift Range	0.1 (10% of height)	Vertical translation range
Shear Range	0.05	Shear intensity (radians)
Brightness Range	(1, 1.4)	Multiplier range for brightness adjustment
Zoom Range	0.05 (5%)	Range for random zooming
Horizontal Flip	True	Random horizontal flipping enabled
Fill Mode	'nearest'	Strategy for filling in newly created pixels

Fig. 6 shows the distribution of the dataset after applying dataset augmentation.

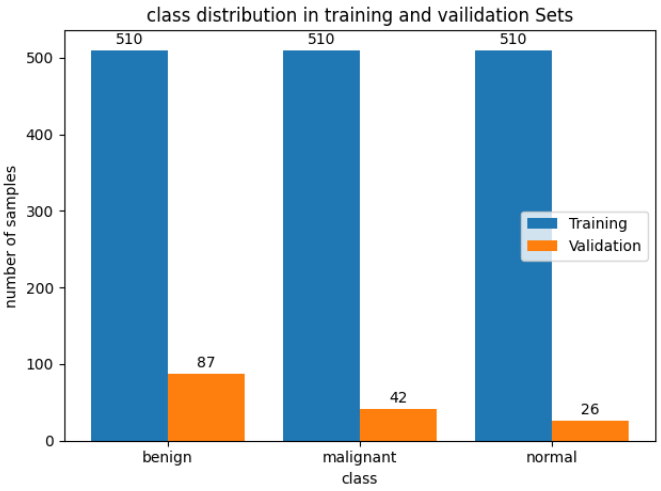

Fig. 6. Class Distribution of training and validation.

C. Classification Results

To identify and categorize breast lesions as benign, malignant, or normal, four classification models were trained using identical preprocessing steps and hyperparameters (50 epochs, batch size of 4, Adam optimizer, and categorical crossentropy loss). Transfer learning was applied to VGG16, InceptionV3, and MobileNet with imagenet weights, while the custom CNN was trained from scratch. The results of the classification task are summarized in Table III.

TABLE III. CLASSIFICATION PERFORMANCE METRICS

Model	Accuracy	Precision (Macro Avg)	Recall (Macro Avg)	F1-score (Macro Avg)
MobileNet	0.98	0.98	0.99	0.98
InceptionV3	0.95	0.93	0.95	0.94
VGG16	0.90	0.94	0.86	0.89
Custom CNN	0.54	0.45	0.37	0.34

MobileNet outperformed all other models, achieving the highest accuracy (98%) along with nearly perfect recall and F1-score across all classes. InceptionV3 followed closely with a 95% accuracy and strong balance between precision and recall. VGG16 showed decent results, particularly for benign and normal classes, but struggled with malignant classification recall. The custom CNN model, trained from scratch, significantly underperformed with an overall accuracy of 54%, highlighting the advantage of using pre-trained models and transfer learning in medical imaging contexts. The classification metrics results are summarized in Fig. 7.

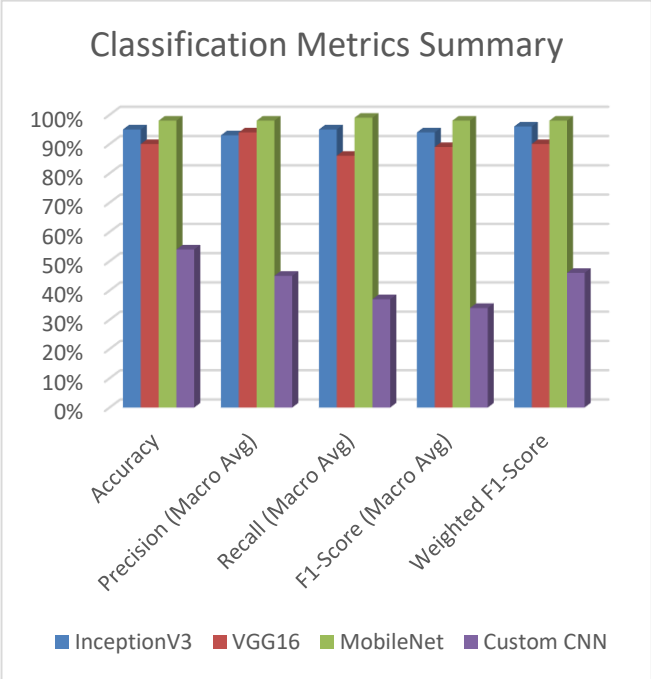

Fig. 7. Summary of classification metrics of four models.

Table IV represents the confusion matrices of the top models (MobileNet and InceptionV3).

TABLE IV. TOP MODELS CONFUSION MATRICES

MobileNet	Pred: Benign	Pred: Malignant	Pred: Normal
Actual: Benign	84	3	0
Actual: Malignant	0	42	0
Actual: Normal	0	0	26
InceptionV3	Pred: Benign	Pred: Malignant	Pred: Normal
Actual: Benign	85	0	2
Actual: Malignant	0	37	5
Actual: Normal	0	0	26

These matrices further emphasize the superiority of MobileNet and InceptionV3 in consistently identifying malignant and benign cases.

D. Segmentation Results

The segmentation module, based on a MobileNet-enhanced U-Net architecture, was evaluated over 50 epochs using a batch size of 8. The loss function combined binary cross-entropy and Dice loss (bce_dice_loss), with the Adam optimizer applied throughout. The segmentation performance is presented in Table V.

TABLE V. SEGMENTATION PERFORMANCE

Metric	Value
Accuracy	0.9648
Dice Coef.	0.8959
Loss	0.6987

The model reached a Dice score of 0.8959, reflecting strong alignment between the generated segmentation masks and the reference annotations. An overall accuracy of 96.48% further emphasizes the model's precision in identifying tumor boundaries. This impressive segmentation capability enhances the reliability of the high classification metrics, demonstrating the robustness of the proposed dual DL framework for breast cancer analysis using ultrasound images Fig. 8 Displays several samples of the results of the images from the dataset that were tested on the proposed model and the samples show good accuracy of the model.

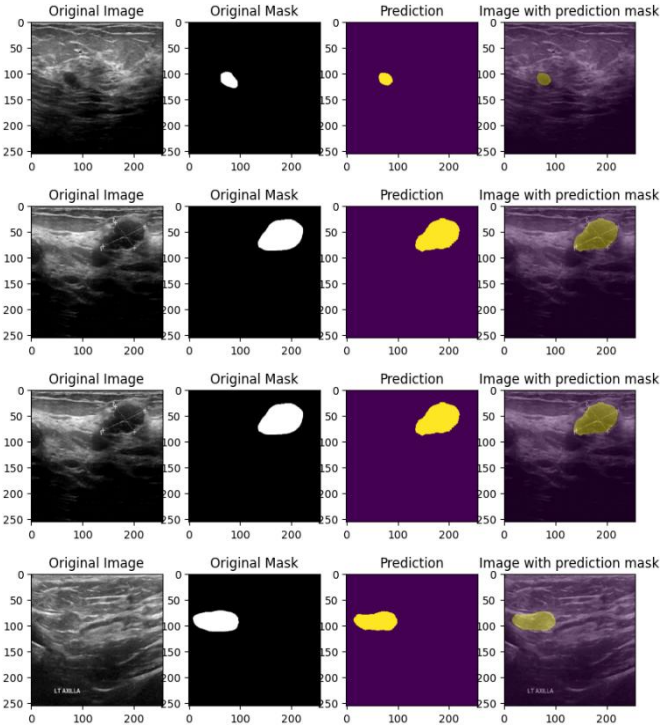

Fig. 8. Several samples were tested on the proposed model and the samples show good accuracy of the model.

E. Discussion Results

The classification and segmentation results of our proposed deep learning framework demonstrate considerable improvements over existing approaches in the literature. Compared to recent studies employing ensemble or modified CNN models for breast cancer detection in ultrasound images, our methodology achieves superior performance in both classification accuracy and segmentation quality. Islam et al. [8] proposed an ensemble of MobileNet and Xception architectures (EDCNN) for classifying breast cancer, reaching an accuracy of 85.69%, an F1-score of 79.39%, precision of 84.00%, and recall of 78.00%. In comparison, our MobileNet model significantly outperformed EDCNN, achieving 98% accuracy, a macro-averaged F1-score of 98%, precision of 98%, and recall of 99%. Similarly, our second-best model, InceptionV3, also surpassed EDCNN, achieving 95% accuracy with a macro F1-score of 94%.

From a segmentation perspective, Islam et al. used a conventional U-Net without detailed segmentation metrics. Our approach, employing a modified U-Net optimized with binary cross-entropy and Dice loss (bce_dice_loss), attained a Dice coefficient of 0.8959 and an overall accuracy of 96.48%. This improvement clearly demonstrates the advantages of optimizing segmentation techniques through carefully selected loss functions and model adjustments.

In the study by Uysal and Köse [10], various CNN architectures including VGG16, ResNet50, and ResNeXt50 were compared for breast cancer classification, with ResNeXt50 achieving the highest accuracy of 85.83%, an F1-score of 87.31%, and AUC of 90%. When benchmarked against these models, our MobileNet architecture outperformed all configurations presented, with accuracy and F1-score improvements exceeding 12 and 10 percentage points, respectively. Table VI provides a detailed comparison of our classification performance relative to these prior studies.

Despite the demonstrated improvements, our research exhibits certain limitations that could guide future work. Firstly, the dataset used, BUSI, is relatively small and imbalanced, potentially limiting the generalizability of our findings. Future studies should consider testing models on larger, more diverse datasets that reflect broader patient demographics and varied imaging conditions. Additionally, our current approach relies significantly on supervised learning, which necessitates substantial manual annotation efforts. Exploring semi-supervised or weakly-supervised learning techniques could further reduce the annotation burden while maintaining or improving model performance.

Another potential limitation is the computational resource requirement. Although our MobileNet-based approach is optimized for lightweight deployment, real-time processing demands could still pose challenges in clinical environments with very limited computational infrastructure. Future research should further investigate model compression techniques, knowledge distillation, or quantization methods to enhance model efficiency and facilitate deployment on lower-resource hardware.

TABLE VI. CLASSIFICATION PERFORMANCE COMPARISON

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
MobileNet (Ours)	98.0	98.0	99.0	98.0
InceptionV3 (Ours)	95.0	93.0	95.0	94.0
VGG16 (Ours)	90.0	94.0	86.0	89.0
Custom CNN (Ours)	54.0	45.0	37.0	34.0
EDCNN (MobileNet + Xception) [8]	85.69	84.0	78.0	79.39
VGG16 [10]	81.11	77.77	70.85	76.90
ResNet50 [10]	85.40	83.20	93.67	87.93
ResNeXt50[10]	85.83	82.92	76.80	87.31

Finally, incorporating multimodal imaging data, such as mammography or MRI, could provide complementary information to further enhance diagnostic accuracy. Investigating fusion methods to integrate multiple imaging modalities represents a promising direction for future research, potentially leading to more robust and clinically applicable diagnostic tools.

V. CONCLUSION

This research paper presents a comprehensive DL framework that integrates image classification and tumor segmentation to enhance breast cancer detection using ultrasound imaging. By leveraging multiple convolutional neural network architectures—including MobileNet, VGG16, InceptionV3, and a custom CNN—for classification, and a MobileNet-optimized U-Net for segmentation, the proposed system demonstrates significant improvements in diagnostic accuracy and spatial localization. Among the evaluated models, MobileNet achieved the highest classification performance with a 98% accuracy and near-perfect precision and recall, while the segmentation module attained a Dice coefficient of 0.8959, indicating strong agreement with ground truth annotations.

The results highlight the effectiveness of combining transfer learning and deep feature extraction in addressing the inherent challenges of medical image analysis, such as limited dataset size and variability in image quality. Furthermore, the use of data normalization and augmentation contributed to enhanced model generalizability, ensuring robustness across diverse imaging conditions.

Ultimately, the dual-function framework developed in this paper offers a reliable, efficient, and interpretable tool that can assist radiologists in the early and accurate diagnosis of breast cancer. By reducing dependence on manual analysis and minimizing diagnostic inconsistencies, the system has the potential to support clinical decision-making and improve patient outcomes. Future work may explore integrating multimodal imaging data and advanced ensemble strategies to further refine diagnostic capabilities and broaden the framework's applicability across diverse clinical settings.

ACKNOWLEDGEMENT

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Research Project under grant number RGP2/455/45.

REFERENCES

[1] R. A. Smith, D. Brooks, V. Cokkinides, D. Saslow and O. W. Brawley, "Cancer screening in the United States, 2013: a review of current American Cancer Society guidelines, current issues in cancer screening, and new guidance on cervical cancer screening and lung cancer screening," CA: a cancer journal for clinicians, vol. 63, no. 2, p. 88–105, 2013.

[2] J. Seely and T. Alhassan, "Screening for Breast Cancer in 2018—What Should We be Doing Today?," Current Oncology, vol. 25, no. s1, pp. 115-124, 2018.

[3] R. Guo, G. Lu, B. Qin and B. Fei, "Ultrasound Imaging Technologies for Breast Cancer Detection and Management: A Review," Ultrasound in Medicine and Biology, vol. 44, no. 1, pp. 37-70, 2018.

[4] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. Li, D. Ni and T. Wang, "Deep Learning in Medical Ultrasound Analysis: A Review," Engineering, vol. 5, p. 261–275, 2019.

[5] E. Güldoğan, H. Ucuzal, Z. Küçükakçalı and C. Çolak, "Transfer Learning-Based Classification of Breast Cancer using Ultrasound Images," Middle Black Sea Journal of Health Science, vol. 7, no. 1, p. 74–80, 2021.

[6] A. Hijab, M. A. Rushdi, M. M. Gomaa and A. Eldeib, "Breast Cancer Classification in Ultrasound Images using Transfer Learning," in 2019 Fifth International Conference on Advances in Biomedical Engineering (ICABME), 2019.

[7] G. Ayana, J. Park, J.-W. Jeong and S.-w. Choe, "A Novel Multistage Transfer Learning for Ultrasound Breast Cancer Image Classification," Diagnostics, vol. 12, no. 135, 2022.

[8] M. R. Islam, M. M. Rahman, M. S. Ali, A. A. N. Nafi, M. S. Alam, T. K. Godder, M. S. Miah and M. K. Islam, "Enhancing breast cancer segmentation and classification: An Ensemble Deep Convolutional Neural Network and U-net approach on ultrasound images," Machine Learning with Applications, vol. 16, p. 100555, 2024.

[9] J. Kim, H. J. Kim, C. Kim, J. H. Lee, K. W. Kim, Y. M. Park, H. W. Kim, S. Y. Ki, Y. M. Kim and W. H. Kim, "Weakly supervised deep learning for ultrasound diagnosis of breast cancer," Scientifc Reports, vol. 11, no. 24382 , 2021.

[10] F. Uysal and M. M. Köse, "Classification of Breast Cancer Ultrasound Images with Deep Learning-Based Models," Engineering Proceedings, vol. 31, no. 8, p. 1–5, 2022.

[11] J. Wei, H. Zhang and J. Xie, "A Novel Deep Learning Model for Breast Tumor Ultrasound Image Classification with Lesion Region Perception," *Current Oncology*, vol. 31, no. 9, pp. 5057-5079, 2024.

[12] C. Aumente-Maestro, J. Díez and B. Remeseiro, "A multi-task framework for breast cancer segmentation and classification in ultrasound imaging," *Computer methods and programs in biomedicine*, vol. 260, 2025.

[13] G. Madhu, A. M. Bonasi, S. Kautish, A. S. Almazyad, A. W. Mohamed, F. Werner, M. Hosseinzadeh and M. Shokouhifar, "UCapsNet: A Two-Stage Deep Learning Model Using U-Net and Capsule Network for Breast Cancer Segmentation and Classification in Ultrasound Imaging," *Cancers (Basel)*, vol. 16, no. 22, p. 3777, 2024.

[14] S. Shilaskar, S. Bhatlawande, M. Talewar, S. Goud, S. Tak, S. Kurian and A. Solanke, "Classification and Segmentation of Breast Tumor Ultrasound Images using VGG-16 and UNet," *Biomedical and Pharmacology Journal* , vol. 18, no. 1, 2025.

[15] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, 2017.

[16] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409.1556, 2014.

[17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, 2016, pp. 2818–2826.

[18] E. H. Houssein, G. M. Mohamed, I. A. Ibrahim and Y. M. Wazery, "An efficient multilevel image thresholding method based on improved heap-based optimizer," *Scientific Reports*, vol. 13, 2023.

[19] Y. Yu, C. Wang, Q. Fu, R. Kou, F. Huang, B. Yang, T. Yang and M. Gao, "Techniques and Challenges of Image Segmentation: A Review," *Electronics*, vol. 12, no. 5, p. 1199, 2023.

[20] S. Basar, M. Ali, G. Ochoa-Ruiz, M. Zareei, A. Waheed and A. Adnan, "Unsupervised color image segmentation: A case of RGB histogram based K-means clustering initialization," *Plos One*, 2020.

[21] S. Jardim, J. António and C. Mora, "Graphical Image Region Extraction with K-Means Clustering and Watershed," *J. Imaging*, vol. 8, no. 6, p. 163, 2022.

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Proc. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), Munich, Germany, 2015, pp. 234–241.

[23] N. Siddique, P. Sidike, C. Elkin, and V. Devabhaktuni, "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications," IEEE Access, vol. 9, pp. 82031–82057, 2021.

[24] G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, "Medical Image Segmentation Based on U-Net: A Review," J. Imaging Sci. Technol., vol. 64, no. 2, pp. 20508-1–20508-12, 2020.

[25] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of Breast Ultrasound Images," Data Brief, vol. 28, p. 104863, 2020.

[26] C. Shorten and T. M. Khoshgoftaar, "A Survey on Image Data Augmentation for Deep Learning," J. Big Data, vol. 6, no. 1, p. 60, 2019.